J Stat Phys (2008) 133: 711–737
DOI 10.1007/s10955-008-9622-z

Interactive Statistical Mechanics and Nonlinear Filtering

Nigel J. Newton

Received: 9 September 2005 / Accepted: 29 August 2008 / Published online: 24 September 2008

Abstract This paper connects non-equilibrium statistical mechanics and optimal nonlinear filtering. The latter concerns the observation-conditional behaviour of Markov *signal* processes, and thus provides a tool for investigating statistical mechanics with partial observations. These allow entropy reduction, illustrating Landauer's Principle in a quantitative way.

The joint process comprising a signal and its nonlinear filter is irreversible in its invariant distribution, which therefore corresponds to a *non-equilibrium* stationary state of the associated *joint* system. Macroscopic entropy and energy flows are identified for this state. Since these are driven by observations *internal* to the system, they do not cause entropy increase, and so the joint system makes statistical mechanical sense in reverse time.

Time reversal yields a *dual* system in which the signal and filter exchange roles. Despite the structural similarity of the original and dual systems, there is a substantial asymmetry in their complexities. This reveals the direction of time, despite the systems being in stationary states that do not produce entropy.

Keywords Dual systems · Information theory · Landauer's principle · Non-equilibrium statistical mechanics · Nonlinear filtering · Partial observations

This work was partially supported by Leverhulme Trust Research Fellowship 2003/0426, by MURI Grant F49620-02-1-0325 (Complex Adaptive Networks for Co-operative Control), by ARO-MURI Grant DAAD19-00-1-0466 (Data Fusion in Large Arrays of Microsensors (sensor web)) and by NSF-ITR Grant CCR-0325774 Collaborative Research: New Approaches to Experimental Design and Statistical Analysis of Genomic and Structural Biological Data from Multiple Sources.

N.J. Newton (✉)
Department of Computing and Electronic Systems, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK
e-mail: njn@essex.ac.uk

N.J. Newton
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

1 Introduction

This paper makes connections between non-equilibrium statistical mechanics and optimal nonlinear filtering. The unifying theme is the theory of Markov processes. These play a central role in the *stochastic dynamics* approach to statistical mechanics, where they are used to model the coarse-grained dynamics of Hamiltonian systems. In this context, the *irreversibility* of a time-homogeneous Markov process in an invariant distribution is associated with a non-equilibrium stationary state, and its large deviations are associated with fluctuations in entropy [26, 28].

Markov processes also play a central role in the theory of optimal nonlinear filtering. This is a branch of signal processing in which a Markov *signal* process is estimated on the basis of partial observations corrupted by noise. The nonlinear filter is a causal device that computes an estimate of the signal at each time t, based on prior statistical knowledge and the history of the observation up to t. This estimate is chosen to minimise an appropriate cost function, such as the mean-squared error. In all but a few special cases, it is impossible to compute such estimates without first computing the observation-conditional distribution of the signal [5], or at least approximating it. This leads to a substantial asymmetry in the dynamical complexities of a signal process and its nonlinear filter.

By considering the nonlinear filter for a process that models the coarse-grained dynamics of a Hamiltonian system, we are able to investigate the statistical mechanics of the latter in the presence of *partial observations*. These allow the reduction of entropy in the manner of Maxwell's demon [30]. Thought of in this way, the nonlinear filter is a demon that reduces entropy by storing partial information on the Hamiltonian system. We develop full information flow models for the nonlinear filter. These provide precise quantitative examples of Landauer's Principle [25] in both 'directions': the supply of information to the filter allows it to reduce entropy; the disposal of information by the filter causes entropy to increase.

The nonlinear filter process (i.e. the stochastic process of conditional distributions) is itself Markov, as is the *joint* process comprising the signal and filter components, and so both can be associated with the coarse-grained dynamics of *abstract* Hamiltonian systems. The system corresponding to the joint process can be decomposed into that corresponding to the filter process and a *conditional* system. The latter is associated with the degrees of freedom of the original system that are not revealed by the observation. The conditional system is shown to exhibit a variant of the Second Law of Thermodynamics in which the (observation-conditional) entropy is a *submartingale*. (This is a stochastic process whose average future value, conditioned on information available in the present, is no smaller than its current value.)

Regardless of whether or not the signal process is reversible, the joint process is not; the flow of information between the signal and filter represents an irreversible component of their interaction. In its invariant distribution the joint process is associated with a stationary non-equilibrium state that exhibits macroscopic flows of entropy and energy. These are driven by the observation mechanism (which is *internal* to the system) rather than by external fields or boundary conditions, and so they are not accompanied by entropy increase. The (abstract) joint system lies on the boundary between systems that do and do not obey the Second Law of Thermodynamics. Because of this, and other symmetry properties derived here, the joint system makes statistical mechanical sense in reverse time. In fact, time reversal yields a *dual* system with the same properties as the original. The original signal and filter processes exchange roles in reverse time, and information and energy flow in the opposite direction; these flows are now caused by the supply of *dual* observation information.

The dual system has one striking feature that distinguishes it from the original—in contrast with the general rule, the dynamics of the dual filter are *simpler* than those of the dual signal. A nonlinear filter has to store all the information it has extracted from the past of the observation that may have relevance to the future of the signal. Because of the effects of nonlinear dynamics and random forcing this typically requires very complex dynamics. That this is not true of the dual system is an indicator of the direction of time that is not dependent on the observation of convergence towards a stationary state, nor on the observation of entropy production in a stationary state.

The paper builds on earlier work appearing in [34], which develops interactive statistical mechanics for the linear Gaussian case. The 'nonlinear' filter is then *linear*, and the conditional distribution it calculates is parametrised by a finite number of statistics (in fact, the conditional mean). Although time reversal is not considered in that paper, the linear Gaussian case it treats is one example which does not share the time asymmetry property mentioned above; the dual to the Kalman-Bucy filter is another Kalman-Bucy filter of identical structure and dimension [36].

The paper considers *continuous time* systems only, although there is no fundamental difference between discrete and continuous time systems in the results obtained. The Markov signal process is assumed to take values in a complete separable metric space, thereby including finite or countably infinite state Markov jump processes, finite- or infinite-dimensional diffusions, and the function spaces in which the nonlinear filters for these processes evolve. In order to aid clarity, two examples are developed in tandem with the general theory—one in which the signal is a finite-state Markov jump process, and another in which it is a multidimensional diffusion process. The partial observations are of finite dimension, and of the 'signal-plus-white-noise' type (although brief mention is made of Poisson counting process observations in the context of the dual to the system with finite-state signal). There is no fundamental problem extending the ideas to observation processes of infinite dimension; however, the finite-dimensional case retains its salient features, and is more accessible to rigorous proof from the nonlinear filtering literature.

Many of the formal results of the paper are based on non-trivial theorems of measure-theoretic probability and stochastic calculus. However, the paper has been written with the non-expert in these fields in mind. Technical conditions have been omitted wherever they shed no light on the underlying ideas; instead, reference is made to appropriate sources.

The material is developed as follows. Section 2 reviews the *stochastic dynamics* approach to statistical mechanics and derives the statistical mechanical laws obeyed by Markov processes in a very general setting. Section 3 introduces nonlinear filtering for these processes, and gives dynamical formulae by which filters can be implemented. It also introduces the information quantities of interest, the information *supply*, *storage* and *dissipation*, and relates them to system parameters. Section 4 develops the statistical mechanics of the interacting signal and filter processes. The abstract statistical mechanical system associated with the joint process comprising the signal and filter is at the heart of this section. Section 4.3 discusses implications for statistical mechanics with partial observations. Section 5 derives the properties of the dual system obtained by time reversal, showing that it retains all the statistical properties of the original. Finally, Sect. 6 discusses the complexity and algebraic structure of filters considered as dynamical machines, and highlights the striking asymmetry in complexity between the dual systems.

2 Markov Processes in Statistical Mechanics

In this section we review the notion that Markov processes can be used to model the coarse-grained dynamics of Hamiltonian systems; this will be the starting point for the material on statistical mechanics with partial observations, and its connections with nonlinear filtering, developed in later sections. The use of stochastic dynamics to model statistical mechanical behaviour has a history going back at least to [22] and [39], where the Shannon entropy was introduced in this context. The Markov property is shown there to arise through coarse-graining. The Markov property is also consistent with a process being a 'vanishingly small' component of the (randomly initialised) phase space variable of a Hamiltonian system, [38]. For recent studies of entropy production and fluctuations in the context of the stochastic dynamics of non-equilibrium systems see, for example, [2, 17, 26, 28].

Throughout the paper, all random variables and stochastic processes will be defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This is a space of *outcomes*, Ω, a σ-field of *events* (subsets of Ω with well-defined probabilities of occurrence), $\mathcal{F}$, and a *probability measure* $\mathbb{P} : \mathcal{F} \to [0, 1]$. For example, the *event* that a particular scalar random variable ξ takes a value exceeding (say) 1.5, $B := \{\omega \in \Omega : \xi(\omega) > 1.5\}$, is a member of $\mathcal{F}$ and is assigned the probability of occurrence $\mathbb{P}(B)$. We consider the evolution of stochastic processes over the finite time interval $[0, T]$. This can be thought of as a 'time window' on processes evolving over longer (potentially infinite) intervals. Appropriate initialisation allows the study of both dynamic and stationary effects.

The following paragraph introduces the Markov processes of interest in this paper. A fair degree of generality is required in order to include the (probability distribution valued) filter processes of Sects. 3 to 5. Technical assumptions, for example on the regularity of the sample paths of the Markov processes, are needed for mathematical rigour. The reader disinterested in such generalities and rigour should consult the two numbered paragraphs below, where simple examples are developed.

Let $(X_t \in \mathbf{X}, t \in [0, T])$ be a time-homogeneous Markov process that takes values in a complete separable metric space $\mathbf{X}$, with metric d_X. For any $t \in [0, T]$ the past and future of X, $(X_s, s \in [0, t])$ and $(X_s, s \in [t, T])$, are independent when conditioned on the present, X_t. We shall assume that the sample paths of X have left and right limits at all $t \in (0, T)$, and are left or right continuous at all $t \in [0, T]$. (This admits continuous diffusion processes and pure jump processes as special cases.) Let $\mathcal{X}$ be the Borel σ-field of subsets of $\mathbf{X}$ (i.e. the smallest family of subsets of $\mathbf{X}$ that is closed under countable unions and intersections, and contains the open sets of $\mathbf{X}$). ($\mathcal{X}$ is an extremely rich collection of subsets $B \subseteq \mathbf{X}$, for which $\mathbb{P}(X_t \in B)$ is well defined.) The statistical properties of X can be obtained from its (time-homogeneous) transition function, $\Pi(t, x, B)$, where $t \in [0, T]$, $x \in \mathbf{X}$ and $B \in \mathcal{X}$:

$$\mathbb{P}(X_t \in B \mid X_r, r \in [0, s]) = \Pi(t - s, X_s, B) \quad \text{for any } 0 \le s < t \le T.$$

For each t, let P_t be the distribution of X_t. We shall assume that this has a density p_t with respect to a (σ-finite) reference measure λ_X, so that, for any $B \in \mathcal{X}$,

$$P_t(B) = \mathbb{P}(X_t \in B) = \int_B p_t(x)\lambda_X(dx). \tag{2.1}$$

We shall also assume that p_t satisfies the following generalised Fokker-Planck (Kolmogorov forward) equation

$$\frac{\partial p_t}{\partial t}(x) = (\mathcal{A}p_t)(x), \tag{2.2}$$

where $\mathcal{A}$ is a linear operator.

Remark 2.1 In many texts on Markov processes the generator and its adjoint are denoted $\mathcal{L}$ (or $\mathcal{A}$), and $\mathcal{L}^*$ (or $\mathcal{A}^*$), respectively. Here, we shall not make direct use of the generator, and label its adjoint $\mathcal{A}$, reserving the asterisk notation for the dual system in Sect. 5.

We shall use two examples of X, throughout the paper, to illustrate and motivate the general theory. These are as follows.

1. *The finite-state process.* In this, $\mathbf{X} = \{1, 2, \ldots, n\}$, d_X is the discrete metric (in terms of which, any element $x \in \mathbf{X}$ is zero distance from itself and unit distance from all other elements), $\mathcal{X}$ is the set of *all* subsets of $\mathbf{X}$ (including $\mathbf{X}$ itself and the null set $\emptyset$), and λ_X is the counting measure (i.e. $\lambda_X(B)$ is the number of elements in B). X is a time-homogeneous Markov jump process taking values in $\mathbf{X}$, and having $n \times n$ rate matrix A. For each t, X_t has probability density $p_t(x) = P_t(\{x\}) = \mathbb{P}(X_t = x)$ with respect to λ_X, and this evolves according to (2.2) with

$$(\mathcal{A}p)(x) = \sum_{\tilde{x}=1}^{n} A_{x,\tilde{x}} p(\tilde{x}). \tag{2.3}$$

2. *The multidimensional diffusion process.* In this, $\mathbf{X} = \mathbb{R}^n$, d_X is the euclidean metric, $\mathcal{X}$ is the σ-field generated by the open hyper-rectangles, and λ_X is Lebesgue (volume) measure. X is a continuous $\mathbb{R}^n$-valued diffusion process with vector-valued *drift* coefficient $b : \mathbb{R}^n \to \mathbb{R}^n$, and positive-semi-definite-matrix-valued *diffusion* coefficient $a : \mathbb{R}^n \to \mathbb{R}^{n\times n}$. We assume that P_0, a and b are sufficiently regular that X_t has a probability density, p_t, for each t, and that this evolves according to (2.2) with

$$(\mathcal{A}p)(x) = \frac{1}{2}\sum_{i,j} \frac{\partial^2 (a_{i,j} p)}{\partial x_i \partial x_j}(x) - \sum_i \frac{\partial (b_i p)}{\partial x_i}(x). \tag{2.4}$$

   Modulo technical conditions, the multidimensional diffusion process will satisfy an Itô stochastic differential equation of the following type:

$$X_t = X_0 + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dV_s, \tag{2.5}$$

   where σ is a matrix square-root of a ($a = \sigma\sigma'$), and V is an n-vector Brownian motion on Ω. (See, for example, [21].)

We shall further assume that X has a unique invariant distribution P_{SS} with density p_{SS}. For example, this is true of the finite-state process if each state is reachable from all other states in a finite number of steps having non-zero rates. For conditions on other processes that lead to unique invariant distributions, see, for example, [4].

For measures λ and μ, and a function f, on a common metric space $\mathbf{S}$, we define:

$$h(\mu \mid \lambda) := \begin{cases} \int_{\mathbf{S}} q(s)\log(q(s))\lambda(ds) & \text{if } \mu \text{ has density } q \text{ with respect to } \lambda \\ +\infty & \text{otherwise,} \end{cases} \tag{2.6}$$

$$\langle f, \lambda \rangle := \begin{cases} \int_{\mathbf{S}} f(s)\lambda(ds) & \text{if the integral exists} \\ +\infty & \text{otherwise.} \end{cases} \tag{2.7}$$

$h(\mu \mid \lambda)$ is the *relative entropy* (Kullback-Leibler divergence) of μ with respect to λ.

We consider a statistical mechanical system, Σ_X, associated with the process X, regarding P_t as the *state* of Σ_X at time t. The *internal energy* $\mathcal{E}_X(P_t)$, *entropy* $\mathcal{S}_X(P_t)$, and *free energy* $\mathcal{F}_X(P_t)$, of Σ_X in this state are defined as follows:

$$\begin{aligned} \mathcal{E}_X(P_t) &:= \langle H_X, P_t \rangle, \\ \mathcal{S}_X(P_t) &:= -h(P_t \mid \lambda_X), \\ \mathcal{F}_X(P_t) &:= \mathcal{E}_X(P_t) - \mathcal{S}_X(P_t), \end{aligned} \tag{2.8}$$

where,

$$H_X(x) := -\log p_{SS}(x). \tag{2.9}$$

(The energy function H_X takes the value $+\infty$ when $p_{SS}(x) = 0$.) For the finite-state process, $\mathcal{S}_X(P_t)$ is the Shannon entropy; for the diffusion process it is a differential entropy defined in terms of Lebesgue (volume) measure in $\mathbb{R}^n$.

The choice of energy function in (2.9) ensures that the invariant distribution P_{SS} is a Gibbs measure for Σ_X, and gives rise to the properties in the following proposition. This is a standard result; however, a short proof is included here for the sake of completeness, and because it applies to a very general case (any Markov process taking values in a complete separable metric space), and this is needed in Sect. 4.

Proposition 2.1

(i) *The unique minimiser of the free energy of the statistical mechanical system Σ_X is the state P_{SS}.*
(ii) $\mathcal{F}_X(P_{SS}) = 0$.
(iii) *The free energy of Σ_X is non-increasing.*

Proof A simple calculation shows that $\mathcal{F}_X(P_t) = h(P_t \mid P_{SS})$, and so parts (i) and (ii) follow from the non-negativity and strict convexity (where finite) of $h(\cdot \mid P_{SS})$, and the fact that $h(P_{SS} \mid P_{SS}) = 0$. For any $0 \le s \le t \le T$, let $P_{s,t}^{(2)}$ be the two-point joint distribution, defined as follows:

$$P_{s,t}^{(2)}(B, C) := \mathbb{P}(X_s \in B, X_t \in C) = \int_B \Pi(t-s, x, C) P_s(dx),$$

where Π is the transition function for X. Let $P_{s,t,SS}^{(2)}$ be this joint distribution in the special case that $P_s = P_{SS}$. It follows from the chain rule of relative entropy (Theorem C.3.1 in [13]), that

$$\begin{aligned} h(P_s \mid P_{SS}) &= h\left(P_{s,t}^{(2)} \mid P_{s,t,SS}^{(2)}\right) \\ &= h(P_t \mid P_{SS}) + \int h\left(\bar{\Pi}(t, s, x, \cdot) \mid \bar{\Pi}_{SS}(t-s, x, \cdot)\right) P_t(dx) \\ &\ge h(P_t \mid P_{SS}), \end{aligned}$$

where $\bar{\Pi}(t, s, x, \cdot)$ is a regular $(X_t = x)$-conditional distribution for X_s under the joint distribution $P_{s,t}^{(2)}$, and $\bar{\Pi}_{SS}(t-s, x, \cdot)$ is the equivalent under the joint distribution $P_{s,t,SS}^{(2)}$. This proves part (iii). □

Remark 2.2 It is quite possible for $\mathcal{F}_X(P_t)$ to be infinite for all t. This can occur, for example, if the evolution of X is deterministic, and results in an atomic invariant measure. However, we shall not be interested in such 'degenerate' cases.

We can consider Σ_X to be one component of a two-component energy-conserving system, that includes a unit-temperature heat bath with which Σ_X interacts. If the entropy of this system is the sum of the entropies of its two components, then any change in this entropy resulting from the evolution of P_t is the *negative* of the corresponding change in $\mathcal{F}_X(P_t)$. Proposition 2.1 thus states that the entropy of the closed system is *maximised* by the state P_{SS}, and *non-decreasing*. For this reason, we shall refer to Proposition 2.1(iii) as a *Second Law* for Σ_X.

Convention on Randomisation If Σ_X corresponds to a physical system in contact with a physical heat bath, then the randomisation of X has its origins in energy exchange with this heat bath. This is the case, for example, with the second-order electrical circuit in [34]. However, we shall also consider here, *abstract* statistical mechanical systems having their origins in the equations of nonlinear filtering. Nevertheless, we still regard energy exchange with a heat bath as being the sole mechanism by which the corresponding Markov process is randomised. (The heat bath in question may include other system components.) This randomising energy exchange can be any mix of the following two types.

- Invisible: energy fluctuates between Σ_X and the heat bath on more finely grained spatial and temporal scales than those revealed by H_X.
- Visible: the only energy exchange is that revealed by H_X.

The choice of λ_X affects the mix of these two types. For example, the choice of Lebesgue (volume) measure with the second-order electrical circuit of [34] results in an energy function corresponding to the physical energy stored in the circuit, and so energy exchange with the heat bath is fully visible. However, if λ_X were chosen to be the invariant distribution P_{SS}, then the energy function would be constant, and so all energy exchange with the heat bath would be of the invisible variety. An important feature of the minimum free energy state P_{SS} is that its entropy cannot be changed by invisible energy exchange.

3 Observations and Nonlinear Filtering

Nonlinear filtering is a sub-discipline of signal processing in which a *signal* is estimated on the basis of partial *observations*. Examples of its application include *automatic speech recognition*, *image processing* and *object tracking*. In the first of these, for example, the evolving configuration of the speech organs is modelled as a 'hidden' Markov process that has to be estimated on the basis of the acoustic signal picked up by a microphone. Nonlinear filters are *causal* systems; the estimates they provide at any particular time depend only on the past and present values of the observation at that time. In the continuous-time setting, the partial observations are often of the *signal-plus-white-noise* type.

We shall consider nonlinear filters for the Markov 'signal' process X of Sect. 2 given *initial* and *running* observations. This will lead to a study of the statistical mechanics of a partially observed version of Σ_X in Sect. 4. The *initial* observation is a random variable, ψ, that is X_0-conditionally independent of X, and such that the ψ-conditional distribution of X_0 is Z_0; i.e., for any $B \in \mathcal{X}$, $\mathbb{P}(X_0 \in B \mid \psi) = Z_0(B)$. In order to avoid any further need to discuss ψ, we shall assume that Z_0 has been computed, and regard it as being a *surrogate* initial observation. The *running* observation takes the form

$$Y_t^r = \int_0^t g(X_s)\,ds + W_t \quad \text{for } t \in [0,T], \tag{3.1}$$

where $g : \mathbf{X} \to \mathbb{R}^d$ is a continuous function, and W is a d-vector Brownian motion, independent of (X, ψ). Equation (3.1) is a rigorous way of writing '$\dot{Y}_t^r = g(X_t) + \dot{W}_t$', which says that $\dot{Y}_t^r$ is an observation of the signal, $g(X_t)$, plus white-noise, $\dot{W}_t$, type. (The problem with this more natural representation is that Brownian motion is not differentiable, and so $\dot{W}$, and hence $\dot{Y}^r$, is not properly defined in an outcome-by-outcome sense.) We assume that the signal component of the running observation satisfies the following *finite energy* condition:

$$\mathbb{E}\int_0^T |g(X_t)|^2\,dt < \infty. \tag{3.2}$$

The *full* observation is the process $(Y_t := (Z_0, Y_t^r), t \in [0,T])$.

The following definition and technical remarks define and discuss the metric spaces in which the filter and observation variables evolve. They are essential for mathematical rigour, but not to an intuitive understanding of what follows, and can thus be skipped by the reader not interested in such details.

Definition 3.1

(i) $(\mathbf{Z}, d_Z)$ is the metric space of probability measures on $\mathcal{X}$ having densities with respect to the reference measure λ_X, where d_Z is the *total variation* metric:

$$\begin{aligned} d_Z(z,\tilde{z}) &:= \sup_{B\in\mathcal{X}}\{|z(B)-\tilde{z}(B)| + |z(\mathbf{X}\setminus B) - \tilde{z}(\mathbf{X}\setminus B)|\} \\ &= 2\sup_{B\in\mathcal{X}}|z(B)-\tilde{z}(B)|. \end{aligned}$$

(ii) For each $t \in [0,T]$, $(\mathbf{Y}_t, d_{Y,t})$ is the metric space $\mathbf{Z}\times C([0,t];\mathbb{R}^d)$, where $C([0,t];\mathbb{R}^d)$ is the space of continuous functions from $[0,t]$ to $\mathbb{R}^d$, and the metric $d_{Y,t}$ is defined as follows:

$$d_{Y,t}((z,y^r),(\tilde{z},\tilde{y}^r)) = d_Z(z,\tilde{z}) + \sup_{s\in[0,t]}\|y_s^r - \tilde{y}_s^r\|.$$

(The observation available at time t, $(Y_s, s \in [0,t])$, is a random variable taking values in $\mathbf{Y}_t$.)

The nonlinear filter for X computes, at each time t, a regular $(Y_s, s\in[0,t])$-conditional distribution for X_t. This is a random variable Z_t that takes values in $\mathbf{Z}$, and has the following properties:

(F1) $Z_t : \Omega \to \mathbf{Z}$ is $(Y_s, s\in[0,t])$-measurable;
(F2) $Z_t(B) = \mathbb{P}(X_t \in B \mid Y_s, s \in [0,t])$.

The first property here means that $Z_t = F_t(Y_s, s\in[0,t])$ for some (measurable) map $F_t : \mathbf{Y}_t \to \mathbf{Z}$. (It is this map that implementations of the filter must compute, or approximate.)

Remark 3.1 (Technical) The filter map F_t can be shown to be *continuous* under mild technical conditions. (See, for example, [6, 7, 11]. These references prove continuity with respect to a weaker topology on $\mathbf{Z}$ than that induced by d_Z. However, the results of [7], which concern the local Lipschitz continuity of $Z_t(B)$ for individual $B \in \mathcal{X}$, can easily be extended to the stronger topology of $(\mathbf{Z}, d_Z)$ since the Lipschitz constants are the same for all B.)

Remark 3.2 (Technical) Since

$$2 \sup_{B\in\mathcal{X}} |z(B) - \tilde{z}(B)| = \int_{\mathbf{X}} |q(x) - \tilde{q}(x)|\lambda_X(dx),$$

where q and $\tilde{q}$ are the densities of z and $\tilde{z}$, $(\mathbf{Z}, d_Z)$ inherits the properties of the space of integrable functions $L_1(\mathbf{X}, \lambda_X)$. In particular $(\mathbf{Z}, d_Z)$ is complete and separable. (See, for example, [12].) The filter process, $(Z_t, t \in [0, T])$, is thus a stochastic process taking values in the complete separable metric space $(\mathbf{Z}, d_Z)$.

In the most general case, the filter variable Z_t can be calculated by means of an abstract version of the Bayes formula, called the Kallianpur-Striebel formula [20]. However, the greatest interest lies in *recursive* formulae for filtering that exploit the Markov nature of X. These typically represent the filter process, Z, as the (continuous) solution of a stochastic differential equation 'driven' by the running observation process, Y^r. Let $(\zeta_t, t \in [0, T])$ be the process of probability densities corresponding to $(Z_t, t \in [0, T])$ (so that $Z_t(B) = \int_B \zeta_t(x)\lambda_X(dx)$). We shall assume that ζ satisfies the following Itô stochastic differential equation:

$$\zeta_t(x) = \zeta_0(x) + \int_0^t (\mathcal{A}\zeta_s)(x)\,ds + \int_0^t \zeta_s(x)\,(g(x) - \langle g, Z_s\rangle)'\,d\nu_s, \tag{3.3}$$

where $\mathcal{A}$ is the linear operator of (2.2) and $(\nu_t, t \in [0, T])$ is the so-called *innovations process*, defined by

$$\nu_t = Y_t^r - \int_0^t \langle g, Z_s\rangle ds. \tag{3.4}$$

Equation (3.3) is a variant of the Fokker-Planck equation (2.2) that includes a nonlinear term depending on the running observation process, Y^r, (through ν). It reduces to (2.2) if there is no information on X in the latter (for example, if g is constant). For the finite-state signal, $\mathcal{A}$ is as defined in (2.3), and (3.3) is a system of n stochastic *ordinary* differential equations, first derived by Wonham [42]. For the multidimensional diffusion signal, $\mathcal{A}$ is as defined in (2.4), and (3.3) is a stochastic *partial* differential equation called the Kushner-Stratonovich equation [23, 24, 40]. The reader interested in rigorous derivations of recursive filtering formulae for various types of signal is referred to [16] or [27].

The connections made in this paper between nonlinear filtering and statistical mechanics involve an *information flow model* for the former. At time t, we have the partial observation $(Y_s, s \in [0, t])$, which is statistically dependent on X_t. The *mutual information*, $I(X_t; (Y_s, s \in [0, t]))$, provides a measure of this dependence—it tells us how much information on X_t we gain through the partial observation. For random variables Θ and Φ taking values in metric spaces and having joint and marginal distributions $\mathbb{P}_{\Theta,\Phi}$, $\mathbb{P}_\Theta$ and $\mathbb{P}_\Phi$, the mutual information is defined as follows:

$$I(\Theta;\ \Phi) = h\left(\mathbb{P}_{\Theta,\Phi} \mid \mathbb{P}_\Theta \otimes \mathbb{P}_\Phi\right), \tag{3.5}$$

where h is the relative entropy defined in (2.6). $I(\Theta;\Phi)$ is an *absolute* measure of information gain; it does not depend on the choice of reference measures on the spaces in which Θ and Φ take values. (Contrast this with the entropy of the signal, as defined in (2.8).) The minimum possible value of $I(\Theta;\Phi)$ is zero, which indicates that Θ and Φ are statistically independent. If Θ and Φ take values in finite sets, then $I(\Theta;\Phi)$ is bounded above by the Shannon entropies of Θ and Φ.

We define information *supply* S, *storage* C, and *dissipation* D processes as follows: for each $t \in [0,T]$,

$$\begin{aligned} S(t) &:= I((X_s, s\in[0,T]); (Y_s, s\in[0,t])),\\ C(t) &:= I((X_s, s\in[t,T]); (Y_s, s\in[0,t])), \qquad (3.6)\\ D(t) &:= S(t) - C(t). \end{aligned}$$

Remark 3.3 (Technical) Since (by definition) the paths of $(X_r, r\in[s,t])$ have left and right limits at all $r\in(s,t)$, and are left or right continuous at all $r\in[s,t]$, they take values in the Skorohod space $D([s,t];\mathbf{X})$. This has metric

$$d_{D,s,t}(x,\tilde{x}) := \inf_{u\in U}\sup_{r\in[s.t]}\{d_X(x_{u(r)},\tilde{x}_r) + |u(r)-r|\},$$

where U is the set of continuous 1–1 functions from $[s,t]$ to $[s,t]$ with $u(s)=s$ and $u(t)=t$. The Skorohod space inherits from $\mathbf{X}$ the property of being a complete separable metric space. (See, for example, [15].) The metric space, $\mathbf{Y}_t$, in which the observation paths take values was introduced in Definition 3.1. The processes in (3.6) can thus be thought of as random variables taking values in these metric (path) spaces.

Remark 3.4 Slightly different definitions of $S(t)$ and $C(t)$ were given in [34]. There, $(X_s, s\in[0,t])$ was used instead of the whole X process in the definition of $S(t)$, and X_t was used instead of the future of X in the definition of $C(t)$. That these are equivalent to the definitions used here follows easily from Proposition 3.1(ii), below. The definitions used here emphasise the fact that the supply and storage represent information derived from Y that is useful for estimating the *future* of X as well as its past and present.

The following proposition states some facts about the joint process (X,Z) that will be used later, and evaluates the information supply, storage and dissipation. A proof is given in Appendix.

Proposition 3.1

(i) *For each* $t\in[0,T]$, $((X_s,Y_s), s\in[0,t])$ *and* $((X_s,Z_s), s\in[t,T])$ *are conditionally independent given* (X_t,Z_t). *In particular, the joint process* (X,Z) *is Markov.*

(ii) *For each* $t\in[0,T]$, $((X_s,Y_s), s\in[0,t])$ *and* $(X_s, s\in[t,T])$ *are conditionally independent given* X_t.

(iii) *For each* $t\in[0,T]$, $(Y_s, s\in[0,t])$ *and* $((X_s,Z_s), s\in[t,T])$ *are conditionally independent given* Z_t. *In particular, the filter process* Z *is Markov.*

(iv) *For each* $t\in[0,T]$, *the information supply, storage and dissipation admit the following representations*:

$$S(t) = C(0) + \frac{1}{2}\mathbb{E}\int_0^t |g(X_s) - \langle g, Z_s\rangle|^2\,ds, \qquad (3.7)$$

$$C(t) = I(X_t; Z_t) = \mathbb{E}h(Z_t \mid P_t), \tag{3.8}$$

$$D(t) = \mathbb{E}I((X_s, s \in [0,t]); (Y_s, s \in [0,t]) \mid X_t), \tag{3.9}$$

where $\langle \cdot, \cdot \rangle$ *is as defined in* (2.7), *and* $I(\Theta; \Phi \mid \Psi)$ *is the* Ψ*-conditional mutual information*

$$I(\Theta; \Phi \mid \Psi) := h\left(P_{\Theta,\Phi|\Psi} \mid P_{\Theta|\Psi} \otimes P_{\Phi|\Psi}\right).$$

Remark 3.5 Results on the *Feller* Markov nature of Z and (X, Z), under slightly stronger conditions, can be found in [3]. Equation (3.7) was derived for the case of the multidimensional diffusion signal in [14].

Equation (3.9) shows that the dissipation is itself an (average) mutual information—that between the observation process $(Y_s, s \in [0,t])$ and the signal process $(X_s, s \in [0,t])$, conditioned on knowledge of X_t. It follows from part (ii) that (3.9) remains valid if we condition on $(X_s, s \in [t,T])$ instead of X_t, and this shows that $D(t)$ is that part of the information on X derived from Y that is of no use in estimating the values that X takes from time t onwards.

The filter can be thought of as a type of *data encoder*. At time t, it partially encodes the observation path $(Y_s, s \in [0,t])$ as the 'statistic' Z_t. Clearly this encoding is *lossy* in the sense that the observation path cannot be recovered from Z_t. Since both $(Y_s, s \in [0,t])$ and Z_t take values in uncountably infinite spaces they both contain an infinite amount of information, and this means that we cannot define a meaningful measure of the *total* information loss arising from the 'encoding'. However, we *can* define a meaningful measure of information loss in the context of a specific estimation problem; this is simply the amount by which the mutual information between the estimand and Z_t is less than that between the estimand and $(Y_s, s \in [0,t])$. For the problem of estimating the *whole* signal process X, the encoding of the filter at time t loses an amount of information $D(t)$, but for the problem of estimating its present and future it is lossless.

Proposition 3.1(iv) shows that the information supply has a well-defined *rate* at time t:

$$\dot{S}(t) = \frac{1}{2}\,\mathbb{E}\,|g(X_t) - \langle g, Z_t\rangle|^2. \tag{3.10}$$

Under various technical conditions on the regularity of the density processes p and ζ the same is true of the dissipation. The rate of change of the storage, $\dot{C}(t)$, can be found by formal application of Itô's rule to $\zeta_t \log(\zeta_t/p_t)(X_t)$, and this shows that the dissipation rate at time t is

$$\dot{D}(t) = \mathbb{E}\left(\frac{\mathcal{A}p_t}{p_t}\log p_t - \frac{\mathcal{A}\zeta_t}{\zeta_t}\log\zeta_t\right)(X_t). \tag{3.11}$$

This is a *Fisher* information quantity. It reveals the sensitivity of the mutual information $C(t)$ to the randomisation in the dynamics of the signal. This was noted in [31] for the multidimensional diffusion signal, where

$$\dot{D}(t) = \frac{1}{2}\,\mathbb{E}\,\nabla\log(\zeta_t/p_t)' a \nabla\log(\zeta_t/p_t)(X_t). \tag{3.12}$$

For the discrete-state signal,

$$\dot{D}(t) = \mathbb{E}\sum_{x,\tilde{x}}\left(\log\frac{\zeta_t(x)}{p_t(x)} - \log\frac{\zeta_t(\tilde{x})}{p_t(\tilde{x})}\right)A_{\tilde{x},x}\zeta_t(x). \tag{3.13}$$

Equations (3.10) and (3.11) show that the supply of information is entirely associated with the second integral in (3.3), and that its dissipation is entirely associated with the first. According to (3.10), $\dot{S}(t)$ is the 'signal-to-noise power ratio' of the running observation Y^r; according to (3.11), $\dot{D}(t)$ is a measure of the rate at which X 'forgets its past'.

The filter can also be thought of as being a *dynamical machine* that receives new observation-derived information at the rate $\dot{S}(t)$, and discards historical information at the rate $\dot{D}(t)$. It stores an amount of information $C(t)$ on X, which is that part of the supply to date useful for estimating the present and future of X.

4 Interactive Statistical Mechanics

This section develops statistical mechanical interpretations of the *interaction* between the signal and filter processes of Sects. 2 and 3. It starts with a *conditional* variant of the statistical mechanical system of Sect. 2. This is shown to be one component of an abstract *joint* system that obeys the statistical mechanical laws of Sect. 2. This system is shown to exhibit macroscopic flows of energy. The results are then interpreted in the context of a *partially observed* variant of the system of Sect. 2.

4.1 The Conditional System $\Sigma_{X|Z}$

The state of the statistical mechanical system of Sect. 2 at time t is the probability measure $P_t \in \mathbf{Z}$, which evolves according to (2.2). The state of the conditional system $\Sigma_{X|Z}$, at time t, is the filter variable $Z_t : \Omega \to \mathbf{Z}$, which fulfils (F1) and (F2), and evolves according to (3.3). Z_t is a random probability measure, and represents a refined version of P_t that takes into account the (random) partial observations available up to time t as well as the *prior* distribution P_t. In what follows, we define an energy function for $\Sigma_{X|Z}$ in such a way that Z_t is the minimum free energy state at time t.

In order to study the statistical mechanics of $\Sigma_{X|Z}$ it is also useful to define other, more general states not sharing this property. Let $(\tilde{Z}_t, t \in [0,T])$ be a stochastic process taking values in $\mathbf{Z}$, that satisfies (F1) for all t, and whose density process, $(\tilde{\zeta}_t, t \in [0,T])$, satisfies (3.3). (This will differ from $(Z_t, t \in [0,T])$ if $\tilde{Z}_0 \neq Z_0$.) The average value of the density, $\mathbb{E}\tilde{\zeta}_t$, corresponds to a state of the original system Σ_X, and satisfies (2.2).

The energy function of $\Sigma_{X|Z}$ is defined as follows:

$$H_{X|Z}(x,t) := -\log \zeta_t(x), \tag{4.1}$$

where $(\zeta_t, t \in [0,T])$ is the filter density process. $H_{X|Z}$ is time-dependent and random through its dependence on ζ_t. The internal energy, entropy and free energy of $\Sigma_{X|Z}$ have similar forms to those of Σ_X, defined in (2.8):

$$\begin{aligned}
\mathcal{E}_{X|Z}(\tilde{Z}_t, t) &:= \langle H_{X|Z}(\cdot,t), \tilde{Z}_t \rangle, \\
\mathcal{S}_{X|Z}(\tilde{Z}_t) &:= \mathcal{S}_X(\tilde{Z}_t) = -h(\tilde{Z}_t \mid \lambda_X), \\
\mathcal{F}_{X|Z}(\tilde{Z}_t, t) &:= \mathcal{E}_{X|Z}(\tilde{Z}_t, t) - \mathcal{S}_{X|Z}(\tilde{Z}_t).
\end{aligned} \tag{4.2}$$

These are random through their dependence on ζ_t and $\tilde{\zeta}_t$. The conditional system obeys the following variants of the statistical mechanical laws of Proposition 2.1.

Proposition 4.1

(i) *The unique minimiser of the free energy of the conditional system* $\varSigma_{X|Z}$ *at time* t *is the state* Z_t.

(ii) $\mathcal{F}_{X|Z}(Z_t, t) = 0$ *for all* t.

(iii) *If* $\mathbb{E}\mathcal{F}_{X|Z}(\tilde{Z}_t, t) < \infty$ *and* $h(\tilde{\Phi}_0 \mid \Phi_0) < \infty$, *where* Φ_0 *and* $\tilde{\Phi}_0$ *are the distributions of* Z_0 *and* $\tilde{Z}_0$, *respectively, then the free energy of* $\varSigma_{X|Z}$, *as its state* $\tilde{Z}_t$ *evolves, is a positive* $(Y_s, s \in [0,t])$-supermartingale.[1]

Proof Parts (i) and (ii) are no more than parts (i) and (ii) of Proposition 2.1 applied to $\tilde{Z}_t$, realisation by realisation. Part (iii) follows from Theorem 2.3 in [8]. □

For each realisation of the observation process Y, the conditional system can be considered to interact with a unit temperature heat bath in the same way as was $\varSigma_X$. As the state $\tilde{Z}_t$ evolves, changes in the entropy of the two-component system comprising $\varSigma_{X|Z}$ and the heat bath are the negative of the corresponding changes in the free energy of $\varSigma_{X|Z}$. With this interpretation, Proposition 4.1(iii) shows that the entropy of this two-component system is a $(Y_s, s \in [0,t])$-submartingale. (The negative of a supermartingale.) The *average* entropy of this two-component closed system is, therefore, non-decreasing. However, the entropy can decrease for individual realisations of the observation process.

We shall refer to Proposition 4.1(iii) as a *Conditional Second Law*. It is somewhat different from the *marginal* Second Law of Proposition 2.1(iii), because of the time-varying nature of $H_{X|Z}$. This allows the energy and entropy of $\varSigma_{X|Z}$ to change as a result of interaction with the filter, and not solely as a result of interaction with the heat bath. The filter controls $H_{X|Z}$ in order to hold $\varSigma_{X|Z}$ in its (time varying) minimum free energy state.

Remark 4.1 The fact that the filter does this is a special example of a very general property of Bayesian estimators developed in [33]. There, the two 'directions' of Bayesian estimation (likelihood function to posterior distribution, and vice-versa) are given dual variational characterisations. In particular, the posterior distribution is characterised as the unique minimiser of *apparent information*. In the present context, this apparent information is the free energy $\mathcal{F}_{X|Z}$.

Remark 4.2 A state $\tilde{Z}_t$ distinct from Z_t may be regarded as the filter variable for an incorrectly initialised filter. (See [8].) The statistical mechanical properties in Proposition 4.1 thus have direct relevance to error sensitivity issues in nonlinear filtering. In particular, the relative insensitivity of nonlinear filters to errors made in the distant past is a consequence of a 'dissipative' law having a statistical mechanical interpretation.

4.2 The Joint System $\varSigma_J$

Proposition 3.1(i) shows that the joint process (X, Z) is Markov, and so, modulo technical conditions ensuring the existence of a joint invariant distribution, it admits the statistical mechanical interpretation of Sect. 2. The state of the resulting *joint* system, $\varSigma_J$, at time t is the joint distribution of X_t and Z_t, $z \otimes \Phi_t$, where Φ_t is the marginal distribution of Z_t.

[1] A $(Y_s, s \in [0,t])$-supermartingale is a conditionally decreasing real-valued stochastic process; i.e. a process $(\eta_t, t \in [0,T])$ that is adapted to $(Y_s, s \in [0,t])$ (for all t, $\eta_t = G_t(Y_s, s \in [0,t])$ for some measurable $G_t : \mathbf{Y}_t \to \mathbb{R}$), and such that, for all $0 \le s \le t \le T$, $\mathbb{E}(\eta_t \mid Y_r, r \in [0,s]) \le \eta_s$.

$z \otimes \Phi_t$ is a probability measure on $\mathcal{X} \times \mathcal{Z}$, where $\mathcal{Z}$ is the Borel σ-field of $(\mathbf{Z}, d_Z)$. In fact, for any $B \in \mathcal{X}$ and $C \in \mathcal{Z}$,

$$\mathbb{P}(X_t \in B, Z_t \in C) = (z \otimes \Phi_t)(B \times C) = \int_C z(B)\Phi_t(dz). \tag{4.3}$$

Remark 4.3 The notation $\otimes$ here is a generalisation of the usual tensor product; it is *defined* by (4.3), and represents the factorisation of a probability measure on a product space into a marginal measure on one space and a regular conditional measure on the other. (This is possible since $(\mathbf{X}, d_X)$ is complete and separable. See, for example, Chap. 1 in [13].) $z \otimes \Phi_t$ is special type of probability measure on $\mathcal{X} \times \mathcal{Z}$ in that the regular z-conditional measure on $\mathcal{X}$ in the factorisation $z \otimes \Phi_t$ is z itself. More general states of the joint system not sharing this property occur in studies of errors in filter dynamics, but will not be considered further in this paper.

We assume that $(Z_t, t \in [0, T])$ has an invariant distribution Φ_{SS}. (See [3] on this issue.) We also assume that Φ_{SS} has a density, ϕ_{SS}, with respect to some reference measure λ_Z on $\mathcal{Z}$. (There may be no obvious candidate for λ_Z like the counting measure for the finite-state signal process, or the volume measure for the multidimensional diffusion signal. However, if all else fails, we can satisfy the above assumption by choosing $\lambda_Z = \Phi_{SS}$.)

The energy function for the joint system is as follows:

$$H_J(x, z) := -\log(q(x)\phi_{SS}(z)), \tag{4.4}$$

where q is the density of z. (According to Definition 3.1, every $z \in \mathbf{Z}$ has a density with respect to λ_X.) Following (2.8) and (4.2), we define the internal energy, entropy and free energy of Σ_J in state $z \otimes \Phi_t$ as follows:

$$\begin{aligned} \mathcal{E}_J(z \otimes \Phi_t) &:= \langle H_J, z \otimes \Phi_t \rangle, \\ \mathcal{S}_J(z \otimes \Phi_t) &:= -h(z \otimes \Phi_t \mid \lambda_X \otimes \lambda_Z), \\ \mathcal{F}_J(z \otimes \Phi_t) &:= \mathcal{E}_J(z \otimes \Phi_t) - \mathcal{S}_J(z \otimes \Phi_t). \end{aligned} \tag{4.5}$$

Proposition 3.1(iii) shows that we can also apply the statistical mechanical interpretation of Sect. 2 to the filter process, Z, alone. The corresponding *filter system*, Σ_Z, has energy function

$$H_Z(z) := -\log \phi_{SS}(z). \tag{4.6}$$

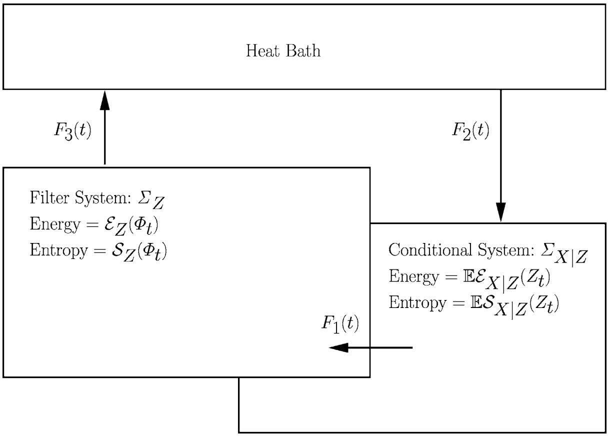
Its state at time t is Φ_t, in which its internal energy, entropy and free energy are as follows:

$$\begin{aligned} \mathcal{E}_Z(\Phi_t) &:= \langle H_Z, \Phi_t \rangle, \\ \mathcal{S}_Z(\Phi_t) &:= -h(\Phi_t \mid \lambda_Z), \\ \mathcal{F}_Z(\Phi_t) &:= \mathcal{E}_Z(\Phi_t) - \mathcal{S}_Z(\Phi_t). \end{aligned} \tag{4.7}$$

The joint energy function admits the following decomposition:

$$H_J(x, Z_t) = H_{X|Z}(x, t) + H_Z(Z_t). \tag{4.8}$$

Fig. 1 The joint system


Heat Bath
$F_3(t)$
$F_2(t)$
Filter System: Σ_Z
Energy = $\mathcal{E}_Z(\Phi_t)$
Entropy = $\mathcal{S}_Z(\Phi_t)$
Conditional System: $\Sigma_{X|Z}$
Energy = $\mathbb{E}\mathcal{E}_{X|Z}(Z_t)$
Entropy = $\mathbb{E}\mathcal{S}_{X|Z}(Z_t)$
$F_1(t)$

Also, according to the chain rule of relative entropy (Theorem C.3.1 in [13]), $\mathcal{S}_J$ can be decomposed as follows,

$$\mathcal{S}_J(z \otimes \Phi_t) = \mathbb{E}\mathcal{S}_{X|Z}(Z_t) + \mathcal{S}_Z(\Phi_t), \tag{4.9}$$

and so Σ_J comprises two subsystems: the filter system, Σ_Z, whose energy depends only on Z_t, and the conditional system, $\Sigma_{X|Z}$, whose energy depends on both X_t and Z_t. The latter is *always* in its minimum free energy state in this decomposition. This minimum free energy is zero, and so $\mathcal{F}_J(z \otimes \Phi_t) = \mathcal{F}_Z(Z_t)$.

Figure 1 shows the decomposition, and identifies the internal energies and entropies of the two subsystems. The original system of Sect. 2, Σ_X, is also a subsystem of Σ_J. It is split into two parts in the decomposition of Fig. 1; one part is the conditional system, and the other is subsumed into Σ_Z. This is reflected in the following splits in its energy and entropy:

$$H_X(x) = H_{X|Z}(x,t) + H_S(x, Z_t), \tag{4.10}$$

$$\mathcal{S}_X(P_t) = \mathbb{E}\mathcal{S}_{X|Z}(Z_t) + C(t), \tag{4.11}$$

where $H_S(x,z)$ $(:= H_X(x) + H_Z(z) - H_J(x,z))$ is the energy *shared* between Σ_X and Σ_Z. Since $\mathcal{F}_{X|Z}(Z_t,t) \equiv 0$,

$$\frac{d}{dt}\,\mathbb{E}\,\mathcal{E}_{X|Z}(Z_t,t) = \frac{d}{dt}\,\mathbb{E}\,\mathcal{S}_{X|Z}(Z_t) = \frac{d}{dt}\mathcal{S}_X(P_t) - \dot{S}(t) + \dot{D}(t). \tag{4.12}$$

In order to identify flows of energy in the joint system we consider the effects of turning off either the dynamics of X, or the observation mechanism at time t; the former can be achieved by temporarily setting $\mathcal{A}$ to zero, and the latter by temporarily setting g to zero. In both cases the parameters of the filtering problem are modified at time t, but the modified Z process retains the property of being the filter for the modified X process; in particular, the modified partially observed system $\Sigma_{X|Z}$ remains in its (modified) minimum free energy state.

Setting $\mathcal{A}$ to zero at time t freezes the signal X, thus disconnecting Σ_X from its heat bath. This freezes $\mathcal{S}_X(P_t)$ and $\mathcal{E}_X(P_t)$, but does not prevent a flow of energy between the two components of Σ_X identified in (4.10) and (4.11). In fact $\mathbb{E}\mathcal{S}_{X|Z}(Z_t)$ and $\mathbb{E}\mathcal{E}_{X|Z}(Z_t,t)$

continue to evolve according to (4.12), but with the first and third terms on the right-hand side set to zero. So

$$\frac{d}{dt}\,\mathbb{E}\,\mathcal{E}_{X|Z}(Z_t,t)\,|_{\mathcal{A}=0} = -\frac{d}{dt}\,\mathbb{E}\,H_S(X_t,Z_t)\,|_{\mathcal{A}=0} = -\dot{S}(t), \tag{4.13}$$

from which we can conclude that there is an observation-driven macroscopic flow of energy from $\Sigma_{X|Z}$ to the shared component (and hence to Σ_Z), of rate $F_1(t) = \dot{S}(t)$.

Setting g to zero at time t disconnects Z from X, thus preventing any further transfer of entropy from the first term to the second term on the right-hand side of (4.11). Furthermore, the optimality of the filter prevents any transfer of entropy in the other direction. (The filter never discards information relevant to the present or future of X.) Thus any change in $\mathcal{S}_{X|Z}(Z_t)$ is entirely due to an interaction between $\Sigma_{X|Z}$ and the heat bath. Since $\Sigma_{X|Z}$ is in a (unit temperature) minimum free energy state, any exchange of entropy with the heat bath must be accompanied by an equal exchange of energy. With g set to zero, the second term on the right-hand side of (4.12) is zero, and so the average rate of flow of energy from the heat bath to $\Sigma_{X|Z}$ is

$$F_2(t) := \frac{d}{dt}\mathcal{S}_X(P_t) + \dot{D}(t). \tag{4.14}$$

As discussed in Sect. 4.1, the time-varying nature of $H_{X|Z}$ allows energy and entropy exchange between $\Sigma_{X|Z}$ and Σ_Z. That this is *not* possible when g is set to zero, is confirmed by the following argument.

$$\frac{\partial}{\partial t}H_{X|Z}\,|_{g=0} = \frac{\mathcal{A}\zeta_t}{\zeta_t},$$

and (assuming that ζ_t is sufficiently regular), for any $B \in \mathcal{X}$,

$$\dot{Z}_t(B)\,|_{g=0} = \int_B (\mathcal{A}\zeta_t)(x)\lambda_X(dx).$$

So $\mathcal{E}_{X|Z}(Z_t,t)$ is differentiable, and

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_{X|Z}(Z_t,t)\,|_{g=0} &= \left\langle \frac{\partial}{\partial t}H_{X|Z}(\cdot,t)\,|_{g=0}, Z_t\right\rangle + \langle H_{X|Z}(\cdot,t), \dot{Z}_t\,|_{g=0}\rangle \\ &= \langle H_{X|Z}(\cdot,t), \dot{Z}_t\,|_{g=0}\rangle. \end{aligned} \tag{4.15}$$

The two terms on the right-hand side here represent interactions with the filter and heat bath, respectively. The fact that the first term is zero shows that, with g turned off, the statistical mechanics of $\Sigma_{X|Z}$ are not affected by the changing energy function, but only by the random evolution of X.

The role of the filter in the joint system is to control the energy function $H_{X|Z}$ so that $\Sigma_{X|Z}$ remains in its (time varying) minimum free energy state. When g is set to zero the filter can achieve this without exchanging energy with $\Sigma_{X|Z}$. However, when g is not set to zero the filter has to extract energy at the average rate $\dot{S}(t)$. In this case $H_{X|Z}$ is not differentiable with respect to t, and (4.15) has to be replaced by a stochastic differential equation, in which the non-zero quadratic variation of $H_{X|Z}(\cdot,t)$ plays a crucial role. In order to cause macroscopic energy changes in $\Sigma_{X|Z}$ in its minimum free energy state, any changes in $H_{X|Z}$ must be large. (In fact of order $\sqrt{\delta t}$ over a small time interval δt.)

When neither $\mathcal{A}$ nor g is set to zero we obtain the three macroscopic energy flows shown in Fig. 1, where

$$F_3(t) = \dot{S}(t) - \frac{d}{dt}\mathcal{E}_Z(\Phi_t) = \frac{d}{dt}\mathcal{S}_X(P_t) + \dot{D}(t) - \frac{d}{dt}\mathcal{E}_J(z \otimes \Phi_t). \tag{4.16}$$

Remark 4.4 In the stationary state, $z \otimes \Phi_{SS}$, all three energy flows have equal rates and, since the subsystems $\Sigma_{X|Z}$ and Σ_Z are then both in unit-temperature stationary states, the three energy flows are accompanied by equal entropy flows. Because of these macroscopic flows, the stationary state is a *non-equilibrium* state. However, it differs from physical non-equilibrium states in that the flows are driven by the mechanism of observation rather than external fields or boundary conditions.

In the stochastic dynamics framework, the non-equilibrium property of a stationary state manifests itself in the *irreversibility* of the associated Markov process. This can be quantified in terms of long-term averages of the relative entropy of the forward and backward *path* distributions of the process. (See, for example, [26], where this average is dubbed a *rate of entropy flow*.) These techniques cannot be applied *directly* to the joint system Σ_J since the relative entropies involved are generally infinite. This is because of the degeneracy of the filter process Z, which is typically of higher dimension than the running observation Y^r that 'drives' it. For example, the filter process for the finite-state signal is an $\mathbb{R}^n$-valued diffusion driven by the $\mathbb{R}^d$-valued process Y^r. If $d < n$, then the forward and backward path distributions of the filter process are typically mutually singular. The situation is even worse for the multidimensional diffusion signal, because the filter is then of *infinite* dimension, but still driven by a finite-dimensional running observation.

This problem is not encountered in [34], where a rate of *interactive* entropy flow for the linear Gaussian case is defined as the difference between the rate of entropy flow of the joint process (X, Z) and those of the marginal processes X and Z. This isolates that part of the irreversibility of the joint process associated with the interaction between its components. The rate of interactive entropy flow turns out to be the sum of the information supply and dissipation rates $\dot{S}$ and $\dot{D}$, a property that is not immediately obvious. This approach can be extended to the nonlinear situation of this paper by means of relaxation arguments. However, we do not pursue this further here. The case of the finite-state signal is developed in depth in [37]. A fully dynamic approach is taken there in the sense that the long-term averages of [26] are replaced by short-term averages (entropic derivatives), which admit processes away from invariant distributions, and even time-inhomogeneous processes. See, also, [36].

The joint system is a type of *perpetual motion machine* in the sense that it exhibits a macroscopic flow of energy without the presence of thermal gradients or external fields, or the increase in entropy these would cause. Of course the joint system is an *abstract* system not a physical system. Any physical realisation of the nonlinear filter would necessarily include components, such as operational amplifiers or digital computers, that created physical entropy in the course of their operation.

4.3 Statistical Mechanics with Partial Observations

If we take the view that observable components of a system are associated with *information* rather than entropy, then the partial observation of Σ_X, inherent in Y, enables the entropy of the former to be reduced. The filter holds an amount $C(t)$ of partial information on Σ_X, thereby reducing its entropy by the same amount, from $\mathcal{S}_X(P_t)$ to $\mathbb{E}\mathcal{S}_{X|Z}(Z_t)$. (See (4.11).)

This illustrates Landauer's Principle (in the reverse sense) in a quantitative way. The filter can be thought of as a *demon* whose aim is to minimise the entropy of Σ_X at each time t. Like Maxwell's demon [30], the filter does this by making use of measurements that are not macroscopic observables.

If we also take the view that observable components of energy are *work*, then the partial observation inherent in Y converts an amount $\mathbb{E}H_S(X_t, Z_t)$ of the energy of Σ_X into work. With this convention, the entire energy of the filter system Σ_Z, including that shared with Σ_X, is work. The remaining component of the energy of Σ_X, $\mathcal{E}_X(P_t) - \mathbb{E}H_S(X_t, Z_t)$, is the average energy of the conditional system, $\mathbb{E}\mathcal{E}_{X|Z}(Z_t, t)$. (See (4.10).) Since $\Sigma_{X|Z}$ is always in its minimum free energy state, this remaining energy can be regarded as *heat*.

In the invariant distribution, the heat and work components of the internal energy of Σ_X do not change. However, the filter continues to convert heat into work as new observation information becomes available. The resulting outflow of heat from $\Sigma_{X|Z}$ is balanced by an equal inflow from the heat bath. This inflow has its origins in energy *fluctuations*. According to the Convention on Randomisation of Sect. 2, X is randomised by fluctuations of energy between Σ_X and its heat bath. (These may be only partially *visible* in the sense of Sect. 2.) During these fluctuations, any energy incoming to Σ_X brings with it *new* entropy and must, therefore, go to the heat component. However, energy outgoing from Σ_X can be from either the heat or work component. Because of the optimality of the filter, the interface between the two components of Σ_X is a perfect 'energy valve'; once heat has become work it cannot revert back to heat without first passing through the heat bath. It is the combined effect of energy fluctuations and this valve that drive the macroscopic flow of energy.

In the invariant distribution, the inflow of energy to the work component of Σ_X, and hence to Σ_Z, is balanced by an equal outflow from Σ_Z back to the heat bath. This outflowing work is associated with degrees of freedom of the filter that no longer bear information about the signal. The filter re-entropises this work (thereby turning it back into heat) by discarding information at the rate $\dot{D}_{SS}$. This illustrates Landauer's Principle (in the forward sense) in a quantitative way.

4.4 The Linear Gaussian Case

The joint system was investigated in [34] for the special case in which X is a linear Gaussian diffusion process and the observation function g is linear. This is a special case of the multidimensional diffusion example, in which the drift coefficient b and observation function g are linear, the diffusion coefficient a is constant, and X_0 and Z_0 are jointly Gaussian. Because of the special properties of Gaussian distributions, the conditional distribution, Z_t, is also Gaussian in this case, for each realisation of $(Y_s, s \in [0, t])$, and so can be parametrised by its mean vector, $\hat{X}_t$ $(:= \mathbb{E}(X_t \mid Y_s, s \in [0, t]))$, and covariance matrix, Q_t $(:= \mathbb{E}((X_t - \hat{X}_t)(X_t - \hat{X}_t)' \mid Y_s, s \in [0, t]))$. The evolution equations of $\hat{X}$ and Q are those of the Kalman-Bucy filter. (See, for example, [10] or [19].) This particular example of 'nonlinear' filtering is, of course, linear. It has found wide application owing to the finite-dimensional nature of the Gaussian subset of $\mathbf{Z}$, which greatly simplifies its implementation. It also possesses other special properties; for example, it turns out that the covariance matrix Q_t does not depend on Y, but only on t, and so all the information stored by the filter at time t is held in the conditional mean vector, $\hat{X}_t$. This takes values in the same space as X_t, and so the essential features of the filter variable Z_t are of the same dimension as the signal X_t. (This is in contrast with the generic multidimensional diffusion case, in which the filter variable is necessarily of infinite dimension.)

5 Time Reversal

Throughout this section we shall assume that X_0 and the (surrogate) initial observation, Z_0, have joint invariant distribution $z \otimes \Phi_{SS}$. The statistical mechanical systems Σ_X, Σ_Z and Σ_J will then remain in their respective minimum free energy states for all t. Because of this, they will satisfy the Second Law of Proposition 2.1(iii) in *reverse*, as well as forward, time.

Consider, for example, the signal system Σ_X. Since the Markov property is time-symmetric, the time-reversed process $(X_{T-t}, t \in [0, T])$ is also Markov. Since $(\mathbf{X}, d_X)$ is complete and separable, $(X_{T-t}, t \in [0, T])$ will have a time-homogeneous transition funtion $\bar{\Pi} : [0, T] \times \mathbf{X} \times \mathcal{X} \to [0, 1]$. Let $(\bar{X}_t, t \in [0, T])$ be a Markov process with this transition function and single-time marginal distributions $(\bar{P}_t, t \in [0, T])$ (not necessarily P_{SS}). Clearly P_{SS} is an invariant distribution for this process, and so $\bar{X}$ can be associated with an (abstract) statistical mechanical system in the manner of Sect. 2. This system has the same energy function as Σ_X, and also the same internal energy, entropy and free energy functionals, $\mathcal{E}_X$, $\mathcal{S}_X$ and $\mathcal{F}_X$. The time-reversed system, $\Sigma_{\bar{X}}$, obeys the Second Law of Proposition 2.1(iii). In the special case that $\Sigma_{\bar{X}}$ starts in the invariant distribution P_{SS} we can, for the sake of convenience, choose $\bar{X}_t = X_{T-t}$ for $t \in [0, T]$. We then have a *single process*, X, that describes both systems Σ_X and $\Sigma_{\bar{X}}$ over the time interval $[0, T]$. Of course, we are not implying that a particular physical system will run backwards in time, but that two different systems can be described by the same Markov process running forwards and backwards in its invariant distribution.

When X is the finite-state process, provided $p_{SS}(x) > 0$ for all $x \in \{1, 2, \ldots, n\}$, it is easy to show that $\bar{X}$ is a Markov jump process with rate matrix $\bar{A}$, defined as follows

$$\bar{A}_{x,\tilde{x}} = A_{\tilde{x},x} \frac{p_{SS}(x)}{p_{SS}(\tilde{x})} \quad \text{if } \tilde{x} \neq x \quad \text{and} \quad \bar{A}_{x,x} = -\sum_{\tilde{x} \neq x} \bar{A}_{\tilde{x},x}. \tag{5.1}$$

When X is the multidimensional diffusion process, it can be shown, under mild technical conditions, that $\bar{X}$ is also an $\mathbb{R}^n$-valued diffusion process with the same diffusion coefficient as X, but drift coefficient, $\bar{b}$, defined as follows:

$$\bar{b} := -b + \mathrm{vec}_i\{\mathrm{div}(a_i)\} + a\nabla \log p_{SS}, \tag{5.2}$$

where a_i is the ith column of a. (See [18] for a very general case admitting degenerate a.) The characterisation of time-reversed Markov processes goes back to [35], which treats the time-homogeneous case in the invariant distribution.

Time reversal can also be applied to the filter process, Z, and the joint process, (X, Z), and leads to corresponding abstract statistical mechanical systems in reverse time. The processes Z and (X, Z) describe reverse-time systems, $\Sigma_{\bar{Z}}$ and $\Sigma_{\bar{J}}$, as well as the forward-time systems Σ_Z and Σ_J. We shall not attempt to characterise the time-reversed transition functions in the general case as this is not required in what follows.

For each $t \in [0, T]$, let

$$X_t^* := Z_{T-t} \quad \text{and} \quad Z_t^* := X_{T-t}. \tag{5.3}$$

Since Z is obtained from Y, parts (i)–(iii) of Proposition 3.1 remain true if Y is replaced by Z. They are then *invariant* under the transformation of (5.3). This shows that we can consider X^* and Z^* as being the signal and filter processes of a *dual* problem. The dual signal, X^*, is a Markov process in its own right evolving on the metric space $(\mathbf{Z}, d_Z)$, and the $\mathbf{X}$-valued process Z^* is a filter for this dual signal in the sense that, for any $t \in [0, T]$,

Z_t^* is a *sufficient statistic* for estimating the present and future of X^* from the past of Z^*. In fact Proposition 3.1(ii) shows that, for any $B \in \mathcal{Z}$,

$$\mathbb{P}\left(X_t^* \in B \mid Z_s^*, s \in [0,t]\right) = \mathbb{P}\left(X_t^* \in B \mid\ Z_t^*\right).$$

Clearly Z_t^* is not the $(Z_s^*, s \in [0,t])$-conditional distribution of X_t^*; (it does not even take values in the space of probability measures on $\mathcal{Z}$). However, like the conditional mean vector $\hat{X}_t$ of the Kalman-Bucy filter, it carries all the information required to construct this conditional distribution. Since the original problem admits an observation-conditional density process (ζ of (3.3)), the same is true of the dual problem; in fact

$$\mathbb{P}\left(X_t^* \in B \mid Z_s^*, s \in [0,t]\right) = \int_B \zeta_t^*(z)\lambda_Z(dz),$$

where

$$\zeta_t^*(z) = \frac{q(Z_t^*)\phi_{SS}(z)}{p_{SS}(Z_t^*)}, \tag{5.4}$$

and q is the density of z. (Equation (5.4) is no more than Bayes' formula applied between X_t^* and Z_t^*.)

The dual filter process, Z^*, is Markov and it easily follows from Proposition 3.1(i) that it is also X^*-conditionally Markov. When X is the finite-state process, Z^* is a Markov jump process with marginal rate matrix $\bar{A}$ of (5.1), and X^**-conditional* rate matrix $\breve{A}(X_t^*)$ at time t, where

$$\breve{A}(z)_{x,\tilde{x}} = A_{\tilde{x},x}\frac{z(\{x\})}{z(\{\tilde{x}\})} \quad \text{if } \tilde{x} \neq x \quad \text{and} \quad \breve{A}(z)_{x,x} = -\sum_{\tilde{x}\neq x}\breve{A}(z)_{\tilde{x},x}. \tag{5.5}$$

(See [37].) When X is the multidimensional diffusion process (and under the conditions of [18]), Z^* is an $\mathbb{R}^n$-valued diffusion process with diffusion coefficient a, and drift coefficient $\bar{b}$, of (5.2). Furthermore, it is X^**-conditionally* an $\mathbb{R}^n$-valued diffusion process with diffusion coefficient a and X_t^*-dependent drift coefficient $\breve{b}(\cdot, X_t^*)$ at time t, where $\breve{b} : \mathbf{X} \times \mathbf{Z} \to \mathbb{R}^n$ is defined as follows:

$$\breve{b}(\cdot, z) = -b + \mathrm{vec}_i\{\mathrm{div}(a_i)\} + a\nabla \log q, \tag{5.6}$$

and q is the density of z. (See Theorem 4.2 in [33].)

So far we have identified dual signal and filter processes, X^* and Z^*, but not a dual *observation* process. One possibility is to consider Z^* itself as being the observation process, in which case there is no processing for the filter to do. It is certainly reasonable to define the dual *initial* observation, ψ^*, to be Z_0^* (just as Z_0 was regarded as a surrogate observation for ψ in Sect. 3). However, there are other interesting possibilities for the dual *running* observation, Y^{*r}. When X is the finite-state process we can define this to be an n-vector of Poisson counting processes with rates that depend on the value of the dual signal process, X^*. For a full development, see [37]. When X is the multidimensional diffusion process, the dual filter process Z^* is a solution of the following Itô stochastic differential equation:

$$Z_t^* = Z_0^* + \int_0^t \breve{b}(Z_s^*, X_s^*)\,ds + \int_0^t \sigma(Z_s^*)\,dW_s^*, \tag{5.7}$$

where σ is as in (2.5) and $(W^*_t, t \in [0, T])$ is an n-dimensional Brownian motion. This equation can be re-expressed as follows:

$$Z^*_t = Z^*_0 + \int_0^t (-b + \mathrm{vec}_i\{\mathrm{div}(a_i)\})(Z^*_s)\, ds + \int_0^t \sigma(Z^*_s)\, dY^{*r}_s, \tag{5.8}$$

where

$$Y^{*r}_t = \int_0^t g^*(Z^*_s, X^*_s)\, ds + W^*_t, \tag{5.9}$$

and, for $x \in \mathbf{X}$ and $z \in \mathbf{Z}$ with density q,

$$g^*(x, z) := (\sigma' \nabla \log q)(x). \tag{5.10}$$

This identifies a dual running observation of the same *signal-plus-white-noise* type as that of the original filter, Y^r.

In both of these examples Y^* is actually defined in terms of Z^*, and thus contains no more 'noise' than Z^*; however it is easy to construct dual observation processes for which this is not the case. The only requirements of Y^* are that:

(O1) Z^*_t is $(Y^*_s, s \in [0, t])$-measurable for all t;
(O2) X^* and $(Y^*_s, s \in [0, t])$ are Z^*_t-conditionally independent for all t.

The first of these requires that the dual filter should be causally derivable from Y^*, the second that any randomness in Y^* that is not in Z^* should bear no additional information about X^*. (See [37] for further discussion.)

As with the original filter, we can identify information supply, storage and dissipation processes for the dual filter: for any $t \in [0, T]$,

$$\begin{aligned} S^*(t) &:= I((X^*_s, s \in [0, T]);\ (Y^*_s, s \in [0, t])), \\ C^*(t) &:= I((X^*_s, s \in [t, T]);\ (Y^*_s, s \in [0, t])), \\ D^*(t) &:= S^*(t) - C^*(t). \end{aligned} \tag{5.11}$$

Remark 5.1 (Technical) The paths of Y^* are assumed here to take values in a metric space, for example, the Skorohod space $D([0, T]; \mathbb{R}^k)$ of vector-valued functions with left and right limits at all points $t \in (0, T)$ that are left or right continuous at all $t \in [0, T]$. The paths of X^* take values in the metric space $C([0, T]; \mathbf{Z})$, which is metrised by the supremum metric.

It follows from (O2) and Proposition 3.1(iii) that

$$C^*(t) = I(X^*_t; Z^*_t) = C(T - t).$$

Furthermore, it follows from (O2), the chain rule of relative entropy and (3.9) that, for any $t \in [0, T]$,

$$S(T) = S^*(T) = S^*(T - t) + D(t) = S(t) + D^*(T - t),$$

so that, for any $0 \le s \le t \le T$,

$$\begin{aligned} S(t) - S(s) &= D^*(T - s) - D^*(T - t), \\ D(t) - D(s) &= S^*(T - s) - S^*(T - t). \end{aligned} \tag{5.12}$$

(These expressions remain valid without any assumptions on dynamics or invariant distributions. See [37].) Over the time interval $[s,t]$ an amount $S(t)-S(s)$ of new information is supplied by the observations of the original problem. By *new* we mean that it relates to a dependency between X_t and Z_t that does not have its origins in a dependency between $(X_r, r\in[0,s])$ and $(Z_r, r\in[0,s])$. Because of this, the dual filter *dissipates* it over the reverse time interval $[T-t, T-s]$. The dissipation of one filter is the supply of its dual.

Since (X,Z) is in the invariant distribution $z\otimes\Phi_{SS}$ the rates of information supply and dissipation of both original and dual filters are equal. For example, when X is the finite-state signal, it follows from (3.10) and (3.13) that

$$\begin{aligned}\dot S_{SS}=\dot D_{SS}=\dot S^*_{SS}=\dot D^*_{SS}&=\frac{1}{2}\,\mathbb{E}\,|g(X_t)-\langle g,Z_t\rangle|^2\\&=\mathbb{E}\sum_{x,\tilde x}\left(\log\frac{\zeta_t(x)}{p_t(x)}-\log\frac{\zeta_t(\tilde x)}{p_t(\tilde x)}\right)A_{\tilde x,x}\zeta_t(x)\\&=\mathbb{E}\sum_x\log\left(\frac{q(x)p_{SS}(Z^*_t)}{q(Z^*_t)p_{SS}(x)}\right)g^*(Z^*_t,X^*_t)_x,\end{aligned}\tag{5.13}$$

where

$$g^*(x,z)_{\tilde z}:=(1-d_X(\tilde x,x))\breve A(z)_{\tilde x,x}.$$

When X is the multidimensional diffusion process, it follows from (3.10) and (3.12) that

$$\begin{aligned}\dot S_{SS}=\dot D_{SS}=\dot S^*_{SS}=\dot D^*_{SS}&=\frac{1}{2}\,\mathbb{E}\,|g(X_t)-\langle g,Z_t\rangle|^2\\&=\frac{1}{2}\,\mathbb{E}\,\nabla\log(\zeta_t/p_t)'a\nabla\log(\zeta_t/p_t)(X_t)\\&=\frac{1}{2}\,\mathbb{E}\,\left|g^*(Z^*_t,X^*_t)-\langle g^*(Z^*_t,\cdot),\zeta^*_t d\lambda_Z\rangle\right|^2,\end{aligned}\tag{5.14}$$

where g^* is as defined in (5.10).

The statistical mechanical properties of the joint process (X,Z), of Sect. 4, clearly apply also to the dual joint process (X^*,Z^*). The *dual joint system*, Σ_{J^*}, thus comprises a *dual filter system*, Σ_{Z^*}, and a *dual conditional system*, $\Sigma_{X^*|Z^*}$. These have energy functions, H_{J^*}, H_{Z^*} and $H_{X^*|Z^*}$, and internal energies and entropies defined in the obvious way. These obey the statistical mechanical laws of Propositions 2.1 and 4.1. The equivalent of (4.12) for the dual joint system is

$$\frac{d}{dt}\,\mathbb{E}\,\mathcal{E}_{X^*|Z^*}(Z^*_t,t)=\frac{d}{dt}\,\mathbb{E}\,\mathcal{S}_{X^*|Z^*}(Z^*_t)=\frac{d}{dt}\mathcal{S}_{X^*}(P^*_t)-\dot S^*(t)+\dot D^*(t).$$

This can be used to find macroscopic flows of energy between the components of Σ_{J^*} by the techniques of Sect. 4.2; these involve temporarily setting $\mathcal{A}^*$ or g^* to zero. Of course, changing a parameter of the dual system in this way will destroy the special nature of the dual filter, whose density ζ^* will no longer depend on Y^* through the simple statistic Z^* as in (5.4), but will evolve with far more complex dynamics. However, its effect on the dual conditional system $\Sigma_{X^*|Z^*}$ will still be to change the latter's energy function, $H_{X^*|Z^*}$, in a way that keeps $\Sigma_{X^*|Z^*}$ in its minimum free energy state, and this is all that is required for the calculation of macroscopic energy flows. These techniques show that there is a dual-observation driven

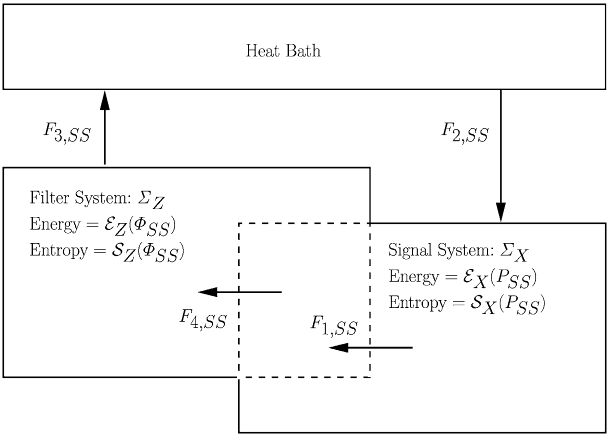

Heat Bath
$F_{3,SS}$
$F_{2,SS}$
Filter System: Σ_Z
Energy = $\mathcal{E}_Z(\Phi_{SS})$
Entropy = $\mathcal{S}_Z(\Phi_{SS})$
Signal System: Σ_X
Energy = $\mathcal{E}_X(P_{SS})$
Entropy = $\mathcal{S}_X(P_{SS})$
$F_{4,SS}$
$F_{1,SS}$

Fig. 2 Full energy flows in the joint system

flow of energy from $\Sigma_{X^*|Z^*}$ to Σ_{Z^*}, and a flow of energy from the heat bath to $\Sigma_{X^*|Z^*}$. Since Σ_{J^*} is always in its stationary state, both of these flows have rate $\dot{S}^*_{SS}$ $(= \dot{S}_{SS})$. The first flow can be interpreted in *forward* time as a *dissipation-driven* flow of energy from the shared component, $\mathbb{E}_{SS}H_S(X_t, Z_t)$, to the 'conditional filter', $\Sigma_{Z|X}$. The latter is the forward time interpretation of $\Sigma_{X^*|Z^*}$ and has internal energy $\mathcal{E}_Z(\Phi_{SS}) - \mathbb{E}_{SS}H_S(X_t, Z_t)$. In forward time, the second flow is from this conditional filter to the heat bath.

Figure 2 shows all the energy flows of Σ_J. The flow into Σ_Z, $F_{1,SS}$, is driven by information supply, and that out of Σ_X, $F_{4,SS}$, by information dissipation. All energy flows are accompanied by equal entropy flows. As is clear from Fig. 2, the conditional system, $\Sigma_{X|Z}$, is part of the heat bath with which the filter system, Σ_Z, interacts. Moreover, it is a part of the heat bath that only *supplies* energy and entropy. Similarly, $\Sigma_{Z|X}$, is part of the heat bath with which Σ_X interacts, and a part that only *absorbs* entropy and energy. These roles are exchanged in the dual system. The joint system has 'energy valves' at the two dashed interfaces in Fig. 2, that prevent energy flow from left to right. Combined with the statistical fluctuations of energy caused by the heat bath, these are what drive macroscopic flows of energy.

6 Complexity Issues

As discussed in Sect. 4.3, the filter system Σ_Z can be thought of as a dynamical machine that makes optimal use of the partial observations Y to extract information and work from the signal system Σ_X. As a rule, this dynamical machine is considerably more complex than the system from which it extracts; for one thing its phase space, $\mathbf{Z}$, is the space of probability measures on that of Σ_X. This complexity is one of the features that make nonlinear filters difficult to implement in practice. An important exception occurs when X is a linear Gaussian diffusion process, and the observation function is linear ($g(x) = Gx$). (See the discussion in Sect. 4.4.) Then the $(Y_s, s \in [0, t])$-conditional mean, $\hat{X}_t$, of the Kalman-Bucy filter is a sufficient statistic for computing Z_t, and since $\hat{X}_t$ evolves according to linear dynamics on $\mathbb{R}^n$, the complexity of Σ_Z is no greater than that of Σ_X. This feature is a property of the algebraic structure of the filtering equation (3.3) in this case.

The filter density process for a multidimensional diffusion signal, ζ of (3.3), does not in general have a finite-dimensional underlying structure. It follows from some important theorems of nonlinear filtering (see [6, 7, 11]) that, modulo technical conditions, there exists a continuous functional, $Q : [0, T] \times \mathbb{R}^n \times \mathbf{Y}_T \to \mathbb{R}^+$, such that for (almost) all (t, x),

$$\zeta_t(x) = \left(\int Q(t, \tilde{x}, (Y_s, s \in [0, T]))\, d\tilde{x} \right)^{-1} Q(t, x, (Y_s, s \in [0, T])), \tag{6.1}$$

where $\mathbf{Y}_T$ is as defined in Definition 3.1. $Q(t, x, (Y_s, s \in [0, T]))$ is an *un-normalised* density process in that its integral over x is not unity. Furthermore, for any $z_0 \in \mathbf{Z}$ having a density q_0, and any *differentiable* $y^r : [0, T] \to \mathbb{R}^d$, $Q(t, x, (z_0, y^r))$ satisfies the following (non-stochastic) partial differential equation:

$$\frac{\partial Q}{\partial t} = \left(\mathcal{A} - \frac{1}{2}|g|^2 \right) Q + (\dot{y}_t^r)' g Q; \qquad Q(0, x, (z_0, y^r)) = q_0(x), \tag{6.2}$$

where $\mathcal{A}$ is the differential operator of (2.4), and g is the observation function of (3.1). This is the so-called *pathwise* version of the Zakai equation of nonlinear filtering [11]. Equation (6.2) is a multi-linear equation involving the $d + 1$ vector fields $(g_i, i = 1, 2, \ldots, d)$ and $\mathcal{A} - \frac{1}{2}|g|^2$. If the Lie algebra, $\mathbf{L}$, generated by these is of finite dimension, and the initial conditional distribution Z_0 is appropriate, then we might expect the nonlinear filter to be expressible in terms of a sufficient statistic that evolves on a finite-dimensional manifold. In the case of the Kalman-Bucy filter, $\mathbf{L}$ is contained in the $(2n + 1)$-dimensional Lie algebra generated by the vector fields $(x_i, \partial/\partial x_i, i = 1, 2, \ldots, n)$ and $\mathcal{A} - \frac{1}{2}|Gx|^2$.

The Kalman-Bucy filter is not the only filter for a multidimensional diffusion process admitting a finite-dimensional implementation. Other examples were discovered by Beneš [1] and Daum [9]. However, such examples are rare. (The reader interested in connections between the algebraic structure and dimension of nonlinear filters is referred to [5, 29, 32] for further information.)

The complexity of nonlinear filters for multidimensional diffusion processes in the general case is a result of the interaction between the vector fields $g_1, g_2, \ldots, g_d$ and $\mathcal{A} - \frac{1}{2}|g|^2$. The multiplicative vector fields g_i are associated purely with the *supply* of information, whereas $\mathcal{A}$ is associated purely with its *dissipation*. It is the interplay between these two mechanisms that underlies the complexity of nonlinear filters in the general case. In order to perform its role, the filter has to store much more information than that relating to the current signal value, X_t. In fact, in total, the filter stores the (typically) *infinite* amount of information $I(Z_t; Z_t)$. Compare this with the information it stores on X_t, $C(t)$; under the finite energy condition (3.2) this is finite. The filter needs to store a large amount of 'management' information in order to know *how* to process incoming information and dissipate redundant information correctly.

The dual system of Sect. 5 is in striking contrast with this general rule of complexity. Here the phase space and dynamics of the filter system, Σ_{Z^*}, are actually simpler than those of the signal system, Σ_{X^*}. This is due to the special relationship between the prior distribution of the dual signal and the observation mechanism, which allows the dual filter density to be expressed in the simple product form of (5.4). The dual supply and dissipation processes interact in a particularly simple way. In some sense, the dual filter does not have as large a 'management overhead' of information as the original filter. For example, if X is the finite-state process and $n = 2$ then the total information stored by the dual filter at time t is $I(Z_t^*; Z_t^*) \le \log(2)$, less than one bit. This is of the same order of magnitude as the information it stores on X_t^*. This asymmetry is an indicator of the direction of time, even if the processes X and Z are in their joint invariant distribution, and even if the joint systems Σ_J and Σ_{J^*} do not produce entropy in this distribution. Even though the original and dual joint systems, Σ_J and Σ_{J*}, have identical statistical mechanical properties, the relative complexity of the signal and filter systems is highly asymmetric. This is not so if the original filter has finite-dimensional algebraic structure; a fact that is undoubtedly related to underlying dynamics that conserve more physical quantities than energy and entropy.

Acknowledgements The author would like to associate this work with Professor Sanjoy Mitter of the Laboratory for Information and Decision Systems (LIDS) at MIT. He would also like to thank Professor Mitter for suggesting this subject area as one fruitful for research, and providing encouragement and support during a number of visits made by the author to LIDS.

Appendix: Proof of Proposition 4.1

We start by introducing some notation. For a stochastic process ϕ taking values in a complete separable metric space Φ, and for $0 \le t \le s \le T$, we denote by $\mathcal{F}^{\phi}_{t,s}$ the σ-field generated by the process $(\phi_r, r \in [t,s])$. If Φ is a linear space over $\mathbb{R}$, we denote by $\mathcal{F}^{\Delta\phi}_{t,s}$ the σ-field generated by the *increments* process $(\phi_r - \phi_t, r \in [t,s])$.

For sub-σ-fields $\mathcal{F}_1, \mathcal{F}_2, \mathcal{G} \subset \mathcal{F}$, we use the notation $CI(\mathcal{F}_1, \mathcal{G}, \mathcal{F}_2)$ to indicate that $\mathcal{F}_1$ and $\mathcal{F}_2$ are $\mathcal{G}$-conditionally independent. Thus, for part (ii) of the proposition, we need to prove that $CI(\mathcal{F}^X_{t,T}, \mathcal{F}^X_{t,t}, \mathcal{F}^{X,Y}_{0,t})$. We shall make frequent use of Proposition 3.2a in [41], which states that

$$CI(\mathcal{F}_1, \mathcal{G}, \mathcal{F}_2 \vee \mathcal{F}_3) \quad \Longleftrightarrow \quad CI(\mathcal{F}_1, \mathcal{G}, \mathcal{F}_2) \quad \text{and} \quad CI(\mathcal{F}_1, \mathcal{G} \vee \mathcal{F}_2, \mathcal{F}_3). \tag{A.1}$$

By hypothesis, for any $t \in [0,T]$, $CI(\sigma(\psi) \vee \mathcal{F}^{\Delta W}_{0,t}, \mathcal{F}^X_{0,0}, \mathcal{F}^X_{0,T})$, and a left-to-right application of (A.1) shows that $CI(\sigma(\psi) \vee \mathcal{F}^{\Delta W}_{0,t}, \mathcal{F}^X_{0,t}, \mathcal{F}^X_{t,T})$. This, together with the Markov property of X, $CI(\mathcal{F}^X_{t,T}, \mathcal{F}^X_{t,t}, \mathcal{F}^X_{0,t})$, and a right-to-left application of (A.1) proves part (ii).

It follows from part (ii) and a left-to-right application of (A.1) that $CI(\mathcal{F}^X_{t,s}, \mathcal{F}^{X,Z}_{t,t}, \mathcal{F}^{X,Y}_{0,t})$ for any $s \in [t,T]$, and since $CI(\mathcal{F}^{X,Y}_{0,t}, \mathcal{F}^{X,Z}_{t,t} \vee \mathcal{F}^X_{t,s}, \mathcal{F}^Z_{t,t})$ (as a result of the second σ-field containing the third), a right-to-left application shows that $CI(\mathcal{F}^X_{t,s} \vee \mathcal{F}^Z_{t,t}, \mathcal{F}^{X,Z}_{t,t}, \mathcal{F}^{X,Y}_{0,t})$. Since $\mathcal{F}^{\Delta W}_{t,s}$ is independent of all the σ-fields here it can be added to the first term to yield, in particular,

$$CI(\mathcal{F}^X_{t,s} \vee \mathcal{F}^Z_{t,t} \vee \mathcal{F}^{\Delta Y}_{t,s}, \mathcal{F}^{X,Z}_{t,t}, \mathcal{F}^{X,Y}_{0,t}). \tag{A.2}$$

We will thus have proved part (i) if we can show that

$$\mathcal{F}^Z_{s,s} \subseteq \mathcal{F}^Z_{t,t} \vee \mathcal{F}^{\Delta Y}_{t,s} \vee \mathcal{N}, \tag{A.3}$$

where $\mathcal{N}$ comprises the null sets of $\mathbb{P}$.

It easily follows from the definition of Z_t that

$$CI(\mathcal{F}^Y_{0,t}, \mathcal{F}^Z_{t,t}, \mathcal{F}^X_{t,t}), \tag{A.4}$$

and since (A.2) implies in particular that $CI(\mathcal{F}^X_{t,s} \vee \mathcal{F}^{\Delta Y}_{t,s}, \mathcal{F}^{X,Z}_{t,t}, \mathcal{F}^Y_{0,t})$, a right-to-left application of (A.1) shows that $CI(\mathcal{F}^X_{t,s} \vee \mathcal{F}^{\Delta Y}_{t,s}, \mathcal{F}^Z_{t,t}, \mathcal{F}^Y_{0,t})$. A left-to-right application to this shows that $CI(\mathcal{F}^X_{t,s}, \mathcal{F}^Z_{t,t} \vee \mathcal{F}^{\Delta Y}_{t,s}, \mathcal{F}^Y_{0,t})$. Now $CI(\mathcal{F}^X_{t,s}, \mathcal{F}^Z_{t,t} \vee \mathcal{F}^Y_{0,s}, \mathcal{F}^{\Delta Y}_{t,s})$ (as a result of the second σ-field containing the third), and so a right-to-left application of (A.1) shows that $CI(\mathcal{F}^X_{t,s}, \mathcal{F}^Z_{t,t} \vee \mathcal{F}^{\Delta Y}_{t,s}, \mathcal{F}^Y_{0,s})$. It thus follows that, for any $B \in \mathcal{X}$,

$$Z_s(B) = \mathbb{E}(\mathbf{1}_B(X_s) \mid \mathcal{F}^Z_{t,t} \vee \mathcal{F}^{\Delta Y}_{t,s}) \quad \text{a.s.}$$

Thus $\sigma(Z_s(B)) \subseteq \mathcal{F}^{Z}_{t,t} \vee \mathcal{F}^{\Delta Y}_{t,s} \vee \mathcal{N}$, and (A.3) follows from Lemma A.5.1 in [13]. This proves part (i).

It easily follows from (A.4) that $CI(\mathcal{F}^{X,Z}_{t,t}, \mathcal{F}^{Z}_{t,t}, \mathcal{F}^{Y}_{0,t})$, which, together with part (i) and a right-to-left application of (A.1), proves part (iii).

It follows from part (ii) and (A.4) that

$$C(t) = I(X_t; (Y_s, s \in [0,t])) = I(X_t; Z_t),$$

which proves (3.8). Furthermore, it follows from part (ii) and a simple variant (that includes the initial observation) of Theorem 7.23 in [27] that

$$S(t) = I((X_s, s \in [0,t]); (Y_s, s \in [0,t])) = \mathbb{E}\log M_t,$$

where

$$M_t = \frac{dZ_0}{dP_0}(X_0)\exp\left(\int_0^t (g(X_s) - \langle g, Z_s\rangle)'\,dW_s + \frac{1}{2}\int_0^t |g(X_s) - \langle g, Z_s\rangle|^2\,ds\right).$$

Since $(g(X_t), t \in [0,T])$ satisfies (3.2) the stochastic integral here has zero mean, (see, for example, [21]), and this proves (3.7).

Let $\xi_t := (X_s, s \in [0,t])$ and $O_t := (Y_s, s \in [0,t])$. Since the paths of X are right-continuous and have left limits ξ_t takes values in the complete separable metric space $D([0,t]; \mathbf{X})$. This fact enables the following factorisations of P_{ξ_t} and $P_{\xi_t|O_t}$ to be made: $P_{\xi_t} = P_{\xi_t|X_t} \otimes P_t$, and $P_{\xi_t|O_t} = P_{\xi_t|X_t,O_t} \otimes Z_t$. It then follows from the chain rule of relative entropy (see, for example, Theorem C.3.1 in [13]) that

$$D(t) = \mathbb{E}h(P_{\xi_t|X_t,O_t}(\cdot, X_t, O_t) \mid P_{\xi_t|X_t}(\cdot, X_t)),$$

and (3.9) follows from the defintion of the conditional mutual information. This completes the proof of part (iv).

References

1. Beneš, V.E.: Exact finite-dimensional filters for certain diffusions with nonlinear drift. Stochastics **5**, 65–92 (1981)
2. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Macroscopic fluctuation theory for stationary non-equilibrium states. J. Stat. Phys. **107**, 635–675 (2002)
3. Bhatt, A.G., Budhiraja, A., Karandikar, R.L.: Markov property and ergodicity of the nonlinear filter. SIAM J. Control Optim. **39**, 928–949 (2000)
4. Borovkov, A.A.: Ergodicity and Stability of Stochastic Processes. Wiley, New York (1998)
5. Chaleyat-Maurel, M., Michel, D.: Des résultats de non existence de filtre de dimension finie. Stochastics **13**, 83–102 (1984)
6. Clark, J.M.C.: The design of robust approximations to the stochastic differential equations of nonlinear filtering. In: Skwirzynski, J.K. (ed.) Communication Systems and Random Process Theory. NATO Advanced Study Institute Series, pp. 721–734. Sijthoff and Noordhoff, Alphen aan den Rijn (1978)
7. Clark, J.M.C., Crisan, D.: On a robust version of the integral representation formula of nonlinear filtering. Probab. Theory Relat. Fields **133**, 43–56 (2005)
8. Clark, J.M.C., Ocone, D., Coumarbatch, C.: Relative entropy and error bounds for filtering of Markov processes. Math. Control Signals Syst. **12**, 346–360 (1999)
9. Daum, F.E.: Exact finite-dimensional nonlinear filters. IEEE Trans. Automat. Contr. **AC-31**, 616–622 (1986)
10. Davis, M.H.A.: Linear Estimation and Stochastic Control. Chapman and Hall, London (1977)
11. Davis, M.H.A.: A pathwise solution of the equations of nonlinear filtering. Theory Probab. Appl. **27**, 167–175 (1983)

12. Doob, J.L.: Measure Theory. Springer, New York (1994)
13. Dupuis, P., Ellis, R.S.: A Weak Convergence Approach to the Theory of Large Deviations. Wiley, New York (1997)
14. Duncan, T.E.: On the calculation of mutual information. SIAM J. Appl. Math. **19**, 215–220 (1970)
15. Ethier, S.N., Kurtz, T.G.: Markov Processes, Characterization and Convergence. Wiley, New York (1986)
16. Fujisaki, M., Kallianpur, G., Kunita, H.: Stochastic differential equations for the nonlinear filtering problem. Osaka J. Math. **9**, 19–40 (1972)
17. Gaveau, B., Schulman, L.S.: Creation, dissipation and recycling of resources in non-equilibrium systems. J. Stat. Phys. **110**, 1317–1367 (2003)
18. Haussmann, U.G., Pardoux, E.: Time reversal of diffusions. Ann. Probab. **14**, 1188–1205 (1986)
19. Jazwinski, A.H.: Stochastic Processes and Filtering Theory. Academic Press, San Diego (1970)
20. Kallianpur, G., Striebel, C.: Estimation of stochastic systems: arbitrary system process with additive white noise observation errors. Ann. Math. Stat. **39**, 785–801 (1968)
21. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus. Springer, Berlin (1991)
22. Kolmogorov, A.N.: A new metric invariant of transitive dynamical systems and automorphisms in Lebesgue spaces. Dokl. Akad. Nauk SSSR **111**, 861–864 (1958)
23. Kushner, H.J.: On the differential equations satisfied by conditinal probability densities of Markov processes, with applications. SIAM J. Control **2**, 106–119 (1962)
24. Kushner, H.J.: Dynamical equations for non-linear filtering. J. Differ. Equ. **3**, 179–190 (1967)
25. Landauer, R.: Dissipation and heat generation in the computing. IBM J. Res. Develop. **5**, 183–191 (1961)
26. Lebowitz, J.L., Spohn, H.: A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics. J. Stat. Phys. **95**, 333–366 (1999)
27. Liptser, R.S., Shiryayev, A.N.: Statistics of Random Processes 1—General Theory. Springer, Berlin (1977)
28. Maes, C.: The fluctuation theorem as a Gibbs property. J. Stat. Phys. **95**, 367–392 (1999)
29. Marcus, S.I.: Algebraic and geometric methods in nonlinear filtering theory. SIAM J. Control Optim. **22**, 817–844 (1984)
30. Maxwell, J.C.: Theory of Heat. Longmans, London (1871)
31. Mayer-Wolf, E., Zakai, M.: On a formula relating the Shannon information to the Fisher information for the filtering problem. In: Korezlioglu, H., Mazziotto, G., Szpirglas, S. (eds.) Filtering and Control of Random Processes. Lecture Notes in Control and Information Sciences, vol. 61, pp. 164–171. Springer, Berlin (1984)
32. Mitter, S.K.: Geometric theory of nonlinear filtering. Outils Modeles Math. Automat. **3**, 37–55 (1983)
33. Mitter, S.K., Newton, N.J.: A Variational approach to nonlinear estimation. SIAM J. Control Optim. **42**, 1813–1833 (2003)
34. Mitter, S.K., Newton, N.J.: Information and entropy flow in the Kalman-Bucy filter. J. Stat. Phys. **118**, 145–176 (2005)
35. Nelson, E.: The adjoint Markov process. Duke Math. J. **25**, 671–690 (1958)
36. Newton, N.J.: Dual Kalman-Bucy filters and interactive entropy production. SIAM J. Control Optim. **45**, 998–1016 (2006)
37. Newton, N.J.: Dual nonlinear filters and entropy production. SIAM J. Control Optim. **46**, 1637–1663 (2007)
38. Propp, M.B.: The thermodynamic properties of Markov processes. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge (1985)
39. Sinai, Ya.G.: On the concept of entropy for a dynamic system. Dokl. Akad. Nauk SSSR **124**, 768–771 (1959)
40. Stratonovich, R.L.: Conditional Markov processes. Theory Probab. Appl. **5**, 156–178 (1960)
41. van Putten, C., van Schuppen, J.H.: Invariance properties of the conditional independence relation. Ann. Probab. **13**, 934–945 (1985)
42. Wonham, W.M.: Some applications of stochastic differential equations to optimal nonlinear filtering. SIAM J. Control **2**, 347–369 (1965)