Recovering hidden dynamical modes from the generalized Langevin equation

THE JOURNAL OF CHEMICAL PHYSICS **145**, 094102 (2016)

Recovering Hidden Dynamical Modes from the Generalized Langevin Equation

Shinnosuke Kawai[1*] and Yusuke Miyazaki[1]

[1] *Department of Chemistry,*
Faculty of Science, Shizuoka University,
836 Ohya, Suruga-ku,
Shizuoka, Japan 422-8529

Abstract

In studying large molecular systems, insights can better be extracted by selecting a limited number of physical quantities for analysis rather than treating every atomic coordinate in detail. Some information may, however, be lost by projecting the total system onto a small number of coordinates. For such problems, the generalized Langevin equation (GLE) is shown to provide a useful framework to examine the interaction between the observed variables and their environment. Starting with the GLE obtained from the time series of the observed quantity, we perform a transformation to introduce a set of variables that describe dynamical modes existing in the environment. The introduced variables are shown to effectively recover the essential information of the total system that appeared to be lost by the projection.

PACS numbers: 05.10.-a, 05.40.-a, 82.20.-w, 34.10.+x

I. INTRODUCTION

With the development of theoretical concepts and computational techniques, chemical physics is continuously expanding its study field into larger and more complex systems of chemical and biological interest. Many such systems comprise a large number of molecules and the phenomena can be regarded as concerted motions generated by many-body interaction among the particles. In order to obtain theoretical understanding of such systems, it is often a good strategy to divide the total system into a *subsystem* and its *environment* rather than to treat all the atomic coordinates equally. The subsystem consists of a small number of selected variables that are of research interest, or are experimentally observable. All the other degrees of freedom are regarded as the environment that surrounds the subsystem. The effect of the environment on the dynamics in the subsystem can be implicitly described as friction and random force. Simply put, the former dissipates the energy of the subsystem into the environment while the latter describes occasional impacts of the environment on the subsystem.

The generalized Langevin equation (GLE)[1–3] is a mathematical description of such situations. A common form of the GLE reads as follows

$$\ddot{q}_1 = -\frac{\partial V_{\mathrm{M}}(q_1)}{\partial q_1} - \int_0^t \gamma(t-t')\dot{q}_1(t')\mathrm{d}t' + \xi(t), \tag{1}$$

where derivation with respect to time t is expressed by the dot over the symbol. Here the coordinate q_1 describes the subsystem chosen from a large system, and $\ddot{q}_1$ is its acceleration. The first term on the right hand side is the mean force acting on q_1 given as the derivative of a function $V_{\mathrm{M}}(q_1)$ which is called the potential of mean force. The second term expresses the frictional force that depends on the velocity $\dot{q}_1$ in the past. The third term $\xi(t)$ is the random force representing the kick from the environment that is determined by the environment and uncorrelated to the initial condition of the subsystem. The mean force represents the interaction within the subsystem and the average effect of the environment. It can be calculated as the force (or acceleration) on the subsystem averaged over all the instants conditioned on the value of q_1. The friction in the GLE appears as delayed response of the environment to the subsystem. It depends on the history of the subsystem, thus making the GLE an integro-differential equation. By using the technique of projection operator, it has been proved[1–3] that any Hamiltonian system can be projected onto its subsystem obeying the GLE. Zwanzig[4] showed that the GLE with a linear friction term

is exact when the environment is a collection of harmonic oscillators and the coupling to the subsystem is bilinear. Cortés *et al.*[5] went further to show that, when the couplings are linear in the subsystem coordinate but nonlinear in the other coordinates, the system still obeys the GLE with linear friction up to the first order in the system-environment coupling strength. The projection operator formalism has also been extended to non-Hamiltonian systems[6].

In the field of molecular science, the Langevin-type formulation was the starting point of the traditional rate theories[7,8]. Kramers[7] derived analytical expressions for one-dimensional barrier-crossing rates by using the Fokker-Planck equation, which is equivalent to the Langevin equation, the short-memory limit of the GLE. The study was followed by Mel'nikov and Meshkov[9] who obtained a formula that connects two limiting behaviors treated by Kramers. Langer[10] gave the multi-dimensional version of the rate formula. Extension to systems having retarded response of the environment, described by the GLE, was given by Grote and Hynes[8]. The useful formula derived by them was used in the analyses of reaction rates in molecular systems[11–13]. The dynamics of the barrier crossing was further studied in terms of the phase space structure in the neighborhood of a saddle point in the energy landscape[14–20]. By using the GLE, it was proved that, even in the existence of thermal noise, there exists a clear structure in the phase space that determines the occurrence of the chemical reaction.

The GLE formulation has found its application also in the time series analysis of meteorological and financial data[21,22]. Another application of the GLE formulation was found in the problem of model reduction[23]. Starting with a rather complicated model for a cellular signaling process, the time series obtained by simulations with the full model was put into the GLE-based time series analysis. The reconstructed GLE provided a reduced one-dimensional model that has the same prediction ability with the original full model. The time series analysis with the Langevin modeling has been applied[24,25] to the data obtained from classical molecular dynamics simulations of protein. The work showed the importance of the multi-dimensionality and the inertial effects (the acceleration term) in the structural dynamics of proteins.

Surely, the reduction of the total system onto a subsystem causes some information to be lost. The choice of the subsystem, that is, the choice of the explicit dynamical variables must therefore be conducted with caution by considering their physical significance and

our research interest. For example, the number of the dynamical variables was chosen carefully by using the principal component analysis in Refs. 24,25. Moreover, even with deliberate choice of the subsystem variables, the nature of the environment affecting the subsystem is hidden in the history-dependent friction term that is rather intractable for intuitive understanding.

To elucidate the dynamical modes that are hidden in the environment but are still interacting with the subsystem, one of the present authors, following a long line of studies,[15,19,26–35] has recently proposed a formulation that converts the GLE into multi-dimensional and memoryless equations of motion[36]. The idea of expressing the GLE by equivalent set of ordinary differential equations dates back to the continued-fraction expansion of the friction kernel introduced by Mori.[26] Grigolini formulated a matrix-form equation of motion for the variables representing the dynamical modes in the environment.[27,29] It was later utilized to perform numerical simulations with GLE and to develop theories of stochastic molecular processes under nonwhite noise.[29,37,38] In the method, new dynamical variables are introduced to convert the GLE to memoryless differential equations. In Ref. 36 the theory was reformulated by starting with multi-exponential form of the friction kernel and provided explicit expression of these new variables in terms of the observed time series $q_1(t)$. In the equations of motion expressed in these new variables, the explicit dependence on the history has been removed at the cost of increasing the number of dynamical variables. The introduced variables can thus be considered to be an effective description of the dynamical modes existing in the environment.

In the present paper, we demonstrate the analysis with a simple model system, with the aim of providing an intuitively clear example of the problem of "information loss" through the reduction onto the subsystem and examining to what extent the GLE-based time series analysis can recover the essential information that appeared to be lost. In Sec. II, a simple model is introduced. Numerical demonstration is given on how the multi-dimensional information is lost by choosing a single coordinate and projecting the total system onto it. In Sec. III, the GLE-based time series analysis is demonstrated on the same model system. We close the paper with a brief concluding remark in Sec. IV

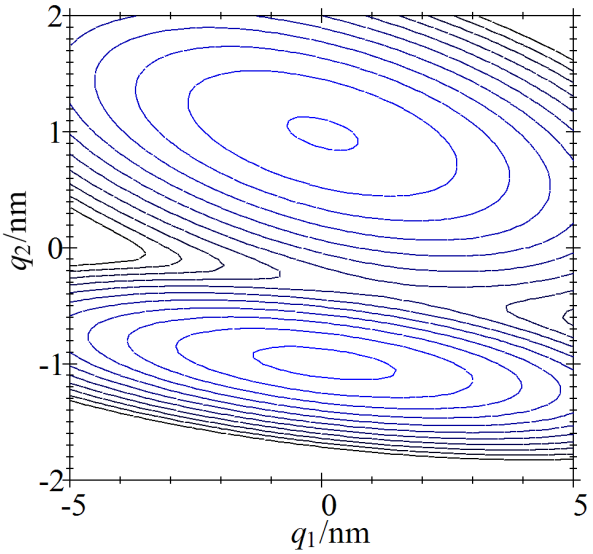

FIG. 1: Potential energy of the model system is shown by contours as a function of the position coordinates q_1 and q_2.

II. MODEL SYSTEM

In studying a system with a large number of atoms or molecules, one often chooses a small number of variables ("subsystem") to analyze. As stated in Sec. I, the main subject of the present study is the problem of "information loss" through this projection of the huge total system onto a low-dimensional subsystem. Let us here examine this information loss by a simple example. Consider a two-dimensional system with the potential energy shown in Fig. 1 as a function of two position coordinates q_1 and q_2. The potential energy surface of the model system is given by

$$\begin{aligned} V(q_1,q_2) =& \frac{1}{2}\left\{V_{\rm A}+V_{\rm B}-\sqrt{(V_{\rm A}-V_{\rm B})^2+b^2}\right\}, \\ V_{\rm A}(q_1,q_2) =& \frac{1}{2}m\omega_{\rm A1}^2 q_1^2+\frac{1}{2}m\omega_{\rm A2}^2\left(q_2-q_{\rm A2}^{\rm eq}\right)^2+mc_{\rm A}q_1\left(q_2-q_{\rm A2}^{\rm eq}\right), \\ V_{\rm B}(q_1,q_2) =& \frac{1}{2}m\omega_{\rm B1}^2 q_1^2+\frac{1}{2}m\omega_{\rm B2}^2\left(q_2-q_{\rm B2}^{\rm eq}\right)^2+mc_{\rm B}q_1\left(q_2-q_{\rm B2}^{\rm eq}\right), \end{aligned} \tag{2}$$

where m is the mass of the particle and the specific values of the parameters $(b,\omega_{\rm A1},\omega_{\rm A2},q_{\rm A2}^{\rm eq},c_{\rm A},\omega_{\rm B1},\omega_{\rm B2},q_{\rm B2}^{\rm eq},c_{\rm B})$ are given in Supplementary Material[39]. The functions $V_{\rm A}$ and $V_{\rm B}$ describe harmonic oscillator potentials centered at $(q_1,q_2)=(0,q_{\rm A2}^{\rm eq})$ and

$(q_1, q_2) = (0, q_{\mathrm{B2}}^{\mathrm{eq}})$, respectively. The potential V satisfies

$$\begin{aligned} V \approx & V_{\mathrm{A}}, \text{ if } V_{\mathrm{B}} - V_{\mathrm{A}} \gg b^2 > 0, \\ V \approx & V_{\mathrm{B}}, \text{ if } V_{\mathrm{A}} - V_{\mathrm{B}} \gg b^2 > 0. \end{aligned} \tag{3}$$

Thus the potential V is almost equal to one of the harmonic potentials in the vicinity of each minimum. And the two harmonic potentials are connected smoothly via the coupling constant b^2.

The equation of motion for the system is given by

$$\begin{aligned} m\ddot{q}_1 = & -\frac{\partial V}{\partial q_1} \\ m\ddot{q}_2 = & -\frac{\partial V}{\partial q_2} - \gamma_2 \dot{q}_2 + \xi_{\mathrm{w}}(t), \end{aligned} \tag{4}$$

where friction and random force have been introduced to the q_2 mode in order to realize a thermalized system. The random force $\xi_{\mathrm{w}}(t)$ is white noise satisfying the following property due to the fluctuation-dissipation theorem:

$$\langle \xi_{\mathrm{w}}(t) \xi_{\mathrm{w}}(t') \rangle = 2 k_{\mathrm{B}} T \gamma_2 \delta(t - t'), \tag{5}$$

where k_{B} is the Boltzmann constant, T the absolute temperature, and δ stands for the Dirac delta function. The specific values of the parameters $(\gamma, k_{\mathrm{B}}, T)$ are given in Supplementary Material[39]. In Supplementary Material we also investigate the case where a white noise is also added to the q_1-coordinate. Numerical integration of the equations of motion (4) was performed by the method of Ref. 40. Trajectories of length 1000 ns were generated with 100 different initial conditions taken from thermal ensemble.

This potential energy surface has two wells in the upper and the lower regions. The system can be trapped in one of the wells for a while and can sometimes make "state transition" from one well to the other. The transition occurs principally along the direction of q_2, while the motion in the q_1 direction is simple oscillation. Note that, for many situations of complex molecular systems, the potential energy surface of the full dimensionality is not available a priori, and it is often difficult to make the very correct choice of the coordinate that is the most "important" for the system. Let us therefore consider the case where we have chosen (wrongly) the horizontal coordinate q_1 as our observable (i.e. subsystem coordinate). Figure 2 shows the potential of mean force along q_1. This curve shows only one well, implying

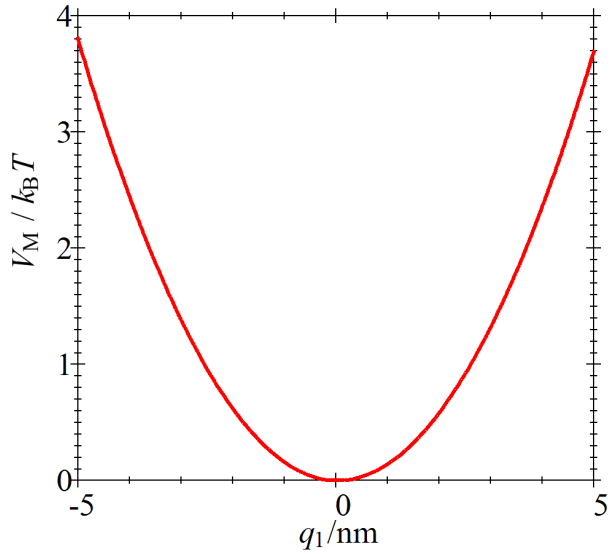

FIG. 2: Potential of mean force along q_1.

that the motion along q_1 is principally a simple oscillation. In other words, the two wells apparent in Fig. 1 has collapsed into one by projection onto the q_1 axis.

Figure 3(a) and (b) show a typical time series of this system. As expected from Fig. 1, the time evolution of q_1 is a simple oscillation. The time series of the other coordinate q_2 in the same time interval shows a state transition from one well ($q_2 \approx -1$ nm) to the other ($q_2 \approx +1$ nm). Remember we are considering the case where we have chosen q_1 as our observable. If we are simply looking by human eyes at the time series $q_1(t)$ as shown in Fig. 3(a), the state transition that the system makes in this time region can hardly be appreciated.

III. RESULTS OF THE GLE-BASED TIME SERIES ANALYSIS

Next let us see what the GLE-based time series analysis can tell us. There is an established method[22–25,41] for obtaining the GLE (that is, the specific functional form of the potential $V_M(q_1)$ of mean force and the friction kernel $\gamma(\tau)$) from the time series data of the observed coordinate $q_1(t)$ only. For the sake of completeness, the method we have used is explained below in Sec. III A together with the numerical result for the friction kernel obtained by the analysis of $q_1(t)$. Then, decomposition of the friction and random force following the procedure of Ref. 36 gives effective dynamical variables to describe the motion existing in

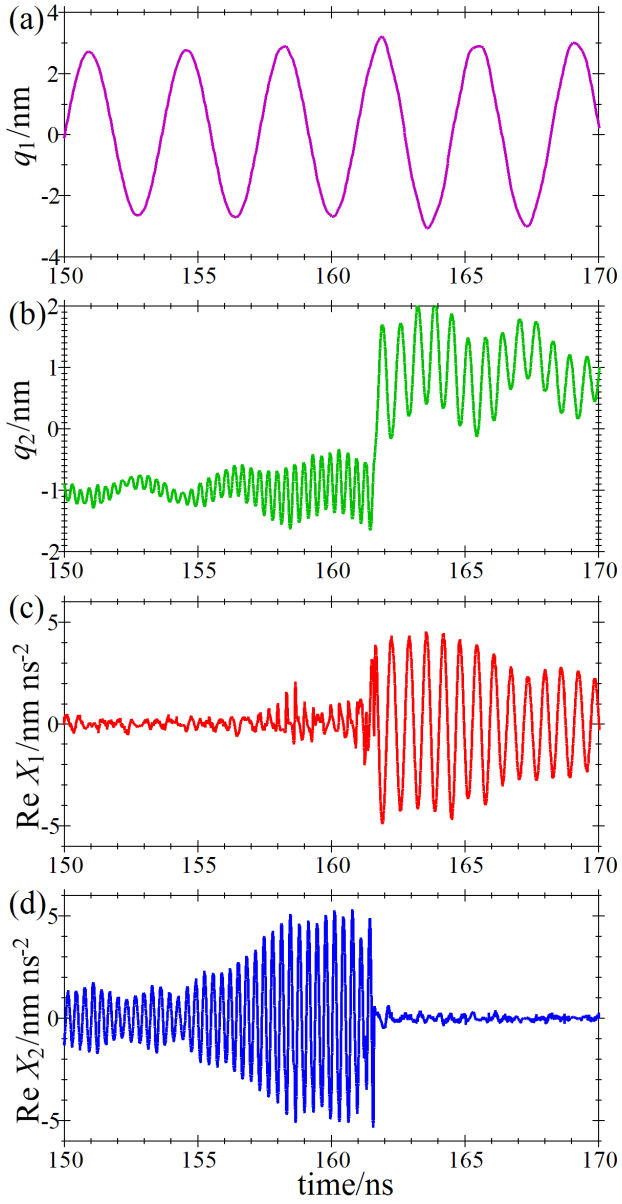

FIG. 3: Typical time evolution of the model system used as an example in the present paper. Panels (a) and (b) show the time series of the position coordinates q_1 and q_2, respectively. Panels (c) and (d) show the time series of the effectively expressed environmental modes obtained by the GLE-based analysis on the time series $q_1(t)$ only.

the environment and interacting with the subsystem. Brief description of the analysis and its results are presented in Sec. III B with more detailed numerical values in Supplementary Material[39].

A. Generalized Langevin Equation from Time Series

1. *Potential of Mean Force*

The definition of the potential of mean force is

$$-\frac{\partial V_{\mathrm{M}}(q_1)}{\partial q_1} = \langle \ddot{q}_1; q_1 \rangle, \tag{6}$$

where the right hand side stands for the average value of the acceleration $\ddot{q}_1$ conditioned on the value of q_1. To calculate the right hand side numerically from time series, we collect all the instants t at which the value of $q_1(t)$ falls in the range $n\Delta q \leq q_1 \leq (n+1)\Delta q$, where Δq is a sufficiently small width and n is an integer. Taking the average value of $\ddot{q}_1(t)$ for these instants estimates the value of $\langle \ddot{q}_1; q_1 \rangle$ at $q_1 = (n+1/2)\Delta q$. This calculation is repeated for all integer values of n in the range of $-600 \leq n \leq +599$, that is, in the range -6 nm $< q_1 < +6$ nm because we used $\Delta q = 0.01$ nm in the present calculation. Outside this region only a very few data points were available in the time series and taking average was numerically unstable. Numerical integration of the mean force with respect to q_1 yields $V_{\mathrm{M}}(q_1)$. The result for the calculated potential of mean force $V_{\mathrm{M}}(q_1)$ along q_1 is shown in Fig. 2.

2. *Friction Kernel*

Multiplying $\dot{q}_1(0)$ on the generalized Langevin equation (1) and taking ensemble average gives

$$\langle \dot{q}_1(0)\ddot{q}_1(t)\rangle = -\left\langle \dot{q}_1(0)\frac{\partial V_{\mathrm{M}}}{\partial q_1}(q_1(t))\right\rangle - \int_0^t \gamma(t-t')\langle \dot{q}_1(0)\dot{q}_1(t')\rangle \mathrm{d}t', \tag{7}$$

because the random force is uncorrelated to the initial condition: $\langle q_1(0)\xi(t)\rangle = 0$. To simplify the notation, we may introduce the following symbols for the correlation functions:

$$\begin{aligned} C(t) :=& \langle \dot{q}_1(0)\dot{q}_1(t)\rangle, \\ F(t) :=& -\left\langle \dot{q}_1(0)\frac{\partial V_{\mathrm{M}}}{\partial q_1}(q_1(t))\right\rangle. \end{aligned} \tag{8}$$

Eq. (7) becomes then

$$\dot{C}(t) = F(t) - \int_0^t \gamma(t-t')C(t')\mathrm{d}t'. \tag{9}$$

Differentiating the both sides with respect to t and integration by parts yield

$$C(0)\gamma(t) = -\ddot{C}(t) + \dot{F}(t) - \int_0^t \gamma(t-t')\dot{C}(t')\mathrm{d}t'. \tag{10}$$

This equation gives $\gamma(t)$ in terms of the correlation functions $C(t)$ and $F(t)$ that can be evaluated from the observed time series of $q_1(t)$. Discretizing the time and writing

$$\begin{aligned} \gamma_j &:= \gamma(j\Delta t), \\ C_j &:= C(j\Delta t), \\ F_j &:= F(j\Delta t), \end{aligned} \tag{11}$$

we can evaluate the integral in Eq. (10) numerically as follows:

$$C_0\gamma_j = -\ddot{C}_j + \dot{F}_j - \Delta t \sum_{i=1}^{j} w_i \gamma_{j-i}\dot{C}_i, \tag{12}$$

where w_j stands for the weights in numerical integration algorithm. Note that γ_j does not appear in the right hand side because $\dot{C}(0) = 0$. Thus, starting with $\gamma_0 = (-\ddot{C}_0 + \dot{F}_0)/C_0$, the value of γ_j for $j = 1, 2, 3, \ldots$ can be successively calculated from the previously calculated values $\gamma_0, \gamma_1, \ldots, \gamma_{j-1}$ through Eq. (12). In the present analysis we have used $\Delta t = 0.1$ ns and we have found that the simple trapezoidal rule ($w_j = 1/2$, and $w_i = 1$ for $1 \le i < j$) is satisfactory in the numerical integration in Eq. (12). The result is shown in Fig. 4.

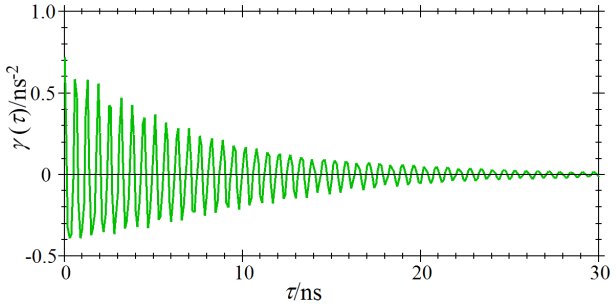

FIG. 4: Friction kernel in the GLE obtained from the time series of q_1.

B. Decomposition of the Friction and Random Force

One of the present authors, following a long line of studies,[15,19,26–35] has recently provided a formulation that converts the GLE into multi-dimensional and memoryless equations of

motion[36]. Briefly, we define new dynamical variables $X_m(t)$ in terms of the observed time series $q_1(t)$ by

$$\begin{aligned} X_m(t) = &- \int_0^t c_m \exp\left(\mathrm{i}\alpha_m(t-t')\right) \dot{q}_1(t')\mathrm{d}t' \\ &+ \int_0^{+\infty} k_m(t-t')\xi(t')\mathrm{d}t', \end{aligned} \tag{13}$$

where complex coefficients α_m and c_m are obtained by fitting the friction kernel to the multi-exponential form: $\gamma(\tau) \approx \sum_m c_m \exp(\mathrm{i}\alpha_m\tau)$. The integration kernel k_m is designed in such a way that the GLE (1) is equivalent to the following multi-dimensional equations of motion:

$$\begin{aligned} \ddot{q}_1 =& -\frac{\partial V_{\mathrm{M}}(q_1)}{\partial q_1} + \sum_m X_m, \\ \dot{X}_m =& \mathrm{i}\alpha_m X_m - c_m\dot{q}_1 + b_m\eta(t), \end{aligned} \tag{14}$$

where b_m is a coefficient obtained from $\{\alpha_m, c_m\}$, and $\eta(t)$ is a white noise. The explicit functional form of k_m can be obtained from $\{\alpha_m, c_m\}$ through the recipe of Ref. 36, which is reproduced in Supplementary Material[39] for the sake of completeness. Compared to Eq. (1), the explicit dependence on the history has been removed in Eq. (14) at the cost of increasing the number of dynamical variables.

In Eq. (14), the effect of the environment (formerly described as friction and random force in Eq. (1)) is expressed in the form of dynamical modes X_m existing in the environment. In the time series analysis, we first have $q_1(t)$ in hand, which is used to obtain $V_{\mathrm{M}}(q_1)$ and $\gamma(\tau)$. Then we can calculate the time series of the random force $\xi(t)$ from Eq. (1), that is,

$$\xi(t) = \ddot{q}_1(t) + \frac{\partial V_{\mathrm{M}}(q_1)}{\partial q_1} + \int_0^t \gamma(t-t')\dot{q}_1(t')\mathrm{d}t'. \tag{15}$$

They are substituted into Eq. (13) to obtain the time series of the effectively expressed environmental modes X_m.

For the present example, the time series of X_m are shown in Fig. 3(c) and (d). By comparing with Fig. 3(b), it is seen that the mode X_1 is activated when the system is in the well at $q_2 \approx +1$ nm, and the mode X_2 is activated when the system is in the other well at $q_2 \approx -1$ nm. State transition is clear in the time series of X_1 and X_2, and this transition corresponds to the real state transition of the original system between the wells demonstrated by q_2.

Note that the time series of X_m are obtained by the GLE-based time series analysis on $q_1(t)$ alone, *without ever looking at* $q_2(t)$. Even when we have no knowledge about $q_2(t)$, the GLE-based time series analysis of the observed variable $q_1(t)$ extracts the dynamical modes existing in the "environment" for q_1. These modes elucidate the state transition that exists in the total system but was not apparent by simply looking at the time series of $q_1(t)$. The result shown here strengthen the validity and usefulness of the Langevin-based time series analysis, which has been proved to be a powerful tool for molecular, biological, meteorological, and even financial systems in previous works[22–25].

IV. CONCLUSION AND OUTLOOK

In order to elucidate the problem associated with the projection of a multi-dimensional system onto a lower-dimensional subsystem, a simple model system was analyzed by pretending to be unaware of its full-dimensional potential energy surface and simply choosing one of the coordinates, which is actually inappropriate to describe the state transition existing in the total system. Time series analysis based on the GLE framework and the effectively expressed environmental modes approach was performed and shown to be capable of recovering the information about the state transition that appeared to have been lost by the projection. Applying the present method to various systems will elucidate essential dynamical modes and state transitions that are important in the total system but may be hidden in the "environment" by our wrong choice of the coordinate to observe.

It may be interesting to comment on a similarity in idea between the present method and the embedding techniques[42,43] used to detect attractors in dynamical systems. With the aim of extracting multi-dimensional information from one-dimensional time series, we introduce the integration over time (Eq. (13)) of the series $q_1(t)$ rather than its value at each instant. Roughly speaking, this integration may be expressed (in the discrete approximation) by a linear combination of $(\ldots, q_1((n-1)\Delta t), q_1(n\Delta t), q_1((n+1)\Delta t), q_1((n+2)\Delta t), \ldots)$ with sufficiently small interval Δt. Thus, in a sense we can say that the present method is performing a linear coordinate transformation from $(\ldots, q_1((n-1)\Delta t), q_1(n\Delta t), q_1((n+1)\Delta t), q_1((n+2)\Delta t), \ldots)$ to $(X_1, X_2, \ldots)$ in the embedding space.

It must be added that the environmental modes X_m in the present example are not q_2 itself. Rather, they describe the oscillation mode in the upper well ($q_2 \approx +1$ nm) and that in

the lower well ($q_2 \approx -1$ nm) separately. This is because these environmental modes are formulated to have constant frequency similarly to the conventional concept of normal modes. In Eq. (14) the real part of α_m gives the frequency for the oscillation of X_m. In this sense, the dynamical variables X_m should be regarded as effectively describing the dynamical variables of the system that exist in the environment and strongly interact with the subsystem, rather than the original atomic coordinates of the environment. It would be interesting to apply the present method to data obtained from molecular dynamics simulations. A good test case may be the ϕ and ψ dihedral angles of a dipeptide in water,[44–48] where we can investigate how much information on $\psi(t)$ can be extracted from the time series of $\phi(t)$, for example, as well as the interaction between the structural dynamics and the surrounding water molecules.[45,47] As in the present example, the environmental coordinates obtained by the analysis would be different from the position coordinate of each atom originally used to describe the total system, but may be effective coordinates describing the collective motions formed by multiple atoms.

This work has been supported by Grant-in-Aid for Scientific Research(B) #24750002 and #93004075 of the Japan Society for the Promotion of Science. The authors thank Prof. Motoyuki Shiga in Japan Atomic Energy Agency for continuous encouragement to this study.

---

* sskawai@shizuoka.ac.jp

[1] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).

[2] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, London, 2001).

[3] K. Kawasaki, J. Phys. A **6**, 1289 (1973).

[4] R. Zwanzig, J. Stat. Phys. **9**, 215 (1973).

[5] E. Cortés, B. J. West, and K. Lindenberg, J. Chem. Phys. **82**, 2708 (1985).

[6] J. Xing and K. S. Kim, J. Chem. Phys. **134**, 044132 (2011).

[7] H. A. Kramers, Physica **7**, 284 (1940).

[8] R. F. Grote and J. T. Hynes, J. Chem. Phys. **73**, 2715 (1980).

[9] V. I. Mel'nikov and S. V. Meshkov, J. Chem. Phys. **85**, 1018 (1986).

[10] J. S. Langer, Ann. Phys. **54**, 258 (1969).

[11] J. J. Ruiz-Pernía, I. Tuñón, V. Moliner, J. T. Hynes, and M. Roca, J. Am. Chem. Soc. **130**, 7477 (2008).

[12] R. P. McRae, G. K. Schenter, B. C. Garrett, Z. Svetlicic, and D. G. Truhlar, J. Chem. Phys. **115**, 8460 (2001).

[13] I. S. Tolokh, G. W. N. White, S. Goldman, and C. G. Gray, Mol. Phys. **100**, 2351 (2002).

[14] T. Bartsch, R. Hernandez, and T. Uzer, Phys. Rev. Lett. **95**, 058301 (2005).

[15] T. Bartsch, T. Uzer, and R. Hernandez, J. Chem. Phys. **123**, 204102 (2005).

[16] T. Bartsch, T. Uzer, J. M. Moix, and R. Hernandez, J. Chem. Phys. **124**, 244310 (2006).

[17] T. Bartsch, J. M. Moix, R. Hernandez, S. Kawai, and T. Uzer, Adv. Chem. Phys. **140**, 191 (2008).

[18] R. Hernandez, T. Uzer, and T. Bartsch, Chem. Phys. **370**, 270 (2010).

[19] S. Kawai and T. Komatsuzaki, Phys. Chem. Chem. Phys. **12**, 15382 (2010).

[20] G. T. Craven and R. Hernandez, Phys. Rev. Lett. **115**, 148301 (2015).

[21] M. Niemann, T. Laubrich, E. Olbrich, and H. Kantz, Phys. Rev. E **77**, 011117 (2008).

[22] D. T. Schmitt and M. Schulz, Phys. Rev. E **73**, 056204 (2006).

[23] A. Mukhopadhyay and J. Xing, ArXiv e-prints (2013), 1304.4908.

[24] R. Hegger and G. Stock, J. Chem. Phys. **130**, 034106 (2009).

[25] N. Schaudinnus, B. Bastian, R. Hegger, and G. Stock, Phys. Rev. Lett. **115**, 050602 (2015).

[26] H. Mori, Prog. Theor. Phys. **34**, 399 (1965).

[27] P. Grigolini, J. Stat. Phys. **27**, 283 (1982).

[28] P. Grigolini, Nuovo Cimento B **63**, 174 (1981).

[29] F. Marchesoni and P. Grigolini, J. Chem. Phys. **78**, 6287 (1983).

[30] S. A. Adelman, J. Chem. Phys. **71**, 4471 (1979).

[31] S. A. Adelman, Adv. Chem. Phys. **44**, 143 (2007).

[32] E. Pollak, H. Grabert, and P. Hänggi, J. Chem. Phys. **91**, 4073 (1989).

[33] C. C. Martens, J. Chem. Phys. **116**, 2516 (2002).

[34] T. Bartsch, J. Chem. Phys. **131**, 124121 (2009).

[35] S. Kawai and T. Komatsuzaki, Phys. Chem. Chem. Phys. **13**, 21217 (2011).

[36] S. Kawai, J. Chem. Phys. **143**, 094101 (2015).

[37] J. E. Straub, M. Borkovec, and B. J. Berne, J. Chem. Phys. **84**, 1788 (1986).

[38] E. Pollak and A. M. Berezhkovskii, J. Chem. Phys. **99**, 1344 (1993).

[39] See Supplementary Material at [URL] for numerical details.

[40] G. Bussi and M. Parrinello, Phys. Rev. E **75**, 056707 (2007).

[41] B. J. Berne and G. D. Harp, Adv. Chem. Phys. **17**, 67 (1970).

[42] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young (Springer, Berlin, 1981), vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381.

[43] R. Hegger and H. Kantz, EPL (Europhysics Letters) **38**, 267 (1997).

[44] J. Hermans, Proc. Natl. Acad. Sci. U.S.A. **108**, 3095 (2011).

[45] V. Cruz, J. Ramos, and J. Martinez-Salazar, J. Phys. Chem. B **115**, 4880 (2011).

[46] H. Jang and T. B. Woolf, J. Comput. Chem. **27**, 1136 (2006).

[47] D. Nerukh and S. Karabasov, J. Phys. Chem. Lett. **4**, 815 (2013).

[48] A. Ljubetič, I. Urbančič, and J. Štrancar, J. Chem. Phys. **140**, 084109 (2014).

Supplementary Material

Recovering Hidden Dynamical Modes from the Generalized Langevin Equation

Shinnosuke Kawai[1] and Yusuke Miyazaki[1]

[1]*Department of Chemistry, Faculty of Science, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, Japan 422-8529*

I. MODEL SYSTEM

The potential energy surface of the model system is given by

$$\begin{aligned} V(q_1,q_2) =& \frac{1}{2}\left\{V_{\mathrm{A}}+V_{\mathrm{B}}-\sqrt{(V_{\mathrm{A}}-V_{\mathrm{B}})^2+b^2}\right\}, \\ V_{\mathrm{A}}(q_1,q_2) =& \frac{1}{2}m\omega_{\mathrm{A}1}^2 q_1^2+\frac{1}{2}m\omega_{\mathrm{A}2}^2\left(q_2-q_{\mathrm{A}2}^{\mathrm{eq}}\right)^2+mc_{\mathrm{A}}q_1\left(q_2-q_{\mathrm{A}2}^{\mathrm{eq}}\right), \\ V_{\mathrm{B}}(q_1,q_2) =& \frac{1}{2}m\omega_{\mathrm{B}1}^2 q_1^2+\frac{1}{2}m\omega_{\mathrm{B}2}^2\left(q_2-q_{\mathrm{B}2}^{\mathrm{eq}}\right)^2+mc_{\mathrm{B}}q_1\left(q_2-q_{\mathrm{B}2}^{\mathrm{eq}}\right), \end{aligned} \tag{S1}$$

where m is the mass of the particle and the other parameters are given by

$$\begin{aligned} b^2/m =& 2500\ \mathrm{nm}^2\ \mathrm{ns}^{-2}, \\ \omega_{\mathrm{A}1}^2 =& 4\ \mathrm{ns}^{-2}, \\ \omega_{\mathrm{A}2}^2 =& 100\ \mathrm{ns}^{-2}, \\ q_{\mathrm{A}2}^{\mathrm{eq}} =& +1\ \mathrm{nm}, \\ c_{\mathrm{A}} =& 10\ \mathrm{ns}^{-2}, \\ \omega_{\mathrm{B}1}^2 =& 4\ \mathrm{ns}^{-2}, \\ \omega_{\mathrm{B}2}^2 =& 400\ \mathrm{ns}^{-2}, \\ q_{\mathrm{B}2}^{\mathrm{eq}} =& -1\ \mathrm{nm}, \\ c_{\mathrm{B}} =& 20\ \mathrm{ns}^{-2}. \end{aligned} \tag{S2}$$

The equation of motion for the system is given by

$$\begin{aligned} m\ddot{q}_1 &= -\frac{\partial V}{\partial q_1} \\ m\ddot{q}_2 &= -\frac{\partial V}{\partial q_2} - \gamma_2 \dot{q}_2 + \xi_{\mathrm{w}}(t), \end{aligned} \tag{S3}$$

where friction and random force have been introduced to the q_2 mode in order to realize a thermalized system. The friction coefficient is $\gamma_2/m = 0.2$ ns^{-1}. The random force $\xi_{\mathrm{w}}(t)$ is white noise satisfying the following property due to the fluctuation-dissipation theorem:

$$\langle \xi_{\mathrm{w}}(t)\xi_{\mathrm{w}}(t')\rangle = 2k_{\mathrm{B}}T\gamma_2\delta(t-t'), \tag{S4}$$

where k_{B} is the Boltzmann constant, T the abolute temperature, and δ stands for the Dirac delta function. We used the value $k_{\mathrm{B}}T/m = 10$ nm^2 ns^{-2}.

II. DECOMPOSITION OF THE FRICTION AND RANDOM FORCE

To extract the X_m coordinates [1], we first calculate the Fourier transform of the friction kernel:

$$\tilde{\gamma}(\omega) := \int_{-\infty}^{+\infty} \gamma(|\tau|) \exp\left(-\mathrm{i}\omega\tau\right) \mathrm{d}\tau. \tag{S5}$$

The numerical result is shown in Fig. S1. This is then fitted to the following rational function:

$$\tilde{\gamma}(\omega) = \frac{A\prod_{n=1}^{N}(\omega-\beta_n)(\omega-\beta_n^*)}{\prod_{m=1}^{M}(\omega-\alpha_m)(\omega-\alpha_m^*)}. \tag{S6}$$

The values of the fitting coefficients A, α_m, and b_m are determined by the least-squares fitting of the numerically obtained $\tilde{\gamma}(\omega)$. The degrees M and N of the denominator and the numerator, respectively, were determined by varying them and monitoring the coefficient of determination. With $M = 4$ and $N = 2$, we have obtained a satisfactory fit (coefficient of determination=0.996). In Fig. S1 the fitted curve is compared with the original $\tilde{\gamma}(\omega)$. The best-fitted values of the coefficients are listed in Table S1. Note $\alpha_3 = -\alpha_1^*$, $\alpha_4 = -\alpha_2^*$ and $\beta_2 = -\beta_1^*$. These coefficients must appear in such pairs because $\tilde{\gamma}(\omega)$ is an even function and takes real values on the real ω-axis [1].

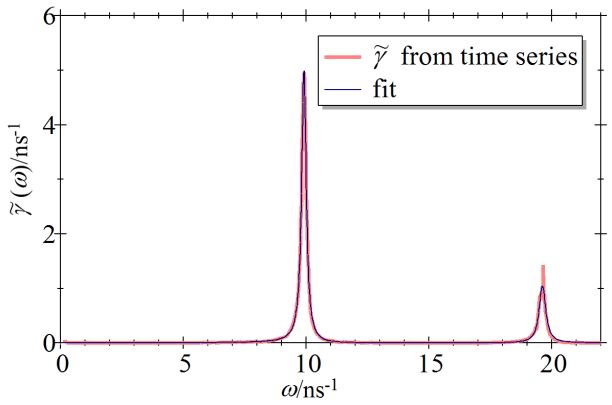

FIG. S1: Fourier transform of the friction kernel and its least-squares fit to the rational function.

TABLE S1: Optimized values of A, α_m, and β_n from the least-squares fitting of $\tilde{\gamma}(\omega)$ to the rational function.

	Re	Im
A/ns^{-5}	1.014272×10^{2}	0
α_1/ns^{-1}	9.921208×10^{0}	1.115815×10^{-1}
α_2/ns^{-1}	1.962735×10^{1}	1.559473×10^{-1}
α_3/ns^{-1}	-9.921208×10^{0}	1.115815×10^{-1}
α_4/ns^{-1}	-1.962735×10^{1}	1.559473×10^{-1}
β_1/ns^{-1}	-1.500655×10^{1}	2.422892×10^{0}
β_2/ns^{-1}	1.500655×10^{1}	2.422892×10^{0}

Eq. (S6) can be rearranged into the following form:

$$\begin{aligned}\tilde{\gamma}(\omega) =& \frac{A\prod_{n=1}^{N}(\omega-\beta_n)(\omega-\beta_n^*)}{\prod_{m=1}^{M}(\omega-\alpha_m)(\omega-\alpha_m^*)} \\ =& \sum_m \left\{\frac{-\mathrm{i}c_m}{\omega-\alpha_m} + \frac{\mathrm{i}c_m^*}{\omega-\alpha_m^*}\right\}. \end{aligned} \tag{S7}$$

The values of c_m can be calculated from A, α_m, and β_n by standard algebra. Through the inverse Fourier transformation, Eq. (S7) gives multi-exponential expression of the friction kernel:

$$\gamma(\tau) = \sum_m c_m \exp(\mathrm{i}\alpha_m|\tau|). \tag{S8}$$

The numerical values of c_m obtained from Table S1 are listed in Table S2. Note $c_3 = c_1^*$ and $c_4 = c_2^*$. Together with $\alpha_3 = -\alpha_1^*$ and $\alpha_4 = -\alpha_2^*$, these ensure that $\gamma(\tau)$ is real-valued.

TABLE S2: Coefficients in the multi-exponential expression of the friction kernel $\gamma(\tau)$.

	Re	Im
c_1/ns^{-2}	2.791906×10^{-1}	-6.341122×10^{-3}
c_2/ns^{-2}	8.418692×10^{-2}	9.492050×10^{-4}
c_3/ns^{-2}	2.791906×10^{-1}	6.341122×10^{-3}
c_4/ns^{-2}	8.418692×10^{-2}	-9.492050×10^{-4}

TABLE S3: Coefficients in the function $\tilde{g}(\omega)$.

	Re	Im
$b_1/\mathrm{ns}^{-3/2}$	-6.344190×10^{-2}	-1.664732×10^{-1}
$b_2/\mathrm{ns}^{-3/2}$	6.344190×10^{-2}	-9.740865×10^{-2}
$b_3/\mathrm{ns}^{-3/2}$	-6.344190×10^{-2}	1.664732×10^{-1}
$b_4/\mathrm{ns}^{-3/2}$	6.344190×10^{-2}	9.740865×10^{-2}

Following the recipe in [1], we next define the following function

$$\begin{aligned}\tilde{g}(\omega) :=& \sqrt{\frac{A}{2}} \frac{\prod_n \mathrm{i}(\omega - \beta_n^*)}{\prod_m \mathrm{i}(\omega - \alpha_m)} = -\sqrt{\frac{A}{2}} \frac{(\omega - \beta_1^*)(\omega - \beta_2^*)}{(\omega - \alpha_1)(\omega - \alpha_2)(\omega - \alpha_3)(\omega - \alpha_4)} \\ =& -\frac{\mathrm{i}b_1}{\omega - \alpha_1} - \frac{\mathrm{i}b_2}{\omega - \alpha_2} - \frac{\mathrm{i}b_3}{\omega - \alpha_3} - \frac{\mathrm{i}b_4}{\omega - \alpha_4},\end{aligned} \tag{S9}$$

where the coefficients b_m can be calculated from A, α_m, and β_n by standard algebra. The numerical values of b_m are listed in Table S3.

The integration kernel k_m in Eq. (13) is given by the inverse Fourier transform of the following function:

$$\tilde{k}_m(\omega) := \frac{-\mathrm{i}b_m}{\omega - \alpha_m} \frac{1}{\tilde{g}(\omega)}. \tag{S10}$$

For $m = 1$, for example,

$$\begin{aligned}\tilde{k}_1(\omega) =& \frac{-\mathrm{i}b_1}{\omega - \alpha_1}(-1)\sqrt{\frac{2}{A}} \frac{(\omega - \alpha_1)(\omega - \alpha_2)(\omega - \alpha_3)(\omega - \alpha_4)}{(\omega - \beta_1^*)(\omega - \beta_2^*)} \\ =& \mathrm{i}b_1 \sqrt{\frac{2}{A}} \frac{(\omega - \alpha_2)(\omega - \alpha_3)(\omega - \alpha_4)}{(\omega - \beta_1^*)(\omega - \beta_2^*)}.\end{aligned} \tag{S11}$$

This can be cast into the following form

$$\tilde{k}_1(\omega) = s_{11}\mathrm{i}\omega + s_{10} + \frac{\mathrm{i}\kappa_{11}}{(\omega - \beta_1^*)} + \frac{\mathrm{i}\kappa_{12}}{(\omega - \beta_2^*)}, \tag{S12}$$

TABLE S4: Coefficients in the function $k_m(\tau)$.

	Re	Im
s_{11}/ns	-8.908690×10^{-3}	-2.337664×10^{-2}
s_{10}	1.849823×10^{-1}	-2.115626×10^{-1}
κ_{11}/ns^{-1}	-2.873538×10^{-1}	-8.215834×10^{-1}
κ_{12}/ns^{-1}	-1.178703×10^{-1}	-3.836437×10^{0}
s_{21}/ns	8.908690×10^{-3}	1.367840×10^{-2}
s_{20}	3.150177×10^{-1}	1.033858×10^{-1}
κ_{21}/ns^{-1}	2.420398×10^{-1}	-3.306199×10^{-1}
κ_{22}/ns^{-1}	1.631843×10^{-1}	2.684234×10^{0}

where the coefficients s_{11}, s_{10}, κ_{11}, and κ_{12} can be calculated from b_1, A, α_2, α_3, α_4, β_1, and β_2. The numerical values of the coefficients are shown in Table. S4. By the inverse Fourier transform, we obtain

$$k_1(\tau) = s_{11}\delta'(\tau) + s_{10}\delta(\tau) + \kappa_{11}\Theta(-\tau)\exp(\mathrm{i}\beta_1^*\tau) + \kappa_{12}\Theta(-\tau)\exp(\mathrm{i}\beta_2^*\tau), \tag{S13}$$

where δ is the Dirac delta function, δ' is its derivative, and Θ is the Heaviside step function:

$$\Theta(\tau) = \begin{cases} 1 & (\tau > 0) \\ 0 & (\tau < 0) \end{cases}. \tag{S14}$$

The convolution of $\xi(t)$ with $k_1(\tau)$ as in Eq. (13) can be calculated by

$$\begin{aligned} &\int_0^{+\infty} k_1(t-t')\xi(t')\mathrm{d}t' \\ &= -s_{11}\dot{\xi}(t) + s_{10}\xi(t) + \kappa_{11}\int_0^{+\infty}\exp(-\mathrm{i}\beta_1^*\tau)\xi(t+\tau)\mathrm{d}\tau + \kappa_{12}\int_0^{+\infty}\exp(-\mathrm{i}\beta_2^*\tau)\xi(t+\tau)\mathrm{d}\tau, \end{aligned} \tag{S15}$$

where we have introduced a change of integration variable $t' \mapsto \tau = t' - t$.

Similarly, k_2 can be expressed as

$$k_2(\tau) = s_{21}\delta'(\tau) + s_{20}\delta(\tau) + \kappa_{21}\Theta(-\tau)\exp(\mathrm{i}\beta_1^*\tau) + \kappa_{22}\Theta(-\tau)\exp(\mathrm{i}\beta_2^*\tau), \tag{S16}$$

The numerical values of the coefficients are shown in Table S4. Note that, due to $b_3 = b_1^*$, $b_4 = b_2^*$, $\alpha_3 = -\alpha_1^*$, $\alpha_4 = -\alpha_2^*$, and $\beta_2 = -\beta_1^*$, we have $k_3(\tau) = k_1(\tau)^*$ and $k_4(\tau) = k_2(\tau)^*$.

Together with $c_3 = c_1^*$ and $c_4 = c_2^*$, it is concluded that $X_3 = X_1^*$ and $X_4 = X_2^*$. Thus X_3 and X_4 are just complex conjugate of X_1 and X_2, respectively. We therefore show only X_1 and X_2 in Fig. 3.

III. CASE OF A FRICTION KERNEL WITH DELTA FUNCTION

Here we consider a case where the white noise is added also on the q_1-direction, that is,

$$\begin{aligned} m\ddot{q}_1 =& -\frac{\partial V}{\partial q_1} - \gamma_1 \dot{q}_1 + \xi_{\mathrm{w}1}(t) \\ m\ddot{q}_2 =& -\frac{\partial V}{\partial q_2} - \gamma_2 \dot{q}_2 + \xi_{\mathrm{w}2}(t), \\ \langle \xi_{\mathrm{w}i}(t)\xi_{\mathrm{w}j}(t')\rangle =& 2k_{\mathrm{B}}T\gamma_i \delta_{ij}\delta(t-t'), \end{aligned} \tag{S17}$$

where δ_{ij} is Kronecker's delta, instead of Eq. (4). In the following calculation we use $\gamma_1/m = \gamma_2/m = 0.2$ ns^{-1}. In this case, reflecting the fact that q_1 is directly exposed to the white noise, the friction kernel $\gamma(\tau)$ in the GLE for q_1 also contains the delta function component:

$$\gamma(\tau) = 2\gamma^{\mathrm{w}}\delta(\tau) + \gamma^{\mathrm{c}}(\tau), \tag{S18}$$

where $\gamma^{\mathrm{c}}(\tau)$ is a smooth function of τ. The coefficient γ^{w} before the delta function is actually the same with γ_1/m in Eq. (S17), but we pretend that we do not know its value *a priori* in the time series analysis. Since the delta function needs a special treatment in the numerical calculation, we make some modification to the formulation in Sec. III A for obtaining the friction kernel. Substituting Eq. (S18) into Eq. (9) yields

$$\dot{C}(t) = F(t) - \gamma^{\mathrm{w}}C(t) - \int_0^t \gamma^{\mathrm{c}}(t-t')C(t')\mathrm{d}t'. \tag{S19}$$

Some fundamental algebra then results in the following expressions:

$$\begin{aligned} \gamma^{\mathrm{w}} =& \frac{F(0) - \dot{C}(0)}{C(0)}, \\ C(0)\gamma^{\mathrm{c}}(t) =& \dot{A}(t) - \gamma^{\mathrm{w}}A(t) - \int_0^t \gamma^{\mathrm{c}}(t-t')K(t')\mathrm{d}t', \end{aligned} \tag{S20}$$

with

$$\begin{aligned} K(t) :=& \dot{C}(t) + \gamma^{\mathrm{w}}C(t), \\ A(t) :=& F(t) - K(t), \end{aligned} \tag{S21}$$

which give γ^{w} and $\gamma^{\mathrm{c}}(t)$ in terms of the correlation functions from time series.

The Fourier transform of γ is given by

$$\tilde{\gamma}(\omega) = 2\gamma^{\mathrm{w}} + \tilde{\gamma}^{\mathrm{c}}(\omega), \tag{S22}$$

which, when expressed in the form of rational function, now contains the same power in the numerator and denominator due to the constant term $2\gamma^{\mathrm{w}}$:

$$\tilde{\gamma}(\omega) = \frac{A\prod_{n=1}^{M}(\omega-\beta_n)(\omega-\beta_n^*)}{\prod_{m=1}^{M}(\omega-\alpha_m)(\omega-\alpha_m^*)}. \tag{S23}$$

Eq. (S9) is accordingly modified to

$$\tilde{g}(\omega) = b_0 - \frac{\mathrm{i}b_1}{\omega-\alpha_1} - \frac{\mathrm{i}b_2}{\omega-\alpha_2} - \frac{\mathrm{i}b_3}{\omega-\alpha_3} - \frac{\mathrm{i}b_4}{\omega-\alpha_4}, \tag{S24}$$

because in the present case we have obtained $M = 4$ by fitting the numerically obtained $\tilde{\gamma}^{\mathrm{c}}(\omega)$. The integration kernels can be given by

$$\begin{aligned} \tilde{k}_0(\omega) =& b_0\frac{1}{\tilde{g}(\omega)}, \\ \tilde{k}_m(\omega) =& \frac{-\mathrm{i}b_m}{\omega-\alpha_m}\frac{1}{\tilde{g}(\omega)} \qquad\qquad (m \geq 1). \end{aligned} \tag{S25}$$

And the X_m variables are

$$\begin{aligned} X_0(t) =& -\gamma^{\mathrm{w}}\dot{q}_1(t) + \int_0^{+\infty} k_0(t-t')\xi(t')\mathrm{d}t', \\ X_m(t) =& -\int_0^t c_m \exp\left(\mathrm{i}\alpha_m(t-t')\right)\dot{q}_1(t')\mathrm{d}t' + \int_0^{+\infty} k_m(t-t')\xi(t')\mathrm{d}t'. \end{aligned} \tag{S26}$$

Fig. S2 shows an example time series of the model given by Eq. (S17) together with the X_m coordinates calculated from $q_1(t)$. Sign of the state transition is still appreciable from the amplitudes of X_m's, though it became less clear due to the existence of the white noise. This may be explained as follows. By definition, the calculation of X_m extracts the environmental dynamical mode with the constant frequency α_m. (Note that, in Eq. (14), α_m appears as the frequency and time constant for the motion of X_m). Since white noise has uniform power spectrum, it contains oscillation of every frequency in a sense. Therefore the components of ξ_{w1} that oscillates with frequency $\mathrm{Re}\alpha_m$ enters into the value of X_m. In molecular systems, a while noise, if there exists, originates from some rapidly decaying motion in the environment. The X_m coordinates would be collective coordinates that partly include such motions.

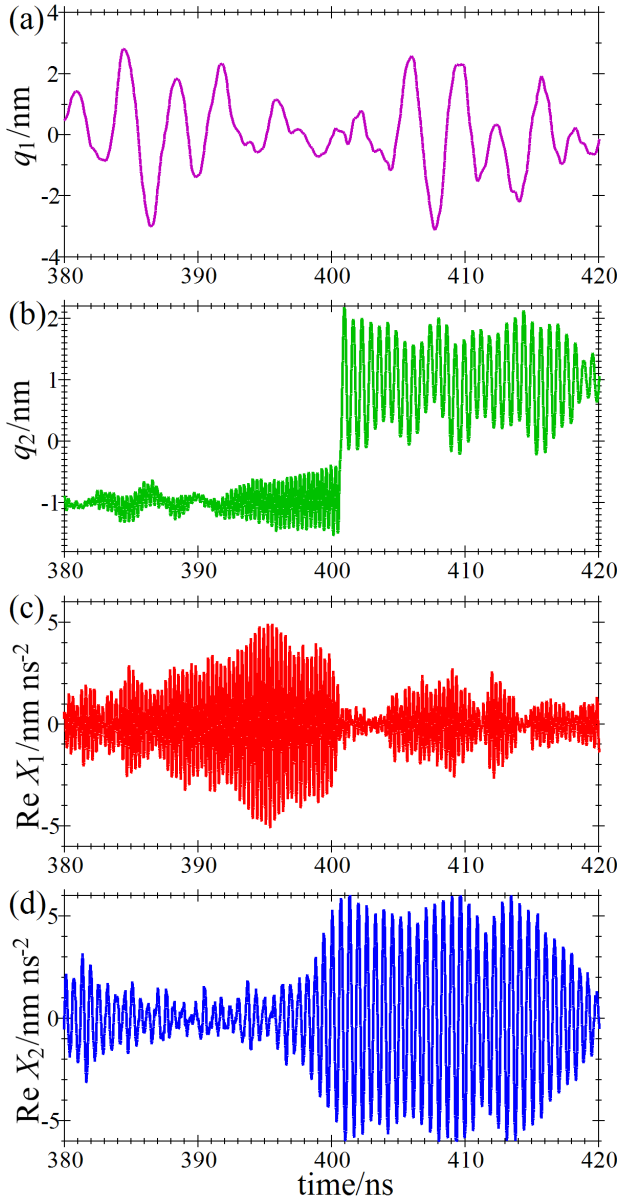

FIG. S2: Typical time evolution of the model system where a white noise is added in the q_1-direction. Panels (a) and (b) show the time series of the position coordinates q_1 and q_2, respectively. Panels (c) and (d) show the time series of the effectively expressed environmental modes obtained by the GLE-based analysis on the time series $q_1(t)$ only.

[1] S. Kawai, J. Chem. Phys. **143**, 094101 (2015).