Independence of the Transient Fluctuation Theorem to Thermostatting Details

Stephen. R. Williams[1], Debra. J. Searles[2] and Denis. J. Evans[1 a]

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

[2]School of Science, Griffith University, Brisbane, Qld 4111 Australia

The Fluctuation Theorems show how macroscopic irreversibility arises from time reversible microscopic dynamics. They have been confirmed in computer simulations and in laboratory experiments. The standard proofs of the Transient Fluctuation Theorems involve the use of time reversible deterministic thermostats to control the temperature of the system of interest. These mathematical thermostats do not occur in Nature. However, since in a gedanken experiment the thermostatting regions can be removed arbitrarily far from the system of interest, it has been argued that the precise details of the thermostat cannot be important and that the resulting Fluctuation Theorems apply to natural systems. In this paper we give a detailed analysis showing how the Fluctuation Theorem is independent of the precise mathematical details of the thermostatting mechanism for an infinite class of fictitious time reversible deterministic thermostats. Our analysis reinforces the implications of the gedanken experiment and implies that thermostats used in the derivations of Fluctuation

---

[a] To whom correspondence should be addressed. e-mail: evans@rsc.anu.edu.au

Theorems are a convenient but ultimately irrelevant device.

PACS numbers: 05.70.Ln, 05.40.-a

Introduction

If one wants to treat nonequilibrium dissipative systems that are maintained at a constant temperature, some form of thermostatting mechanism must be employed. A natural procedure would be to surround the nonequilibrium system of interest with an extremely large inert block of heat conducting material - the thermal reservoir. If the heat capacity of the reservoir is much greater than that of the system of interest, then we can expect that the composite system may relax to a nonequilibrium quasi steady state in which the rate of temperature rise for the composite system is so small that it can be regarded as being zero. This approach is mathematically and computationally complex.

Twenty years ago Hoover et al. [1] and Evans [2] independently but simultaneously developed time reversible deterministic thermostats to enable convenient and efficient computer simulations of thermostatted dissipative systems. These thermostats do not exist in Nature but nonequilibrium statistical mechanics has been used to prove that under specific circumstances thermodynamic properties and transport coefficients computed from simulations using these thermostats are essentially exact [3].

The development of fictitious mathematical thermostats and algorithms for simulating transport coefficients of nonequilibrium thermodynamic systems has lead to an enormous advancement of nonequilibrium statistical mechanics [3]. These two developments have allowed the mathematical apparatus of dynamical systems theory to be brought to bear on statistical mechanics.

One of the major achievements in this field was the discovery [4] and subsequent proof [5, 6] of Fluctuation Theorems. These theorems showed for the first time how macroscopic irreversibility arises from time reversible microscopic equations of motion. A number of experimental tests have confirmed the predictions of these Fluctuation Theorems in the laboratory [7, 8].

The standard proofs of the Transient Fluctuation Theorems (TFTs) [6]involve the use of time reversible deterministic thermostats to control the temperature of the system of interest. These mathematical thermostats do not occur in Nature. In the present paper we give a detailed analysis showing how the Fluctuation Theorem is independent of the precise mathematical details of the thermostatting mechanism for an infinite class of fictitious time reversible deterministic thermostats. Our analysis reinforces the view that the thermostats used in the derivations of Fluctuation Theorems are a convenient but ultimately irrelevant device [6].

Theory

We denote the phase space $\boldsymbol{\Gamma} \equiv (\mathbf{q}, \mathbf{p})$ of our system to consist of $2DN$ variables, $\mathbf{q}$ being the position vector, $\mathbf{p}$ the momentum vector, D the Cartesian dimension of the system and N the number of particles. We divide the system into N_w wall or reservoir particles, which are thermostatted, and are in contact with N_s particles which are not thermostatted. These latter particles comprise the nonequilibrium system of interest. They experience a dissipative field, $\mathbf{F}_e$ which does work on the system of interest driving it away from equilibrium. This work is gradually converted (irreversibly) into heat. This heat is removed by the thermostatted wall particles, allowing a nonequilibrium steady state to develop at sufficiently long times after the application of the dissipative field.

Consider the family of thermostats which, following Gauss' Principle of Least Constraint [3, 9, 10], maintain the μ^{th} moment of the momentum distribution for the wall particles,

$$g(\mu) = \sum_{i=1}^{N_w} (\mathbf{p}_i \cdot \mathbf{p}_i)^{\frac{1}{2}\mu} = \sum_{i=1}^{N} s_i (\mathbf{p}_i \cdot \mathbf{p}_i)^{\frac{1}{2}\mu} = g_0(\mu), \tag{1}$$

as a constant of the motion. In Eq. (1) $\mathbf{p}_i$ denotes the momentum vector for the i^{th} particle and $s_i = 1 \;\; for \;\; i \le N_w$ while $s_i = 0 \;\; for \;\; i > N_w$ where we choose μ to be any real number with the provision $\mu > 1$ for $D = 1$ and $\mu > 1 - D/2$ for all higher dimensions, thus ensuring the averages of all variables used in this paper converge. Using Gauss' Principle of Least Constraint [3, 9, 10] we obtain the equations of motion for the system subject to the thermostatting constraint $g(\mu)$,

$$\dot{\mathbf{q}}_i = \mathbf{p}_i / m + (1 - s_i)\mathbf{C}_i(\Gamma) \cdot \mathbf{F}_e$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i(\mathbf{q}) + (1 - s_i)\mathbf{D}_i(\Gamma) \cdot \mathbf{F}_e - (\mathbf{p}_i \cdot \mathbf{p}_i)^{\frac{1}{2}\mu - 1} s_i \alpha \mathbf{p}_i \tag{2}$$

$$\alpha = \frac{\sum_{i=1}^{N_w} (\mathbf{p}_i \cdot \mathbf{p}_i)^{\frac{1}{2}\mu - 1} \mathbf{p}_i \cdot \mathbf{F}_i}{\sum_{i=1}^{N_w} (\mathbf{p}_i \cdot \mathbf{p}_i)^{\mu - 1}}.$$

In this equation the dissipative field $\mathbf{F}_e$ couples to the system via the dyadic phase functions $\mathbf{C}_i(\Gamma)$ and $\mathbf{D}_i(\Gamma)$. The phase space compression factor for the full system is defined as [3]

$$\Lambda(\Gamma) = \frac{\partial}{\partial \Gamma} \cdot \dot{\Gamma}, \tag{3}$$

and determines the evolution of an infinitesimal volume element of phase space δV_Γ surrounding a trajectory $\Gamma(t)$,

$$\delta V_\Gamma(\Gamma(t)) = \exp\left[\int_0^t \Lambda(\Gamma(s))\, ds\right] \delta V_\Gamma(\Gamma(0)). \tag{4}$$

Assuming in the absence of the thermostatting particles that the phase space compression factor is zero (AIΓ) [3] and ignoring terms involving the momentum dependence of α (which is of $O(1/N_w)$ relative to the terms involving $\mathbf{p}_i$ explicitly) we obtain the phase space compression factor for the μ^{th} thermostat

$$\Lambda\left(\Gamma,\mu\right)=-\alpha(D+\mu-2)\sum_{i=1}^{N_w}\left(\mathbf{p}_i\cdot\mathbf{p}_i\right)^{\frac{1}{2}\mu-1}. \quad (5)$$

If the ratio N_w/N_s is large enough, the perturbation caused to the wall particles by the nonequilibrium system of interest can be regarded as infinitesimal. In this limit, the thermostatting particles can be described by the appropriate equilibrium distribution.

Under these conditions we will now show that the space compression factor is solely determined by the heat transfer rate from the total system by the thermostat, and the absolute temperature of that thermostat. Furthermore the phase space compression factor is independent of the momentum moment, μ, that is constrained by the thermostat. The rate of change of the internal energy may be split into an adiabatic (or work) component $\dot{W}\left(t\right)$ and a heat transfer $\dot{Q}_l\left(t\right)$ component,

$$\dot{H}\left(t\right)=\dot{W}\left(t\right)+\dot{Q}_l\left(t\right)$$

$$\dot{Q}_l\left(t\right)=\frac{-\alpha\left(t\right)}{m}\sum_{i=1}^{N_w}\left(\mathbf{p}_i\cdot\mathbf{p}_i\right)^{\frac{1}{2}\mu} \quad (6)$$

$$\dot{W}\left(t\right)=-\sum_{i=N_w+1}^{N}\left[\frac{\mathbf{p}_i}{m}\cdot\mathbf{D}_i-\mathbf{F}_i\cdot\mathbf{C}_i\right]\cdot\mathbf{F}_e\equiv-\mathbf{J}(\Gamma)V\cdot\mathbf{F}_e.$$

The term $\dot{Q}_l\left(t\right)$ is the change in energy due to the thermostat. It involves only the reservoir particles, as the particles in the system of interest are not thermostatted. In

Eq. (6) $\mathbf{J}$ is the so-called dissipative flux and V is the volume of the system of interest [3]. We will show that for sufficiently large N_w,

$$\dot{Q}_l - k_B T \Lambda(\Gamma, \mu) = 0, \tag{7}$$

for all $\mu > 2 - D$. Note that for a system with a large reservoir undergoing reversible heat transfer Eq. (7) shows that the phase space compression factor is related to the entropy lost from the thermostatting reservoir region

$$\dot{Q}_l / T = \dot{S}_{res} = k_B \Lambda\left(\Gamma, \mu\right), \ \forall \quad \mu > 2 - D. \tag{8}$$

The kinetic temperature of the reservoir is given by

$$k_B T = \frac{1}{DmN_w} \sum_{i=1}^{N_w} \mathbf{p}_i \cdot \mathbf{p}_i. \tag{9}$$

Since the reservoir is very large compared to the dissipative system of interest we can assume that in the reservoir region the momentum is distributed according to an equilibrium Maxwell-Boltzmann distribution and that the sum $\sum_{i=1}^{N_w} \mathbf{p}_i \cdot \mathbf{p}_i$ may be treated as an average $N_w \left\langle \mathbf{p}_i \cdot \mathbf{p}_i \right\rangle$. It is convenient to define,

$$I(a,b) \equiv \int_0^\infty dx\, x^a \exp\left(-bx^2\right) = \frac{1}{2} b^{-(1+a)/2} \Gamma\left(\left(1+a\right)/2\right), \ \forall \ a > -1, \tag{10}$$

where Γ denotes the gamma function, and then one can show that

$$\dot{Q}_l = \frac{-\alpha N_w I(D+\mu-1,\beta/2m)}{mI(D-1,\beta/2m)}, \tag{11}$$

and

$$\Lambda = \frac{-\alpha(D+\mu-2)N_w I(D+\mu-3,\beta/2m)}{I(D-1,\beta/2m)}. \tag{12}$$

Using Eqs. (11) and (12) and the relation $I(a,b)=(a-1)I(a-2,b)/2b$ it is easy to prove Eq. (7).

We can now derive the equilibrium distribution function for a large system of particles where the μ^{th} moment of the momentum distribution is constrained using the dynamics defined by Eq. (2). The Liouville equation may be written as [3]

$$\frac{d}{dt}\ln f\left(\Gamma,t\right) = -\Lambda\left(\Gamma\right), \tag{13}$$

which when combined with Eq. (7) and the fact that the system is at equilibrium ($\dot{W}=0$ thus $\dot{Q}\left(t\right)=\dot{H}\left(t\right)$) gives

$$\frac{d}{dt}\ln f\left(\Gamma\left(t\right)\right) = -\frac{\dot{H}\left(\Gamma\left(t\right)\right)}{kT}. \tag{14}$$

Upon integrating both sides with respect to time we obtain the distribution function

$$f_\mu(\Gamma)=\frac{\exp\left[-\beta H(\Gamma)\right]\delta\left(g(\mu,\Gamma)-g_0(\mu)\right)}{\int d\Gamma\exp\left[-\beta H(\Gamma)\right]\delta\left(g(\mu,\Gamma)-g_0(\mu)\right)},\tag{15}$$

where $g_0(\mu)$ is the value to which the μ^{th} moment is constrained, and $\beta=1/k_BT$ is the inverse thermal energy. In the case of the Gaussian isokinetic thermostat ($\mu=2$) the kinetic degrees of freedom are distributed microcanonically and the configurational degrees of freedom are distributed canonically, thus we have a clear link with equilibrium thermodynamics. For our system involving thermostatted wall particles this distribution function will be correct for the whole system given that there are enough thermostatted particles to make Eq. (7) valid: here $g(\mu,\Gamma)$ is only a function of the wall particle momenta.

After some rather tedious algebra one can prove that for systems in *D* Cartesian dimensions that the μ^{th} moments of the momenta are related to the equilibrium thermodynamic temperature by the equation,

$$g_0(\mu)=N_w\frac{\Gamma\left((\mu+D)/2\right)}{\Gamma(D/2)}\left(2mk_BT\right)^{\frac{1}{2}\mu}.\tag{16}$$

Since the distribution function for the thermostatted system is known we can apply the TFT to these systems. The dissipation function appearing in the TFT is defined as [6],

$$\bar{\Omega}_t t = \int_0^t ds\, \Omega\left(\Gamma\left(s\right)\right) \equiv \ln\left[\frac{f\left(\Gamma\left(0\right),0\right)}{f\left(\Gamma\left(t\right),0\right)}\right] - \int_0^t \Lambda\left(\Gamma\left(s\right)\right) ds. \qquad (17)$$

The TFT then gives the probability ratio

$$\frac{p\left(\bar{\Omega}_t = A\right)}{p\left(\bar{\Omega}_t = -A\right)} = \exp\left[At\right]. \qquad (18)$$

We may now combine Eqs. (7), (15), and (17) to show that when the number of degrees of freedom in the walls is large and much larger than the number of degrees of freedom in the system of interest,

$$\begin{aligned} \bar{\Omega}_t t &= \beta\left(\Delta H(t) - \Delta Q_l(t)\right) \\ &= \beta \Delta W(t) = -\beta \int_0^t ds\, \mathbf{J}(\Gamma(s)) V \cdot \mathbf{F}_e, \quad \forall \ \mu > 2 - D. \end{aligned} \qquad (19)$$

Substituting into Eq (18) gives

$$\frac{p\left(-\beta \bar{\mathbf{J}}_t \cdot \mathbf{F}_e = A\right)}{p\left(-\beta \bar{\mathbf{J}}_t \cdot \mathbf{F}_e = -A\right)} = \exp\left[AVt\right], \quad \forall \ \mu > 2 - D. \qquad (20)$$

Eq (20) is the central result of this paper.

Simulation

We test this theory by carrying out simulations using two thermostats, the usual Gaussian isokinetic thermostat $\mu = 2$ and the higher constrained moment of $\mu = 4$. We choose an application of the TFT that has been investigated in the past by both simulation and experiment [7]. A particle, immersed in a fluid, is held by a stationary harmonic well (optical trap in the experiment [7]) at equilibrium. At an arbitrary time $t = 0$ the trap suddenly moves at a constant velocity. In the large thermostat limit we have just proved that the dissipation function is independent of the thermostatting moment μ. The dissipation function has been given before [7] and is

$$\bar{\Omega}_t t = \beta \int_0^t ds\, \mathbf{v}_{opt} \cdot \mathbf{F}_{opt}, \quad \forall \;\; \mu > 2 - D. \tag{21}$$

Computer simulations were carried out in 2 dimensions using a WCA potential, the fluid number density was $\rho = 0.4$, the number of fluid particles in the system of interest was $N_s = 32$ and the number of thermostatted wall particles was $N_w = 28$ for the $\mu = 2$ thermostat while for the $\mu = 4$ thermostat systems with $N_w = 28$ and $N_w = 112$ were simulated. For both thermostats 2×10^5 trajectories were computed to form the averages with the trajectories being of duration $t = 5.0$, in dimensionless time units, all other details were as reported by Wang et al. [7].

For all systems the kinetic temperature of the fluid particles increased, on average, by ~2% over the full duration of the trajectory, while the kinetic temperature of the wall particles did not change. For the $\mu = 4$ thermostats the nominal dimensionless wall temperature was set at $T = 1$ by fixing the constraint Eq. (1) to the value specified by

Eq. (16). This resulted in an average kinetic temperature for the wall particles of T = 1.01 for the N_w = 112 system and T = 1.03 for the N_w = 28 system.

A direct test of Eq. (20) is plotted in Figure 1 for the time $t = 5.0$ where the trajectory finishes. This shows the logarithm of the probability ratio of observing a trajectory which has had a positive value for the time integral of the dissipation function relative to its negative, as a function of the integral value. For a dissipative field coupled with a large volume of particles or for long times this probability ratio diverges to infinity recovering the second law of thermodynamics. The figure shows excellent agreement between the data for both thermostats and the plotted prediction of Eq. (20). Note that in the case of the $\mu = 4$ thermostat only the large N_w = 112 system results are shown here.

The TFT Eq. (18) may be reduced in detail to obtain the integrated fluctuation theorem (ITFT).

$$\frac{p\left(\bar{\Omega}_t < 0\right)}{p\left(\bar{\Omega}_t > 0\right)} = \left\langle \exp\left(-\bar{\Omega}_t t\right)\right\rangle_{\bar{\Omega}_t > 0}. \tag{22}$$

The LHS is the probability of observing a negative value for the dissipation function $\bar{\Omega}_t$ divided by the probability of observing a positive value. The RHS is an ensemble average formed from the set of trajectories which have a positive value for the dissipation function at time t. In figure 2 we plot both sides of Eq. (22) which have been calculated directly from our simulation data for both thermostats (again in the case of the $\mu = 4$ thermostat only the large N_w = 112 system results are shown). In both cases we observe excellent agreement.

In the case of the smaller N_w = 28, $\mu = 4$ thermostat system, systematic differences between the fluctuation relation that is derived for large thermostatting regions and the simulation results are observed and this is shown in Fig. 3 for the case of the integrated fluctuation theorem. The systematic difference is a result of the distribution function for the wall particles' momentum being somewhat non-Gaussian due to both the heat flow from the system of interest and the constraint Eq. (1). As the number of wall particles is increased the intrinsic heat flow into the wall will decrease as will the effect of the constraint on its momentum distribution. If the field $\mathbf{F}_e$ acted on all the system of interest's particles we would expect this deviation to be larger. Here the heat would flow into the wall more directly with a stronger effect on the momentum distribution of the particles at the walls edge. Again this difference would become insignificant for a system with a very large number of wall particles.

Conclusion

In this paper we have shown that if we study a nonequilibrium system which is in contact with a thermostat then when the number of degrees of freedom in the thermostat is large and also large compared to the number of degrees of freedom in the nonequilibrium system, the Transient Fluctuation Relation is insensitive to the details of the thermostatting mechanism. We have shown this both theoretically and numerically for a class of time reversible deterministic thermostats that fix various moments of the momentum distribution. In this large thermostat limit the TFT is independent of the precise moment that the thermostat fixes.

These results are thus consistent with the previous gedanken arguments that for large thermostats the Fluctuation Theorem is insensitive to the precise details of the thermostatting mechanism and although the Fluctuation Theorems may be derived using fictitious (*i.e.* unnatural) thermostats, the Theorems nevertheless apply to natural systems.

One of the key results of this paper is Eq. (7). In a literal sense this equation says that there is an instantaneous relationship between the instantaneous phase space compression factor for the composite system and the instantaneous energy lost from the composite system due to the thermostat and the absolute temperature of the thermostat. This happens for any thermostat among the infinite family of thermostats considered here as long as the thermostat is sufficiently large that the thermostatting region can be regarded as being at equilibrium. Because the thermostat can be regarded as at equilibrium, the energy lost to the thermostat can be regarded as *heat*. Furthermore the heat loss divided by the absolute temperature is precisely the thermodynamic entropy lost from the composite system through the boundary region to the thermostat. Again this relation is independent of the precise details of the thermostat.

Since Eq. (7) is true regardless of the mathematical form of the thermostat we propose the following conjecture. Consider a nonequilibrium dissipative system (the system of interest) embedded in an (initially) equilibrium nondissipative Hamiltonian system. This system of interest has a fixed number of particles and a fixed dissipative field F_e is applied for $t > 0$. Assume the initial equilibrium thermodynamic temperature of the embedding system is *T*. Very far from the system of interest, beyond the embedding system, there may be some form of μ thermostatting region or the Hamiltonian embedding system may simply continue on forever, at temperature *T*.

If we consider a composite system (system of interest plus a surrounding sphere of the embedding system) of radius r, then for a fixed time $t > 0$, as *r* increases the embedding system located in the spherical shell region at $r \pm dr$, will be closer and closer to thermodynamic equilibrium. We can now apply Eq. (7) to relate the phase space compression of the composite (system of interest plus embedding) of radius *r*, $\Lambda_r(\Gamma)$, to the entropy lost from the composite system at radius *r*, to the surrounding equilibrium (embedding) system, $\dot{S}_r$:

$$\lim_{r\to\infty} \Lambda_r(\Gamma) = \dot{S}_r \tag{23}$$

This equation is expected to be true regardless of whether ultimately at very large distances the composite system contains a time reversible deterministic thermostat or whether the Hamiltonian embedding system continues indefinitely. This equation relates entropy loss through an equilibrium boundary to the phase space compression of the phase space of the system enclosed within that boundary. In this sense phase space compression can occur in purely Hamiltonian (sub) systems. This occurs in spite of the obvious fact that in the full phase space of any Hamiltonian system the phase space compression is identically zero [11].

We acknowledge financial support from the Australian Research Council and thank Joanne Bright for useful discussions.

[1] W. G. Hoover, A. J. C. Ladd, and B. Moran, Phys. Rev. Lett. **48**, 1818 (1982).
[2] D. J. Evans, J. Chem. Phys. **78**, 3297 (1983).
[3] D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids.* (Academic, London, 1990). Also available at: http://rsc.anu.edu.au/~evans/evansmorrissbook.htm
[4] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).
[5] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).
[6] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).
[7] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89** (2002).
[8] D. M. Carberry, J. C. Reid, G. M. Wang, E. M. Sevick, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **92**, 140601 (2004).
[9] R. Kalaba and R. Xu, Amer. Math. Monthly **102**, 821 (1995).
[10] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies: With an Introduction to the Problem of Three Bodies.* (Dover, New York, 1944).
[11] D. J. Evans and L. Rondoni, J. Stat. Phys. **109**, 895 (2002).

Figures

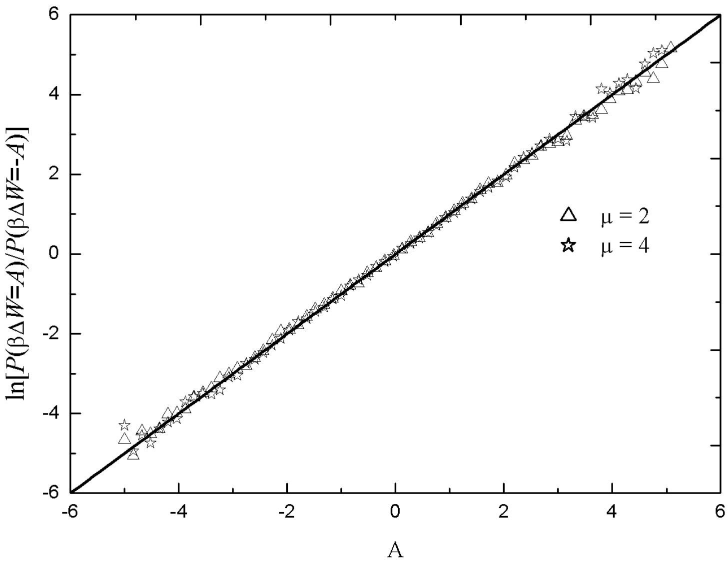

Figure 1. TFT data

The $\mu = 2$ thermostat simulation has 28 wall particles while the $\mu = 4$ thermostat simulation has 112 wall particles.

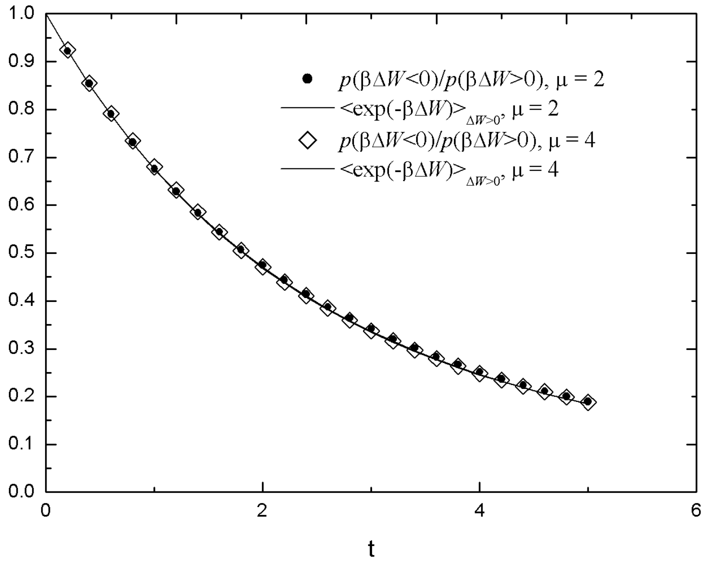

Figure 2 IFT data

The $\mu = 2$ thermostat simulation has 28 wall particles while the $\mu = 4$ thermostat simulation has 112.

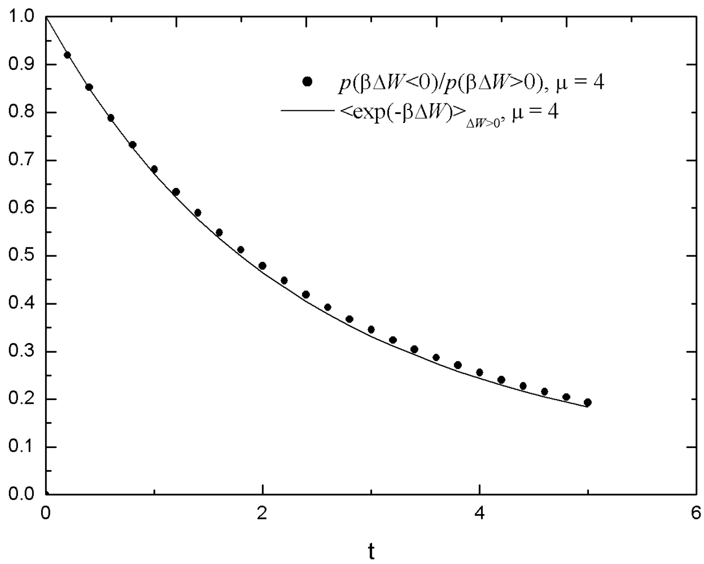

Figure 3. IFT data.

The simulation has 28 wall particles with a $\mu = 4$ thermostat. A small systematic disagreement between the theory and the simulation may be seen.