Notes on Monte Carlo Methods and Nonequilibrium Thermodynamics

Scott Cheng-Hsin Yang [*]
(Dated: December 11, 2009)

This report covers most of the main points I learned throughout the reading course. Its main function is to remind myself of what I have learned. It is somewhat meant to be read with the references.

The report's focus is on ideas rather than technical details. Presentation of calculation steps means that I found them enlightening and possibly useful.

I. MONTE CARLO METHODS

Monte Carlo are generally used for two purposes: 1) generating samples from a target distribution and 2) performing integration that is impossible analytically and unrealistically costly with quadrature. The latter is always a problem for high-dimensional systems.

Reference:

> David J. C. Mackay. **Information Theory, Inference, and Learning Algorithms**. Cambridge University Press, NY. 2003. Chapter 29, pg. 357–359.

A. Central Concept

The central concept of the Monte Carlo integration is that an integral $\int P(x)Q(x)dx$ (note that all integrals can be expressed in this form) can be estimated as an average of "random evaluation" of $P(X)$. The Xs are sampled from $Q(x)$, a target distribution. Basically, Monte Carlo integration comes down to doing evaluation at points that contribute most to the integral. The convergence rate of the estimator depends on the samples number and the correlation between subsequent evaluations.

Remarks:

1. The estimator asymptotically converges to the real value.
2. For independent samples, the central limit theorem governs the standard deviation of the estimator.

Reference:

> Malvin H. Kalos and Paula A. Whitlock. **Monte Carlo Methods**. 2nd Ed. Wiley-Blackwell, Weinheim. 2008. Section 2.6, pg. 21–22.

B. Transformation Method

The method transforms a uniform distribution of samples into a target distribution whose integral is analytically invertible. The key concept is that the areas under the uniform and target distribution match each other so that probability is conserved over the domain. Formula can be found everywhere.

C. Rejection Sampling

When the target distribution $Q(x)$ is too complicated to integrate (and invert) but evaluable at all points, one can use to the rejection method for sampling instead. The key idea is that one samples from an easy-to-sample trial distribution $R(x)$ and accept the sample with a probability $Q(x)/R(x)$.

Remarks:

1. Trial distribution needs to shadow target distribution everywhere.
2. In practice, one samples a X_i from the trial distribution $R(x)$ and a corresponding Y_i from a uniform distribution with domain $[0, R(x)_{max}]$. Samples are accepted if $Y_i \leq Q(X_i)$.
3. This method requires knowledge about the global structure of the target, which is often not available.

Reference:

> David J. C. Mackay. **Information Theory, Inference, and Learning Algorithms**. Cambridge University Press, NY. 2003. Chapter 29, pg. 364.

D. Importance Sampling

When performing Monte Carlo integration, the target distribution $Q(x)$ is often too difficult to sample. One can then sample X from an easy-to-sample trial distribution $R(x)$, evaluate $P(X)$, then weight the $P(X)$ with the ratio $Q(X)/R(X)$. The key idea is that the importance of the contribution of the evaluation is adjusted by the ratio.

Remarks:

1. This ratio between the target and the trial distribution comes up in different flavors everywhere for computing acceptance probability. It is a very simple yet important concept.

Reference:

[*]Electronic address: scotty@sfu.ca

Dann Frenkel. **Statistical Mechanics for Computer Simulators**. Material for Boulder Summer School 2002. Section 5.1, pg. 14–16. http://boulder.research.yale.edu/Boulder-2002/Lectures/Frenkel/simu.pdf

E. Metropolis-Hasting Algorithm

The Metropolis scheme is a simple yet very powerful algorithm for sampling from any target distribution $Q(x)$ that is evaluable at all coordinates. It is basically a biased random walk: a walker takes a trial move from X to X'; the trial move is accepted with a probability $min(1, Q(X')/Q(X)$. This is it. It works because the acceptance rule satisfies detailed balance and makes $Q(x)$ a stationary distribution. In less technical language, if N random walkers forms a distribution $R(x) = Q(x)$, then the evolution of the random walkers will keep being distributed the same way; if $R(x) \neq Q(x)$, the acceptance rule will bias $R(x)$ toward $Q(x)$.

Remarks:

1. This is guaranteed to work for infinite steps.
2. The samples from the trajectory is correlated. Be careful when using them for integration purposes.
3. See the first paragraph of the following section on Simulated Annealing for limitations.
4. Good Monte Carlo simulation and good molecular dynamics simulation are equivalent in terms of statistical properties. Both generate numerical trajectories that are close to the REAL ones.

Extensions:

1. The Rosenbluth scheme of composite moves is a useful extension for simulating linked random walkers (e.g., a polymer).
2. The Metropolis algorithm naturally connects with statistical mechanics. The connection inspired many extensions of the algorithm: sampling from different ensembles (isobaric-siothermal, grand canonical, Gibbs, and etc.), virtual moves for calculating chemical potentials, and etc.
3. Another extension is sequential Monte Carlo, or particle filter, which inserts a bootstrap resampling step to the distribution of random walkers. The technique is good for estimating Bayesian marginal distribution.

Reference:

Malvin H. Kalos and Paula A. Whitlock. **Monte Carlo Methods**. 2nd Ed. Wiley-Blackwell, Weinheim. 2008. Section 3.7, pg. 64–70.

David J. C. Mackay. **Information Theory, Inference, and Learning Algorithms**. Cambridge University Press, NY. 2003. Chapter 29, pg. 367–370.

Dann Frenkel. **Statistical Mechanics for Computer Simulators**. Material for Boulder Summer School 2002. Sections 5–11, pg. 13–55.

Dann Frenkel. Material for Boulder Summer School 2006. http://boulder.research.yale.edu/Boulder-2006/Lectures/Frenkel_BoulderNotes.pdf

Google Arnaud Doucet and Nando de Freitas from UBC.

F. Simulated Annealing

In practice, trial moves in the Metropolis scheme usually have small stepsize to ensure reasonable acceptance probability. The Metropolis walkers are thus local exploiters. In theory, with infinite moves, the distribution of the walkers' positions will eventually converge to any target distribution $Q(x)$. However, when the $Q(x)$ is multimodal, convergence can be very slow due to local traps. To deal with this problem, the walkers (at least some of them) should both be global explorers and local exploiters while satisfying detailed balance. Here, I show three popular working techniques: simulated annealing, parallel tempering, and genetic algorithm.

Inspired by the physical process of annealing, one introduces a temperature, and the modified target distribution becomes $e^{\ln Q(x) T^{-1}}$, where T is the temperature. Walkers still move according to the Metropolis scheme, just with the modified target distribution. At high T, the walkers explore a nearly uniform distribution and are not trapped. As T decreases quasi-statically (annealing), the walkers will at every T be distributed according to the modified $Q(x)$. When $T = 1$, one obtains $Q(x)$.

Remarks:

1. In principle, this method always works if one starts at sufficiently high T. Then, the method is also independent of initial distribution.
2. Quasi-static cooling can take long. A popular simple cooling schedule takes the form $T_{k+1} = \alpha T_k$, where α is a constant less but close to 1.
3. Faster-than-quasi-static cooling can result in "defects," which is equivalent to local traps.

Reference:

Phil Gregory. **Bayesian Logical Data Analysis for the Physical Sciences**. Cambridge University Press, NY. 2005. Section 11.4.1, pg. 296.

Peter J. M. van Laarhoven and E. H. L. Aarts. **Simulated Annealing: Theory and Application**. D. Riedel Publishing Comapany, Dordretcht, Holland. 1987. pg. 59–71.

G. Parallel Tempering

The walkers are separated into m different "parades." The target distribution (same form as in simulated annealing) of each parade is at a different temperature. The m parades proceed in parallel: those at high T explore, while those at low T exploit. The crucial step is that, sometimes, two walkers from different parades are allowed to swap positions and walk in the new parade. Looking at only the $T = 1$ parade, this is effectively the same as exploiting walkers taking occasional exploring moves. The hidden trick that allowed swapping moves is the formation of a new target distribution that is the product of all the parade's target distributions.

Remarks:

1. In simulated annealing, all walkers do the same kind of walk. Here, the walkers split tasks and exchange information. If sampling partial features of $Q(x)$ is rewarding, parallel tempering is much faster than simulated annealing.

2. Performance of parallel tempering is somewhat dependent on the initial distribution of the parades (especially the $T = 1$ one). Bad initial distributions need many swaps to correct for and can thus be slow.

3. Another tempering method is called simulated tempering. Here, transition to any temperature is allowed with a predetermined probability. The extended target distribution is the product of the original target distribution times the predetermined temperature distribution.

Reference:

Artur B. Adib. The Theory behind tempered Monte Carlo Methods. 2005. `arturadib.googlepages.com/tempering.pdf`

H. Genetic Algorithm

Most genetic algorithms are used for finding global extremum rather than sampling. To adopt normal genetic algorithms for sampling, one just needs to construct an appropriate target distribution and accept moves with the Metropolis rule. In the case below, the full target distribution with N walkers is the product of all the walkers' distributions: $Q(x_1)Q(x_s)\cdots Q(x_N)$.

The special move here is large-scale swapping, or in genetic language, crossover. As a simple illustration, consider splitting the N walkers into two equal groups (representing two genes). The crossover move swaps the positions of the walkers in the two genes pairwise (with the Metropolis rule), i.e., $i \leftrightarrow N/2+i$ for $i = 1, 2, ..., N/2$. Then, one compares the fitness, a quantitative measure of resemblance to $Q(x)$, between the original (parent) and the crossed (child) and keep the one with the higher fitness. Mutation moves can be replaced by translational moves.

Remarks:

1. As for all genetic algorithms, the method works if "feature building blocks" are present. For example, given that groups of walkers have found all the peaks of $Q(x)$ but are not proportioned rightly, genetic algorithms can evolve the walkers to the right proportion faster than temperature-based methods. In practice, genetic walkers are really "jumpers" that hop from peak to peak.

2. To obtain feature blocks, one should start with position-sorted walkers, so the walkers can arrange themselves into groups.

3. On size of target distribution space: If there are N walkers, n spatial spots, n_T temperature spots, and m parades, then, the "sizes" of the full target distributions of the different methods are: simulated annealing $\rightarrow n_x$, simulated tempering $\rightarrow n_x \times n_T$, parallel tempering $\rightarrow n^m$, and genetic algorithm $\rightarrow n^N$. Methods involving swapping have much bigger sizes because swapping moves require swappers to have their own identity, while walkers do not. This is similar to the idea of distinguishable and indistinguishable particle.

Reference:

Tuomas J. Lukka and Janne V. Kujala. Using Genetic Operators to Speed up Markov Chain Monte Carlo Integration. *Monte Carlo Methods and Applications*: **8**, pg. 51–72, Ingent, 2002. `users.jyu.fi/~jvkujala/papers/gomcmc.ps`

I. Profitable Future Exercises

1. Testing convergence and efficiency numerically.

2. General techniques for analytical calculation of convergence rate.

3. Novel moves and hybrid models.

II. NONEQUILIBRIUM THERMODYNAMICS

I reviewed classical thermodynamics, Hamiltonian formulation, and equilibrium statistical mechanics using my undergraduate textbooks:

Ashley H. Carter. **Classical and Statistical Thermodynamics**. Prentice-Hall, New Jersey. 2001. Chapters 1, 3, 6, 7.

L. D. Landau and E. M. Lifshitz. **Mechanics: Volume 1 of *Course of Theoretical Physics***. 3rd ed. Butterworth-Heinenann, Oxford. 2001. Chapter VII, pg. 131–132.

Walter Greiner, Ludwig Neise, and Hosrt Stöcker. **Thermodynamics and Statistical Mechanics**. Springer-Verlag, New York. 1995. Chapter 2, pg. 43–46. Chapters 5–7, 9.

The following discussion of nonequilibrium thermodynamics focuses on systems that are randomly perturbed. Even when starting and ending with the same states, the system evolves differently for every realization.

A. Markov-chain Framework

In this section, I follow Ritort's review. The method was first developed by Crooks.

Consider a system (with configuration C) embedded in a bath whose properties are controlled by a parameter λ (often called external control). By changing λ non-quasi-statically, the system can be driven away from equilibrium. The system's response to the change in λ is probabilistic. A Markov-chain framework is used to describe these probabilistic changes. Key ingredients include:

1. A distribution of configurations $P_\lambda(C)$ that describes the system totally.

2. A transition probability $T_\lambda(C \to C')$ that governs how the system changes from $C \to C'$, given a particular λ.

3. Microscopic reversibility or local detailed balance

$$\frac{T_\lambda(C \to C')}{T_\lambda(C' \to C)} = \frac{P^{eq}_\lambda(C')}{P^{eq}_\lambda(C)}$$

that asserts that the transition probability would drive the system to equilibrium if λ stops changing. This also means time reversal invariance.

4. A path probability

$$P(\Gamma) = \prod_{k=0}^{M-1} T_{\lambda_k}(C_k \to C_{k+1})$$

(the product of transition probabilities) that describes a single path in phase space.

5. Ensemble path average

$$\langle A \rangle = \sum_\Gamma A(\Gamma) P_{\lambda_o}(C_o) P(\Gamma)$$

that generates thermodynamic quantities.

Within this framework, wonders happen when defining the total change in entropy as

$$\begin{aligned} \Delta S_{tot}(\Gamma) &= \Delta S_m(\Gamma) + \Delta S \\ \Delta S_m(\Gamma) &= \sum_{k=0}^{M-1} \ln\left[\frac{P^{eq}_{\lambda_k}(C_{k+1})}{P^{eq}_{\lambda_k}(C_k)}\right] \\ \Delta S &= \ln\left[\frac{P_{\lambda_o}(C_o)}{P_{\lambda_M}(C_M)}\right], \end{aligned}$$

where ΔS_m (related to the path) is the entropy change of the medium, and ΔS (related to the boundaries) is the entropy change of the system. The two wonders are:

1. Integral fluctuation theorem. Computing the ensemble path average of $\exp(\Delta S_{tot})$, one finds

$$\langle \exp(-\Delta S_{tot}) \rangle = 1.$$

2. Detailed fluctuation theorem. Using a δ-function to pick a specific path in the ensemble path average, one can find

$$\frac{P_F(\Delta S_{tot})}{P_R(-\Delta S_{tot})} = \exp(\Delta S_{tot}),$$

where F and R means moving along the forward and the exact reverse path.

Further, when using equilibrium boundary distributions, $P^{eq}_{\lambda_o}(C_o)$ and $P^{eq}_{\lambda_M}(C_M)$, and the first law, $W = \Delta V + Q$:

1. the integral fluctuation theorem leads to Jarzynki's nonequilibibrium work relation

$$\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta F),$$

where ΔF is the equilibrium free energy difference between the two boundary distributions.

2. the detailed fluctuation theorem leads to Crooks fluctuation theorem,

$$\frac{P_F(W)}{P_R(-W)} = \exp\left[\beta(W - \Delta F)\right].$$

The two relations are powerful because they connect nonequilibrium quantities with equilibrium quantities in closed form and are valid arbitrary far from equilibrium. An application of the two relations is discussed later.

The results above help clarify thermodynamic concepts:

1. Interpretation of irreversibility: from the detailed fluctuation theorem, every process is, precisely speaking, reversible. It's just that the reverse process is exponentially unlikely to happen.

2. The entropy production from the system to the bath (or simply the entropy change in the bath) is path dependent (excluding the boundaries), while the entropy change of the system depends only on the boundaries.

3. Work is related to changing the external control λ while fixing the system's configuration. If λ is changed in a way that the system prefers (based on the ratio $P_{\lambda'}(C)/P_\lambda(C)$) to remain at C, work was done by the system; otherwise, work is done on the system (or in the case of equal likelihood, no work is done).

4. Heat is related to changing the system's configuration while fixing λ. If the system goes to a more favourable configuration (based on $P_\lambda(C')/P_\lambda(C)$) or a lower energy state, heat is dissipated to the bath.

5. Any step that changes the system's configuration and the external parameter together can be broken into a work part, a heat part, and a boundary (internal energy) part.

Remarks:

1. The Markov-chain is only an assumption and probably cannot describe all nonequilibrium processes.

2. Jarkynski first derived his nonequilibrium work relation using Hamiltionian evolution which is deterministic. This reflects the fact that, although the system takes on probabilistic trajectories, the average path is a deterministic one. In fact, from the use of equilibrium distributions to construct transition probabilities, the average path is (I think) the quasi-static equilibrium path.

3. Using Jensen's inequality on integral fluctuation theorem gives the second law, on Jarzynski's relation gives $W \geq \Delta F$.

Reference:

Felix Ritort. Nonequilibrium fluctuations in small systems: from physics to biology. *Advances in Chemical Physics*: **137**, 31–123. 2008. `http://www.ffn.ub.es/ritort/PAPERS/RITORT.ACP07.pdf`

Gavin E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys Rev E*: **61**, 2361-2366. 2000.

B. Langevin-dynamics Framework

The framework starts with describing the dynamics of thermodynamic systems with the Langevin equation. Sekimoto described this connection, "If a Langevin equation represents the balance of forces on a system, then the Langevin dynamics conserves the energy of the system plus the surrounding heat bath. This is essentially the first law of thermodynamics...."

To describe ensemble averages, one resorts to the corresponding Fokker-Planck equation that describes the evolution of the distribution of states. The evolving distribution allows one to define a path.

In more detail, following Seifert, path representation is constructed with the following steps: discretize the Langevin equation. Form the joint probability distribution of noise with the now explicit time index. Substitute the velocity and force terms for the noise. This is equivalent to a change of coordinate and results in a Jacobian factor. And done! Now, all of the results derived with ensemble path average in the Markov-chain framework can be derived here.

The difference between this approach and the above are mainly two-fold:

1. Langevin easily includes non-conservative forces. This makes Langevin thermodynamics the natural language to describe nonequilibrium steady states (e.g., a circuit made out of a battery and a resistance, where a part of the current always dissipates as heat). Interesting, even though the nonconservative forces break microscopic reversibility, the fluctuation theorems obtained here have the same forms as the above. This implies a more general kind of local detailed balance.

2. The explicit dynamic formulation allows for the use of techniques from mechanics. An example is Seifert's use of the Euler-Lagrange equations to find the optimal finite-time protocol that minimizes work done. Another example is the finding of entropy production rate and the entropy's equations of motion.

To manipulate the equations of motions in this framework, it is fundamental to know how to go from Langevin to Fokker-Planck. I follow Schwabl:

1. The probability of the system's being in state α at time t can be written as an ensemble average of delta functions:

$$P(\alpha, t) = \langle \delta(\alpha - x(t)) \rangle .$$

2. Total time derivative of $P(\alpha, x)$ is zero because of conservation of probability. The partial time derivative is

$$\partial_t P(\alpha, t) = -\partial_\alpha \langle \delta(\alpha - x(t)) \dot{x}(t) \rangle = -\partial_\alpha j(\alpha, t),$$

where j is called the probability current. Note that taking the derivative involves first acting on the state α, which is a value picked by us and constant in time. Then, the derivative acts on x, which is governed by the Langevin dynamics and time-dependent. The chain rule reflects the fact that $P(\alpha, t)$ is composed of the trajectories' occupying α and is only time-dependent because $x(t)$ is.

3. The spatial derivative of $j(\alpha)$ characterize the difference between the probability flux into and out of α. Such spatial difference in flux changes the probability in time.

4. Since $\dot{x}$ has a force and a noise part, j also has the corresponding two parts. The force part shifts all the trajectories $x(t)$ in determinisitic ways and can be taken out of the average. This gives rise to the drift term $F(\alpha,t)P(\alpha,t)$, where F is the force that has a conservative and a non-conservative part.

5. The noise part randomly adds Δx to the trajectories $x(t)$, where Δx is usually drawn from a Gaussian distribution with zero mean and variance $2D$, D being the diffusion coefficient. The evaluation of the average requires integration by part, giving rise to a $D\partial_\alpha P(\alpha,t)$ part in j. This is called the diffusion term.

6. The flux's dependence on a spatial derivative (difference in α occupancy) reflects the the randomness of diffusion process, while the flux's dependence on $P(\alpha)$ (individual α occupancy) reflects deterministic motion.

7. Mathematically, the spatial derivative comes from integration by part. Therefore, it is possible to have coloured noise distributions that result in multiple derivatives. Thinking about these possibilities should be intriguing.

Gardiner detailed formally how to go from Fokker-Plank to Langevin. At this point, topics along the line of equations of motion can be further pursuit.

Remarks:

1. Interpretations of work, heat, and entropy production are the same as the Markov-chain framework.

2. One can recast everything in classical thermodynamics into the Langevin framework. This should be insightful to work through.

Reference:

Ken Sekimoto. Langevin equation and thermodynamics. *Progress of Theoretical Physics Supplement*: 130. 1998.

Udo Seifert. Lecture Notes: Stochastic Thermodynamics. `http://www.theo2.physik.uni-stuttgart.de/institut/seifert/b5_seifert_web.pdf`.

Franz Schwabl. **Statistical Mechanics**. 2nd ed. Springer-Verlag, Berlin. 2006. pg. 416–417.

Crispin Gardiner. **Stochastic Methods: A Handbook for The Natural and Social Sciences**. 4th ed. Springer-Verlag, Berlin. 2009. pg. 42–55.

C. Application: Free Energy Extraction

A direct application of Jarzynski's and Crooks' relations is the extraction of equilibrium free energy difference ΔF from nonequilibrium processes, where a fraction of the work done on the system is dissipated as heat. The two relations are first applied to single-molecule pulling, where fluctuation is non-negligible and external control well defined. The goal of these first tests was to extract the ΔF between the folded and unfolded states of a molecule from non-quasi-static pulling.

The basic setup is the following: prepare a RNA (e.g., a hairpin) that has a folded and an unfolded state. Attached each end of the RNA to a piece of complementary DNA. Biotinylate each DNA piece to a bead. One bead is attached to a pipette for pulling; the other bead is trapped optically for force measurement. The external control is the end-to-end distance z of the RNA, which can be measured by a CCD camera. Different loading (pulling) rates are used to test whether the relations are truly independent of control protocols. Work done on the system is computed by finding the area under the force-extension curve.

Lipardt et al. computed the total work done to unfold the RNA for each pull. The work values are averaged according to the Jarzynski relation to obtain the ΔF of unfolding. They also performed quasi-static pulling to experimentally obtain ΔF and showed that Jarzynski's relation worked. Properties on the distribution of dissipated work (total work $-\Delta F$) for unfolding include:

1. For small z (5 nm), the distribution is delta function like and centers around zero, regardless of loading rates.

2. For medium z (15 nm), the distribution is nearly Gaussian, regardless of loading rates. Increasing loading rate right-shifts and widens the distribution. The changes reflect that both the average and the range of dissipation increase with faster loading rate.

3. When the distribution is Gaussian, the system is near equilibrium, and the fluctuation-dissipation theorem ($\Delta F =< W > -\beta\sigma^2/2$) holds.

4. For large z (25 nm), the distribution is skewed with a long left tail. Increasing loading rate right-shifts the mode and elongates the tail. The shapes of these distributions are strongly constrained by the integral fluctuation theorem.

Collin et al. computed the total work for both the unfolding and refolding processes. Properties of these two work distributions and their relationships (all of which match Crooks' prediction) include:

1. For non-reversible pulling, the unfolding work distribution is clearly to the right (larger work values) of the refolding distribution, implying hysteresis.

2. The unfolding work distribution is always wider than the refolding distribution. Their complementary shapes are strongly constrained by the detailed fluctuation theorem.

3. The work value at which the two distributions are equal is ΔF, regardless of loading rates.

4. The method is precise enough to distinguish between the unfolding energy of a wild-type ($\approx$ 25 basepairs) and its single-basepair mutant.

Remarks on validity:

1. To use the Jarzynski relation, the probed system has to start from but not end with equilibrium. If one starts with a nonequilibrium state, the energy difference computed will be *something* but not the equilibrium ΔF.

2. When using the Crooks relation, similar ideas hold: both the forward and backward processes require the system to be initially at equilibrium.

3. The system should have access to all possible configurations (ergodic).

4. The application of the Jarzynski relation boils down to sample rare events of small work values, as these are weighted exponentially more. For processes with large dissipation, many more realizations are needed to obtain the rare events. Crooks fluctuation theorem suffers less from this because it uses entire distributions. It is, however, still true that large dissipation implies that the most important region — region where the forward and backward distributions cross — is from rare events.

Reference:

Jan Liphardt, Sophie Dumont, Steven B. Smith, Ignacio Tinoco Jr., and Carlos Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*: **296** 1832–1835. 2002.

D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr, and C. Bustamante. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*: **437** 231–234. 2005.