Proceedings of the 2007 American Control Conference
Marriott Marquis Hotel at Times Square
New York City, USA, July 11-13, 2007

The Statistical Mechanics of Fluctuation-Dissipation and Measurement Back Action

Henrik Sandberg, Jean-Charles Delvenne, and John C. Doyle

Abstract— **In this paper, we take a control-theoretic approach to answering some standard questions in statistical mechanics. A central problem is the relation between systems which appear macroscopically dissipative but are microscopically lossless. We show that a linear macroscopic system is dissipative if and only if it can be approximated by a linear lossless microscopic system, over arbitrarily long time intervals. As a by-product, we obtain mechanisms explaining Johnson-Nyquist noise as initial uncertainty in the lossless state, as well as measurement back action and a trade-off between process and measurement noise.**

I. Introduction

The derivation of thermodynamics as a theory of large systems which are microscopically governed by fundamental laws of physics (Newton's laws or quantum physics) has a large literature and tremendous progress for over a century within the field of statistical physics. See for instance [1] for a physicist's account of statistical mechanics. Nevertheless, from a control theorist's perspective, there are inadequacies in the existing treatment both with the level of mathematical rigor, and the applicability to far-from-equilibrium systems, particularly when subject to complex regulatory mechanisms. Substantial work has already been done in formulating various results of classical thermodynamics in a more mathematical framework (e.g. [2]–[6] is a small sample), but statistical mechanics has received much less comparable attention. This paper focuses on simple problems in statistical mechanics in which the issue of rigor can be pursued, but aims also to set the stage for broader applicability.

In particular, we construct a simple and clear control-theoretic modeling framework in which the only assumptions on the nature of the physical systems are conservation of energy and causality and all systems are of finite dimension and act on finite time horizons. We construct high-order lossless systems that approximate low-order dissipative systems in a systematic manner, and prove that a linear model is dissipative if and only if it is arbitrarily well approximated by lossless causal linear systems over an arbitrary long time horizon. We show how the error between the systems depend on the number of states in the approximation and the length of the time horizon. Since human experience is based on a finite window of space and time, we argue that no human can directly distinguish between a low-order macroscopic dissipative system and its high-order lossless approximation.

The lossless systems studied here are consistent with classical physics, since they conserve energy, are causal, and are time reversible. Uncertainty in their initial state gives a simple explanation of the *Johnson-Nyquist noise* that can be observed at a macroscopic level. We also derive some well-known results from statistical mechanics, including the *fluctuation-dissipation theorem*. As a further application, we study the implications of these results for an idealized measurement device, and exhibit a back-action effect, that there is no precise measurement without perturbation on the measured system, that arises naturally in a purely classical setting.

We hope this paper is a step towards building a framework for understanding fundamental limitations in control and estimation that arise due to the physical implementation of measurement and actuation devices. We defer many important and difficult issues here such as how to actually model measurement devices realistically. It is also clear that this framework would benefit from a behavioral setting [7]. However, for the points we make with this paper, a conventional input-output setting with only regular interconnections is sufficient. Aficionados will easily see the generalizations, the details of which might be an obstacle to readability for others. Perhaps the most glaring unresolved issue is how to best motivate the introduction of stochastics. In conventional statistical mechanics, a stochastic framework is taken for granted, whereas we aim to explain if and when stochastics arise naturally, and in this we are only partially successful.

The organization of the paper is as follows: In Section II, we define the class of linear lossless/causal systems. In Section III, we derive lossless/causal approximations of memoryless dissipative systems and obtain Johnson-Nyquist noise. In Sections IV and V, we discuss interconnections of systems and introduce an idealized measurement device with back action. Finally, in Section VI we generalize the procedure from Section III to a class of linear dissipative systems with memory, and in Section VII obtain the fluctuation-dissipation theorem.

II. Lossless/Causal Linear Systems

In this paper, we consider linear systems in the form

$$\begin{aligned} \dot{x}(t) &= Jx(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, \\ y(t) &= B^T x(t), \end{aligned} \tag{1}$$

H. Sandberg is supported by the Hans Werthén foundation and a post-doctoral grant from the Swedish Research Council.

H. Sandberg and J.C. Doyle are with California Institute of Technology, Control and Dynamical Systems, M/C 107-81, Pasadena, CA 91125, USA. {henriks,doyle}@cds.caltech.edu

J.-C. Delvenne is with Imperial College, Institute for Mathematical Sciences, 53 Prince's Gate, South Kensington, London, SW7 2PG, UK. jc.delvenne@imperial.ac.uk. This research was partly supported by the FNRS and the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office. The scientific responsibility rests with its authors.

where $J = -J^T$ and (J, B) is controllable. It is assumed that the input $u(t)$ and the output $y(t)$ are scalars. We define the internal energy of (1) as

$$U(x(t)) \triangleq \frac{1}{2}x(t)^T x(t).$$

We argue these systems have desirable "physical" properties. These properties are losslessness and causality.

Lossless [8], [9] means that the internal energy satisfies

$$\frac{dU(x(t))}{dt} = x(t)^T \dot{x}(t) = y(t)u(t) \triangleq w(t),$$

where $w(t)$ is the *work rate* on the system. If there is no work done on the system, $w(t) = 0$, then the internal energy $U(t)$ is constant and conserved. If there is work done on the system, $w(t) > 0$, the internal energy increases. The work, however, can be extracted again, $w(t) < 0$, since the energy is conserved and the system is controllable. Conservation of energy is a common assumption on microscopical models in statistical mechanics [1].

Causal here means that there is no direct term between the input u and the output y. This means that there is no instantaneous reaction of the system. Also this is a reasonable physical assumption.

Definition 1: Systems (1) that satisfy the above assumptions are simply called *lossless/causal systems*.

Later we will seek approximations of dissipative systems in the class of lossless/causal systems.

The lossless/causal systems are rather abstract but have properties that we argue are reasonable from a physical point of view, as illustrated by the following example.

Example 1: The inductor-capacitor circuit in Fig. 1 with $u = i$ and $y = v_1$ can be modeled by the lossless/causal system

$$\dot{x} = \begin{pmatrix} 0 & -1/\sqrt{C_1 L_1} & 0 \\ 1/\sqrt{C_1 L_1} & 0 & -1/\sqrt{L_1 C_2} \\ 0 & 1/\sqrt{L_1 C_2} & 0 \end{pmatrix} x + \begin{pmatrix} 1/\sqrt{C_1} \\ 0 \\ 0 \end{pmatrix} u,$$

$$y = \begin{pmatrix} 1/\sqrt{C_1} & 0 & 0 \end{pmatrix} x, \quad x^T = \begin{pmatrix} \sqrt{C_1}v_1 & \sqrt{L_1}i_1 & \sqrt{C_2}v_2 \end{pmatrix},$$

$$U = \frac{1}{2}x^T x = \frac{1}{2}(C_1 v_1^2 + L_1 i_1^2 + C_2 v_2^2), \quad w = yu = v_1 i.$$

In fact, all linear minimal lossless single-input–single-output system with supply rate $y(t)u(t)$ can be written in the form (1), see [9, Theorem 5].

III. Lossless/Causal Approximations of Dissipative Memoryless Systems

In this section, we see how dissipative models, models where energy disappears, can be approximated by the lossless/causal systems. We start with simple memoryless models, which give rise to heat baths and Johnson-Nyquist noise.

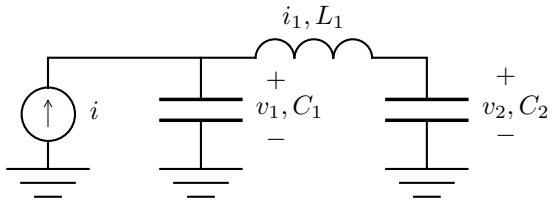

Fig. 1. The inductor-capacitor circuit in Example 1.

A. Dissipative memoryless systems

Many times macroscopic systems, such as resistors, can be modeled approximately by simple input-output relations

$$y(t) = ku(t), \tag{2}$$

where k is a scalar. If $k > 0$, the system is dissipative since we can never extract any work. This is because the work rate is always positive, $w(t) = y(t)u(t) = ku(t)^2 \geq 0$, for all t and u. Hence, (2) is neither lossless nor causal. Next, we show how we can approximate (2) arbitrarily well with a lossless/causal system over *finite*, but arbitrarily long, time horizons.

First, choose a time interval of interest, $[0, \tau]$, and rewrite (2) using a convolution integral

$$y(t) = \int_0^\tau k\delta(t-s)u(s)ds, \tag{3}$$

when u is at least continuous and has compact support on $[0, \tau]$, and δ is the Dirac distribution. Let us call τ the *recurrence time* of the model. The recurrence time interval contains all the time instants where we perform experiments on the model, and can be very long. Over this time interval, the system is equally well modeled by the impulse response $\kappa(t) = \sum_{l=-\infty}^{\infty} k\delta(t - l2\tau)$ which is a 2τ-periodic distribution. $\kappa(t)$ can be expanded in a Fourier series with convergence in the sense of distributions:

$$\kappa(t) \sim \frac{k}{2\tau} + \sum_{l=1}^{\infty} \frac{k}{\tau} \cos l\omega_0 t, \tag{4}$$

where $\omega_0 = \pi/\tau$. Define the truncated Fourier series by $\kappa_N(t) \triangleq k/(2\tau) + \sum_{l=1}^{N}(k/\tau)\cos l\omega_0 t$. We can split $\kappa_N(t)$ into its causal and anti-causal parts:

$$\kappa_N(t) \triangleq \kappa_N^c(t) + \kappa_N^{ac}(t)$$
$$\kappa_N^c(t) = 0 \quad (t < 0), \quad \kappa_N^{ac}(t) = 0 \quad (t \geq 0).$$

We can realize the causal part $\kappa_N^c(t)$ as the impulse response of a lossless/causal system of order $2N+1$ with the matrices

$$J_N = \begin{bmatrix} 0 & \Omega_N & 0 \\ -\Omega_N^T & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \Omega_N = \mathrm{diag}\{\omega_0, 2\omega_0, \ldots, N\omega_0\},$$

$$C_N = \sqrt{\frac{k}{\tau}} \begin{pmatrix} 1 & \ldots & 1 & 0 & \ldots & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}, \quad B_N = C_N^T. \tag{5}$$

That the series (4) converges in the sense of distributions means that for all smooth u of compact support on $[0, \tau]$ we have that

$$ku(t) = \lim_{N\to\infty} \int_0^\tau \left(\kappa_N^{ac}(t-s) + \kappa_N^c(t-s)\right) u(s)ds.$$

A closer study of the two terms in the integral reveals that

$$\lim_{N\to\infty} \int_0^\tau \kappa_N^{ac}(t-s)u(s)ds = \frac{1}{2}ku(t+),$$
$$\lim_{N\to\infty} \int_0^\tau \kappa_N^{c}(t-s)u(s)ds = \frac{1}{2}ku(t-),$$

because of the anti-causal/causal decomposition. Hence, since u is continuous, we can model $y(t) = ku(t)$ with only the causal part if we normalize the causal part with a factor two.

We identify the lossless/causal approximation of (2) with a linear operator $K_N : \mathcal{C}^2(0,\tau) \to \mathcal{C}^2(0,\tau)$:

$$y_N(t) = K_N u(t): \quad y_N(t) = \int_0^t 2\kappa_N^c(t-s)u(s)ds.$$

It is realized by the triple $(J_N, \sqrt{2}B_N, \sqrt{2}C_N)$. We can bound the approximation error as seen in the following proposition.

Proposition 1: Assume that $u \in \mathcal{C}^2(0,\tau)$ and $u(0) = 0$. Let $y(t) = ku(t)$, $k > 0$, and $y_N(t) = K_N u(t)$. Then

$$|y(t) - y_N(t)| \leq \frac{2k\tau}{\pi^2 N}\left(|\dot{u}(t)| + |\dot{u}(0)| + \|\ddot{u}\|_{L_1[0,t]}\right),$$

for t in $[0,\tau]$.

Proof: We have that $y(t) - y_N(t) = \sum_{l=N+1}^{\infty}(2k/\tau)\int_0^t \cos l\omega_0(t-s)u(s)ds$, $t \in [0,\tau]$. We have changed the order of summation and integration because this is how the value of the series is defined in distribution sense. We proceed by using repeated integration by parts on each term in the series. We have $\int_0^t \cos l\omega_0(t-s)u(s)ds = [\int_0^t \sin l\omega_0(t-s)\dot{u}(s)ds]/(l\omega_0) = [\dot{u}(t) - \dot{u}(0)\cos l\omega_0 t - \int_0^t \cos l\omega_0(t-s)\ddot{u}(s)ds]/(l^2\omega_0^2)$. Hence, we have the bound $|y(t) - y_N(t)| \leq (2k/\tau)\sum_{l=N+1}^{\infty}\left(|\dot{u}(t)| + |\dot{u}(0)| + \int_0^t |\ddot{u}(s)|ds\right)/(l^2\omega_0^2)$. Since $\sum_{l=N+1}^{\infty} 1/l^2 \leq 1/N$, we can establish the bound in the proposition. ∎

The proposition shows that by choosing N sufficiently large, we can approximate the memoryless model (2) as well as we like with a lossless/causal system, if inputs are smooth. It is a reasonable assumption that inputs, such as voltages, are smooth since we usually cannot change them arbitrarily fast due to physical limitations. Physically, we can think of $2N+1$ as the number of degrees of freedom in a resistor. This is usually a number with the size of Avogadro's number, $N \approx 10^{23}$. Then the recurrence time τ can be very large without a significant error. This explains how the dissipative model (2) is consistent with a physics based on energy conserving systems.

B. Initial conditions in K_N

The general solution to the lossless/causal approximation K_N is

$$y_N(t) = \sqrt{2}B_N^T e^{J_N t}x(0) + \int_0^t 2\kappa_N^c(t-s)u(s)ds, \quad (6)$$

where J_N and B_N are defined in (5), and $x(0)$ is the initial state. It is the second part of the solution that approximates $ku(t)$. The first part, the homogeneous solution, is not desired in the approximation, but is always present for a linear dynamical system. Next, we study the influence of this term.

Proposition 1 suggests that we will need a system of incredibly high order to approximate the dissipative system (2) on a reasonably long time horizon. When dealing with systems of such extremely high dimensions, it is reasonable to assume that the exact initial state $x(0)$ is not known. Therefore, we will take a statistical approach to study its influence.

We have that

$$\mathbf{E}y_N(t) = \sqrt{2}B_N^T e^{J_N t}\mathbf{E}x(0) + \int_0^t 2\kappa_N^c(t-s)u(s)ds,$$

if the input u is deterministic and known. The covariance function for $y_N(t)$ is then

$$\begin{aligned} R_{y_N}(s,t) &\triangleq \mathbf{E}[y_N(t) - \mathbf{E}y_N(t)][y_N(s) - \mathbf{E}y_N(s)] \\ &= 2B_N^T e^{J_N t}Xe^{-J_N s}B_N, \quad (7) \end{aligned}$$

where X is the covariance of the initial state,

$$X \triangleq \mathbf{E}[x(0) - \mathbf{E}x(0)][x(0) - \mathbf{E}x(0)]^T. \quad (8)$$

In Section III-D, we discuss how it is reasonable to choose X. The arguments are information theoretical and physical in nature. Both arguments result in an equipartition-type statement that result in the concept of temperature. For now, let us only define the notion of temperature of a lossless/causal system.

Definition 2 (Temperature): A lossless/causal system with deterministic input has temperature T (T is scalar) if

$$R_y(s,t) = T \cdot B^T e^{J(t-s)}B.$$

If X commutes with J_N and admits B_N as an eigenvector with eigenvalue T, (7) satisfies Definition 2 and we have (in the sense of distributions)

$$R_{y_N}(s,t) \to 2Tk\delta(t-s), \quad t,s \in [0,\tau], \quad N \to \infty. \quad (9)$$

A stochastic signal with this property is called *white noise*.

C. Johnson-Nyquist noise

From Proposition 1, (6), and (9) we obtain the following proposition.

Proposition 2: In the limit when $N \to \infty$, the lossless/causal system K_N, given by (6), converges to

$$y_\infty(t) = ku(t) + \sqrt{2Tk}w(t), \quad t \in [0,\tau], \quad (10)$$

when it has temperature T. The signal $w(t)$ is stochastic white noise of unit intensity. The input $u(t)$ should satisfy the assumptions of Proposition 1.

Definition 3 (Heat bath): A system (10) is called a *heat bath* of strength k, temperature T, and recurrence time τ.

Hence, in the limit, the uncertainty in the initial state of the microscopic lossless/causal system K_N is transformed into white noise added to the output of the macroscopic model (2). This is a generalization of Johnson-Nyquist noise of resistors, see [10], [11]: It is a fact that careful measurements of the voltage across a resistor reveal that there is noise that depends on the resistance and temperature. Usually this noise is modeled by stochastic white noise. The noise is often explained using methods from statistical mechanics and circuit theory. See, for example, [1]. Here we obtain exactly

the same result using lossless/causal systems and a suitable definition of temperature.

Remark 1: That Proposition 2 indeed leads to the standard form of the Johnson-Nyquist noise of a resistor can be seen as follows: We have $v = Ri$ from Ohm's law. Assume that $i = 0$ and study the variance of $v(t)$ through a low-pass filter of bandwidth B. Then we have, since $|\hat{R}_w(j\omega)|^2 = 1$ (white noise), $\mathbf{E}v(t)^2 = \int_{-B}^{B} 2TR|\hat{R}_w(j\omega)|^2 d\omega = 4TRB$, which is usually how Johnson-Nyquist noise is presented. Notice that Boltzmann's constant here should be included in the temperature T. It is also interesting to notice that the factor two in the noise intensity $2TR$ in our derivation originates from the causal/anti-causal decomposition in the construction of K_N. A very different argument is used in the derivation in [1].

D. Equipartition of energy

In this section, we discuss how the covariance of the initial state $x(0)$ of K_N, defined in (8), should be chosen. This discussion leads up to the definition of temperature, Definition 2. The first argument is information theoretical, and the second argument has a more physical flavor. As mentioned in the introduction, how to properly motivate the introduction of the stochastic element is not easy. Here we just give two arguments whose consequences are compatible with macroscopic observations, if Johnson-Nyquist noise is modeled by stochastic white noise. Neither of the arguments is entirely convincing, and we hope to return to these issues elsewhere.

MaxEnt argument: The first argument is based on the MaxEnt principle, due to Jaynes [12], [13]. This means that we should assign the distribution of $x(0)$ that maximizes the Shannon entropy of the distribution subject to all known constraints. The procedure is justified because it leads to the least biased guess. Assume that the expected internal energy of the initial state is E:

$$E = \mathbf{E}\frac{1}{2}x(0)^T x(0) = \frac{1}{2}\mathbf{E}x(0)^T \mathbf{E}x(0) + \frac{1}{2}\mathrm{Tr}X.$$

Maximization of the Shannon entropy subject to this constraint leads to a distribution of $x(0)$ that is Gaussian with mean zero and with covariance matrix

$$X = \frac{2E}{2N+1} \cdot I_{2N+1}.$$

If we define the temperature T as $2E/(2N+1)$ and use this X in (7), we see that the covariance function of y_N satisfies the requested relation in Definition 2. This means that the energy is distributed equally between all degrees of freedom. We have equipartition. The temperature is the expected amount of energy (up to a factor two) of each degree of freedom. This coincide with the usual notion of temperature in physics.

White noise argument: Assume that K_N had temperature zero a long time back, i.e., $x(-h) = 0$ where h is a large number. We will be more precise about the size of h later. We start our experiment at time $t = 0$ and wonder what a reasonable assumption on the initial state $x(0)$ is. Let us now assume that K_N has been subject to low-intensity white noise over the time interval $[-h, 0]$, i.e., $\mathbf{E}u(t)u(s) = (i/h)\delta(t-s)$, $\mathbf{E}u(t) = 0$, where i is an intensity constant. One can say that K_N has been weakly connected to an even larger heat bath for a long time.

In the end, we want to compute R_{y_N} as defined in (7), and it is of interest to compute X. We have

$$\begin{aligned} X = \mathbf{E}x(0)x(0)^T &= \frac{2i}{h}\int_{-h}^{0} e^{-J_N s} B_N B_N^T e^{J_N s} ds \\ &= \frac{2i}{h}\int_{-h}^{0} \frac{k}{\tau} \begin{pmatrix} \cos\omega_0 s \\ \cos 2\omega_0 s \\ \vdots \\ \sin\omega_0 s \\ \sin 2\omega_0 s \\ \vdots \\ 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \cos\omega_0 s \\ \cos 2\omega_0 s \\ \vdots \\ \sin\omega_0 s \\ \sin 2\omega_0 s \\ \vdots \\ 1/\sqrt{2} \end{pmatrix}^T ds. \end{aligned}$$

Notice that if $h = 2\tau$ we have that $X = (ik/\tau)I_{2N+1}$. This is the amount of time the white noise needs to excite all the modes equally. When $h > 2\tau$ we can use that

$$\begin{aligned} \lim_{h\to\infty} \frac{1}{h}\int_{-h}^{0} \cos k\omega_0 s \ \cos l\omega_0 s \, ds &= \frac{1}{2}\delta_{k-l} \\ \lim_{h\to\infty} \frac{1}{h}\int_{-h}^{0} \sin k\omega_0 s \ \cos l\omega_0 s \, ds &= 0. \end{aligned}$$

Hence we have that $X \to (ik/\tau)I_{2N+1}$, when $h \to \infty$, and from (7) that

$$R_{y_N}(s,t) = 2B_N^T e^{J_N t} X e^{-J_N s} B_N = \frac{ik}{\tau} 2B_N^T e^{J_N(t-s)} B_N.$$

According to Definition 2, the temperature of K_N is $T = ik/\tau$.

IV. Interconnections

Definition 4: The *physical interconnection* of the lossless/causal system (J_1, B_1, B_1^T) to the lossless/causal system (J_2, B_2, B_2^T) is given by

$$\begin{aligned} \frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} J_1 & -B_1 B_2^T \\ B_2 B_1^T & J_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u \\ y &= B_1^T x_1. \end{aligned}$$

The physical interconnection is still lossless/causal. The interconnection makes physical sense if one studies interconnections of circuit or mechanical models, for example. It is also a neutral interconnection, as defined in [8]. Motivated by this definition, and that we in Section III showed that the lossless/causal system $(J_N, \sqrt{2}B_N, \sqrt{2}C_N)$ converges to a heat bath, we make the following definition.

Definition 5: The *physical interconnection* of the lossless/causal system (J, B, B^T) to a heat bath of strength k, temperature T, and recurrence time τ, is given by

$$\begin{aligned} \dot{x}(t) &= (J - kBB^T)x(t) + Bu(t) - B\sqrt{2kT}w(t) \\ y(t) &= B^T x(t), \end{aligned} \tag{11}$$

for $t \in [0, \tau]$, where w is stochastic white noise of unit intensity.

Notice that even though (J, B, B^T) is lossless, when connected to the heat bath, (11) looks dissipative since the eigenvalues of $J - kBB^T$ have negative real parts. This is essentially a *Langevin equation*.

V. Back Action of Linear Measurements

As a simple application of the results in Section III and the definitions in Section IV, consider the problem of measuring the output $y(t)$ of the lossless/causal system (J, B, B^T). For this purpose, we define an idealized measurement device

$$y_m(t) = k_m y(t), \tag{12}$$

where $k_m > 0$ is a scalar, and the signal $y_m(t)$ is such that we can read it out perfectly. With such a measurement device, we can also read out the output $y(t) = y_m(t)/k_m$ perfectly.

Now we construct a slightly less idealized measurement device by replacing (12) by a lossless/causal approximation of (12). This is a more physical device, as argued before. According to Section III, we obtain

$$y_m(t) = k_m y(t) + \sqrt{2k_m T_m} w(t), \tag{13}$$

in the limit if the initial state of the measurement device is not perfectly known. T_m is the temperature of the device, and (13) is essentially a heat bath. If we make a physical interconnection of (J, B, B^T) to (13), we obtain

$$\begin{aligned} \dot{x}(t) &= (J - k_m BB^T)x(t) - B\sqrt{2k_m T_m} w(t), \\ \hat{y}(t) &\triangleq y_m(t)/k_m = B^T x(t) + \sqrt{\frac{2T_m}{k_m}} w(t), \end{aligned} \tag{14}$$

using (13) and Definition 5, where $\hat{y}(t)$ is an estimate of $y(t)$. Acting on the system (14) we have

$$\begin{aligned} &\text{process noise:} && p(t) \triangleq \sqrt{2k_m T_m} w(t) \\ &\text{measurement noise:} && m(t) \triangleq \sqrt{\frac{2T_m}{k_m}} w(t). \end{aligned}$$

The measurement device generates process noise and dissipation. This is called *back action* of measurements. This is a well-known phenomenon in quantum physics. Here we obtain a similar effect based on lossless/causal approximations and using physical interconnections. Also notice that it holds that

$$\mathbf{E} p(t) m(s) = 2T_m \delta(t - s). \tag{15}$$

The cross-covariance between process and measurement noise is independent of the amplification k_m of the measurement device. For large k_m, we get a good estimate of y, but on the other hand, the process noise gets large. Hence, there is a trade-off. It is only the temperature T_m of the measurement device that controls the trade-off in (15).

VI. Lossless/Causal Approximations of Dissipative Systems with Memory

In this section, we generalize the procedure from Section III to dissipative systems that have memory. We consider strictly stable linear causal systems G with impulse response g. Their input-output relation is given by

$$y(t) = \int_0^t g(t - s) u(s) ds. \tag{16}$$

The system (16) is dissipative with respect to the work rate $w(t) = y(t)u(t)$ if $\int_0^T y(t)u(t)dt \geq 0$, for all $T \geq 0$ and admissible $u(t)$. An equivalent condition, see [14], is that the transfer function is positive real

$$\operatorname{Re} \hat{g}(j\omega) \geq 0 \quad \text{for all} \quad \omega. \tag{17}$$

Here $\hat{g}(j\omega)$ is the Fourier transform of $g(t)$.

The following theorem shows that the system (16) is dissipative if and only if it can be approximated arbitrarily well by a lossless/causal system over any finite time horizon $[0, \tau]$.

Theorem 1: Assume that G is a linear (causal) system with impulse response g, such that $g \in L_1 \cap L_2(0, \infty)$ and $\dot{g} \in L_1(0, \infty)$. Then G is dissipative if and only if for all $\epsilon > 0$ and $\tau > 0$ there is a lossless/causal linear system G_τ with impulse response g_τ such that

$$\|g - g_\tau\|_{L_2[0,\tau]} \leq \epsilon. \tag{18}$$

Proof: See appendix. ■

Notice that Theorem 1 shows that a large class of dissipative systems (macroscopic systems) can be approximated by the lossless/causal systems we introduced in Section II.

VII. Fluctuation-dissipation Theorem

If a lossless/causal system satisfies Definition 2, then by definition we have

$$R_y(s, t) = T \cdot B^T e^{J(t-s)} B.$$

This can be said to be the *fluctuation* of the system. The response of the lossless/causal system to an impulse $u(t) = \delta(t)$ is

$$B^T e^{Jt} B.$$

If the lossless/causal system approximates a dissipative system over $[0, \tau]$, see Theorem 1, then the impulse response decays over this time interval. This represents the *dissipation* of the system. The expressions of the fluctuation and dissipation are equal up to a constant, the temperature T. This is a property that can be observed in physical systems close to equilibrium (and hence can be linearized).

VIII. Conclusions

In this paper, we defined the class of lossless/causal systems and used them to approximate dissipative systems. We obtained an if and only if characterization and gave explicit error bounds that depend on the time horizon and the order of the approximations. When applied to memoryless models, we saw that Nyquist-Johnson noise (macroscopic measurable noise) can be explained by uncertainty in the initial state of a lossless/causal approximation of very high order. We also saw that using these techniques, it was relatively easy to obtain a back-action effect of measurements. This gave rise to a trade-off between process and measurement noise.

Acknowledgment

The authors would like to thank Ben Recht for many helpful suggestions and comments.

REFERENCES

[1] G. H. Wannier, *Statistical Physics*. New York: Dover Publications, 1987.

[2] R. W. Brockett and J. C. Willems, "Stochastic control and the second law of thermodynamics," in *Proceedings of the IEEE Conference on Decision and Control*, San Diego, California, 1978, pp. 1007–1011.

[3] S. K. Mitter and N. J. Newton, "Information and entropy flow in the Kalman-Bucy filter," *Journal of Statistical Physics*, vol. 118, pp. 145–176, 2005.

[4] W. M. Haddad, V. S. Chellaboina, and S. G. Nersesov, *Thermodynamics: A Dynamical Systems Approach*. Princeton University Press, 2005.

[5] M. Barahona, A. C. Doherty, M. Sznaier, H. Mabuchi, and J. C. Doyle, "Finite horizon model reduction and the appearance of dissipation in Hamiltonian systems," in *Proceedings of the 41st IEEE Conference on Decision and Control*, vol. 4, 2002, pp. 4563–4568.

[6] D. S. Bernstein and S. P. Bhat, "Energy equipartition and the emergence of damping in lossless systems," in *Proceedings of the 41st IEEE Conference on Decision and Control*, Las Vegas, Nevada, 2002, pp. 2913–2918.

[7] J. W. Polderman and J. C. Willems, *Introduction to Mathematical Systems Theory — A Behavioral Approach*. Springer, 1997.

[8] J. C. Willems, "Dissipative dynamical systems part I: General theory," *Archive for Rational Mechanics and Analysis*, vol. 45, pp. 321–351, 1972.

[9] ——, "Dissipative dynamical systems part II: Linear systems with quadratic supply rates," *Archive for Rational Mechanics and Analysis*, vol. 45, pp. 352–393, 1972.

[10] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical Review*, vol. 32, pp. 97–109, 1928.

[11] H. Nyquist, "Thermal agitation of electrical charge in conductors," *Physical Review*, vol. 32, pp. 110–113, 1928.

[12] E. T. Jaynes, "Information theory and statistical mechanics i," *Physical Review*, vol. 106, pp. 620–630, 1957.

[13] ——, "Information theory and statistical mechanics ii," *Physical Review*, vol. 108, pp. 171–190, 1957.

[14] J. Slotine and W. Li, *Applied nonlinear control*. Upper Saddle River, New Jersey: Prentice Hall, 1991.

APPENDIX

Proof of Theorem 1: We first show the 'if' direction. Assume the opposite: There are lossless approximations that satisfy (18) even though G is not dissipative. If G is not dissipative, we can find an input $u(t)$ over an interval $[0,T]$ such that $\int_0^T y(t)u(t)dt = -K_1 < 0$, i.e., we extract energy from G even though its initial state is zero. Call $\|u\|_{L_1[0,T]} = K_2$ and $\|u\|_{L_2[0,T]} = K_3$. For any $\tau > T$ and $\epsilon > 0$ we thus have $\int_0^T (y_\tau(t) - y(t))u(t)dt \leq \epsilon K_2 K_3$, by the assumption that lossless approximations G_τ exist and using the Cauchy-Schwarz inequality. But the lossless approximation satisfies $\int_0^T y_\tau(t)u(t)dt = \frac{1}{2}x_\tau(T)^T x_\tau(T)$, since $x_\tau(0) = 0$. Hence, $-\int_0^T y(t)u(t)dt = K_1 \leq \epsilon K_2 K_3 - \frac{1}{2}x_\tau(T)^T x_\tau(T) \leq \epsilon K_2 K_3$. But since ϵ can be made arbitrarily small, this leads to a contradiction.

To prove the 'only if' direction we explicitly construct a G_τ that satisfies (18). We first need to make some definitions. Let $C \triangleq (2/\pi)\|\dot{g}\|_{L_1}$. Also define $\delta(t) \triangleq \int_t^\infty |g(s)|ds$ that is a continuously decreasing function that satisfies $\lim_{t\to\infty}\delta(t) = 0$. We need that the recurrence time τ is such that

$$\delta(\tau) \leq \frac{\epsilon^2}{8C}. \tag{19}$$

If the chosen τ does not satisfy (19), we can without loss of generality increase it to the smallest τ that satisfies (19). It is assumed that this has been done in the following.

The model G_τ we construct is based on a truncated version of the impulse response $g_{N,\tau}(t)$ where

$$g_{N,\tau}(t) = \frac{a_0}{2} + \sum_{k=1}^{N} a_k \cos\frac{k\pi t}{\tau}, \quad t \in [0,\tau],$$
$$a_k = \frac{2}{\tau}\int_0^\tau g(t)\cos\frac{k\pi t}{\tau}dt$$
$$\|g_{N,\tau}\|^2_{L_2[0,\tau]} = \frac{\tau}{4}a_0^2 + \frac{\tau}{2}\sum_{k=1}^{N} a_k^2 \leq \frac{\tau}{2}\sum_{k=0}^{N} a_k^2.$$

Assume that τ is fixed as in (19). Next pick the smallest N such that

$$\|g - g_{N,\tau}\|_{L_2[0,\tau]} \leq \frac{\epsilon}{2}. \tag{20}$$

Such an N always exist since $g \in L_2$ and the cos-terms are a basis in $L_2[0,\tau]$.

Define $\hat{g}_{N,\tau}(j\omega) \triangleq \int_0^\tau g(t)e^{-j\omega t}dt$, and notice that $a_k = (2/\tau)\mathrm{Re}\,\hat{g}_{N,\tau}(jk\pi/\tau)$. We have that $|\mathrm{Re}\,\hat{g}(j\omega) - \mathrm{Re}\,\hat{g}_{N,\tau}(j\omega)| = \left|\mathrm{Re}\int_\tau^\infty g(t)e^{-j\omega t}dt\right| \leq \|g\|_{L_1[\tau,\infty)} = \delta(\tau) \leq \epsilon^2/(8C)$, for all ω. Since, $\mathrm{Re}\,\hat{g}(j\omega) \geq 0$ for all ω by (17), we have $a_k \geq -\epsilon^2/(4C\tau)$. We need a second bound on a_k that bounds the rate of decay to zero. It holds that $a_k = (-2/\tau)\int_0^\tau \dot{g}(t)\frac{\tau}{k\pi}\sin\frac{k\pi t}{\tau}dt$, and thus $|a_k| \leq C/k$, independent of τ. Taken together, the bounds give that possibly strictly negative a_k, call them a_k^-, must satisfy

$$|a_k^-| \leq \min\left\{\frac{\epsilon^2}{4C\tau}, \frac{C}{k}\right\}. \tag{21}$$

Next, define $g_{N,\tau}(t) \triangleq g^+_{N,\tau}(t) + g^-_{N,\tau}(t)$, where $g^-_{N,\tau}(t)$ contains all the terms in $g_{N,\tau}(t)$ with strictly negative Fourier coefficients, $a_k = a_k^-$. Notice that $g^+_{N,\tau}$ can be realized with a linear lossless/causal system, compare with (5). We can now bound the worst-case L_2-norm of $g^-_{N,\tau}$. Using (21) we have

$$\|g^-_{N,\tau}\|^2_{L_2[0,\tau]} \leq \frac{\tau}{2}\sum_{k=0}^{N}(a_k^-)^2 \leq \sum_{k=0}^{\lfloor\frac{4C^2\tau}{\epsilon^2}\rfloor}\frac{\tau}{2}\frac{\epsilon^4}{16C^2\tau^2}$$
$$+\sum_{k=\lfloor\frac{4C^2\tau}{\epsilon^2}\rfloor+1}^{\infty}\frac{\tau}{2}\frac{C^2}{k^2} \leq \frac{4C^2\tau}{\epsilon^2}\frac{\tau\epsilon^4}{32C^2\tau^2} + \frac{\epsilon^2}{4C^2\tau}\frac{\tau C^2}{2} = \frac{\epsilon^2}{4},$$

independent of how large N is.

A lossless/causal approximation that satisfies the bound (18) is now given by $g_\tau(t) = g^+_{N,\tau}(t)$, where τ and N were fixed in (19) and (20). This is because the triangle inequality gives

$$\|g - g^+_{N,\tau}\|_{L_2[0,\tau]} \leq \|g - g_{N,\tau}\|_{L_2[0,\tau]} + \|g^-_{N,\tau}\|_{L_2[0,\tau]}$$
$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This concludes the proof.