Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

ThPI26.11

Linear-Quadratic-Gaussian Heat Engines

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

Abstract— In this paper, we study the problem of extracting work from heat flows. In thermodynamics, a device doing this is called a heat engine. A fundamental problem is to derive hard limits on the efficiency of heat engines. Here we construct a linear-quadratic-Gaussian optimal controller that estimates the states of a heated lossless system. The measurements cool the system, and the surplus energy can be extracted as work by the controller. Hence, the controller acts like a Maxwell's demon. We compute the efficiency of the controller over finite and infinite time intervals, and since the controller is optimal, this yields hard limits. Over infinite time horizons, the controller has the same efficiency as a Carnot heat engine, and thereby it respects the second law of thermodynamics. As illustration we use an electric circuit where an ideal current source extracts energy from resistors with Johnson-Nyquist noise.

I. INTRODUCTION

A central problem in thermodynamics is the study of hard limits on the efficiency of heat engines. Using the laws of thermodynamics, one can derive an upper bound on the efficiency and construct a heat engine, the *Carnot heat engine*, that achieves this bound. The bound is simple, $1 - T_{\text{cold}}/T_{\text{hot}}$, where T_{cold} and T_{hot} are the temperatures of the heat sources the engine can exchange heat with. Carnot heat engines are covered in many books, for example in [1]. The Carnot heat engine is a theoretical construct that operates infinitely slowly and in quasi equilibrium. Hence, the basic theory does not answer how well a heat engine can do over finite time intervals. To study this, we have to put more assumptions on the environment and the engine. See, for example, [2], for a physicist's treatment of the problem.

For a control engineer it is natural to study these problems in a dynamical systems setting. Previous control-theoretic approaches to analysis of efficiency of heat engines, [3], [4], have assumed linear parameter-varying systems and used open-loop control strategies. It has been shown that the Carnot engine efficiency can be achieved in this setting over infinite time horizons. Hard limits for the finite-time case are still unknown, however.

In this paper, we assume linear time-invariant systems, and use measurements and Linear-Quadratic-Gaussian (LQG)

optimal control theory. This let us easily compute the optimal solution, both over finite and infinite time intervals. We see that as the time interval tends to infinity, we achieve the Carnot heat engine efficiency. In the physics literature, a heat engine that uses measurements is often called a *Maxwell's demon*. A similar approach is used in [5]. There, however, the Kalman-Bucy filter is in focus. We also put more assumptions on the models here. Most notably, the assumptions about lossless and dissipative models. This yields simple expressions with clear interpretation. Also, one can argue that physical systems are best modeled using lossless and dissipative models, see [6], [7], and also [8]. Other control-theoretic approaches to various problems of thermodynamics are found in [9]–[11].

The hard limits we obtain are dependent on the idealizations we make. Here lies a trade-off: Too simple models and the hard limits are trivial. Too complex models and no analytical treatment is possible. The models we use here are highly idealized, yet the hard limits are nontrivial. The idealizations we make include that computation and amplification in the controller and actuator are free. The measurement device is modeled by a linear dissipative system with noise. This captures some back action effects from measurements [8]. An interesting problem for future work is to develop good models of active elements. These should include noise and energy losses. Such active elements can then be used in the models of actuator and the measurement devices, and would give more realistic hard limits.

The organization of the paper is as follows: In Section II, an electric circuit is used to present the problem. In Section III, the model setup is given. It is essentially identical to the one used in [8]. In Sections IV and V, optimal open-loop control and measurement strategies are given, respectively. In Section VI, the optimal closed-loop control strategy is given. Finally, in Section VII, we return to the electric circuit and discuss the meaning of the results.

The optimal-control results we use in this paper are more or less standard, and can be found in references such as [6], [7], [12]–[14].

H. Sandberg is now at Royal Institute of Technology (KTH), School of Electrical Engineering, Automatic Control, Osquldas väg 10, SE-100 44 Stockholm, Sweden. hsan@ee.kth.se. He was before at California Institute of Technology and his postdoctoral work there was supported by grants from the Hans Werthén foundation and the Swedish Research Council.

J.C. Doyle is at California Institute of Technology, Control and Dynamical Systems, M/C 107-81, Pasadena, CA 91125, USA. doyle@cds.caltech.edu

J.-C. Delvenne is at Imperial College, Institute for Mathematical Sciences, 53 Prince's Gate, South Kensington, London, SW7 2PG, UK. jc.delvenne@imperial.ac.uk

II. MOTIVATING EXAMPLE

In Fig. 1, a circuit with a lossless impedance Z, three resistors, and an ideal current source, is shown. We assume that the resistor R_h has temperature T_h, the resistor R_m has temperature T_m, and the resistor R_a has no temperature. The *Johnson-Nyquist noise* [15], [16] in the resistors R_h and R_m will "heat up" the impedance Z. The resistance R_m is the resistance of a *passive* (voltage) measurement device. Its noise will add heat to the circuit and give measurement
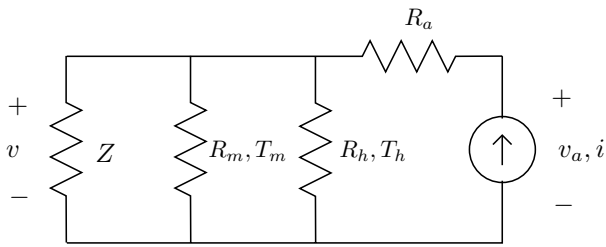
Fig. 1. The circuit that is used as motivating example.

noise. The resistance R_h models losses of energy to, and noise from, the environment and the current source (these are lumped together here). The resistance R_a is introduced to model energy losses that occur when the current i is applied.

The noisy resistors will provide uncertain energy ("heat energy") to the states of Z. The problem addressed in this paper is the following: Given noisy voltage measurement of v, how should we choose the current i so as to maximize the extracted (expected) amount of energy $-\int_0^t v_a(s)i(s)ds$? When this problem is solved, we have constructed a heat engine that transfers heat energy from the resistors into "useful energy" in the current source. We return to and solve this example in Section VII.

III. PRELIMINARIES

Consider linear systems in the form

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

where $J = -J^T$ (anti-symmetric) and (J, B) is controllable. It is assumed that the input $u(t)$ and the output $y(t)$ are scalars (single-input–single-output [SISO] system). We define the energy of (1) as $E(t) = \frac{1}{2}x(t)^T x(t)$. The system (1) is *lossless* [6], [7], [14] which means that the rate of change of energy satisfies $\dot{E}(t) = x(t)^T \dot{x}(t) = y(t)u(t) =: w(t)$, where $w(t)$ is the *work rate*. All lossless (minimal) linear SISO systems with work rate $w(t)$ can be written in the form (1), as seen in [7, Theorem 5]. In the following, all lossless systems are assumed to take the form (1). The impedances Z we consider are assumed to be lossless.

Assume now that the initial state $x(0)$ of a lossless system has a Gaussian distribution with mean $\bar{x}_0 = \mathbf{E}x(0)$ and covariance $X_0 = \mathbf{E}[x(0) - \bar{x}_0][x(0) - \bar{x}_0]^T$. Assuming a deterministic input $u(t)$, the mean and the covariance of $x(t)$ propagate as [12]

$$\dot{\bar{x}}(t) = J\bar{x}(t) + Bu(t), \quad \bar{x}(t_0) = \bar{x}_0, \quad \bar{y}(t) = B^T \bar{x}(t),$$
$$\dot{X}(t) = JX(t) + X(t)J^T, \ X(t_0) = X_0, \ X(t) \in \mathbb{R}^{n \times n}.$$

The *expected* energy of the system is given by

$$\bar{E}(t) := \mathbf{E}E(t) = E_o(t) + U(t),$$
$$E_o(t) = \frac{1}{2}\bar{x}(t)^T \bar{x}(t), \quad U(t) = \frac{1}{2}\text{Tr} X(t).$$

We call $U(t)$ the *internal energy* of the system and $E_o(t)$ the *organized energy* of the system.

Remark 1: In [4], the term "mechanical energy" is used instead of the term organized energy. By the latter we also

mean to use electrical energy, such as the one stored in a capacitor. One can think of organized energy as the macroscopic kinetic and potential energy. If there is no uncertainty, all energy is organized energy. The internal energy contains the energy related to the (microscopic) fluctuations around the mean state.

When a deterministic input $u(t)$ is applied, the variance $X(t)$ (and thus the internal energy $U(t)$) is not affected, and we have the rate of change of expected energy

$$\frac{d\bar{E}(t)}{dt} = \dot{E}_o(t) + \underbrace{\dot{U}(t)}_{=0} = \bar{y}(t)u(t) = \mathbf{E}y(t)u(t) =: \bar{w}(t),$$

for (1), where $\bar{w}(t)$ is the expected work rate on the system.

A stochastic input does change the internal energy. Stochastic inputs are provided by heat baths in this paper. A heat bath is defined as follows.

Definition 1 (Heat bath [8]): Let

$$y(t) = ku(t) - \sqrt{2kT}n(t), \quad (2)$$

where $n(t)$ is Gaussian white noise with covariance function $\mathbf{E}n(t)n(s) = \delta(t - s)$. Then (2) is an input-output model of a *heat bath* with admittance $k > 0$ and temperature $T > 0$.

A heat bath can be realized by a resistor with Johnson-Nyquist noise [15], [16].

A *physical interconnection* [8] of a heat bath (2) and a lossless system (1) has the dynamics

$$\dot{x}(t) = J_k x(t) + Bu(t) + B\sqrt{2kT}n(t), \quad (3)$$

$J_k := J - kBB^T$. The system (3) can be a model of a lossless impedance Z connected to a resistor $R = 1/k$ of temperature T, for example. The interconnected system (3) is a *Langevin equation*, and is asymptotically stable.

Proposition 1: The matrix $J - kBB^T$ is Hurwitz.

The mean and variance for the system (3) propagate as

$$\dot{\bar{x}}(t) = J_k\bar{x}(t) + Bu(t), \quad \bar{x}(t_0) = \bar{x}_0, \quad \bar{y}(t) = B^T\bar{x}(t),$$
$$\dot{X}(t) = J_k X(t) + X(t)J_k^T + 2kTBB^T, \quad X(t_0) = X_0. \quad (4)$$

The rate of change of organized energy for (3) is given by $\dot{E}_o(t) = \bar{w}(t) - d(t)$, $d(t) = \frac{1}{2}k\bar{y}(t)^T\bar{y}(t)$, where $d(t)$ is the spontaneous dissipation rate of organized energy into the heat bath. The rate of change of internal energy is $\dot{U}(t) = q_{\text{in}} - q_{\text{out}}(t)$, where $q_{\text{in}} = kT \text{Tr}[BB^T]$ and $q_{\text{out}}(t) = k\text{Tr}[BB^T X(t)]$. The terms q_{in} and $q_{\text{out}}(t)$ are the inflow and outflow of heat to the lossless system. The heat bath increases or decreases the internal energy $U(t)$ depending on how large the covariance $X(t)$ is.

Remark 2 ("$\dot{U} = w + q$"?): Notice that work and heat are stored in two different quantities, E_o and U. The reason is that we have restricted ourselves to linear time-invariant models. In [3], [4], both work and heat enter in the rate of change of internal energy. Work is then extracted from the internal energy by periodically varying the matrices J_k and B over time. The way we extract work from internal energy here is by measurements on the system. Thereby we increase the amount of organized energy, see Section V. In physics, a device doing this is called a Maxwell's demon.

The lossless system naturally strives towards thermal equilibrium with the heat bath.

Proposition 2: The solution $X(t)$ of (4) satisfies $X(t) \rightarrow T \cdot I_n$, $U(t) \rightarrow \frac{1}{2}nT$, as $t \rightarrow \infty$, where I_n is the $n \times n$ identity matrix.

This means that in thermal equilibrium, there is *energy equipartition*: All state elements of the lossless system have an expected energy of $T/2$. In equilibrium statistical mechanics, energy equipartition is instead often shown using the Maxwell-Boltzmann distribution, see, for example, [1].

IV. EXTRACTION OF WORK IN OPEN LOOP

In this section, we consider the problem of extracting as much work as possible from the system (3). The input $u(t)$ is generated by an actuator device. An actuator with energy losses is defined next.

Definition 2: A *lossy actuator* with admittance $k_a > 0$ provides work to its environment at the rate $w_a(t) = w(t) + u(t)^2/k_a$.

Hence, for the actuator to provide the work rate $w(t)$ to the system, its work rate must be $w_a(t) \geq w(t)$. A lossy actuator can be realized by a current source in series with a resistor $R_a = 1/k_a$.

Let

$$\bar{W}_{\bar{x}}(t) := \max_{u(\bar{x})} \mathbf{E}W_a(t) = -\min_{u(\bar{x})} \mathbf{E} \int_0^t w_a(s)ds$$

denote the maximum of the expected work extracted from (3) using a lossy actuator, with respect to all control strategies where the input $u(t)$ is based on the mean state trajectory $\bar{x}(t)$. The input $u(t)$ is deterministic, since it does not depend on the stochastic process $n(t)$. Hence, $\mathbf{E}w(t) = \bar{y}(t)^T u(t) = \bar{x}(t)^T Bu(t)$.

The optimal control problem can be formulated as

$$\max_{u(\bar{x})} \mathbf{E}W_a(t) = -\min_u \int_0^t \bar{x}(s)^T Bu(s) + u(s)^2/k_a ds \quad (5)$$

subject to $\dot{\bar{x}}(t) = J_k\bar{x}(t) + Bu(t)$, $\bar{x}(0) = \bar{x}_0$.

This is a linear-quadratic optimal control problem with a well-known solution. See, for example, [14, Section 5.4] where similar equations are used to prove the necessity part of the positive real lemma.

Theorem 1: The work-extraction problem (5) has the value

$$\bar{W}_{\bar{x}}(t) = \frac{1}{2}\bar{x}_0^T R_t(0)\bar{x}_0,$$

and $\bar{W}_{\bar{x}}(t)$ is an increasing function of t. The matrix-valued function $R_t(s)$ satisfies the differential Riccati equation

$$-\dot{R}_t(s) = R_t(s)J_k + J_k^T R_t(s)$$
$$+ \frac{k_a}{2}(R_t(s) - I_n)BB^T(R_t(s) - I_n), \quad R_t(t) = 0,$$

$R_t(s) \in \mathbb{R}^{n \times n}$. Furthermore, the optimal input is given by the control strategy $u(s) = -L(s)\bar{x}(s)$, where $L(s) = \frac{k_a}{2}B^T(I_n - R_t(s))$.

Remark 3: Notice that the optimal control is an open-loop strategy since it can be computed before it is applied to (3).
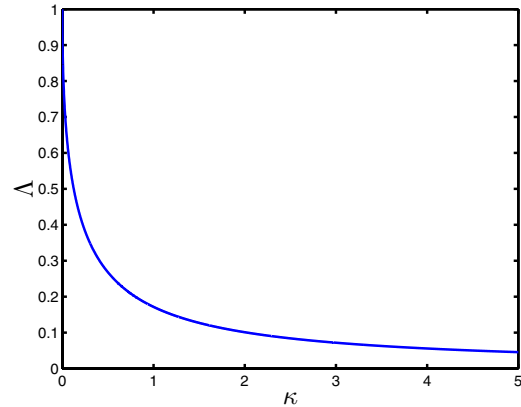

Fig. 2. The actuator efficiency Λ plotted as a function of $\kappa = k/k_a$.

As the time horizon t increases, we can extract more work from the system. This follows since $R_t(s)$ is a decreasing function of s, and the dynamics is time invariant. Hence, it is of interest to study the limit $R_\infty = \lim_{t \rightarrow \infty} R_t(s)$, for fixed s. The limit R_∞ is independent of s and is the stabilizing solution to the algebraic Riccati equation

$$0 = R_\infty J_k + J_k^T R_\infty + \frac{k_a}{2}(R_\infty - I_n)BB^T(R_\infty - I_n),$$

see [14, Lemma 5.4] and [7, Lemma 2]. The solution R_∞ has the simple diagonal structure

$$R_\infty = \Lambda(\kappa) \cdot I_n, \quad \Lambda(\kappa) = 1 + 2\kappa - 2\sqrt{\kappa(1+\kappa)}, \quad (6)$$

where $\kappa := k/k_a$ is the ratio between the admittances of the heat bath and the actuator.

Definition 3: The scalar quantity $\Lambda(\kappa)$ is called the *actuator efficiency*.

The strictly decreasing function $\Lambda(\kappa)$ is plotted in Fig. 2. We have that $\Lambda(\kappa) \rightarrow 1$, $\kappa \rightarrow 0$, and $\Lambda(\kappa) \rightarrow 0$, $\kappa \rightarrow \infty$. Using the structure of R_∞ and the increasing property of $\bar{W}_{\bar{x}}(t)$, the following corollary to Theorem 1 can be derived.

Corollary 1: It holds that $\bar{W}_{\bar{x}}(t) \leq E_o(0)$ for all t, and $\lim_{t \rightarrow \infty} \bar{W}_{\bar{x}}(t) = E_o(0)\Lambda(\kappa)$, where $E_o(0) = \frac{1}{2}\bar{x}_0^T\bar{x}_0$ is the initial organized energy.

Thus, when the losses in the actuator are small ($k_a \gg k$), we can extract almost all the initial organized energy using the optimal open-loop strategy in Theorem 1. When the admittance of the heat bath is large ($k \gg k_a$), then almost all initial organized energy dissipates into the heat bath. Notice that the actuator efficiency decays quickly with increasing κ.

Remark 4: Notice that in the infinite-time limit the influence of the dynamics (J, B) on the efficiency disappears. This shows that the framework agrees with regular thermodynamics where dynamics usually is not considered. There changes are often assumed to be done infinitely slowly and in quasi equilibrium.

Example 1: Consider the circuit in Fig. 1. Assume the lossless system is a capacitor of capacitance C ($J = 0$ and $B = 1/\sqrt{C}$), $k_a = 1/R_a$, and $k = 1/R_h + 1/R_m$. Define

$\Lambda_+ = 1 + 2\kappa + 2\sqrt{\kappa(1+\kappa)}$ and $\lambda = \sqrt{k(k+k_a)}/C$. Then

$$R_t(s) = \Lambda \frac{1 - \exp 2\lambda(s-t)}{1 - (\Lambda/\Lambda_+)\exp 2\lambda(s-t)},$$

and it is seen that $R_t(s)$ converges exponentially fast towards Λ as the time horizon t tends to infinity.

V. MEASUREMENTS AND COOLING

In this section, we introduce measurements and see how they can be used to increase the efficiency of the open-loop control strategy in Section IV. A measurement device is modeled as follows.

Definition 4: A *measurement device* with admittance $k_m > 0$ and temperature $T_m > 0$ gives the measurement signal $y_m(t) = y(t) - \sqrt{2T_m/k_m}m(t)$ where $m(t)$ is Gaussian white noise with covariance function $\mathbf{E}m(t)m(s) = \delta(t-s)$.

The physical interconnection [8] of the lossless system, a heat bath with admittance k_h and temperature T_h, and the measurement device is given by

$$\dot{x}(t) = J_{k_h+k_m}x(t) + Bu(t)$$
$$+ B\sqrt{2k_hT_h}n(t) + B\sqrt{2k_mT_m}m(t), \quad (7)$$
$$y_m(t) = B^T x(t) - \sqrt{2T_m/k_m}m(t). \quad (8)$$

The Gaussian noise $n(t)$ and $m(t)$ are assumed uncorrelated. Using the substitutions

$$k = k_h + k_m, \quad T = \frac{k_h T_h + k_m T_m}{k_h + k_m}, \quad (9)$$

we see that (7) can be written in the form (3). Hence, the results in the previous sections remain true for (7). The measurement device acts as a second independent heat bath. Notice, however, that the noise $m(t)$ enters both as process noise and measurement noise. This is called "measurement back action" in [8].

Let

$$Y_t = \{y_m(s) : t_0 \le s \le t\}$$

denote a measurement sequence. With some abuse of notation, we sometimes also let Y_t denote a stochastic variable that takes its values in the set of all possible measurement sequences. Let $\mathbf{E}[f|Y_t]$ denote the conditional mean of f given a particular measurement sequence Y_t. The law of total expectation gives that $\mathbf{E}_{Y_t}\mathbf{E}[f|Y_t] = \mathbf{E}f$.

Define the estimate $x_m(t)$ as the conditional expected value of the state $x(t)$, $x_m(t) = \mathbf{E}[x(t)|Y_t]$. As is well known, see, for example [12], [13], the estimate $x_m(t)$ has many good properties. We have the following identities

$$\mathbf{E}_{Y_t}x_m(t) = \mathbf{E}x(t) = \bar{x}(t),$$
$$\mathbf{E}_{x,Y_t}[x_m(t) - x(t)]x_m(t)^T = 0, \quad (10)$$
$$\mathbf{E}\{[x_m(t) - x(t)]x_m(t)^T|Y_t\} = 0,$$

and

$$X_m(t) := \mathbf{E}_{x,Y_t}[x_m(t) - x(t)][x_m(t) - x(t)]^T$$
$$= \mathbf{E}\{[x_m(t) - x(t)][x_m(t) - x(t)]^T|Y_t\} \quad (11)$$

is both the unconditional and conditional estimation error covariance (since the system is linear and the noise is

Gaussian). Given an estimate $x_m(t)$, we can redistribute the amounts of internal energy $U(t)$ and organized energy $E_o(t)$. Using the identities (10) and (11), we have the conditional and unconditional system energies

$$\mathbf{E}[E(t)|Y_t] = \frac{1}{2}x_m(t)^T x_m(t) + U_m(t),$$
$$\bar{E}(t) = \mathbf{E}E(t) = \mathbf{E}_{Y_t}\mathbf{E}[E(t)|Y_t] = E_{o,m}(t) + U_m(t),$$

where

$$E_{o,m}(t) := \frac{1}{2}\mathbf{E}_{Y_t}x_m(t)^T x_m(t) = E_o(t) + U(t) - U_m(t),$$
$$U_m(t) := \frac{1}{2}\text{Tr}X_m(t).$$

It follows that $U_m(t) \le U(t)$ and $E_{o,m}(t) \ge E_o(t)$. The terms $E_{o,m}$ and U_m are the organized energy and internal energy, respectively, given the measurements.

Assume now that $t_0 < 0$, and that we have access to the measurement sequence Y_0. We now want to choose a control signal u over the time interval $[0, t]$ so as to maximize $\mathbf{E}W_a(t)$, just as in Section IV. The only difference is that we have access to more information about the system. Denote by $\max_{u(Y_0)}\mathbf{E}W_a(t)$ the maximum of the expected work extracted with respect to all control strategies based on Y_0. We have the following update to Theorem 1 and Corollary 1.

Proposition 3: It holds that

$$\bar{W}_{Y_0}(t) := \max_{u(Y_0)}\mathbf{E}W_a(t) = \frac{1}{2}x_m(0)^T R_t(0)x_m(0),$$

and

$$\mathbf{E}_{Y_0}\bar{W}_{Y_0}(t) \le E_{o,m}(0), \quad \mathbf{E}_{Y_0}\lim_{t\to\infty}\bar{W}_{Y_0}(t) = E_{o,m}(0)\Lambda(\kappa).$$

Proposition 3 shows that measurements Y_0 increase the amount of work we can extract, and that it is the organized energy $E_{o,m}(0)$ instead of $E_o(0)$ that gives the upper bound.

As is well known, see, for example [12], the Kalman-Bucy filter gives the estimate $x_m(t)$. The Kalman-Bucy filter for (7)–(8) is given in the following theorem.

Theorem 2: The estimate $x_m(t)$ and the error covariance $X_m(t)$ satisfy

$$\dot{x}_m(t) = J_k x_m(t) + Bu(t) + K_m(t)(y_m(t) - B^T x_m(t)),$$
$$x_m(t_0) = \bar{x}_0,$$

and $\dot{X}_m(t) = J_k X_m(t) + X_m(t)J_k^T + 2kTBB^T - \frac{k_m}{2T_m}(X_m(t) - 2T_mI_n)BB^T(X_m(t) - 2T_mI_n)^T$, where $X_m(t_0) = X_0$, and $K_m(t) = \frac{k_m}{2T_m}(X_m(t) - 2T_mI_n)B$.

We also have the following corollary, which should be compared to Proposition 2.

Corollary 2: The estimation error covariance $X_m(t)$ satisfies $X_m(t) \to T_M \cdot I_n$ as $t \to \infty$, where the scalar T_M is given by

$$T_M = -2\kappa_m T_m + 2\sqrt{\kappa_m T_m(\kappa_m T_m + T_h)} \le T, \quad (12)$$

and $\kappa_m := k_h/k_m$ is the ratio of the admittances of the heat bath and the measurement device.

If we by temperature mean the steady-state value of the uncertain amount of energy per degree of freedom, then it makes sense to call T_M a temperature of the system. At least this is true for a person who has access to the estimate $x_m(t)$. Since $T_M \leq T$, the system can be said to be cooled by the measurements. We make the following definition.

Definition 5: The temperature T is called the *physical temperature* and T_M the *measured temperature* of the system.

VI. EXTRACTION OF WORK IN CLOSED LOOP

A disadvantage with the open-loop method for work extraction that is suggested in Sections IV and V is that it is only initial organized energy $E_o(0)$ or $E_{o,m}(0)$ that is extracted. The reason is that no measurements after time 0 are used. Therefore, no work can be extracted from the heat inflow after time 0. In this section, we use measurements to continuously transfer heat inflow to work. Using the well-known *separation theorem*, see for example [12, Theorem 7.1], the optimal solution is given by combining the work extraction method in Section IV and the measurement method in Section V.

Denote by $\max_{u(Y_t)} \mathbf{E} W_a(t)$ the maximum of the expected work extracted with respect to all control strategies based on Y_t (put $t_0 = 0$). The separation theorem says the maximum is achieved using the feedback

$$u(t) = -L(t)x_m(t), \qquad (13)$$

where $L(t)$ is given in Theorem 1 and $x_m(t)$ in Theorem 2. The work output $\max_{u(Y_t)} \mathbf{E} W_a(t)$ can be calculated explicitly by applying Theorem 7.1 in [12]. We obtain the following theorem.

Theorem 3: The feedback (13) achieves

$$\bar{W}_{Y_t}(t) := \max_{u(Y_t)} \mathbf{E} W_a(t) = \frac{1}{2}\bar{x}_0^T R_t(0)\bar{x}_0 + \frac{1}{2}\text{Tr}[R_t(0)X_0]$$

$$+ \int_0^t kT \cdot \text{Tr}[R_t(s)BB^T] - \text{Tr}[L(s)^T L(s)X_m(s)]/k_a ds, \qquad (14)$$

where $R_t(s)$, $L(s)$, and $X_m(s)$ are given in Theorems 1 and 2, and the physical temperature T and admittance k are given in (9).

The first term on the right-hand side of (14) is equal to $\bar{W}_{\bar{x}}(t)$, see Theorem 1. The second term shows that by measurements we can retrieve some work from the initial internal energy $U(0) = \frac{1}{2}\text{Tr}X_0$, since $R_t(0) \leq \Lambda \cdot I_n \leq I_n$. The third term shows how much of the heat inflow $q_{\text{in}} = kT \cdot \text{Tr}[BB^T]$ over the time interval $[0,t]$ we can extract as work. Notice that $R_t(s) \leq \Lambda \cdot I_n \leq I_n$ appears as a time-varying weight. The sizes of the initial three terms depend on how much losses there are in the actuator ($R_t(s)$). These three terms represent how much work that can be extracted from the system with perfect state estimation, $x_m(t) = x(t)$. The fourth term is negative, and its size depends on the measurements. The smaller the estimation error $X_m(t)$ is, the more work we get out.

It is illustrating to let the time horizon tend to infinity. Since there is a continuous supply of heat q_{in} from the heat bath, we have to normalize with the horizon length to obtain finite quantities.

Theorem 4: Define the expected work extraction rate by $w_{\text{out}} = \lim_{t \to \infty} \frac{1}{t}\bar{W}_{Y_t}(t)$. Then

$$\frac{w_{\text{out}}}{q_{\text{in}}} = \left(1 - \frac{T_M}{T}\right) \cdot \Lambda(\kappa), \qquad (15)$$

and therefore $w_{\text{out}}/q_{\text{in}} \to 1 - T_M/T$, as $\kappa \to 0$, where T is the physical temperature, T_M the measured temperature (12), and Λ the actuator efficiency.

Proof: First notice the identity

$$\frac{(1 - \Lambda(\kappa))^2}{4\kappa} = \Lambda(\kappa). \qquad (16)$$

Using Theorem 3, (16), and the limit expressions $\lim_{t \to \infty} R_t(s) = \Lambda(\kappa) \cdot I_n$ and $\lim_{s \to \infty} X_m(s) = T_M \cdot I_n$, we have $w_{\text{out}} = [kT\Lambda(\kappa) - k_a T_M(1 - \Lambda(\kappa))^2/4]\text{Tr}[BB^T] = k(T - T_M)\Lambda(\kappa)\text{Tr}[BB^T]$. From Section III we have $q_{\text{in}} = kT\text{Tr}[BB^T]$, and the result follows. ∎

For the transient behavior of the closed-loop controlled system ("the LQG heat engine") in Theorem 3, the dynamics (J, B) of the lossless system is important, but in the infinite-time limit it does not matter for the efficiency, as seen in (15).

If we interpret the measured temperature T_M as the temperature of a heat sink, and T as the temperature of a heat source, then (15) is identical to the Carnot engine efficiency [1], up to the actuator efficiency factor $\Lambda(\kappa)$. According to Carnot's theorem, no heat engine can exceed the Carnot engine efficiency. Thus the LQG heat engine achieves this upper bound when the actuation loss tends to zero (using proper temperature interpretation).

The Carnot heat engines in [3], [4], also achieve the Carnot efficiency, but these engines operate infinitely slowly and use quasi-equilibrium arguments. These engines use open-loop control and two physical heat sources of different temperatures. The LQG heat engine is really only connected to one physical heat source (of temperature T). In [3], [4], it is shown that such a heat engine cannot continuously deliver work using open-loop control strategies ($w_{\text{out}} = 0$). In fact, we have also seen that here in Theorem 1, since $\lim_{t \to \infty} \frac{1}{t}\bar{W}_{\bar{x}}(t) = 0$. However, using closed-loop control strategies, we have shown that work can be delivered continuously ($w_{\text{out}} > 0$, Theorem 4). The LQG heat engine works like a Maxwell's demon.

VII. MOTIVATING EXAMPLE REVISITED

In this section, we use the LQG heat engine for the circuit example in Section II, see Fig. 1. Z is the lossless system, for example a capacitor, and $k_h = 1/R_h$, $k_a = 1/R_a$, and $k_m = 1/R_m$. The current source is the actuator with input $u(t) = i(t)$. The work rate of the current source is $w_a(t) = v_a(t)i(t) = v(t)i(t) + i(t)^2 R_a$, compare with Definition 2. The voltage over Z, the output $y(t) = v(t)$, is applied to the voltage measurement device R_m. The measurement signal is
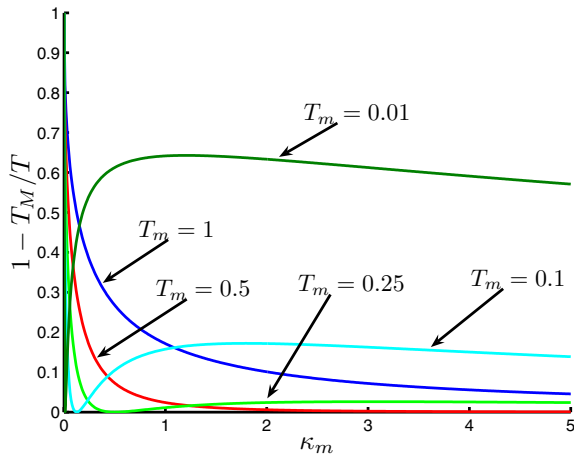
Fig. 3.   The LQG heat engine steady-state efficiency $1 - T_M/T$ plotted as function of $\kappa_m = k_h/k_m$, for five different temperatures T_m of the measurement device, when $T_h = 1.0$.

$y_m(t) = v(t) + \sqrt{2T_m R_m}\, m(t)$, where T_m is the temperature of the resistor R_m.

In Fig. 3, the steady-state efficiency $1 - T_M/T$ (assume no actuation losses, $R_a = 0$ and $\Lambda = 1$) of the engine is plotted as a function of $\kappa_m = k_h/k_m = R_m/R_h$, for five different temperatures T_m of the measurement device. For $T_m = 1$ and 0.5, the curves are monotonically decreasing functions of κ_m. For $T_m = 0.25$, 0.1, and 0.01, the curves decrease until they become zero, and then increase again until they reach a maximum. Based on these two different behaviors, we say that the LQG heat engine can operate in two different modes. We call the first mode the "back action" mode and the second mode the "classical" mode.

In the back action mode, it is the measurement device that is the main provider of energy. This mode works for all temperatures T_m, but best for high T_m and small κ_m. For $2T_m > T_h$, it is the only mode of operation, since the $1 - T_M/T$ curves are then monotonically decreasing. The reason for this threshold is the measurement/process noise cross-covariance term $X_M - 2T_m I_n$ in Theorem 2. In Fig. 3, it is seen that as T_m drops from 1.0 to 0.01, a smaller κ_m is needed to achieve the same efficiency. Hence, when T_m is small, we essentially need to disconnect the system from the other heat source (R_h large) for the engine to work in this mode. The back action mode works since the measurements contain a lot of information about the process noise.

When $\kappa_m \approx 1$ and T_m is small ($2T_m < T_h$), the relative amount of heat coming from the measurement device is small. Yet we can extract energy using the classical mode. Based on the voltage measurements we can estimate the state of Z well and extract the energy provided by R_h. This effect can be seen in Fig. 3 for $T_m = 0.25$, $T_m = 0.1$, and $T_m = 0.01$. The efficiency increases with κ_m up to a maximum that lies around $\kappa_m \approx 1$. The classical mode only works better with decreasing T_m, since the state estimate keeps getting better.

VIII. Conclusions

We formulated the problem of extracting work from a heat flow in a linear systems setting, and solved it using LQG optimal control methods. This gave hard limits on how much work that can be extracted, both over finite and infinite time intervals. Up to the actuator efficiency factor, the LQG heat engine achieves the Carnot heat engine efficiency, using proper temperature definitions. Hence, the heat engine respects the second law of thermodynamics. We applied the results to an electric circuit, and identified two modes of operation where a current source can extract energy from warm resistors.

An interesting problem for future research is how to implement the Kalman-Bucy filter and the optimal controller using physical components such as resistors, inductors, capacitors, and possible active elements. This would lead to additional noise and losses, and other hard limits, as discussed in the introduction. For this, the careful modeling of the Kalman-Bucy filter in [5], and the abstract and undetailed view on how linear systems respect the second law of thermodynamics in [4] may be useful.

References

[1] G. H. Wannier, *Statistical Physics*.   Dover Publications, 1987.
[2] B. Andresen, R. Berry, A. Nitzan, and P. Salamon, "Thermodynamics in finite time. I. The step-Carnot cycle," *Physical Review A*, vol. 15, pp. 2086–2093, 1977.
[3] R. W. Brockett and J. C. Willems, "Stochastic control and the second law of thermodynamics," in *Proceedings of the IEEE Conference on Decision and Control*, San Diego, California, 1978, pp. 1007–1011.
[4] J.-C. Delvenne, H. Sandberg, and J. C. Doyle, "Thermodynamics of linear systems," in *Proceedings of the European Control Conference*, Kos, Greece, July 2007.
[5] S. K. Mitter and N. J. Newton, "Information and entropy flow in the Kalman-Bucy filter," *Journal of Statistical Physics*, vol. 118, pp. 145–176, 2005.
[6] J. C. Willems, "Dissipative dynamical systems part I: General theory," *Archive for Rational Mechanics and Analysis*, vol. 45, pp. 321–351, 1972.
[7] ——, "Dissipative dynamical systems part II: Linear systems with quadratic supply rates," *Archive for Rational Mechanics and Analysis*, vol. 45, pp. 352–393, 1972.
[8] H. Sandberg, J.-C. Delvenne, and J. C. Doyle, "The statistical mechanics of fluctuation-dissipation and measurement back action," in *Proceedings of the American Control Conference*, New York City, New York, July 2007.
[9] W. M. Haddad, V. S. Chellaboina, and S. G. Nersesov, *Thermodynamics: A Dynamical Systems Approach*.   Princeton University Press, 2005.
[10] M. Barahona, A. C. Doherty, M. Sznaier, H. Mabuchi, and J. C. Doyle, "Finite horizon model reduction and the appearance of dissipation in Hamiltonian systems," in *Proceedings of the 41st IEEE Conference on Decision and Control*, vol. 4, 2002, pp. 4563–4568.
[11] D. S. Bernstein and S. P. Bhat, "Energy equipartition and the emergence of damping in lossless systems," in *Proceedings of the 41st IEEE Conference on Decision and Control*, Las Vegas, Nevada, 2002, pp. 2913–2918.
[12] K. J. Åström, *Introduction to Stochastic Control Theory*.   Dover Publications, 2006.
[13] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*.   Dover Publications, 2005.
[14] B. D. O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis: A Modern Systems Theory Approach*.   Dover Publications, 2006.
[15] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical Review*, vol. 32, pp. 97–109, 1928.
[16] H. Nyquist, "Thermal agitation of electrical charge in conductors," *Physical Review*, vol. 32, pp. 110–113, 1928.