Verification of time-reversibility requirement for systems satisfying the Evans-Searles Fluctuation Theorem

Emil Mittag, Denis J. Evans and Stephen R. Williams

Research School of Chemistry, Australian National University, Canberra ACT 0200, Australia

The Evans-Searles Fluctuation Theorem (ESFT) has been shown to be applicable in the near and far from equilibrium regimes for systems with both constant and time-dependent external fields. The derivations of the ESFT have assumed that the external field has a definite parity under a time-reversal mapping. In the present paper we confirm that the time-reversibility of the system dynamics is a necessary condition for the ESFT to hold. The manner in which the ESFT fails for systems that are not time-reversible is presented and results are shown which demonstrate that systems which fail to satisfy the ESFT may still satisfy the Crooks relation.

In 1876, Joseph Loschmidt was the first to observe that since mechanics is time-reversible, events that happen in one direction should be equally likely to occur in the opposite direction, given the application of a time-reversal mapping on the velocities of the particles in the system.[1] In order to try to reconcile this view with that of Boltzmann's and the second law of thermodynamics, much discussion ensued.[2] It was more than one hundred years later that this paradox was finally resolved mathematically with the discovery of the fluctuation theorem, which was first demonstrated numerically by Evans, Cohen and Morriss in 1993[3] and proved by Evans and Searles in 1994.[4]

The Evans-Searles fluctuation theorem (ESFT) is a mathematical expression that gives the probability that second-law violating events may be observed for small systems at short time-scales. The mathematical form of the ESFT is

$$\frac{p_F(\overline{\Omega}_t = A)}{p_F(\overline{\Omega}_t = -A)} = \exp(At), \tag{1}$$

where

$$\overline{\Omega}_t t \equiv \int_0^t ds\ \Omega(\mathbf{\Gamma}(s)) = \ln\left[\frac{f(\mathbf{\Gamma}(0),0)}{f(\mathbf{\Gamma}(t),0)}\right] - \int_0^t ds\ \Lambda(\mathbf{\Gamma}(s)). \tag{2}$$

is the general dissipation function[5, 6], $p_F(x)$ is the probability of observing an event x, $f(\mathbf{\Gamma}(0),0)$ is the phase-space distribution function at a point $\mathbf{\Gamma}(p(0),q(0)) = \mathbf{\Gamma}(0)$ in phase

space at time zero. $f(\mathbf{\Gamma}(t), 0)$ is the distribution about a point $\mathbf{\Gamma}(t)$, also at time zero. Λ is the phase space compression factor.[7] Close to equilibrium the average of the dissipation function is to leading order in the external field, equal to the irreversible entropy production calculated using linear irreversible thermodynamics.

The ESFT is important for several reasons: it is one of the few mathematical expressions that is valid arbitrarily far from equilibrium, its validity has been confirmed in several laboratory experiments,[8–11] it can be used to derive the Green-Kubo relations, it can be used to derive the Second Law inequality[12], it is closely related to another important mathematical relation in nonequilibrium statistical mechanics, the Jarzynski Equality[13], and it is valid in many ensembles.[14] Since it was first proved, much work has been done to extend the ESFT. Both steady-state and transient versions of the theorem have been studied.[6] Other work done with the ESFT has shown it to be valid for stochastic systems[15, 18–20] and a quantum version of the theorem has been derived.[16, 17]

A similar relation derived by Crooks[19, 20], relates the probability, p_F, of the dissipation function taking on a positive value, A, for a trajectory under the influence of an external field in the forward direction $F_e(t)$, $0 < t < \tau$, where τ is the maximum averaging time, to the probability, p_R, of the dissipation function taking on a value, $-A$, for a trajectory in the reverse direction with the waveform of the field reversed, $F_e(\tau - t)$, $0 < t < \tau$,

$$\frac{p_F(\overline{\Omega}_t = A)}{p_R(\overline{\Omega}_t = -A)} = \exp(At). \tag{3}$$

Note that this relation is only valid if the free energy difference for the forward and reverse states is zero. In this case, the Crooks work function is equal to the general dissipation function and the Crooks relation (CR) takes the form shown in Equation (3). For a more detailed discussion on the Crooks relation, the interested reader is referred to [19, 20].

It has been shown that the ESFT is valid for non-equilibrium systems with constant[5] and time-dependent external fields[21, 22] when the system is time-reversible. The primary difference between the ESFT, Equation 1, and the CR is that for the CR a time-reversal mapping of the external field takes place, while for the ESFT no such mapping is applied.

In this paper we confirm that the time-reversibility requirement is a necessary condition for the ESFT to hold. Using a test system we show that if the system is not time-reversible, and therefore not ergodically consistent, the ESFT is not satisfied. We also show that the Crooks relation is still verified for systems that are not time-reversible, as expected.

I. TIME REVERSIBILITY

The time reversibility requirement is intrinsic to the ESFT derivation[6] since time averages of the general dissipation function are calculated from trajectories in phase space that are related to one another via a time-reversal mapping, ie a time-reversal symmetry for the system exists.

In Equation 1, we see that the denominator considers the probability of observing time-averaged values of the general dissipation function that have a sign opposite to that predicted by the second law. In order for this to be possible, there must be a finite probability of finding trajectories in phase space whose response averaged over the length of the trajectory have such values. From the derivation of the ESFT[6], we know that phase trajectories must exist in pairs (ie be time-reversal maps of one another) if the system is time reversible. We emphasise that the ESFT is a mathematical expression involving the averaging of the general dissipation function over trajectories where the external field has a definite parity under a time-reversal mapping, while the Crooks relation involves the explicit mapping of the external field (ie $F_e(t) = F_e(\tau - t)$, $0 < t < \tau$) when computing the denominator as compared to the numerator of Equation (3).

Consider a trajectory that is propagated from a point $\mathbf{\Gamma}(0)$ in phase space for a time t, which ends at a phase point $\mathbf{\Gamma}(t)$. At the endpoint of the trajectory we can apply a time reversal mapping, such that $M^T[\mathbf{\Gamma}(t)] = \mathbf{\Gamma}^*(0)$, where M^T is a time reversal mapping that changes the signs of all of the system momenta, i.e. $M^T[\mathbf{\Gamma}(p,q)] = \mathbf{\Gamma}(-p,q)$. In Equation 2 we see that the distribution functions about the starting and ending points of a trajectory in phase space are used to define the general dissipation function. It is therefore imperative that there be a finite probability that one can find the ending points of phase trajectories at time zero in order to be able to find starting points for trajectories that are time-mapped equivalents of trajectories that would generate time averages of the general dissipation function in the numerator of Equation 1.

Systems which satisfy the relation $f(\mathbf{\Gamma}(t), 0) \neq 0$, which indicates that there must be a finite probability of finding trajectory endpoints in phase space at time zero, are referred to as ergodically consistent.[6]

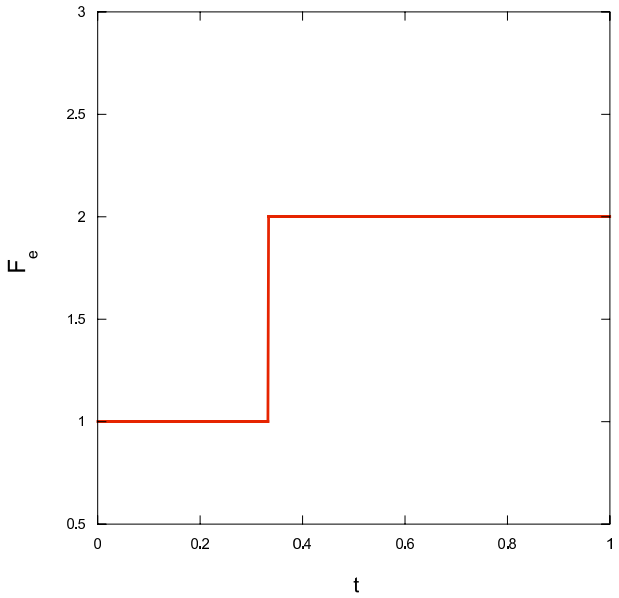

FIG. 1: The waveform of a time-dependent external field that has no parity. The system to which this field is applied is therefore non-time-reversible.

II. NUMERICAL SIMULATIONS

In order to ensure that the test system is not time-reversible, we use a waveform as shown in Figure 1. This waveform is not symmetric over the length of the transient trajectory, which results in the dynamics being non-time-reversible, since upon application of the time-reversal mapping at the end of the trajectory of length τ, $M^T(F_e(t)) \neq \pm F_e(\tau - t)$, where $F_e(t)$ is the time-dependent external field. The external field for the waveform shown has the mathematical form

$$F_e(t) = \begin{cases} 1.0 & t < \frac{\tau}{3} \\ 2.0 & t \geq \frac{\tau}{3}. \end{cases} \tag{4}$$

The phase-space distribution function of our system at time $t = 0$ (the equilibrium distribution) is $f(\mathbf{\Gamma}, 0) \sim \exp(-\beta[H_0 + \frac{1}{2}Q\zeta^2])$, where Q is the effective mass of the heat bath, $\beta = \frac{1}{k_B T} = \frac{2K}{dN} + \mathcal{O}(\frac{1}{N})$, ζ is the Nosé-Hoover thermostat multiplier, k_B is Boltzmann's constant, K is the kinetic energy, d is the system dimension and N is the number of particles.

Colour diffusion is used for the non-equilibrium dynamics of our system. The system Hamiltonian is $H(\mathbf{\Gamma}) = H_0(\mathbf{\Gamma}) + F_c(t) \sum_{i=1}^{N} c_i x_i$, where $c_i = (-1)^i$ is the colour field coupling constant, $F_c(t)$ is the time-dependent colour field of the system and $H_0(\mathbf{\Gamma}) = p_i^2/2m + \Phi(\mathbf{q})$ is the system's internal energy. Here $\Phi(\mathbf{q})$ is the potential that models the potential energy

of the inter-particle interactions. We use the WCA potential[23] to model these interactions, i.e. $\Phi(\mathbf{q}) = \sum_{i=1}^{N-1} \sum_{j>i}^{N} \phi(|q_i - q_j|)$, $\phi(q) = 4\epsilon(q^{-12} - q^{-6})$ if $q < 2^{1/6}$ or zero otherwise.

The equations of motion for the system are

$$\begin{aligned} \dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m} \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \mathbf{i}c_i F_e(\phi) - \zeta \mathbf{p}_i \\ \dot{\zeta} &= \frac{1}{Q}\left[\sum \frac{p_i^2}{m} - (g+1)k_B T\right] \\ \dot{\phi} &= \omega, \end{aligned} \tag{5}$$

where $\mathbf{F}_i = \frac{\partial \Phi(\mathbf{q})}{\partial \mathbf{q}_i}$, $\phi(t+P) = \phi(t)$, $g = 6N + \mathcal{O}(1)$ and ω is the frequency of the external time-dependent field.

The system Hamiltonian is $\dot{H}_0^{ad} \equiv -JVF_e$[7], where the superscript "ad" denotes that the time derivative is taken in the absence of a thermostat. Here the dissipative flux, J, is given by $J = V^{-1} \sum_{i=1}^{N} c_i p_{xi}$. We substitute the distribution function for the system into Equation (2) to get the form of the general dissipation function for this system, $\overline{\Omega}_t = -\beta \frac{1}{t} \int_0^t ds\ J(s)F_e(s)V = -\beta \overline{[J(t)F_e(t)]}_t V$

III. RESULTS AND DISCUSSION

Figure 2 shows a histogram of the ratio of logarithm probabilities of positive and negative values for the general dissipation function (i.e. the logarithm of the left hand side of equation 1). For systems that have external fields with definite parity, such a histogram is approximately Gaussian. Here, however, we see that the the distribution is far from Gaussian. If we take points on this graph that are equidistant from the y-axis, they correspond to conjugate values of the general dissipation function (i.e. A and $-A$ from Equation 1). Clearly the histogram bins to the far right side of the plot have more samples than they would if the curve were Gaussian. This departure from a Gaussian distribution is indicative that the system is not time-reversible, since the current trace in bins equidistant from the y-axis and of equal width are not related to one another via a time-reversal mapping, as shown in Figure 3. In Reference [22] it was shown that for time-reversible systems, the current traces for histogram bins that are equidistant from the y-axis and of equal width

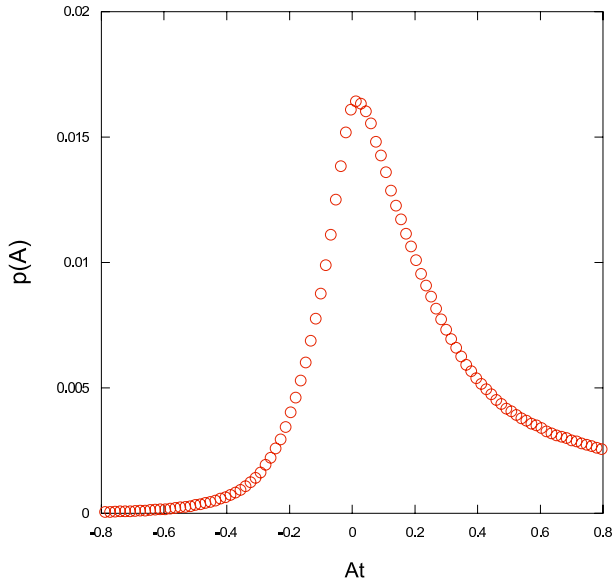

FIG. 2: Histogram of the results of a test of the ESFT for a system with an external field of no parity, which is thus not time-reversible.

are time-reversal maps of one another.

Figure 4 shows the results of a test of the ESFT. The main figure is a plot of the deviation of the numerical results from the predicted results with error bars. The simulation data were accumulated from 20 simulations, each of 5 million trajectories. The standard error was calculated from these 20 simulations and used for the error bars shown in the figure. The standard error is defined by $\sigma/\sqrt{N_r}$, where σ is the standard deviation and N_r is the number of simulations runs, which is 20 in this case. The results indicate that within the error bars, the numerical results do not converge to the predicted one. The points that do not fall exactly on what appears to be a regular curve are due to insufficient sampling. If the simulation is allowed to run for a longer period of time, such points gradually converge to the curve as more samples are accumulated in the associated histogram bin of the histogram shown in Figure 2. The inset of Figure 4 shows the numerical and predicted ESFT curves. The line is that predicted by the ESFT and the circles are the computational results. We observe that there is a slight deviation of the numerical results to what is predicted.

Figure 5 shows a plot of the test of the Crooks relation. The large plot in the figure shows the deviation of the numerical test of CR from what is predicted, along with the associated error bars calculated from 20 runs as described in the discussion of the ESFT results, above. Note that the deviation is nearly zero in the x-axis range around zero, which corresponds

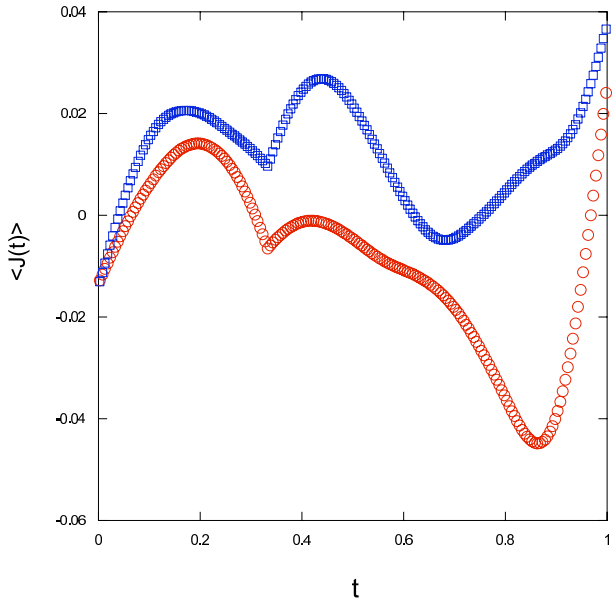

FIG. 3: Current response in two histogram bins of equal width immediately to the left (circles) and right (squares) of the y-axis of Figure 2. If the system were time-reversible, the current traces in these two bins would be related to one another via the time-reversal mapping, which is not the case for this system.

to a location near the centre of the histogram shown in Figure 2. Since this part of the histogram has the highest number of samples, the averages calculated in the histogram bins near the y-axis are approximately equal to the values that are predicted by the CR. As one moves further away from the peak of the histogram, the histogram bins have fewer samples and the this lack of statistics means that rare events can cause a deviation in the numerical results from what is predicted. As more samples are accumulated, the numerical results will approach the predicted one and the curve shown in Figure 5 will approach the x-axis over its entire range.

IV. CONCLUSIONS

The results of the numerical tests show that the ESFT is not satisfied by systems with external fields with no parity, such as that used in the test system of this study. Figure 3 demonstrated that there is no time-reversible relation between the currents for conjugate bins of the histogram shown in Figure 2. It was shown in Reference [22] that for systems with external fields of definite parity such a relation between currents exists. Since there

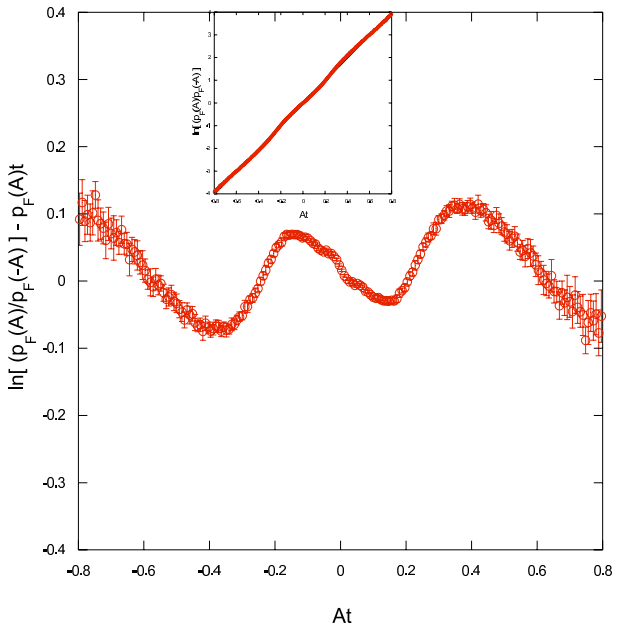

FIG. 4: A test of the ESFT. The curve shows the difference between the numerically obtained results and that predicted by the ESFT. The inset shows the predicted line and numerical results (circles). The deviation that is difficult to see in the inset is clearly visible in the difference plot. It is clear that the ESFT is not satisfied for this system.

FIG. 5: A test of the Crooks relation. The curve shows the difference between the numerically obtained results and that predicted by the Crooks relation. The inset shows the predicted line and numerical results (circles). There is almost no difference between the predicted and observed curves, as expected. The small deviation between the two that is visible is due to statistics and vanishes as more samples are obtained in the simulation. Clearly the CR is satisfied for this system.

is no relation between the two responses of these bins, we can conclude that there is a difficulty in sampling initial phase points that will lead to trajectories that are related to one another via the time-reversal mapping. The inability to sample such starting points from our distribution indicates that the system is not ergodically consistent. Since the systems of our previous study in Reference [22] were ergodically consistent, one can only conclude that this ergodic inconsistency of the test system used in this work is due to the form of the external field, since the system dynamics of the two studies is identical.

Figure 4 shows that the ESFT is not satisfied for the system we tested. The error bars in that figure indicate that no matter how long the simulation is allowed to run the results

will never converge to the theoretical prediction. However, Figure 5 clearly shows that the Crooks relation is satisfied by the numerical results obtained. The reason that the CR is satisfied and the ESFT is not lies in the derivation of the CR, which entails the time-reversal mapping of the waveform of the external field. This has the effect that with the time-reversal mapped field the system is able to sample phase points that are unreachable with the unmapped field. By combining the results from the forward and reverse field simulations, as required by the CR, one is able to test the CR in a similar way as to if one had performed a time-reversible simulation that is ergodically consistent and therefore able to sample all initial phase points that lead to trajectories that are conjugate pairs. We therefore conclude that systems without time-reversal symmetry cannot satisfy the ESFT because they are not ergodically consistent.

---

[1] J. Loschmidt, *Sitzungsber. der kais. Akad. d. W. math. naturw. II*, **73**, 128 (1876)

[2] Broda, E. *Ludwig Boltzmann: Man—Physicist—Philosopher* Woodbridge: Ox Bow Press, 1983

[3] Evans, D.J.; Cohen, E.D.G.; Morriss, G., *Phys. Rev. Lett.*, **71**, 2401 (1993)

[4] Evans, D.J.; Searles, D.J. *Phys. Rev. E*, **50**, 1645 (1994)

[5] Evans, D.J.; Searles, D.J. *Phys. Rev. E*, **52**, 5839 (1995)

[6] Evans, D.J.; Searles, D.J. *Avd. Phys.*, **51**, 1529 (2002)

[7] Evans, D.J.; Morriss, G. *Statistical Mechanics of Nonequilibrium Liquids*, London: Academic Press, 1990

[8] Wang, G.M.; Sevick, E.M.; Mittag, E.; Searles, D.J., Evans, D.J. *Phys. Rev. Lett.*, **89**, 050601 (2002)

[9] Reid, J.C.; Carberry, D.M.; Wang, G.M.; Sevick, E.M.; Evans, D.J.; Searles, D.J. *Phys. Rev. E*, **70**, 016111 (2004)

[10] Canberry, D.M.; Reid, J.C.; Wang, G.M.; Sevick, E.M.; Searles, D.J.; Evans, D.J. *Phys. Rev. Lett.*, **92**, 140601 (2004)

[11] Liphardt, J.; Dumont, S.; Smith, S.B.; Tinoco, I.; Bustamante, C. *Science*, **296**, 1832 (2002)

[12] Williams, S.R.; Evans, D.J.; Mittag, E *submitted for publication*

[13] Jarzynski, C. *Phys. Rev. Lett.*, **78**, 2690 (1997)

[14] Searles, D.J.; Evans, D.J. *J. Chem. Phys.*, **113**, 3503 (2000)

[15] Kurchan, J. *J. Phys. A*, **31**, 3719 (1998)

[16] Kurchan, J. *xxx.lanl.gov/abs/cond-mat/0007360* (2001)

[17] Monnai, T. *Phys. Rev. E*, **72**, 027102 (2005)

[18] Lebowitz, J.L.; Spohn, H. *J. Stat. Phys.*, **95**, 333 (1999)

[19] Crooks, G. *Phys. Rev. E*, **60**, 2721 (1999)

[20] Crooks, G. *Phys. Rev. E*, **61**, 2361 (2000)

[21] Evans, D.J.; Searles, D.J. *Phys. Rev. E*, **53**, 5808 (1996)

[22] Mittag, E.; Evans, D.J. *Phys. Rev. E*, **67**, 026113 (2003)

[23] Weeks, J.W.; Chandler, D.; Andersen, H.C. *J. Chem. Phys.*, **54**, 5237 (1971)