

Information and Entropy Flow in the Kalman-Bucy Filter ^{*}

Sanjoy K. Mitter [†] Nigel J. Newton [‡]

Abstract

We investigate the information theoretic properties of Kalman-Bucy filters in continuous time, developing notions of information supply, storage and dissipation. Introducing a concept of *energy*, we develop a physical analogy in which the unobserved signal describes a statistical mechanical system interacting with a heat bath. The abstract ‘universe’ comprising the signal and the heat bath obeys a non-increase law of entropy; however, with the introduction of partial observations, this law can be violated. The Kalman-Bucy filter behaves like a Maxwellian demon in this analogy, returning signal energy to the heat bath without causing entropy increase. This is made possible by the steady supply of new information.

In a second analogy the signal and filter interact, setting up a stationary non-equilibrium state, in which energy flows between the heat bath, the signal and the filter without causing any overall entropy increase. We introduce a *rate of interactive entropy flow* that isolates the statistical mechanics of this flow from marginal effects. Both analogies provide quantitative examples of Landauer’s Principle.

KEYWORDS: Information Theory, Landauer’s Principle, Non-Equilibrium Statistical Mechanics, Statistical Filtering.

1 Introduction

In this article we study continuous-time Kalman-Bucy filters from information theoretic and statistical mechanical viewpoints. The information flows

^{*}This work was partially supported by MURI Grant F49620-02-1-0325 (Complex Adaptive Networks for Cooperstone Control), and by ARO-MURI Grant DAAD19-00-1-0466 (Data Fusion in Large Arrays of Microsensors (sensor web)).

[†]Department of Electrical Engineering and Computer Science, and Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. (mitter@mit.edu)

[‡]Department of Electronic Systems Engineering, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. (njn@essex.ac.uk) (Tel:int + 44 1206 872932)

for such filters are identified and these strongly resemble the *entropy* flows of non-equilibrium statistical mechanics, which occur when a statistical mechanical system is held away from its equilibrium state by an interaction with an exogenous system. (See, for example, [6] or [7].)

By introducing a concept of energy, we construct a physical analogy for the Kalman-Bucy filter, in which the partially observed signal interacts with a heat bath. The interaction forces the signal towards a stationary state, which maximises the entropy of the ‘universe’ comprising the signal and the heat bath. Whatever the initial state might be, it is impossible for this entropy to decrease at any stage during this convergence, and so our abstract universe obeys a law akin to the Second Law of Thermodynamics. However, this law can be broken in the presence of partial observations: the entropy of the abstract universe can be reduced (at least temporarily) at any rate up to that of the information supply from the observations. We show, in the analogy, that the filter behaves like a Maxwellian *demon* [15], extracting energy from the signal and returning it to the heat bath, thus ‘cooling’ the signal. The filter acts as a *heat pump* in this analogy but, unlike those of real physical heat pumps, its operations cause no overall increase in entropy. This is made possible by the steady supply of new observations.

In a second physical analogy, the *joint* system, comprising the signal and filter, interacts with a heat bath. We identify ‘conditional signal’ and filter subsystems, and show that energy flows around a loop comprising these subsystems and a heat bath. In the stationary state of this system, this energy flow occurs with no change in the overall entropy. Thus the system in the second physical analogy is a type of perpetual motion machine, and follows a type of non-equilibrium statistical mechanics. We use recent techniques in this field (see, for example, [2], [7] and [12]) to quantify the entropy flows in this system, and introduce a concept of *interactive entropy flow* to isolate the interaction of the components from their internal, autonomous non-equilibrium mechanics.

Our research is partly inspired by the doctoral thesis of Michael Propp [20], written under the direction of the first author. In this, an input-output view of a thermodynamic system is constructed by associating a Markov process with the system and then defining forces and fluxes for this process. A dissipation inequality, analogous to that of Willems [22] is then derived for this process. There, as in recent developments in non-equilibrium statistical mechanics and the results presented here, time reversibility plays an essential role. These ideas were also applied in [20] to the study of electrical networks involving Nyquist-Johnson resistors. (See, also, the related work in [4].)

Our work is also connected with ideas on the thermodynamics of computation, which have received much attention in recent decades. (See [1]

for a review article.) Because they can be investigated in the context of simple abstract universes comprising a few components, our physical analogies for the Kalman-Bucy filter provide precise, quantitative examples of Landauer's Principle, [11]. This states that, under certain circumstances, entropy can be increased by the erasure of information. Of course it is not our aim here to investigate the feasibility (or otherwise) of thermodynamically efficient computing machines.

The results in this article concern stable, time-homogeneous systems. They are generalised to a wider class of linear systems in [17], and to certain types of nonlinear system in [18]. These articles also investigate the effects of filter errors.

The specific problem addressed here is that of the evolution of a linear, partially-observed, Gaussian system over the finite time interval, $[-T, T]$, for some $0 < T < \infty$. A *symmetric* interval is chosen in order to lighten the notation when time reversal enters the theory; a *finite* interval is chosen since it admits the study of transient effects. The 'steady state' of the system can be investigated by means of appropriate initialisation. (Our model includes an 'initial' observation, which allows the signal and filter to be initialised in any consistent state.)

All random variables and processes will be defined on the 'filtered' probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \in [-T, T]), \mathbb{P})$. The signal, X , is an \mathbb{R}^n -valued process defined by the following Itô equation:

$$X_t = \xi + \int_{-T}^t AX_s ds + V_t \quad \text{for } t \in [-T, T]. \quad (1)$$

Here A is an $n \times n$ matrix, all eigenvalues of which have negative real parts, ξ is an \mathcal{F}_{-T} -measurable, \mathbb{R}^n -valued, Gaussian random variable with mean zero and positive-definite covariance matrix P_i , and $(V_t, \mathcal{F}_t, t \in [-T, T])$ is an n -vector Brownian motion with quadratic covariation

$$d[V, V']_t = \Sigma_V dt,$$

for some positive-definite $n \times n$ matrix Σ_V . V can be thought of as being the integral of a 'vector-valued white noise process with covariance matrix Σ_V '. (The reader is referred to [10] for the basic definitions and results of the Itô calculus.)

Of course X , thus defined, is a zero-mean vector Gaussian process with covariance matrix, $P(t) = \mathbf{E}X_t X_t'$, satisfying

$$P(t) = P_i + \int_{-T}^t (AP(s) + P(s)A' + \Sigma_V) ds.$$

The observation comprises two parts: the *initial* observation

$$Y^i = \xi + \zeta, \quad (2)$$

where ζ is an \mathcal{F}_{-T} -measurable, \mathbb{R}^n -valued, Gaussian random variable, independent of ξ , with mean zero and positive-definite covariance matrix M ; and the \mathbb{R}^n -valued *running* observation,

$$Y_t^r = \int_{-T}^t \Sigma_W X_s ds + W_t \quad \text{for } t \in [-T, T], \quad (3)$$

where Σ_W is a positive-semi-definite $n \times n$ matrix, and $(W_t, \mathcal{F}_t, t \in [-T, T])$ is an n -vector Brownian motion, independent of V , having quadratic co-variation

$$d[W, W']_t = \Sigma_W dt.$$

Remark 1.1. For the purposes of estimating X , Y^r is equivalent to the process

$$\tilde{Y}_t^r = \int_{-T}^t \Gamma X_s ds + \tilde{W}_t \quad \text{for } t \in [-T, T],$$

where Γ is an $n \times n$ matrix for which $\Gamma' \Gamma = \Sigma_W$, and $(\tilde{W}_t, \mathcal{F}_t, t \in [-T, T])$ is an n -dimensional, *standard* Brownian motion, independent of V .

A convenient way of combining Y^i and Y^r into a single observation process is to set

$$Y_t = CY^i + Y_t^r \quad \text{for } t \in [-T, T],$$

for some non-singular $n \times n$ matrix C . For an appropriately chosen C , Y could be thought of as being part of a running observation of the form (3) extending over the interval $(-\infty, T]$. In this interpretation Y^i would summarise the partial observations of the process X over times prior to $-T$. Since C plays no role in the information content of Y we economise on notation by choosing it to be the identity matrix; thus we shall use the *composite* observations process:

$$Y_t = Y^i + \int_{-T}^t \Sigma_W X_s ds + W_t \quad \text{for } t \in [-T, T]. \quad (4)$$

The Kalman-Bucy filter for X given Y is a recursive formula for calculating the conditional distribution of X_t given $(Y_s, s \in [-T, t])$, which is, of course, Gaussian. The *covariance form* of the filter propagates the mean, \hat{X} , and covariance matrix, Q , of the conditional distribution as follows:

$$\begin{aligned} \hat{X}_{-T} &= P_i(P_i + M)^{-1}Y^i, \\ d\hat{X}_t &= (A - Q(t)\Sigma_W) \hat{X}_t dt + Q(t) dY_t, \\ Q(-T) &= (P_i^{-1} + M^{-1})^{-1}, \\ \dot{Q}(t) &= AQ(t) + Q(t)A' + \Sigma_V - Q(t)\Sigma_W Q(t). \end{aligned} \quad (5)$$

(See, for example, [5] or [9].) Since P_t and M are positive definite, the inverse matrices here are well defined, and $P(t)$, $Q(t)$ and $P(t) - Q(t)$ are all positive definite for all t .

We denote by $(\mathcal{F}_t^Y, t \in [-T, T])$ the filtration generated by Y , and by ν the associated innovations process: for $t \in [-T, T]$,

$$\begin{aligned}\mathcal{F}_t^Y &= \sigma(Y_s, s \in [-T, t]) \\ \nu_t &= Y_t - \hat{X}_{-T} - \int_{-T}^t \Sigma_W \hat{X}_s ds.\end{aligned}\tag{6}$$

Of course, with this definition, (ν_t, \mathcal{F}_t^Y) is an n -vector Brownian motion with the same quadratic covariation as W , and non-zero initial value having the n -variate Gaussian distribution $N(0, M - Q(-T))$. (See, for example, [5].) (Here, and in what follows, we denote the multi-variate Gaussian distribution with mean vector μ and covariance matrix Σ by $N(\mu, \Sigma)$, and its density by $n(\mu, \Sigma)$.)

2 A Physical Analogy for the Signal.

In this section we explore the notion that the signal, X , of (1) can be thought of as describing an abstract statistical mechanical system. Under the conditions placed on A and Σ_V in the introduction X has the unique invariant distribution, $N(0, P_{SS})$, with positive-definite covariance matrix satisfying the following algebraic equation

$$AP_{SS} + P_{SS}A' + \Sigma_V = 0.\tag{7}$$

(See, for example, Section 5.6 in [10].)

The essential extra ingredient for the physical analogy is a concept of *energy* for the signal, and so we begin by defining a *Hamiltonian* for the signal:

$$H_X(x) = \frac{1}{2}x'P_{SS}^{-1}x.\tag{8}$$

For a probability measure μ on $(\mathbb{R}^n, \mathcal{B}^n)$, the *average energy*, $\mathcal{E}(\mu)$, *entropy*, $\mathcal{S}(\mu)$, and *free energy*, $\mathcal{F}(\mu)$, are then defined as follows:

$$\begin{aligned}\mathcal{E}(\mu) &= \int H_X(x)\mu(dx) \\ \mathcal{S}(\mu) &= - \int \log \left(\frac{d\mu}{d\lambda}(x) \right) \mu(dx) \quad \text{if the integral exists} \\ &\quad -\infty \quad \text{otherwise,} \\ \mathcal{F}(\mu) &= \mathcal{E}(\mu) - \mathcal{S}(\mu),\end{aligned}\tag{9}$$

where λ is Lebesgue (volume) measure. It can easily be shown by a variational argument that the free energy \mathcal{F} is minimised by the invariant distribution $N(0, P_{SS})$.

At time t , the average energy $E_X(t)$, entropy $S_X(t)$ and free energy $F_X(t)$ of the signal X are as follows:

$$\begin{aligned} E_X(t) &= \mathcal{E}(N(0, P(t))) = \frac{1}{2} \text{tr} (P(t) P_{SS}^{-1}), \\ S_X(t) &= \mathcal{S}(N(0, P(t))) = \frac{n}{2} (1 + \log(2\pi)) + \frac{1}{2} \log |P(t)|, \\ F_X(t) &= \frac{1}{2} \text{tr} (P(t) P_{SS}^{-1}) - \frac{n}{2} (1 + \log(2\pi)) - \frac{1}{2} \log |P(t)|. \end{aligned} \quad (10)$$

As t increases, energy flows into the signal at an average rate

$$\dot{E}_X(t) = \text{tr} (A(P(t) - P_{SS}) P_{SS}^{-1}), \quad (11)$$

which causes its entropy to increase at rate

$$\dot{S}_X(t) = \text{tr} (A(P(t) - P_{SS}) P(t)^{-1}). \quad (12)$$

(Of course, both of these rates could be negative, corresponding to an average *outflow* of energy.) It then easily follows that

$$\begin{aligned} \dot{F}_X(t) &= -\frac{1}{2} \text{tr} ((P_{SS}^{-1} - P(t)^{-1}) \Sigma_V (P_{SS}^{-1} - P(t)^{-1}) P(t)) \\ &\leq 0. \end{aligned}$$

In fact this ‘non-increase’ property of the free energy is true whatever the distribution of X_{-T} ; the positive definiteness of Σ_V ensures that, for every $t > -T$, X_t has a smooth density $p(\cdot, t)$ satisfying the Kolmogorov forward (Fokker-Planck) equation

$$\frac{\partial p}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} ((Ax)_i p) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} ((\Sigma_V)_{ij} p),$$

from which it follows that

$$\begin{aligned} \frac{d}{dt} \mathcal{F}(p(\cdot, t)) &= -\frac{1}{2} \int (P_{SS}^{-1} x + \nabla \log p)' \Sigma_V (P_{SS}^{-1} x + \nabla \log p) p(x, t) dx \\ &\leq 0. \end{aligned}$$

The process X can be thought of as describing the evolution of an abstract statistical mechanical system subject to random exogenous forces, which add or remove energy in order to drive the system towards its invariant distribution, and so minimise its free energy.

We follow the nomenclature of statistical mechanics by referring to invariant distributions as *stationary equilibrium states* or *stationary non-equilibrium states* according to the net *flow* of entropy at the invariant distribution, an equilibrium state being one for which this flow is zero. (See, for example, [2], [6] or [12].) In order to quantify entropy flow for the process X we first define its *entropy production*. This involves the time-reversed dynamics of X . For each $t \in [-T, T]$, let

$$\begin{aligned}\bar{X}_t &= X_{-t}, \\ \bar{\mathcal{F}}_t^X &= \sigma(\bar{X}_s, s \in [-T, t]), \\ \bar{A}(t) &= -A - \Sigma_V P(-t)^{-1}, \\ \bar{V}_t &= \bar{X}_t - \bar{X}_{-T} - \int_{-T}^t \bar{A}(s) \bar{X}_s ds.\end{aligned}\tag{13}$$

Proposition 2.1 (i) The process $(\bar{V}_t, \bar{\mathcal{F}}_t^X, t \in [-T, T])$ is an n -dimensional Brownian motion with quadratic covariation

$$d[\bar{V}, \bar{V}']_t = \Sigma_V dt.\tag{14}$$

(ii) For each $t \in [-T, T]$,

$$\bar{\mathcal{F}}_t^X = \sigma(\bar{X}_{-T}, (\bar{V}_s, s \in [-T, t])).\tag{15}$$

Proof. It follows from (1) and the definition of \bar{X} that for $s \leq t$

$$\mathbf{E} \bar{X}_s \bar{X}_t' = \exp(-A(t-s))P(-t).$$

Straightforward calculations now show that, for any $r \leq s \leq t$,

$$\mathbf{E} \bar{X}_r (\bar{V}_t - \bar{V}_s)' = 0,$$

and that (14) holds, and so \bar{V} is an independent-increments, Gaussian process, independent of \bar{X}_{-T} , and with the quadratic covariation of (14). It is thus a Brownian motion with this quadratic covariation with respect to the filtration it generates. It now follows from the definition of \bar{V} that

$$\bar{X} = \bar{\Phi}(\bar{X}_{-T}, \bar{V}),$$

where $\bar{\Phi}$ is the strong solution of the following Itô equation:

$$d\bar{X}_t = \bar{A}(t)\bar{X}_t dt + d\bar{V}_t,\tag{16}$$

and this establishes part (ii). This, and the independence of \bar{X}_{-T} and \bar{V} establish part (i). •

For each $-T \leq t - \epsilon < t + \epsilon \leq T$, let $\Pi_{t+\epsilon|t}^X$ and $\Pi_{t-\epsilon|t}^X$ be the X_t -conditional distributions of the processes $(X_u, u \in [t, t + \epsilon])$ and $(\bar{X}_u, u \in [-t, -t + \epsilon])$, respectively. Since the diffusion matrix Σ_V is positive definite it follows from the Cameron-Martin-Girsanov theory (see, for example, Chapter 6 in [13]) that $\Pi_{t+\epsilon|t}^X$ and $\Pi_{t-\epsilon|t}^X$ are mutually absolutely continuous probability measures with Radon Nikodym derivative (relative density)

$$\begin{aligned} \frac{d\Pi_{t+\epsilon|t}^X}{d\Pi_{t-\epsilon|t}^X}(X) &= \exp \left(\int_t^{t+\epsilon} X'_s (A - \bar{A}(s - 2t))' \Sigma_V^{-1} dV_s \right. \\ &\quad \left. + \frac{1}{2} \int_t^{t+\epsilon} X'_s (A - \bar{A}(s - 2t))' \Sigma_V^{-1} (A - \bar{A}(s - 2t)) X_s ds \right). \end{aligned}$$

Thus we may define the *rate of entropy production* of X at time $t \in (-T, T)$ as

$$\begin{aligned} R_X(t) &:= \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \mathbf{E} \log \left(\frac{d\Pi_{t+\epsilon|t}^X}{d\Pi_{t-\epsilon|t}^X}(X) \right) \\ &= \frac{1}{2} \text{tr} \left((A - \bar{A}(-t))' \Sigma_V^{-1} (A - \bar{A}(-t)) P(t) \right). \end{aligned} \tag{17}$$

Remark 2.1 $R_X(t)$ measures the *degree of time-asymmetry* of the process X at time t . Imagine a game in which one player secretly ‘cuts out’ the small segment of a sample path of X in the interval $(t - \epsilon, t + \epsilon)$, tosses a coin, reversing the time direction of the segment if ‘heads’ occurs, and then shows the other player the segment, asking whether or not it has been reversed. $R_X(t)$ is a measure of the average degree of ease with which the second player could answer correctly.

Remark 2.2 $R_X(t)$ would be infinite if Σ_V were singular, since it would then be possible, with probability one, for the second player in the game described above to distinguish time directions. For example, consider the case in which

$$A = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad \Sigma_V = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

The direction of time could be distinguished with probability one, here, from a comparison of the signs of $X_{t,2} - X_{t,1}$ and of the slope of $X_{t,2}$ at time t .

Remark 2.3 The rate of entropy production of the time-reversed process \bar{X} at time t is the same as that of the forward process X at time $-t$.

Remark 2.4 $R_X(t)$ is non-negative. It is zero if and only if X is in its invariant distribution at time t ($P(t) = P_{SS}$) and X is *self-adjoint* in the sense that

$$P_{SS} A' = A P_{SS}, \tag{18}$$

in which case $\bar{A}(t) = A$ for all t , and the dynamics of X are identical in both time directions.

Remark 2.5 R_X is *homeomorphism invariant*. If f is a continuous, one-to-one mapping from \mathbb{R}^n to \mathbb{R}^n , then it induces new probability measures on the space of continuous functions from $[t, t + \epsilon]$ to \mathbb{R}^n corresponding to $\Pi_{t+\epsilon|t}^X$ and $\Pi_{t-\epsilon|t}^X$. These can be used to define the rate of entropy production of the process $f(X_t)$. This is, of course, equal to $R_X(t)$.

We can now define the entropy flow of X (possibly away from its invariant distribution) as the difference between its rate of entropy production and its rate of change of entropy:

$$\Phi_X(t) := R_X(t) - \dot{S}_X(t). \quad (19)$$

Thus the entropy production comprises two parts: one that drives the process towards its stationary (minimum free energy) state, and another that represents net entropy flow. If X is self-adjoint in the sense of (18), then this flow is zero in the stationary state, and the latter is called an equilibrium state; otherwise it is called a non-equilibrium state.

In a more general context, stationary states of statistical mechanical systems are minimisers of free energies of the form

$$\mathcal{F}(\mu) = \mathcal{E}(\mu) - T\mathcal{S}(\mu),$$

where T is the *temperature* of the stationary state. (See, for example, [6] or [8].) Thus, we can consider our abstract system X as being in contact with a *heat bath* at unit temperature that supplies or removes heat in order to drive the system towards the stationary state $N(0, P_{SS})$. During this convergence, the entropy of X may increase or decrease according to the value of the covariance, $P(t)$. Of course, the entropy of the heat bath, $S_H(t)$, is also changed by this interaction. (Our heat bath is *idealised* in the sense that it can supply or absorb any finite amount of energy without suffering a temperature change.)

If we consider the combination of the signal and heat bath as a closed system in which energy is conserved, then, since the heat bath has unit temperature,

$$S_H(t) = K - E_X(t)$$

for some constant K , and it easily follows that

$$\frac{d}{dt}(S_X(t) + S_H(t)) = -\dot{F}_X(t) \geq 0.$$

Thus, the rate of increase of entropy for the ‘universe’ comprising the signal and the heat bath is non-negative, which shows that it obeys a law of non-decrease of entropy similar to the Second Law of Thermodynamics. The

fact that the temperature of the equilibrium state is unity is a consequence of the way in which H_X was defined in (8).

When X is in its stationary state there is *on average* no flow of energy between the heat bath and the signal. However, for *individual outcomes* of X there is a continuous exchange of energy back and forth between these components. It is these *fluctuations* that cause some energy to *loop* in the presence of observations. The invariant distribution represents a type of *dynamic equilibrium*.

In a well known paradox of statistical mechanics due to Maxwell, [15], a *demon* is able to make heat flow from a box containing a low temperature gas into an adjacent box containing a gas at higher temperature, thus (apparently) reducing the entropy of the system and violating the Second Law of Thermodynamics. It does this by observing the molecules of both gases in the vicinity of a (closable) hole connecting the two boxes. When a molecule of the cool gas with unusually high kinetic energy approaches the hole, the demon opens it allowing the molecule through. It does likewise when a molecule of the hot gas with unusually *low* kinetic energy approaches the hole from the other side. In fact it is generally accepted that this does not violate the Second Law since, in carrying out its role, the demon is not only reducing the entropy of the system of gases but also *erasing* the information held in the system's observational state. According to Landauer's Principle this erasure, in so much as it causes irreversibility, involves entropy increase in other components of the Universe. (See, for example, [11] or [1].) The demon can avoid irreversibility by retaining copies of all the observational state measurements it has used in performing its role. We show in the following section that the existence of the observation process Y of (4) allows an entropy reduction of the demonic type in the 'universe' comprising the signal and the heat bath.

3 The Role of Observations.

We begin this section by evaluating the information flows that occur in the Kalman-Bucy filter. Let $C(t)$ be the mutual information between X_t and $(Y_s, s \in [-T, t])$:

$$C(t) := I(X_t; (Y_s, s \in [-T, t])).$$

This can be thought of as being the observation-derived information on X_t *stored* by the filter at time t . Since $C(t)$ is a *mutual* information, it is not dependent on any underlying reference measure (such as Lebesgue (volume) measure). It therefore has *absolute* meaning, unlike quantities such as the signal entropy, $S_X(t)$ of (10). (For example, $C(t) = 0$ would imply that the observations up to time t were completely useless for estimating X_t .) Also, since \hat{X}_t is a sufficient statistic for the conditional distribution of X_t , $C(t)$

is also the mutual information between X_t and \hat{X}_t . Of course, $C(t)$ is also *homeomorphism invariant* in the sense that, for any continuous, one-to-one maps from \mathbb{R}^n to \mathbb{R}^n , f_1 and f_2 , the mutual information between $f_1(X_t)$ and $f_2(\hat{X}_t)$ is also $C(t)$.

It now easily follows that

$$\begin{aligned} C(-T) &= \frac{1}{2} \log |P_i + M| - \frac{1}{2} \log |M|, \\ \dot{C}(t) &= \frac{1}{2} \text{tr}(\Sigma_W Q(t)) - \frac{1}{2} \text{tr}(\Sigma_V(Q(t)^{-1} - P(t)^{-1})), \end{aligned} \tag{20}$$

from which it is tempting to think of the information *supply* to the store up to time t as being

$$S(t) = \frac{1}{2} \log |P_i + M| - \frac{1}{2} \log |M| + \frac{1}{2} \int_{-T}^t \text{tr}(\Sigma_W Q(s)) ds, \tag{21}$$

and the information *dissipated* from the store up to time t as being

$$D(t) = \frac{1}{2} \int_{-T}^t \text{tr}(\Sigma_V(Q(s)^{-1} - P(s)^{-1})) ds. \tag{22}$$

The following lemma justifies this interpretation. The result is not new; for example, it is a multi-dimensional, linear version of Lemma 16.9 in [14]. (See also [21].) A sketch proof is included here for the sake of completeness.

Lemma 3.1 Let S and D be as defined in (21) and (22), and, for any $s \leq t$ let $C_p(s, t)$ be the mutual information between the *paths* $(X_r, r \in [s, t])$ and $(Y_r, r \in [-T, t])$; then

$$C_p(s, t) = S(t) - D(s). \tag{23}$$

Proof. Let \mathbb{P}^R and \mathbb{P}^M be the probability measures on \mathcal{F} , defined by the following Girsanov transformations. (See, for example, Chapter 6 in [13])

$$\begin{aligned} \frac{d\mathbb{P}^R}{d\mathbb{P}} &= \frac{n(0, I)(Y^i)}{n(\xi, M)(Y^i)} \exp \left(- \int_{-T}^T X'_t dY_t + \frac{1}{2} \int_{-T}^T X'_t \Sigma_W X_t dt \right) \\ \frac{d\mathbb{P}^M}{d\mathbb{P}^R} &= \frac{n(0, P_i + M)(Y^i)}{n(0, I)(Y^i)} \exp \left(\int_{-T}^T \hat{X}'_t dY_t - \frac{1}{2} \int_{-T}^T \hat{X}'_t \Sigma_W \hat{X}_t dt \right). \end{aligned} \tag{24}$$

It easily follows from elementary manipulations of n -variate Gaussian distributions and the Cameron-Martin-Girsanov theory that neither transformation in (24) alters the distribution of X . However, the first transformation renders Y a Brownian motion, independent of X , having the same

quadratic covariation as W , and non-zero initial value with distribution $N(0, I)$. (\mathbb{P}^R is the *reference* probability of linear and nonlinear filtering.) The second transformation restores the original marginal distribution to Y , while retaining the independence of X and Y . (This follows from the innovations representation of Y in (6)). Thus, under \mathbb{P}^M , X and Y are independent but have the same marginal distributions they had under \mathbb{P} .

It now follows that

$$\begin{aligned} C_p(s, t) &= C(s) + \mathbf{E} \log \left(\mathbf{E}^M \left(\frac{d\mathbb{P}}{d\mathbb{P}^M} \middle| \mathcal{F}_t \right) \left(\mathbf{E}^M \left(\frac{d\mathbb{P}}{d\mathbb{P}^M} \middle| \mathcal{F}_s \right) \right)^{-1} \right) \\ &= C(s) + \frac{1}{2} \int_s^t \text{tr}(\Sigma_W Q(r)) dr \\ &= S(t) - D(s), \end{aligned}$$

as claimed. •

$S(-T)$ is the information gain on the whole process X arising from the initial observation Y^i , and $S(t) - S(-T)$ is the information gain arising through the running observation Y^r between times $-T$ and t . Like $C(t)$, $C_p(s, t)$ is an *absolute* information quantity, not depending on any underlying reference measure. We can think of $C_p(s, t)$ as being the information stored by a *path* estimator that has access to $(Y_r, -T \leq r \leq t)$ but has no interest in the values of X prior to time s . If s increases but t remains constant, the path estimator dissipates this stored information at rate $\dot{D}(s)$; the dissipation process represents observation-derived information that was useful for estimating the past of X , but is of no use in estimating its future.

We now take the view that entropy is simply *unobservable* information. Thus, if the signal process X of (1) were completely unobservable, its entropy at time t would be $S_X(t)$, as defined in (10). However, this is reduced in the presence of the partial observations $(Y_s, s \in [-T, t])$ to

$$\begin{aligned} S_{X|Y}(t) &= \mathcal{S}(N(\hat{X}_t, Q(t))) \\ &= S_X(t) - C(t). \end{aligned} \tag{25}$$

We cannot allow *perfect* observations of X_t since these would convert an infinite amount of entropy into stored information and create mathematical difficulties. These difficulties are avoided in (4) by the non-degeneracy conditions on the observation noise terms ζ and W .

With the addition of observations, we could modify our two-component universe, to include a third physical component represented by the observation process Y . This would involve a modified heat bath that accounted for the observation noise, W , as well as that in the signal (V). We defer this approach until Section 4. For the moment we note that the signal

energy can be split (at least conceptually) into two components, as follows:

$$H_X(X_t) = \frac{1}{2}(X_t - \hat{X}_t)' P_{SS}^{-1} (X_t + \hat{X}_t) + \frac{1}{2} \hat{X}_t' P_{SS}^{-1} \hat{X}_t.$$

Since the second of these is completely determined by the observations up to time t , it is available to a *demon* having access to Y . Thus the average energy of the signal, $\mathcal{E}(N(0, P(t)))$, can be split into two parts: that available to the demon, $\mathcal{E}(N(0, P(t) - Q(t)))$, which we shall call *work*, and that remaining, $\mathcal{E}(N(0, Q(t)))$, which we shall call *heat*. In this sense the signal is *cooled* by the observations. If the observations were to be turned off at time t (which could be achieved by setting Σ_W to zero) then the heat component of the signal energy would converge towards the steady-state value of $S_X(t)$ in much the same way as $S_X(t)$ itself. The cooled signal has entropy $S_{X|Y}(t)$, and this is less than that of the uncooled signal by the quantity of information stored by the filter, as shown by (25).

The signal now interacts with the heat bath in exactly the way it did in the absence of observations. However, the interaction now sub-divides into two sub-interactions: one between the cooled signal and the heat bath, and one between the demon and the heat bath. During fluctuations, both sub-components of the signal can *lose* energy to the heat bath, but only the cooled signal can *gain* energy from it. This is because energy coming from the heat bath has entropy associated with it. (The signal gains energy from the heat bath through the small fluctuations of the Brownian motion, V , and these are completely unpredictable from \mathcal{F}_t^Y .) Of course, there is also an interaction between the sub-components of the signal: as t increases, the demon continues to ‘extract’ work from the cooled signal as new observations become available. The combination of these effects causes three energy *flows*, as follows.

Flow 1: Heat Bath to Cooled Signal. The average rate of flow of energy can be found from the rate of change of energy of the cooled signal with the work extraction process ‘turned off’. This can be achieved by temporarily setting Σ_W to zero.

$$\begin{aligned} \dot{E}_1(t) &= \frac{d}{dt} \mathcal{E}(N(0, Q(t)))|_{\Sigma_W=0} \\ &= \text{tr} (A(Q(t) - P_{SS}) P_{SS}^{-1}). \end{aligned}$$

Flow 2. Cooled Signal to Demon. The demon continues to receive new information, which allows it to extract work from the cooled signal at an average rate of

$$\begin{aligned} \dot{E}_2(t) &= \dot{E}_1(t) - \frac{d}{dt} \mathcal{E}(N(0, Q(t))) \\ &= \frac{1}{2} \text{tr} (Q(t) \Sigma_W Q(t) P_{SS}^{-1}). \end{aligned}$$

Flow 3. Demon to Heat Bath. As described above, the demon loses energy to the heat bath during fluctuations, but gets none back. This results in a net flow of energy from the demon to the heat bath with average rate

$$\begin{aligned}\dot{E}_3(t) &= \dot{E}_2(t) - \frac{d}{dt} \mathcal{E}(N(0, P(t) - Q(t))) \\ &= \text{tr} \left(A(Q(t) - P(t)) P_{SS}^{-1} \right).\end{aligned}$$

The net average rate of outflow of energy from the heat bath is thus

$$\dot{E}_1(t) - \dot{E}_3(t) = \dot{E}_X(t),$$

which is unaltered by the existence of observations. The three energy flows are shown in Figure 1.

The rates of change of entropy and information are as follows.

In the Cooled Signal. The entropy is raised by the inflow of energy (Flow 1) and lowered by the outflow (Flow 2). The net rate of change is

$$\dot{S}_{X|Y}(t) = \dot{S}_X(t) + \dot{D}(t) - \dot{S}(t).$$

In the Demon. The demon has associated with it an amount of *information* $C(t)$, but no entropy. This corresponds with the notion that the energy available to it is work. $C(t)$ is increased at rate $\dot{S}(t)$ by the supply of new information, and reduced at rate $\dot{D}(t)$ by the dissipation of historical information.

In the Heat Bath. Since the net average rate of change of energy in the heat bath is unaltered by the existence of observations, so is its rate of change of entropy.

The term $\dot{D}(t)$ in the equation for $\dot{S}_{X|Y}(t)$ is the excess rate of entropy increase of the cooled signal (as compared with the uncooled signal) and is caused by the increased rate of inflow of energy from the heat bath: $\dot{E}_1(t) - \dot{E}_X(t)$. Thus the filter can be seen to be *entropically efficient* in the sense that it dissipates information at exactly the rate of this (unavoidable) entropy increase. If the filter dissipated at a higher rate, it would cause an additional increase in the entropy of the whole system, illustrating Landauer's Principle; if it dissipated at a lower rate, it would retain more information than strictly needed for estimating the future of X . In order not to cause unnecessary entropy increase, the filter must retain all information that is not held as entropy in other parts of the 'universe'.

The rate of change of entropy of the universe with observations differs from that of the original universe by $\dot{D}(t) - \dot{S}(t)$. If, during convergence towards the invariant distribution, $\dot{S}(t) > \dot{D}(t)$ for some t , then it is possible for the entropy of the universe to decrease at time t . For example, this is

the case at $t = -T$ if the signal is initialised in its invariant distribution, $N(0, P_{SS})$, and the filter is initialised with near total ‘ignorance’, $M \gg P_{SS}$.

In the stationary state the overall rate of change of entropy is the same as that in the original universe (zero), but energy circulates around a loop comprising the heat bath, the cooled signal and the demon. The demon is analogous to a perfect *heat pump* that cools the signal, returning the extracted energy to the heat bath, and doing this with no increase in entropy. It maintains the cooled signal at a temperature lower than that of the heat bath, and this causes an inflow of heat (Flow 1), with a resultant entropy increase. However, the entropy increase is countered by the steady supply of new information, which, arising as it does *within* the universe, constitutes a matching entropy decrease. This illustrates Landauer’s Principle in reverse. The entropic efficiency of the Kalman-Bucy filter is a special case of the information conserving properties of Bayesian estimators investigated in [16]. These issues are developed further in [17].

Because of the circulation of energy, we might expect the stationary state of the universe with observations to be a *non-equilibrium* state, even if the signal process X is self adjoint. To make these ideas precise we introduce, in the next section, a second statistical mechanical analogy in which the filter is a separate physical component capable of holding energy in its own right.

4 Interactive Statistical Mechanics

The physical analogy, developed in Section 2 for the signal alone, can also be applied to the *joint*, $2n$ -dimensional process (X, \hat{X}) . In order to define entropy production for this we require the observation noise covariance matrix, Σ_W , to be *strictly* positive definite. The joint process then has the invariant distribution $N(0, P_J)$, with covariance matrix

$$P_J = \begin{bmatrix} P_{SS} & P_{SS} - Q_{SS} \\ P_{SS} - Q_{SS} & P_{SS} - Q_{SS} \end{bmatrix}, \quad (26)$$

where, Q_{SS} is the stationary covariance matrix of the filter; this satisfies the algebraic Riccati equation:

$$AQ_{SS} + Q_{SS}A' + \Sigma_V - Q_{SS}\Sigma_W Q_{SS} = 0;$$

the Hamiltonian for (X, \hat{X}) is

$$H_J(x, \hat{x}) = \frac{1}{2} \begin{bmatrix} x' & \hat{x}' \end{bmatrix} P_J^{-1} \begin{bmatrix} x \\ \hat{x} \end{bmatrix}. \quad (27)$$

The joint process can be considered as describing a statistical mechanical system interacting with a unit temperature heat bath in the same way as was X in Section 2. This interaction forces the system towards its stationary state thus maximising the entropy of the universe comprising the joint process and the heat bath. The stationary state in this analogy is a *non-equilibrium* state, regardless of whether or not the signal, X , is self-adjoint. This is because of the interaction between the two components, X and \hat{X} . We first investigate this interaction when the system is in its stationary state.

The joint process is a $2n$ -vector, Gaussian process with drift coefficient $f(\theta) = A_J\theta$ and diffusion matrix Σ_J , where

$$A_J = \begin{bmatrix} A & 0 \\ Q_{SS}\Sigma_W & A - Q_{SS}\Sigma_W \end{bmatrix} \quad \text{and} \quad \Sigma_J = \begin{bmatrix} \Sigma_V & 0 \\ 0 & Q_{SS}\Sigma_W Q_{SS} \end{bmatrix}.$$

At each time t , (X_t, \hat{X}_t) has mean zero, and the covariance matrix P_J of (26). It can be expressed in time-reversed form as a $2n$ -vector Gaussian process with drift coefficient $\bar{A}_J\theta$, and diffusion matrix Σ_J , where

$$\bar{A}_J = -A_J - \Sigma_J P_J^{-1}.$$

(This follows from Proposition 2.1.) The rate of entropy production for the joint process can be found in the same way as was R_X in Section 2. In fact

$$R_J = \frac{1}{2} \text{tr} \left((A_J - \bar{A}_J)' \Sigma_J^{-1} (A_J - \bar{A}_J) P_J \right). \quad (28)$$

Since this is a rate of entropy production in a stationary state, it is also the joint rate of entropy *flow* in this state.

The key to isolating the *interactive* component of this flow is the fact that both X and \hat{X} are *autonomously* Markov. This is clearly true of X , but also true of \hat{X} since the latter can be expressed autonomously via the innovations process of (6):

$$d\hat{X}_t = A\hat{X}_t dt + Q_{SS} d\nu_t \quad \text{for } t \in [-T, T]. \quad (29)$$

The rate of entropy flow in X alone, R_X , is given by (17) in the stationary state, and that in \hat{X} alone, $R_{\hat{X}}$, is given by the following:

$$R_{\hat{X}} = \frac{1}{2} \text{tr} \left((A - \bar{A}_{\hat{X}})' (Q_{SS}\Sigma_W Q_{SS})^{-1} (A - \bar{A}_{\hat{X}}) (P_{SS} - Q_{SS}) \right), \quad (30)$$

where

$$\bar{A}_{\hat{X}} = -A - Q_{SS}\Sigma_W Q_{SS} (P_{SS} - Q_{SS})^{-1}.$$

We can now define a rate of *interactive entropy flow*:

$$\begin{aligned}
 R_I &:= R_J - R_X - R_{\hat{X}} \\
 &= \frac{1}{2} \text{tr}(\Sigma_W Q_{SS}) + \frac{1}{2} \text{tr}(\Sigma_V (Q_{SS}^{-1} - P_{SS}^{-1})) \\
 &= \dot{S}_{SS} + \dot{D}_{SS},
 \end{aligned} \tag{31}$$

where \dot{S}_{SS} and \dot{D}_{SS} are the steady-state values of the information supply and dissipation processes of Section 3. (Of course, these are equal.) The rate of interactive entropy flow is thus the total flow rate of information to and from the information store.

Since X and \hat{X} are Markov processes in their own right, they separately describe statistical mechanical systems interacting with unit temperature heat baths. The *marginal* interactions are governed by the Hamiltonians H_X of (8) and $H_{\hat{X}}$, defined by:

$$H_{\hat{X}}(\hat{x}) = \frac{1}{2} \hat{x}' (P_{SS} - Q_{SS})^{-1} \hat{x}. \tag{32}$$

We can also identify the *conditional* Hamiltonians

$$\begin{aligned}
 H_{X|\hat{X}}(x, \hat{x}) &= H_J(x, \hat{x}) - H_{\hat{X}}(\hat{x}) \\
 H_{\hat{X}|X}(\hat{x}, x) &= H_J(x, \hat{x}) - H_X(x).
 \end{aligned}$$

The Hamiltonian of the joint system can be expressed as the sum of three components:

$$H_J(x, \hat{x}) = H_{X|\hat{X}}(x, \hat{x}) + e_C(x, \hat{x}) + H_{\hat{X}|X}(\hat{x}, x), \tag{33}$$

where e_C is a component of energy common to the signal and the filter (defined by (33)). The sum of the first two components in (33) is the Hamiltonian of the signal, H_X , and the sum of the last two components is that of the filter, $H_{\hat{X}}$.

$H_{X|\hat{X}}$ can be expressed in the following form:

$$H_{X|\hat{X}}(x, \hat{x}) = \frac{1}{2} (x - \hat{x})' Q_{SS}^{-1} (x - \hat{x}),$$

and is, therefore, determined by the ‘conditional signal’, $\tilde{X} := X - \hat{X}$. This is also a Markov process in its own right, and evolves according to the Itô equation

$$d\tilde{X}_t = (A - Q(t)\Sigma_W)\tilde{X}_t dt + dV_t - Q(t) dW_t. \tag{34}$$

It describes a statistical mechanical system with Hamiltonian

$$H_{\tilde{X}}(\tilde{x}) = \frac{1}{2} \tilde{x}' Q_{SS}^{-1} \tilde{x},$$

that also interacts with a unit temperature heat bath.

The joint statistical mechanical system, described by (X, \hat{X}) , can thus be thought of as comprising two ‘physically distinct’ subsystems: the conditional signal, with state variable \tilde{X}_t and Hamiltonian $H_{\tilde{X}}$, and the filter, with state variable \hat{X}_t and Hamiltonian $H_{\hat{X}}$. By ‘physically distinct’, we mean that the subsystems satisfy three conditions: (i) their state variables are autonomously Markov; (ii) energy is additive—the Hamiltonian of the joint system is the sum of the Hamiltonians of the two subsystems; and (iii) entropy is additive—since \tilde{X}_t and \hat{X}_t are independent, the entropy of the joint system is the sum of the entropies of the subsystems.

The conditional signal has average energy

$$E_{\tilde{X}}(t) = \mathbf{E}H_{\tilde{X}}(\tilde{X}_t) = \frac{1}{2}\text{tr}(Q(t)Q_{SS}^{-1});$$

and this evolves as follows:

$$\dot{E}_{\tilde{X}}(t) = \frac{1}{2}\text{tr}(\Sigma_V Q_{SS}^{-1}) + \text{tr}(AQ(t)Q_{SS}^{-1}) - \frac{1}{2}\text{tr}(Q(t)\Sigma_W Q(t)Q_{SS}^{-1}). \quad (35)$$

It forms one component of the average signal energy, $E_X(t)$ (as defined in (10)), the other component of which is the average common energy, $\mathbf{E}e_C(X_t, \hat{X}_t)$. The evolution of $E_X(t)$ is not affected by the presence of the filter; in particular, it would not be changed if Σ_W were set to zero, and so we may conclude that the third term on the right-hand side of (35) represents an energy flow from the conditional signal to the common energy, and hence to the filter. This is a delicate point. In trying to identify circular flows of energy between three or more subsystems, we must break one of the connections between subsystems and observe the increase in energy of the ‘upstream’ component, or the decrease in energy of the ‘downstream’ component. However, in changing a parameter of the system, we must be careful that we do not alter dynamical aspects of the system other than that intended. Clearly, setting Σ_W to zero not only disconnects the filter from the signal but also alters its interaction with the heat bath. Thus, observing the filter energy with Σ_W set to zero will not reveal the energy inflow from the signal. However, observing the energy of the conditional signal in the same circumstances does. The fact that the marginal statistical mechanics of the signal are not affected by the value of Σ_W is crucial here.

The filter itself has average energy $E_{\hat{X}}(t)$, given by

$$E_{\hat{X}}(t) = \mathbf{E}H_{\hat{X}}(\hat{X}_t) = \frac{1}{2}\text{tr}((P(t) - Q(t))(P_{SS} - Q_{SS})^{-1});$$

and this evolves as follows:

$$\begin{aligned} \dot{E}_{\hat{X}}(t) &= \frac{1}{2}\text{tr}(Q(t)\Sigma_W Q(t)(P_{SS} - Q_{SS})^{-1}) \\ &\quad + \text{tr}(A(P(t) - Q(t))(P_{SS} - Q_{SS})^{-1}). \end{aligned}$$

We can thus identify the following three energy flow rates:

Flow 4. Heat Bath to Conditional Signal.

$$\dot{E}_4(t) = \frac{1}{2} \text{tr}(\Sigma_W Q_{SS}^{-1}) + \text{tr}(A Q(t) Q_{SS}^{-1});$$

Flow 5. Conditional Signal to Filter.

$$\dot{E}_5(t) = \frac{1}{2} \text{tr}(Q(t) \Sigma_W Q(t) Q_{SS}^{-1});$$

Flow 6. Filter to Heat Bath.

$$\begin{aligned} \dot{E}_6(t) &= \frac{1}{2} \text{tr}(Q(t) \Sigma_W Q(t) (Q_{SS}^{-1} - (P_{SS} - Q_{SS})^{-1})) \\ &\quad - \text{tr}(A(P(t) - Q(t))(P_{SS} - Q_{SS})^{-1}). \end{aligned}$$

In the stationary state, all three energy flows have the common rate

$$\dot{E}_{SS} = \frac{1}{2} \text{tr}(\Sigma_W Q_{SS}) = \dot{S}_{SS}.$$

Since all three components of the universe have unit temperature, the energy flows are accompanied by equal entropy flows. In particular, the energy flow from the conditional signal to the filter is associated with an entropy flow of the same rate as that of the information supply, identified in Section 3. The flow of energy in the stationary state is not driven by temperature gradients and does not cause any increase in overall entropy, and so is ‘physically reversible’.

The physical analogy here is distinct from that of Section 3 in that the energy flows of the latter are driven by the supply of observation information, whereas the flows here are driven by the nature of the interaction between X and \hat{X} , and do not depend on any distinction being made between entropy and observable information.

The joint Hamiltonian can also be expressed as the sum of that of the signal, H_X , and that of the ‘conditional filter’, $H_{\hat{X}|X}$. The latter can be expressed in the form

$$H_{\hat{X}|X}(\hat{x}, x) = \frac{1}{2} \check{x}' (Q_{SS}^{-1} - P_{SS}^{-1})^{-1} \check{x},$$

where

$$\check{x} = P_{SS}^{-1} x - Q_{SS}^{-1} (x - \hat{x}).$$

This is determined by the process $\check{X} := P_{SS}^{-1} X - Q_{SS}^{-1} \hat{X}$.

In the stationary state, \check{X} is a Markov process in its own right, and its value at time t is independent of that of X , and so we can also decompose

the joint system into subsystems associated with X and \check{X} . (Note, however, that this is not, in general, true away from the stationary state.) Thus, in the stationary state, we can identify a flow of energy from the signal to the conditional filter in the same way that the flow of energy from the conditional signal to the filter was identified above. (This involves setting Σ_V to zero.) This energy flow has the same rate as the others, \dot{E}_{SS} , and leads to the symmetrical system shown in Figure 2. The joint system has two ‘internal’ energy flow points (ie. points of flow not involving the heat bath), each of which has an associated entropy flow of rate \dot{S}_{SS} ; the sum of these is equal to the rate of interactive entropy flow R_I , as defined in (31).

In the presence of observations, we can make the distinction between entropy and information that was made in Section 3. The entropy of the joint system in the presence of observations then becomes that of the conditional signal,

$$S_{\check{X},SS} = \frac{n}{2}(1 + \log(2\pi)) + \frac{1}{2} \log |Q_{SS}|,$$

and its information content becomes

$$S_{\hat{X},SS} = \frac{n}{2}(1 + \log(2\pi)) + \frac{1}{2} \log |P_{SS} - Q_{SS}|.$$

This is made up of two components: the information stored on the signal

$$C_{SS} = \frac{1}{2} \log |P_{SS} Q_{SS}^{-1}|,$$

and the ‘residual’ information

$$S_{\check{X},SS} = \frac{n}{2}(1 + \log(2\pi)) + \frac{1}{2} \log |Q_{SS} (Q_{SS}^{-1} - P_{SS}^{-1}) Q_{SS}|.$$

As in Section 3, we call the energy of the conditional signal *heat*, and that of the filter *work*. Heat is converted into work at rate \dot{E}_{SS} by the arrival of new observation information. It is converted back into heat when it is returned to the heat bath by the filter. The filter uses its information dissipation process, which has exactly the correct rate, to provide the necessary entropy. Once again, this provides a quantitative example of Landauer’s Principle.

Many of the properties of entropy production, discussed in the remarks following its definition in Section 2, are inherited, in the stationary state, by interactive entropy flow. In particular, the interactive entropy flow of the time-reversed joint process (X, \check{X}) is the same as that of the forward-time joint process. It turns out that the components of the time-reversed joint process can be thought of as being the signal and filter processes of a *dual*

problem, in which information supply and dissipation exchange roles. The second physical analogy for this dual problem is then the physical reversal of that of the original. These duality ideas are developed further for linear filters in [17], and for nonlinear filters in [18].

5 Conclusions

This article has explored the information flows associated with continuous-time Kalman-Bucy filters, and connected them with the entropy flows occurring in non-equilibrium statistical mechanical systems. It has shown via a physical analogy that a law of non-decrease of entropy need not apply to such systems in the presence of observations that continue to supply new information.

Like other Bayesian filters, the Kalman-Bucy filter is information conserving in the manner described in [16], and, because of this, it is also entropically efficient in the physical analogy: it achieves the maximum possible reduction in entropy from a given supply of observations, and stores no more information than is strictly necessary to do this. The entropic efficiency manifests itself in a second analogy as physically reversible dynamics for the system described by the joint signal-filter process.

The fact that observations can reduce entropy is, essentially, an ‘inverse Landauer Principle’. The Kalman-Bucy filter provides quantitative examples of both ‘regular’ and ‘inverse’ Landauer’s Principles. These ideas can also be applied to linear, time-inhomogeneous, degenerate systems and to nonlinear systems. (See [17] and [18].) The physical analogies they provide are particularly useful in the nonlinear case since they form the basis of an information theoretic Lyapunov theory for nonlinear filters. We believe that the results should be of independent interest to the statistical physics research community.

References

- [1] C. H. BENNETT, *The thermodynamics of computation—a review*, Int. J. Theoretical Physics, 21 (1982), pp. 905–940.
- [2] L. BERTINI, A. DE SOLE, D. GABRIELLI, G. JONA-LASINIO AND C. LANDIM, *Macroscopic fluctuation theory for stationary non-equilibrium states*, Journal of Statistical Physics, 107 (2002), pp. 635–675.
- [3] B. Z. BOBROVSKY AND M. ZAKAI, *A lower bound on the estimation error for certain diffusion processes*, IEEE Trans. Information Theory, 22 (1976), pp. 45–52.

- [4] R. W. BROCKETT AND J. C. WILLEMS, *Stochastic control and the second law of thermodynamics*, Proceedings of the 17th IEEE Conference on Decision and Control, San Diego CA, IEEE (1979), pp. 1007–1011.
- [5] M. H. A. DAVIS, *Linear Estimation and Stochastic Control*, Chapman and Hall, 1977.
- [6] G. GALLAVOTTI, *Statistical Mechanics—A Short Treatise*, Springer-Verlag, 1999.
- [7] B. GAVEAU AND L. S. SCHULMAN, *Creation, dissipation and recycling of resources in non-equilibrium systems*, Journal of Statistical Physics, 110 (2003), pp. 1317–1367.
- [8] H-O. GEORGII, *Gibbs Measures and Phase Transitions*, de Gruyter, 1988.
- [9] A. H. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.
- [10] I. KARATZAS AND S. SHREVE, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, 1991.
- [11] R. LANDAUER, *Dissipation and heat generation in the computing process*, IBM J. Research and Development, 5 (1961), pp. 183–191.
- [12] J. L. LEBOWITZ AND H. SPOHN, *A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics*, J. Statistical Physics, 95 (1999), pp. 333–366.
- [13] R. S. LIPTSER AND A. N. SHIRYAYEV, *Statistics of Random Processes 1—General Theory*, Springer-Verlag, 1977.
- [14] R. S. LIPTSER AND A. N. SHIRYAYEV, *Statistics of Random Processes 2—Applications*, Springer-Verlag, 1978.
- [15] J. C. MAXWELL, *Theory of Heat*, Longmans, London, 1871.
- [16] S. K. MITTER AND N. J. NEWTON, *A Variational approach to Nonlinear Estimation*, SIAM J. Control Optim., 42 (2003), pp. 1813–1833.
- [17] N. J. NEWTON, *The interactive statistical mechanics of linear filters*, in preparation.
- [18] N. J. NEWTON, *Dual nonlinear filters and interactive entropy production*, in preparation.

- [19] O. PENROSE, *Foundations of Statistical Mechanics*, Pergamon, Oxford, 1970.
- [20] M. B. PROPP, *The Thermodynamic Properties of Markov Processes*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1985.
- [21] E. MAYER-WOLF AND M. ZAKAI, *On a formula relating the Shannon information to the Fisher information for the filtering problem*, in *Filtering and Control of Random Processes*, H. Korezlioglu, G. Mazziotto, S. Szpirglas (eds.), *Lecture Notes in Control and Information Sciences* 61, Springer, 1984, pp. 164–171.
- [22] J. C. WILLEMS, *Dissipative dynamical systems, part i: general theory*, *Arch. Rational. Mech. Anal.*, 45 (1972), pp. 321–351.