*Annals of Biomedical Engineering*, Vol. 24, pp. 513–536, 1996
Printed in the USA. 

0090-6964/96 $10.50 + .00


# Nonlinear Analysis of Biological Systems Using Short M-Sequences and Sparse-Stimulation Techniques

HAI-WEN CHEN, CHERYL J. AINE, ELAINE BEST, DOUG RANKEN, REID R. HARRISON, EDWARD R. FLYNN, and C. C. WOOD

Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM

**Abstract**—The m-sequence pseudorandom signal has been shown to be a more effective probing signal than traditional Gaussian white noise for studying nonlinear biological systems using cross-correlation techniques. The effectiveness is evidenced by the high signal-to-noise (S/N) ratio and speed of data acquisition. However, the "anomalies" that occur in the estimations of the cross-correlations represent an obstacle that prevents m-sequences from being more widely used for studying nonlinear systems. The sparse-stimulation method for measuring system kernels can help alleviate estimation errors caused by anomalies. In this paper, a "padded sparse-stimulation" method is evaluated, a modification of the "inserted sparse-stimulation" technique introduced by Sutter, along with a short m-sequence as a probing signal. Computer simulations show that both the "padded" and "inserted" methods can effectively eliminate the anomalies in the calculation of the second-order kernel, even when short m-sequences were used (length of 1023 for a binary m-sequence, and 728 for a ternary m-sequence). Preliminary experimental data from neuromagnetic studies of the human visual system are also presented, demonstrating that the system kernels can be measured with high signal-to-noise (S/N) ratios using short m-sequences.

**Keywords**—Volterra-Wiener approach, Volterra and Wiener kernels, Random versus pseudorandom signals, Anomalies of kernel estimation, Inserted method, Padded method, EEG and MEG studies

## INTRODUCTION

Several approaches to studying unknown biological systems consider the system initially as a "black box." In one approach, knowledge of the input-output (I/O) behaviors of the system are acquired by identifying the system's I/O transfer function. For example, the Laplace transfer function of a linear system is directly related to the I/O relationship of the system described by a linear differential equation. Unlike analyses of simple linear systems, analyses for nonlinear systems present a greater challenge, since high-order transfer functions are required for full characterization of a nonlinear system's I/O behaviors. One method used to characterize a nonlinear system (which is generally time-invariant or stationary, causal, and continuous), expresses the I/O relationship in terms of Volterra or Wiener functional expansions (25,37,39,45, 47,48). Such expansions are fully specified by a set of Volterra (or Wiener) kernels that can be estimated from the I/O measurements of the system under study. The system kernels provide a characterization of the system I/O relationship, and thus allow one to predict the response of a physical or physiological system to an arbitrary stimulus. (See, for example, 9,13,15,16,18,19,25,28,29,36,41,45,46 for physical and physiological interpretations of system kernels.) Furthermore, the measured system kernels can provide information about the internal structure of the practical system under study (8,10,17), which suggests a number of potential applications in the areas of system modeling and identification, and system fault detection and diagnosis (see 10 for a detailed review). For example, we are interested in characterizing the I/O relationship of populations of neurons in cortical areas of human brain, obtained from noninvasive electromagnetic measurements. Neuronal responses can be modeled as circuitry, and nonlinear analysis may ultimately allow one to characterize both the I/O relationship and the structure of neural system circuitry, using system kernels.

As discussed above, accurate kernel measurements are crucial to studies of practical nonlinear systems. Although methods are well established for identifying lower-order kernels, the measurement of high-order kernels is time consuming, prone to noise contamination, and presents an obstacle to efficient system modeling. As pioneers in this area, Lee and Schetzen in 1965 developed a practical method for measuring the system kernels using Gaussian

*Acknowledgment*—This work was supported by a grant from NEI (EY08610) and by the United States Department of Energy (DOE) Contract W-7405-ENG-36.

Address correspondence to Hai-Wen Chen, Biophysics Group, MS-D454, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.

*(Received 17Apr95, Revised 17Jul95, Revised 31Oct95, Accepted 24Jan 96)*

white noise as the input signal (23). One practical disadvantage of this method is the requirement of having a long record length as input to approximate true white noise. Since their seminal work, improved methods for measuring the higher-order kernels have been developed using deterministic or pseudorandom probing signals (6,7,11,12,19,21,22,27,28,43–46). For example, Sutter (43,44) has shown that deterministic binary m-sequences are quite effective for studying nonlinear visual systems. Benardate and Victor (6) have recently proposed an improved "hybrid" m-sequence technique using a sum of m-sequences as the input, which provides some control over anomalies in kernel measurements and permits the measurement of the main diagonal slice of the second-order binary kernel.

In this paper, a refinement of Sutter and Klein's (19,44) approach (*i.e.*, the "inserted" sparse-stimulation method and a relatively long m-sequence as a probing signal) is presented, which uses the "padded" sparse-stimulation method for measuring system kernels and a short m-sequence as a probing signal. The advantages for using short input signals are: (a) fast kernel calculations and (b) short data acquisition times. The present simulation results demonstrate (a) when short m-sequences are used as probing signals (a 1023 length for binary and a 728 length for ternary), both the padded and inserted sparse-stimulation methods can help alleviate the estimation errors, caused by anomalies, by increasing the anomaly-free area (AFA) of measured second-order kernels; and (b) when using sparse-stimulation, within the AFA, the errors of kernel measurements using the simpler, traditional cross-correlation method, are as low as those using the inverse-repeat method.

Preliminary experimental data from neuromagnetic studies of the human visual system are presented, demonstrating an application of this approach in basic biological research. Neuromagnetic responses were evoked by light-modulated, binary short m-sequences (1023 length) and were measured using highly sensitive SQUID (superconducting quantum interference device) sensors placed around the head surface. Preliminary analyses show that cross-correlations with high S/N ratio can be obtained using this method. Since the experimental recording period is short (20–30 s), using short m-sequences, this approach may be useful for studying mild nonstationary systems and systems that can be considered stationary only within a certain brief period of time [*e.g.*, short recording sessions are preferred for human subjects in EEG (electroencephalography) and MEG (magnetoencephalography) studies to avoid artifacts such as eye and head movements, *etc.*]

The results of this paper are organized into four sections: the m-sequence pseudorandom signal, sparse stimulation, experimental results of visual MEG studies, and general discussion.

## THE M-SEQUENCE PSEUDORANDOM SIGNAL

### *The Volterra-Wiener Approach and Properties of M-Sequences*

The Volterra-Wiener approach characterizes the I/O behaviors of a nonlinear system using functional expansions. The system under study is generally time-invariant, causal, and continuous, and can have single-input single-output (SISO), multi-input single-output (MISO), or multi-input multi-output (MIMO) configurations (10,25,37,39,45,47,48; see also 35, pp. 126–130, for a tutorial introduction of the Volterra -Wiener approach and its applications for EEG studies). For example, given an SISO nonlinear system with certain restrictions (25,37,39), Volterra theory (47) enables one to express the relationship between the system input $x(t)$ and output $y(t)$ as a series of multiple convolution integrals:

$$y(t) = k_0 + \int_0^\infty k_1(\tau)x(t-\tau)d\tau + \int_0^\infty \int_0^\infty k_2(\tau_1,\tau_2)\, x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \cdots + \int_0^\infty \overset{n}{\cdots} \int_0^\infty k_n(\tau_1, \cdots ,\tau_n)\, x(t-\tau_1)\cdots x(t-\tau_n)d\tau_1 \cdots d\tau_n + \cdots, \quad (1)$$

where the function $k_n(\tau_1, \cdots ,\tau_n)$ is the $n$th-order Volterra kernel, which can be assumed to be symmetric without loss of generality (25,37,39). The arguments $\tau_1, \cdots, \tau_n$ are in time units (same as $t$). The Volterra functional expansion represents a linear system as a special case when all of the kernels higher than the first-order are identically zero. In this case, the first-order kernel represents the impulse response function of the linear system. Sandberg (38) has recently proved that a Volterra or Volterra-like I/O functional expansion exists for an important large class of nonlinear systems.

In general, the Volterra kernels are difficult to measure, since the Volterra functionals of different orders are nonorthogonal to one another. Wiener (48) proposed an orthogonalization procedure for the Volterra functionals using a Gaussian white noise as the system stimulus. Based on Wiener's results, Lee and Schetzen (23) developed a practical cross-correlation technique for measuring the Wiener kernels using Gaussian white noise as the input signal. They have shown that the $n$th-order cross-correlation between the output $y(t)$ and the input $x(t)$ of a nonlinear system is proportional to the $n$th-order Wiener kernels and thus the stimulus-response cross-correlations can be used as the measurements of the Wiener kernels.

In practice, analog experimental data are digitized, and thus it is more convenient to express the auto- and cross-correlations in discretized forms:

$$\phi_{xx\cdots x}(\tau_1, \cdots ,\tau_n) = \frac{1}{2M-1} \sum_{i=-(M-1)}^{M-1} x(i-\tau_1)\cdots x(i-\tau_n), \quad (2)$$

and

$$\phi_{yx\cdots x}(\tau_1, \cdots, \tau_n) = \frac{1}{M}\sum_{i=0}^{M-1} y(i)x(i-\tau_1)\cdots x(i-\tau_n),$$

where $M$ is the length of data set $y(i)$ or $x(i)$. In addition, one may only need to calculate half of the symmetrical auto-correlation; and for stationary processes, the second-order autocorrelation can be reduced to a one-dimensional (1D) function, *i.e.*,

$$\phi_{xx}(\tau) = \frac{1}{M}\sum_{i=0}^{M-1} x(i)x(i-\tau). \quad (4)$$

Practically speaking, one cannot obtain true white noise of infinite length. Therefore, one major disadvantage of the Lee-Schetzen method is the requirement of having a long record length as input to approximate true white noise. The second-order autocorrelation of a true white noise would be a δ function. An ideal δ function would have a large value at $t = 0$, and identical zeros at all other points along the time axis. Fig. 1a shows how the record length of the Gaussian white noise influences the measurements of the second-order autocorrelations that were calculated using Eq. 4 with 50 shifts along the $\tau$ axis. The Gaussian white noise with variance $\sigma^2 = 1.6$ was generated using the function "gasdev" supplied in the book *Numerical Recipes in C* (33). The results show that when the record length $M$ of the Gaussian white noise is 20,000, the measured autocorrelation is a good approximation of a δ function. However, when $M$ was reduced to 728, large nonzero values of the autocorrelation at $\tau \neq 0$ resulted, which can cause estimation errors [the so-called nonorthogonal errors (19,25,44)]. The nonorthogonal estimation errors caused by the imperfections of the autocorrelation vary inversely with the square root of the record length (25).

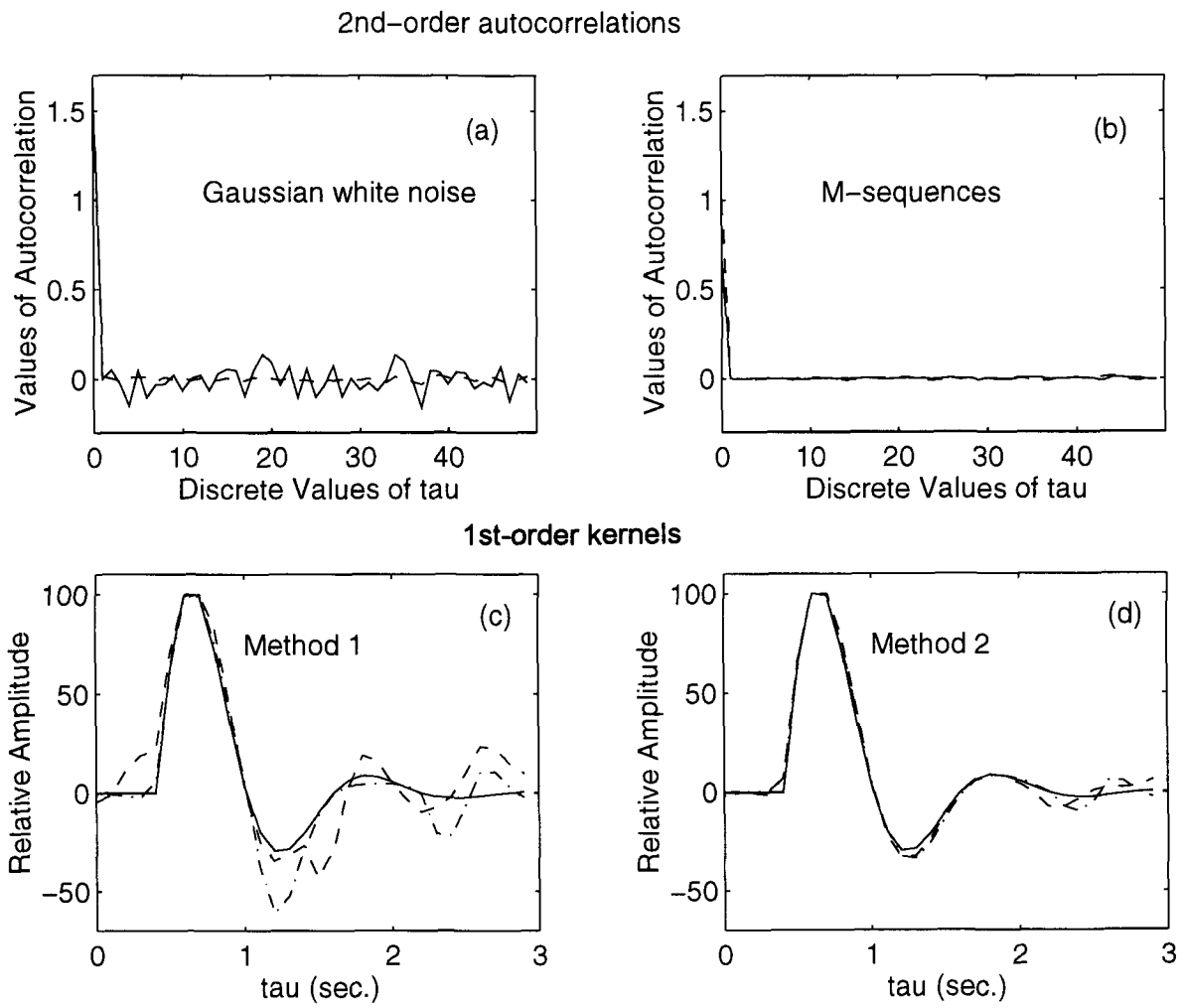


**FIGURE 1. (a) Second-order autocorrelations of Gaussian white noise. Solid line: length 728. Dashed line: length 20,000. (b) Second-order autocorrelations of short m-sequences. Solid line: ternary m-sequence with length 728. Dashed line: binary m-sequence with length 1023. (c and d) The first-order binary and ternary kernels using Methods 1 and 2. Solid lines: theoretical first-order kernels. Dashed lines: estimated first-order binary kernels. Dash-dotted lines: estimated first-order ternary kernels.**

*Properties of m-Sequences.* M-sequences, also called maximum length sequences (14), can be generated easily by a shift-register with an "exclusive OR" feedback configuration. The m-sequence can have multiple levels, binary (two levels) and ternary (three levels) sequences are used most often. For a binary m-sequence, each stage in the register has two states (0 or 1), and the state of each stage shifts to its neighboring stage recursively by feeding back a modulo 2 sum of selected stages (14,25,44). For a ternary m-sequence, the stages in the register have three states (0, 1, or 2), and the states of the register are shifted by feeding back a modulo 3 sum of selected stages (14, pp. 111). When the locations of the feedback stages are selected appropriately, the generated sequence will reach maximum length before repeating. That is, within the maximum length (a full cycle of m-sequence), the state string of stages in the register will never repeat. This is why the deterministic m-sequence is more efficient and less redundant than stochastic white noise. The maximum length is $2^n - 1$ for a binary m-sequence, and $3^n - 1$ for a ternary m-sequence, where $n$ is the number of stages in the register.

The second-order autocorrelations of a binary m-sequence with length 1023 ($2^{10} - 1$) and a ternary m-sequence with length 728 ($3^6 - 1$) are shown in Fig. 1b. For satisfying the zero-mean condition, the 0 and 1 values in the binary m-sequence are converted to $-1$ and 1, respectively, and the 0, 1, and 2 values in the ternary m-sequence are converted to $-1$, 0, and 1, respectively. The mean value is 0.00097 ($= 1/1023$) for the binary m-sequence, and 0.0014 ($= 1/728$) for the ternary m-sequence. The results in Fig. 1b show that the autocorrelation functions are good approximations of the $\delta$ function even for *short* m-sequences. To quantitatively evaluate the deviations of the autocorrelations from the true $\delta$ function, we introduce a percentage mean-square-error index (MSEI) as defined below:

$$\text{MSEI} = \frac{\sqrt{\dfrac{1}{L-1}\sum_{\tau=1}^{L}\phi_{xx}^2(\tau)}}{\phi_{xx}(0)}, \qquad (5)$$

where $L$ is the shift along the $\tau$ axis. Note that different values of $L$ may result in different MSEI values, even for the same sequence. In general, the length $L$ should be larger than the system memory $m$ (See Eq. 8a for the definition of $m$). It is obvious from Eq. 5 that for a true $\delta$ function MSEI = 0; the larger the MSEI value, the more the autocorrelation deviates from the true $\delta$ function. At $L = 50$, the MSEI value is 4.35% for Gaussian white noise with a length of 728 (solid curve in Fig. 1a); the MSEI value is 0.65% for Gaussian white noise with a length of 20,000 (dashed curve in Fig. 1a); the MSEI value is 0.65% for ternary m-sequence with a length of 728 (solid curve in Fig. 1b); and the MSEI value is 0.57% for binary m-sequence with a length of 1023 (dashed curve in Fig. 1b). Given the equivalence of MSEI values for short m-sequences (728 samples) and long stochastic Gaussian white noise (20,000 samples), a full-cycle short m-sequence can effectively eliminate the autocorrelation noise. It is important to use a full-cycle m-sequence, since the autocorrelation of a sequence of 1023 samples extracted from a full-cycle m-sequence with length 2047 or longer will be much worse than a full-cycle m-sequence with length 1023.

### *The Anomalies*

The properties of autocorrelations with different orders for Gaussian white noise are the basis for measuring the Wiener kernels using cross-correlation techniques (6,23,25). The property of the second-order autocorrelation is related to the identification of linear systems and the first-order kernel of nonlinear systems (6,14,23,25), whereas the properties of the higher-order autocorrelations are related to the identification of the second- and higher-order kernels of nonlinear systems (6,23,25). Since m-sequences have second-order autocorrelations that are similar to those for real white noise, as shown in Fig. 1b, they are widely used for identifying linear systems. However, higher, even-order autocorrelations of m-sequences are quite different from those of Gaussian white noise. Serious "anomalies" (estimation errors) result in the second- and higher-order kernels using m-sequences as system inputs. The anomalies manifest as biases in kernel estimation. For example, spurious bumps (I-spikes) may appear in the estimated kernels (see the simulated examples in Fig. 3, c and d). Previous studies (3,4,6,24,25,34) have attributed these estimation errors to the anomalies and imperfections in the higher-order autocorrelations of the m-sequence. For example, Srebro and Wright (40) have shown serious anomalies when using short ternary m-sequences for VEP (visual evoked potential) studies; Benardete and Victor (6) have quantitatively shown how the higher-order autocorrelations of m-sequences cause anomalies. Recently, Sutter (43,44) proposed a more detailed explanation for the occurrence of these estimation biases for the binary m-sequence case. Based on the useful deterministic property of the binary m-sequence, *i.e.*, that all of the auto-products of a binary m-sequence are still the same sequence up to a shift, Sutter proposed a method to calculate the higher-order binary kernels using only a first-order cross-correlation with a long shift. Sutter suggested that the origin of the anomaly problem for the binary m-sequence is caused by the overlap of higher-order kernel slices. He demonstrated the quantitative condition under which anomalies will not occur:

$$n \geq m - d, \tag{6a}$$

where $n$ is the stage number of the register generating the m-sequence (*i.e.*, the sequence length is $2^n - 1$); $m$ is the system memory number (*e.g.*, if the system memory is 300 ms, the sampling interval of the response is 20 ms, then $m = 300/20 = 15$); and $d$ is the system delay [Sutter (44) refers to $m - d$ as the "input window"]. Therefore, the condition in Eq. 6a is quite restrictive. For example, to study a system with a system memory number $m = 30$ and a delay $d = 5$, one would have to use an m-sequence with a length of $2^{25} - 1 = 33{,}554{,}431$. Obviously, increasing the sequence length is not a practical way to avoid the anomalies. In the next section, we will demonstrate that the sparse-stimulation technique can effectively reduce the anomalies even for short m-sequences. It is also worth noting that Eq. 6a was derived using a deterministic binary approach with Walsh transforms, which is only suitable for binary sequences. In this paper, we also examine ternary sequences and thus use the cross-correlation approach with *short* input m-sequences. When the computation is within the size of the Walsh array, the deterministic binary approach and the cross-correlation approach are almost equivalent to one another (see 44 for a detailed discussion comparing these two approaches). In the deterministic binary approach, the system delay $d$ has to be chosen in advance so that the Walsh array only needs to cover the input window. In the cross-correlation approach, one does not need to know the system delay in advance and thus Eq. 6a should be modified as

$$n \geq m. \tag{6b}$$

*Simulation Results*

In this section, we use computer-simulated examples to show: (a) how well one can measure the first-order kernel for single-input and multi-input nonlinear systems using short m-sequences; (b) the seriousness of the anomaly problems when measuring the second-order kernel; (c) a comparison of two methods for kernel calculations [one is the traditional Lee-Schetzen cross-correlation method (Method 1), and the other is the inverse-repeat method (Method 2)]; and (d) a comparison of measured binary and ternary kernels. Simulation results comparing Method 1 with Method 2 are shown in Fig. 1, c and d. System simulations were performed with an IBM compatible PC 486 computer using Personal VisSim (Visual Solutions, Inc. Westford, MA, USA), a commercial software package. The accuracy of the measured kernels were evaluated using a simple LN (Wiener) model whose theoretical first- and second-order kernels are readily available for comparison. For example, the first-order kernel of an LN model is proportional to the impulse response function of the linear subsystem in this model, and the second-order kernel is proportional to the outer product of the impulse response function. The block diagram programming chart for simulating an LN model is illustrated in Fig. 2. A half-square nonlinearity was used for the N operator (the *half-square* box in the chart), the same as the simulation in (10):

$$y(t) = N[u(t)] = \left[\frac{|u(t)| + u(t)}{2}\right]^2, \tag{7}$$

where $u(t)$ and $y(t)$ are the input and output of the $N$ operator, respectively. The half-square is neither an even nor an odd function, therefore, the system has both even- and odd-order kernels. The linear subsystem (the $h(t)$ box in the chart) can be chosen either in discrete form using a $Z$ transfer function (as used in the simulation in 10), or in continuous form using a Laplace transfer function. The two forms result in similar accuracy, with the discrete form being slightly better when using discrete inputs. The continuous form will be more suitable when using continuous inputs, as shown in the simulation cases in the next section. The linear subsystem in Fig. 2 is in continuous form, with parameters $h(t) = 4e^{-2.2(t-0.4)} \sin[5.2(t - 0.4)]$ for $t \geq 0.4$ and $h(t) = 0$ for $t < 0.4$, whose Laplace transfer function is

$$H(s) = \frac{130e^{-0.4s}}{s^2 + 4.4s + 31.88}. \tag{8}$$

The simulation was conducted using the Euler integration algorithm with step size 0.01 s (Fig. 2). The *input.dat* box in Fig. 2 represents a binary m-sequence with length 1023 (Plot 1 in Fig. 2). The input and output sampling intervals, $\Delta t_s$ and $\Delta t_r$, were equal in this case (0.1 s); therefore the simulated duration was 102.3 s. The simulation took 21 sec in real time to finish. Two methods for measuring the system kernels have been discussed in (10). In Method 1, the original Lee-Schetzen method, the first- and second-order kernels were calculated by cross-correlating the *input.dat* and the *output.dat* data files. In Method 2, also referred to as the "inverse-repeat method" (6,19,44), the experiment is supposed to be conducted twice to obtain the two responses to inputs $x(t)$ and $-x(t)$. However, the simulation shown in Fig. 2 was conducted only once because two identical LN models were used. The second-order kernel was calculated by cross-correlating the *input.dat* and the *output_e.dat* files, and the first-order kernel was calculated by cross-correlating the *input.dat* and the *output_o.dat* files. (Note that the main diagonal slice of the second-order binary kernel is not measurable using the cross-correlation technique, since in calculation of the diagonal slice there is a squaring of the binary input.)

The estimated first-order cross-correlations (the first-order binary kernel) using Methods 1 and 2 were plotted as

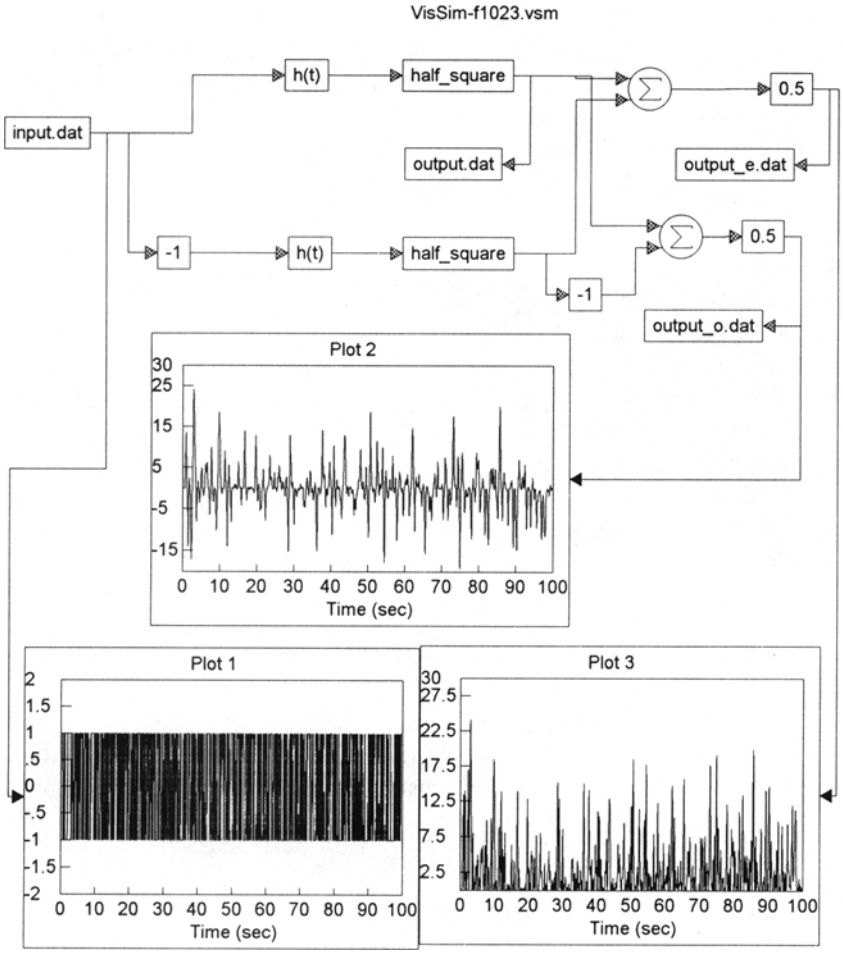


**FIGURE 2. A block diagram programming chart simulating an LN model. The *input.dat* box represents a binary m-sequence with length 1023, which is shown in Plot 1. The *output.dat* box represents the LN model's response (output) to the m-sequence input. The *output_o.dat* and *output_e.dat* boxes represent the odd-order and even-order output components, respectively (Plots 2 and 3). The *h(t)* and *half_square* boxes represent the linear subsystem and the half-square N operator of the LN model, respectively.**

dashed curves in Fig. 1, c and d, respectively. The ternary kernels were estimated using a ternary m-sequence with length 728 as the *input.dat* file in Fig. 2, and the system parameters were the same as the binary case discussed above. The sampling interval $\Delta\tau$ was 0.1 s and the number of samples along the $\tau$ axis was 30. The curve values are almost damped to zero at the end of the $\tau$ axis ($\tau = 3$). According to Eq. 6, the effective system memory number $m$ can be approximated as 30, and the system delay $d$ is 4. Each curve in Fig. 1, c and d, has been normalized to have a positive peak value of 100 units. It is obvious from Fig. 1, c and d, that Method 2 (Fig. 1d) results in smaller kernel estimation errors than Method 1 (Fig. 1c). The advantage of Method 2 is that it eliminates leakage between the even- and odd-order kernels. However, the drawback of Method 2 is that the experiment must be conducted twice. The results above were obtained using the usual white-noise approach (*i.e.*, $\Delta t_s = \Delta t_r$). In the next section, we will show that, when using the sparse-stimulation technique (*i.e.*, $\Delta t_s > \Delta t_r$), these two methods result in similar estimation accuracy and thus the simpler Method 1 can be used for practical applications.

Fig. 3, a and b, shows the theoretical second-order kernels with and without the main diagonal slice, respectively. The sampling interval $\Delta\tau_1$ (or $\Delta\tau_2$) was 0.1 s and the number of samples along the $\tau_1$ and $\tau_2$ axes was 30. The estimated second-order cross-correlations using Method 2 are plotted in Fig. 3c (using a 1023 binary m-sequence as input, $n = 10$) and Fig. 3d (using a 728 ternary m-sequence as input, $n = 6$). Both plots (Fig. 3, c and d) show strong anomalies, *i.e.*, there are many spurious bumps (I-spikes) as compared with Fig. 3b. According to Eq. 6b, we would need a binary m-sequence input with length $2^{30} - 1$ to avoid the anomalies. In other words, the estimated second-order kernel can only have an anomaly-free area of ($0 \leq \tau_1 \leq n\Delta\tau_1$, $0 \leq \tau_2 \leq n\Delta\tau_2$), which is satisfied by the measured second-order kernel in Fig. 3c, where no anomalies occurred within the area ($0 \leq \tau_1 \leq 1$ s, $0 \leq \tau_2 \leq 1$ s). (We have tested this conclusion using other binary m-sequence lengths (*e.g.*, 2047) and

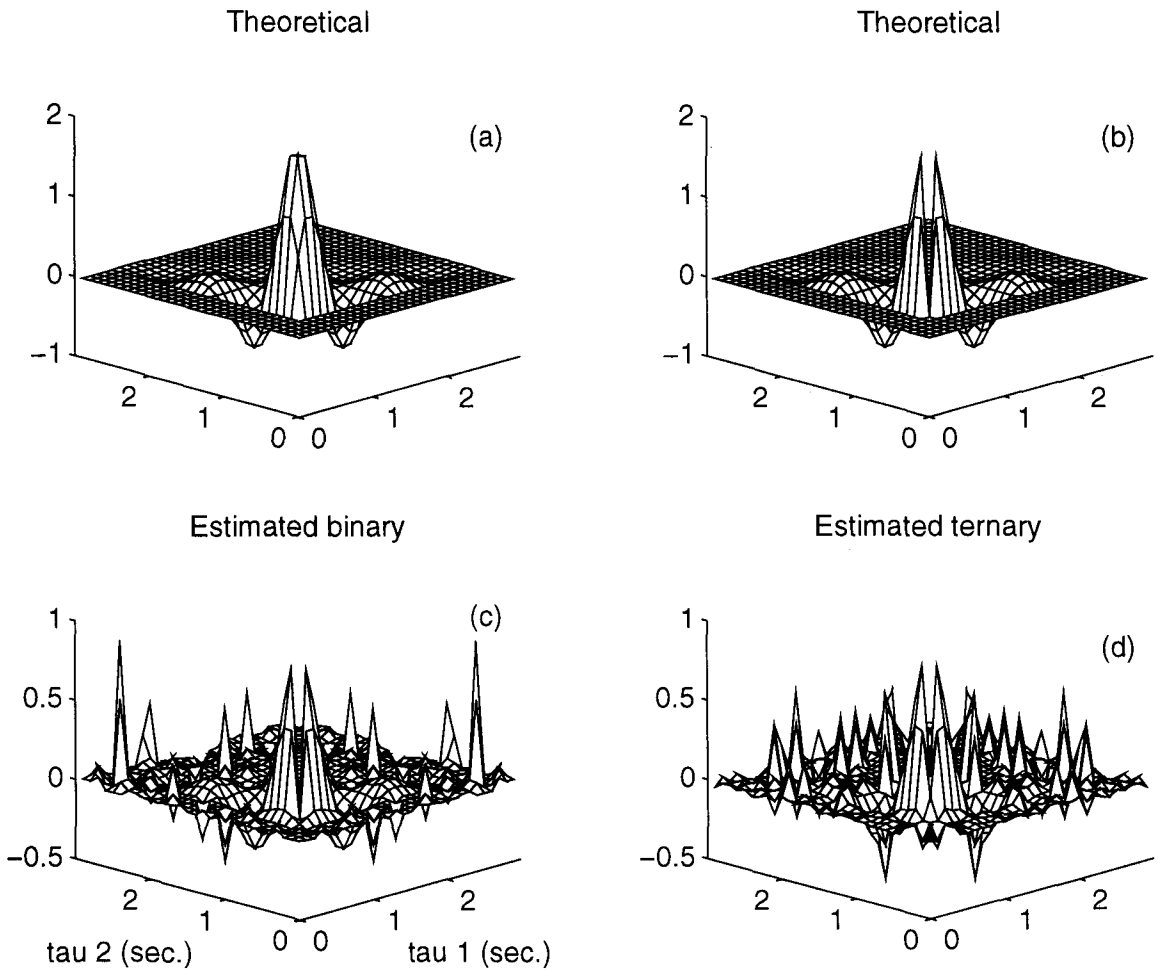


**FIGURE 3. (a and b) Theoretical second-order kernels with and without the main diagonal slice. (c) An estimated second-order binary kernel using a 1023 binary m-sequence as input (Method 2). (d) An estimated second-order ternary kernels using a 728 ternary m-sequence as input (Method 2).**

other system parameters.) Note that the main diagonal slice of the estimated second-order binary kernel has near-zero values, *i.e.*, the main diagonal values of the second-order kernel cannot be obtained from the second-order cross-correlation.

## SPARSE-STIMULATION

### *The Zero-Order Hold*

So far we have discussed m-sequences generated in a discrete form. However, in practice, many system inputs are applied in continuous form. The technique applied most frequently to convert a discrete input to a continuous one is the zero-order hold (14,25,40,41), which holds the discrete input values constant for a period of $\Delta t_s$ (the sampling interval). The converted continuous binary and ternary m-sequences are shown in Plots 1 and 2 in Fig. 4, respectively. They are generated by cascading a zero-order hold to the output of the binary m-sequence (with 1023 length, the *b*1023.*dat* box in Fig. 4) or the ternary m-sequence (the *t*728.*dat* box in Fig. 4). The duration $T$ of the hold was set at 0.1 s, the same as $\Delta t_s$ for the discrete m-sequences in Fig. 4. (Note: The computer-generated hold in Fig. 4 results in an ideal waveform with transition time between states $(-1, 0, +1)$ being zero, *i.e.*, $T_t = 0$. However, in practice, the hold's function is performed by physical digital-to-analog transducers, which will give rise to a certain transition time caused by some inevitable inertia of the physical transducers. Nevertheless, the estimation errors may be negligible if $1/T_t > f_{system}$, where $f_{system}$ is the frequency bandwidth of the system under study (25). See (14) and (25) for detailed evaluations of the estimation errors caused by reversible and irreversible transition times.) The Laplace transfer function of the zero-order hold is (30,32)

$$HD(s) = \frac{1 - e^{-Ts}}{s}, \tag{9}$$

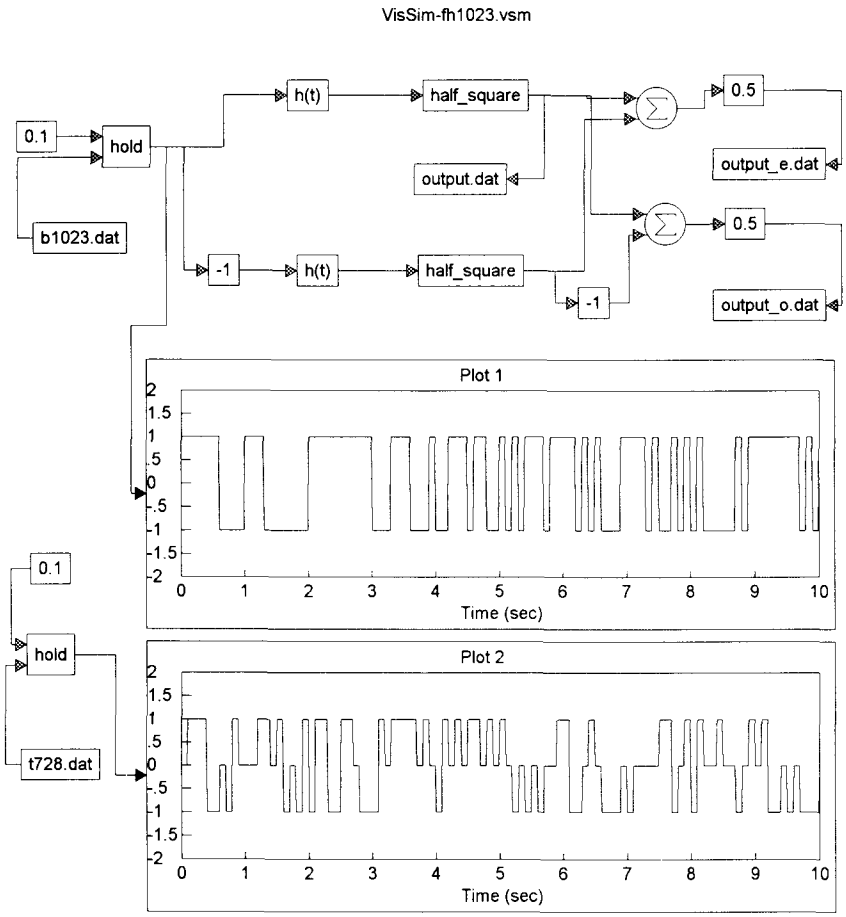


**FIGURE 4. Block diagram programming chart simulating an LN model using continuous m-sequences. The continuous m-sequences are generated by cascading a zero-order hold (the *hold* box) to the output of the binary m-sequence (with 1023 length, the *b1023.dat* box) or the ternary m-sequence (the *t728.dat* box). The duration $T$ of the hold was set to 0.1 s (the 0.1 box), the same as $\Delta t_s$ for the discrete m-sequences ($\Delta t_s = \Delta t_r$). The converted continuous binary and ternary m-sequences are shown in Plots 1 and 2, respectively.**

where $T$ is the duration of the hold (and $T = \Delta t_s$). For $s = j\omega$ in Eq. 9, we obtain the Fourier transfer function for the zero-order hold:

$$HD(\omega) = T \frac{\sin(\omega T/2)}{\omega T/2} e^{-Tj\omega/2}. \qquad (10)$$

As shown in Fig. 1b, the autocorrelations of the short m-sequences are very close to a $\delta$ function, the Fourier transfer function of which is a constant. Therefore, the frequency properties of the continuous m-sequence will be solely determined by Eq. 10. By taking the discrete Fourier transform of the two autocorrelations of the discrete m-sequences in Fig. 1b, the power density spectra of these two short discrete m-sequences (with lengths of 1023 and 728) were computed and plotted in Fig. 5a, which can be approximated as constants. The two ends of the $x$-axis ($\pm 0.5 f_s$, where $f_s = 1/\Delta t_s$) are the negative and positive Nyquist frequencies. The power density spectrum of a continuous m-sequence (a discrete m-sequence cascaded by a hold), which is also in discrete form, has been derived by Davies (14, pp. 67):

$$\Phi_{xx}(f) = \frac{a^2(N+1)}{Nf_s} \left[\frac{\sin(\pi f/f_s)}{\pi f/f_s}\right]^2, \qquad (11)$$

where $N$ is the length of the m-sequence, $a$ is the physical intensity unit of the m-sequence, *i.e.*, the two levels $-a$ and $+a$ for the binary case; the three levels $-a$, 0, and $+a$ for the ternary case. The function $[\sin(\pi f/f_s)/(\pi f/f_s)]^2$ was plotted in Fig. 5b, which shows that the zero-order hold has a low-pass filtering property. The power density spectrum reaches its first zero values at $f = f_s$ and has a value of $a^2(N+1)\Delta t_s/N$ at low frequencies and falls by approximately 3 *db* at the frequency $f = f_s/3$ Hz. Therefore, the effective frequency bandwidth of a continuous m-sequence is from $f = f_s/N$ to $f = f_s/3$. It is obvious that the sampling interval $\Delta t_s$ ($= 1/f_s$) of the discrete m-sequences (and thus the duration $T$ of the hold) is a crucial parameter for determining the bandwidth of continuous m-sequences and the impact of low-pass filtering (by the

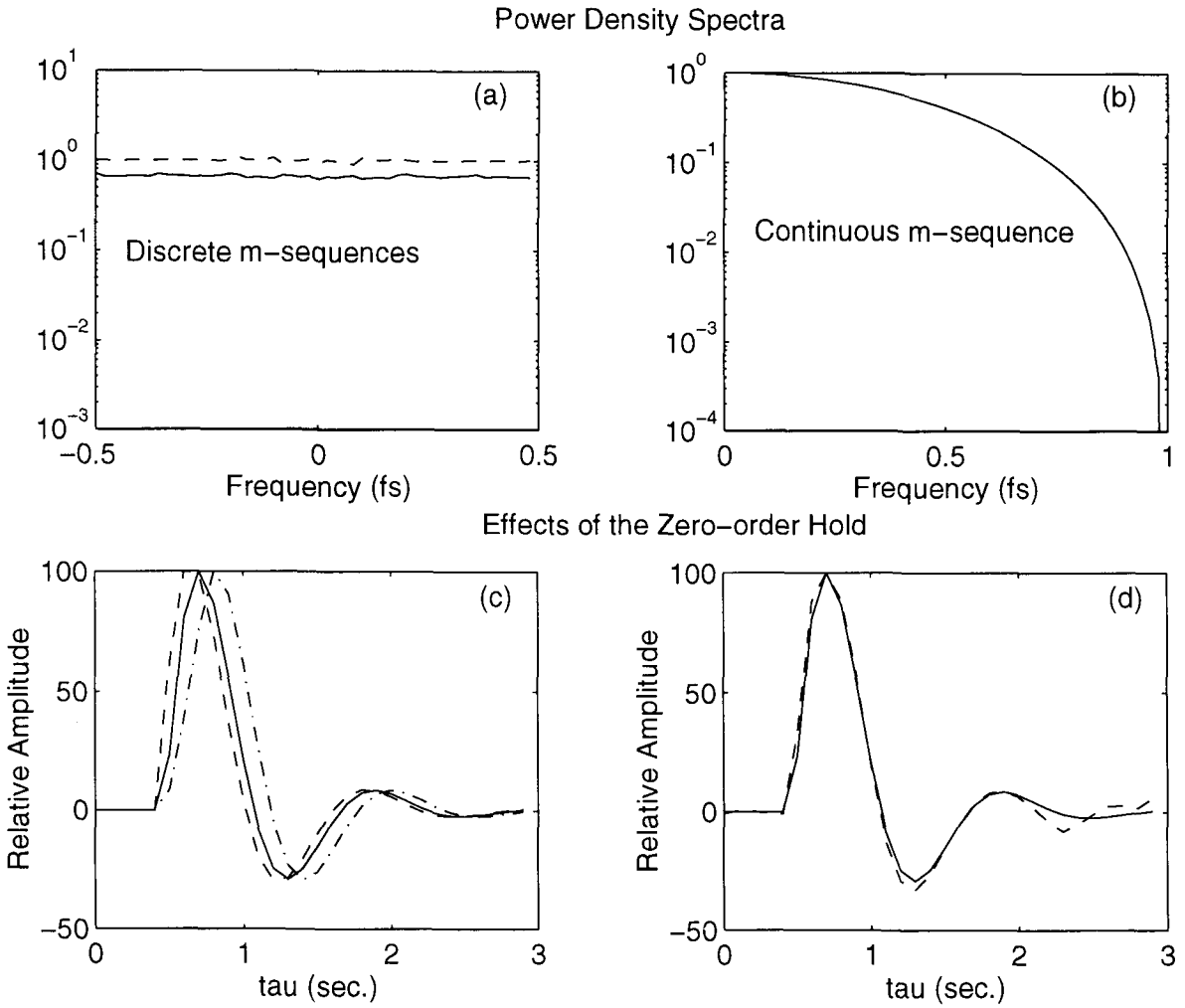


**FIGURE 5. Discrete versus continuous m-sequences and effects of the zero-order hold. (a) Solid line: The power density spectrum of the discrete ternary m-sequence (728). Dashed line: The power density spectrum of the discrete binary m-sequence (1023). (b) Solid line: The function $[\sin(\pi f/f_s)/(\pi f/f_s)]^2$ of the power density spectrum of continuous m-sequences (Eq. 11). The unit of the frequency axes is $f_s = \Delta t_s$ Hz in (a) and (b). (c) Simulation showing the effect of low-pass filtering by the zero-order hold with different durations ($T$): dashed line: $T = 0$ (no hold); solid line: $T = 0.1$ s; and dash-dotted line: $T = 0.3$ s. (d) The solid line is replotted from the solid curve in (c). The dashed line is the estimated first-order kernel using Method 2 from the simulation in Fig. 4.**

hold) on the resulting system kernels. As shown in Fig. 5b, if $f_s$ is chosen so that $f_s/3 > f_{system}$, then the effect of low-pass filtering may be negligible. However, as discussed later, sparse-stimulation techniques may be used to avoid anomalies occurring in higher-order kernel measurements. In this case, $\Delta t_s$ is chosen to be larger than $\Delta t_r$; consequently, the frequency bandwidth may be narrower than that of the system under study. Therefore, it will be important to recover the true system kernels from the low-pass filtering caused by the zero-order hold. This configuration can be considered as a linear subsystem (the hold) cascaded in front of the system under study. The simulation results in Fig. 5c show how different durations ($T$) of the hold can change the system's impulse response function (*i.e.*, effects of low-pass filtering). The curves are normalized and plotted in Fig. 5c for comparison as dashed ($T = 0$), solid ($T = 0.1$s), and dash-dotted ($T = 0.3$ s) curves. From Fig. 5c it was found that when increasing $T$ ($=0$, 0.1, and 0.3 s) the resulting impulse response shifts towards the right-hand side.

According to the Volterra kernel expressions for cascade nonlinear systems (25, pp. 393, 37), since we know the Fourier transfer function of the hold as shown in Eq. 10, the true first- and second-order Volterra kernels can be expressed as deconvolution relationships of the measured kernels and the hold:

$$K_1(\omega) = \frac{K'_1(\omega)}{HD(\omega)}, \text{ and } K_2(\omega_1\omega_2) = \frac{K'_2(\omega_1\omega_2)}{HD(\omega_1)HD(\omega_2)}, \tag{12}$$

where $K_1'(\omega)$ and $K_2'(\omega_1,\omega_2)$ are the measured first- and second-order Volterra kernels, and $K_1(\omega)$ and $K_2(\omega_1,\omega_2)$ are the recovered true system Volterra kernels. The results in Eq. 12 can be easily generalized for higher-order Volterra kernels. According to the results in Ref. 23, section 10 (25,39), the results in Eq. 12 can be extended for Wiener kernels by simply substituting the Volterra kernels in Eq. 12 with the corresponding Wiener kernels.

The simulation in Fig. 4 was designed to confirm the results in Eq. 12. The simulation settings were the same as those presented above (Simulation Results) (Fig. 2); the only difference was the additional hold ($T = 0.1$ s) cascaded between the discrete binary 1023 m-sequence and the system. According to Eq. 12, the measured first-order kernel will be proportioned to the convolution of the hold ($T = 0.1$ s) and the true first-order kernel, *i.e.*, the solid curve in Fig. 5c. The measured first-order kernel using the inverse-repeat method (Method 2) and the solid curve in Fig. 5c were normalized and plotted in Fig. 5d as dashed and solid curves, respectively. These two curves are quite proportional to each other.

### *Sparse-Stimulation*

All of the simulations conducted above used a usual white-noise approach (25), *i.e.*, $\Delta t_s = \Delta t_r$. Sparse-stimulation refers to the situation when a stimulus is sampled less often than the response (19), *i.e.*, $\Delta t_s > \Delta t_r$. The mathematical foundation of the white-noise approach using the sparse-stimulation technique and its applications have been developed and presented in (19,44). There are several advantages to using sparse-stimulation:

1. For many experimental situations there are fewer restrictions in choosing $\Delta t_r$ than $\Delta t_s$. For example, in our own MEG studies discussed in the next section, the sampling rate of the data acquisition system for the responses can be as high as 40 kHz; however, the maximum rate of presentation for visual stimuli is 30 Hz because of the video card. The sparse-stimulation technique allows one to select $\Delta t_r$ (independently of $\Delta t_s$) based on the frequency bandwidth of the responses and the required resolution for the measured kernels (the kernel resolution is determined by $\Delta t_r$). The choice of $\Delta t_s$ may be based on the limitations of the equipment.

2. For a pseudorandom (or colored noise) input, such as a continuous m-sequence discussed above, the power level of the input may be low if the input covers a large frequency bandwidth. In this case, the input may not effectively drive the system under study if the system has high thresholds, resulting in noisy measurements. As indicated in Eq. 11, if we increase $\Delta t_s$ ($= 1/f_s$), the power level of the m-sequences will be higher, even though the frequency bandwidth will decrease.

3. The most important advantage of sparse-stimulation techniques is the reduction of anomalies that occur in higher-order kernel measurements. As discussed in the last section, to increase the anomaly-free area, we have to increase m-sequence length, which is not a practical way to reduce anomalies. In this section, we will show that the anomaly-free area of the second-order kernel can be increased by increasing the sparse-stimulation ratio (SSR), even for short m-sequences. The SSR is defined as $\text{SSR} = \Delta t_s/\Delta t_r$.

Simulations of sparse-stimulation have been conducted using the same simulation settings outlined above (The Zero-Order Hold) (Fig. 4), except for the duration of the hold. The hold duration was changed from $T = 0.1$ s to $T = 0.3$ s (*i.e.*, $\Delta t_s = 0.3$ s), while still using the same $\Delta t_r$ ($= 0.1$ s). Therefore, the simulation time ($1023 \times 0.3 = 306.9$ s) was three times longer than that used for the simulation in Fig. 4. The length of the output files was 3069 (*versus* the 1023 length of the input file). Figure 6 shows these differences graphically by comparing the traditional stimulation case ($\Delta t_s = \Delta t_r$) for a 1023 binary m-sequence (Fig. 6a) with the sparse-stimulation case (SSR = 3) for the same m-sequence (Fig. 6b).

There are two methods for calculating the kernels from sparse-stimulation data (*e.g.*, the data obtained from the simulation discussed above (SSR = 3)), the previously proposed "inserted" method (19,44), and the "padded" method proposed in this paper. That is, these two methods are based on the same stimulus-response data (Fig. 6b). First we discuss the inserted method to demonstrate the advantages of the sparse-stimulation technique, which can help reduce the anomalies. A discussion of the padded method follows. The effects of low-pass filtering on the measured system kernels associated with these two methods will also be discussed.

*The Inserted Method.* A binary m-sequence with a length of 1023 (input $x(t)$) was used to obtain the simulation results shown in Fig. 6b. The output files ($y(t)$) obtained had a length of 3069, as shown in Eq. 13:

$$x(t) = \underbrace{[1, -1, -1, 1, 1, -1, \cdots]}_{1023}, \text{ and}$$
$$y(t) = \underbrace{[d_1, d_2, d_3, \cdots d_{3069}]}_{3069}, \tag{13}$$

where, $t = 0, 1, 2, \cdots$. However, the traditional cross-correlation method (25) requires the same length for the input and output files. To satisfy this requirement, we can divide the output data file $y(t)$ into three subfiles, each with a length of 1023. The first subfile $y_1(t)$ extracts data from the original output file at data numbers $1, 4, 7, \cdots$; the second subfile $y_2(t)$ extracts data from the original

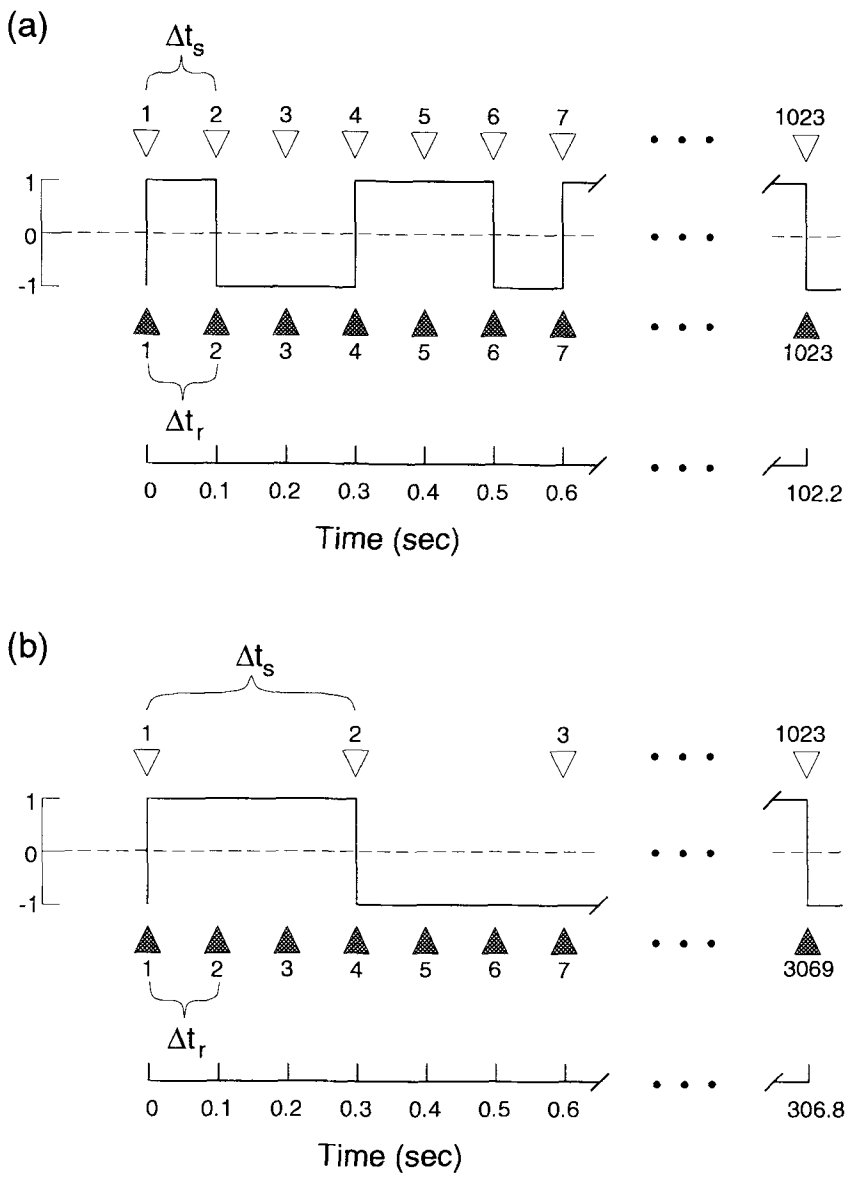


**FIGURE 6. Traditional stimulation and sparse-stimulation for a binary m-sequence. (a) Traditional stimulation using a 1023 binary m-sequence. Both input (m-sequence) file and the output file are of 1023 length ($\Delta t_s = \Delta t_r = 0.1$ s); the stimulation time is 102.3 s. (b) Sparse-stimulation using the same m-sequence. The input file is still of 1023 length, but the output file is of 3069 length ($\Delta t_s = 3 \times \Delta t_r = 0.3$ s); the simulation time is 306.9 s.**

output file from data numbers 2, 5, 8, · · · ; and the third subfile $y_3(t)$ contains data from 3, 6, 9, · · · , as shown in Eq. 14:

$$y_1(t) = y(3t) = \underbrace{[d_1, d_4, \cdots, d_{3067}]}_{1023},$$

$$y_2(t) = y(3t + 1) = \underbrace{[d_2, d_5, \cdots, d_{3068}]}_{1023},$$

and

$$y_3(t) = y(3t + 2) = \underbrace{[d_3, d_6, \cdots, d_{3069}]}_{1023}. \quad (14)$$

Therefore, we can calculate three subkernel sets (*e.g.*, the first-order subkernels: $\phi_{y_1x}(\tau)$, $\phi_{y_2x}(\tau)$, and $\phi_{y_3x}(\tau)$) by cross-correlating each of the three subfiles with the same input file $x(t)$. Then we can insert the three estimated subkernel sets into one kernel set, according to the relative order of the subfiles. For example, for the first-order kernel, let

$\phi_{y_1x}(\tau) = [i_1, i_2, i_3, \cdots]$, $\phi_{y_2x}(\tau) = [j_1, j_2, j_3, \cdots]$, and $\phi_{y_3x}(\tau) = [k_1, k_2, k_3, \cdots]$.

Then, $\phi_{y_1x}(\tau)$ will be placed at data numbers 1, 4, 7, · · · , in the inserted first-order kernel $k_1(\tau)$; and so on for $\phi_{y_2x}(\tau)$ (at (2, 5, 8, · · ·)) and $\phi_{y_3x}(\tau)$ (at (3, 6, 9, · · ·)), as shown in Eq. 15:

$$k_1(\tau) = [i_1, j_1, k_1, i_2, j_2, k_2, i_3, j_3, k_3, \cdots]. \quad (15)$$

Note that the resolution is $\Delta t_s$ for the three subkernels, but is $\Delta t_\tau$ for the inserted kernel.

From the procedure (inserted method) discussed above, it is clear why sparse-stimulation can reduce the anomalies. As discussed in the last section, the anomaly-free area for each of the three subkernels is $n = 10$ for the input sequence length of 1023, and thus the anomaly-free area for the *inserted* kernels will be $3 \times 10 = 30$. Therefore, the sparse-stimulation technique can reduce the anomalies much more effectively than a method that increases the length of the input sequences. It is generally true that when using sparse-stimulation Eq. 6b can be modified as

$$n \geq m/\text{SSR}, \quad \text{or} \quad \text{AFA} = n \times \text{SSR}, \qquad (16)$$

where AFA denotes an anomaly-free area. The first-order, inserted kernel from the simulation of sparse-stimulation using the inverse-repeat method (Method 2) was calculated and plotted in Fig. 7a as the dash-dotted curve. The theoretical first-order kernel was plotted in Fig. 7a as a solid curve. The inserted kernel shifted toward the right-hand side; this shift is mainly caused by the low-pass filtering effect of the hold. In the inserted method, the subkernels ($\phi_{y_1x}(\tau)$, $\phi_{y_2x}(\tau)$, and $\phi_{y_3x}(\tau)$) are the convolutions of the true kernel and the holds for $T$ are 0.3, 0.2, and 0.1 s, respectively. Therefore, the true kernels can be recovered using Eq. 12. As shown in Fig. 5c, the effect of the hold for $T = 0.3$ s on the subkernels related to the first subfile will be the strongest regarding the shift of the measured inserted kernels. The first-order inserted kernel and the dash-dotted curve in Fig. 5c (the hold with $T = 0.3$ s) are replotted in Fig. 7b as dash-dotted and solid curves, respectively. They have similar shifts toward the right-hand side. Although we can obtain the full first-order kernel (with the sampling interval $\Delta t_r$) using the

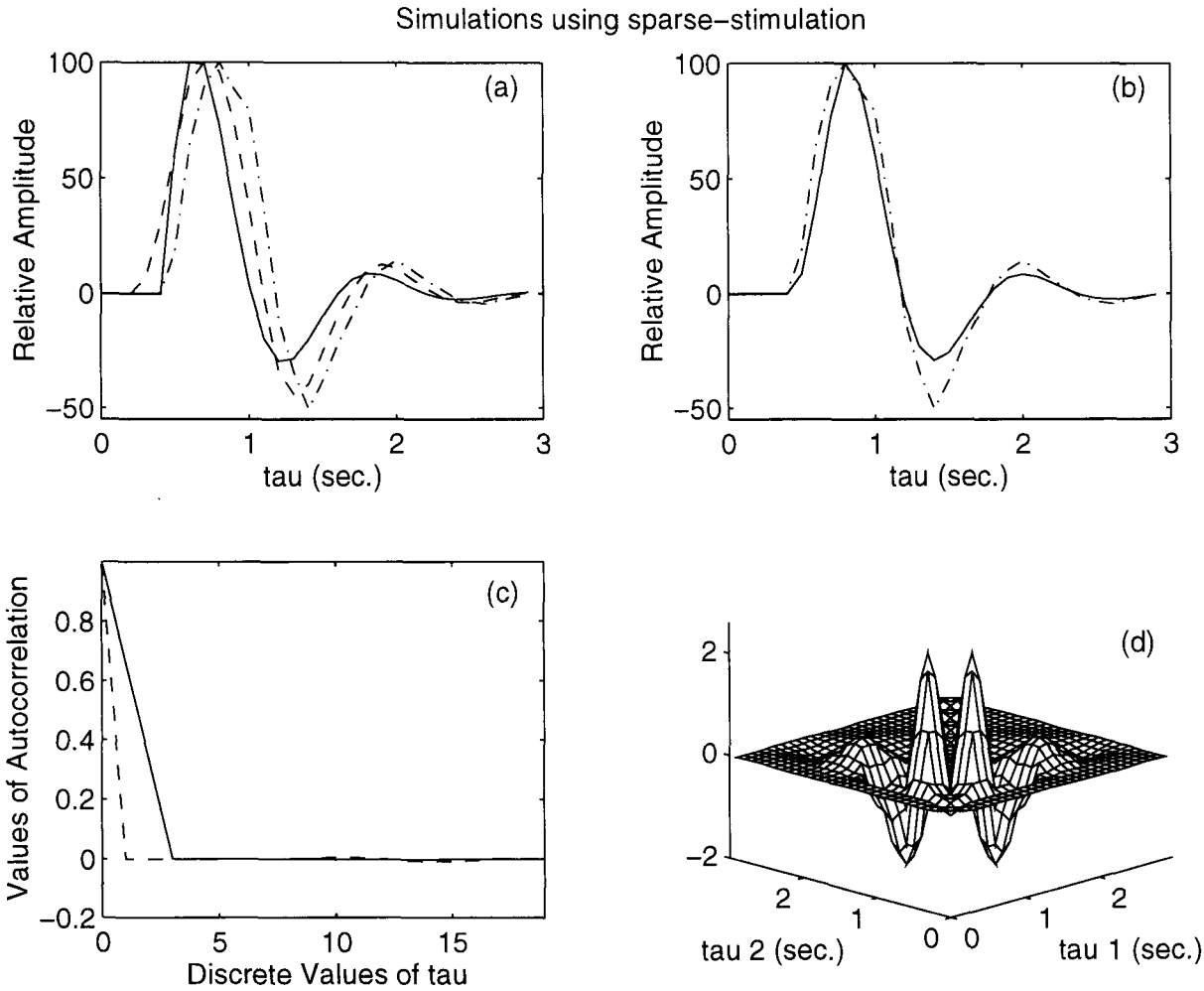


**FIGURE 7. Simulation results of sparse-stimulation (SSR = 3). (a) Solid line: the theoretical first-order kernel. Dashed line: the estimated first-order padded kernel. Dash-dotted line: the estimated first-order inserted kernel. (b) Solid line: the response of the true first-order kernel convolved with a zero-order hold ($T = 0.3$ s, replotted from Fig. 5c). Dash-dotted line: the estimated first-order inserted kernel. (c) Autocorrelations of the original binary m-sequence with length 1023 (dashed line, replotted form Fig. 1b) and the padded (by 3) binary m-sequence with length 3069 (solid line). (d) The estimated second-order binary padded kernel using Method 2 from the simulation of sparse-stimulation (SSR = 3).**

inserted method, this method does not permit us to obtain the entire second- and higher-order kernels. For example, let

$$\phi_{y_1xx}(\tau_1,\tau_2) = i_{\tau_1,\tau_2},\ \phi_{y_2xx}(\tau_1,\tau_2) = j_{\tau_1,\tau_2},\quad \text{and}$$
$$\phi_{y_3xx}(\tau_1,\tau_2) = k_{\tau_1,\tau_2},$$

where $\tau_1$ (and $\tau_2$) $= 0,1,2, \cdots$, and $\phi_{y_1xx}(\tau_1,\tau_2)$, $\phi_{y_2xx}(\tau_1,\tau_2)$, and $\phi_{y_3xx}(\tau_1,\tau_2)$ are the second-order subkernels by cross-correlating each of the three subfiles ($y_1(t)$, $y_2(t)$, and $y_3(t)$) with the same input file $x(t)$. Then, the second-order inserted kernel will be

$$k_2(\tau_1,\tau_2) =$$

$$\begin{bmatrix}
— & — & — & i_{1,0} & — & — & i_{2,0} & — & — & i_{3,0} & \cdots \\
— & — & — & — & j_{1,0} & — & — & j_{2,0} & — & — & \cdots \\
— & — & — & — & — & k_{1,0} & — & — & k_{2,0} & — & \cdots \\
i_{0,1} & — & — & — & — & — & i_{2,1} & — & — & i_{3,1} & \cdots \\
— & j_{0,1} & — & — & — & — & — & j_{2,1} & — & — & \cdots \\
— & — & k_{0,1} & — & — & — & — & — & k_{2,1} & — & \cdots \\
i_{0,2} & — & — & i_{1,2} & — & — & — & — & — & i_{3,2} & \cdots \\
— & j_{0,2} & — & — & j_{1,2} & — & — & — & — & — & \cdots \\
— & — & k_{0,2} & — & — & k_{1,2} & — & — & — & — & \cdots \\
i_{0,3} & — & — & i_{1,3} & — & — & i_{2,3} & — & — & — & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}, \tag{17}$$

where the symbols "—" indicate gaps. It is clear from Eq. 17 that we can only calculate reduced second-order kernel slices along the diagonal direction with the interval of $\Delta t_s$, which may cause problems in acquiring sufficient information about the nonlinear system under study. (See 44 for a detailed discussion on this issue.)

*The Padded Method.* To demonstrate the padded method, the same input ($x(t)$) (with a length of 1023) and output ($y(t)$) (with a length of 3069) data files used in the demonstration of the inserted method are used again. Instead of dividing the output file into three subfiles, we can pad the input file $x(t)$ (with length 1023; see Eq. 13) by 3, so that the input file will have a length of 3069, the same as the output files $y(t)$. That is, each datum in the original 1023 file was repeated two additional times:

$$x_p(t) = \underbrace{[1,1,1,-1,-1,-1,-1,-1,-1,1,1,1, \cdots]}_{3069}. \tag{18}$$

The resulting padded file $x_p(t)$ (length of 3069) will be the same as the data file sampled from the continuous m-sequence curve in Fig. 6b with a sampling interval of 0.1 s (the solid triangle series). Therefore, all of the system kernels can be obtained (with the sampling interval $\Delta t_r$) by cross-correlating the output data file $y(t)$ with the padded input file $x_p(t)$. The first- and second-order cross-correlations were calculated (using Method 2) and plotted in Fig. 7, a (dashed curve) and d, respectively.

The padded input file is no longer a real m-sequence, and thus the autocorrelation of the padded m-sequence will not be a $\delta$ function. The autocorrelations of the original m-sequence $x(t)$ (binary, 1023) and the padded m-sequence $x_p(t)$ (3069) are plotted in Fig. 7c as dashed and solid curves, respectively. The autocorrelation of the real m-sequence becomes zero at data point 2 (the discrete $\tau$ value of 1 in Fig. 7c), and the autocorrelation of the padded m-sequence (3069) becomes zero at data point 4 (the $\tau$ value of 3). In general, for an autocorrelation of an m-sequence padded by $l$ (=SSR), the first data value reaching zero will be at the data number $l + 1$ (*i.e.*, the base of the triangle-shaped autocorrelation will be increased when $l$ increases). This fact has also been noted by Klein (19). The imperfection of the autocorrelation of the padded m-sequence will cause deviations of the estimated kernels from the true kernels using the cross-correlation technique. Therefore, it is important to find a tractable way to recover the true system kernels from the filtering caused by the padded effect. In their seminal work (23), Lee and Schetzen discussed measurements of Wiener kernels with nonwhite (colored) noise inputs. The padded m-sequence can be considered as a nonwhite input filtered by an equivalent linear system (representing the padding effect) cascaded in front of the system. According to (23,25,39), if the power density spectrum of a nonwhite input is factorable, we can always obtain the transfer function of the cascaded equivalent linear system $\Phi(\omega)$. Then, similar to Eq. 12, the true first- and second-order kernels can be recovered as

$$K_1(\omega) = \frac{K_1'(\omega)}{\Phi(\omega)},\ \text{and}\ K_2(\omega_1,\omega_2) = \frac{K_2'(\omega_1,\omega_2)}{\Phi(\omega_1)\Phi(\omega_2)}. \tag{19}$$

(Note: The power density spectra of the nonwhite noise inputs with physical origin will always be factorable. The padded m-sequence may not have physical origin and thus their power spectra may not be factorable. Nevertheless, we may use a similar power spectrum that is factorable to approximate that of the padded m-sequence, or just simply conduct accurate simulations to estimate $\Phi(\omega)$ from the estimated and the true first-order kernels, as shown in Eq. 19.)

It is interesting that the triangle-shaped autocorrelation of the padded m-sequence (Fig. 7c) is the same as the impulse response function of a first-order hold. The associated transfer function is given by (32, pp. 525)

$$H(\omega) = \frac{1}{T}\left[\frac{\sin(\omega T/2)}{\omega/2}\right]^2, \tag{20}$$

where $T$ is equal to half of the triangle base of the autocorrelation, *i.e.*, $T$ = SSR × $\Delta t_r$. (For example, for the padded m-sequence autocorrelation in Fig. 7c, $T$ = 0.3 s.) We have used $H(\omega)$ as $\Phi(\omega)$ in Eq. 19 to recover the true system kernels. The results show that the true kernels can be recovered from the estimated padded kernels. These results will be examined further in the General Discussion.

In the padded method, there is a low-passing filtering effect caused by the zero-order hold, in addition to the padded effect $\Phi(\omega)$ (see Fig. 6b). Even though $\Delta t_s$ = 0.3 s, the discretized interval of the padded m-sequence is essentially $\Delta t_r$ (or 0.1 s in this case). The hold converts the discrete padded m-sequence to a continuous one with a duration of $\Delta t_r$. As discussed above, in practice, $\Delta t_r$ can be chosen to be short enough to cover the frequency bandwidth of the system. In this case, the filtering effect, caused by the hold, may be negligible compared with the padding effect. The estimated first-order padded kernel, using Method 2, is shifted to the right less than the first-order inserted kernel (compare the dashed curve (the padded kernel) with the dash-dotted curve (the inserted kernel) in Fig. 7a). That is, the padded kernel has less estimation deviation than the inserted kernel for a short m-sequence input. From our quantitative analysis based on simulation results, for SSR = 3 the PMSE (percentage minimum squared error) is 14.61% for the first-order inserted kernel and is 6.51% for the first-order padded kernel. For SSR = 6 the PMSE is 22.96% for the first-order inserted kernel and is 9.64% for the first-order padded kernel.

The second-order padded kernel using Method 2 was calculated and plotted in Fig. 7d. This figure indicates that all of the spurious bumps that occurred in Fig. 3c are no longer in the system memory area (system memory number $m$ = n × SSR = 30). Therefore, Eq. 16 appears to hold for the padded method using binary m-sequences. Intuitively, since $\Delta t_s$ of the m-sequence has been increased three (= SSR) times, the AFA related to the system memory will also be increased by 3. Accordingly, from Eq. 16, a practical criterion for selecting $n$ and $\Delta t_s$ is expressed as

$$n\Delta t_s \geqslant m\Delta t_r, \tag{21}$$

where $m\Delta t_r$ is the effective memory of the system under study.

In summary, both the inserted and padded methods are compared using the same simulated data set. The padded kernels may be easier to calculate or may have lower estimation errors for short m-sequences. It is important to note, as suggested by one of the reviewers, that the padded kernels may not contain more information than the inserted kernels, since the padding effect may be similar to a smoothing operation applied to the inserted method.

*Method 1 vs Method 2.* As shown in Figs. 1c and 1d, for cases where $\Delta t_s \simeq \Delta t_r$, the kernel estimated using Method 2 (inverse-repeat method) is more accurate than that using Method 1 (original Lee-Schetzen method). The estimated first-order padded kernels using sparse-stimulation for Methods 1 and 2 are plotted in Fig. 8a as dashed and dash-dotted curves, respectively. They are almost identical to each other (*i.e.*, they appear as a single curve). The estimated second-order padded kernel using Method 1 is plotted in Fig. 8b, which is also similar to the one in Fig. 7d using Method 2, except for the few spurious bumps occurring at the edge of the system memory ($m$ = 30). These anomalies may be caused by the odd, higher-order kernels (*e.g.*, the third-order). The results in Figs. 7d, 8a, and 8b indicate that within the anomaly-free area, method 2 has no advantage over Method 1 when using the sparse-stimulation technique. This suggests that in practice, we can use the simpler Method 1. This conclusion is corroborated by the Experimental Results in the next section. (Note that in Figs. 7d and 8b small bumps remain along the sides of the main diagonal, which may be caused by the discontinuity of the main diagonal slice with its neighboring slices.) A simulation with parameters similar to those in Figs. 4 and 6b was also conducted using a ternary m-sequence (728 length). The first- and second-order padded kernels (by 3) were calculated and plotted in Figs. 8c and 8d (the solid curve). The dashed curve in Fig. 8d is the first-order padded kernel using the binary m-sequence (1023 length) as the system input, which was previously plotted in Figs. 7a and 8a. The two curves in Fig. 8d are quite similar to each other, suggesting that the padded effect has the same impact on binary and ternary m-sequences. In Fig. 8c a few small spurious bumps occurred at about $\tau_1$ (or $\tau_2$) = 2 s (*i.e.*, $m$ = 20), suggesting that Eqs. 16 and 21 may also hold true for the padded method using ternary m-sequences (*i.e.*, AFA = $n$ × SSR = 6 × 3 = 18.)

## EXPERIMENTAL RESULTS OF VISUAL MEG STUDIES

Stimuli utilized in EEG and MEG studies have typically been classified as "transient" or "steady-state" (35, Section 1.3). Transient stimuli are presented at rates that allow the response to return to baseline levels before the presentation of the next stimulus. Steady-state stimuli are presented at a single frequency or rate that is typically much faster than rates utilized for transient stimulation. Consequently, steady-state stimulation offers certain advantages over transient stimulation in terms of speed of data acquisition and good signal-to-noise ratio. However, if the neural response is nonlinear and does not have an opportunity to return to baseline levels before the presentation of the next stimulus, the temporal dynamics of the

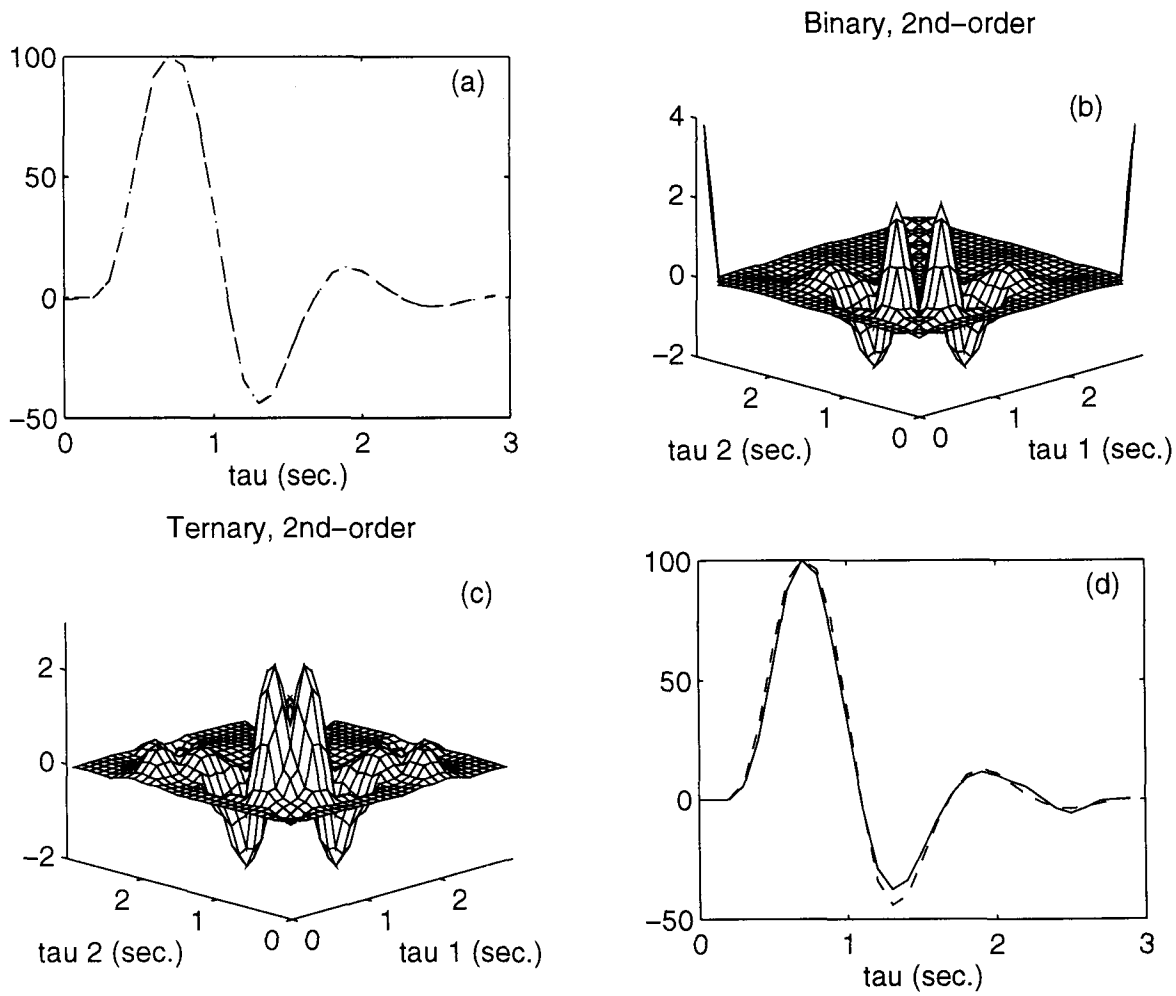


**FIGURE 8. Binary and ternary padded kernels. (a) The estimated first-order binary padded kernels using Methods 1 (dashed line) and 2 (dash-dotted line) from the simulation of sparse-stimulation. (b) The estimated second-order binary padded kernel using Method 1 from the simulation of sparse-stimulation. (c) The second-order ternary padded kernel using a ternary m-sequence with length 728 as the input. (d) Solid line: the estimated first-order ternary padded kernel. Dashed line: the estimated first-order binary padded kernel, which was previously plotted in Figs. 7a and 8a.**

response may become lost altogether (*i.e.*, the resulting composite response may contain components from several cycles of the stimulus). The m-sequence pseudorandom stimulus is a more general form of steady-state stimuli with broad-band frequencies. Its frequency bandwidth is from $f_s/N$ to $f_s/3$, where $f_s$ and $N$ are the sampling frequency and the length of the m-sequence, respectively. Therefore, theoretically, if the frequency bandwidth of the m-sequence stimulus is a superset of the frequency bandwidth of the system under study, the temporal dynamics of the system's response will not be lost.

A number of VEP studies have applied m-sequence stimuli to evoke electrical responses in humans (5,44). These studies, as discussed previously, used inserted sparse-stimulation techniques and relatively long m-sequences (*e.g.*, $2^{15} - 1$). They demonstrated that the sparse-stimulation technique can reduce second-order kernel estimation errors caused by anomalies. The simulation results presented in this paper suggest that the padded sparse-stimulation method can also aid in alleviating estimation errors caused by anomalies, even for short m-sequences. The primary goal of the experimental studies is to assess the utility of using the padded sparse-stimulation method and short m-sequence stimuli in visual evoked MEG studies. Three reasons for examining m-sequences as stimuli, in general, are: (a) m-sequences are statistically independent of (random or periodic) output noise and may therefore enhance the S/N ratio; (b) because different m-sequences are orthogonal to one another, responses to each stimulus component of simultaneous multi-input stimuli may be recovered, thereby allowing faster data acquisition; and (c) results from m-sequence studies, along with nonlinear systems analysis, may provide information about the system's nonlinear dynamics. Thus far, m-sequences have been shown to be effective for studying linear and nonlinear dynamics of the VEP using EEG mea-

surements (5,40,41,43,44). In this section, we present preliminary experimental MEG data using short m-sequence stimuli and padded sparse-stimulation techniques. The S/N enhancement is evaluated in this section by comparing the S/N ratio per unit time for both transient and m-sequence methods.

*Methods*

Evoked neuromagnetic fields were recorded with a BTi 7-channel SQUID-coupled gradiometer system in a magnetically shielded chamber. Three subjects with normal vision participated in this study. They viewed computer-generated visual displays projected (by a Sharp XG-2000U LCD projector, Japan) on a rear-projection screen via a system of mirrors. A circular achromatic patch of light, 1° in size, was located 3° from the point of fixation in the lower-right visual field. This stimulus was modulated at 30 Hz above and below the background mean luminance (0.85 $\mu W/cm^2$) by a short m-sequence. The highest (or lowest) intensity of the stimulus was at its center, with intensity varying radially as a sine function (0.25 cycles) from the central point; consequently, the edge of the stimulus had the same intensity as the background mean luminance (reducing the edge effect). The modulation depth (contrast) at the center of the stimulus was about 100%. The m-sequence length was 1023 (binary) for each trial, and thus the duration of each trial was $1023 \times \Delta t_s$ ($=34.1$ s). Subjects rested for 10s after each of 10 binary m-sequence trials. The MEG responses were digitized at 200 Hz. The sparse-stimulation ratio (SSR) was 6.667 ($=200/30$). In the previous section,we only discussed the situation where SSR was an integer number. In the case where SSR is not an integer number, portions of the original m-sequence data values will be repeated (padded) six times, whereas others will be repeated seven times, according to the occurrence of the input data samples (with interval $\Delta t_s$) related to the occurrence of the output data samples (with interval $\Delta t_r$). (Note that our data acquisition system can only digitize the response at discrete rates: 100 Hz, 200 Hz, 500 Hz, 1 kHz, *etc.*, which results in the use of a noninteger SSR (6.67). It would be better to use an integer SSR, if possible, to avoid some possible artifacts due to beating.)

For comparison, a transient stimulus (circular sinusoid) in which image intensity varied radially as a sine function from a central point (1,2), was presented at the same location as the m-sequence stimuli. The size of the stimulus was also equated (1°). Stimulus duration was 266 ms, with an average presentation rate of 1 Hz. Each block of trials contained 25 stimulus presentations. Subjects rested for about 10 s after each three blocks of trials. Data were digitized from 100 ms before stimulus onset to 500 ms after stimulus onset.

The seven sensors, spaced 2 cm apart, were arranged in a hexagonal array (sensor 1 was in the center, with the other six surrounding it). Responses were recorded from one dewar location (seven-sensor array), the location that produced a maximum response to the transient stimulus between 100–150-ms poststimulus. Once the location of the maximum response was determined, responses to m-sequences were also recorded. (For one subject (CA), we recorded responses evoked by transient and m-sequence stimuli from 13 different dewar locations to obtain neuromagnetic field distributions spanning occipital, temporal, and parietal cortex.)

*Results*

A comparison of MEG responses evoked by binary m-sequences and transient stimuli is presented in this section. In the transient studies, neuromagnetic responses evoked by the transient stimulus are shown for one subject (HWC) in Fig. 9. Averaged waveforms (105 individual neural responses) from each of the seven sensor locations are shown, along with a graph showing the digitized stimulus values (the lower right portion in Fig. 9). The physical unit of the responses is magnetic field intensity in femtoTesla (fT). In the m-sequence studies, the first- and second-order (the eighth subdiagonal slice) binary kernels, calculated by cross-correlation techniques from each trial using the padded method, are shown in the first and second columns of Fig. 10, respectively (only results from four sensors are shown). As discussed previously (25), the physical units for the first- and second-order kernels are, respectively,

$$\frac{\text{Unit}_{\text{response}}}{\text{Unit}_{\text{stimulus}}}, \quad \text{and} \quad \frac{\text{Unit}_{\text{response}}}{(\text{Unit}_{\text{stimulus}})^2}. \tag{22}$$

The physical units for the estimated first- and second-order kernels for the MEG data are 0.85 $fT/(\mu W/cm^2)$ and $0.85^2$ $fT/(\mu W/cm^2)^2$. The number 0.85 is the calibration factor, since we used the values $-1$ and $+1$ as the two states of the binary m-sequence for calculating the kernels.

As can be seen in Fig. 10 (columns 1 and 2, averages of 10 trials), the measured first-order cross-correlations are weak compared with the second-order cross-correlations (similar results were obtained for all three subjects), suggesting that the measured visual magnetic responses are strongly nonlinear for the present type (unpatterned) of m-sequence stimulus used. (It is worth noting that our recent results show that both strong first- and second-order kernels can be measured when using in-phase sinusoidal target stimuli modulated by m-sequences.) One way to evaluate the noise level for the cross-correlation results is to use a noncausal cross-correlation method (17). The noncausal cross-correlation was obtained by cross-correlating the stimulus and the

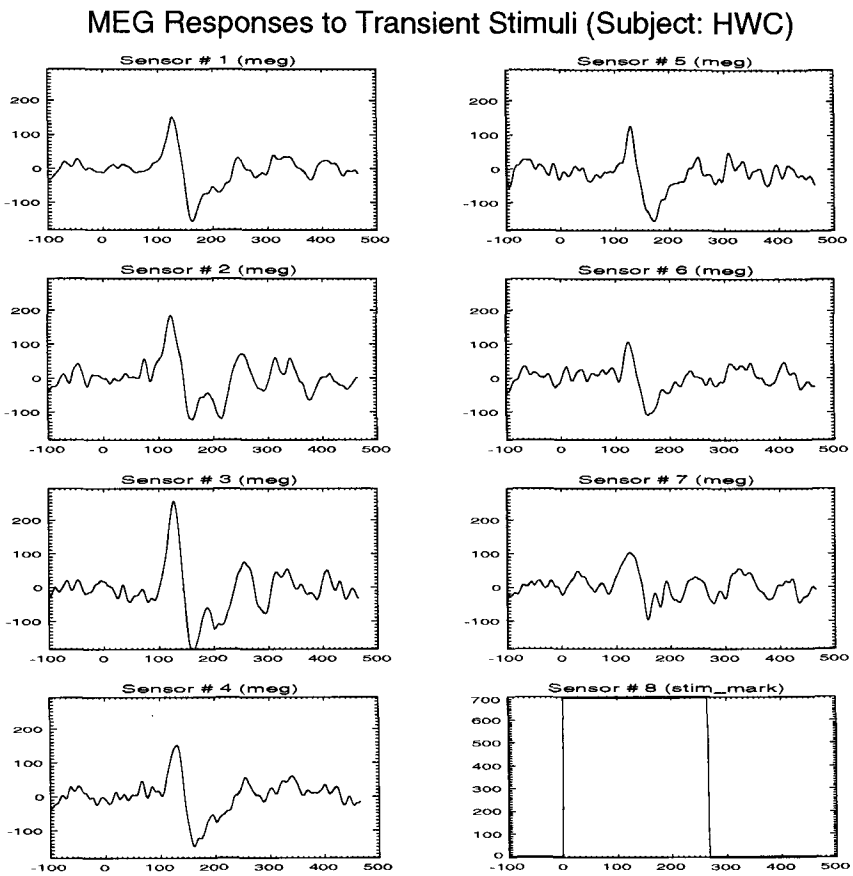


**FIGURE 9. Experimental results using a transient stimulus. Averaged neuromagnetic responses recorded from one subject using a seven-sensor array. The time course of the stimulus was digitized and is shown in the lower-right hand corner (from −100 ms to 500 ms, subject HWC).**

response by allowing the response to precede the stimulus, as shown in Fig. 10, for the duration from −100 ms to 0 ms. This duration is equivalent to the prestimulus period for the transient study shown in Fig. 9. Another way to evaluate the noise level is to conduct control experiments occluding the m-sequence and transient stimuli from the subject's view. Fig. 10 (column 3) shows results from such a control experiment for one subject (HWC). In the control experiment, the subject looked at the fixation point on a blank mean luminance screen (without the stimulus). All other variables were constant across studies. Fig. 10 (column 3) shows second-order (the eighth subdiagonal slice) binary kernels (averages of 10 trials) for the four sensors computed from the control data. Noise levels obtained from this method (occluded stimuli) were similar to those determined using the noncausal cross-correlation shown in Fig. 10 (columns 1 and 2, from −100 ms to 0). The control data shown in Fig. 10 (Column 3) also confirm that the response signals evoked by the m-sequence stimuli in column 2 of Fig. 10 cannot be attributed to equipment artifacts or environmental noise.

S/N ratios were calculated from m-sequence and transient responses using the ratio between the peak-to-peak (maximum-to-minimum) signal value and the standard deviation (STD) of the noise. For the noncausal cross-correlation method, the noise level was estimated using the data in the noncausal duration ($t < 0$), and both causal and noncausal periods were used for estimating the noise level for the control method. For example, for m-sequence stimuli, the STDs of the noise calculated from a single trial recorded by sensor 3 for two different intervals of time were 2.87 (column 4 of Fig. 10, from −100 ms to 0), and 2.95 (column 3 of Fig. 10, from −100 ms to 400 ms). For transient stimuli (Fig. 9), data in the prestimulus period (from −100 ms to 0) were used to estimate the noise level. For example, for one block of trials (35 repeated presentations), the STD across the prestimulus period was calculated (ensemble averaging). The resulting S/N ratios (average of 35 presentations, running time $T \approx 35$ s) were 7.77, 5.17, and 5.02 for the three subjects (HWC, WEN, and CA), respectively, whereas the S/N ratios calculated from m-sequence responses (one trial using a binary m-sequence with length 1023, running time $T = 34.1$ s) were 16.34, 17.52, and 10.46 for the same subjects using the noncausal method. Therefore, the S/N ratio from the m-sequence study was 2–3 times higher than the transient study, given similar lengths of total stimulus presentation time. Regan (35) reports typical S/N ratios for averaged VEPs ranging between 2 and 6. The results discussed above demonstrate that cross-correlation techniques can significantly increase the S/N ratio for MEG studies.

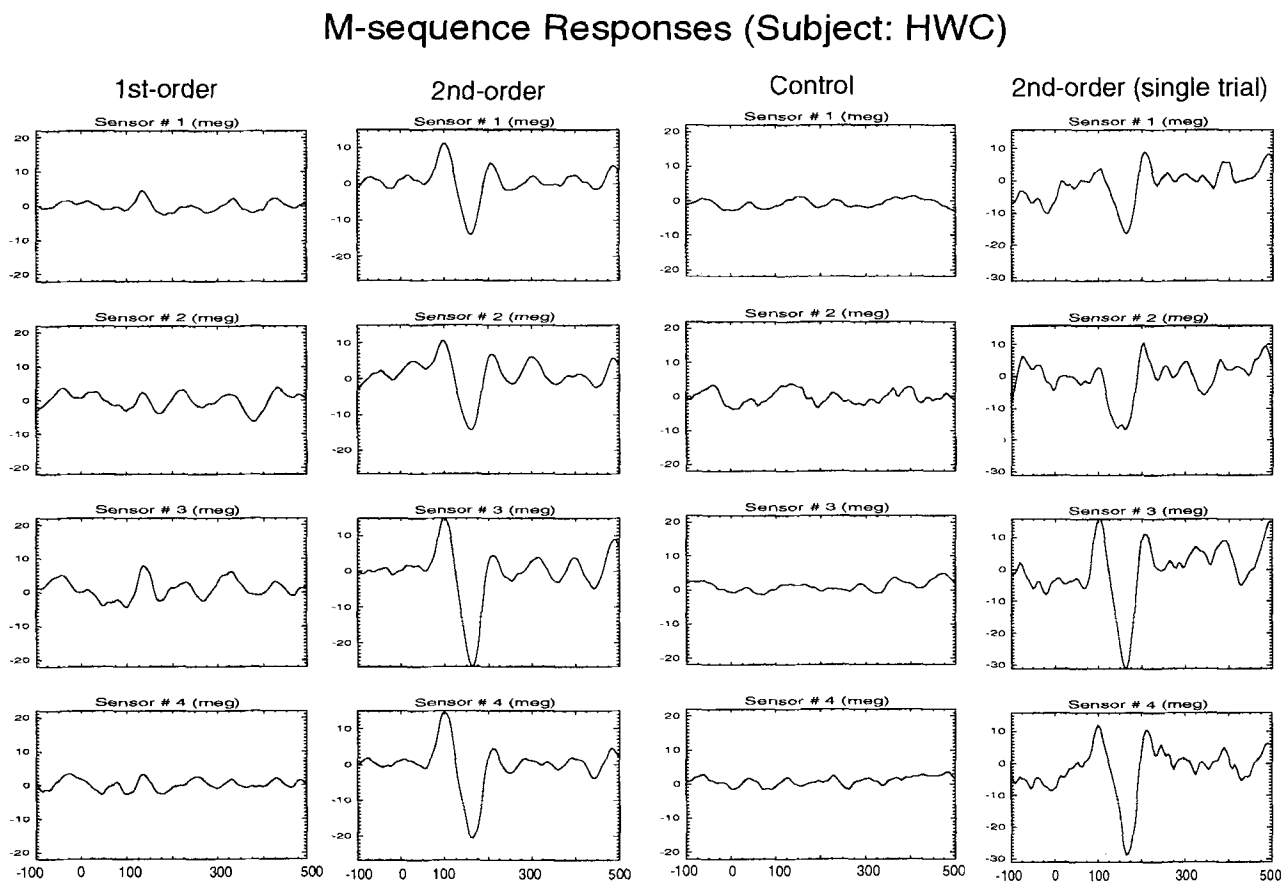


**FIGURE 10. Experimental results for subject HWC using m-sequence stimuli (padded method). Column 1: the first-order cross-correlations (average of 10 trials) between the neuromagnetic responses of the seven-sensor array and a short binary m-sequence (1023) stimulus (from −100 ms to 500 ms). Results from four sensors are shown. Column 2: the eighth subdiagonal slices of the second-order cross-correlations (average of 10 trials) for a short binary m-sequence (1023) stimulus (from −100 ms to 500 ms). Column 3: results from a control experiment evaluating noise levels. The eighth subdiagonal slices of the second-order cross-correlations (average of 10 trials). Column 4: the second-order cross-correlation slices for one of 10 experimental trials, the averaged data of which were shown in column 2.**

The calculated S/N ratios from this study were estimated from one m-sequence trial (column 4 of Fig. 10) to match total stimulus presentation time between m-sequence and transient studies. Clearly, S/N ratios will be higher for data averaged across 10 trials (column 2 of Fig. 10) than for one trial (column 4). But, the waveforms in column 4 from one of the individual trials were surprisingly similar to those in column 2, although they were noisier. Similar trends were noted for the nine other trials. In general, an average of five to six (repeated) trials has S/N ratios similar to those obtained from the average of 10 trials; therefore, we have been routinely acquiring six repeated trials for ongoing studies. Figure 11 shows the second-order (the eighth subdiagonal slice) binary kernels (averages of 10 trials, from −250 ms to 750 ms for all seven sensors) from another subject (CA). As discussed in the last section, the padded method allows one to obtain the entire second- and higher-order kernels. The entire second-order padded kernels for subject CA, obtained from sensors 4 and 5, are plotted in Fig. 12, a and b, respectively. According to Eq. 16, the anomaly-free area for this case will be AFA = 10 × 6.67 = 66.7 (×5 ms) (=333.5 ms). Figure 13 displays empirical neuromagnetic field distributions at a single instant in time for subject CA, evoked by transient (Fig. 13a) and m-sequence stimuli (Fig. 13b). (A time series of field distributions was constructed by measuring field amplitudes, relative to the prestimulus baseline, at 5-ms intervals across all sensor locations (7-sensor array × 13 dewar positions = 91 sensor locations).) Note that the lower portions of the two field distributions have similar features.

### *Discussion*

The experimental results show that relatively strong evoked neuromagnetic responses can be measured using padded sparse-stimulation techniques and short m-sequence stimuli. The time courses of the waveforms in Fig. 9 (or the cross-correlations in Figs. 10 and 11) across the seven different sensor locations are not proportional to one

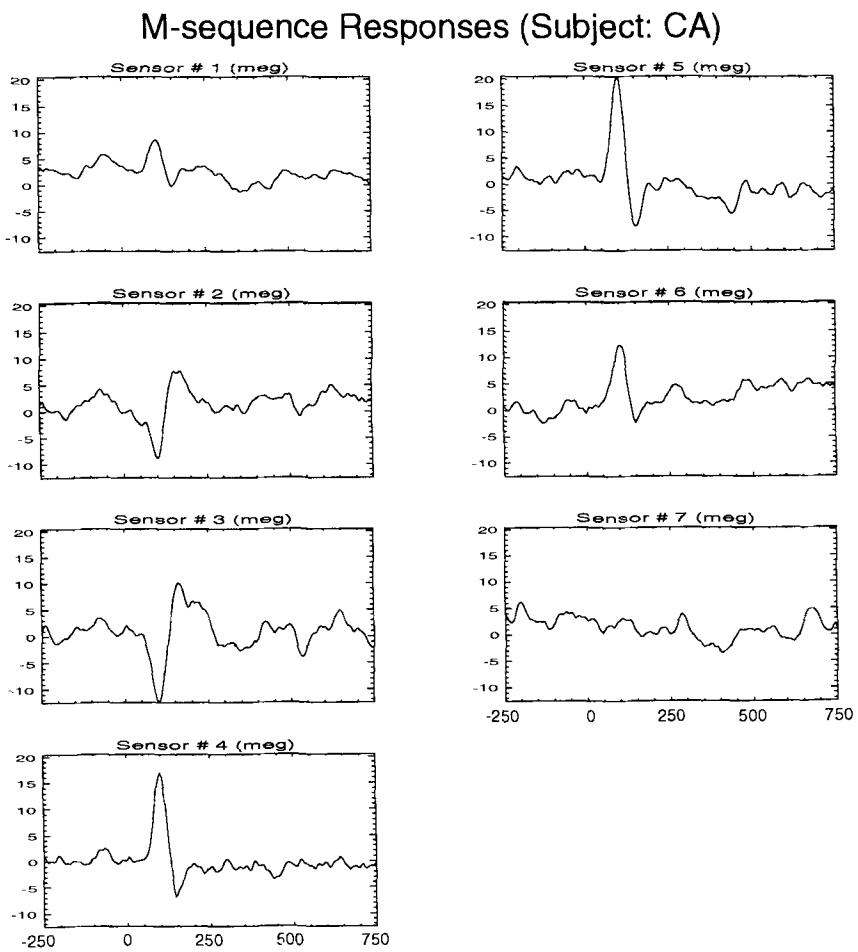


**FIGURE 11. Experimental results for subject CA using m-sequence stimuli (padded method). The eighth subdiagonal slices of the second-order cross-correlations (average of 10 trials) between the neuromagnetic responses of the seven-sensor array and a short binary m-sequence (1023) stimulus (from −250 ms to 750 ms). Subject CA.**

another, suggesting multiple sources. (Five sources were identified during the initial 80–165-ms interval poststimulus in previous studies using transient stimuli (1,2)). As shown in Figs. 9 and 10 (column 2), the amplitude distributions among the seven SQUID sensor recordings agree well between the studies (*i.e.*, transient and m-sequence) for subject HWC. For example, both studies showed the highest response amplitude in sensor 3. Similar results were obtained from the other two subjects. However, there are several discrepancies between the waveforms in Figs. 9 and 10. First, the maximum peaks in Fig. 9 occurred later than those in Fig. 10, presumably because of the long duration of the transient stimuli (266 ms). This shift is equivalent to the low-pass filtering effect caused by a zero-order hold (see Fig. 5c). Second, the onset latencies of the maximum peaks noted in Fig. 10 are shorter than those in Fig. 9, which was most likely caused by the low-pass filtering of the padding effect, as evidenced by the simulated results in Fig. 7a (see the dashed and solid curves in Fig. 7a). (Note: both low-pass filtering effects can also produce changes in peak widths, as shown in Figs. 5c and 7a.) Futhermore, differences in the waveforms (Fig. 9 *vs* 10) are also likely due to the fact that the waveforms of the transient responses (Fig. 9) contain higher-order nonlinearities, rather than the second-order kernel (the second-order nonlinearity) as in Fig. 10. Nevertheless, the low-pass filtering effects should not have a differential effect on the neuromagnetic field distributions evoked by m-sequence stimuli. In other words, the low-pass filtering effects may affect the temporal waveforms across sensors, but these effects should not have a differential affect on the spatial source parameters (*i.e.*, location and orientation) derived from the field distributions, using nonlinear minimization techniques. As shown in Fig. 13 (a and b), the lower portions of the two field distributions, in the occipital region of visual cortex, are quite similar to each other. Some discrepancies between the field distributions, however, are expected, since the second-order kernels calculated from m-sequence stimuli should represent only a portion of the total distribution evoked by transient stimuli.

The results from the MEG experimental section indicate that the S/N ratio of responses evoked by m-sequences can be 2–3 times higher than responses evoked by transient stimuli. The reasons for this enhanced S/N ratio are twofold:

1. Since m-sequences, generalized steady-state stimuli, do not allow the response to return to baseline levels, more

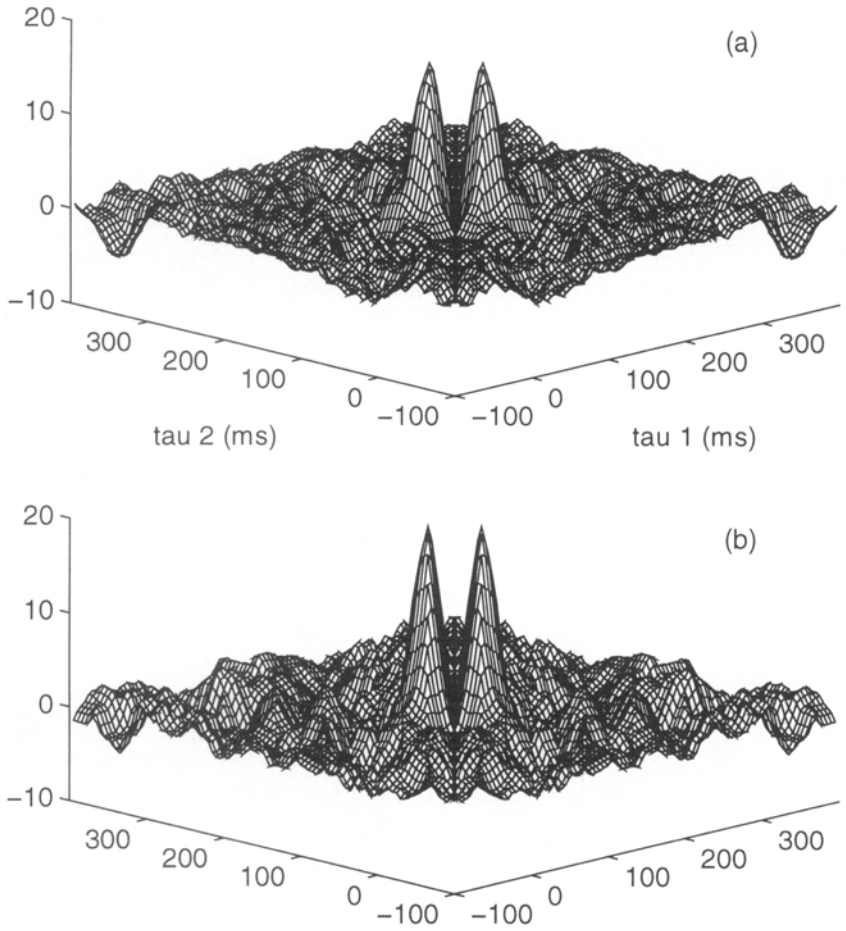


**FIGURE 12. Second-order cross-correlations for subject CA. (a) The second-order cross-correlation from sensor 4. (b) The second-order cross-correlation from sensor 5.**

stimuli may be presented during an interval equivalent to that used in acquiring responses to transient stimuli. For example, responses were acquired for 34.1 s for a 1023 binary m-sequence. For an equivalent acquisition interval for transient stimuli, only 9.31 s ($=0.266 \times 35$ for a block of 35 trials) of this interval is actually used in data acquisition.

2. In addition, m-sequences are statistically independent of output noise; consequently, the noise is filtered out by the cross-correlation process.

In EEG and MEG studies, a portion of the output noise may be associated with periodic spontaneous responses in addition to random noise. The orthogonal property of m-sequences may help separate the evoked component from the spontaneous component of the response (*e.g.*, the evoked response may be estimated by convoluting the measured system kernels with the input m-sequence, and the spontaneous response may be obtained by subtracting the estimated evoked response from the system response). We are currently investigating this property.

## GENERAL DISCUSSION

One useful property of m-sequences is that different m-sequences are orthogonal to one another. This property allows for the separation of response components related to each individual input evoked by the simultaneous multi-input stimulation, thereby permitting faster data acquisition times. The greater the number of stimulus presentations (multiple inputs), the greater the savings in experimental time. For example, the VEP study (5) used 64 orthogonal m-sequences simultaneously at different visual field locations to stimulate the retina. Much longer experimental times would be required if the retina were stimulated sequentially. Another advantage of using simultaneous stimulation is that one can measure the cross-kernels (also called mutual-kernels in 44), which reflect the nonlinear interaction between the different input pathways (9,10,25).

The traditional Lee-Schetzen approach (23) uses Gaussian white noise as the input signal; therefore, long record lengths of the input are required to approximate the true white noise. The primary disadvantages of the Lee-Schetzen approach in practice are twofold: (a) experimental studies with longer data records will require much longer times to calculate the higher-order kernels; and (b) this approach cannot be applied to physiological systems and biological preparations, which can be considered stationary only within certain short periods of time. In the

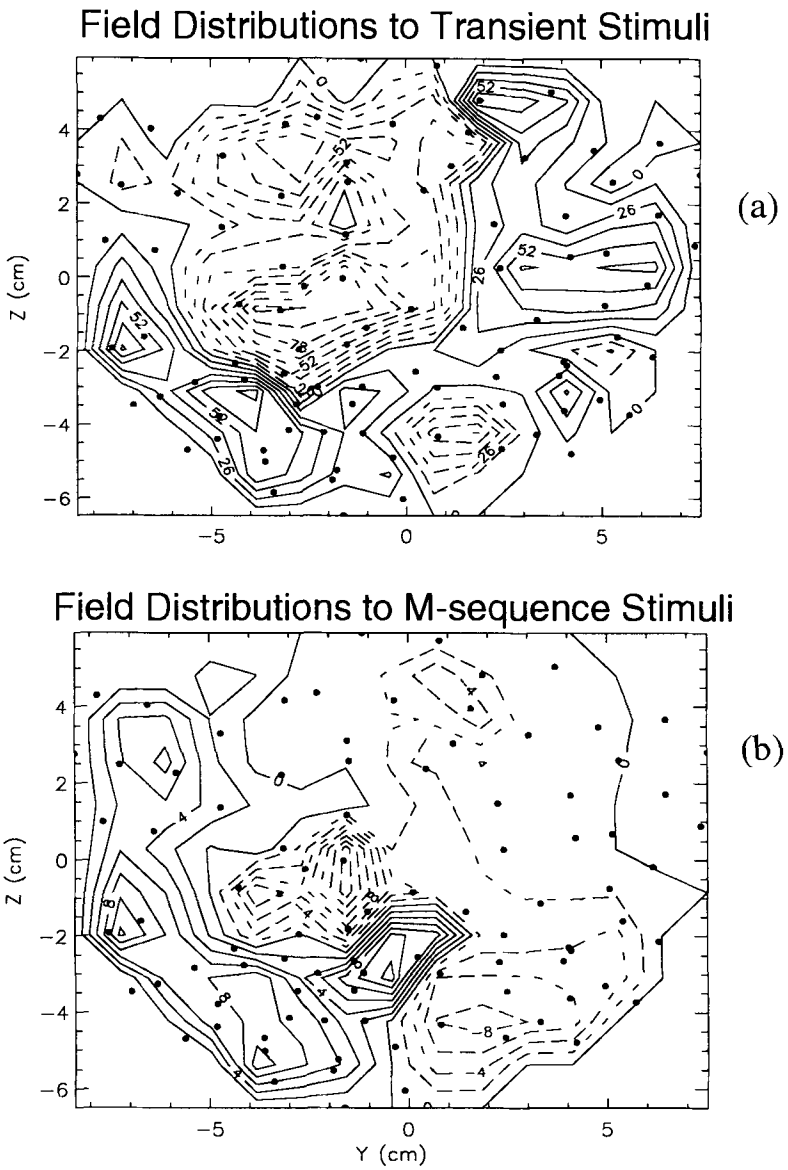


**FIGURE 13. Neuromagnetic field distributions for subject CA. These maps are *Y–Z* surface projections viewed from the back of the head. The midline is at *y*= 0; negative *y* values represent the left hemisphere. The positive (solid) contour curves represent flux emerging from the head and negative (dashed) curves represent re-entering flux. The black dots represent sensor locations. (a) Responses (of latency 115 ms) to transient stimuli. (b) Second-order cross-correlations (latency 105 ms, eighth subdiagonal slices, average of six trails) calculated from responses to binary m-sequence stimuli. The latencies were chosen at around the maximum peak of the waveforms. Differences in the peak values (10–20 ms) were caused by the low-pass filtering effects, as examined in the text (Discussion).**

present study, short m-sequences were used as input signals, which significantly reduced the experimental time (*e.g.*, the presentation of a 1023 binary m-sequence at a rate of 30 Hz took 34.1 s, which can be further reduced to one-half when presented at a rate of 60 Hz). The combination of short m-sequence techniques with short recording periods may also allow one to study mild nonstationary nonlinear systems by measuring the system kernels at different periods to investigate the time-variant properties of the kernels (see 26 for a discussion of time-variant kernels). Furthermore, the 20–30-s recording period for short m-sequences is quite suitable for EEG and MEG studies using human subjects where short trial periods are preferred to avoid potential artifacts caused by eye blink, fatigue, and body (muscle) movements, *etc.* If long m-sequences (*e.g.*, $2^{15} - 1 = 32{,}767$) are used, as in (5), then the long m-sequence should be parceled into smaller segments (*e.g.*, 16), which must be linked together later using linear interpolation techniques for kernel calculations (5). Another advantage for using short m-sequences is that one can repeat the same m-sequence several times but still keep the experiment within a reasonable duration, as demonstrated in the experiments in this paper. The averaged kernels may provide additional statistical measures for evaluating the experimental data (*e.g.*, the standard deviation of the averaged data). It is worth noting that for a multi-input system with a large number of inputs (*e.g.*, 64 inputs were used in (5)), one must use long m-sequences to ensure the orthogonality between inputs, as discussed in detail in the next paragraph.

From an experimental standpoint, the shorter the stimulus sequence, the better. However, short m-sequences may not reveal all of the system properties, if the system under study has long memory and many higher-order ker-

nels as discussed in (5,19,43–45). Although a long m-sequence was used for the VEP study (5), all 64 visual inputs used the same m-sequence with a time delay of 512 ($=2^{15}/64$) between each input. Therefore, the equivalent sequence length is 512 for each input. As discussed previously (5,19,44), for the memory range of typical VEP responses, a length of 512 is adequate for measuring the first- and second-order self kernel slices as well as some cross-kernel slices if the third- and higher-order kernels are negligible. For two m-sequences (1023 length) with a time delay of 512, the equivalent sequence length is 512 for each input. The simulation shown in Fig. 4 and the experimental results using single inputs (Figs. 10–12) used an m-sequence with length of 1023 for the single input, and thus the equivalent sequence length is also 1023. Since the memory range of MEG responses is similar to that of EEG responses, a binary m-sequence with length 1023 should be sufficient for measuring the first- and second-order kernel slices from MEG responses. Note that Sutter (44) presented a practical method for calculating the size of the anomaly-free area (average "distance" between kernel slices; (44)) for binary m-sequences. For example, consider a system with a first-order kernel and a second-order kernel with 20 slices, *i.e.*, a total of 21 kernel slices for this system. For a binary m-sequence with 1023 stimulus epochs, the average "distance" is about 1023/21 = 48. However, when the system has a higher-order nonlinearity such as the third and fourth order, the size of the AFA will be much smaller, as evidenced by the simulated results in Fig. 3, c and d. (The system simulated in Fig. 3, c and d, has a half-square nonlinearity that contains a strong third-order kernel (10).) In general, the third-order nonlinearity may contaminate the first-order kernel, and the second- and fourth-order nonlinearities may contaminate the second-order kernel (6,44). Nevertheless, when using sparse-stimulation, as shown in Eq. 16 and Fig. 7d, the size of the AFA can be increased by a factor of SSR.

The simulation results shown in Fig. 8b indicate strong spurious spikes occurring outside the anomaly-free area, caused by the third- and higher-order kernel slices. However, such spikes were not evident in the MEG responses evoked by stimuli located 3° from the fovea. For example, as shown in Fig. 11, there were no strong bumps outside the anomaly-free area (from 333.5 ms to 750 ms). These results suggest that the measured MEG responses may be dominated by the second-order nonlinearity, whereas the third and higher orders may be weak and negligible.

As discussed previously, sparse-stimulation ($\Delta t_s > \Delta t_r$) has several advantages over the usual approach ($\Delta t_s = \Delta t_r$). For example, $\Delta t_r$ can be selected independently of $\Delta t_s$; a higher power level of the input may be selected by increasing $\Delta t_s$; and most importantly, the anomalies may be reduced. As discussed previously, there are two methods for calculating the system kernels from sparse-stimulation data: the inserted method and the padded method. The estimated kernels derived from both methods deviate from the true kernel values because of the low-pass filtering (blurring) of the zero-order hold and the padding effects as shown in Fig. 7a. As discussed above (The Padded Method), the measurement errors caused by the low-pass filtering effects are low when the SSR is small (*e.g.*, the PMSE is 6.51% for the first-order padded kernel when SSR = 3). Therefore, depending on the required estimation accuracy for the system under study, the estimated kernels may be used as approximations of the true kernels when SSR is small. For a large SSR, the true kernels should be recovered from the estimated kernels using inverse-filtering (deblurring) techniques, as shown in Eqs. 12 and 19 (also see the next paragraph for a more detailed discussion on this issue). Nevertheless, in some cases, even for a large SSR, the blurred inserted and padded kernels may still be used as substitutes for the true kernels. For example, the blurred inserted and padded kernels could theoretically be used without bias to localize different neural sources for EEG/MEG studies, since the blurring effects should have an equal affect on the generators of the signal.

Although theoretically the true kernels can be recovered by using Eqs. 12 and 19, the development of tractable inverse-filtering techniques presents a great challenge, since the deblurring process can introduce significant noise. Our preliminary results show that the true first-order kernel can be recovered from the estimated first-order padded kernel by using both Eqs. 12 and 19, *i.e.*, $K_1(\omega) = K_1'(\omega)/[HD(\omega)\Phi(\omega)]$, where $HD(\omega)$ and $\Phi(\omega)$ are described in Eqs. 10 and 20, respectively. The duration of the hold (Eq. 10) is $T = \Delta t_r$. For example, based on the estimated first-order padded kernel of a simulation using the inverse-filtering technique, the PMSE was reduced from 6.51% to 1.64% for an SSR = 3, the PMSE was reduced from 9.64% to 2.36% for an SSR = 6, and the PMSE was reduced from 16.05% to 3.10% for an SSR = 10. However, the algorithms did not work as well for the second-order binary kernels, since the gap along the main diagonal slice of the binary kernel introduces significant high-frequency noise. Further work is needed for developing tractable ways to deblur higher-order kernels.

Many applications of the Volterra-Wiener approach have been devoted to studying nonlinear dynamics of physiological and biological systems using invasive single neuron electrophysiological recording techniques. For example, the first-order kernels have been used to study color processing in lateral geniculate nucleus (36) and to study spatial information processing in cortical simple cells (9,18). The second-order kernels have been associated with nonlinear contributions to spatial information

processing and motion selectivity by retinal ganglion cells (13) and cortical complex cells (15,16), and have been associated with binocular filtering by complex cells (31). However, the invasive recording techniques are mainly applied in animal studies, not human studies. We hope that the Volterra-Wiener approach can be successfully applied to studying the nonlinear dynamics of human neural systems using noninvasive EEG and MEG techniques. However, there are important differences between the single-unit studies and the EEG/MEG studies. In single-unit studies, the location of the recording neuron in cortex is known. But with EEG/MEG studies, one is dealing with a more ambiguous problem, since the signals represent an unknown composition of neural sources (generators), weighted by their distance, orientation, and temporal properties, *etc.* We are continuing to investigate the value of nonlinear systems analysis for noninvasive EEG and MEG recordings.

## REFERENCES


1. Aine, C. J., S. Supek, and J. S. George. Temporal dynamics of visual-evoked neuromagnetic sources: Effects of stimulus parameters and selective attention. *Int. J. Neurosci.* 80:79–104, 1995.
2. Aine, C. J., S. Supek, J. S. George, D. Ranken, J. Lewine, J. Sanders, E. Best, W. Tiee, E. R. Flynn, and C. C. Wood. Retinotopic organization of human visual cortex: Departures from the classical model. *Cereb. Cortex* 6:354–361, 1996.
3. Barker, H. A., and T. Pradisthayon. High-order autocorrelation functions of pseudorandom signals based on m sequences. *Proc. IEE.* 117:1857–1863, 1970.
4. Barker, H. A., S. N. Obidegwu, and T. Pradisthayon. Performance of antisymmetric pseudorandom signals in the measurement of 2nd-order Volterra kernels by cross-correlation. *Proc. IEE* 119:353–362, 1972.
5. Baseler, H. A., E. E. Sutter, S. A. Klein, and T. Carney. The topography of visual invoked response properties across the visual field. *EEG Clin. Neurophysiol.* 90:65–81, 1994.
6. Benardete, E. A., and J. D. Victor. An extension of the m-sequence technique for the analysis of multi-input nonlinear systems. In: Advanced Methods of Physiological System Modeling, Vol. 3, edited by V. Z. Marmarelis. Plenum Press, New York, 1994, pp. 87–110.
7. Boyd, S., Y. S. Tang, and L. O. Chua. Measuring Volterra kernels. *IEEE Trans. Circuits Syst.* 30:571–577, 1983.
8. Chen, H.-W., L. D. Jacobson, and J. P. Gaska. Structural classification of multi-input nonlinear systems. *Biol. Cybern.* 63:341–357, 1990.
9. Chen, H.-W., L. D. Jacobson, J. P. Gaska, and D. A. Pollen. Cross-correlation analyses of nonlinear systems with spatiotemporal inputs. *IEEE Trans. Biomed. Eng.* 40:1102–1113, 1993.
10. Chen, H.-W. Modeling and identification of parallel nonlinear systems: Structural classification and parameter estimation methods. *Proc. IEEE* 83:39–66, 1995.
11. Chua, L. O., and Y. Liao. Measuring Volterra kernels (II). *Int. J. Circuit Theor. Appl.* 17:151–190, 1989.
12. Chua, L. O., and Y. Liao. Measuring Volterra kernels III: How to estimate the highest significant order. *Int. J. Circuit Theor. Appl.* 19:189–209, 1991.
13. Citron, M. C., and V. Z. Marmarelis. Applications of minimum-order Wiener modeling to retinal ganglion cell spatiotemporal dynamics. *Biol. Cybern.* 57:241–247, 1987.
14. Davies, W. D. T. System Identification for Self-Adaptive Control. New York: Wiley-Interscience, 1970, 394 pp.
15. Emerson, R. C., J. R. Bergen, and E. H. Adelson. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Res.* 32:203–218, 1992.
16. Gaska, J. P., L. D. Jacobson, H.-W. Chen, and D. A. Pollen. Space-time spectra of complex cell filters in the Macaque monkey: a comparison of results obtained with pseudo-white noise and grating stimuli. *Vis. Neurosci.* 11:805–821, 1994.
17. Jacobson, L. D., J. P. Gaska, H.-W. Chen, and D. A. Pollen. Structural testing of multi-input linear-nonlinear cascade models for cells in macaque striate cortex. *Vision Res.* 33:609–626, 1993.
18. Jones, J. P., and L. A. Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58:1187–1211, 1987.
19. Klein, S. Optimizing the estimation of nonlinear kernels, In: Nonlinear Vision, edited by R. B. Pinter and B.Nabet (eds.). Boca Raton: CRC Press, 1992, pp. 109–170.
20. Korenberg, M. J. Identification of biological cascades of linear and static nonlinear systems. *Proc. Midwest Symp. Circuit Theor.* 18.2:1–9, 1973.
21. Korenberg, M. J., S. B. Bruder, and P. J. McIlroy. Exact orthogonal kernel estimation from finite data records: Extending Wiener's identification of nonlinear systems. *Ann. Biomed. Eng.* 16:201–214, 1988.
22. Korenberg, M. J. Parallel cascade identification and kernel estimation for nonlinear systems. *Ann. Biomed. Eng.* 19:429–455, 1991.
23. Lee, Y. W., and M. Schetzen. Measurement of the Wiener kernels of a nonlinear system by cross-correlation. *Int. J. Control.* 2:237–254, 1965.
24. Marmarelis, V. Z. Random versus pseudorandom test signals in nonlinear-system identification. *Proc. IEE* 125:425–428, 1978.
25. Marmarelis, P. Z., and V. Z. Marmarelis. Analysis of Physiological Systems: The White Noise Approach. New York: Plenum Press, 1978, 504 pp.
26. Marmarelis, V. Z. Practicable identification of nonstationary nonlinear systems. *Proc IEE* 128:211–214, 1981.
27. Marmarelis, V. Z. Identification of nonlinear biological systems using Laguerre expansions of kernels. *Ann. Biomed. Eng.* 21:573–589, 1993.
28. Marmarelis, V. Z., K. H. Chon, Y. M. Chen, D. J. Marsh, and H.-H. Holstein-Rathlou. Nonlinear analysis of renal autoregulation under broadband forcing conditions. *Ann. Biomed. Eng.* 21:591–603, 1993.
29. Naka, K.-I., H. M. Sakai, and N. Ishii. Generation and transformation of second-order nonlinearity in catfish retina. *Ann. Biomed. Eng.* 16:53–64, 1988.
30. Ogata, K. Modern Control Engineering, Englewood Cliffs, NJ: Prentice-Hall, 1970, 960 pp.
31. Ohzawa, I., R. D. Freeman, and G. C. DeAngelis. Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science* 249:1037–1041, 1990.
32. Oppenheim, A. V., A. S. Willsky, and I. T. Young. Sig-

nals and Systems. Englewood Cliffs, NJ: Prentice-Hall, 1983.
33. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. Numerical Recipes in C. Cambridge: Cambridge University, 1988, 1992, pp. 274–328, 347–608.
34. Ream, N. Nonlinear identification using inverse-repeat m-sequences. *Proc. IEE* 117:213–218, 1970.
35. Regan, D. Human Brain Electrophysiology—Evoked Potential and Evoked Magnetic Fields in Science and Medicine. New York: Elsevier Science Publishing Co., 1989, pp. 1–482.
36. Reid, R. C., and R. M. Shapley. Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature* 356:716–718, 1992.
37. Rugh, W. J. Nonlinear System Theory: The Volterra/ Wiener Approach. Baltimore: Johns Hopkins University Press, 1981.
38. Sandberg, I. W. The mathematical foundations of associated expansions for mildly nonlinear systems. *IEEE Trans. Circuits Syst.* 30:441–445, 1983.
39. Schetzen, M. The Volterra and Wiener Theories of Nonlinear Systems. New York: Wiley, 1980.
40. Srebro, R., and W. W. Wright. Pseudorandom sequences in the study of evoked potentials. *Ann. N.Y. Acad. Sci. 388:98–112, 1982.*
41. Srebro, R. An analysis of the VEP to luminance modulation and of its nonlinearity. *Vision Res* 32:1393–1404, 1992.
42. Sutter, E. A revised conception of visual receptive fields based upon pseudorandom spatiotemporal pattern stimuli. In: Proceedings of the First Symposium on Testing and Identification of Nonlinear Systems, edited by G. D. McCann and V. Z. Marmarelis. Pasadena: CIT, 1975, pp. 353–365.
43. Sutter, E. E. A practical nonstochastic approach to nonlinear time-domain analysis. In: Advanced Methods of Physiological System Modeling, edited by V. Z. Marmarelis. Los Angeles: USC, 1987, pp. 303–315.
44. Sutter, E. A deterministic approach to nonlinear systems analysis. In: Nonlinear Vision, edited by R. B. Pinter and B. Nabet. Boca Raton: CRC Press, 1992, pp. 171–220.
45. Victor, J. D. Nonlinear systems analysis in vision: Overview of kernel methods. In:Nonlinear Vision, edited by R. B. Pinter and B. Nabet. Boca Raton: CRC Press, 1992, pp. 1–37.
46. Victor, J. D., and B. W. Knight. Nonlinear analysis with an arbitrary stimulus ensemble. *Q. Appl. Math.* 74:113–136, 1979.
47. Volterra, V. Theory of Functionals and of Integral and Integro-differential Equations. New York: Dover Publications, 1959.
48. Wiener, N. Nonlinear Problems in Random Theory. New York: The Technology Press of MIT and Wiley, 1958.