Generalized Jarzynski Equality under Nonequilibrium Feedback Control

Takahiro Sagawa[1] and Masahito Ueda[1,2]

[1]*Department of Physics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan*
[2]*ERATO Macroscopic Quantum Control Project, JST, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan*

The Jarzynski equality is generalized to situations in which nonequilibrium systems are subject to a feedback control. The new terms that arise as a consequence of the feedback describe the mutual information content obtained by measurement and the efficacy of the feedback control. Our results lead to a generalized fluctuation-dissipation theorem that reflects the readout information, and they can be experimentally tested using small thermodynamic systems. We illustrate our general results by introducing an "information ratchet", which can transport a Brownian particle in one direction and extract positive work from the particle.

Introduction.—Since the 1990s, the field of nonequilibrium statistical mechanics has witnessed remarkable progress so that nonequilibrium dynamics can now be described in terms of equalities [1,2]. These equalities have been experimentally verified using small thermodynamic systems such as biomolecules or colloidal particles [3]. A prominent example of such equalities is the Jarzynski equality [1]:

$$\langle e^{-\beta(W-\Delta F)} \rangle = 1, \qquad (1)$$

where $\langle W \rangle$ is the statistical average of work W performed on a thermodynamic system, ΔF is the free-energy difference, and $T \equiv (k_B \beta)^{-1}$ is the temperature of the initial canonical distribution. Equality (1) is satisfied even when the final state of the system is far from equilibrium, i.e., even when we drive the system from $t = 0$ to τ by changing the external parameter λ from $\lambda(0)$ to $\lambda(\tau)$ with a finite speed, where λ describes, for example, the volume of the gas or the frequency of an optical tweezer. The left-hand side of (1) involves all orders of cumulants of W. In fact, the second law of thermodynamics [4]

$$\langle W \rangle \geq \Delta F \qquad (2)$$

and the fluctuation-dissipation theorem result from the first and second cumulants of W, respectively [1].

Furthermore, recent advances in active control and precision measurement of small thermodynamic systems present new possibilities in nonequilibrium physics of small systems. In particular, feedback control enhances our controllability of small thermodynamic systems [5,6], and plays a crucial role in biological and artificial nanomachines [7]. In addition to such engineering applications, feedback control of thermodynamic systems has also been a subject of active research in terms of the foundation of the second law of thermodynamics. In fact, it is well understood that the role of "Maxwell's demon" can be characterized as a feedback controller for thermodynamic systems [8,9].

Suppose that we perform a measurement on a stochastic thermodynamic system at time t_m. Let Γ_m be the phase-space point of the system at that time, $P[\Gamma_m]$ its probability, and y the measurement outcome. We assume that the measurement can involve a measurement error which is characterized by the conditional probability $P[y|\Gamma_m]$ of obtaining outcome y on condition that the state of the system is Γ_m. For example, if a Gaussian noise is induced in the measurement, the conditional probability is given by $P[y|\Gamma_m] = (2\pi N)^{-1/2} \exp[-(y - \Gamma_m)^2/2N]$ with $N > 0$ being the variance of the noise. The probability of obtaining outcome y is given by $P[y] = \int d\Gamma_m P[y|\Gamma_m]P[\Gamma_m]$. The information obtained by the measurement can be characterized by the mutual information [10], $\langle I \rangle \equiv \int d\Gamma_m dy P[y|\Gamma_m]P[\Gamma_m]I[\Gamma_m, y]$ with $I[\Gamma_m, y] \equiv \ln(P[y|\Gamma_m]/P[y])$. If we perform a feedback control, the control protocol of parameter λ depends on the outcome y after t_m, which we write as $\lambda(t; y)$. The introduction of the feedback control requires us to generalize the second law of thermodynamics (2) by including the mutual information $\langle I \rangle$ obtained by the feedback controller (or the "demon") [9]:

$$\langle W \rangle \geq \Delta F - k_B T \langle I \rangle. \qquad (3)$$

Thus, the work that needs to be performed on a thermodynamic system can be lowered by feedback control.

Now the crucial question is, Is it possible to generalize the Jarzynski equality (1) in the presence of feedback control such that we can obtain more detailed information about nonequilibrium dynamics than inequality (3), as is the case for the original Jarzynski equality? In this Letter we answer the question in the affirmative.

First main result.—The generalized Jarzynski equality involves a term of information on the left-hand side:

$$\langle e^{-\beta(W-\Delta F)-I} \rangle = 1, \qquad (4)$$

which will be proved later. We note that ΔF may depend on y if $\lambda(\tau; y)$ does. Our result is applicable to classical

stochastic processes that satisfy the local detailed balance (or the detailed fluctuation theorem) [2]. Therefore, our result can be applied to a broad class of active control of small nonequilibrium systems.

The first cumulant of Eq. (4) straightforwardly reproduces inequality (3) because of the concavity of the exponential function. If all of the stochastic variables are Gaussian, the second cumulant leads to a generalized fluctuation-dissipation theorem including the term of the mutual information:

$$\langle \sigma + I \rangle = \tfrac{1}{2}[\Delta(\sigma + I)]^2, \tag{5}$$

where $\sigma \equiv \beta(W - \Delta F)$ is the dissipation of work (or the entropy production), and $[\Delta(\sigma + I)]^2 \equiv \langle (\sigma + I)^2 \rangle - \langle \sigma + I \rangle^2$ is the variance of the sum of work and mutual information. Therefore, the more information we obtain, the less dissipation the system will suffer.

Second main result.—If we measure the left-hand side of the original Jarzynski equality (1) in the presence of feedback control, the right-hand side is expected to differ from unity. Let us write it as γ:

$$\langle e^{-\beta(W - \Delta F)} \rangle = \gamma. \tag{6}$$

The crucial point is that we can directly measure γ by using backward control protocols, and that γ characterizes the efficacy of feedback control. Thus the left-hand and right-hand sides of Eq. (6) can be measured by the independent procedures.

We now discuss the properties of γ. We first note that the control protocol of λ depends on measurement outcome y at time $t > t_m$ with feedback control. In particular, if the number of possible outcomes is finite and given by M, we have M types of protocols $\lambda(t; y)$ in the forward process. Corresponding to each of them, we perform a backward protocol $\lambda^\dagger(t; y) \equiv \lambda(\tau - t; y)$, which depends on y only in $0 \leq t < \tau - t_m$, by starting with the initial canonical distribution corresponding to parameter $\lambda^\dagger(0; y)$. We stress that we do not perform any feedback control in the backward processes. Instead, we respectively drive the system depending on the forward outcome y. We then perform a measurement during the backward processes at time $\tau - t_m$, and obtain outcome y'. Let $P_{\lambda^\dagger(t;y)}[y']$ be the probability of obtaining outcome y' with control protocol $\lambda^\dagger(t; y)$, which is normalized as $\int dy' P_{\lambda^\dagger(t;y)}[y'] = 1$ for all y. We then write the time-reversal y as y^*; if we only measure the momentum of the system, then $y_i^* = -y_i$; if we only measure the position of the system, then $y_i^* = y_i$. For a special case of $y' = y^*$, we use the notation $P_{\lambda^\dagger(t;y)}[y^*]$, which is not necessarily unity. Then we can show that γ is given by

$$\gamma = \int dy P_{\lambda^\dagger(t;y)}[y^*]. \tag{7}$$

As discussed in detail later, to prove Eq. (7), we assume that the conditional probability satisfies $P[y^*|\Gamma_m^*] =$

$P[y|\Gamma_m]$. Here, Γ_m^* is the time-reversal of a phase-space point Γ_m. For example, if $\Gamma_m = (\boldsymbol{r}, \boldsymbol{p})$ with position $\boldsymbol{r}$ and momentum $\boldsymbol{p}$, then $\Gamma_m^* = (\boldsymbol{r}, -\boldsymbol{p})$. Physically, γ is the sum of the probabilities of obtaining the time-reversed outcomes with time-reversed protocols. Without feedback control, we have $\gamma = 1$ because $P_{\lambda^\dagger(t)}[y^*]$ would then reduce to a single probability distribution.

The validity of Eq. (6) can be tested experimentally by measuring the left-hand side and the right-hand side independently; we can measure W and ΔF with forward processes, and determine γ by repeatedly performing time-reversed protocols $\lambda^\dagger(t; y)$ for all possible outcomes y. Once the validity of Eq. (6) has been confirmed, we can estimate the feedback efficacy γ by only measuring W and ΔF with forward protocols.

We note that the effect of feedback control can be amplified by Eq. (6); a small amount of work that satisfies $W < \Delta F$ can make an exponentially large contribution to the left-hand side of Eq. (6). In particular, with feedback control, a situation can occur in which Eq. (1) is violated while inequality (2) is still satisfied. We will discuss such an example later. In such a situation, the feedback control only affects cumulants higher than the first order $\langle W \rangle$.

We next discuss the relationship between mutual information I and parameter γ. Let $C[X] \equiv \ln\langle e^{-X} \rangle$ be the cumulant generating function of a probability variable X. From Eqs. (4) and (6), and identity $\langle e^{-I} \rangle = 1$, we have $C[\sigma + I] - C[\sigma] - C[I] = -\ln\gamma$. The left-hand side of this equality characterizes the correlation between σ and I, and therefore we find that γ is a measure of the correlation between the dissipation and the information. In particular, if the joint distribution of σ and I is Gaussian, we have

$$\langle \Delta\sigma\Delta I \rangle = -\ln\gamma, \tag{8}$$

where $\langle \Delta\sigma\Delta I \rangle \equiv \langle \sigma I \rangle - \langle \sigma \rangle\langle I \rangle$. While I only characterizes the information obtained by the measurement, γ characterizes how efficiently we use the obtained information with feedback control. When γ is large, we efficiently make the dissipation σ smaller by using the obtained information I, i.e., if we possess more information I, the dissipation σ is lesser. We note that I only depends on the measurement; however γ depends both on the measurement and the feedback protocol.

Examples.—As an illustrative example, we consider Eq. (4) for the Szilard engine [8]. The Szilard engine is a single-molecule ideal gas controlled by Maxwell's demon. The gas is initially in thermodynamic equilibrium with a heat bath at temperature T. We partition the box into two boxes of equal volume. We then perform a measurement on the system to determine which box the molecule is in; the measurement outcome is "left" or "right." By this measurement, we gain(1 bit $= \ln 2$ nat) of information. When the outcome is right, we remove the left box and quasistatically move the right one to the left. Finally, we expand the box to the right, and the state of the system S returns to the

initial state. During the entire process, we extract $k_B T \ln 2$ of work from the system with no free-energy change (i.e., $\Delta F = 0$). Since $W = k_B T \ln 2$ holds for all trajectories in the quasistatic limit and $I = \ln 2$ holds for both left and right, we find that Eq. (4) holds for the case of the Szilard engine, that is, $\exp[-\beta(-k_B T \ln 2) - \ln 2] = 1$.

The backward process of the Szilard engine is described as follows. The gas is initially in thermodynamic equilibrium, and we quasistatically compress the box to the left. The following step bifurcates into two branches depending on the measurement outcome of the forward process. If the outcome is left, we do not move the box, and measure the position of the molecule. Clearly, the outcome must be left with unit probability: $P_{\lambda^\dagger(t;\text{left})}(\text{left}) = 1$. On the other hand, if the outcome is right, we quasistatically move the box to the right, and measure the position of the molecule. The outcome must be right with unit probability: $P_{\lambda^\dagger(t;\text{right})}(\text{right}) = 1$. Finally, we remove the partition of the box and let the gas freely expand. We then obtain $\gamma = P_{\lambda^\dagger(t;\text{left})}(\text{left}) + P_{\lambda^\dagger(t;\text{right})}(\text{right}) = 2$. Therefore, we find that Eq. (6) holds as $\exp[-\beta(-k_B T \ln 2 - 0)] = 2$.

We next discuss a model of "information ratchet." We consider a one-dimensional Brownian particle in a harmonic potential. Suppose that the particle is initially at thermal equilibrium in a potential $V_X(x) \equiv k(x - X)^2/2$, where X is the center position of the potential. We then measure the position x of the particle and obtain outcome y. We assume that the measurement involves a Gaussian noise whose probability distribution is $p(y - x) = (2\pi N)^{-1/2} \exp[-(y - x)^2/2N]$. The joint probability of x and y with potential $V_X(x)$ is then given by

$$p_X(x, y) = \frac{1}{2\pi\sqrt{SN}} \exp\left(-\frac{(x - X)^2}{2S} - \frac{(y - x)^2}{2N}\right), \quad (9)$$

where $S \equiv (k\beta)^{-1}$. Immediately after the measurement, we perform the following feedback control [see also Fig. 1(a) for the case of $X = 0$]: if $y \geq X + L$ with $L > 0$ being a constant, then we switch the potential to $V_{X+2L}(x)$; if $y < X + L$, we do nothing. We next wait for the relaxation of the particle. When the probability distribution of the particle becomes a thermal equilibrium one, we repeat the same feedback protocol by replacing X by $X + 2L$. By repeatedly performing this protocol, the average position of the particle moves to the right. We note that $\Delta F = 0$ holds for this process. This one-way transportation of the particle looks like a ratchet model [5]. However, the distinctive feature of the present model is that we do not need any asymmetry of the potential shape to drive the particle in one direction. Moreover, we can even extract positive work during this transport if the measurement errors are sufficiently small, as discussed below. Information obtained from the measurements enables one-way transportation driven by feedback contro, and therefore we call this model information ratchet.
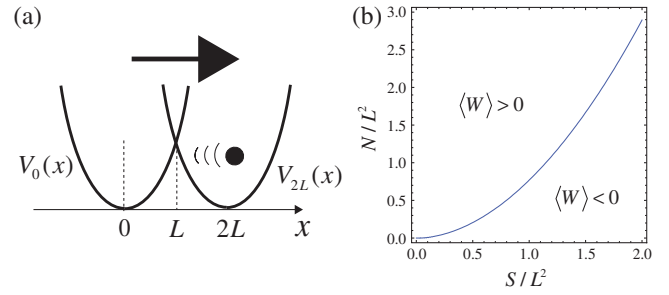
(a)

(b)

FIG. 1 (color online). (a) Schematic of the information ratchet with $X = 0$, with which we transport a Brownian particle to the right solely by means of a feedback control. (b) Regimes of $\langle W \rangle < 0$ and $\langle W \rangle > 0$ on the (S, N) plane. The conventional second law (2) is violated only in the regime of $\langle W \rangle < 0$ with a large S/N ratio, whereas the original Jarzynski equality (1) is not satisfied on the entire region.

We discuss the energetics of the information ratchet for a single step. The work performed on the particle at x is given by $W(x) \equiv V_{X+2L}(x) - V_X(x)$ if $y \geq X + L$, and 0 if $y < X + L$. Therefore, the average work for each step is

$$\beta\langle W \rangle = \int_{-\infty}^{\infty} dx \int_{X+L}^{\infty} dy W(x) p_X(x, y)$$
$$= \frac{L^2}{S} \text{erfc}\left(\frac{L}{\sqrt{2(S + N)}}\right) - \frac{2L}{\sqrt{2\pi(S + N)}} e^{-L^2/2(S+N)},$$

(10)

where $\text{erfc}(t) \equiv (2/\sqrt{\pi}) \int_t^{\infty} e^{-t'^2} dt'$. We can show that $\langle W \rangle < 0$ holds for all S if $N \to 0$, implying that we can extract a positive work during the one-way transportation if the measurement is error-free. Figure 1(b) shows the regimes of $\langle W \rangle < 0$ and $\langle W \rangle > 0$ on the (S, N) plane. The mutual information is determined by the S/N ratio: $\langle I \rangle = (1/2) \ln(1 + S/N)$. Since $e^{-I} = p_X(x)p_X(y)/p_X(x, y)$ holds with $p_X(x) \equiv \int_{-\infty}^{\infty} dy p_X(x, y)$ and $p_X(y) \equiv \int_{-\infty}^{\infty} dx p_X(x, y)$, we obtain $\langle e^{-\beta W - I} \rangle = \int_{-\infty}^{\infty} dx \times \int_{-\infty}^{X+L} dy 1 p_X(x) p_X(y) + \int_{-\infty}^{\infty} dx \int_{X+L}^{\infty} dy e^{-\beta W(x)} p_X(x) \times p_X(y) = 1$, and therefore Eq. (4) is satisfied in this model. We can also show that $\langle e^{-\beta W} \rangle = \int_{-\infty}^{\infty} dx \times \int_{-\infty}^{X+L} dy 1 p_X(x, y) + \int_{-\infty}^{\infty} dx \int_{X+L}^{\infty} dy e^{-\beta W(x)} p_X(x, y) = \text{erfc}[-L/\sqrt{2(S + N)}]$, and $\gamma = \int_{-\infty}^{X+L} dy p_X(y) + \int_{X+L}^{\infty} dy p_{X+2L}(y) = \text{erfc}[-L/\sqrt{2(S + N)}]$. Therefore, Eq. (6) is also satisfied. The efficacy parameter satisfies $\gamma > 1$ for all (S, N) as long as $L > 0$ so that the particle is transported to the right. In addition, γ is a monotonically decreasing function of N, and $\gamma \to 1$ holds with $N \to \infty$ which implies that feedback control does not work at all because of an infinite amount of error. We note that $\langle W \rangle$ is positive for a regime with a small S/N-ratio even though $\gamma > 1$ always holds. In this regime, the second law (2) is satisfied while the Jarzynski equality (1) is violated, as mentioned before.

Proof of the main results.—Let $\Gamma^{\dagger}(t) \equiv \Gamma^{*}(\tau - t)$ be the time-reversed trajectory of $\Gamma(t)$. With control protocols $\lambda(t; y)$ and $\lambda^{\dagger}(t; y)$, we denote the probability densities of trajectories $\Gamma(t)$ and $\Gamma^{\dagger}(t)$ as $\mathcal{P}_{\lambda(t;y)}[\Gamma(t)]$ and $\mathcal{P}_{\lambda^{\dagger}(t;y)}[\Gamma^{\dagger}(t)]$, respectively. These are respectively normalized as $\int \mathcal{P}_{\lambda(t;y)}[\Gamma(t)]\mathcal{D}[\Gamma(t)] = 1$ and $\int \mathcal{P}_{\lambda^{\dagger}(t;y)}[\Gamma^{\dagger}(t)]\mathcal{D}[\Gamma^{\dagger}(t)] = 1$, where $\mathcal{D}[\Gamma(t)] = \mathcal{D}[\Gamma^{\dagger}(t)]$. It has been well-established that without any feedback control the local detailed balance holds for any control protocol [2], which is given by $e^{-\sigma} = \mathcal{P}_{\lambda^{\dagger}(t;y)}[\Gamma^{\dagger}(t)]/\mathcal{P}_{\lambda(t;y)}[\Gamma(t)]$ with protocol $\lambda(t; y)$, where y is fixed.

The joint distribution of $\Gamma(t)$ and y is given by $P[y|\Gamma_m]\mathcal{P}_{\lambda(t;y)}[\Gamma(t)]$. Noting that $e^{-I} = P[y]/P[y|\Gamma_m]$, we have $\langle e^{-\sigma-I} \rangle = \int dy \mathcal{D}[\Gamma(t)]P[y|\Gamma_m]\mathcal{P}_{\lambda(t;y)}[\Gamma(t)] \times e^{-\sigma[\Gamma(t)]}P[y]/P[y|\Gamma_m] = \int \mathcal{D}[\Gamma(t)]dy \mathcal{P}_{\lambda^{\dagger}(t;y)}[\Gamma^{\dagger}(t)]P[y] = 1$, which proves Eq. (4).

To prove Eq. (7), we use the assumption of the time-reversal symmetry of the measurement, $P[y^{*}|\Gamma_m^{*}] = P[y|\Gamma_m]$. The joint distribution of $\Gamma^{\dagger}(t)$ and y' under the protocol $\lambda^{\dagger}(t; y)$ is given by $P[y'|\Gamma_m^{*}]\mathcal{P}_{\lambda^{\dagger}(t;y)}[\Gamma^{\dagger}(t)]$ so that $P_{\lambda(t;y)}[y'] = \int \mathcal{D}[\Gamma^{\dagger}(t)]P[y'|\Gamma_m^{*}]\mathcal{P}_{\lambda^{\dagger}(t;y)}[\Gamma^{\dagger}(t)]$ for arbitrary y and y'. Therefore we obtain

$$\langle e^{-\sigma} \rangle = \int dy \mathcal{D}[\Gamma(t)]P[y|\Gamma_m]\mathcal{P}_{\lambda(t;y)}[\Gamma(t)]e^{-\sigma[\Gamma(t)]}$$

$$= \int dy \mathcal{D}[\Gamma^{\dagger}(t)]P[y^{*}|\Gamma_m^{*}]\mathcal{P}_{\lambda^{\dagger}(t;y)}[\Gamma^{\dagger}(t)]$$

$$= \int dy P_{\lambda^{\dagger}(t;y)}[y^{*}] \qquad (11)$$

which proves Eq. (7).

In conclusion, we have generalized the Jarzynski equality to situations in which we perform feedback control on a nonequilibrium dynamics. The first generalization (4) includes the mutual information, and it leads to the generalized second law (3) and the generalized fluctuation-dissipation theorem (5) corresponding to the first and second cumulants, respectively. The second generalization (6) includes the efficacy parameter γ, which can be determined by backward processes, and it characterizes the efficacy of feedback, as shown in (8). We have also illustrated the equalities by the Szilard engine and an information ratchet. We note that our results are consistent with the conventional second law of thermodynamics because of the energy cost of the controller [8].

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

[2] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993); G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995); G. E. Crooks, Phys. Rev. E **60**, 2721 (1999); C. Maes, J. Stat. Phys. **95**, 367 (1999); C. Maes, J. Math. Phys. (N.Y.) **41**, 1528 (2000); C. Jarzynski, J. Stat. Phys. **98**, 77 (2000); T. Hatano and S.-I. Sasa, Phys. Rev. Lett. **86**, 3463 (2001); T. Harada and S.-I. Sasa, Phys. Rev. Lett. **95**, 130602 (2005); R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007); A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. E **78**, 011107 (2008); S. Vaikuntanathan and C. Jarzynski, arXiv:0909.3457.

[3] G. M. Wang *et al.*, Phys. Rev. Lett. **89**, 050601 (2002); J. Liphardt *et al.*, Science **296**, 1832 (2002); E. H. Trepagnier *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 15 038 (2004); D. Collin *et al.*, Nature (London) **437**, 231 (2005); S. Toyabe *et al.*, Phys. Rev. E **75**, 011122 (2007).

[4] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (John Wiley & Sons, New York, 1985), 2nd ed.

[5] F. J. Cao, L. Dinis, and J. M. R. Parrondo, Phys. Rev. Lett. **93**, 040603 (2004); B. J. Lopez *et al.*, Phys. Rev. Lett. **101**, 220601 (2008).

[6] K. H. Kim and H. Qian, Phys. Rev. E **75**, 022102 (2007); F. J. Cao and M. Feito, Phys. Rev. E **79**, 041118 (2009); M. Bonaldi *et al.*, Phys. Rev. Lett. **103**, 010601 (2009); H. Suzuki and Y. Fujitani, J. Phys. Soc. Jpn. **78**, 074007 (2009).

[7] R. D. Astumian, Science **276**, 917 (1997); V. Serreli *et al.*, Nature (London) **445**, 523 (2007); E. R. Kay, D. A. Leigh, and F. Zerbetto, Angew. Chem. **119**, 72 (2007).

[8] J. C. Maxwell, *Theory of Heat* (Appleton, London, 1871); L. Szilard, Z. Phys. **53**, 840 (1929); *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, edited by H. S. Leff and A. F. Rex (Princeton University Press, New Jersey, 2003); K. Maruyama, F. Nori, and V. Vedral, Rev. Mod. Phys. **81**, 1 (2009); T. Sagawa and M. Ueda, Phys. Rev. Lett. **102**, 250602 (2009).

[9] T. Sagawa and M. Ueda, Phys. Rev. Lett. **100**, 080403 (2008).

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 1991).