An automated algorithm for the generation of dynamically reconstructed trajectories

C. Komalapriya,[1,2,a)] M. C. Romano,[3,4] M. Thiel,[3] N. Marwan,[2] J. Kurths,[3,5,2] I. Z. Kiss,[6] and J. L. Hudson[7]
[1]*Interdisciplinary Centre for Dynamics of Complex Systems, University of Potsdam, 14476 Potsdam, Germany*
[2]*Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany*
[3]*Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom*
[4]*Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, United Kingdom*
[5]*Institute of Physics, Humboldt University-Berlin, 12489 Berlin, Germany*
[6]*Department of Chemistry, Saint Louis University, Missouri 63103, USA*
[7]*Department of Chemical Engineering, University of Virginia, Virginia 22904, USA*

(Received 15 September 2009; accepted 6 December 2009; published online 3 February 2010)

The lack of long enough data sets is a major problem in the study of many real world systems. As it has been recently shown [C. Komalapriya, M. Thiel, M. C. Romano, N. Marwan, U. Schwarz, and J. Kurths, Phys. Rev. E **78**, 066217 (2008)], this problem can be overcome in the case of ergodic systems if an ensemble of short trajectories is available, from which dynamically reconstructed trajectories can be generated. However, this method has some disadvantages which hinder its applicability, such as the need for estimation of optimal parameters. Here, we propose a substantially improved algorithm that overcomes the problems encountered by the former one, allowing its automatic application. Furthermore, we show that the new algorithm not only reproduces the short term but also the long term dynamics of the system under study, in contrast to the former algorithm. To exemplify the potential of the new algorithm, we apply it to experimental data from electrochemical oscillators and also to analyze the well-known problem of transient chaotic trajectories. © *2010 American Institute of Physics*. [doi:10.1063/1.3279680]

Many data sets that are measured in laboratories or that are observed by monitoring natural systems are either short or contain gaps. In such cases, deciphering the characteristics of the underlying system by conventional time series analysis techniques might not be possible, as many of these techniques require temporally continuous long data sets. An algorithm, based on the concept of recurrence, has been proposed to overcome the problem of short data sets or missing values. The original method generates long artificial phase space trajectories—called dynamically reconstructed trajectories (DRTs)—from a collection of short data sets that have been observed at different instances of time. In this paper, we present a substantially improved algorithm that operates with a single parameter. The new algorithm not only performs better than the former one, but also reproduces the short and the long term dynamics of the underlying system closely. Furthermore, we demonstrate how to automate the algorithm. The applicability of the proposed automated algorithm is validated with experimental chemical oscillator data. Moreover, we apply it to numerically characterize the properties of chaotic saddles by generating an artificial long trajectory (LT) from an ensemble of transient chaotic trajectories.

I. INTRODUCTION

Many natural systems are difficult to monitor in a continuous way. Here restrictions mainly arise due to existing physical or experimental limitations, technical or data storage problems, and the cost factor.[1,2] On the other hand, there are systems which exhibit interesting behavior for a very brief period of time. A good example of the latter case is transient chaos, where the trajectories starting from different initial conditions stay in the vicinity of a nonattracting chaotic saddle for a short time, before escaping to a final attractor.[3–7] Nevertheless, it is often possible to observe an ensemble of short trajectories by starting the experiments or the simulations at different, usually random, initial conditions or by recording observations discontinuously. However, many conventional time series analysis techniques are not suitable for these cases as they usually work only if long and continuous sets of data are available.

We have recently shown[8] that it is in fact possible to circumvent this problem by utilizing the basic concepts of chaos, Poincaré recurrences, and ergodic theory.[9–11] The key idea is to piece together the short trajectories of the ensemble in a dynamically appropriate way to overcome the discontinuities or the gaps present in the data. The process results in a LT, which by construction replicates the dynamics of the underlying system, and is hence called DRT. Any time series analysis method can then be applied to this *synthetic long continuous* trajectory in order to extract the required

a)Electronic mail: komala@agnld.uni-potsdam.de.

information about the underlying system. This solves a highly relevant problem in data analysis.

In order to apply the algorithm proposed in Ref. 8, we have to find optimal values for two parameters (threshold ε and jumping probability p), so that the dynamics of the underlying systems is correctly reproduced. This can be a rather tedious and computationally expensive task. In this paper, we present an optimized algorithm that overcomes this problem. The new approach has only one parameter and as a result, the computational effort is significantly reduced. This allows us, moreover, to automate the algorithm, making it even more useful for the analysis of real world data. Furthermore, we show that the long term dynamics is considerably better reproduced than with the former algorithm.

The organization of this paper is as follows. In Sec. II we present the improved algorithm that generates the long dynamical replicants. In Sec. III we list the measures that are used to assess the quality of the generated DRTs. In Sec. IV we perform sensitivity studies with respect to the original and the new algorithms depending on (i) the single algorithmic parameter, (ii) the characteristics of the given ensemble of short trajectories, and (iii) observational noise. We then show how to completely automate the algorithm in Sec. V. The validity of the new algorithm is demonstrated by applying it to experimental data, as well as to a problem of transient chaos in Secs. VI and VII, respectively. The results are summarized in Sec. VIII.

II. DYNAMICS AND DISCONTINUITIES—DRT ALGORITHM

Assume that we have an ensemble of M d-dimensional short trajectories from an ergodic system, each of length N, i.e., $\vec{\tilde{x}}_i^j \in \mathbb{R}^d$, where $i=1,\ldots,N$ and $j=1,\ldots,M$. If the number of short trajectories M is large enough, then the ensemble as a whole will embody, although discontinuously, the complete attracting set of the underlying system. From this ensemble, it is possible to compute reasonable estimates of fractal measures, such as the correlation dimension.[12–14] However, an accurate estimation of dynamically invariant measures such as Lyapunov exponents, entropies,[10,14] or other recently proposed measures of complexity[15] is in general not possible with this temporally discontinuous data.

Therefore, we introduced an algorithm to reconstruct a long synthetic trajectory, called DRT, from a given ensemble of short trajectories.[8] A DRT imitates the dynamical properties of a LT of the underlying system and thus it can be used to estimate dynamical measures of an underlying system. The main idea of this algorithm is based on the fundamental recurrence property of the ergodic systems. Recurring states are used to bridge the gaps between different sections to generate a long synthetic trajectory. The algorithm builds a DRT by jumping with probability p from a point within a short trajectory to the future of one of its neighbors in the phase space, which is present within the same or in a different short trajectory.

The proposed algorithm, thus, requires two parameters to generate a DRT: (i) a threshold ε that determines the recurring system states and (ii) a parameter p, called jumping probability, which triggers the algorithm to jump at appropriate points. Generally, ε should be small enough to avoid large errors due to the jumps and large enough to avoid recursive use of certain segments of short trajectories. The parameter p should be, on the one hand, small enough to avoid frequent jumps and utilize as much information as possible from a single short trajectory and, on the other hand, large enough to prevent redundant use of certain parts of short trajectories. To determine the optimal values of ε and p, it is necessary to study the errors intrinsic to DRTs depending on both ε and p. However, this is a tedious task when the data set under investigation is rather large or when one has many ensembles of short data sets from different systems for comparative studies.

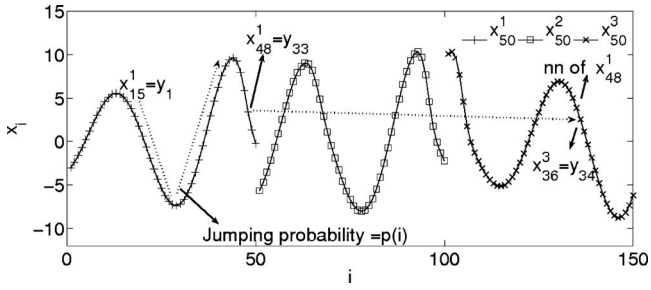

FIG. 1. The DRT algorithm: x_{50}^1, x_{50}^2, and x_{50}^3 represent the x-component of three short trajectories of length $N=50$ that are assembled and concatenated one after another. In this example, the point y_1 is the randomly chosen first point of the DRT belonging to the x^1. Since $\lim_{i\to 1} p(i)\approx 0$ and $\lim_{i\to N} p(i)\to 1$, the algorithm stays in the same short trajectory until x_{48}^1, before making a jump to the future of one of the nearest neighbors of x_{48}^1.

Therefore, it is desirable to reduce the number of algorithmic parameters. In order to do this, we modify the parameter p from being a constant throughout the short trajectory to a function $p(i)$, where i denotes the position within a short trajectory, i.e., $i=1,\ldots,N$. The function $p(i)$ is called jumping probability function (JPF). The JPF $p(i)$ is chosen to increase monotonously between 0 and 1 within a short segment, i.e., $\lim_{i\to 1} p(i)=0$ and $\lim_{i\to N} p(i)=1$. This reduces the errors in the generated DRTs that arise due to too frequent jumps. In other words, rather than jumping from every point of a single short trajectory with the same probability p, the algorithm is made to jump with a low probability at the beginning of the short trajectory, and with a higher probability toward the end of the short trajectory. Numerical studies conducted for three JPFs, namely, $e^{-(N-i)}$, $e^{-(N-i)^2/N}$, and $N^{-(N-i)/N}$, show that the exponential function of the type

$$p(i) = e^{-(N-i)^2/N}, \quad i = 1,\ldots,N \tag{1}$$

is a good choice for the reconstruction process. In comparison to others, the above JPF [Eq. (1)] performed better in terms of certain characteristic measures (to be described in Sec. III) and, hence, we will illustrate the results of the new algorithm by using this $p(i)$.

For a given ensemble of short trajectories, $\vec{\tilde{x}}_i^j$ with $i=1,\ldots,N$ and $j=1,\ldots,M$, the new algorithm to generate a DRT consists of the following steps (Fig. 1).

(1) Concatenate the ensemble of short trajectories $\vec{\tilde{x}}_i^j$ with $i=1,\ldots,N$ and $j=1,\ldots,M$ to generate $\vec{x}_k$, where $k=1,\ldots,L$ and $L=N\times M$.

(2) Determine the set of all neighbors of every point of $\vec{x}_k$ for a given threshold ε, i.e., compute the set of neighbors $\{\vec{x}_q \in B(\vec{x}_k)\}$, where $B(\vec{x}_k)$ is the neighborhood of $\vec{x}_k$.
(3) The first point $\vec{y}_1$ of the DRT is chosen randomly from the ensemble, i.e., $\vec{y}_1 = \vec{x}_l$ with $1 \le l \le L$.
(4) The next point of the DRT is either $\vec{x}_{l+1}$ with probability $1-p(l)$ or the future $\vec{x}_{q+1}$ of a randomly chosen neighbor of $\vec{x}_l$ with probability $p(l)$.
(5) The last step is repeated until we get a DRT of a desired length L_D.

If at some point of the DRT generation process, the algorithm reaches the end point of a short trajectory that does not have any neighbors, then the DRT generation process is restarted. However, if the number T of trials necessary to generate a DRT of the desired length exceeds a critical value, say $T=20\,000$, then the process is aborted assuming that the chosen threshold ε is not appropriate to generate a DRT from the given set of short trajectories.

Note that the new algorithm has just one parameter. Thus, with the modified algorithm it is enough to investigate only the effect of ε on the quality of the DRTs. This is one of the crucial advantages of the new algorithm over the former one. In Sec. IV we show that the new algorithm not only generates better DRTs, but is also successful in generating synthetic LTs for a wide range of thresholds. Additionally, we demonstrate that the new algorithm is more robust to noise compared with the formerly published one. Results of the numerical studies, which have been carried out with respect to the algorithmic parameter ε and the experimentally set values of the ensemble, namely, M and N, clearly illustrate these facts (Sec. IV).

Automatizing the process of the DRT generation is now a rather straightforward procedure. Automatization can be achieved by systematically varying the threshold ε within a certain range. A further better procedure to automate the algorithm is to generate DRTs by first defining the neighborhood of every phase space point in terms of a normalized measure called recurrence rate (RR) (see Sec. V and also Ref. 15), and then systematically varying it between 0 and 1. Like this, we reduce the arbitrariness in defining the threshold range. Before presenting the results of the numerical analysis in Sec. IV, we first enlist and describe the measures that we use in our subsequent studies to assess the quality of the DRTs and the performance of the algorithm.

III. CHARACTERIZING MEASURES

The quality of the DRTs is analyzed in terms of their dynamical properties, first by using some model systems. The generated DRTs are compared with LTs of the underlying system, obtained by integrating or iterating the system equations. The linear and the nonlinear measures that are used to compare the quality of the DRTs with that of the LTs are (i) the autocorrelation function c_τ,[14] (ii) the mutual information function I_τ,[16] (iii) the mean diagonal line length of the recurrence plot (RP) $\bar{D}$,[15] and (iv) the Rényi entropy of second order K_2.[12,17] The autocorrelation function c_τ reflects the linear correlations of a signal at lag τ. Its nonlinear generalization is the time delayed mutual information I_τ The mean diagonal line length $\bar{D}$ is a complexity measure estimated from the RP (Ref. 15) of a given signal; it measures the deterministic nature of the underlying system. Unlike these measures, K_2 is a dynamically invariant measure. It is a lower bound of the sum of all positive Lyapunov exponents, and hence it quantifies the predictability of a given system.

The quality tests are performed by first generating 100 realizations of DRTs (each of length L_D) from an ensemble of M short trajectories, each of length N. The above measures are then computed for each of these 100 DRTs and are compared with that of 100 realizations of the LTs (which are also of length L_D) from the same underlying system. In the case of c_τ and I_τ, we estimate the mean relative error as follows:

$$R_{c/I} = \frac{1}{\tau_{\max}} \sum_{\tau=1}^{\tau=\tau_{\max}} \frac{|\mu_\tau - \mu'_\tau|}{\mu'_\tau}, \tag{2}$$

where μ_τ represents the mean of the absolute value of the autocorrelation/mutual information function at lag τ for the ensemble of DRTs and μ'_τ that for the LTs. In the case of the mean diagonal line $\bar{D}$ of the RP and the Rényi entropy K_2, we compute the relative error as follows:

$$R_{\bar{D}/K_2} = \frac{|\mu - \mu'|}{\mu'}. \tag{3}$$

Here, μ and μ' represent the mean value of $\bar{D}/K_2$ estimated from the ensemble of DRTs and LTs, respectively.

There are two fundamental factors that might cause deviations between the dynamics of the DRTs and the LTs: (i) the error due to jumping, in other words, the impact of ε, and (ii) the redundancy problem that arises due to the repeated use of certain parts of short trajectories. To properly quantify the deviations in the dynamics of the DRTs originating from (ii), we define a new measure called redundancy factor (r_F). Let $\tilde{\vec{y}}_j \in \{\vec{y}_i | i=1,\ldots,L_D\}$ be the subset of redundantly used (used more than once) points of a DRT with $j=1,\ldots,L'_D$ and $L'_D < L_D$. Then r_F is defined as follows:

$$r_F = \frac{\sum_{j=1}^{L'_D} \nu(\tilde{\vec{y}}_j)}{L'_D}, \tag{4}$$

where $\nu(\tilde{\vec{y}}_j)$ denotes the total number of times the point $\tilde{\vec{y}}_j$ has been used in the reconstructed DRT. When the DRT has no redundantly used points, then $\nu(\tilde{\vec{y}}_j)=0\,\forall\, j$, and hence $r_F=0$. On the other hand, if the DRT is generated by repeatedly using a few phase space points, then $r_F \to L_D/2$. The mean redundancy factor $\langle r_F \rangle$ estimated from an ensemble of DRTs generated for a particular threshold will serve as a coarse estimator of the redundancy in the generated realizations of DRTs.

The DRT algorithm relies on the chance that a point with at least few neighbors will be chosen as a jumping point. The successful reconstruction of a DRT will also depend upon the probability that the neighbors of the chosen jumping point have a suitable future to jump to. These issues become critical when we are dealing with a smaller ensemble of short trajectories or when using smaller thresholds. Thus, often for smaller ensembles and for certain (usually smaller) thresh-

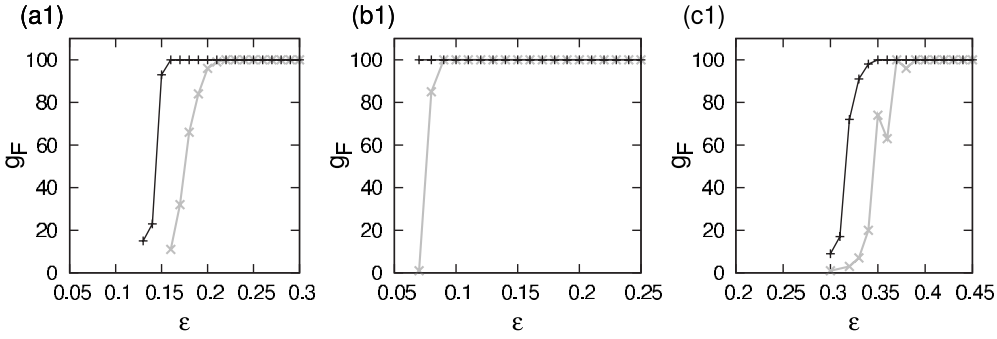

FIG. 2. The DRT generation factor g_F in dependence on the threshold ε, corresponding to ensembles (a1) $N=50$, $M=200$; (a2) $N=50$, $M=1000$; and (c1) $N=10$, $M=1000$. The black lines (with plus signs) and the gray lines (with cross signs) correspond to the DRTs that are obtained using the modified algorithm and the formerly published algorithm, respectively. The minimum thresholds (ε) from which the former algorithm generates the desired 100 realizations of DRTs (i.e., $g_F=100$) are: 0.220 (a1), 0.090 (b1), and 0.370 (c1), respectively. The same values for the new algorithm are as follows: 0.160 (a1), 0.070 (b1), and 0.350 (c1).

olds ε, the algorithm either fails to generate the desired 100 realizations of DRTs or it generates DRTs with higher values of redundancy (i.e., a crucial short trajectory is missing, and hence the algorithm tends to repetitively use certain available information).

The latter problem can be rather easily identified with the help of the redundancy measure $\langle r_F \rangle$. In order to take the former effect into account, we introduce another measure called the DRT generation factor g_F. It is defined as the total number of DRTs generated from a given ensemble of short trajectories at a particular threshold ε. If ε is chosen appropriately for the given ensemble, then $g_F=100$. Otherwise, the generation of DRTs is not always successful and we obtain $g_F<100$. Thus, the DRT generation factor g_F establishes the performance of the algorithm for a given threshold.

IV. SENSITIVITY STUDIES

Based on the measures established in Sec. III, we now evaluate the performance of the former and the improved algorithm. Note that the previous algorithm corresponds to the following JPF:

$$p(i) = \begin{cases} p & \text{if } 1 \le i \le N-1 \\ 1 & \text{if } i = N. \end{cases} \tag{5}$$

In contrast to Eq. (1), the jumping probability in Eq. (5) is set to be a constant for all but the last point of the short segment. In Secs. IV A–IV C, we show the results of the former algorithm with $p=0.05$, since this value has been shown to generate rather good quality DRTs.[8] It is worth mentioning that we have also compared the original algorithm with the new one for other values of p. However, since the outcome of the comparison was qualitatively the same, only the results for $p=0.05$ are presented in this paper.

The sensitivity studies are performed by considering the chaotic Rössler oscillator[18]

$$\dot{x} = -y - z, \quad \dot{y} = x + 0.2y, \quad \dot{z} = 0.2 + (x - 5.7)z, \tag{6}$$

and systematically varying the threshold ε, the length N, and the number M of short trajectories. We also investigate the robustness of the two algorithms with respect to noise. We carried out similar investigations by using some other prototypical dynamical systems like the Hénon map[19] the Rössler oscillator in the nonphase coherent chaotic funnel regime,[20] or the Lorenz oscillator.[21] In spite of the substantial topological differences of the attractors of these systems,[22] the results obtained from the sensitivity studies are qualitatively the same.

A. Sensitivity studies with respect to ε, N, and M

Integrating the Rössler system [Eq. (6)] with an integration step of 0.01 and a sampling rate of 20, we produce three ensembles of short trajectories: (i) $N=50$ and $M=200$, (ii) $N=50$ and $M=1000$, and (iii) $N=10$ and $M=1000$. While the first ensemble is used to investigate the influence of the threshold ε, the second and the third ensembles are utilized to study the influence of the number M and the length N of the short trajectories. We generate ensembles of DRTs (each of length $L_D=5000$), corresponding to the original and the improved algorithm, from all the three ensembles for a range of different thresholds ($0.01 \le \varepsilon \le 3.0$).

In all three cases, the new algorithm generates 100 realizations of the DRTs ahead of the original algorithm (see Fig. 2). This indicates that the new algorithm has a better jumping criterion when compared with the original one, and hence promotes the reconstruction process at smaller values of the threshold. Note that when the number of short trajectories of the ensemble is rather small, the new algorithm generates 100 realizations of DRTs for much smaller values of the threshold in comparison to the former one. As we will see later in Sec. IV B, this helps substantially to reproduce the long term dynamics of the system. The new algorithm will therefore be of great use for practical applications, since obtaining a sufficient number of short trajectories from experiments is not always possible.

The mean redundancy factor $\langle r_F \rangle$ calculated for both algorithms is shown in Figs. 3(a1)–(c1). The values of $\langle r_F \rangle$ are generally higher at smaller thresholds for both algorithms because for very small values of the threshold ε the number of neighbors to which the algorithm can jump is very low. As a result, the algorithm tends to use certain parts of the short trajectories recursively for the reconstruction process. The mean redundancy factor $\langle r_F \rangle$ also depends on the ensemble

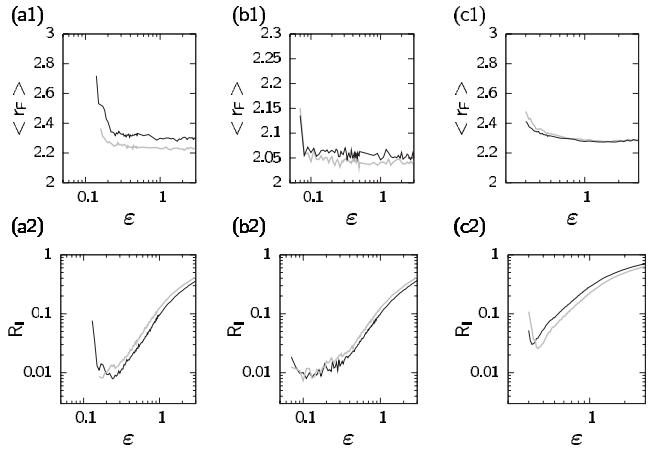

FIG. 3. The DRT generation factor g_F and the error in the mutual information function (R_I) in dependence of the threshold ε. The plots correspond to the DRTs generated from ensembles (a1, a2) $N=50$; $M=200$, (b1, b2) $N=50$; $M=1000$, and (c1, c2) $N=10$; $M=1000$, respectively. The black and the gray lines represent the DRTs obtained using the new algorithm and the former algorithm, respectively.

size L. The larger the ensemble, the larger is the number of neighbors for a fixed ε, and hence the smaller is the mean redundancy factor [compare Fig. 3(b1) with Figs. 3(a1) and (c1)].

If the length N of the short trajectories is large, then the mean redundancy factor of the trajectories generated with the new algorithm is higher compared with the former one [Figs. 3(a1) and (b1)]. This is because in contrast to the original algorithm, the modified algorithm is designed to jump toward the end of a short trajectory, which in turn causes the algorithm to redundantly use certain short trajectories for reconstruction. If N is small, both algorithms are forced to jump often, and as a result, both algorithms spend less time on a single short trajectory [Fig. 3(c1)]. Hence, when N is small, the estimates of $\langle r_F \rangle$ are rather similar. However, according to the nonlinear measures, e.g., mutual information I_τ, the opposite is true. Numerical studies show that in general, higher values of redundancy are correlated with smaller errors in the nonlinear measures like I_τ [Figs. 3(a2)–(c2)]. This is because with the modified algorithm, the average time spent in a single short trajectory is higher than with the original algorithm. To conclude, we can say that the performance of the algorithms in terms of redundancy factor and error measures depends upon the short trajectory length N.

From the results of relative error in I_τ [Figs. 3(a2)–(c2)] we can also conclude that (i) regardless of the algorithm used, the general tendency is that the error measure R_I increases with ε due to the larger jumps made by the algorithm for higher values of ε (this trend is also observed for the error measures R_c and $R_{\bar{D}}$); (ii) there exists, however, an intermediate range of ε for which the error R_I in the mutual information is minimal. This is because the DRTs generated for very low values of ε deviate from their original LTs due to the effect of redundancy [see Figs. 3(a1)–(c1)]. Thus, in terms of I_τ, both algorithms tend to generate better DRTs for intermediate values of ε.

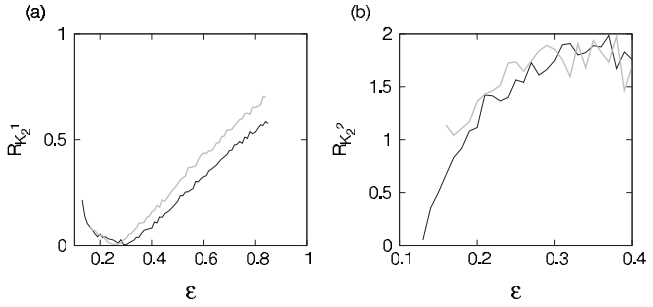

FIG. 4. Error in the (a) short and the (b) long time scale dynamics of the Rössler oscillator in dependence of the threshold ε and with respect to the modified (black line) and the original algorithm (gray line).

B. Reproduction of short and long term dynamics

In Sec. IV A, we have discussed the performance of the DRT algorithm in terms of the measure I_τ. However, for rigorous theoretical applications, invariant measures, such as Lyapunov exponents or entropies, are usually preferred. Estimation of these measures requires an extensive amount of continuous data sets which are often not available. In order to critically compare the performance of the original and new algorithm, we now analyze the dynamically invariant measure—Rényi entropy of second order K_2. Here we investigate how well the DRTs generated by both algorithms reproduce the short and the long term dynamics of the Rössler oscillator by estimating two different K_2 values, K_2^1 and K_2^2, respectively.[17,23]

We consider DRTs generated using both the improved and the previously published algorithms from an ensemble of short trajectories with $N=50$ and $M=200$. Figure 4 shows the error estimates in the short ($R_{K_2^1}$) and the long ($R_{K_2^2}$) term dynamics of the two algorithms and for various values of ε [Eq. (3)]. For both K_2^1 and K_2^2, the modified algorithm performs better than the former algorithm for almost all values of the threshold ε. Moreover, similar to the error in the mutual information function [Fig. 3(a2)], the error in the short term dynamics has a minimum in both algorithms at a threshold of about $\varepsilon=0.25$ [Fig. 4(a)]. Again, the cause for the occurrence of this minimum is the redundant use of certain parts of short trajectories for very low values of ε. In contrast, the error in the long term dynamics of the DRTs increases monotonously with the threshold ε [Fig. 4(b)], indicating that moderately higher values of redundancy do not affect the reproduction of the long term dynamics.

C. Robustness with respect to noise

In this section we investigate the influence of observational noise on the DRT algorithm, since data contamination by noise is inevitable in most real world systems. We expect that noise will have an effect on the process of reconstruction of DRTs, as noise may considerably change the neighborhood of a phase space point.[24] Hence, we analyze the robustness of the algorithm by applying it to ensembles of noisy short trajectories.

In order to demonstrate the performance of the original and the new algorithm in the presence of noise, we consider an ensemble of short trajectories from the Rössler oscillator

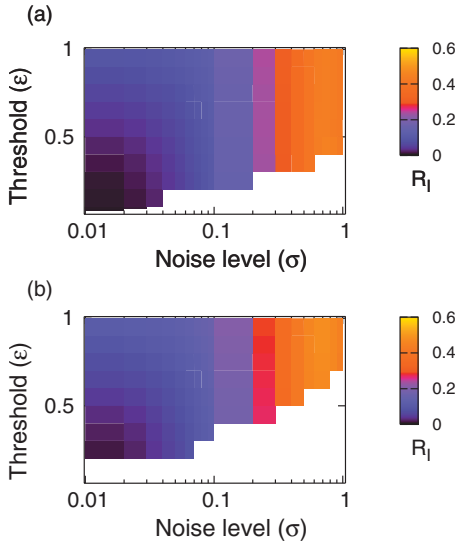

FIG. 5. (Color) Estimates of the error R_I in the mutual information function calculated with respect to the JPFs (a) modified algorithm and (b) original algorithm in the $\sigma-\varepsilon$ parameter space. The blue regions of the plots represent the parameter values that correspond to the lower error and the yellow regions to that having higher error in R_I. The white regions of the plots correspond to the parameter values for which $g_F=0$.

[Eq. (6)] with $N=50$ and $M=1000$. Gaussian white noise with standard deviation $S_{\text{noise}}=\sigma S_j$ is added to each of the three components of the Rössler short trajectories. Here, S_j is the standard deviation of the jth component of the concatenated trajectory and σ is the noise level. Ensembles of DRTs, each of length $L_D=5000$, were generated from this noise corrupted ensemble by using both algorithms for a range of thresholds ($0.06<\varepsilon<1.0$). The measures described in Sec. III were estimated from the ensembles of DRTs. Note that the relative errors were computed by comparing the DRTs with an ensemble of noisy LTs of the Rössler oscillator (i.e., LTs contaminated with the same level of noise σ). The above steps were repeated for different noise levels by varying σ between 0.01 and 1.0.

Figure 5 shows the estimates of R_I with respect to the original [Fig. 5(b)] and the improved [Fig. 5(a)] algorithms in the σ versus ε parameter space. The white regions of the plots correspond to the parameter values for which the algorithm did not manage to generate any DRTs, i.e., where $g_F=0$. Generally, as the level of noise in the signal increases, the range of the thresholds for which the algorithm can generate DRTs decreases. As we can see from Fig. 5, the white region is significantly smaller for the modified algorithm when compared with the formerly published one, indicating that the new algorithm performs better in the presence of noise.

To summarize this section, we can state that with the proposed changes, we achieve a modest improvement in the performance of the algorithm in terms of the relative error measures (e.g., with respect to R_c, R_I, $R_{\bar{D}}$, $R_{K_2^1}$, and $R_{K_2^2}$). The improved algorithm, however, promotes generation of DRTs at smaller thresholds than the original algorithm, especially when the number of short trajectories constituting the ensemble is rather low. Although the DRTs generated with the improved algorithm have in general a higher redundancy when compared with the original algorithm, their dynamics are closer to the original LTs. Furthermore, analysis with respect to noise suggests that the modified algorithm performs substantially better. Taking also into consideration that by means of the improved algorithm we can also reduce the number of parameters to just one, we can state that the modified algorithm is superior to the former one.

In the next step, we utilize a method that is widely used in the field of synchronization analysis[15] to automate the process of DRT generation and thereby simplify its application.

V. AUTOMATION OF THE ALGORITHM BY USING FIXED AMOUNT OF NEAREST NEIGHBORS

In the improved algorithm, the only parameter that has to be determined is the threshold ε. Practically it is necessary to vary ε in a certain range and study then its influence on a dynamical measure, such as the mean diagonal line $\bar{D}$, in order to determine an optimal value of ε for the reconstruction. The interval of ε over which the study can be conducted might obviously differ from one system to another due to the different phase space diameter of the attractor under investigation. Therefore, the range of thresholds that need to be analyzed is usually estimated by some "rules of thumb" given in the literature.[15,25]

This arbitrariness in determining the threshold range can be avoided by fixing the number of nearest neighbors in the second step of the modified algorithm (see Sec. II), rather than by fixing the threshold ε. By fixing the number of nearest neighbors of each state space point to be L_r, we fix the RR of every single point, and hence that of the whole ensemble as $\text{RR}=L_r/L$.[15] Since the RR is a normalized measure that varies from 0 to 1, the algorithm can be automated to generate DRTs by varying the RR in a certain range that is the same for all systems. For example, for an ensemble of size $L=50\,000$, we suggest to use a range of about 0.0001–0.03. In this case, RR=0.0001 would mean that the number of nearest neighbors is fixed to 5 and RR=0.03 will correspond to fixing the number of nearest neighbors as 1500. As we can anticipate, choosing a further lower value of RR ($\text{RR}<0.0001$) will cause higher redundancy in the resulting DRTs or failure of the algorithm to generate a DRT. On the other hand, opting for a higher value of RR ($\text{RR}>0.03$) will increase the error due to bigger jumps. Thus, one should choose a suitable range of RR depending on the size of the ensemble.

The phase space projection of a DRT obtained from an ensemble of short trajectories of the Rössler oscillator with $N=50$ and $M=1000$ using the new algorithm for RR=0.001 is shown in Fig. 6(a). As we can see, the phase portrait of the DRT clearly resembles that of an original LT of the Rössler oscillator. Figure 6(b) shows the variations in the relative error measures depending on RR. As expected, the relative error measures computed with respect to the autocorrelation function, mutual information, and mean diagonal line of the RP increase rapidly for $\text{RR}>0.006$. Such a response of the relative error measures is similar to and is also about the same order of magnitude as that observed when ε is increased [Figs. 3(a2), (b2), and (c2)].

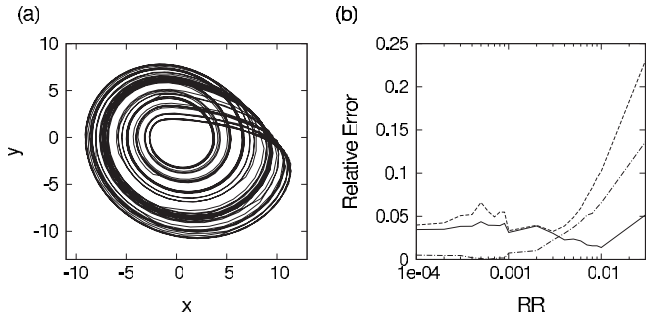

FIG. 6. (a) Phase space projection of a DRT constructed using the new algorithm from an ensemble of short trajectories ($N=50$, $M=1000$) of the Rössler oscillator for RR=0.001. (b) Error in autocorrelation (solid line), mutual information (dashed line), and mean diagonal line of RP (dot-dashed line) in dependence of RR.

These results show that fixing the number of nearest neighbors, rather than ε, does not worsen the results. Thus, they clearly justify the use of RR for generating DRTs instead of ε, and thereby allow the automation of the algorithm. The proposed automation process will strongly facilitate the application of the new improved algorithm to experimental data.

VI. APPLICATION TO ELECTROCHEMICAL DATA

In this section we test the performance of the improved algorithm reproducing both the short and the long time scale dynamics by applying it to univariate experimental data from electrochemical oscillators. The data were obtained by measuring the current from an electrode that is immersed in sulfuric acid. This system, measured with a sampling rate of 200 Hz, has been shown to display chaotic dynamics.[26–28] The phase portrait of the long time series [Fig. 7(a)] is obtained by using Taken's time delay embedding.[29] The embedding dimension estimated using the false nearest neighbor method is 4, and the embedding delay used for the reconstruction estimated from the autocorrelation function is 26.[8,14] The estimates of K_2 corresponding to the short and long term dynamics of the attractor are $K_2^1=1.9092\pm0.020$ and $K_2^2=0.2972\pm0.016$.

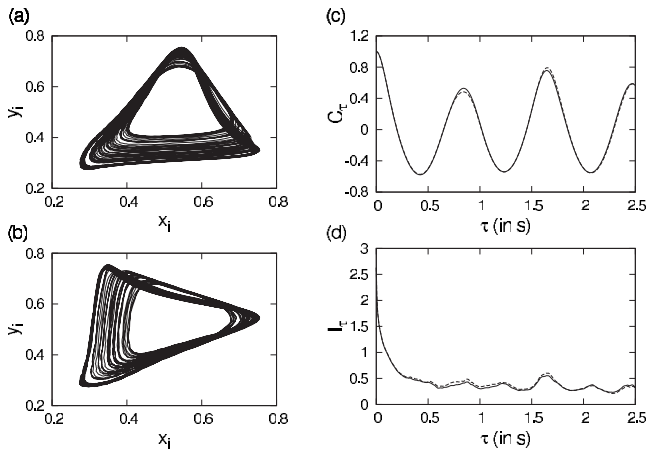

FIG. 7. Phase portraits of the original long time series (a) of the chemical oscillator and that of the (b) DRTs generated using the new algorithm for a RR of 0.0001. The (c) autocorrelation and the (d) mutual information function of the original long time series (solid line) and that of the DRTs (dashed line).

Next we apply the new algorithm to an ensemble consisting of $M=329$ short trajectories, each of length 200 (corresponding to a duration of 1 s). Each of the short trajectories is then embedded with dimension 4 and delay 26 (corresponding to a duration of 0.13 s). As discussed in Ref. 8, in order to determine the embedding parameters from the ensemble of short trajectories, one can concatenate the short trajectories and estimate both the embedding dimension and the delay from the long concatenated time series. This approach is valid if the length of the short time series is larger than the correlation time of the underlying system. Applying the new algorithm, a DRT of length 10 000 is generated using a fixed RR=0.0001. The phase space of the DRT clearly resembles that of the original LT as well as the estimates of the autocorrelation function and mutual information functions [Figs. 7(b)–7(d)]. Since a mere eyeball comparison of the phase portraits might sometimes be deceiving and does not really give evidence that the algorithm indeed reconstructs the dynamical replicant, we further proceed to calculate the entropy estimates from the generated DRT. The estimates of K_2^1 and K_2^2 computed from the DRTs are 1.8724 ± 0.025 and 0.4216 ± 0.023, respectively. Both these values are rather close to the ones obtained using the LTs. It is worth noting that the estimates of K_2^1 and K_2^2 obtained with the original algorithm were 1.6188 ± 0.071 and 0.5416 ± 0.137, respectively, showing a higher discrepancy with respect to the values obtained for the LTs.

These results clearly show the potentials of the improved algorithm in reproducing the dynamics of an underlying system, when we have a sufficient amount of information available although discontinuously. However, note that sometimes short trajectories that are obtained experimentally might not be sufficient enough to generate DRTs that reflect the underlying system's dynamics, e.g., if the length of the short trajectories is smaller than the correlation time of the original trajectory. This issue needs to be resolved in the future.

VII. APPLICATION TO TRANSIENT CHAOS

Now we exemplify the applicability of our algorithm to a classical problem of transient chaos. Transient chaotic dynamical systems have trajectories that exhibit chaotic behavior for a rather brief period of time before settling onto a final state. The primary reason for the occurrence of such a phenomenon is the existence of chaotic saddles in phase space, which have a fractal structure along their stable and unstable manifolds. Chaotic saddles, which typically arise during crisis, attract the trajectories starting at nearby initial conditions, causing them to exhibit chaotic behavior before escaping through the unstable direction to some other attractor.[4–7] Transient chaos has been found to occur in a wide range of low- and high-dimensional dynamical systems, playing a key role in numerous physical phenomena, such as chaotic scattering and particle transport in hydrodynamical flows.[30–33] Furthermore, chaotic transients are also observed in various experimental systems such as chemical oscillators, electric power systems, and ecological systems.[34]

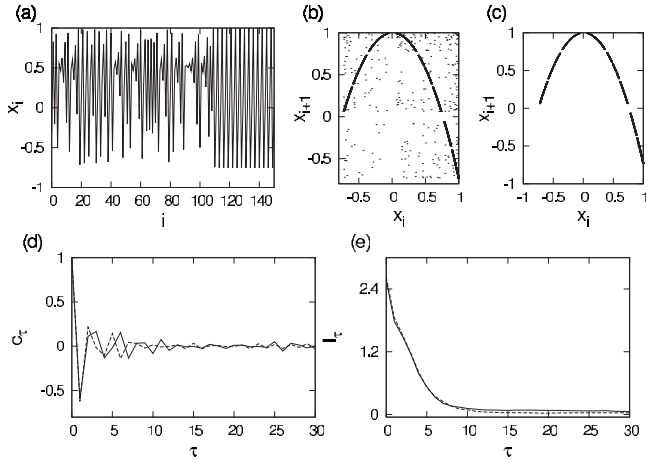

FIG. 8. (a) A short segment of the x-component of the quadratic map displaying a clear transition to a three period cycle. The phase projection of (b) the ensemble of transiently chaotic trajectories and (c) the DRT generated with the new algorithm at RR=0.0002. The (d) autocorrelation function and the (e) mutual information function of the DRT (solid line) and a long time series of the chaotic logistic map (dashed line).

The extensive occurrence of transient chaos has kindled the investigation of the properties of chaotic saddles by either generating long numerical trajectories on the chaotic saddles[35–38] or by direct investigation of an ensemble of short transient trajectories.[13,39] While the latter methods might not give all the necessary details about the saddle, many of the former methods require an explicit knowledge about the system equations to generate LTs. Jánosi *et al.*,[36] however, adopted a purely numerical approach to the problem, analogous to ours. However, their procedure involves loss of valuable data to reconstruct a LT.

We now apply our algorithm to an ensemble of short transient chaotic trajectories obtained from a quadratic map to reconstruct a LT on the chaotic saddle.

The quadratic map[5,6]

$$x_{n+1} = 1 - ax_n^2 \tag{7}$$

is one of the simplest cases where transient chaos has been observed.[6] For $a=1.7548$, the trajectories starting from random initial conditions behave chaotically for a short time interval, before settling into a period-3 orbit [Fig. 8(a)]. Note that for a given system and random initial conditions the duration of transients is exponentially distributed.[6] Since, in general, it is possible to apply our algorithm to an ensemble of short trajectories which do not have the same length, we now generate a DRT from the ensemble of transients, whose lengths are exponentially distributed.

First, an ensemble of short transients was obtained by iterating the map starting at random initial conditions that were uniformly distributed in the interval $0 \leq |x_0| \leq 1.0$. The iterations were stopped when the phase space trajectories reached the three period cycle, and the first 15 and the last 75 points of transients were omitted from this analysis. The ensemble consisted of $M=360$ transients, each of whose length was greater than 2. The average length of the short trajectories in the ensemble was 31.98. The semiattractor corresponding to such an ensemble is shown in Fig. 8(b).

Applying the modified DRT algorithm, a DRT of length 5000 was generated for RR=0.0002. The phase space projection of this DRT is shown in Fig. 8(c). As we see, the phase portrait of the DRT closely resembles the one of the logistic map. Moreover, the estimated autocorrelation function and the mutual information function reproduce the ones of the chaotic logistic map [Figs. 8(d) and 8(e)]. Furthermore, the value of the Rényi entropy K_2 calculated from the DRTs is 0.472 ± 0.004, which is very close to the estimated value of the Lyapunov exponent of the semiattractor, $\lambda = 0.489 \pm 0.0003$,[6] validating the obtained result.

VIII. CONCLUSION

In this paper we have addressed one crucial problem of the analysis of real world data, namely, the lack of long enough data sets. Insufficient amount of continuous data often hinders the application of many time series analysis techniques. Here we have presented a substantially improved version of an earlier proposed algorithm to generate DRTs that imitate the underlying system dynamics. The modified algorithm has only one parameter, allowing its complete automation and reducing the computation time needed for its application. We have extensively compared the improved algorithm with the original one and have shown that the new algorithm performs better when compared with the formerly published approach with respect to linear and nonlinear measures that characterize the dynamics of the underlying system. Furthermore, the improved algorithm reproduces both the short term and the long term dynamics of the system under investigation better than the original algorithm.

We have shown the applicability of the improved DRT algorithm to experimental data from chemical oscillators and to the problem of transient chaotic trajectories, thereby exemplifying the potential of our algorithm.

ACKNOWLEDGMENTS

C.K. thanks the Virtual Institute-Pole Equator Pole (PEP), NATO projects, and the Microgravity Application Program/Biotechnology from ESA (Contract No. 14592) for their financial support. M.C.R. would like to acknowledge the Scottish Universities Life Science Alliance (SULSA) for financial support. M.T. would like to acknowledge the RCUK academic fellowship from EPSRC. N.M. would like to thank the DFG graduate school 1364 (Shaping Earth's Surface in a Variable Environment) for the financial support. J.K. would like to thank SFB 555 (C4) for financial support. J.L.H. acknowledges the support from the National Science Foundation.

[1] E. Ullner, A. Koseska, J. Kurths, E. Volkov, H. Kantz, and J. G.-Ojalvo, Phys. Rev. E **78**, 031904 (2008).
[2] V. S. Pande, I. Baker, J. Chapman, S. P. Elmer, S. Khaliq, S. M. Larson, Y. M. Rhee, M. R. Shirts, C. D. Snow, E. J. Sorin, and B. Zagrovic, Biopolymers **68**, 91 (2003).
[3] J. A. Yorke and E. D. Yorke, J. Stat. Phys. **21**, 263 (1979).
[4] C. Grebogi, E. Ott, and J. A. Yorke, Phys. Rev. Lett. **48**, 1507 (1982).
[5] C. Grebogi, E. Ott, and J. A. Yorke, Physica D **7**, 181 (1983).
[6] H. Kantz and P. Grassberger, Physica D **17**, 75 (1985).
[7] T. Tél, *Directions in Chaos* (World Scientific, Singapore, 1990), Vol. 3.
[8] C. Komalapriya, M. Thiel, M. C. Romano, N. Marwan, U. Schwarz, and J. Kurths, Phys. Rev. E **78**, 066217 (2008).

[9]H. Poincaré, Acta Math. **13**, 1 (1890).
[10]E. Ott, *Chaos in Dynamical Systems*, 2nd ed. (Cambridge University Press, Cambridge, 2002).
[11]G. Robinson and M. Thiel, Chaos **19**, 023104 (2009).
[12]P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).
[13]M. Dhamala, Y.-C. Lai, and E. J. Kostelich, Phys. Rev. E **64**, 056207 (2001).
[14]H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. (Cambridge University Press, Cambridge, 2004).
[15]N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, Phys. Rep., Phys. Lett. **438**, 237 (2007).
[16]A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[17]M. Thiel, M. C. Romano, P. L. Read, and J. Kurths, Chaos **14**, 234 (2004).
[18]O. E. Rössler, Phys. Lett. A **57**, 397 (1976).
[19]M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[20]*Encyclopedia of Nonlinear Sciences* (Taylor & Francis, London, 2005).
[21]E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[22]R. Gilmore, Rev. Mod. Phys. **70**, 1455 (1998).
[23]Nonhyperbolic dynamical systems are generally characterized by the existence of multiple main time scales (Refs. 17 and 40). In the chaotic Rössler system, there exist two main time scales that characterize the dynamics. While the first time scale describes the short term dynamics of the attractor and is linked to the amplitude fluctuations of the system, the second one reflects the long term dynamics of the system and is related to the system's phase diffusion or relaxation behavior. The presence of two time scale dynamics can be identified either by analyzing the decay of autocorrelation function's envelope (Ref. 40) or the diagonal line distributions of a RP (Ref. 17).
[24]M. Thiel, M. C. Romano, J. Kurths, R. Meucci, E. Allaria, and F. T. Arecchi, Physica D **171**, 138 (2002).
[25]S. Schinkel, O. Dimigen, and N. Marwan, Eur. Phys. J. Spec. Top. **164**, 45 (2008).
[26]I. Z. Kiss and J. L. Hudson, Phys. Chem. Chem. Phys. **4**, 2638 (2002).
[27]I. Z. Kiss and J. L. Hudson, Phys. Rev. E **64**, 046215 (2001).
[28]I. Z. Kiss, W. Wang, and J. L. Hudson, Phys. Chem. Chem. Phys. **2**, 3847 (2000).
[29]F. Takens, *Dynamical Systems and Turbulence* (Springer-Verlag, Berlin, 1981), Vol. 898, pp. 366–381.
[30]J. P. Crutchfield and K. Kaneko, Phys. Rev. Lett. **60**, 2715 (1988).
[31]B. Eckhardt and A. Mersmann, Phys. Rev. E **60**, 509 (1999).
[32]E. Ott and T. Tél, Chaos **3**, 417 (1993).
[33]Z. Toroczkai, G. Károlyi, Á. Péntek, T. Tél, and C. Grebogi, Phys. Rev. Lett. **80**, 500 (1998).
[34]M. Dhamala and Y.-C. Lai, Phys. Rev. E **59**, 1646 (1999).
[35]H. E. Nusse and J. A. Yorke, Physica D **36**, 137 (1989).
[36]I. M. Jánosi, L. Flepp, and T. Tél, Phys. Rev. Lett. **73**, 529 (1994).
[37]P. Moresco and S. P. Dawson, Physica D **126**, 38 (1999).
[38]D. Sweet, H. E. Nusse, and J. A. Yorke, Phys. Rev. Lett. **86**, 2261 (2001).
[39]M. Dhamala, Y.-C. Lai, and E. J. Kostelich, Phys. Rev. E **61**, 6485 (2000).
[40]V. S. Anishchenko, T. E. Vadivasova, G. A. Okrokvertskhov, and G. I. Strelkova, Phys. Usp. **48**, 151 (2005).