Measurements of a physical or biological system result in a time series, 
\( s(t) = s(t_0 + n \tau_t) = s(n) \), sampled at intervals of \( \tau_t \) and initiated at \( t_0 \). When a signal can be represented as a superposition of sine waves with different amplitudes, its characteristics can be adequately described by Fourier coefficients of amplitude and phase. In these circumstances, linear and Fourier-based methods for extracting information from the signal are appropriate and powerful. However, the signal may be generated by a nonlinear system. The waveform can be irregular and continuous and broadband in the frequency domain. The signal is noise-like, but is deterministic and may be chaotic. More information than the Fourier coefficients is required to describe the signal \([1, 2]\).

This article describes methods for distinguishing chaotic signals from noise, and how to utilize the properties of a chaotic signal for classification, prediction, and control.

A measured signal that has an irregular time series and a continuous, broadband power spectrum is distinguishable from noise and can occur within the dynamics of a few degree-of-freedom dynamical system. To describe such a signal with the infinite number of Fourier coefficients is not appropriate. One goal of this article is to describe a class of practical techniques for dealing with such signals. These techniques enable one to easily:

- Determine directly from measured data how many degrees of freedom are operating to produce the signal,
- Determine the predictability of the underlying dynamics from these observations,
- Make simple, useful predictive models,
- Identify parameters in a system of (nonlinear) equations devised to describe the original source of the signals, and
- Devise control strategies for nonlinear processes.

Although these methods were developed primarily from studies of chaotic systems, they also are useful for characterizing and exploiting all types of signals.

Data from a nonlinear circuit \([3]\) (Fig. 1) will be used to illustrate the general points of the article. The differential equations for this circuit are:

\[
\begin{align*}
\frac{dx(t)}{dt} &= y(t) \\
\frac{dy(t)}{dt} &= -x(t) - \delta y(t) + \varepsilon(t) \\
\frac{dz(t)}{dt} &= \gamma [\alpha f(x(t)) - \varepsilon(t)] - \delta y(t)
\end{align*}
\]

(1)

where \(x(t)\) and \(z(t)\) are the voltages across the capacitors, \(C\) and \(C'; \alpha\) is the gain of the nonlinear converter at \(x = 0\); and \(y(t) = J(t)(L/C)^{1/2}\). Time \(t\) has been normalized with respect to \((LC)^{1/2}\). The nonlinear response of the converter, \(N\), is approximated well by the piecewise nonlinear function...
\[
\begin{cases}
0.528 & \text{if } x \leq -12 \\
0.528 & \text{if } x \geq 12 \\
x(1-x^2) & \text{if } -12 < x < 12
\end{cases}
\]

The other parameters of the model are dependent on the linear feedback loop through
\[
\gamma = \sqrt{\frac{L}{RC}}, \quad \delta = r \left( \frac{C}{L} \right)^{\frac{1}{3}}, \quad \sigma = \frac{C}{C'}.
\]

For circuit elements with \( R = 3.98 \Omega, \) \( r = 36.1 \Omega, \) \( C = 23.01 \mu F, \) \( C' = 334 \mu F, \) \( L = 152.6 \) mH, and \( \alpha = 24.24, \) the output is the time series of Fig. 2 where the sampling rate is 50 kHz (\( \tau_s = 0.02 \) ms).

This is an example of a signal arising from the deterministic dynamics of a three-degree-of-freedom system. The power spectrum (Fig. 3) is broad and continuous and the autocorrelation function (Fig. 4) rapidly goes to zero. Using methods that will be described in detail in the next section, the dynamical structure of this signal can be captured using a three-dimensional state space created from the measured voltage \( x(n) = x(t_n + n\tau_s) \) and its time lags, \( T. \) Using a time lag of 15 samples (\( T = 15\tau_s = 0.3 \) ms) three-dimensional vectors are created:
\[
y(n) = [x(n), x(n + T), x(n + 2T)],
\]

This article describes how to transform the measurements, \( x(n), \) to the geometric object associated with the source of the signal and how to extract dynamical information from that object. Knowledge of the dynamical system or the governing differential equations is not required to analyze the signal. A more detailed explanation of the underlying theory and methods is in reference [4]. The software used for this analysis is called “csp” (cspW for Windows, cspX for UNIX) [5].

**The State Space and Attractor Reconstruction**

The state of a dynamical system at any time can be specified by a state-space vector where the coordinates of the vector are the independent degrees of freedom of the system. For the circuit of Fig. 1, the state of the system is given by \( y(n) = [x(n), y(n), z(n)] \) at any time, \( n = t_n + n\tau_s. \) Generally, the number of first-order differential equations describing the system determines the number of independent components in \( y(n). \) The “embedding theorem” [1, 2, 6] establishes that, when there is only a single measured quantity from a dynamical system, it is possible to reconstruct a state space that is equivalent to the original (but unknown) state space composed of all the dynamical variables.

The embedding theorem [7-9] states that if the system produces orbits in the original state space that lie on a geometric object of dimension \( d_A \) (which need not be integer), then the object can be unambiguously seen without any spurious intersections of the orbit in another space of integer dimension \( d > 2d_A, \) or larger, comprised of coordinates that are (almost) arbitrary nonlinear transformations of original state-space coordinates. The absence of intersections in the second space means the orbit is resolved without ambiguity when \( d \) is large enough. Overlaps of the orbit may occur in lower dimensions and the ambiguity at the intersections destroys the possibility of predicting the evolution of the system.

In a dissipative system, the geometric object to which orbits go in time is called the system attractor. If \( d_A \) is not integer, it is called a strange attractor (after Ruelle [10]), and the system is chaotic. It is the attractor and the motion of system orbits on it, not the projection of those motions onto the observation axis of the measurements, where prediction, classification, control, and other signal-processing tasks can be carried out without ambiguity. That is, it is necessary to go to a higher-dimensional space to do signal processing when the signal comes from a nonlinear system that may exhibit irregular, chaotic motions.

Almost any set of \( d \) coordinates is equivalent by the embedding theorem. Each set is a different way of unfolding the attractor from its projection onto the observations. Formally, an autonomous system producing orbits \( x(t) \) through the dynamics is

**Glossary**

*Chaos* - irregular but deterministic motion. Characterized by a continuous, broadband Fourier spectrum. Possible only in a three- or more-dimensional nonlinear system of differential equations or a two- or more-dimensional nonlinear discrete time map.

*Dissipative System* - A system with sources and sinks of energy.

*Embedding Dimension* - The number of independent geometric coordinates in state space needed to capture the behavior of an attractor. Always an integer.

*Attractor* - The set of points in state space visited by a signal trajectory after transients are gone.

*Strange Attractor* - An attractor set of noninteger dimension. Associated with nonperiodic motions.

*Reconstructed State Space* - The space, constructed solely from measurements, that serves as a proxy for the full multidimensional dynamical system state space. It is constructed using observations and their time delays in this article.

*Trajectory* - The path that a signal vector follows through state space.

*Lyapunov Exponent* - A statistical quantity that quantifies uncertainty about the future state of the system. It is the average (over all points on the attractor) rate at which nearby trajectories separate while moving on an attractor.
\[
\frac{dx(t)}{dt} = F(x(t)),
\]
and the output is \(s(t) = h(x(t))\). \(x\) is an \(n\)-dimensional vector, and \(s(t)\) is typically a one-dimensional output signal. With mild restrictions on the functions \(F(x)\) and \(h(x)\), any independent set of quantities related to \(s(t)\) can serve as the coordinates for a state space for the system. Time derivatives of \(s(t)\) are the natural set of independent coordinates. But, when the signal is sampled in discrete time, the derivatives are a high-pass filter and therefore emphasize errors and noise in the measurements. The first derivative of \(s(t)\) is approximated by

\[
s(n) = \frac{s(t_n + (n + 1)τ) - s(t_n + nτ)}{τ},
\]

which is also a high-pass filter.

Equation 6 suggests another natural set of coordinates for the state space. The signal \(s(n)\) and its time delays are the ingredients in the approximations to the time derivatives of \(s(n)\). The time-delay values of \(s(n)\) are new information that enters the approximation of each derivative. Using the observed signal and its time delays avoids the emphasis on errors and noise associated with high-pass filter approximations to the time derivatives and requires no computation on the observations themselves. This set of coordinates is realized by forming the vectors

\[
y(n) = [s(n), s(n-T), s(n-2T), ..., s(n-d_T-1)(T)],
\]

where a \(d_T\)-dimensional state space is constructed, and each component of the vector \(y(n)\) is separated in time by \(T\) (\(T\) is an integer, as is \(d_T\)). An example of this reconstruction for a simple periodic signal is shown in Fig. 6. In this case, the attractor is a closed loop that can be embedded in a two-dimensional space (although three are used for the display—this signal is over embedded). To do this embedding, a time delay, \(T\), and the number of components, \(d_T\), in the vector \(y(n)\) are needed.

**Choosing a Time Delay**

When the signal is represented by Eq. (7), assurance is needed that each component of the vector is providing new information about the signal source at a given time. The dynamical difference between the components is achieved by the evolution of the signal source for a time \(T\) (\(T\) is an integer, as is \(d_T\)). An example of this reconstruction for a simple periodic signal is shown in Fig. 6. In this case, the attractor is a closed loop that can be embedded in a two-dimensional space (although three are used for the display—this signal is over embedded). To do this embedding, a time delay, \(T\), and the number of components, \(d_T\), in the vector \(y(n)\) are needed.
\[
\log_2 \left[ \frac{P_{AB}(a_j, b_k)}{P_A(a_j)P_B(b_k)} \right]
\]

where \( P_A(a) \) is the probability of measurement \( a \), \( P_B(b) \) is the probability of measurement \( b \), and \( P_{AB}(a, b) \) is the joint probability of measurements \( a \) and \( b \). For these computations, the measurements \( a_j \) are the \( s(n) \), and the \( b_k \) are the \( s(n + T) \). These are deterministic signals, so the probability is the distribution of values of the variables as a function of time. \( P_A(a) \) is the normalized histogram of the \( a_j \) observations and \( P_B(b) \) is the normalized histogram for the \( b_k \). The joint histogram is \( P_{AB}(a, b) \).

Thus, the suitable nonlinear correlation function is the mutual information of the measurements, averaged over all measurements

\[
I(T) = \sum_{a_j, b_k} P_{AB}(a_j, b_k) \log_2 \left[ \frac{P_{AB}(a_j, b_k)}{P_A(a_j)P_B(b_k)} \right]
\]

\( I(T) \geq 0 \) for all \( T \), so there is not a zero value analogous to the familiar linear correlation function, which is sometimes advocated for selecting the time delay. An alternative prescription for nonlinear correlation is to use the value of \( T \) at the first minimum of \( I(T) \) to create the vectors \( y(n) \) [12, 13]. For the circuit data, the average mutual information evaluated using 215 samples is shown in Fig. 7 where the first minimum is at \( T = 15 \).

The choice of the first minimum of \( I(T) \) is a prescription, not a rigorous fixed choice. In general, the prescription provided by the average mutual information often appears to yield a shorter time delay than the first zero crossing of the autocorrelation function. The danger in using the longer time delay is that the components of the vector can become independent of each other and subsequent calculations may not be valid (this can happen with the Lyapunov exponents, for example). Thus, an even safer procedure is to evaluate the important dynamical quantities, such as the Lyapunov exponents, over a range of time delays.

**Choosing a Global Unfolding Dimension**

The embedding theorem [7, 8] provides a sufficient dimension in which the orbits of the system are no longer crossing each other and the temporal sequence of points is without ambiguity. The data may not require this large a dimension, and by following the general sense of the embedding theorem there is another way to choose the number of coordinates [14]. For any point on the attractor, \( y(n) \), one can ask whether its nearest neighbor in a state space of dimension \( d \) is there for dynamical reasons or is instead projected into the neighborhood because the dimension is too small. That neighbor is then examined in dimension \( d + 1 \) by simply adding another coordinate to \( y(n) \) using the time-delay construction. If the nearest neighbor is a true neighbor, meaning it arrived there dynamically, it will remain a neighbor in this larger space. If that nearest neighbor is there because dimension \( d \) is too small, it will move away from \( y(n) \) as dimensions are added. When the number of these false nearest neighbors becomes zero, the attractor has been unambiguously unfolded because crossings of the orbit have been eliminated. The dimension in which false neighbors disappear is the dimension necessary for viewing the data. This minimum embedding dimension, \( d_f \), is less than or equal to the sufficient dimension found from the embedding theorem.

Desirable neighbors lie close by virtue of the evolving dynamics of the system. In practice, however, there are four types of neighbors, only one of which are true neighbors:

- **True neighbors** that result from the dynamics of the system. These neighbors are acceptable and can be used for analysis, prediction, and control.
- **False nearest neighbors** result from a projection of the attractor into an insufficient dimension, as discussed above.
Oversampling neighbors arise from using too high a sampling rate. This causes the next point on the orbit to always be identified as the nearest neighbor and misses the recurrent visits to the neighborhood of the point in question.

Quantized neighbors that arise from rounding error, for example, from an A/D converter.

False nearest neighbors are the result of using too low an embedding dimension. An ambiguity results where the trajectory crosses itself. At such a crossing, one cannot tell which path the trajectory should follow, in the absence of other information. The consequences for prediction algorithms and analysis methods that rely on the behavior of close neighbors can be disastrous. Fortunately, false neighbors are easy to detect and are eliminated by selecting an appropriate embedding dimension.

Oversampled neighbors result from a high sampling rate or may arise when a low-pass filter has been applied to data. Subsequent samples may be so close in time that they are closer than neighbors that are more distant in time, but are also close in state space (the true neighbors). Oversampled neighbors are easily eliminated. When searching for neighbors, one simply requires that they have some reasonable temporal separation. Because the sample number is a proxy for time, it is easy to require that candidate neighbors be separated by a reasonable number of samples, usually the time delay.

The final type of neighbor poses a severe problem that is not easily resolved. Because A/D converters use a finite number of bits, a rounding error occurs. This may happen when a converter with a small number of bits is used, or if the range of signal activity is small relative to the range expected by the converter. Suppose there are two floating-point numbers, known to six significant digits, such as 128.123456 and 128.123123. If, in the collection or conversion process, these are rounded to 128.123 they become the same number. When this happens, then there is a probability that low-dimensional vectors formed from the numbers will also have this ambiguity. Vectors that are actually distinct in state space all collapse to a single vector. There is no ideal way to fix this, although the effects can be mitigated.

It is easy to sort the points in dimension $d$ by how close they are to $y(n)$ [15]. We use Euclidean distance, but any distance measure will work. Once sorted, it is straightforward to examine the nearest neighbor of every point, and thus identify any false neighbors in the selected dimension. As the dimension is increased from $d = 1$ to $d = 2, \ldots$, the percentage of false nearest neighbors should go to zero at the dimension where the attractor is globally unfolded [14]. This is the minimum global embedding dimension, $d_g$. A residual of false nearest neighbors is caused by contamination of the data by a high-dimensional signal that is conveniently called “noise.” Noisy data have nonzero false neighbors because the attractor is blurred. As $d$ increases, the noise component slowly begins to predominate. Thus, there is a strong motivation not to over embed noisy signals.

Figure 8 is the percentage of false nearest neighbors out of 65,536 samples from the nonlinear circuit. The data vectors for the nonlinear circuit are constructed in dimension $d = 1, 2, \ldots, 10$ using the time delay of $T = 15$ chosen from the prescription by average mutual information. It is clear that the false nearest neighbors disappear at $d_g = 3$. Once there are no false nearest neighbors left, further unfolding of the attractor has no effect. These data are relatively noise free, so no rise in false nearest neighbors is seen as $d$ increases beyond $d_g$. However, the correlated noise and the EKG data show the residual that is characteristic of noise.

Local Properties of the Dynamics

The dynamics associated with the circuit develop locally on the attractor through some discrete-time map or through differential equations. The number of degrees-of-freedom, $d_f$, governing the dynamics is less than, or equal, to the global embedding dimension $d_g$. $d_f$ can be greater than $d_g$, because the coordinate system is created solely from measured data. This coordinate system may twist and fold the attractor relative to its geometry in the original, but unknown, coordinates of the dynamical system. However, if the dynamical system has only $d_f$ degrees-of-freedom locally, that information can be used to classify the dynamics and construct a framework for predictive models and control strategies. For example, the attractor corresponding to periodic motion is a closed loop in the state space (local dimension $d_g = 1$) although generically it may only be unfolded in three-dimensional space ($d_g = 3$).

![Reconstructed attractors for a variety of systems.](image)
Local Dynamical Dimension

The embedding theorem is concerned only with global properties of the dynamics. Our method for identifying the local dynamics goes beyond those safe confines and addresses the important issue of predictability. While computing distances between points on the attractor in dimension \( d_L \), the quality of predictive models is examined in dimension \( d_L = 1, 2, \ldots \), until the number of coordinates required to evolve the observations locally in reconstructions state space is established.

The model-making is straightforward. For a neighborhood of state-space vectors around a point, \( y(n) \), on the attractor, its \( N_g \) neighbors, \( y^{(r)}(\vec{n}) \), \( r = 1, 2, \ldots, N_g \), are found \((\vec{n})\) is used to note that the index for the neighbors is not the same as the index \( n \) for \( y(n) \). A local model that takes these neighbors of \( y(n) \) to their respective points, \( y^{(r)}(\vec{n} + 1) \), at the next sample is,

\[
y(r; \vec{n} + 1) = \sum_{m} \Phi_m (y^{(r)}(\vec{n})) c(\vec{n}, m),
\]

where \( \Phi_m (x) \) are a selected set of basis functions used to interpolate among the data points \( y^{(r)}(\vec{n}) \), and the \( c(\vec{n}, m) \) are local coefficients determining the local map from time “\( n \)” to time “\( n + 1 \)”. Note that \( y(r; \vec{n} + 1) \) is not necessarily the \( r \)-th nearest neighbor of \( y(n + 1) \).

The coefficients \( c(\vec{n}, m) \) are determined by using a least-squares criterion, e.g., by minimizing

\[
\sum_{r=1}^{N_g} \left| y(r; \vec{n} + 1) - \sum_{m} \Phi(y^{(r)}(\vec{n})) c(\vec{n}, m) \right|^2.
\]

The predictions are made using the known basis functions and the local coefficients. In the present case the issue is how these predictions behave as a function of both the local dimension \( d_L \) of the vectors in the model and the number of neighbors \( N_g \) used in determining the coefficients. The neighbors are chosen in \( d_g \), but the predictions depend on the local dimension. When a measure of the quality of the predictions becomes independent of \( d_L \) and of \( N_g \), the local dimension of the dynamics has been determined. The quality of the predictions is quantified by determining how many of the trajectories associated with the \( N_g \) neighbors remain within some fraction of the attractor size for a number of samples forward in time (the prediction limit), which is usually set to \( T \) or \( 1/2T \) [16]. A “bad” prediction is one where a trajectory diverges from the original neighborhood by more than this fraction before it reaches the prediction limit.

In Fig. 9 the percentage of “bad” predictions made using data from the nonlinear circuit are plotted as a function of local dimension (on the horizontal axis) and as a function of \( N_g \) for \( N_g = 20, 40, 60, \) and 80. Predictions were made in 5,000 neighborhoods using local polynomials as a basis set, and the separation fraction was 0.13 of the RMS size of the attractor. The results for these data indicate that \( d_L = 3 \), which is consistent with the known origin of the data.

It is important to determine the local dynamics of the signal source to be able to predict and model the future behavior of a dynamical system. Because \( d_L \leq d_g \), the determination of an embedding space in which the attractor can be unfolded does not provide the information needed for prediction and modeling.

Finally, we note that an enormous amount of effort has gone into determining the fractal dimension, \( d_g \), of a complex signal, but that (generally noninteger) dimension can always be determined in the integer dimension just greater than \( d_g \). This need not be as large as \( d_g \), in which one can compute distances between points on an attractor, nor need it be equal to \( d_L \), for which one can build a model.

Lyapunov Exponents

Nonlinear systems that have irregular temporal behavior may be chaotic, which means that the system is unstable on its attractor and within the basin of attraction for that attractor. The orbits of such systems are said to be “sensitive to initial conditions.” From a signal-processing standpoint, “sensitive to initial conditions” is the same as stating that nearby trajectories on the attractor will diverge at an exponential rate. Since, in reality, the initial conditions can only be specified with some finite precision, two trajectories that start from almost identical initial conditions will move apart at an exponential rate. This rate, and hence the predictability of the system, is described by the largest of the Lyapunov exponents of the system. Predictability is quantified by examining the behavior of a small initial perturbation, \( w(0) \), made to an orbit, \( y(n) \), of the system. Assuming a dynamical rule, \( y(n + 1) = F(y(n)) \), describes the system, then

\[6. The time series of a periodic signal and its reconstructed attractor in three-dimensional embedding space. Note how this signal actually only occupies two of the three dimensions and, because it is not chaotic, the trajectory closes on itself.]
\[ y(n + 1) + w(n + 1) = F(y(n) + w(n)) = F(y(n)) + DF(y(n)) \cdot w(n) + \text{Order}(w(n)^2) \] (12)

or in the limit as \( w(n) \rightarrow 0 \)

\[ w(n + 1) = DF(y(n)) \cdot w(n) = DF(y(n)) \cdot DF(y(n - 1)) \cdots DF(y(0))w(0) = DF^N(y(0)) \cdot w(0). \] (13)

The Jacobian matrix of the dynamics is

\[ DF(x)_{ab} = \frac{\partial F_a(x)}{\partial x_b}; \quad a, b = 1, 2, \ldots, d. \] (14)

where \( DF^N(x) \) is the composition of \( N \) Jacobians calculated at iterations of a starting point, \( x \).

The stability of the system for the small perturbation \( w(0) \) is determined by the eigenvalues of the matrix \( DF^N(y(n)) \) after \( L \) samples along the orbit after the perturbation. Oseledec [17] analyzed this eigenvalue problem, which is a generalization of the familiar linear stability problem for perturbations about a fixed point or equilibrium orbit. The matrix

\[ OSL(L, x) = \left[ DF^L(x) \cdot DF^L(x)^T \right]^\frac{1}{2} \] (15)

is well defined because the product of \( DF^2 \) and its transpose is orthogonal. As \( L \) becomes large (in a formal sense, as \( L \rightarrow \infty \)) the eigenvalues of this matrix have been shown to exist for almost all \( x \) along an orbit and are independent of \( x \). The eigenvalues are also unchanged under a smooth nonlinear transformation of the coordinate system, which means that the dynamical system can be classified with these eigenvalues. Regardless of the coordinate system, the computed values will be the same. The eigenvalues are \( e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_d} \) in a \( d \)-dimensional system where by convention \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \) are the global Lyapunov exponents [18].

If any of the \( \lambda_a \) are positive, then the orbit is unstable. Dissipation in the dynamics requires \( \sum_{a=1}^{d} \lambda_a < 0 \), so that volumes in state space shrink as time progresses. This leads to the manifestation of the instability on the attractor, although it remains a compact geometric object in the state space. If any of the \( \lambda_a \) are greater than zero, the system is chaotic by definition. If one of the \( \lambda_a \) is zero, then the dynamical system producing the signal can be described by a set of ordinary differential equations. It is possible for a discrete time dynamical system to have a zero Lyapunov exponent, but the parameter values for the system would be exceptional.

The \( \lambda_a \) are evaluated by numerically determining the \( DF(x) \) locally in state space from the local predictive maps described above. The linear term in these maps gives \( DF(x) \). The composition of the Jacobians can be diagonalized by a recursive decomposition method, even though the product is ill conditioned. The values of the Lyapunov exponents, \( L \), samples forward (or backward) in time after a perturbation at \( x \) can be determined from \( OSL(L, x) \). These are the local Lyapunov exponents, \( \lambda_a(L, x) \) [19-21], and their average over many initial conditions, \( x_a \), are denoted \( \lambda_a(L) \).

The \( \lambda_a(L) \) converge to the global exponents, \( \lambda_a \), as a power of \( L \). It is useful to display the \( \lambda_a(L) \) as a function of \( L \) to identify the zero exponent, if it exists, as well as characterize the general trend of all exponents. Fig. 10 shows the average local Lyapunov exponents for the example circuit that were evaluated from knowledge of the scalar measurement, \( x(t) \), alone. The near zero global exponent indicates that a set of three ordinary differential equations is required for the description of the data.

This method directly evaluates the \( \lambda_1 \) exponents. If a dimension of \( \lambda_1 \) were used for the computations, there would be \( \lambda_1 < 0 \) false Lyapunov exponents with little sense of which were the true exponents and which were false. In principle, the extraneous exponents should be negative and represent the orbit moving from some point in state space onto the attractor. In practice, this rarely happens. One can either select \( \lambda_1 \) using the false nearest neighbor algorithm or, if there is enough noise-free data, evaluate the \( \lambda_1(L) \) forward and backward, which produces \( \lambda_1 \) exponents that change sign and \( \lambda_1 \) exponents that do not change sign. The correct \( \lambda_1 \) exponents are the ones that change sign [22, 23].

If the sum of the exponents is greater than zero, the results are meaningless. This can occur when the signal migrates through state space instead of remaining in a bounded region (a situation typical of much financial data), if the signal is nonstationary, or if the trajectories are not coherent for some reasonable period of time. The latter may result when the underlying system is stochastic. In this case, close trajectories may even go in opposite directions.

The sum of the Lyapunov exponents may also be greater than zero if too large a time delay is selected, in which case the observations that make up the vector \( y(n) \) have become independent of \( x \). Thus, an examination of the sum of the exponents is necessary to ensure the interpretation of the largest exponent is valid. For example, one prescription for selecting the time delay is to use the first zero-crossing of the autocorrelation function. For the example nonlinear circuit, Fig. 4 leads to a choice of \( T = 100 \). But, if this time delay is used to compute the Lyapunov exponents, the sum of the exponents is positive and the largest exponent, despite the fact it is invalid, is almost 20 times larger than the value that results from using \( T = 15 \). The implication for prediction is that the if the larger time delay were accepted, the horizon to which one could predict would be greatly reduced (or nonexistent). The prescription provided by Fraser [12, 13] does not always generate the smallest value of \( \lambda_1 \) (which provides the greatest predictability), but in our experience, it never overestimates the time delay.

The Lyapunov exponents are perhaps the most important quantifying measure of chaotic motions. Of all the
classifying quantities, they are the only ones that are truly sensitive to changes in dynamics, which often occur without a change in the dimension. The Lyapunov exponents also provide a way to quantify the efficacy of prediction and control.

Using the full spectrum of Lyapunov exponents, a fractal dimension of the attractor, called the Lyapunov dimension, can be defined [24]. This is the dimension of a ball in state space that neither grows nor shrinks as the dynamical system evolves. A line segment in the space grows as $e^{\lambda_1 t}$, while a full N-volume shrinks as $e^{(\lambda_1 + \lambda_2 + \cdots + \lambda_N)t}$. If $K$ is the largest integer for which $\sum_{i=1}^{K} \lambda_i > 0$, then the Lyapunov dimension is

$$D_L = K + \frac{\sum_{i=1}^{K} \lambda_i}{|\lambda_{i+1}|}. \quad (16)$$

In some cases, this is the same as the so-called information dimension.

There are many fractal dimensions and an enormous effort has been expended in producing strategies to establish them [25, 26]. Fractal dimensions have not proved useful for signal processing because fractal dimension estimates depend on accurate measurements of the distance between near neighbors at the smallest scale of the attractor. But, the attractor is blurred at these scales by noise, contamination, and A/D quantizing errors [27]. Further, fractal dimension estimates contribute no information needed for signal analysis. The minimum state space needed to unfold the system dynamics, the critical parameter for processing, can be determined from other more robust algorithms. The existence of chaos (or noise on a linear signal) is established by the Lyapunov exponent.

**Correlated Noise**

In general, correlated noise has a broadband power spectrum, colored or white, that is similar to the power spectrum of a chaotic system. Because correlated noise may arise in real physical systems, it is important to make a distinction between a signal generated by a deterministic system and noise. This is not an easy problem.

If signals from a known stochastically driven source are examined, water surface waves generated by a randomly driven wave maker (Fig. 5), for example, then one discovers that there can be a well defined and reasonable minimum of the average mutual information. The global embedding dimension is also well defined and comparable to what intuition suggests is reasonable for surface waves. For the randomly generated waves, $d_L$ is seven (Fig. 8), which is what has also been observed in true ocean waves [28]. And, $d_L$ for both the generated and open-ocean waves is similar.

To discriminate between a signal generated by a deterministic chaotic system and colored noise, the method of surrogate data has been proposed [29]. The chaotic characteristics should differ for a genuine chaotic signal and a signal that has a similar power spectrum but is known to be correlated noise a priori. The method generates the surrogate data by randomizing the phases of the Fourier coefficients. However, even if the dynamical characteristics such as the embedding dimension and the Lyapunov exponents are different for the surrogate and original data, it is not conclusive evidence for low-dimensional chaos. In the latter, colored noise can also have partially correlated phases that will make it look different from totally uncorrelated colored noise.

From an applied-signal-processing standpoint, a practical definition of the difference between a chaotic signal and correlated noise is the horizon for which predictions can be made with some confidence. This is the issue addressed by the largest Lyapunov exponent. If the largest Lyapunov exponent is very large, it indicates a rapid loss of predictability. If a metric is the time (number of samples) it takes for neighboring trajectories to separate to say, 1/2 of the mean attractor size, and that number is on the order of the time it takes for a single trajectory to move across a comparable distance, then there is a sense that predictions will be futile and the signal can be treated as noise.

**Other Examples**

We have examined many other signals including electrocardiograms (Fig. 5(b)), machine tools (Fig. 5(d)), lasers, ambient noise in the ocean [27], ocean waves [28], water-level fluctuations in a large lake [30], hydrodynamic turbulence [31], and ship hull model motions (Fig. 5(g)). Other applications that we are examining include the following.

**Signal Detection**

We are investigating how the methods described here, as well as other algorithms, can be used to improve signal detection. A significant loss of detector performance is...
typically due to the mismatch between the noise statistics assumed in a matched filter and the true statistics of environmental noise and clutter. Depending on the context, the loss may be on the order of 10 dB or more. A better understanding of the true nature of noise can be used to improve signal-detector performance by mitigating the mismatch between the assumed and the true noise. Prototypes of methods to do this have been constructed and have demonstrated significant gains over existing processors.

Electrocardiograms (EKGs)
The electrophysiology of the heart is complex because it involves interactions among multiple physiological variables including autonomic and central nervous regulation, hemodynamic and humoral effects, and other extrinsic variables. Yet, the instrumentation for measurements are often limited to electrocardiograms from skin-mounted sensors. This is precisely the type of situation where the tools described in this article are most powerful. The system is a multivariate nonlinear system, yet it is possible to collect only one type of signal.

An additional difficulty arises from the nature of the collection. The period over which the system is stationary is short (literally, how long will a human subject remain motionless while data are collected?), the sensors are not calibrated, and there is a great deal of variability among subjects. A further complication is the dearth of data, especially from healthy subjects. People without health problems rarely undergo extensive cardiac testing where high-quality data can be collected.

Nonetheless, there are some consistent results. The strange attractor of an EKG is shown in Fig. 5(b). It is typical of resting-state EKGs and, along with other cases, has a global embedding dimension of five and a local dynamical dimension of four. In fact, for the cases we have examined, the \( d_g \) is almost always five and the \( d_L \) is almost always four. Our experiments to date include subjects who are undergoing a stress test for a specific pathology and these dimension measures vary little during the test. For almost all cases, including one extreme pathology, dimension measures remain relatively constant. The one nonlinear characterization of the data that does change, as a preliminary finding, is the spectrum of Lyapunov exponents.

The extreme pathology is ventricular fibrillation. The dimensions still appear to be \( d_g = 5 \) and \( d_L = 4 \), but the largest Lyapunov exponent is different from the other cases. In fact, given considerations of the short periods over which such data can be collected and instrumentation calibration and noise, it appears that a distinguishing hallmark of this life-threatening condition is that the largest Lyapunov exponent can be considered to be near zero—the system is not chaotic. Thus, while the statistical base is still small, there is evidence that knowledge about the diversity of chaos is beneficial and that it has practical applications in classifying cardiac behavior.

Radar Backscatter
In a companion article in this issue, Simon Haykin reviews some of his research on radar backscattering from the ocean surface. The IPX X-band radar was land mounted 25 to 30 meters above sea level, which is a typical height for a ship-mounted radar. The antenna was fixed, so that the system operated in a dwell mode and illuminated a fixed patch of the ocean surface. The tempo-
A water-depth predictions in shallow navigable channels is important because of the need for accurate short-term search. The largest Lyapunov exponent is about 0.14 (Fig. 11).

The grazing-angle radar is from a low-dimensional system used to extract that signal from its background. This reflects chaotic behaviors and the configuration of neural networks where the embedding dimension is five (Fig. 8) and the local dynamical dimension of four shows that, in general, stations exposed to the open ocean or five recorded was due entirely to the motion of the ocean surface. Haykin's article addresses what appears to be the natural connection between the dimensionality of a chaotic signal and the configuration of neural networks used to extract that signal from its background. Haykin's research [32] has shown that backscatter from a low-grazing-angle radar is from a low-dimensional system where the embedding dimension is five (Fig. 8) and the largest Lyapunov exponent is about 0.14 (Fig. 11).

### Ocean Water Levels

Classification and prediction of ocean water levels is important because of the need for accurate short-term water-depth predictions in shallow navigable channels. Optimum siting of tide stations is also a long-standing problem. It is desirable to minimize the number of stations in a region, but a complete description and record of the tides along all sections of the coast is imperative. An analysis of data from seven east coast U.S. tides stations shows that, in general, stations exposed to the open ocean record observations that have an observed embedding dimension of five and a local dynamical dimension of four or five [33]. However, there are large variations in the Lyapunov exponents and, for example, the largest Lyapunov exponent for Charleston, South Carolina, whose attractor is shown in Fig. 5, is about 0.62 while at Ft. Pulaski, Georgia, it is about 0.78. This demonstrates how behaviors can vary within a fixed dimensionality. Interestingly, for stations inside a complex estuary (Chesapeake Bay), the embedding dimension increased to six and the Lyapunov exponents are much larger (3.91 at Baltimore, Maryland). This follows intuition in that more variables ought to be active inside the estuary and that behavior ought to be more complex. Fort Pulaski and Charleston will be examined in the next section as an example of how one sensor's measurements can be used to predict the behavior of a signal from another.

### Prediction and Control

Characterizing nonlinear systems by the state-space analysis of observed data provides a framework for performing important tasks. These include making predictions based on an observed set of data from the past [34], predicting one "hard" dynamical variable from the observation of an "easy" variable using past simultaneous measurement of both variables [35], and controlling dynamical systems to some more regular, or possibly to an even more irregular orbit [36]. We briefly discuss each of these and mention other applications of the framework established by the nonlinear analysis described above. These methods have also been used to devise secure communications systems [37], but these applications will not be covered in this article.

### Prediction

Using a single, scalar measurement from a nonlinear dynamical system, one can reconstruct its attractor from the methods described earlier. If the source of the signal is stationary, then the attractor remains the same whenever the observations begin, even though the actual orbit is "sensitive to initial conditions" and is uncorrelated with any past measurements. One can use the statistically stationary properties of the attractor to provide an accurate "black box" means for predicting future evolution of an orbit when the past has been well sampled. It is a "black box" procedure because no knowledge of the underlying physics of the source is required. While it provides accurate predictive power, there is a limit on the depth to which one can take this approach because of the inherent loss of predictability in a chaotic system.

The method for nonlinear estimation is similar to linear autoregressive (AR) and Kalman filter techniques except that estimation is done in the spatial and temporal domain of the attractor, rather than the frequency or traditional time domain of the signal. As with AR and Kalman filters, an explicit model of the system is not required. If every neighborhood of the attractor has been well visited, then the evolution of points from one state-space neighborhood to the next neighborhood can be mapped. Knowledge of a local numerical map that moves the trajectories from neighborhood to neighborhood provides a way to place any newly observed point, y(n), into the neighborhood nearest to it and then project the trajectory.
forward from that neighborhood to predict $y(n+1)$. The process is then repeated to move $y(n+1)$ to $y(n+2)$, and so on. The power of the method is limited by the growth of errors, as dictated by the largest Lyapunov exponent $\lambda_1$, which in time $t = 1/\lambda_1$ destroys the possibility of further prediction. This is actually a limit on any method of nonupdated prediction, including direct use of the equations of motion.

To build local maps that describe the evolution of neighborhoods, we start with an observed point, $y(n)$, on the attractor. In a neighborhood with $N$ neighbors $y'_{(i)}(\vec{n})$ with $\vec{n} = 1, 2, \ldots, N$, the (unknown) vector field, $F(y)$, that evolves points on the attractor $y(n+1) = F(y(n))$ can be expanded as

$$F(y(n)) = \sum_{m=1}^{M} c(m) \Phi_m(y(n)) \quad (17)$$

in terms of $M$ basis functions, $\Phi_m(y)$. Radial basis functions, sigmoidal functions associated with neural networks, and polynomials are widely used choices for the $\Phi_m(y)$ [34]. The mathematical problem being addressed is that of interpolation within neighborhoods of a multidimensional space populated by observed data points. There are arguments for choosing any of these alternatives, and there are also certainly other choices with excellent justification.

An expansion that involves both the choice of basis function $\Phi_m(y)$ and $M$ is

$$y(r;\vec{n}+1) = \sum_{m=1}^{M} c_k(m) \Phi_m(y'(r)(\vec{n})), \quad (18)$$

where $y(r;\vec{n}+1)$ is the point to which the neighbor $y'(r)(\vec{n})$ goes in one sample period ($T_s$). The coefficients, $c_k(m)$, are chosen by a least-squares method that minimizes the error in this model of the dynamics. So minimizing

$$\sum_{r=1}^{N} \left| y(r;\vec{n}+1) - \sum_{m=1}^{M} c_k(m) \Phi_m(y'(r)(\vec{n})) \right|^2, \quad (19)$$

gives the coefficients $c_k(m), m = 1, 2, \ldots, M$ associated with the local map $F(x)$, at time $n$. This is a standard linear problem. One of the advantages of this technique is that the coefficients can be computed as needed, so the computation burden is low.

For each new data point, $z$, its evolution can be predicted by first locating the nearest point in the baseline data (which provides the coefficients $c_k(m)$). Call it $y(P)$. Then the point that follows $z$ is $z_1$:

$$z_1 = \sum_{m=1}^{M} c_{Q}(m) \Phi_m(z), \quad (20)$$

and the point $z_2$ that follows $z_1$ is determined by the neighborhood in which $z_1$ falls. If the nearest point to $z_1$ in the baseline set is $y(P)$, then

$$z_2 = \sum_{m=1}^{M} c_{P}(m) \Phi_m(z_1). \quad (21)$$

The number of points $z \rightarrow z_1 \rightarrow z_2 \rightarrow \ldots$ that can be predicted accurately is limited by the largest global Lyapunov exponent $\lambda_1$.

Figures 12 to 15 show predictions for the example circuit with time horizons from five to 40 samples (0.10 ms to 0.80 ms). These predictions are produced using 20,000 samples as a baseline set. In each case the samples from 30,000 to 31,600 samples are given as initial conditions and predictions are made 5, 12, 25, and 40 samples ahead. The predictions are compared with the
measured value at each sample. The prediction, observed values, and errors in prediction are plotted in each figure. As expected, the errors grow as the prediction horizon is extended from five samples to 40, and while there are regions where the errors are small, even for predicting 40 samples ahead, the overall degradation of predictability is apparent. A quantitative estimate of the growth of errors in this kind of prediction strategy is given in [38].

There is a notable similarity of these nonlinear methods and some linear methods. Global linear methods derive model coefficients from training data to use as a map that moves \( z(n) \) to \( z(n+1) \), based on some number of prior observations, such as autoregression [39]:

\[
z(k) = \sum_{i=1}^{p} b_i z_{k-i} + r_k
\]

where the coefficients \( b_i \) are computed from the autocorrelation coefficients of the training data using the Yule-Walker equations:

\[
\begin{bmatrix}
\varphi_{xx}(0) & \varphi_{xx}(1) & \ldots & \varphi_{xx}(p-1) \\
\varphi_{xx}(1) & \varphi_{xx}(0) & \ldots & \varphi_{xx}(p-2) \\
\vdots & \vdots & \ddots & \vdots \\
\varphi_{xx}(p-1) & \varphi_{xx}(p-2) & \ldots & \varphi_{xx}(0)
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_p
\end{bmatrix} = \begin{bmatrix}
\varphi_{xx}(1) \\
\varphi_{xx}(2) \\
\vdots \\
\varphi_{xx}(p)
\end{bmatrix}
\]

where \( \varphi_{xx}(i) \) are the autocorrelation coefficients. The evaluation of these terms using \( p \) prior observations or estimates is straightforward.

State-space estimation derives coefficients from previously observed data for a map that moves \( y(n) \) to \( y(n+1) \) (note the similarity between Eqs. (20) and (22)). “Disembedding” by taking the first component of the vector \( y(n+1) \) yields the estimate of \( z(n+1) \).

The restrictions on linear methods require that there must be significant Fourier coefficients to compute the \( b_i \). This is true regardless of the specific technique: AR (autoregression), ARMA (AR moving average), or Kalman filter. The assumptions behind linear estimation hold that the signals are composed of sinusoids and superposition can be applied. If there are no significant features in the Fourier domain, then linear methods cannot work. If the signal is chaotic, then the residual \( r_k \) grows to the magnitude of the observed signal within a few samples. A primary advantage of estimation in state space is that points on the orbit that are distant in time may be close in the state space, hence knowledge of longer period dynamics is incorporated into the predictions. Thus, the estimator may have a lower residual, \( r_k \), over longer prediction periods.

**Predicting One Time Series From Another (Virtual Sensor)**

One of the remarkable implications of the embedding theorem is that any measured variable can be used to produce coordinates for a state space that fully captures the dynamics of the signal source. This means that for two measured variables, \( V_A(t) = V_A(t_0 + n\tau_A) = V_A(n) \) and \( V_B(t) = V_B(t_0 + n\tau_B) = V_B(n) \), either one or a mixture of the two can be used to produce a state space for the dynamics. If variable \( A \) is used to create data vectors in dimension \( d_E \), then

\[
y_A(n) = [V_A(n), V_A(n - T_A), V_A(n - 2T_A), \ldots, V_A(n - (d_E - 1)T_A)]
\]

If \( d_E \) is large enough, then no further coordinates are required. The embedding theorem requires that any other aspect of the dynamical system, such as the measurement \( V_B(n) \), is determined by these vectors via

\[
V_B(n) = f_B(y_A(n)).
\]

The scalar function \( f_B(\bullet) \) is not known in general, but may be determined by expanding \( f(y(n)) \) in some convenient basis set

\[
f_B(y(n)) = \sum_{m=1}^{M} c_m \Phi_m(y(n)).
\]

Neighborhoods of simultaneous observations may also be used—the baseline data again—to determine the coefficients \( c_m \) locally in state space. This means that from a set of simultaneous observations of the dynamical variables “A” and “B,” “A” can be predicted from “B” or vice-versa, which can eliminate the need to make both measurements in the future [35].

There are two obvious applications for this method. One situation is the case where one of the measurements
is difficult and expensive, while the other measurement is easy and inexpensive. In such a case, simultaneous measurements of both variables would be made for some baseline period. Subsequently, the easy measurement can be used to predict the hard one. As long as the signal source is stationary, the local maps determined from the simultaneous measurements eliminate the need for the expensive observations. Alternatively, cross-predictions can be used to monitor the reliable operation of various sensors. A real-world case is power generation plants where both plant operating conditions and the sensor systems must be monitored and rapid assessment of anomalous sensor readings must be made. Traditionally, the monitoring of instrument performance is done using redundant monitors and comparison to analytical estimates of the process. In most practical cases, however, analytical knowledge of the system behavior is unavailable.

The methods of nonlinear system identification described in this article are intended for autonomous dynamical systems. With some modifications, they can also be applied to so-called input-output systems [40]. These systems do not reside on a low-dimensional attractor as they are driven by external (possibly high-dimensional) forces. Still, the internal nonlinear structure of the system can be unfolded using time-delayed reconstruction of the state vector based on scalar measurements from one or more meters.

As a demonstration of this method for plant and instrument monitoring, we analyzed multiple channel data from a pressurized water reactor power plant that was recorded for 8 hours on 28 January 1997 at a 2 second sampling interval. Two measurements, steam flow and pressurizer water level, were chosen for the cross-predictions. Using the pressurizer data, the time delay for embedding was determined to be 50 samples (100 s) based on the average mutual information, and the embedding dimension was determined to be \( d_E = 4 \). Fig. 16 shows the measured steam flow, the measured pressurizer level, and the predictions of the flow that were computed from the pressurizer level measurements. Thus, the unreliable measurement of flow rate can be monitored (or replaced) by a nonlinear predictor based on simultaneous measurements of the reliable pressurizer water level. This type of assessment is critical for complex systems where it is essential to rapidly discriminate between true systems anomalies and sensor systems failures.

A second application of the method is to identify measuring sensors that are “redundant” in the sense that the information gathered from one is accurate enough to allow the information gathered by the second to be deduced. An interesting, nontrivial example of this situation is the potential redundancy of tide stations that measure ocean and estuary water levels. The problem of water-level characterization and prediction in coastal and shallow waters is very difficult because there are many nonlinear forces and mechanisms that act in addition to the seemingly regular “tides.” Local meteorology and bathymetry often have an influence that exceeds the magnitude of the tidal constituents.

Because of the expense of operating tide stations, there is considerable interest in finding methods that can reduce the number of active tide stations, while improving the safety of marine navigation. Numerical models that solve the hydrodynamic equations of motion (the inherently nonlinear Navier-Stokes equations) have not been approved for operational use for shallow coastal waters because of the precision required for navigation, the large number of forces, the complex bathymetry, the large number of sensors that are needed, the small grid size that is required, and the computation expense. Further, predictive numerical models of water levels are dependent on other models for predictions of forces such as the local wind fields.

Ocean water level data at tide stations exposed to the open ocean, when analyzed by the methods in this article, have \( d_E = 5 \) and \( d_v = 4 \) or 5. Across surprisingly large distances, the present measurements of one station can be replaced by a reconstructed state space made from measurements of a baseline period at another station and the function of Eqs. 25 and 26 (obviously, data must also be collected at the first station during the baseline period). Fig. 17 is the predicted water levels at the tide station at Ft. Pulaski, Georgia, based on observations at the tide station located at Charleston, South Carolina.

Simultaneous measurements of the water levels at both stations were made every six minutes throughout 1994 [41] for a total of 87,600 samples. Predictions of water levels at Ft. Pulaski were made starting at 00:06 on 1 January 1995 and continued throughout 1995. Fig. 17 displays the predictions and actual measurements for Ft. Pulaski from 04:06 on 15 April 1995 through 04:06 on 25 April 1995 along with the observations at Charleston and the prediction errors. During the period shown, the range of water-level variations at Ft.
Pulaski was 2.16 meters, while at Charleston it was 1.63 meters. In predicting Ft. Pulaski from Charleston, we used a four-dimension space ($d_2$ for Charleston) and a first-order local linear polynomial predictor. The variance in the prediction error was 0.0029 m$^2$ for all of 1995, which is a substantial improvement over a good statistical correlation method that yielded a variance of 0.0050 m$^2$. In another case, water levels at Key West, Florida, were predicted from Charleston, some 940 kilometers distant, with an average error of 11%, compared to 25% for the statistical correlation [42].

**Control of Chaotic Motions**

Chaos has little appeal to engineers seeking regular, predictable environments in which to accomplish various tasks. However, the desire to push further into interesting parameter regimes to accomplish a task faster or more thoroughly (such as optimum mixing) may stress the dynamical system in a way that results in instabilities associated with chaotic motions of the system. To transform this unwanted chaotic motion into motion that is more regular and predictable, one might try to use small variations in system parameters to control the chaotic motion. Indeed, if one could, using small forces, move the chaotic system into a part of state space where the stable parts of the dynamical motions would dominate, then the exponential contractions associated with these stable directions would produce regular orbits.

There is evidence, but no theorems, that within the motions of a system moving along a strange attractor there are, in principle, an infinite number of periodic orbits, each of which is unstable in the uncontrolled, chaotic system. Each of these unstable periodic orbits is a solution to the equations of motion of the system and it should be possible, using small forces, to drive the chaotic system to a selected periodic orbit and stabilize it by a judicious choice of forces. Because any individual chaotic trajectory comes close to each of these unstable periodic orbits [3], the required forces should be small.

The first proposal to address this problem came from Ott, Grebogi, and Yorke (OGY) who suggested a method to drive chaotic motions to the stable directions associated with unstable periodic orbits [43]. That is, their method moves poles lying outside the unit circle all the way to the origin in eigenvalue space [44]. An alternative, less demanding view based on optimal control theory seeks to drive the poles back into the unit circle. Both work with small forces, and each is briefly described.

The OGY method is most easily described by considering the observations made in discrete time intervals (coinciding with the period of the goal orbit) that can be unfolded in a space of dimension two. In this space, the dynamical variable is a two-dimensional vector $y(n)$, and the target is a fixed point $Y_F$ satisfying

$$y_F(u = 0) = F(y_F(u = 0), u = 0),$$

where the $u$ are parameters of the system. We presume that the $u$ can be controlled with external forces that remain small ($u = 0$ always). The dynamics are described by

$$y(n + 1) = F(y(n), u(n)).$$

A rule is needed for $u(n)$ that places the point $y(n + 1) - y_F$ on the stable direction of the local Jacobian $DF(y_F)$. This is assured by the following control requirement [43], which is correct when $u(n) = 0$.

$$u(n) = \frac{\lambda_u}{\lambda_{u-1}} \left[ \frac{(y(n) - y_F) \cdot f_u}{g \cdot f_u} \right],$$

where $f_u$ is the left eigenvector of $DF(y_F)$ associated with the unstable eigenvalue $\lambda_u$ that lies outside the unit circle.

$$g = \frac{\partial y_F(u)}{\partial u}$$

describes how the system varies with the control forces. There is a substantial amount of experimental verification of this basic OGY formula [45], yet there are several problematic aspects of the overall strategy. First, the formula requires the orbit to be in the neighborhood of the fixed point $y_F$ before control is applied. This can take some time, but there is a clever way to work around the problem using a method called targeting [46]. Unfortunately, targeting requires detailed knowledge of the global stable and unstable manifold structure of the dynamics, which is difficult to acquire. The second problem with the OGY strategy is that it is difficult to apply in higher dimensions, although there are ways to work around this, and they seem to perform adequately [47].

The final problem with the method is that it requires a lot
of knowledge about the dynamical system. The eigenvectors and eigenvalues of the local Jacobian matrix $DF(y)$ are needed in addition to $g$. However, all of these quantities can be evaluated in state space reconstructed from a single variable and its time delays.

An alternative to the strong requirement of the OGY method to place the unstable eigenvalues of the Jacobian matrix at the origin is to make sure they lie within the unit circle. This guarantees stabilization and requires much less detailed knowledge of the local dynamics. If this can be accomplished by an optimum control method, then there is no longer the requirement to be near the targeted unstable periodic orbit or to push unstable eigenvalues to the origin. Only knowledge of the approximate local dynamics and the variation of the vector field with respect to the control forces is required. This enables control over chaotic instabilities and allows rapid forcing of the chaotic orbits to a selected unstable periodic orbit [48].

There are at least two advantages of the dynamical systems approach to control of nonlinear systems. First, the fact that control can be done in the reconstructed state space means that detailed, accurate original state-space models of the dynamics are not required because these models can be learned well enough from observations of evolution on the attractor. Second, control of unstable periodic orbits of the chaotic system can be achieved with small control forces. The presence of these unstable periodic orbits is new to the discussion of control. We have not discussed how to find these unstable periodic orbits, but the method is well documented in the literature (see, for example, [49]). There is substantial literature on nonlinear control [50] and these developments hardly replace or displace that information. The methods described here are valuable because they add previously absent dimensions to the body of literature.

**Concluding Remarks**

The methods discussed in this article are tools for

- characterizing,
- classifying,
- predicting,
- and controlling nonlinear systems that are intractable with traditional linear tools. The examples in this article are novel applications and solve problems that have heretofore defied solution, often having been dismissed as "noise."

This article demonstrates the practical value of analyzing nonlinear signals in time-domain state space reconstructed from measurements. In many cases, the issue of whether the signal of interest is, in a technical sense, chaotic, is unimportant relative to the ability to exploit it using these algorithms. We anticipate that these techniques will integrate rapidly into the general toolkit of signal-processing techniques as an adjunct to linear techniques that are powerful in their own right.

Nonlinear signal processing is a young field and has just started to move from the laboratory to the real world. But a promising beginning has been established and software is available to accomplish the tasks discussed here.

Because deterministic, broadband signals with limited predictability have been dismissed as "noise" for so long, there are numerous and important applications for new tools that can process such signals. Applications that are being developed include:

- **Control of nonlinear systems**
  - mechanical vibration (machine tools, for example)
  - biological (cardiology)
  - lasers

- **Synchronization for communication**
  - private communications
  - noise-like emissions

- **Generalizations of cross-prediction**
  - (virtual sensors)
  - power-plant monitoring
  - ocean water levels
  - economics
  - physiology

![Image](image_url)

**16. Predicting a time series from a different type of sensor by predicting steam flow from the pressurizer level at the Pressurized Water Reactor. Predictions of flow were made using a model in a four-dimensional space constructed from pressurizer level data and its time delays (10,000 baseline samples and time delay of 100 s were used). Panel (a) shows actual flow data, (b) actual level data, and (c) the predicted flow.**
3. Predicting a time series from another spatially distant sensor. Data on ocean water level was simultaneously measured at Ft. Pulaski, Georgia, and Charleston, South Carolina, every six minutes for the year 1994. Predictive models of the water level at Ft. Pulaski were made in a $d = d_w = 4$ dimensional space constructed from Charleston water levels and their time delays. This yielded predictions for the Ft. Pulaski water levels based on observations of Charleston water levels for every six minutes of 1995. Measurements and predictions of Ft. Pulaski water levels, measurements of Charleston water levels, and errors in the Ft. Pulaski predictions are shown for 15 April 1995 through 25 April 1995.

Pathology detection and classification
- rotating machinery
- cardiology
- neurology

Improved signal detection
- noise model construction for matched filters
- explicit signal models
- improved spectral noise equalizers

There are numerous issues that need to be resolved and many improvements to be made. For example, one would like to know how much data and computation are required to evaluate the local Lyapunov exponents in $d$-dimensions and, most importantly, how much error is associated with any finite amount of data in the presence of a finite noise contamination level. The prescription of average mutual information should be proved and its connection to the Lyapunov exponents has yet to be explored. Similar questions can be asked of each of the statistical quantities in this article and in the sizable literature. These are just a few of the fertile research issues remaining.

In no small way, the excitement comes from the thought that the best applications have not yet been discovered. Perhaps they will be found through the use of a new tool or, inadvertently, during an attempt to find a solution to a seemingly mundane problem.

Acknowledgments
This work was supported in part by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Engineering and Geosciences, under grants DE-FG03-90ER14138 and DE-FG03-95ER14516, in part by the Office of Naval Research (Contract N00014-91-C-0125), in part from the National Science Foundation under grant NSF:NCR(FDP) NCR-9612250, and in part by Energy Power Research Institute (EPRI) under Contract WO8015-11. We are grateful to N.F. Rulkov for supplying us with the circuit data for our standard example, to EPRI and South Carolina Electric & Gas Company for the power-plant data, and to W. Scherer and M. Earle for the data from the NOAA tide stations. The ocean water level study was supported in part by the National Ocean Service, National Oceanic and Atmospheric Administration through U.S. Navy contract N00014-95-C-6004 with Neptune Sciences, Inc.

Henry D.I. Abarbanel is a Professor at the Department of Physics and the Marine Physical Laboratory of the Scripps Institution of Oceanography, both part of the University of California, San Diego, in La Jolla, California. Ted W. Frison is the President of Randle, Inc., in Great Falls, Virginia. Lev Sh. Tsimring is a Research Physicist at the Institute for Nonlinear Science at the University of California, San Diego, in La Jolla, California.

References