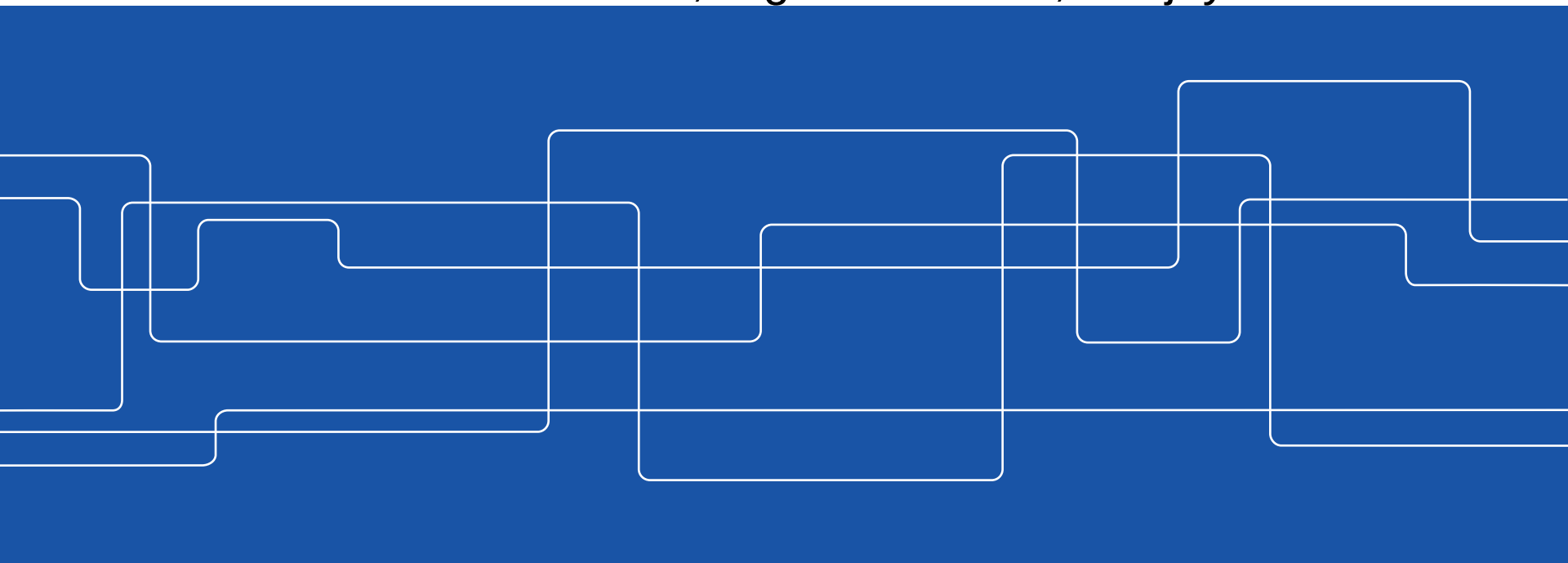
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Security and Limitations of Cyber-Physical Systems

Lecture 4: Implementation Costs and Information Flow in Kalman-Bucy Filters

Henrik Sandberg,
Jean-Charles Delvenne, Nigel J. Newton, Sanjoy K. Mitter



Course Outline

- Monday (8:30-10:30):
  - Lecture 1 (HS): Introduction, data attacks against non-dynamic systems, power network monitoring, security index, graph min cut

- Tuesday
  - 8:30-12:30:
    - Lecture 2 (HS): Attack space for cyber-physical systems: DoS, undetectable, stealth, covert, bias, replay attacks
    - Lecture 3 (AT): Defense mechanisms, risk management, anomaly detectors, watermarking
  - 15:30-16:30:
    - Exercise session (Graph min cut, security index)

- Wednesday (8:30-10:30):
  - **Lecture 4 (HS): Physical limits of control implementations**
  - Exercise session (open discussion, Q & A)

ROYAL INSTITUTE OF TECHNOLOGY

Key References for Lecture

- Henrik Sandberg, Jean-Charles Delvenne, Nigel J. Newton, Sanjoy K. Mitter: "Maximum work extraction and implementation costs for nonequilibrium Maxwell's demons". Physical Review E, 90, 042119, 2014.

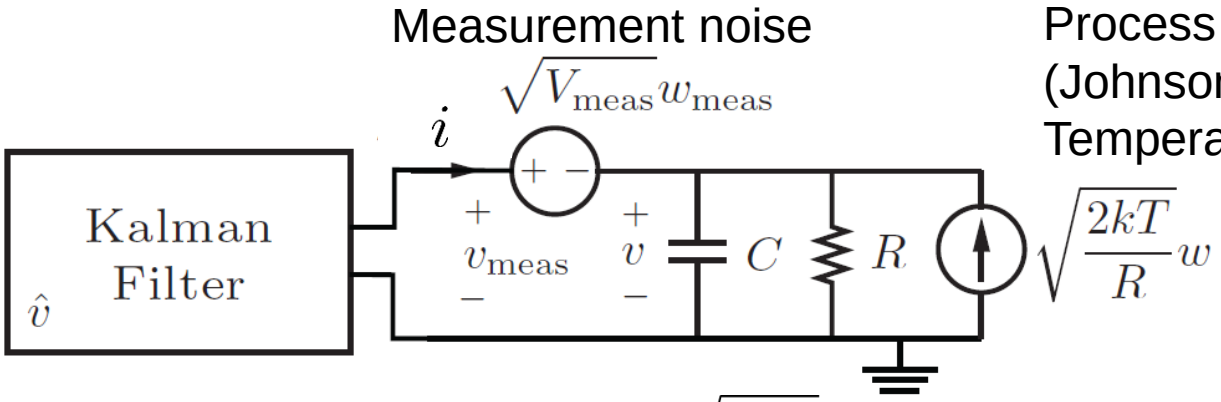
- Henrik Sandberg, Jean-Charles Delvenne, Nigel J. Newton, Sanjoy K. Mitter: "Thermodynamic Costs in Implementing Kalman-Bucy Filters". In Proceedings of the 52nd Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, October 2014.

Outline

- Motivating example
- Thermodynamics of the Kalman-Bucy filter
- Relation to the information flow and Landauer's principle
- (Maxwell's demon)

[Sandberg *et al*., Allerton, 2014]

Motivating Example and Problem Formulation

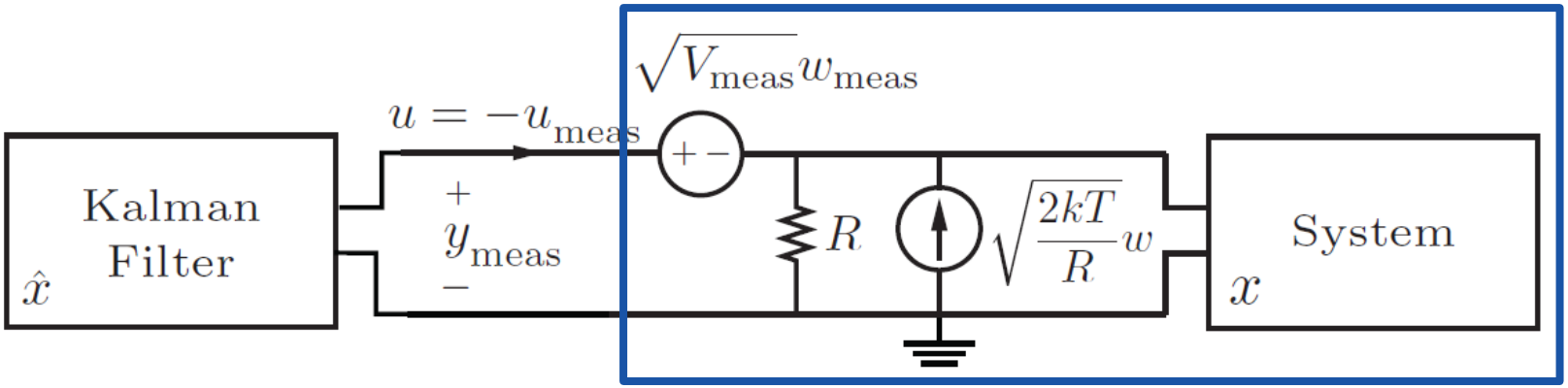


$$\dot{v} = -\frac{1}{RC}v + \frac{1}{R}i + \sqrt{\frac{2kT}{R}}w$$

$$v_{\text{meas}} = v + \sqrt{V_{\text{meas}}}w_{\text{meas}}$$

1. Is there a lower bound on external power supply to a physical implementation of the filter?
2. What is a simple cheap exact physical implementation of the Kalman-Bucy filter?

More General Class of Systems to Measure



Linear passive system (cf. Darlington synthesis)

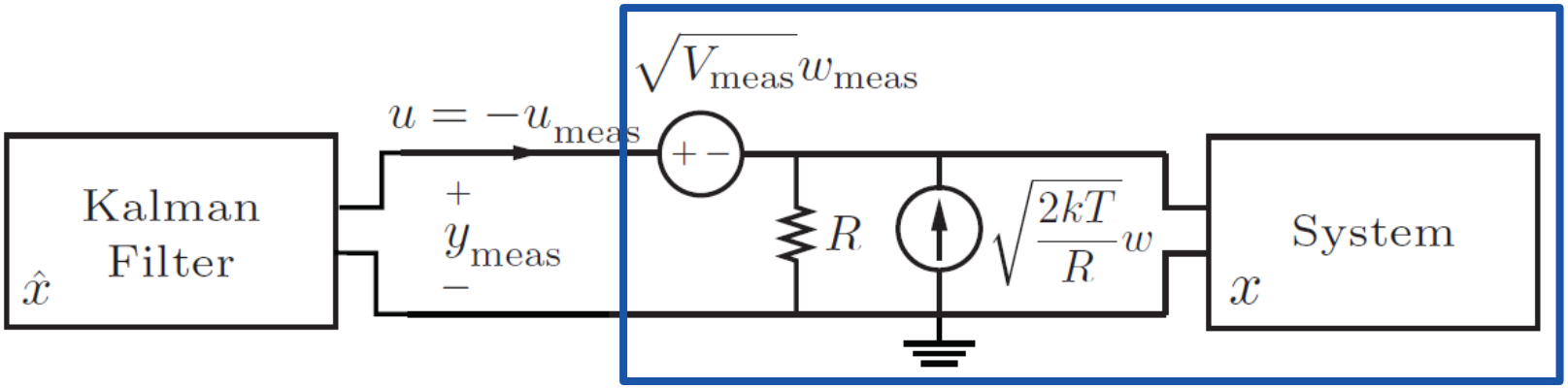
$$\dot{x} = (J - GBB^T)Mx + Bu + B\sqrt{2kTG}w$$

$$y = B^T Mx$$

$$y_{\text{meas}} = B^T Mx + \sqrt{V_{\text{meas}}}w_{\text{meas}}$$

$$(J = -J^T, \quad G := 1/R, \quad M = M^T > 0)$$

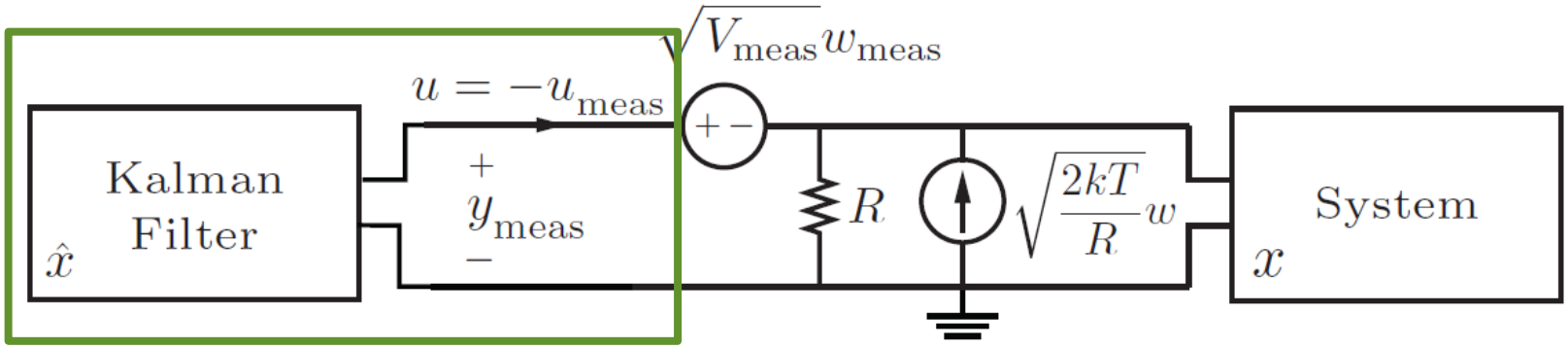
More General Class of Systems to Measure



Signal-to-noise ratio (SNR):

$$\frac{(\text{process noise [V]})^2}{(\text{measurement noise [V]})^2} \equiv \frac{2kT}{GV_{\text{meas}}} =: \sigma$$

Kalman-Bucy Filter



$$\frac{d}{dt}\hat{x} = (J - GBB^T)M\hat{x} + Bu + K(y_{\text{meas}} - B^T M\hat{x})$$

Lemma: Kalman gain is

$$K = (\sqrt{1+\sigma} - 1)GB \equiv g_K B$$

Kalman-Bucy Filter is Passive

Assumption: Admit linear back action/feedback current

$$\underbrace{u_{\text{meas}}}_{\text{current}} = -u = g\underbrace{B^T M\hat{x}}_{\text{voltage}}, \quad g = \text{gain} \in (0,\infty), \text{ free parameter}$$

Theorem: A realization of Kalman-Bucy filter is

$$\frac{d}{dt}\hat{x}_s = (J - ZB_sB_s^T)M\hat{x}_s + B_s y_{\text{meas}}$$

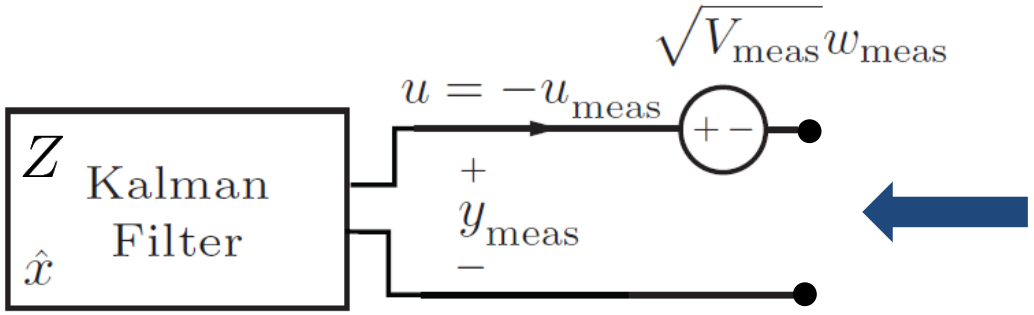
$$u_{\text{meas}} = B_s^T M\hat{x}_s$$

$$\hat{x}_s = \sqrt{g/g_K}\hat{x}$$

with effective resistance

$$Z = \frac{\sqrt{\sigma+1}}{(\sqrt{\sigma+1}-1)g} + \frac{1}{(\sqrt{\sigma+1}-1)G} > 0$$

Effective Temperature of Kalman-Bucy Filter



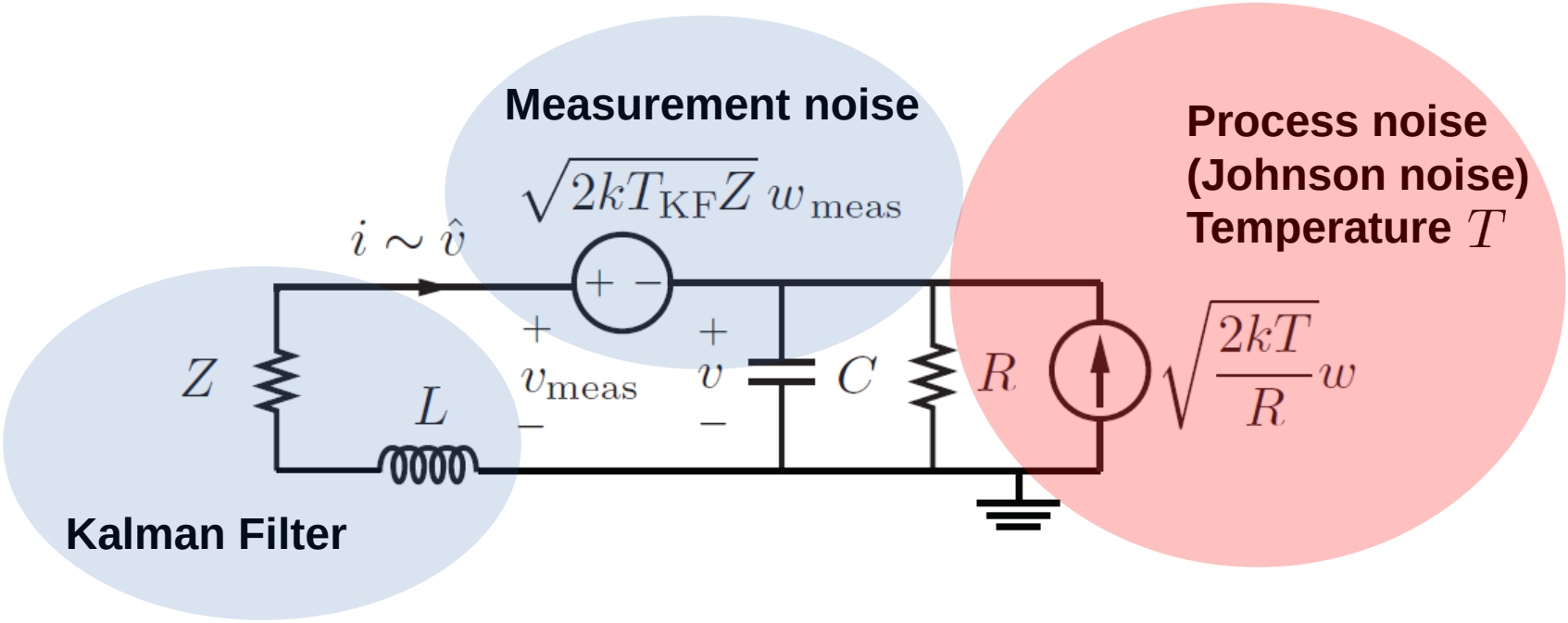
Fluctuation-dissipation theorem: $2kT_{\text{KF}}Z = V_{\text{meas}}$

Effective filter temperature:

$$T_{\text{KF}} = \frac{T}{\sqrt{1+\sigma}+1}\frac{g}{g+g_K+G} < \frac{1}{2}T$$

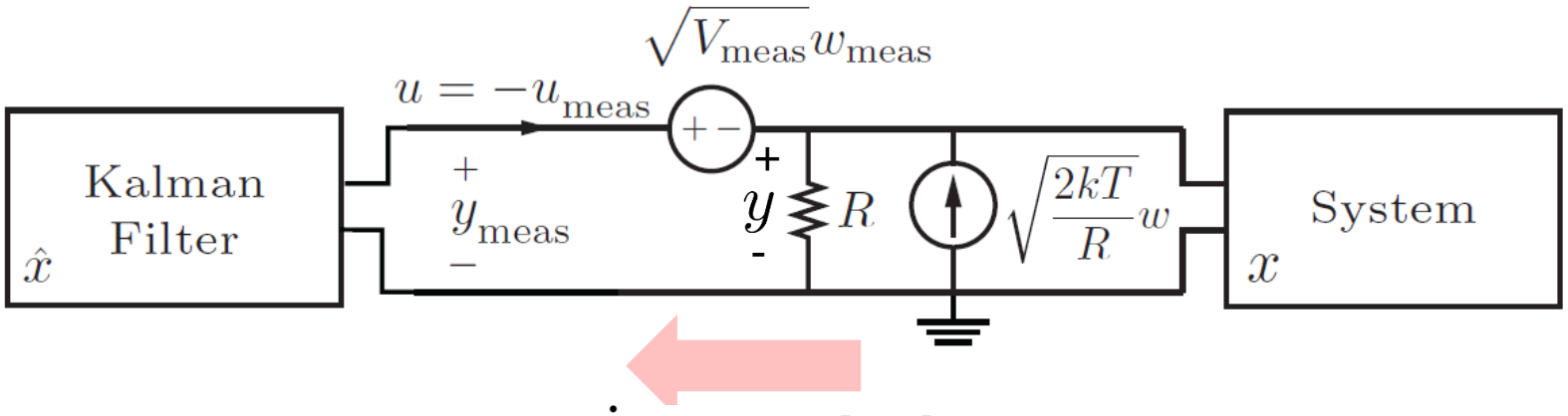
Insight: As back action/feedback tends to zero, filter temperature tends to zero

Exact Filter Circuit for Motivating Example



- Externally equivalent implementation using passive non-zero temperature components
- $T_{KF} < \frac{1}{2}T$ (non-equilibrium thermodynamic system)

Heat Flow



$$\dot{Q} := -\mathbf{E}[uy]$$

Expected stationary energy flow from system to filter:

$$\dot{Q} := -\mathbf{E}[uy] = \frac{g}{g+G}\dot{Q}_{\max}$$

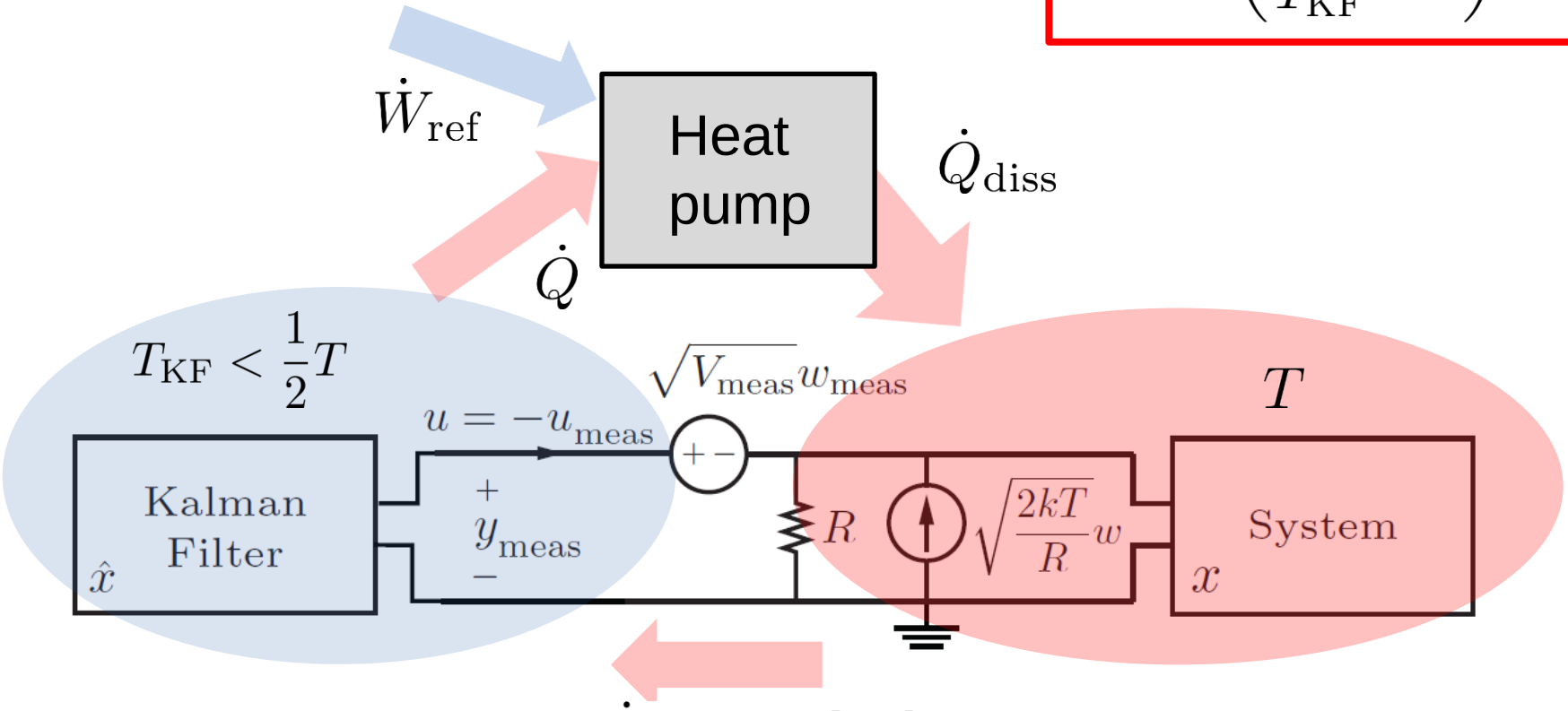
Maximum possible energy flow ($g \to \infty$):

$$\dot{Q}_{\max} := kGB^T MBT \frac{\sqrt{1+\sigma}-1}{\sqrt{1+\sigma}+1}$$

Heat and Work Flows

2nd Law (Carnot):

$$\dot{W}_{\mathrm{ref}} \geq \left(\frac{T}{T_{\mathrm{KF}}} - 1 \right) \dot{Q}$$



$$\dot{Q} := -\mathbf{E}[uy]$$

(Back action heat flow)

Power Supply Required by the 2nd Law

$$\dot{W}_{\mathrm{ref}} \geq \kappa T(1+\sigma-\sqrt{1+\sigma})\left(1-\frac{\sqrt{1+\sigma}}{1+\sqrt{1+\sigma}}\mathrm{BA}\right)$$

Heat conductivity of system

$$kGB^TMB$$

Normalized back action

$$\frac{g}{g+G}$$

- Kalman-Bucy filter with no feedback: $\mathrm{BA} \to 0$
- Minimum-variance controller: $\mathrm{BA} \to 1$

Power Supply Required by the 2nd Law

$$\dot{W}_{\mathrm{ref}} \geq \kappa T(1+\sigma-\sqrt{1+\sigma})\left(1-\frac{\sqrt{1+\sigma}}{1+\sqrt{1+\sigma}}\mathrm{BA}\right)$$

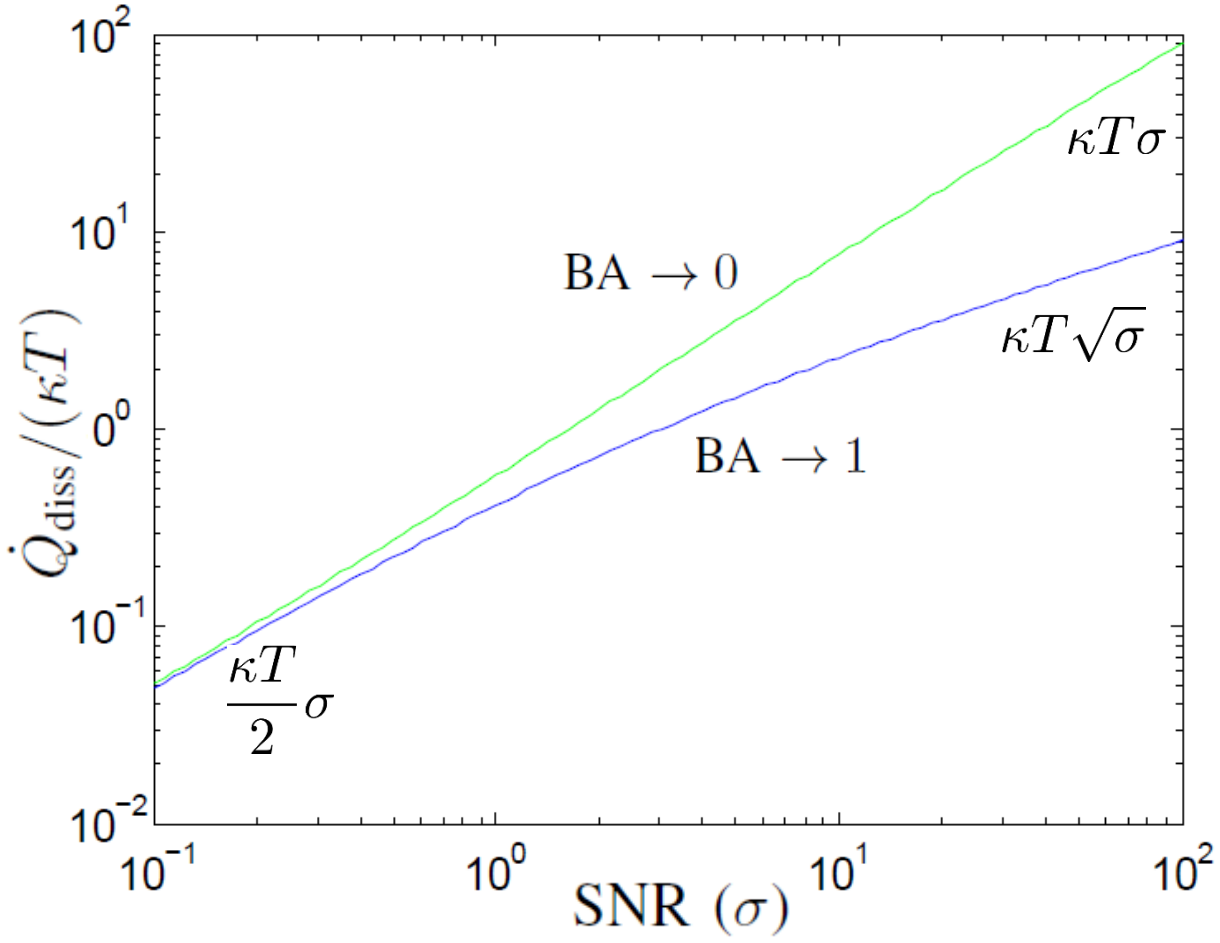
No feedback (BA→0)

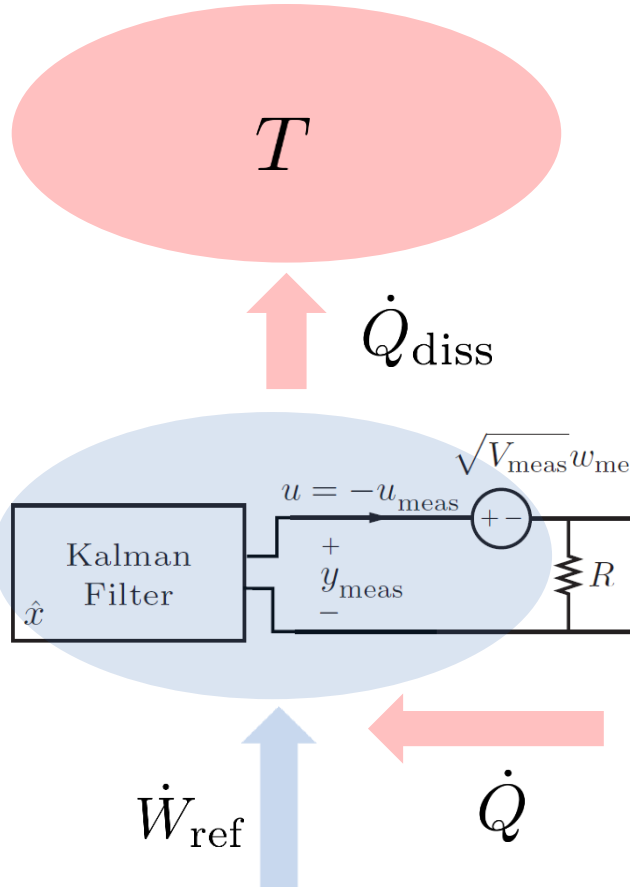
$$\dot{W}_{\mathrm{ref}} \geq \kappa T(1+\sigma-\sqrt{1+\sigma}) \approx \begin{cases} \dfrac{\kappa T}{2} \times \mathrm{SNR}\ \text{(low SNR)} \\ \kappa T \times \mathrm{SNR}\ \text{(high SNR)} \end{cases}$$

Min-variance control (BA→1)

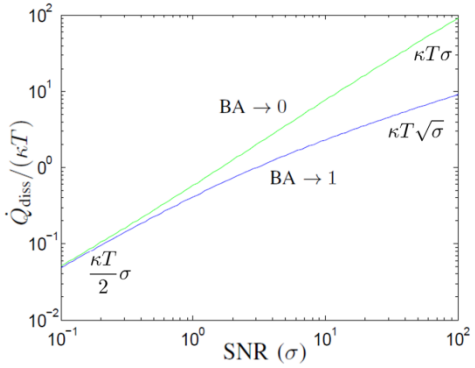
$$\dot{W}_{\mathrm{ref}} \geq \kappa T \frac{1+\sigma-\sqrt{1+\sigma}}{1+\sqrt{1+\sigma}} \approx \begin{cases} \dfrac{\kappa T}{4} \times \mathrm{SNR}\ \text{(low SNR)} \\ \kappa T \times \sqrt{\mathrm{SNR}}\ \text{(high SNR)} \end{cases}$$

Trade-Off: Power Dissipation vs. SNR



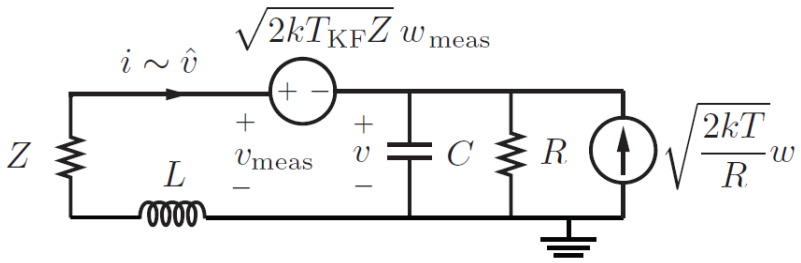


Observations



- Large back action/feedback ⇒ smaller power dissipation (“energy harvesting”)
- **Explanation**: Temperature ratio T/T_{KF} smaller ⇒ 2nd law less restrictive
- Difference significant for high SNRs
- No back action/feedback costs a factor $\sqrt{\mathrm{SNR}}$ more than min-variance control in high SNR regime

Example: Power Supply for RC Circuit



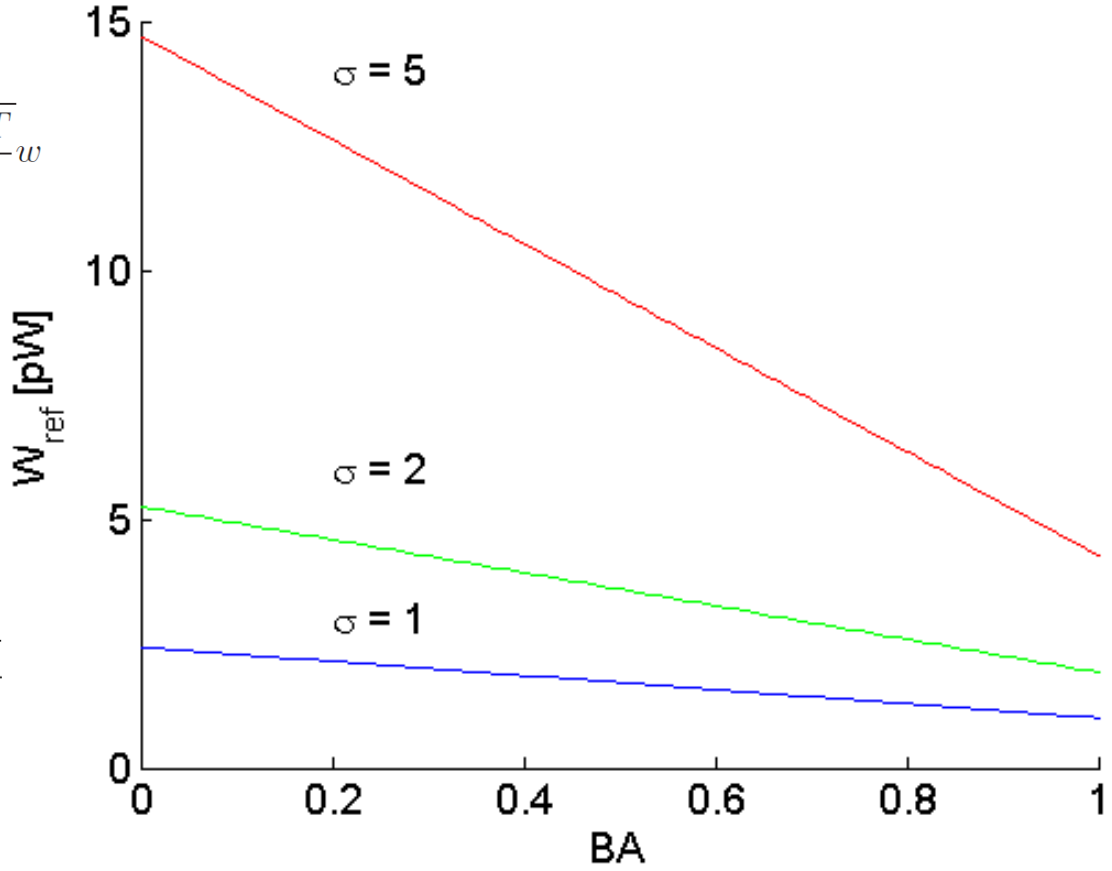
$$R = 1\,\Omega$$

$$C = 1\,\mathrm{nF}$$

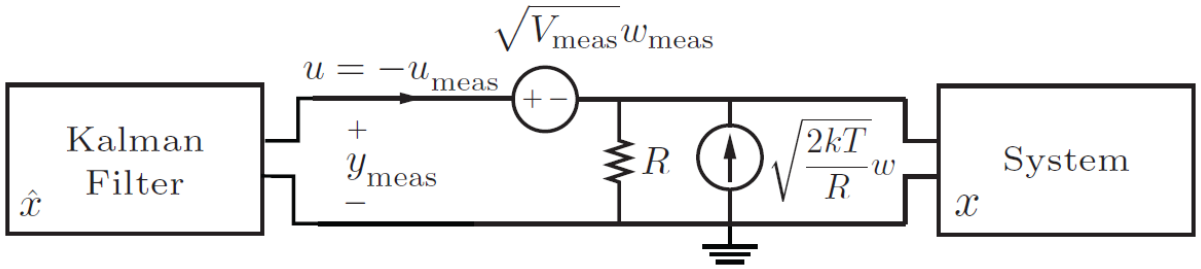
$$k = 1.38 \cdot 10^{-23}\,\mathrm{J/K}$$

$$\kappa = \frac{k}{RC} = 1.38 \cdot 10^{-14}\,\mathrm{W/K}$$

$$T = 300\,\mathrm{K}$$



Landauer's Principle and Directed Information

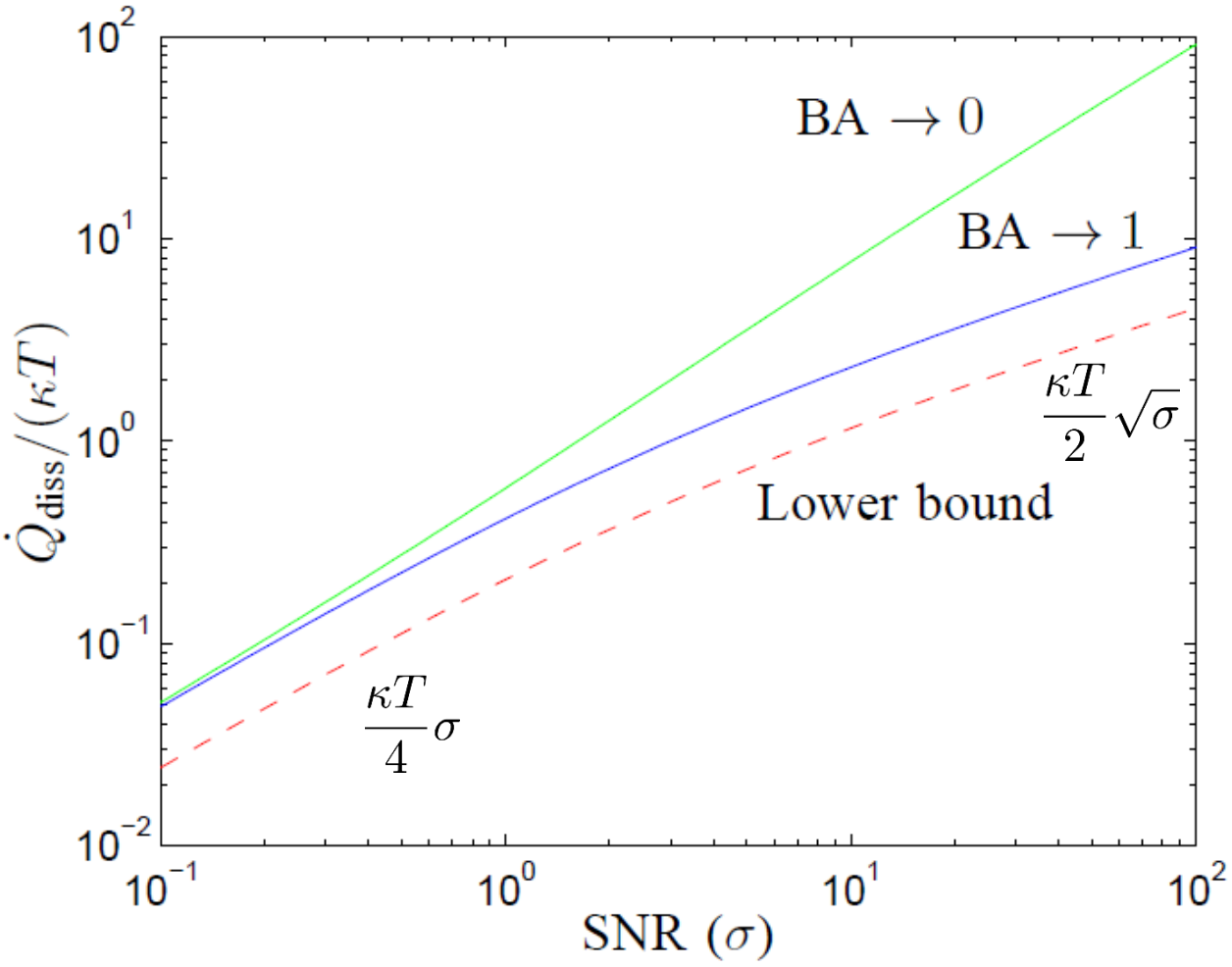


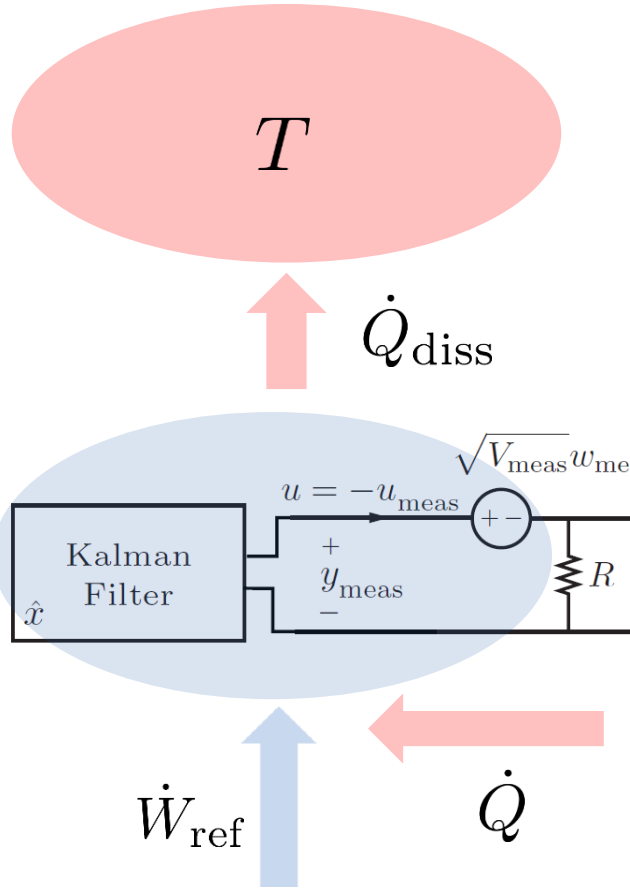
- **Landauer's principle (1961):** Need to spend at least work $kT \ln 2$ to erase one bit of information
- Directed information flow (system to filter):

$$\dot{I}_c := \frac{d}{dt} I((w_0^t, x(0)); (y_{\text{meas}})_0^t)$$

$$\dot{W}_{\text{memo}} \geq kT\dot{I}_c = \frac{\kappa T}{2}(\sqrt{1+\sigma} - 1)$$

Absolute Lower Bound Compared to Physical Implementations





Observations

- “Passive” implementation at least a factor 2 more dissipation than required by lower bound

- **Possible explanations:**
  - Landauer’s principle holds for **infinitely slow erasure**. Here finite erasure rate, which costs more
  - Factor 2 often appears in maximum-power relations. Compare with impedance matching
  - Directed information rate is a **lower bound** on **entropy rate** of memory in filter. Entropy rate can be a factor 2 larger

[Sandberg *et al*., Phys. Rev. E, 2014]

Outline

- Motivating example
- Thermodynamics of the Kalman-Bucy filter
- Relation to the information flow and Landauer's principle
- **(Maxwell's demon)**

James Clerk Maxwell (1831-1879)

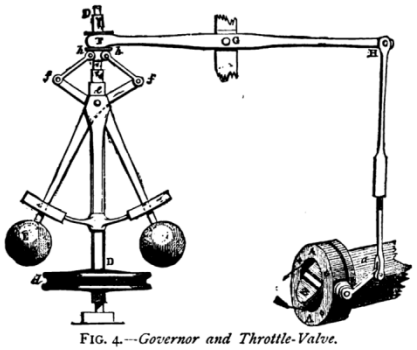
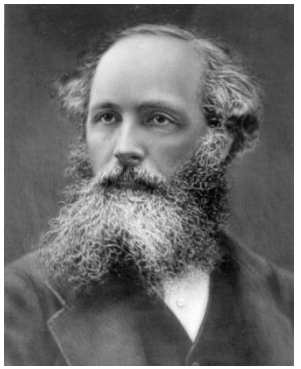


FIG. 4.—*Governor and Throttle-Valve.*



ON GOVERNORS

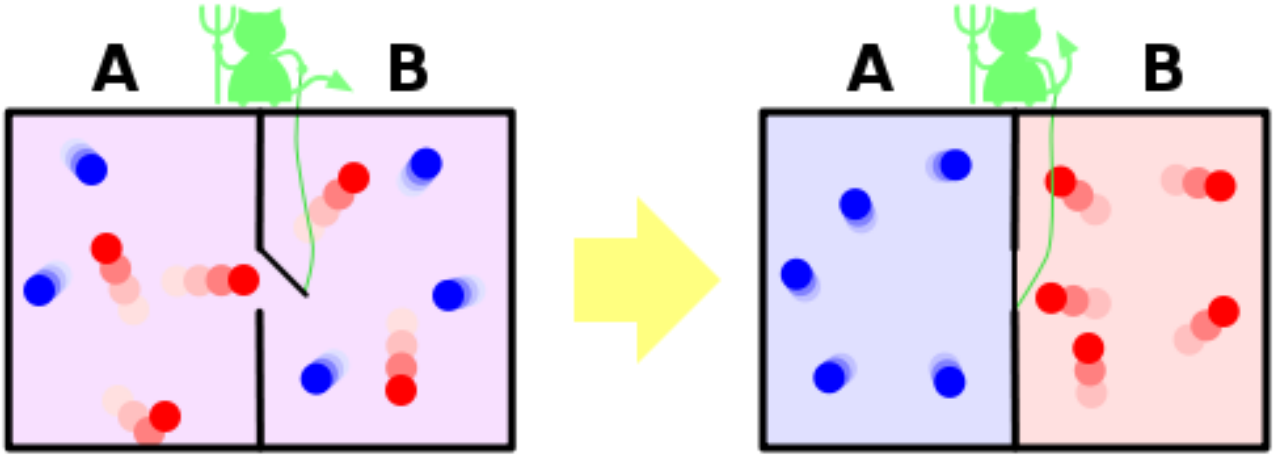
J.C. MAXWELL

From the Proceedings of the Royal Society, *No.100, 1868.*

A GOVERNOR is a part of a machine by means of which the velocity of the machine is kept nearly uniform, notwithstanding variations in the driving-power or the resistance.

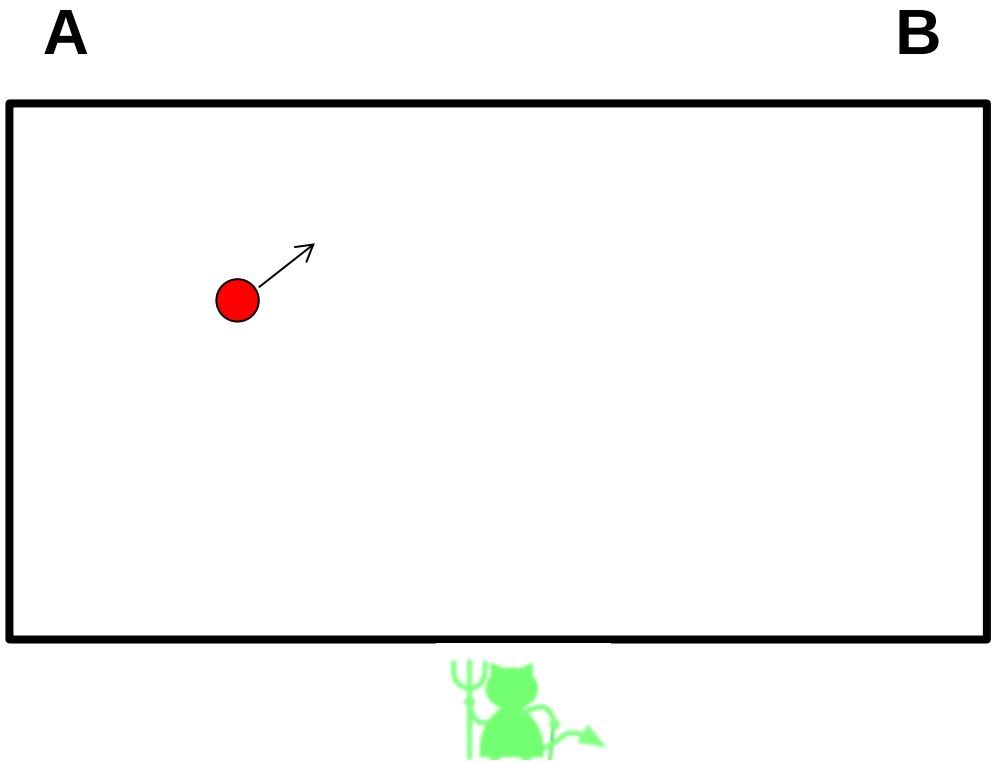
Most governors depend on the centrifugal force of a piece connected with a shaft of the machine. When the velocity increases, this force increases, and either increases the pressure of the piece against a surface or moves the piece, and so acts on a break or a valve.

Maxwell's Demon (1867, 1871)



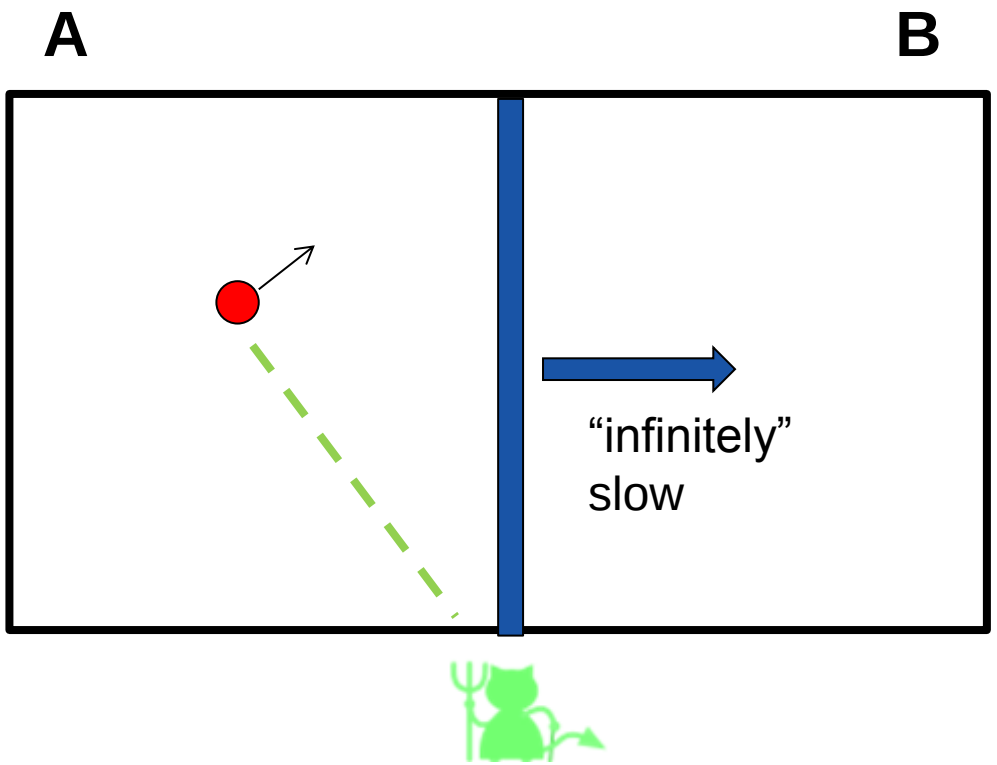
"… Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower molecules to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics ..."

Szilard's Engine (1929)



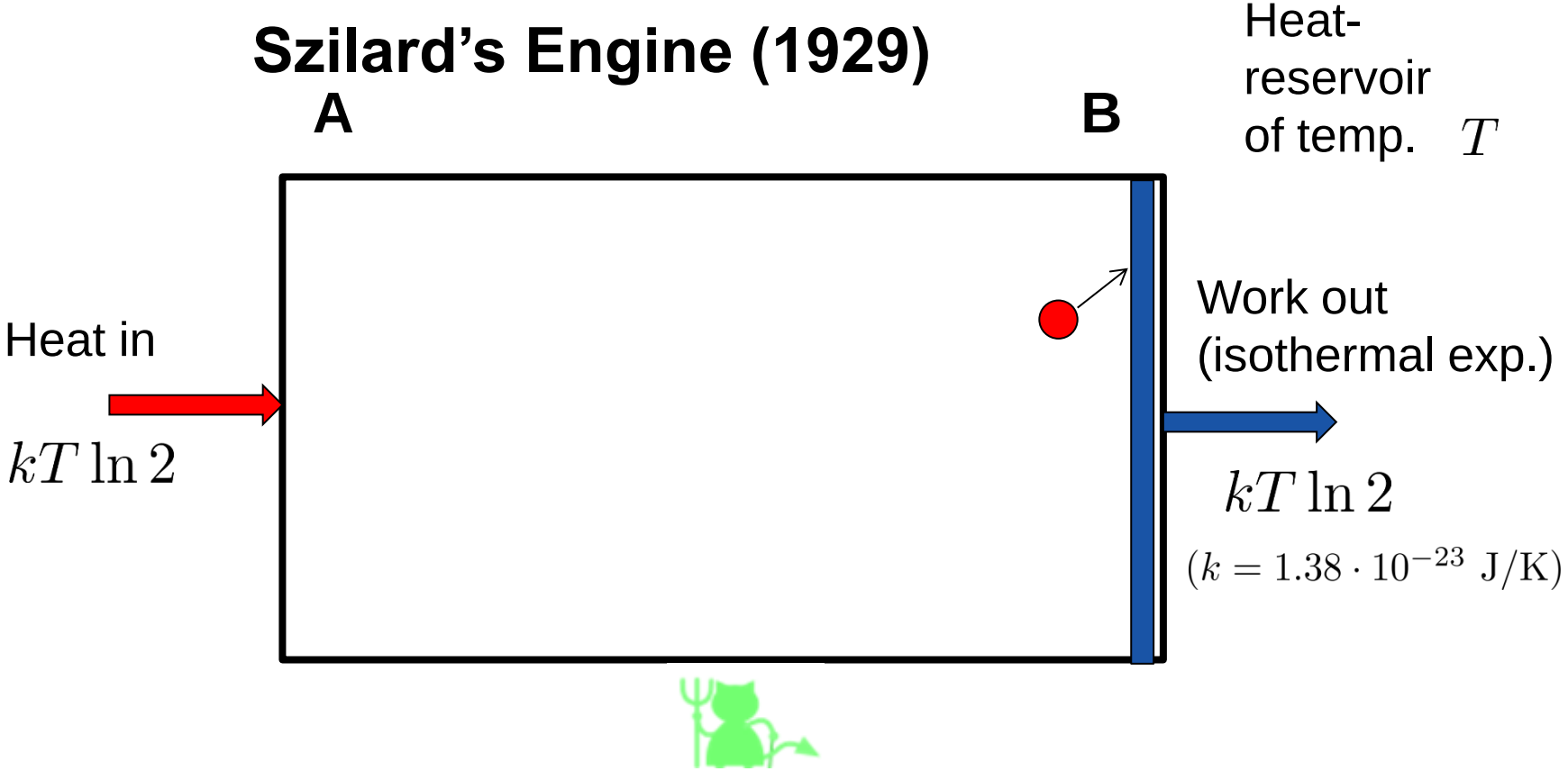
Heat-reservoir of temp. T

Szilard's Engine (1929)



Heat-reservoir of temp. T

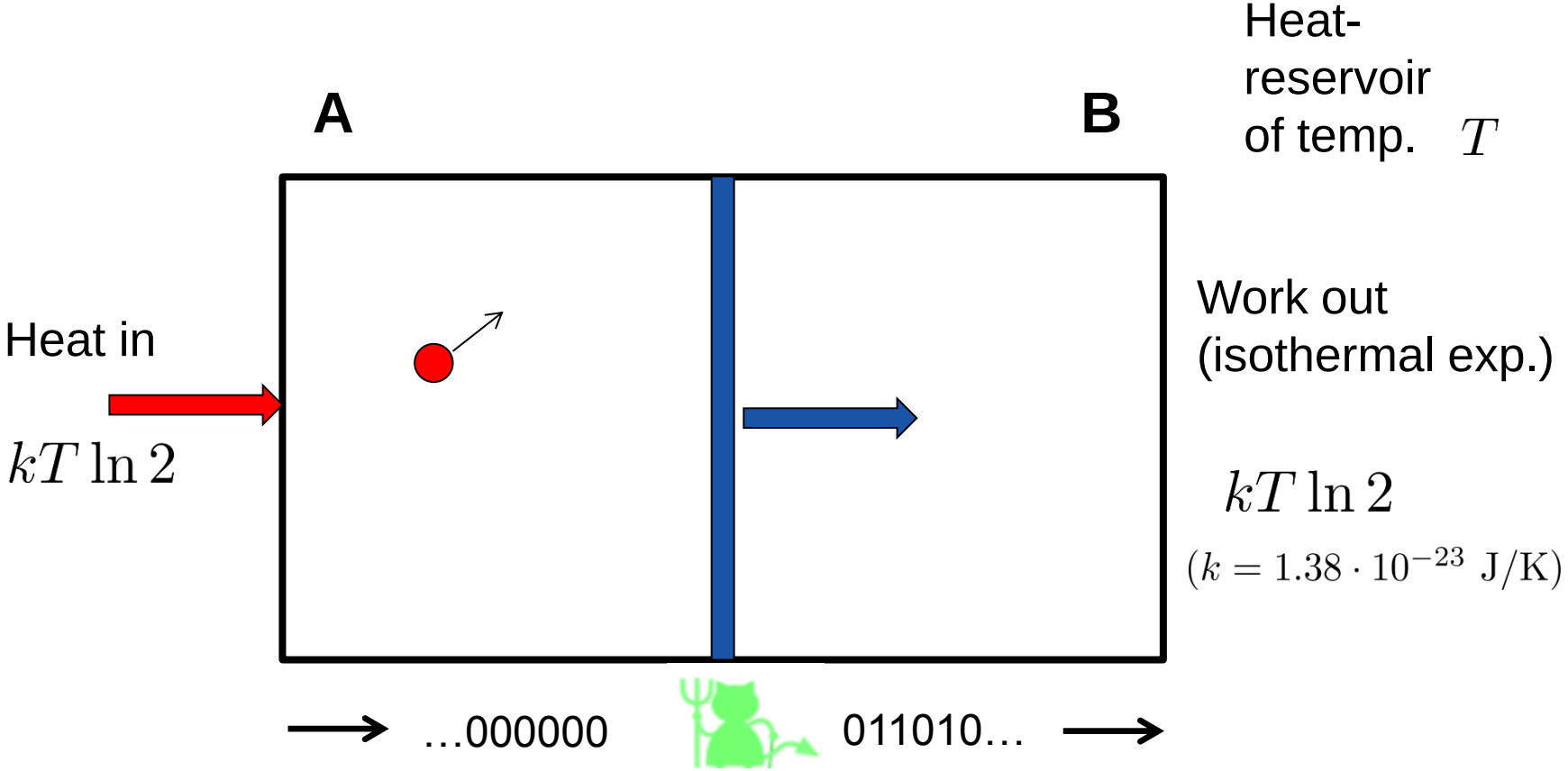
Szilard's Engine (1929)



Seemingly in violation of the Kelvin-Planck statement of the 2nd law:

"It is impossible to construct an engine which will work in a complete cycle, and produce *no effect* except the raising of a weight and the cooling of a heat-reservoir." (Planck)

Solution by Landauer, Penrose, and Bennett



Demon obtains information (1 bit/cycle), which is stored in its memory.

Landauer's principle:
Erasing 1 bit of information requires at least the work $kT \ln 2$
(IBM J. Res. Dev., 1961)

Summary

- Class of systems with “passive” Kalman-Bucy filters found. Passive but active cooling required (unless we own a cold heat bath…)
- Back action/feedback reduces required power supply
- Physical implementations are a factor 2 away from Landauer’s lower bound. In fact optimal implementation?
- Possible applications: stochastic thermodynamics, synthetic biology, energy harvesting…