Mechanical Intelligence

A live textbook on information in mechanisms

Zach J. Patterson

21 November 2026

Table of contents

Preface **5**
How the book is organized 5
How the book changes 5
What you can rely on 6
Reproducing the book 6
How to cite 6
Acknowledgments 7

I Part I. Information in mechanisms **8**

1 Information in mechanisms **9**
1.1 The idea in plain language 9
1.2 Setup and notation 12

2 The energy metric **15**
2.1 What whitening does 15
2.2 Invariance and the noise law 17
2.2.1 Exact invariance under linear changes of coordinates 17
2.2.2 Why this noise: maximum entropy at fixed energy 20

3 The dynamics in whitened coordinates **24**

4 Bodies with different numbers of coordinates **28**
4.1 The duplicate-readout artifact 28
4.2 A general constraint-limit theorem 29
4.3 The SEA: Definition 4.1 holds, with explicit rates 30
4.4 How the noise scale enters: two regimes 34

5 Attribution and morphological computation **37**
5.1 A consistent noise model for subsets of coordinates 37
5.2 Morphological computation MC_W 39

6 Continua, gears and floating bases **42**
6.1 Continuum and soft bodies: independence from the mesh 42
6.2 Gear ratios and transmission design 44
6.3 Floating bases: when a singular metric loses nothing 45
6.4 Further uses 46

II Part II. Is energy the right choice? **48**

7 Why energy **49**
7.1 What maximum entropy does and does not decide 49
7.2 Five requirements, and a uniqueness theorem . . 49
7.3 Free motion cannot create information 53
7.4 The alternatives compared 55

8 A thermodynamically consistent body: Langevin noise **59**

9 Open problems **65**
9.1 What energy whitening does and does not capture 65
9.2 Limitations 66

III Research log **68**

Research notes **69**

2026-11-21: Energy is the right noise model, and morphology routes information **70**

Research roadmap **72**
How we work 72
Threads 73
Findings log 75
Decisions log 76
Paper candidates 77

Changelog **78**
v0.1 (2026-11-21) 78

Appendices 79

A Status of the literature claims 79

References 82

Preface

This is a live textbook on how mechanisms carry and process information about the world: how much a robot's body knows about the forces acting on it, how that knowledge depends on its design, and how much of it reaches the sensors a controller can use. It starts from 1, which compared actuator transmissions in legged robots with information-theoretic measures, and it grows as new results come in.

How the book is organized

Part I builds the main tool, the *energy metric*. It explains why the body's state is measured in units of energy before noise is added, proves what that buys (invariance under any change of coordinates, the right behaviour when a spring becomes rigid, results that do not depend on a finite-element mesh), and works through the leg designs of the running example. Part II asks whether energy is the right choice: it proves that energy is the only noise model with a short list of physical properties, compares the alternatives, and builds a thermally consistent body in which a design's job is to route information to its sensors. The research log holds dated notes, the roadmap and the changelog.

How the book changes

The book is not planned in advance. New results first appear as dated research notes in the research log. A note moves into a chapter once its result is checked, its proofs are labelled, and it connects to the existing material; the note stays as the dated record. The roadmap lists the open threads and what happens

next on each, and the changelog lists what changed in each release.

Every page opens with its status:

- *exploratory*: a new result, checked by code but not yet connected to the rest;
- *draft*: connected and proved, still being revised;
- *stable*: unlikely to change except for corrections.

What you can rely on

Every number in the text is computed by the Python package `mibook` in this book's repository, and a test in `tests/` pins it, so a number that drifts fails the build. Figures are generated by code cells when the book is built. A proof that is not complete is titled "Proof sketch" and says what it assumes. Every citation is marked by whether its source was read in full or only through abstracts and citing works (Appendix A).

Reproducing the book

The repository is github.com/CyPhiLab/mechintelligence. With uv installed:

```
uv sync
uv run pytest
uv run quarto render
```

The computations depend on the analysis library from zpatty/quadruped-mech-info, pinned to a fixed commit.

How to cite

Cite the version you read, since pages change between releases:

Patterson, Z. J. *Mechanical Intelligence: A live textbook on information in mechanisms.* Version 0.1, 2026. https://mechintelligence.cyphilab.com/

Acknowledgments

The text and code were developed with the help of Claude Code. Every result is checked by the code in the repository.

Part I

Part I. Information in mechanisms

1 Information in mechanisms

Status: stable · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

1.1 The idea in plain language

This section gives the whole idea without proofs. Later chapters make each sentence precise.

The running example. A robot leg presses on the ground, and an unknown force pushes back. We ask how much the leg's motion tells us about that force. Two leg designs recur throughout the book, both from 1 with parameters from Wensing et al. (2017). One has a *rigid* (proprioceptive) transmission, with the motor geared directly to the joint. The other is a *series-elastic actuator* (SEA), with a spring of stiffness k_s between motor and joint. As k_s grows, the SEA turns into the rigid leg.

A deterministic mechanism carries infinite information. The leg obeys deterministic equations. If we knew them exactly and could read the state with perfect precision, we could run them backward and recover the force exactly. Recovering a real number exactly takes infinitely many bits. Strictly, then, every mechanism carries infinite information about the force, and comparing two designs gives "∞ vs ∞." To get a finite answer we have to say how precisely the state can be known, which means adding noise. The noise defines the resolution, and the answer depends on it.

Relabeling changes nothing. If you transform the observed quantity in a way that can be undone, the information stays the same. Converting meters to millimeters, rotating axes and

multiplying the state by any invertible matrix are all examples. Anything learnable from the original is learnable from the transformed version by undoing the transformation first, and the reverse holds too.

Whitening, $x = \mathcal{E}^{1/2} x_q$, is such a transformation, so on its own it changes nothing. It matters because of the order of operations: first transform, then add noise of the same size in every direction. Adding equal noise *after* the transformation is the same as adding a particular, unequal noise *before* it. So "energy whitening" is a statement about what the noise on the physical state looks like.

What the noise looks like. Take a mass m on a spring of stiffness k, with position q and velocity $\dot{q}$. Whitening gives $x = (\sqrt{k}\, q, \sqrt{m}\, \dot{q})$. Its squared length $kq^2 + m\dot{q}^2$ is twice the stored energy, and every component is measured in $\sqrt{\text{J}}$. Adding noise of standard deviation σ to both components means, back in physical units:

- position noise of standard deviation $\sigma/\sqrt{k}$;
- velocity noise of standard deviation $\sigma/\sqrt{m}$.

Stiff springs make position precise, and heavy masses make velocity precise, because both blurs carry the same energy, $\frac{1}{2}\sigma^2$. The rule being imposed is that two states differing by less than a fixed amount of energy cannot be told apart.

Geometrically, the noise has the constant-energy ellipses as its contour lines (Figure 1.1(a)). The alternative, "add the same noise to each coordinate," asserts that an error of 10^{-4} rad in an angle and 10^{-4} rad/s in a velocity are equally hard to detect (Figure 1.1(b)). The two quantities do not even share units, and the assertion changes when the units do.

Why this noise: maximum entropy. Suppose the only thing we are willing to commit to is the noise's average energy. Choosing a distribution means making assumptions: a sharply peaked distribution claims to know a lot, and a lopsided one claims some directions matter more. The standard measure of how few assumptions a distribution makes is its *entropy*, and Jaynes' *principle of maximum entropy* says to choose, among

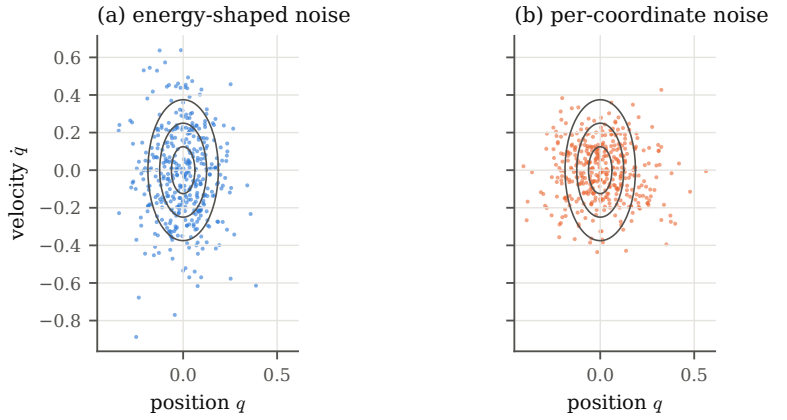


Figure 1.1: A mass on a spring ($k = 4$, $m = 1$, illustrative units). Grey curves are constant-energy ellipses. (a) Energy-shaped noise has the same ellipses as its contours: every perturbation of equal energy is equally likely. (b) Per-coordinate noise is round in whatever units the coordinates happen to use.

all distributions consistent with what you know, the one with the largest entropy.

Among all noise laws with a given average energy, the largest-entropy one is $p \propto \exp(-\text{energy}/T)$. For quadratic energy this is a Gaussian with covariance $T\,\mathcal{E}^{-1}$, which is the noise above with $T = \sigma^2$. In whitened coordinates the constraint says only that the average squared length is fixed, which treats every direction alike, and the most even way to spread a fixed average squared length is the round Gaussian.

Gibbs and equipartition. The formula $p \propto \exp(-\text{energy}/T)$ is the Gibbs (Boltzmann) distribution, the law of a system in equilibrium with a heat bath. It implies *equipartition*: every quadratic term in the energy holds $\frac{1}{2}k_BT$ on average. For the mass on a spring, $\langle q^2 \rangle = k_BT/k$ and $\langle \dot{q}^2 \rangle = k_BT/m$, which are the variances above. Energy noise therefore blurs the body as if it were jiggling thermally at temperature σ^2.

That temperature is a resolution written in energy units. The value used throughout, $\sigma^2 = 10^{-4}$ J, is about 10^{16} times room-temperature k_BT, so it has nothing to do with real heat.

Physics fixes only the *shape* of the noise; its *size* mostly sets the zero of the bit scale.

Why this matters for comparing designs. The SEA has an extra coordinate, the motor angle, which moves in lockstep with the joint angle once the spring is stiff.

- *With per-coordinate noise*, the motor angle becomes a second, independently noisy reading of the same signal. That is worth about half a bit per force mode, 8.47 bits here, so a stiff SEA appears "smarter" than the rigid leg it has physically become.
- *With energy noise*, a direction of motion earns information only through the energy it stores, so relabeling or duplicating coordinates never creates bits.

Is energy the best choice? Part II takes this up (Chapter 7, Chapter 8). Maximum entropy justifies the noise's *shape* once energy is chosen, but it does not choose energy. That choice has a separate justification: energy is the only noise metric with five physical properties, the decisive one being that free motion can never create information. Even then, a noise model is not yet a measure of computation, and a thermodynamically consistent "Langevin body" shows what morphology can and cannot do.

1.2 Setup and notation

The mechanism. A linear(ized) mechanism with n generalized coordinates $q \in \mathbb{R}^n$ obeys

$$H\ddot{q} + D\dot{q} + Kq = J^\top f(t), \tag{1.1}$$

where:

- $H = H^\top \succ 0$ is the mass matrix;
- $K = K^\top \succeq 0$ is the stiffness, including the ground spring $J^\top k_i J$;
- $D = D^\top \succeq 0$ is the damping;
- J maps generalized velocity to foot velocity, so f enters through $J^\top$.

The state is $x_q = (q, \dot{q}) \in \mathbb{R}^{2n}$, and Equation 1.1 reads $\dot{x}_q = Ax_q + Bf$ with

$$A = \begin{bmatrix} 0 & I \\ -H^{-1}K & -H^{-1}D \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ H^{-1}J^\top \end{bmatrix}. \tag{1.2}$$

The energy metric. As in Patterson (2026, Eq. 9), define

$$\mathcal{E} = \begin{bmatrix} K & 0 \\ 0 & H \end{bmatrix}, \qquad x = \mathcal{E}^{1/2} x_q. \tag{1.3}$$

Here $\mathcal{E}^{1/2}$ is the symmetric positive semidefinite square root (`mechinfo.mechanics.sym_sqrt`). Because $\mathcal{E}$ is block diagonal, $\mathcal{E}^{1/2} = \mathrm{diag}(K^{1/2}, H^{1/2})$. The calligraphic $\mathcal{E}$ leaves $\mathbb{E}$ free for expectation. When $K \succ 0$ we call the metric *nondegenerate*.

The channel. The unknown force over a window $[0, T]$ is expanded in d basis functions, $f(t) = \sum_{j=1}^{d} w_j \varphi_j(t)$ with $w \sim \mathcal{N}(0, \sigma_w^2 I_d)$. The state is sampled at $t_k = k/f_s$, $k = 0, \dots, N_t - 1$. A readout matrix C is applied, and i.i.d. noise $n_k \sim \mathcal{N}(0, \sigma^2 I)$ is added:

$$y_k = C x_q(t_k) + n_k, \qquad y = Mw + n. \tag{1.4}$$

Here M stacks the responses: column j of M is the sampled readout of the response to φ_j. Write $X_k \in \mathbb{R}^{2n \times d}$ for the matrix whose column j is the state at t_k under input φ_j, so that $M = [CX_0; \dots; CX_{N_t - 1}]$. The energy readout is $C = \mathcal{E}^{1/2}$, with noise scale σ in $\sqrt{\mathrm{J}}$ (`passive_info.information.SIGMA_ENERGY`).

Lemma 1.1 (Gaussian channel)**.** *For Equation 1.4,*

$$I(w; y) = \tfrac{1}{2} \log\det(I_d + \gamma M^\top M), \qquad \gamma = \sigma_w^2 / \sigma^2,$$

and the right-hand side depends on C only through the $d \times d$ Gram matrix

$$G \;=\; \gamma M^\top M \;=\; \gamma \sum_k X_k^\top \left(C^\top C\right) X_k. \tag{1.5}$$

Proof.

1. Given w, y is Gaussian with covariance $\sigma^2 I$, so $h(y \mid w) = \frac{1}{2}\log\det(2\pi e\,\sigma^2 I)$.
2. Unconditionally, y is Gaussian with covariance $\sigma^2 I + \sigma_w^2 M M^\top$, so $h(y) = \frac{1}{2}\log\det(2\pi e(\sigma^2 I + \sigma_w^2 M M^\top))$.
3. Subtracting, $I(w;y) = h(y) - h(y \mid w) = \frac{1}{2}\log\det(I + \gamma M M^\top)$.
4. Sylvester's identity, $\det(I+PQ) = \det(I+QP)$, replaces $MM^\top$ by $M^\top M$.
5. Finally, $M^\top M = \sum_k (CX_k)^\top (CX_k) = \sum_k X_k^\top C^\top C X_k$.

□

Throughout, "bits" means $\mathrm{bits}(G) = \frac{1}{2}\log_2\det(I+G)$ (`mechinfo.mechanics.bits_from_gram`).

The question. Lemma 1.1 shows that the readout enters only through $C^\top C$, so whitening by $\mathcal{E}^{1/2}$ amounts to declaring $C^\top C = \mathcal{E}$. The following chapters ask why this is the right declaration, what it buys, and where it fails.

Models and settings. Unless stated otherwise, the numbers come from the clamped one-leg model of 1, in its rigid form ($q = \theta$) and its SEA form ($q = (\theta_j, \theta_m)$), with force band $B_w = 100$ Hz ($d = 17$ modes), window $T = 85$ ms and sampling rate $f_s = 4$ kHz. `mibook.checks` computes every number, and the tests in `tests/` pin them.

2 The energy metric

Status: stable · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

2.1 What whitening does

Lemma 2.1 (Only the noise model matters)**.** *Let S be an invertible matrix and let X be any random vector.*

(a) $I(w; SX) = I(w; X)$.
(b) Without noise, the information is infinite: if $M \neq 0$, then $I(w; Mw) = \infty$.
(c) $I(w;\, \mathcal{E}^{1/2}x_q + n) = I(w;\, x_q + \mathcal{E}^{-1/2}n)$ *when $\mathcal{E} \succ 0$. Hence isotropic noise on the whitened state is* the same channel *as noise $\nu = \mathcal{E}^{-1/2}n \sim \mathcal{N}(0, \sigma^2\mathcal{E}^{-1})$ added to the physical state.*

Proof.

(a) The map $X \mapsto SX$ is a bijection, with inverse S^{-1}. By the data-processing inequality, $I(w; SX) \leq I(w; X)$ because SX is computed from X. Equally, $I(w; X) \leq I(w; SX)$ because $X = S^{-1}(SX)$ is computed from SX. So the two are equal.

(b) For any $\varepsilon > 0$, the vector $Mw + \varepsilon n$ is computed from Mw and an independent n. Data processing therefore gives $I(w; Mw) \geq I(w; Mw + \varepsilon n)$. By Lemma 1.1 the right-hand side is $\frac{1}{2}\log\det(I + \sigma_w^2\varepsilon^{-2}M^\top M)$. At least one eigenvalue of $M^\top M$ is positive, so this tends to ∞ as $\varepsilon \to 0$.

(c) Apply (a) with $S = \mathcal{E}^{-1/2}$ to the observation $\mathcal{E}^{1/2}x_q + n$. That gives $x_q + \mathcal{E}^{-1/2}n$. The covariance of $\mathcal{E}^{-1/2}n$ is $\mathcal{E}^{-1/2}(\sigma^2 I)\mathcal{E}^{-1/2} = \sigma^2\mathcal{E}^{-1}$.

□

What this means physically

The trajectory of a deterministic mechanism determines the force exactly, so without a declared resolution the bit count is infinite. That infinity comes from the analysis, and says nothing about the body. Whitening changes no information by itself. Its whole content is one statement: the body's state is known to within a Gaussian blur of covariance $\sigma^2\mathcal{E}^{-1}$. In communication-theory terms, $\mathcal{E}^{1/2}$ is the *noise-whitening filter* for that colored noise. Applying it before the isotropic-noise formula is the standard way to compute capacity with colored noise (Cover and Thomas 2006).

Lemma 2.2 (Energy identity)**.** $|x|^2 = x_q^\top \mathcal{E} x_q = q^\top K q + \dot{q}^\top H \dot{q} = 2(V + T_{\text{kin}})$*, which is twice the stored mechanical energy. Every component of* x *therefore has units of* $\sqrt{\text{J}}$*.*

Proof. $|x|^2 = x_q^\top \mathcal{E}^{1/2}\mathcal{E}^{1/2}x_q = x_q^\top \mathcal{E} x_q$. Expanding the block-diagonal form gives the rest. □

Lemma 2.3 (Every square root gives the same bits)**.** *Let* C *be any readout with* $C^\top C = \mathcal{E}$*. Examples are the symmetric root* $\mathcal{E}^{1/2}$*, the upper Cholesky factor, and any* $U\mathcal{E}^{1/2}$ *with* $U^\top U = I$*. Then every such* C *produces the same Gram matrix Equation 1.5, and so the same bits. Conversely, if* C *is square and invertible with* $C^\top C = \mathcal{E}$*, then* $C = U\mathcal{E}^{1/2}$ *for an orthogonal* U*.*

Proof. The first claim is immediate from Equation 1.5, which contains C only as $C^\top C$.

For the converse, put $U = C\mathcal{E}^{-1/2}$. Then $U^\top U = \mathcal{E}^{-1/2}C^\top C\mathcal{E}^{-1/2} = \mathcal{E}^{-1/2}\mathcal{E}\,\mathcal{E}^{-1/2} = I$, so U is orthogonal. □

Remark 2.1 (Relation to statistical whitening). Lemma 2.3 is the mechanical version of the rotational freedom of whitening noted by Kessy et al. (2018, Eq. (4)): every whitening matrix has the form $W = Q_1\Sigma^{-1/2}$ with Q_1 orthogonal. Among these, the symmetric (ZCA-Mahalanobis) choice $W = \Sigma^{-1/2}$ (their Eq. (8)) is the unique one that minimizes $\mathbb{E}|z - x|^2$ (their Eqs. (13) and (14)). With $\Sigma = \sigma^2\mathcal{E}^{-1}$ that whitener is $\mathcal{E}^{1/2}/\sigma$.

The symmetric root used throughout is therefore the one that keeps whitened coordinates as close as possible to the original ones, though the bits do not depend on the choice. `embodied.sim_profile_bits` uses $\sum_k X_k^\top \mathcal{E} X_k$ directly, and `sim_info` uses $\mathcal{E}^{1/2}$; by Lemma 2.3 the two agree.

Numerical check

The symmetric root, the Cholesky factor, the direct Gram, and physical noise $\sigma^2\mathcal{E}^{-1}$ (block-diagonal over time) all give 71.77253 bits for the SEA at $k_s = 10^4$. They agree to 10^{-13} (`check_square_roots`, `check_noise_equivalence`).

2.2 Invariance and the noise law

2.2.1 Exact invariance under linear changes of coordinates

Suppose the same mechanism is described in new coordinates $\tilde{q}$, related by $q = T\tilde{q}$ with T constant and invertible. Examples: radians to degrees; joint and motor angles to joint angle and spring deflection $\delta = \theta_j - \theta_m$; a motor angle measured before or after a gear.

Theorem 2.1 (Coordinate invariance)**.** *Under $q = T\tilde{q}$, the energy-whitened Gram matrix, and hence every bit count computed from it, is* identical *in the two descriptions. This includes chain-rule band decompositions, which are functions of sub-blocks of G.*

By contrast, for per-coordinate noise with a fixed diagonal weight Λ (i.e. $C^\top C = \Lambda$ in both descriptions), the Gram

changes from $\sum_k X_k^\top \Lambda X_k$ *to* $\sum_k X_k^\top \mathcal{T}^{-\top}\Lambda\mathcal{T}^{-1}X_k$, *where* $\mathcal{T} = \mathrm{diag}(T, T)$. *This differs unless* $\mathcal{T}$ *preserves* Λ.

Proof. Step 1: how the matrices transform. Kinetic energy, potential energy, Rayleigh dissipation and virtual work are scalars, so they must take the same value in both descriptions:

- $\frac{1}{2}\dot{q}^\top H\dot{q} = \frac{1}{2}\dot{\tilde{q}}^\top (T^\top HT)\dot{\tilde{q}}$;
- $\frac{1}{2}q^\top Kq = \frac{1}{2}\tilde{q}^\top (T^\top KT)\tilde{q}$;
- $\frac{1}{2}\dot{q}^\top D\dot{q} = \frac{1}{2}\dot{\tilde{q}}^\top (T^\top DT)\dot{\tilde{q}}$;
- $f \cdot J\delta q = f \cdot (JT)\delta\tilde{q}$.

So $\tilde{H} = T^\top HT$, $\tilde{K} = T^\top KT$, $\tilde{D} = T^\top DT$ and $\tilde{J} = JT$. Equivalently, substitute $q = T\tilde{q}$ into Equation 1.1 and multiply on the left by $T^\top$.

Step 2: same trajectories. If $q(t)$ solves Equation 1.1 from rest, then $\tilde{q}(t) = T^{-1}q(t)$ solves the transformed equation from rest. Solutions of linear ODEs are unique, so this is *the* transformed trajectory. Hence $\tilde{X}_k = \mathcal{T}^{-1}X_k$ for every k.

Step 3: same Gram. The new metric is $\tilde{\mathcal{E}} = \mathrm{diag}(\tilde{K}, \tilde{H}) = \mathcal{T}^\top \mathcal{E}\,\mathcal{T}$. Therefore

$$\sum_k \tilde{X}_k^\top \tilde{\mathcal{E}}\tilde{X}_k = \sum_k X_k^\top \mathcal{T}^{-\top}\mathcal{T}^\top \mathcal{E}\,\mathcal{T}\mathcal{T}^{-1}X_k = \sum_k X_k^\top \mathcal{E}X_k.$$

The same computation with Λ in place of $\tilde{\mathcal{E}}$ gives the second claim. □

What this means physically

The energy inner product $\langle x_q, x'_q\rangle_{\mathcal{E}} = x_q^\top \mathcal{E}x'_q$ between two states is a physical quantity: (twice) a cross-energy. It does not care what you called the coordinates. A per-coordinate noise budget asserts that "one unit" of motion in each coordinate is equally distinguishable, and that assertion changes when the coordinates do.

Example 2.1 (Radians versus degrees)**.** Relabel both SEA angles in degrees ($T = \frac{\pi}{180}I$) and keep a per-coordinate

noise of fixed numerical size. Then $\mathcal{T}^{-\top}\Lambda\mathcal{T}^{-1} = c\Lambda$ with $c = (180/\pi)^2 \approx 3283$, so the Gram is multiplied by c.

If all $d = 17$ modes are well excited, the high-SNR formula of Proposition 4.3 predicts a gain of $\frac{1}{2} \cdot 17 \cdot \log_2 c = 99.29$ bits. The exact gain is $\frac{1}{2}\sum_i \log_2 \frac{1+c\lambda_i}{1+\lambda_i} = 99.258$ bits, and the code returns $180.62 - 81.37 = 99.258$ bits. The energy-whitened count is 71.77253 bits in both cases (Figure 2.1).

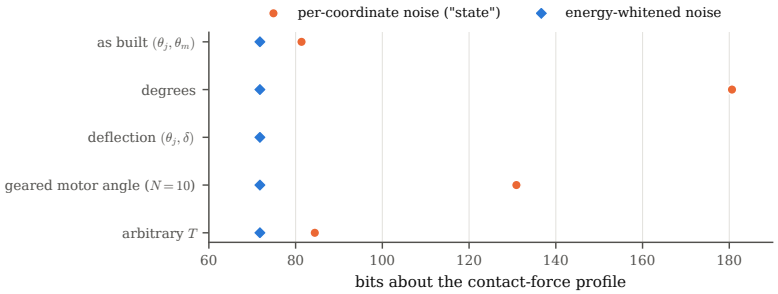


Figure 2.1: The same SEA ($k_s = 10^4$) described in five sets of generalized coordinates. Energy whitening gives 71.77253 bits every time; the spread is 3×10^{-14}. A per-coordinate noise budget gives anything from 71.6 to 180.6 bits (`check_invariance`).

Remark 2.2 (Scope: nonlinear coordinates, and what Shapley values inherit).

1. *Nonlinear reparametrizations.* For $q = \phi(\tilde{q})$, the kinetic energy transforms exactly: $\tilde{H}(\tilde{q}) = D\phi^\top H(\phi(\tilde{q}))D\phi$, because $H(q)$ is a Riemannian metric. The stiffness Hessian transforms tensorially only at an equilibrium, where $\nabla V = 0$; elsewhere a term $\sum_i \partial_i V \, \nabla^2 \phi_i$ appears.
2. *Consequence for the simulator.* The MuJoCo leg freezes $\mathcal{E}$ at the stance pose (`embodied._sim_metric`). That is exact to first order in the deflection. A natural refinement is to whiten velocities with $H(q(t))$ along the trajectory. This is exactly invariant under any smooth reparametrization and costs one `mj_fullM` call per sample.
3. *Shapley values.* The *total* bits are coordinate-free, but a Shapley attribution *over coordinates* is not: its play-

ers are the coordinates. Attributing over (θ_j, δ) gives different shares from attributing over (θ_j, θ_m). Shares are meaningful when the players are physically meaningful (sensor sites, modes); see Chapter 5.

Remark 2.3 (Relation to geometric identification). Lee et al. (2020) pursue the same goal in a different space. As we understand it from secondary sources, they replace Euclidean distances between inertial parameters by a coordinate-invariant Riemannian metric on positive-definite pseudo-inertia matrices. That makes identification independent of the body frame.

Energy whitening does the analogous thing on the *state* space. A physically defined positive-definite form ($\mathcal{E}$, rather than the pseudo-inertia) fixes the geometry, so that results do not depend on the coordinates (Theorem 2.1). The analogy is conceptual: the two metrics live on different spaces and serve different purposes.

2.2.2 Why this noise: maximum entropy at fixed energy

Theorem 2.2 (Maximum entropy). *Let $\mathcal{E} \succ 0$ be $m \times m$ and $\sigma > 0$. Among all probability densities p on $\mathbb{R}^m$ with mean energy budget $\mathbb{E}_p[\nu^\top \mathcal{E} \nu] = m\sigma^2$, the differential entropy is uniquely maximized by $g = \mathcal{N}(0, \sigma^2 \mathcal{E}^{-1})$.*

Proof. Step 1: the log-density of g.

$$\log g(\nu) = -\tfrac{1}{2} \log \det(2\pi\sigma^2 \mathcal{E}^{-1}) - \frac{\nu^\top \mathcal{E} \nu}{2\sigma^2}.$$

Step 2: average it under any admissible p. The constraint fixes $\mathbb{E}_p[\nu^\top \mathcal{E} \nu] = m\sigma^2$, so

$$-\mathbb{E}_p[\log g] = \tfrac{1}{2} \log \det(2\pi\sigma^2 \mathcal{E}^{-1}) + \tfrac{m}{2}.$$

This number does not depend on which admissible p we chose. In particular it equals $-\mathbb{E}_g[\log g] = h(g)$, since g is itself admissible: $\mathbb{E}_g[\nu^\top \mathcal{E} \nu] = \operatorname{tr}(\mathcal{E}\, \sigma^2 \mathcal{E}^{-1}) = m\sigma^2$.

Step 3: Gibbs' inequality.

$$0 \le \mathrm{KL}(p \,\|\, g) = \mathbb{E}_p[\log p] - \mathbb{E}_p[\log g] = -h(p) + h(g).$$

So $h(p) \le h(g)$, with equality if and only if $p = g$ almost everywhere. □

What this means physically

If the only thing we are willing to assert about the body's resolution is *how much energy* an indistinguishable perturbation carries, the least committal noise law (Jaynes 1957, sec. 2) is $\mathcal{N}(0, \sigma^2 \mathcal{E}^{-1})$. In whitened coordinates this is isotropic noise, the model used throughout the book.
Jaynes treats the discrete case (his Eqs. (2-4) and (2-10), with the Boltzmann distribution in Sec. 3), and his footnote 6 sets the continuous case aside. The continuous argument above is the standard extension (Cover and Thomas 2006, Ch. 12). One caveat: it uses Lebesgue measure on $(q, \dot{q})$. For constant H this differs from Liouville measure on (q, p) only by the constant factor $\det H$.

Corollary 2.1 (Gibbs, equipartition, and the effective temperature)**.**

1. *The Gibbs density at temperature Θ for the quadratic energy $\frac{1}{2} x_q^\top \mathcal{E} x_q$ is $\propto \exp(-x_q^\top \mathcal{E} x_q / 2 k_B \Theta)$, which is $\mathcal{N}(0, k_B \Theta\, \mathcal{E}^{-1})$.*
2. *Energy noise is therefore the equilibrium fluctuation law at effective temperature $k_B \Theta_{\text{eff}} = \sigma^2$.*
3. *Each whitened coordinate carries mean energy $\frac{1}{2}\sigma^2$, which is equipartition.*

What this means physically

We anchor $\sigma = \sqrt{k_i}\, r\, \sigma_\theta = 10^{-2}\sqrt{\mathrm{J}}$. That makes σ^2 twice the ground-spring energy stored by one encoder tick of foot travel, $\sigma^2 = 10^{-4}$ J.
Room-temperature $k_B T$ is 4.1×10^{-21} J, so Θ_{eff} is about 10^{16} times room temperature. The "temperature" is a *resolution* written in energy units, not thermal noise. Only its *shape* $\mathcal{E}^{-1}$ is fixed by the physics; its *scale* sets the

zero of the bit axis. Proposition 4.3 shows when design comparisons do not depend on that scale.

Proposition 2.1 (Fluctuation-dissipation balance)**.** *Let $K \succ 0$, and drive Equation 1.1 by white force noise ξ in place of $J^\top f$, with intensity $\mathbb{E}[\xi(t)\xi(s)^\top] = 2k_B\Theta\, D\,\delta(t-s)$. Then $P = k_B\Theta\,\mathcal{E}^{-1}$ is a stationary covariance: it solves the Lyapunov equation $AP + PA^\top + B_\xi(2k_B\Theta D)B_\xi^\top = 0$, where $B_\xi = [0; H^{-1}]$.*

Proof. Step 1. With $P = k_B\Theta\,\mathrm{diag}(K^{-1}, H^{-1})$, direct multiplication gives

$$AP = k_B\Theta \begin{bmatrix} 0 & H^{-1} \\ -H^{-1} & -H^{-1}DH^{-1} \end{bmatrix}.$$

Step 2. Adding its transpose, the off-diagonal blocks cancel:

$$AP + PA^\top = k_B\Theta\ \mathrm{diag}(0,\ -2H^{-1}DH^{-1}).$$

Step 3. The noise term is $B_\xi(2k_B\Theta D)B_\xi^\top = k_B\Theta\ \mathrm{diag}(0,\ 2H^{-1}DH^{-1})$, which cancels Step 2 exactly. □

What this means physically

Kubo (1966) ties the intensity of thermal forcing to the dissipation. Proposition 2.1 is the matrix statement for a linear mechanism: if the body is "thermalized" by noise matched to its own dampers, its state fluctuates with covariance $\propto \mathcal{E}^{-1}$.

For the observation noise of Section 1.2 this is only an analogy: that noise is added to the readout, and σ^2 is not tied to D, so only the shape matches. The morphological-computation metric of Section 5.2 uses the dynamical version: `channel.gibbs_pair` chooses the process noise so that $\mathcal{E}^{-1}$ is exactly stationary. (Residual of the Lyapunov equation on both leg models: 1.4×10^{-16}, `check_fdt`.)

Remark 2.4 (The Gaussian assumption is conservative). Diggavi and Cover (2001, Lemma II.2), a result they attribute to Ihara, states: if the input is Gaussian, then among all additive

noises with a given covariance the Gaussian one minimizes the mutual information.

The prior on w is Gaussian, and noise of covariance $\sigma^2\mathcal{E}^{-1}$ maps to noise of covariance $\sigma^2 I$ under the invertible map $\mathcal{E}^{1/2}$ (Lemma 2.1(a)). So the bits of Lemma 1.1 are a *lower bound* on the information under *any* noise law with the same energy-shaped covariance.

3 The dynamics in whitened coordinates

Status: stable · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

In whitened coordinates the equations of motion separate into a part that conserves energy and a part that dissipates it. That structure is behind several later results: the information-preservation property of Proposition 7.1, the stationarity used for MC_W in Section 5.2, and the uniform energy bound in the proof of Proposition 4.2.

Theorem 3.1 (Port-Hamiltonian form)**.** *Let $K \succ 0$, and let $z = \mathcal{E}^{1/2} x_q$. Then $\dot{z} = \tilde{A} z + \tilde{B} f$ with*

$$\tilde{A} = \mathcal{E}^{1/2} A \, \mathcal{E}^{-1/2} = \underbrace{\begin{bmatrix} 0 & \Omega \\ -\Omega^\top & 0 \end{bmatrix}}_{\mathcal{J} = -\mathcal{J}^\top} - \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & H^{-1/2} D H^{-1/2} \end{bmatrix}}_{\mathcal{R} = \mathcal{R}^\top \succeq 0}, \qquad \Omega = K^{1/2} H^{-1/2}, \qquad \tilde{B} = \begin{bmatrix} 0 \\ H^{-1/2} J^\top \end{bmatrix}. \tag{3.1}$$

Consequently:

- *(a)* Power balance: $\frac{d}{dt} \frac{1}{2} |z|^2 = -\dot{q}^\top D \dot{q} + (J\dot{q})\, f$. *The rate of change of energy is the power delivered at the foot minus the power dissipated.*
- *(b)* Contraction: $\|e^{\tilde{A} t}\|_2 \le 1$ *for* $t \ge 0$. *Free motion never increases the whitened norm.*
- *(c)* Natural frequencies: *the singular values of Ω are the undamped natural frequencies ω_i, the square roots of the eigenvalues of $K\phi = \omega^2 H \phi$.*
- *(d)* Modes are rotations: *if $D = 0$, each mode is a rotation at rate ω_i in a two-dimensional plane of z-space. The radius of the circle is $\sqrt{2 \times (\text{modal energy})}$.*

Proof. Computing $\tilde{A}$. With $\mathcal{E}^{1/2} = \mathrm{diag}(K^{1/2}, H^{1/2})$, multiply out $\mathcal{E}^{1/2} A \mathcal{E}^{-1/2}$ block by block using Equation 1.2:

- the $(1,1)$ block is $K^{1/2} \cdot 0 \cdot K^{-1/2} = 0$;
- the $(1,2)$ block is $K^{1/2} \cdot I \cdot H^{-1/2} = \Omega$;
- the $(2,1)$ block is $H^{1/2}(-H^{-1}K)K^{-1/2} = -H^{-1/2}K^{1/2} = -\Omega^\top$;
- the $(2,2)$ block is $H^{1/2}(-H^{-1}D)H^{-1/2} = -H^{-1/2}DH^{-1/2}$.

Also $\tilde{B} = \mathcal{E}^{1/2}B = [0; H^{1/2}H^{-1}J^\top] = [0; H^{-1/2}J^\top]$. The matrix $\mathcal{R}$ is positive semidefinite because it is congruent to $D \succeq 0$.

(a) Write $z = (z_q, z_v)$, so $z_v = H^{1/2}\dot{q}$. Then $\frac{d}{dt}\frac{1}{2}|z|^2 = z^\top \tilde{A} z + z^\top \tilde{B} f$. Now $z^\top \mathcal{J} z = 0$ because $\mathcal{J}$ is skew. The remaining terms are $-z_v^\top H^{-1/2} D H^{-1/2} z_v = -\dot{q}^\top D \dot{q}$ and $z_v^\top H^{-1/2} J^\top f = (J\dot{q}) f$.

(b) Take $f = 0$ in (a): $\frac{d}{dt}|z|^2 = -2z^\top \mathcal{R} z \le 0$. So $|e^{\tilde{A}t} z_0| \le |z_0|$ for every z_0.

(c) $\Omega\Omega^\top = K^{1/2}H^{-1}K^{1/2} = K^{1/2}(H^{-1}K)K^{-1/2}$ is similar to $H^{-1}K$. The eigenvalues of $H^{-1}K$ are the ω_i^2.

(d) Take the singular value decomposition $\Omega = P\Sigma U^\top$. In the coordinates $(P^\top z_q, U^\top z_v)$ the matrix $\mathcal{J}$ becomes $\begin{bmatrix} 0 & \Sigma \\ -\Sigma & 0 \end{bmatrix}$. This decouples into 2×2 blocks $\begin{bmatrix} 0 & \omega_i \\ -\omega_i & 0 \end{bmatrix}$, each generating a rotation at rate ω_i. The squared radius in block i is that mode's share of $|z|^2$, which is twice its energy by Lemma 2.2.

□

Numerical check

For the rigid leg and SEAs with $k_s \in \{70, 10^3, 10^4, 10^5, 10^6\}$:

- the symmetric part of $\tilde{A}$ equals $-\mathcal{R}$ to 7×10^{-16} relative error;
- the largest eigenvalue of $\tilde{A} + \tilde{A}^\top$ is $\le 3 \times 10^{-14}$;
- the singular values of Ω equal $2\pi \times$ `modes_hz()`. For example, at $k_s = 10^4$ they are 482.087 and 1197.605

rad/s (76.7 and 190.6 Hz).

(`check_port_hamiltonian`.)

Relation to the literature.

- *Port-Hamiltonian systems.* Equation 3.1 is the input-state-output port-Hamiltonian form $\dot{z} = [\mathcal{J} - \mathcal{R}]\,\partial\mathcal{H}/\partial z + gu$, $y = g^\top \partial\mathcal{H}/\partial z$ of Schaft and Jeltsema (2014, Eq. (4.10)), with Hamiltonian $\mathcal{H} = \frac{1}{2}|z|^2$ and $g = \tilde{B}$. Its natural output, $y = \tilde{B}^\top z = J\dot{q}$, is the foot velocity: the power conjugate of the unknown force.
  - Energy whitening is the change of coordinates that turns the storage function into the identity.
  - The storage function $\frac{1}{2}x_q^\top \mathcal{E} x_q$ is the classical one of dissipativity theory (Willems 1972).
- *Modal analysis and sensor placement (Li et al. 2007).* The velocity block $z_v = H^{1/2}\dot{q}$ is the mass weighting behind the *modal kinetic energy* (MKE) criterion for sensor placement. The eigenvectors of $\Omega^\top\Omega = H^{-1/2}KH^{-1/2}$ are the mass-normalized modes $H^{1/2}\phi_i$. The input in modal coordinates is the modal participation $\phi_i^\top J^\top f$.
  - As we understand it from secondary sources (Appendix A), the cited work relates effective independence (EfI) and MKE through iteration and re-orthonormalization, without showing that the two coincide under mass whitening.
  - EfI is the Fisher/D-optimal criterion of Kammer (1991). It assumes isotropic noise in *physical* coordinates.
  - The analogy is with MKE's mass weighting. It covers the kinetic block only, not the stiffness block.
- *Gramians and balancing (Moore 1981).* Moore (1981) observed that principal components of the controllability Gramian depend on coordinates: under $x \mapsto Sx$ the Gramian transforms as $W \mapsto SWS^\top$. He removed the dependence by *balancing*, i.e. choosing coordinates in which the controllability and observability Gramians are equal.

– Energy whitening removes the same dependence differently. By Proposition 2.1, it is the coordinate system in which the controllability Gramian of the *thermal* input is isotropic.
– The Gram matrix Equation 1.5 is a finite-horizon, input-restricted, Gramian-like object whose output inner product is fixed by $\mathcal{E}$.
– Balanced and energy coordinates generally differ. One does not justify the other.

4 Bodies with different numbers of coordinates

Status: stable · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

The SEA has one more coordinate than the rigid transmission. As $k_s \to \infty$ the two mechanisms become physically identical, so any sound information measure must give them the same value in that limit. This chapter shows that energy whitening does, that per-coordinate noise does not, and by exactly how much.

4.1 The duplicate-readout artifact

Proposition 4.1 (A free second look)**.** *Suppose an observation is duplicated with* independent *noise of the same size, i.e. the readout C is replaced by $[C; C]$. Then $G \mapsto 2G$, and*

$$\text{bits}(2G) - \text{bits}(G) = \tfrac{1}{2} \sum_i \log_2 \frac{1 + 2\lambda_i}{1 + \lambda_i},$$

where the λ_i are the eigenvalues of G. Each term increases with λ_i toward $\frac{1}{2}$, so the gain tends to $\frac{1}{2}$ bit per well-excited mode.

Proof. By Equation 1.5, $[C; C]^\top [C; C] = 2C^\top C$, so the Gram doubles. Then $\det(I + 2G)/\det(I + G) = \prod_i (1 + 2\lambda_i)/(1 + \lambda_i)$. The function $\lambda \mapsto \frac{1+2\lambda}{1+\lambda}$ increases from 1 toward 2. □

Numerical check

On the rigid leg with the per-coordinate readout, duplicating C adds 8.4716 bits. The formula gives 8.4716, and

$\frac{1}{2} \times 17$ modes $= 8.5$ (`check_duplicate`).

4.2 A general constraint-limit theorem

Let $K_\kappa = K_0 + \kappa K_c$ with $K_c \succeq 0$. Here κ is the stiffness of a stiff internal element, such as a series spring or a stiff joint. Let $V \in \mathbb{R}^{n \times n_V}$ have columns spanning $\ker K_c$: the motions that do not stretch the stiff element. Pick any complement W, so that every configuration has a unique decomposition $q = V\theta + W\delta$. The *constrained* ("locked") mechanism has coordinates θ and matrices

$$H_V = V^\top H V, \qquad K_V = V^\top K_0 V, \qquad D_V = V^\top D V, \qquad J_V = JV,$$

with energy metric $\mathcal{E}_V = \mathrm{diag}(K_V, H_V)$.

Lemma 4.1 (The energy form restricts to the locked mechanism, exactly)**.** *For every κ, every pair of locked states $x_V = (V\theta, V\dot{\theta})$ and $x'_V = (V\theta', V\dot{\theta}')$ satisfies*

$$\langle x_V,\, x'_V \rangle_{\mathcal{E}_\kappa} = \langle (\theta, \dot{\theta}),\, (\theta', \dot{\theta}') \rangle_{\mathcal{E}_V}.$$

A per-coordinate weight Λ instead restricts to $V^\top \Lambda V$, which counts separately every original coordinate that moves with θ.

Proof. For the configuration part, $(V\theta)^\top K_\kappa (V\theta') = \theta^\top V^\top K_0 V\theta' + \kappa\, \theta^\top V^\top K_c V\theta'$, and $K_c V = 0$ because V spans $\ker K_c$. For the velocity part, $(V\dot{\theta})^\top H (V\dot{\theta}') = \dot{\theta}^\top H_V \dot{\theta}'$. The claim about Λ is the same computation with Λ in place of $\mathcal{E}_\kappa$. □

Definition 4.1 (Assumption: Singular-perturbation convergence)**.** Let $x_k^{(\kappa)} = x_{V,k} + x_{\perp,k}$ be the decomposition of the sampled state, with $x_\perp = (W\delta, W\dot{\delta})$. For each input mode j and each sample k, as $\kappa \to \infty$:

(i) $(\theta, \dot{\theta})$ converges to the locked mechanism's response;
(ii) $\|x_{\perp,k}\|_{\mathcal{E}_\kappa} \to 0$.

Theorem 4.1 (Energy whitening commutes with the constraint limit)**.** *Under Definition 4.1,* $G_\kappa \to G_{\text{locked}}$ *and hence* $\text{bits}(G_\kappa) \to \text{bits}(G_{\text{locked}})$*. Under a per-coordinate weight* Λ*, the limit is instead the locked mechanism observed with weight* $V^\top \Lambda V$*.*

Proof. Fix two input modes i, j. Entry (i, j) of G_κ/γ is $\sum_k \langle x_k^i,\, x_k^j \rangle_{\mathcal{E}_\kappa}$. Expand each state as $x = x_V + x_\perp$:

$$\langle x^i,\, x^j \rangle_{\mathcal{E}_\kappa} = \underbrace{\langle x_V^i,\, x_V^j \rangle_{\mathcal{E}_\kappa}}_{(1)} + \underbrace{\langle x_V^i,\, x_\perp^j \rangle_{\mathcal{E}_\kappa} + \langle x_\perp^i,\, x_V^j \rangle_{\mathcal{E}_\kappa}}_{(2)} + \underbrace{\langle x_\perp^i,\, x_\perp^j \rangle_{\mathcal{E}_\kappa}}_{(3)}.$$

- By Lemma 4.1, term (1) is the locked-mechanism inner product of $(\theta^i, \dot\theta^i)$ and $(\theta^j, \dot\theta^j)$. By Definition 4.1(i) it converges to the corresponding entry of G_{locked}/γ, because the inner product is continuous.
- By Cauchy-Schwarz in the $\mathcal{E}_\kappa$ inner product (a semi-inner product suffices), $|(2)| \le \|x_V^i\|\|x_\perp^j\| + \|x_\perp^i\|\|x_V^j\|$. Here $\|x_V\|$ stays bounded, being convergent by (i) and Lemma 4.1, and $\|x_\perp\| \to 0$ by (ii). So (2) tends to 0.
- Likewise $|(3)| \le \|x_\perp^i\|\|x_\perp^j\| \to 0$.

The sums over k are finite, so every entry converges. Finally, $G \mapsto \log\det(I + G)$ is continuous on positive semidefinite matrices. The Λ statement follows from the second half of Lemma 4.1 by the same argument, provided the off-subspace motion vanishes in the Λ norm too. □

4.3 The SEA: Definition 4.1 holds, with explicit rates

For the clamped SEA, $q = (\theta_j, \theta_m)$:

- $H = \text{diag}(m_l r^2, I_{\text{rot}})$;
- $K_\kappa = k_i r^2 e_1 e_1^\top + k_s \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right]$, so $\kappa = k_s$;
- $D = b_i r^2 e_1 e_1^\top + b_s \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right]$ with $b_s = 2\zeta_s \sqrt{k_s I_{\text{rot}}}$;
- $J = (-r, 0)$.

Take $V = (1, 1)^\top$ and $W = (0, -1)^\top$, so that $\theta = \theta_j$ and $\delta = \theta_j - \theta_m$ is the spring deflection. Then:

- $H_V = m_l r^2 + I_{\text{rot}}$, $K_V = k_i r^2$, $D_V = b_i r^2$ and $J_V = -r$, which are exactly the rigid transmission's matrices;
- $\|x_\perp\|^2_{\mathcal{E}_\kappa} = k_s \delta^2 + I_{\text{rot}} \dot{\delta}^2$.

The redundant direction's whitened coordinates are therefore $\sqrt{k_s}\,\delta$ and $\sqrt{I_{\text{rot}}}\,\dot{\delta}$. Physically, $\sqrt{k_s}\,\delta = \tau_s/\sqrt{k_s}$: *the transmitted spring torque divided by* $\sqrt{k_s}$. The torque stays bounded, because it has to carry the load, so this coordinate vanishes like $k_s^{-1/2}$.

Proposition 4.2 (Rates for the SEA)**.** *Let the force f be bounded and integrable on $[0, T]$, which covers the zero-order-hold inputs used by the code, and let the system start from rest. Write $m_j = m_l r^2$, define the reduced inertia $I_{\text{red}} = m_j I_{\text{rot}}/(m_j + I_{\text{rot}})$, and set $\zeta' = \zeta_s \sqrt{I_{\text{rot}}/I_{\text{red}}}$, which is assumed to be < 1 (here $\zeta' = 0.115$). Then there is a constant A_*, independent of k_s, such that for all $t \in [0, T]$*

$$\sqrt{k_s}\,|\delta(t)| \le \frac{I_{\text{rot}} A_*}{\zeta' \sqrt{1-\zeta'^2}}\, k_s^{-1/2}, \qquad \sqrt{I_{\text{rot}}}\,|\dot{\delta}(t)| \le \frac{I_{\text{rot}}^{3/2} A_*\, c_2(\zeta')}{\sqrt{I_{\text{red}}}}\, k_s^{-1/2},$$

and $\theta_j \to \theta_{\text{rigid}}$, $\dot{\theta}_j \to \dot{\theta}_{\text{rigid}}$ uniformly on $[0, T]$. Here $c_2(\zeta) = (1 + \zeta/\sqrt{1-\zeta^2})/\zeta$. Hence Definition 4.1 holds, and by Theorem 4.1 the SEA's energy-whitened bits converge to the rigid transmission's.

Proof. The proof has five steps. Each uses only the two equations of motion and elementary bounds.

Step 1: a uniform energy bound. By Theorem 3.1(a) and Cauchy-Schwarz, $\frac{d}{dt}\frac{1}{2}|z|^2 \le |z|\,|\tilde{B}|\,|f|$. Whenever $|z| > 0$ this says $\frac{d}{dt}|z| \le |\tilde{B}||f|$. Integrating from rest,

$$|z(t)| \le \beta \int_0^T |f|\,dt =: Z_*, \qquad \beta = |\tilde{B}| = |H^{-1/2} J^\top| = r/\sqrt{m_j}.$$

What matters is that β involves only H and J, *neither of which depends on k_s*. Two consequences:

- $|z|^2 \ge k_i r^2 \theta_j^2$, so $|\theta_j| \le Z_*/(\sqrt{k_i}\, r)$;
- $|z|^2 \ge m_j \dot{\theta}_j^2$, so $|\dot{\theta}_j| \le Z_*/\sqrt{m_j}$.

Both bounds are uniform in k_s.

Step 2: the centre of inertia moves slowly. Add the two rows of the equations of motion. The spring and damper torques are internal, so they cancel:

$$m_j\ddot{\theta}_j + I_{\text{rot}}\ddot{\theta}_m = -rf - k_i r^2\theta_j - b_i r^2\dot{\theta}_j.$$

Define the centre-of-inertia acceleration $\bar{a} = (m_j\ddot{\theta}_j + I_{\text{rot}}\ddot{\theta}_m)/(m_j + I_{\text{rot}})$. By Step 1,

$$|\bar{a}| \le A_* := \frac{r\|f\|_\infty + k_i r^2\,\Theta_* + b_i r^2\,\Omega_*}{m_j + I_{\text{rot}}}, \qquad \Theta_* = \frac{Z_*}{\sqrt{k_i}\,r}, \quad \Omega_* = \frac{Z_*}{\sqrt{m_j}},$$

which is again uniform in k_s.

Step 3: the deflection is a stiff oscillator driven by $\bar{a}$. The motor row reads $I_{\text{rot}}\ddot{\theta}_m = k_s\delta + b_s\dot{\delta}$. Write $\ddot{\theta}_m = \bar{a} - \frac{m_j}{m_j+I_{\text{rot}}}\ddot{\delta}$, which is an identity from the definitions of $\bar{a}$ and δ. Substituting and rearranging gives

$$I_{\text{red}}\,\ddot{\delta} + b_s\dot{\delta} + k_s\delta = I_{\text{rot}}\,\bar{a}(t), \qquad \delta(0) = \dot{\delta}(0) = 0. \tag{4.1}$$

This is a damped oscillator with mass I_{red}, stiffness k_s, natural frequency $\omega' = \sqrt{k_s/I_{\text{red}}}$ and damping ratio $b_s/(2\sqrt{k_s I_{\text{red}}}) = \zeta'$, a constant.

Step 4: impulse-response bounds. The impulse response of $m\ddot{x} + b\dot{x} + kx = u$ with $\zeta < 1$ is $g(t) = e^{-\zeta\omega t}\sin(\omega_d t)/(m\omega_d)$, where $\omega_d = \omega\sqrt{1-\zeta^2}$. Bounding $|\sin| \le 1$ and integrating,

$$\int_0^\infty |g|\,dt \le \frac{1}{m\omega_d\,\zeta\omega} = \frac{1}{k\,\zeta\sqrt{1-\zeta^2}}.$$

Similarly $|\dot{g}| \le e^{-\zeta\omega t}(1 + \zeta/\sqrt{1-\zeta^2})/m$, so $\int|\dot{g}| \le c_2(\zeta)/(m\omega) = c_2(\zeta)/\sqrt{km}$.

The solution of Equation 4.1 is the causal convolution $\delta = g * (I_{\text{rot}}\bar{a})$, and $\dot{\delta} = \dot{g} * (I_{\text{rot}}\bar{a})$ because $g(0) = 0$. Young's inequality, $\|g * u\|_\infty \le \|g\|_1\|u\|_\infty$, then gives

$$|\delta| \le \frac{I_{\text{rot}}A_*}{k_s\,\zeta'\sqrt{1-\zeta'^2}}, \qquad |\dot{\delta}| \le \frac{I_{\text{rot}}A_*\,c_2(\zeta')}{\sqrt{k_s I_{\text{red}}}}.$$

Multiplying the first by $\sqrt{k_s}$ and the second by $\sqrt{I_{\mathrm{rot}}}$ gives the stated rates.

Step 5: the slow coordinate converges. Let $\bar{\theta} = (m_j\theta_j + I_{\mathrm{rot}}\theta_m)/(m_j + I_{\mathrm{rot}}) = \theta_j - \frac{I_{\mathrm{rot}}}{m_j+I_{\mathrm{rot}}}\delta$. The summed equation of Step 2 can be rewritten as

$$(m_j+I_{\mathrm{rot}})\ddot{\bar{\theta}}+b_i r^2\dot{\bar{\theta}}+k_i r^2\bar{\theta} = -rf+e(t), \qquad e = -\tfrac{I_{\mathrm{rot}}}{m_j+I_{\mathrm{rot}}}(k_i r^2\delta+b_i r^2\dot{\delta}).$$

With $e = 0$ this is exactly the rigid transmission's equation. By Step 4, $\sup|e| = O(k_s^{-1/2})$. The difference $\bar{\theta} - \theta_{\mathrm{rigid}}$ is the rigid system's response to e from rest. Its impulse response g_{r} and $\dot{g}_{\mathrm{r}}$ are integrable on $[0,T]$, so $\sup|\bar{\theta} - \theta_{\mathrm{rigid}}| \le \|g_{\mathrm{r}}\|_{L^1[0,T]} \sup|e| \to 0$, and likewise for the derivative. Finally $\theta_j = \bar{\theta} + O(\delta)$ and $\dot{\theta}_j = \dot{\bar{\theta}} + O(\dot{\delta})$. □

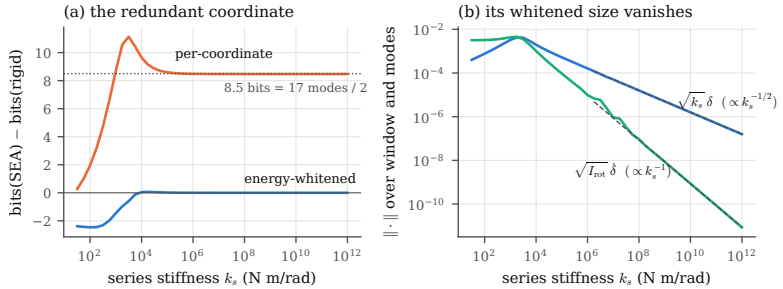


Figure 4.1: (a) Bits of the SEA minus bits of the rigid leg ($B_w = 100$ Hz, $d = 17$). Energy-whitened bits converge to the rigid leg's (the gap is 2.3×10^{-5} bits at $k_s = 10^8$). Per-coordinate bits converge to 8.4716 bits above it, exactly the duplicate-readout gain of Proposition 4.1. (b) The whitened size of the redundant direction, summed over samples and input modes. $\sqrt{k_s}\,\delta$ falls with fitted slope -0.500, as Proposition 4.2 predicts. $\sqrt{I_{\mathrm{rot}}}\,\dot{\delta}$ falls with slope -1.009, faster than the bound, because $\bar{a}$ is smooth and one more integration by parts gains another factor $k_s^{-1/2}$.

Why per-coordinate noise fails, exactly. For the readout $\Lambda = \mathrm{diag}(1, 1, \omega_{\mathrm{ref}}^{-2}, \omega_{\mathrm{ref}}^{-2})$ (the code's `"state"`), $V^\top\Lambda_q V = 2$ and

$V^\top \Lambda_v V = 2/\omega_{\mathrm{ref}}^2$. These are *twice* the rigid leg's per-coordinate weights. By Theorem 4.1 the stiff SEA is therefore the rigid leg observed twice with independent noise, and by Proposition 4.1 that is worth $\frac{1}{2}\sum_i \log_2 \frac{1+2\lambda_i}{1+\lambda_i} = 8.4716$ bits.

The numerical limit in Figure 4.1(a) is 8.4716 bits, the same number to four decimals. The analysis library has long pinned this excess, as 8.46 bits, in `test_a_locked_series_spring_is_informationally_a_rigid_transmission`; the argument above derives it.

The finite-k_s excess is real. Away from the limit, the spring stores energy ($\sqrt{k_s}\,\delta$ peaks near $k_s \approx 3 \times 10^3$, Figure 4.1(b)), and a second mode enters the force band.

- At $B_w = 200$ Hz, the band used in 1, the energy-whitened gap is +4.74, +7.03 and +8.57 bits at $k_s = 70$, 10^3 and 10^4 (second mode at 190.6 Hz), and −0.02 bits at 10^5. This is the peak in Fig. 1 of

  1.

- These bits are carried by energy the body really stores. By Lemma 2.2 the whitened size of the spring direction is $\sqrt{k_s}\,\delta$, the square root of twice the energy the spring stores. At finite k_s that energy is not zero, so the spring direction is a genuine observation, not a copy. As $k_s \to \infty$ it vanishes, and with it the extra bits (Theorem 4.1).

4.4 How the noise scale enters: two regimes

Proposition 4.3 (Low- and high-SNR asymptotics)**.** *Write the Gram as $G = \sigma^{-2} G_0$, where the noise-free part $G_0 = \sigma_w^2 \sum_k X_k^\top \mathcal{E} X_k$ has eigenvalues $\lambda_1 \geq \ldots \geq \lambda_d \geq 0$, r of them positive. In this proposition only, write $\gamma = \sigma^{-2}$ for the precision.*

(a) Low SNR (energy regime): $\mathrm{bits}(G) \leq \frac{\gamma \operatorname{tr} G_0}{2 \ln 2}$, *with equality to first order in γ. Moreover*

$$\frac{\gamma \operatorname{tr} G_0}{2 \ln 2} = \frac{1}{\sigma^2 \ln 2} \sum_k \mathbb{E}_w[\mathcal{V}(t_k)],$$

where $\mathcal{V}(t)$ is the stored mechanical energy. At low SNR, the bits are *an energy budget: expected stored energy summed over samples, in units of $\sigma^2 \ln 2$.*

(b) High SNR (counting regime): $\mathrm{bits}(G) = \frac{r}{2}\log_2 \gamma + \frac{1}{2}\sum_{i\le r}\log_2 \lambda_i + o(1)$ *as* $\gamma \to \infty$.

(c) Consequence for design comparisons. *If designs A and B have the same number r of excited modes, their gap tends to $\frac{1}{2}\log_2(\det_+ G_{0,A}/\det_+ G_{0,B})$, which is* independent of *σ. If $r_A \neq r_B$, the gap grows like $\frac{1}{2}(r_A - r_B)\log_2 \gamma$, and the noise scale decides it.*

Proof.

(a) Write bits $= \frac{1}{2}\sum_i \log_2(1+\gamma\lambda_i)$. Use $\log_2(1+u) \le u/\ln 2$, with equality to first order at $u = 0$. Then compute $\operatorname{tr} G_0 = \sigma_w^2 \sum_k \operatorname{tr}(X_k^\top \mathcal{E} X_k) = \sum_k \mathbb{E}_w |\mathcal{E}^{1/2} X_k w|^2$, since $\mathbb{E}[ww^\top] = \sigma_w^2 I$. By Lemma 2.2 the last expression is $2\sum_k \mathbb{E}_w \mathcal{V}(t_k)$.

(b) For $\lambda_i > 0$, $\log_2(1+\gamma\lambda_i) - \log_2(\gamma\lambda_i) = \log_2(1+1/(\gamma\lambda_i)) \to 0$. Zero eigenvalues contribute nothing.

(c) Subtract the expansions in (b).

□

Numerical check

Rigid leg vs SEA at $k_s = 10^4$ ($B_w = 100$ Hz):

- both Grams have full rank 17, with condition numbers 26.5 and 29.4;
- the predicted high-SNR gap is $\frac{1}{2}\log_2$ det-ratio $=$ 0.0494 bits;
- the computed gap is 0.0494 bits for every $\sigma/\sigma_{\mathrm{ref}} \le 1$, peaks at 0.057 in the crossover, and decays to 0 at low SNR, as (a) requires (Figure 4.2).

Across this range the absolute bits move from 191 to 1 while the gap stays put. The analysis library pins the same behaviour in `test_the_comparison_does_not_ride_on_the_precision_anchor`.

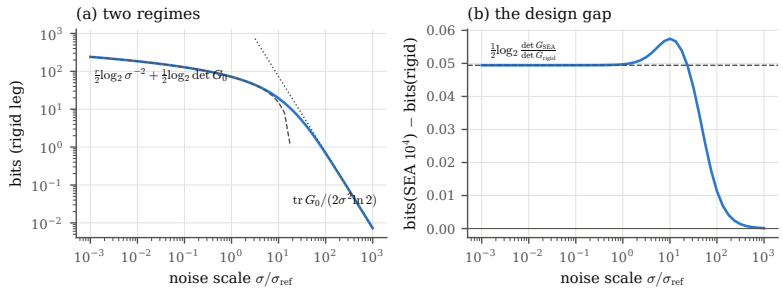


Figure 4.2: (a) Bits of the rigid leg against the noise scale, with the high-SNR counting asymptote (dashed) and the low-SNR energy asymptote (dotted) of Proposition 4.3. (b) The SEA(10^4) minus rigid gap equals the σ-independent log-determinant ratio throughout the high-SNR regime.

Remark 4.1 (Sampling rate). Noise is i.i.d. per sample, so one might fear that raising f_s adds bits for free, as in Proposition 4.1. Whether it does depends on how the force prior scales with f_s.

- *The convention used here is consistent.* The DCT modes are unit-norm over the samples (`mode_basis`), so a coefficient's force amplitude scales as $f_s^{-1/2}$, while the number of samples scales as f_s. The Gram becomes a Riemann sum for $\int_0^T x(t)^\top \mathcal{E} x(t)\, dt$, and bits converge as f_s grows: 71.62 → 71.72 → 71.77 bits at 2, 4 and 8 kHz (rigid leg).
- *The inconsistent alternative.* Holding the force waveform fixed instead adds 8.5 bits per doubling: 63.18 → 71.72 → 80.24. This is the duplicate-readout artifact in time.
- *The design gap is unaffected* in both conventions (equal rank, Proposition 4.3(c)).
- Results should therefore state that the φ_j are orthonormal over samples, or equivalently give the force prior in continuous L^2 (`check_sampling`).

5 Attribution and morphological computation

Status: stable · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

5.1 A consistent noise model for subsets of coordinates

A Shapley attribution, as in 1 [Eq. 10], assigns bits to coordinates through the game $\nu(S) = I(W; X_S)$. That needs a noise model for every subset X_S. The analysis library (`spectrum.coordinate_bits`) uses the marginal $(\sigma^2\mathcal{E}^{-1})_{SS}$ with a pseudo-inverse, and this section explains why that is the right choice.

Definition 5.1 (Observing a subset)**.** The noise lives on the *body*, not on a sensor: $\tilde{x}_q = x_q + \nu$ with $\nu \sim \mathcal{N}(0, \sigma^2\mathcal{E}^{-1})$ independently at each sample. Observing a subset S of coordinates means observing $P_S\tilde{x}_q$, where P_S selects rows. Its noise covariance is $N_S = \sigma^2(\mathcal{E}^{-1})_{SS}$.

Proposition 5.1 (Properties of the subset game)**.** *With $\nu(S)$ the information in $\{P_S\tilde{x}_q(t_k)\}_k$:*

(a) Consistency: *ν(all coordinates) equals the energy-whitened total.*
(b) Monotonicity: *$S \subseteq S'$ implies $\nu(S) \leq \nu(S')$. Hence every Shapley value is nonnegative, and they sum to the total (efficiency).*

(c) Exact duplicates are worth one copy: *if coordinate b's signal* and *noise are almost surely equal to coordinate a's, then $\nu(S \cup \{a,b\}) = \nu(S \cup \{a\}) = \nu(S \cup \{b\})$ for every S. Hence $\phi_a = \phi_b$, and the pair together earns what one of them earns.*

(d) Independent duplicates fuse by precision: *if $y_a = v + n_a$ and $y_b = v + n_b$ with n_a, n_b independent of each other and of v, with variances s_a^2 and s_b^2, then (y_a, y_b) carries exactly the information of one observation $v + \hat{n}$ with $1/\operatorname{Var}\hat{n} = 1/s_a^2 + 1/s_b^2$.*

Proof.

(a) With S everything, $N_S = \sigma^2 \mathcal{E}^{-1}$ is invertible. The Gaussian formula with colored noise uses the weight $N_S^{-1} = \mathcal{E}/\sigma^2$, which is Equation 1.5 with $C^\top C = \mathcal{E}$.

(b) $P_S \tilde{x}_q$ is a sub-vector of $P_{S'} \tilde{x}_q$, hence a function of it, and data processing applies. Every marginal contribution $\nu(S \cup \{i\}) - \nu(S)$ is therefore ≥ 0, so every Shapley value, being a weighted average of these, is ≥ 0. Efficiency is a standard property of the Shapley value.

(c) If $y_b = y_a$ almost surely, then (y_S, y_a, y_b) and (y_S, y_a) are functions of each other, so they carry the same information; the same holds with a and b swapped. The two players are interchangeable in every coalition, so the symmetry axiom gives $\phi_a = \phi_b$.

On the code side: the noise covariance of the pair is singular, $s^2 \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$. Its pseudo-inverse $\frac{1}{4s^2} \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$, applied to the signal (m, m), gives m^2/s^2, which is exactly the single-copy value. The pseudo-inverse thus computes the correct information on the range of the noise.

(d) Let $\hat{v} = (y_a/s_a^2 + y_b/s_b^2)/(1/s_a^2 + 1/s_b^2)$ and $u = y_a - y_b = n_a - n_b$.

- The map $(y_a, y_b) \mapsto (\hat{v}, u)$ is invertible.
- u does not involve v.
- $\operatorname{Cov}(\hat{v}, u) \propto s_a^2/s_a^2 - s_b^2/s_b^2 = 0$, so the jointly Gaussian pair $(\hat{v} - v, u)$ is independent, and u is independent of $(v, \hat{v})$.

Therefore $I(v; y_a, y_b) = I(v; \hat{v}, u) = I(v; \hat{v})$, and $\hat{v} = v + \hat{n}$ with $\operatorname{Var} \hat{n} = (1/s_a^2 + 1/s_b^2)^{-1}$. □

The stiff SEA has one redundancy of each kind.

- *Positions are exact duplicates in the limit.* The position noise is $\sigma^2 K_\kappa^{-1}$. As $k_s \to \infty$, the correlation of the noises on θ_j and θ_m goes to 1: it is 0.08, 0.30, 0.71, 0.995 and 0.99995 at $k_s = 70$, 10^3, 10^4, 10^6 and 10^8. The signals coincide too. By (c) the two positions earn equal Shapley credit, 22.948 and 22.947 bits at $k_s = 10^8$, and together they earn what the rigid leg's single angle earns (66.836 bits, both).
- *Velocities are independent duplicates.* H is diagonal, so the velocity noise $\sigma^2 H^{-1}$ is *uncorrelated*, with precisions $m_l r^2/\sigma^2$ and I_{rot}/σ^2. By (d) they fuse into one observation with precision $(m_l r^2 + I_{\text{rot}})/\sigma^2$, which is *exactly the rigid leg's velocity precision*, H_V/σ^2. Numerically, the pair gives 51.216 bits, as does the rigid leg's velocity.

  The players are *not* symmetric: $\dot{\theta}_j$ carries three times the precision of $\dot{\theta}_m$ ($m_l r^2 = 0.03$ vs $I_{\text{rot}} = 0.01$). Its Shapley share is correspondingly larger, 15.60 vs 10.23 bits.

1 [Fig. 2] states that the coordinates "are redundant at high stiffness, and so give equal Shapley contributions." That holds for positions; for velocities the credit is split by inertia (Figure 5.1). Both mechanisms are the restriction property of Lemma 4.1 at work on subsets.

5.2 Morphological computation MC_W

Zahedi and Ay (2013) define $\mathrm{MC}_W = I(W'; W \mid A)$ for discrete variables estimated from visited states. They do not assume an equilibrium distribution (we could read this only through secondary sources; Appendix A).

The linear version in the analysis library (`channel.morphological_bits`) adds a modelling assumption of its own. It takes the state law

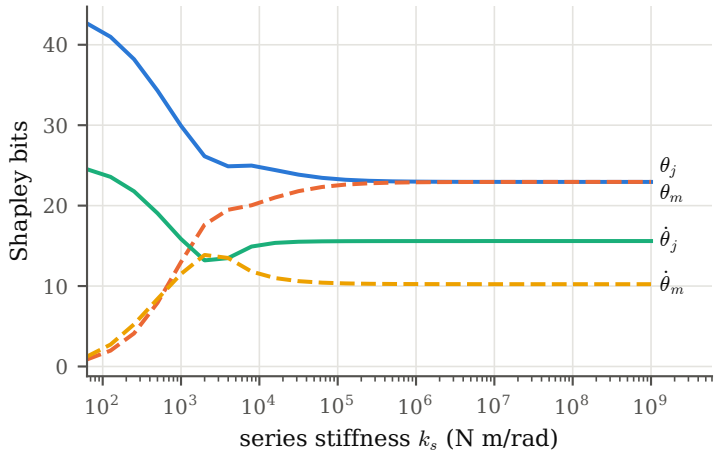


Figure 5.1: Shapley bits per SEA coordinate against k_s ($B_w =$ 100 Hz). Positions converge to equal shares, because their noise becomes perfectly correlated. Velocities converge to unequal shares fixed by the inertia ratio, because their noise stays independent and fuses by precision (Proposition 5.1).

to be the Gibbs law $\mathcal{E}^{-1}$ and chooses the one-step process noise $Q = P - A_d P A_d^\top$ so that P is stationary, with $A_d = e^{A\Delta t}$.

Proposition 5.2 (Closed form, and when it exists)**.** *Let* $\tilde{A}_d = \mathcal{E}^{1/2} A_d \mathcal{E}^{-1/2}$ *with singular values* s_i*.*

(a) $Q \succeq 0$ *if and only if* $\|\tilde{A}_d\|_2 \leq 1$*, and this always holds by Theorem 3.1(b), since* $\tilde{A}_d = e^{\tilde{A}\Delta t}$*.*

(b) Without action conditioning,

$$\mathrm{MC}_W = I(x_{k+1}; x_k) = -\tfrac{1}{2} \log_2 \det\left(I - \tilde{A}_d \tilde{A}_d^\top\right) = -\tfrac{1}{2} \sum_i \log_2(1 - s_i^2).$$

Proof. Mutual information is invariant under the invertible maps $x \mapsto \mathcal{E}^{1/2} x$ (Lemma 2.1(a)), so we may work in whitened coordinates.

(a) There P becomes I, and Q becomes $\tilde{Q} = I - \tilde{A}_d \tilde{A}_d^\top$. $\tilde{Q} \succeq 0$ exactly when every singular value of $\tilde{A}_d$ is at most 1.

(b) In whitened coordinates, $\tilde{x}_{k+1} = \tilde{A}_d \tilde{x}_k + \tilde{n}$ with $\operatorname{Cov} \tilde{x}_k = I$ and $\operatorname{Cov} \tilde{n} = \tilde{Q}$. Hence $\operatorname{Cov} \tilde{x}_{k+1} = \tilde{A}_d \tilde{A}_d^\top + \tilde{Q} = I$, and

$$I = \tfrac{1}{2} \log_2 \frac{\det \operatorname{Cov} \tilde{x}_{k+1}}{\det \operatorname{Cov} \tilde{n}} = -\tfrac{1}{2} \log_2 \det \tilde{Q}.$$

□

What this means physically

MC_W measures how close the whitened one-step map is to an *isometry*. Each lightly damped direction ($s_i \to 1$) contributes $-\frac{1}{2}\log_2(1-s_i^2)$, which is large. In plain terms, MC_W rewards *undissipated stored energy*: memory.

For the SEA: an extra mode with series damping ratio $\zeta_s = 0.1$ adds a nearly isometric direction. The largest s_i is 0.99999985 at $k_s = 70$ and 0.99995 for the rigid leg. So MC_W is 22.22 bits (SEA, $k_s = 70$) vs 8.56 bits (rigid), falling to 9.61 bits at $k_s = 10^6$. The closed form matches the code to 10^{-9} (`check_mcw_closed_form`).

This is the mechanism behind the tension reported in 1 [Sec. VI]. $I(W;X)$ rewards *transmitting force into* the state; MC_W rewards *keeping* energy in the state. A compliant, lightly damped body does more of the second and, at stiff contact, less of the first. The closed form also explains why MC_W diverges as $\Delta t \to 0$ ($s_i = 1 - O(\zeta_i \omega_i \Delta t)$) and why it requires $\mathcal{E} \succ 0$ (Section 6.3).

6 Continua, gears and floating bases

> **Status:** stable · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

6.1 Continuum and soft bodies: independence from the mesh

Soft robots and compliant links are modelled by finite elements, where the number of coordinates is a numerical choice. A sound information measure must converge as the mesh is refined.

Theorem 6.1 (Mesh independence)**.** *Let a linear elastic body be semi-discretized by conforming finite elements, with consistent mass H_N and stiffness K_N, and let $u_N(t)$ be the finite-element field. Then the energy-whitened Gram is*

$$G_N^{ij} = \gamma \sum_k \Big[a(u_{N,k}^i, u_{N,k}^j) + m(\dot{u}_{N,k}^i, \dot{u}_{N,k}^j) \Big],$$

where $a(\cdot,\cdot)$ is the continuum strain-energy bilinear form and $m(\cdot,\cdot)$ the continuum kinetic-energy (mass) form. Suppose the finite-element solutions converge in the energy norm at each sample time. This is classical for semi-discrete approximations of damped second-order hyperbolic problems with sufficiently regular data (Dupont 1973; Baker 1976; Geveci 1984); we assume *it here. Then $G_N \to G_\infty$ and the bits converge.*

By contrast, per-node isotropic noise has Gram entries $\approx \frac{N}{L} \times$ (L^2 inner product). It grows linearly in N, so by Proposition 4.3(b) the bits grow like $\frac{r}{2}\log_2 N$: about $\frac{r}{2}$ bits per mesh doubling.

Proof. The identity. With $u_N = \sum_i U_i \psi_i$ in the nodal basis, the finite-element matrices are defined by $U^\top K_N U' = a(u_N, u'_N)$ and $\dot{U}^\top H_N \dot{U}' = m(\dot{u}_N, \dot{u}'_N)$. Substituting into Equation 1.5 with $C^\top C = \mathcal{E}_N$ gives the displayed identity. The nodes have disappeared: the Gram is a sum of *continuum* energy inner products of the finite-element fields.

Convergence. Each entry is a finite sum (over finitely many k) of inner products, which are continuous in the energy norm. So energy-norm convergence of each field implies convergence of each entry. Then apply continuity of $\log\det(I + \cdot)$.

The per-node estimate. For a uniform mesh of spacing $h = L/N$ and smooth fields, $\sum_i U_i U'_i \approx h^{-1} \int uu'\, dx$, a Riemann sum. □

Remark 6.1 (Cylindrical noise is harmless here). In the continuum limit, isotropic noise in energy coordinates is white noise on the energy Hilbert space. That is a cylindrical measure, not a countably additive one. It does no harm here: the signal $\{Mw\}$ spans at most d dimensions, and Lemma 1.1 shows that the information depends only on the $d \times d$ matrix $M^\top M$. This is the noise-side counterpart of the requirement, stressed in function-space Bayesian inversion, that priors and inner products be defined so as to survive refinement:

- MCMC proposals that are robust under mesh refinement (Cotter et al. 2013);
- finite-element inner products weighted by the mass matrix (Bui-Thanh et al. 2013, sec. 3.3);
- see also Stuart (2010).

Numerical check

A fixed-free axial rod (length 0.2 m, $EA = 10^3$ N, $\rho A = 0.1$ kg/m, so the first mode is at 125 Hz) with consistent mass, stiffness-proportional damping and a tip force ($B_w = 400$ Hz, $d = 68$), for $N = 2, \ldots, 128$ elements (Figure 6.1, `check_rod`):

- energy-whitened bits converge, 297.93 → 302.109, changing by 0.004 between $N = 64$ and 128;

- per-node isotropic bits grow by 33.2 bits per doubling, against $\frac{r}{2} = 34$ predicted;
- a fixed tip sensor (a physical point sensor) also converges, which shows that the problem is noise *per coordinate*, not discretization.

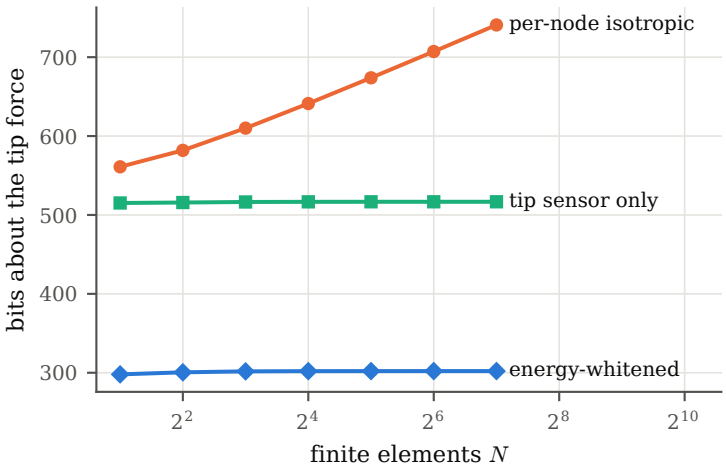


Figure 6.1: Mesh refinement of a finite-element rod. Energy whitening gives a mesh-independent answer. Per-node noise is the duplicate-readout artifact of Proposition 4.1 at scale: every refinement adds "free second looks."

6.2 Gear ratios and transmission design

Relabelling a motor angle on the input side of a gear as an angle on the output side is a reparametrization: $T = \mathrm{diag}(1, 1/N)$ in the SEA layout. By Theorem 2.1, energy-whitened bits do not move. Per-coordinate bits jump from 81.37 to 130.91 bits (Figure 2.1).

Changing the gear ratio is a different matter: it changes the reflected inertia $N^2 I_{\mathrm{rot}}$, which is a physical change in H. That change enters $\mathcal{E}$, and it correctly changes the bits. The energy metric separates bookkeeping from design, which is what

a sweep over rotor inertia (such as Table I of 1) needs.

6.3 Floating bases: when a singular metric loses nothing

For an unclamped leg or a free-flying quadruped, K is singular: rigid-body *positions* store no energy, while $H \succ 0$ still weighs velocities. The whitened readout $\mathcal{E}^{1/2}$, with eigenvalues clipped at zero by `sym_sqrt`, is blind to zero-energy positions. Does that lose information?

Proposition 6.1 (No loss when force and ground share the Jacobian)**.** *Suppose:*

(i) $K \succeq k_i J^\top J$ *with* $k_i > 0$*, where* J *is the Jacobian through which the unknown force acts;*
(ii) $D\phi = 0$ *for every* $\phi \in \ker K$*. Physically, internal and ground dampers act only on relative or foot motion.*

Then, from rest, the trajectory never enters $\ker K$*. More precisely,* $\phi^\top H q(t) = 0$ *for all* $\phi \in \ker K$ *and all* t*. On the subspace where the motion lives,* $\mathcal{E}$ *is positive definite, so nothing is lost.*

Proof. Step 1: zero-energy modes are not forced. For $\phi \in \ker K$, $0 = \phi^\top K \phi \geq k_i (J\phi)^2$, so $J\phi = 0$.

Step 2: their coordinates stay at zero. Let $\eta = \phi^\top H q$. Multiply Equation 1.1 on the left by $\phi^\top$:

$$\ddot{\eta} = (J\phi)^\top f - (D\phi)^\top \dot{q} - (K\phi)^\top q = 0.$$

With $\eta(0) = \dot{\eta}(0) = 0$, we get $\eta \equiv 0$.

Step 3: $\mathcal{E}$ is definite on the motion. By Step 2, $q(t)$ lies in the H-orthogonal complement of $\ker K$. Any nonzero q in that complement is not in $\ker K$, so $q^\top K q > 0$ there. □

Numerical check

On the analysis library's free bench (trunk plus leg, $q = (y_b, \theta)$), the energy-whitened and `"kinetic"` ($\omega_{\text{ref}}^2 H$) read-

outs give the same bits, 71.709.
The equality has a reason: the motion lives on $\mathrm{span}(H^{-1}J^\top)$. On that line, K restricts to $k_i(JH^{-1}J^\top)^2$ and H restricts to $JH^{-1}J^\top$. Their ratio is $k_i/m_e = \omega_{\mathrm{ref}}^2$, where m_e is the effective mass at the foot, so the two metrics agree on the motion.
A counterexample violates (i) by pushing on the trunk instead of the foot. The zero mode is then forced ($\phi_0^\top e_1 = -0.17$), and the singular metric loses $71.86 - 49.66 = 22.2$ bits relative to the regularized one (`check_floating`).

What this means physically

The quadruped experiments of 1 [Sec. V] apply random pushes w_{ext} to the calves of a floating quadruped. Those pushes do excite zero-stiffness base positions, so an energy metric built from joint springs alone would discard base-position information. Three remedies:

1. keep the base velocities (always weighted, since $H \succ 0$) and accept that base position is gauge;
2. regularize the position block with $\omega_{\mathrm{ref}}^2 H$ (the code's `"kinetic"` readout) and report sensitivity to ω_{ref};
3. use the potential of gravity and contact springs at the operating point.

The Zahedi-Ay MC_W of Section 5.2 has no Gibbs law when $\mathcal{E}$ is singular, which is why `morphological_bits` correctly refuses the free bench.

6.4 Further uses

- *Sensor placement.* Definition 5.1 turns "where should the encoder go?" into a Bayesian D-optimal design over physical sensor sites, with noise attached to the body. This is the energy-consistent analogue of effective independence (Kammer 1991; Papadimitriou 2004). Joint-side vs motor-side sensing on the SEA (Figure 5.1) is the simplest instance.

- *Variable-stiffness and parallel-elastic actuators, compliant feet.* All of these add or remove coordinates as parameters vary. Theorem 4.1 guarantees continuity across the degenerate limits.
- *Learned estimators.* For InfoNCE, whitening does not change the true information (Lemma 2.1(a)), but it gives the critic dimensionally homogeneous inputs in $\sqrt{\mathrm{J}}$. Per-feature z-scoring is a data-dependent diagonal metric: it is not basis-free, and it changes with the force prior. Noise must be added *after* whitening, as `sim_info` does.

Part II

Part II. Is energy the right choice?

7 Why energy

Status: draft · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

Part I took the energy metric as given and showed what it does. This chapter asks whether it is the right choice for measuring mechanical intelligence, and the next asks what it leaves out.

7.1 What maximum entropy does and does not decide

The proof of Theorem 2.2 never uses the fact that $\mathcal{E}$ is an energy. For *any* positive-definite form W, the same three steps show that the maximum-entropy law under $\mathbb{E}[\nu^\top W \nu] = m\sigma^2$ is $\mathcal{N}(0, \sigma^2 W^{-1})$. So maximum entropy turns a choice of *what is held fixed* into a noise law. It does not make that choice.

For example, constraining the mean *action*, i.e. each mode's energy divided by its frequency, gives the action-weighted law. That is the shape of quantum zero-point noise (Remark 7.2). The real modeling assumption is "resolution is measured in energy." The rest of this section justifies that assumption from properties a noise model should have.

7.2 Five requirements, and a uniqueness theorem

Definition 7.1 (Mechanisms and noise rules)**.** A mechanism is assembled from single-parameter elements:

- *springs*: stiffness k_e acting on a relative coordinate $s_e = a_e^\top q$;
- *inertias*: a positive-definite inertia matrix M_m acting on a velocity $u_m = B_m \dot{q}$;
- *dampers*: coefficient b_e acting on $\dot{s}_e$.

Hence $K = \sum_e k_e a_e a_e^\top$, $H = \sum_m B_m^\top M_m B_m$ and $D = \sum_e b_e a_e a_e^\top$. A *noise rule* assigns to every mechanism a positive-semidefinite quadratic form $W(q, \dot{q})$, the precision of the body's resolution: noise $\mathcal{N}(0, \sigma^2 W^{-1})$.

The requirements, each with the reason it matters:

- (U1) *Coordinate-free.* W transforms with the coordinates ($\tilde{W} = \mathcal{T}^\top W \mathcal{T}$ under $q = T\tilde{q}$), including rescalings of each element's own coordinate. Otherwise relabeling changes the answer (Theorem 2.1).
- (U2) *Local and universal.* W is a sum of one term per element. Each term depends only on that element's parameter and kinematics, through a function fixed for each element type. Adding a component somewhere must not change how precisely another component's state is known; this is what makes components compose, stiff limits behave (Theorem 4.1) and meshes converge (Theorem 6.1). Universality says the rule cannot know which spring is which.
- (U3) *No outside scales.* No dimensional constant other than the element's own parameter may appear: no reference frequency, length or time. Otherwise the result depends on an arbitrary choice; the `"kinetic"` readout's ω_{ref} can flip the sign of the SEA-rigid comparison (`test_the_ground_decides_whether_series_compliance_helps_or_hurts`).
- (U4) *Time-reversal symmetric.* $W(q, -\dot{q}) = W(q, \dot{q})$. Resolution should not depend on the direction of motion.
- (U5) *Unchanged by free motion.* For every mechanism without dampers, W is constant along every solution of $H\ddot{q} + Kq = 0$. Proposition 7.1 shows this is exactly the requirement that free motion never creates information.

Theorem 7.1 (Uniqueness of the energy metric)**.** *A noise rule satisfies (U1) to (U5) if and only if* $W = c\,(q^\top K q + \dot{q}^\top H \dot{q}) =$

$c\,x_q^\top \mathcal{E} x_q$ for a constant $c > 0$, i.e. it is the energy metric up to the overall noise scale.

Proof. The energy metric satisfies all five.

- (U1) is Theorem 2.1.
- (U2) to (U4) hold by inspection: $k_e s_e^2$ and $u_m^\top M_m u_m$ are energies.
- (U5) holds because energy is conserved without damping: $\frac{d}{dt}(q^\top K q + \dot{q}^\top H \dot{q}) = 2\dot{q}^\top (Kq + H\ddot{q}) = 0$.

Conversely, step 1: a spring. By (U2) its term is a quadratic form in its own kinematics $(s, \dot{s})$ with coefficients depending only on k: $\alpha_1(k)s^2 + \alpha_2(k)s\dot{s} + \alpha_3(k)\dot{s}^2$.

- (U4) gives $\alpha_2 = 0$.
- Rescaling the spring's coordinate, $s = \lambda\tilde{s}$, turns the stiffness into $\tilde{k} = \lambda^2 k$. By (U1) the term must keep its value: $\alpha_1(\lambda^2 k)\tilde{s}^2 = \alpha_1(k)\lambda^2\tilde{s}^2$ for every $\lambda > 0$, so $\alpha_1(k) = \alpha\, k$ for a dimensionless constant α. Likewise $\alpha_3(k) \propto k$.
- But $k\dot{s}^2$ has units of energy times time squared. Turning it into an energy needs a time scale, which (U3) forbids. So $\alpha_3 = 0$, and the spring's term is $\alpha\, k_e s_e^2$.

Step 2: an inertia. Its kinematics is a velocity u, and a position-dependent term would need a time scale (U3). So its term is $u^\top F(M)u$ for some matrix function F.

- (U1), applied to a change of the element's velocity coordinates $u = T\tilde{u}$ (which sends $M \mapsto T^\top M T$), requires $F(T^\top M T) = T^\top F(M) T$ for every invertible T.
- Taking $T = M^{-1/2}$ gives $F(M) = M^{1/2} F(I) M^{1/2}$.
- Taking T orthogonal gives $F(I) = T^\top F(I) T$ for every orthogonal T, so $F(I) = \beta I$ (a matrix that commutes with all rotations is a multiple of the identity).
- Hence the term is $\beta\, u_m^\top M_m u_m$.

Step 3: a damper. The same rescaling argument makes any term proportional to b. The candidates are $b\, s\dot{s}$, $b\, s^2$ and $b\, \dot{s}^2$:

- $b\, s\dot{s}$ is odd in time, which (U4) forbids;

- $b\,s^2$ has units of energy times time, and $b\,\dot{s}^2$ is a power; turning either into an energy needs a time scale, which (U3) forbids.

Dampers therefore contribute nothing.

Step 4: free motion fixes the ratio. By universality ((U2)) the constants α and β are the same for every element, so $W = \alpha\, q^\top K q + \beta\, \dot{q}^\top H \dot{q}$. Along any solution of $H\ddot{q} = -Kq$,

$$\tfrac{d}{dt} W = 2\alpha\, \dot{q}^\top K q + 2\beta\, \dot{q}^\top H \ddot{q} = 2(\alpha - \beta)\, \dot{q}^\top K q.$$

For any mechanism with a spring, some state has $\dot{q}^\top K q \neq 0$, so (U5) forces $\alpha = \beta =: c$. □

Remark 7.1 (What the axioms do and do not exclude).

- *Elements are single constitutive laws.* A Kelvin-Voigt element (spring and damper as one element) has an internal time scale b/k. If it were treated as one element, (U3) would allow a term $(b^2/k)\dot{s}^2$. Treating springs and dampers as separate elements, as their additive force laws suggest, removes this.
- *Universality matters.* Without it, (U5) still forces $\alpha = \beta$, but only within each mechanically connected component. Two disconnected bodies could carry different noise scales.
- *Dissipation drops out by itself.* No assumption singled out damping; the axioms exclude it from the metric.

Numerical check

The identity $\frac{d}{dt} W = 2(\alpha - \beta)\dot{q}^\top K q$ holds to 4×10^{-14} on random SEA states (`check_uniqueness_flow`). Over 20 ms of undamped SEA motion, the deviation of $\Phi^\top W \Phi$ from W is 4×10^{-13} for the energy and the frequency-weighted forms of Section 7.4, but 6.4 for per-coordinate noise and 1.7 for the kinetic metric (relative to $\max|W|$; `check_flow_invariance`).

7.3 Free motion cannot create information

The fifth requirement looks like a technicality. It is the most important one for attributing computation to a body.

Proposition 7.1 (Information is created only by work)**.** *Store information about a random variable Y in the body's state, $x(0)$, and let the body move freely ($f = 0$). Observe $x(t)$ through physical noise $\nu \sim \mathcal{N}(0, \sigma^2 W^{-1})$ independent of Y, and let $I_t = I(Y; x(t) + \nu)$. Write $\Phi(t)$ for the state-transition matrix.*

- *(i) If $D = 0$ and $\Phi^\top W \Phi = W$ (for example $W = \mathcal{E}$), then $I_t = I_0$ for all t.*
- *(ii) If $W = \mathcal{E}$ and $D \succeq 0$, then I_t is non-increasing in t.*
- *(iii) If $D = 0$ and $\Phi(t)^\top W \Phi(t) \neq W$ for some t, then there is a Gaussian scalar Y with $I_t > I_0$: free motion creates apparent information.*

Consequently, among quadratic noise models, (U5) holds if and only if *free conservative motion never creates information.*

Proof. Work in the coordinates $z = W^{1/2} x$, where the noise is $\mathcal{N}(0, \sigma^2 I)$ and the flow is $\tilde{\Phi} = W^{1/2} \Phi W^{-1/2}$. Note that $\tilde{\Phi}^\top \tilde{\Phi} = W^{-1/2} \Phi^\top W \Phi W^{-1/2}$.

(i) If $\Phi^\top W \Phi = W$, then $\tilde{\Phi}$ is orthogonal. The observation $\tilde{\Phi} z(0) + n$ is an invertible transformation of $z(0) + \tilde{\Phi}^\top n$, and $\tilde{\Phi}^\top n \sim \mathcal{N}(0, \sigma^2 I)$ has the same law as n. So the two observations carry the same information about Y (Lemma 2.1(a)).

(ii) With $W = \mathcal{E}$, $\tilde{\Phi}(s) = e^{\tilde{A} s}$ is a contraction, $\|C\|_2 \leq 1$, by Theorem 3.1(b). Fix t and $s \geq 0$, and let $C = e^{\tilde{A} s}$, so that $z(t + s) + n = C z(t) + n$. Let $n_1 \sim \mathcal{N}(0, \sigma^2 I)$ and $n_2 \sim \mathcal{N}(0, \sigma^2 (I - C C^\top))$ be independent; the second covariance is valid because $\|C\| \leq 1$. Then $C(z(t) + n_1) + n_2$ has noise covariance $\sigma^2 (C C^\top + I - C C^\top) = \sigma^2 I$, so it has the same joint law with Y as $C z(t) + n$. It is obtained from the earlier observation $z(t) + n_1$ by further processing plus independent noise. By the data-processing inequality, $I_{t+s} \leq I_t$.

(iii) Let $R = \tilde{\Phi}^\top \tilde{\Phi} \neq I$; it is symmetric positive definite. Liouville's theorem gives $\det \Phi(t) = e^{t \operatorname{tr} A} = 1$, because $\operatorname{tr} A = -\operatorname{tr}(H^{-1}D) = 0$ when $D = 0$. Hence $\det R = \det(\Phi)^2 = 1$. A positive-definite matrix with determinant 1 that is not the identity has an eigenvalue $\lambda_{\max} > 1$ (eigenvalues whose product is 1 cannot all be ≤ 1 unless all equal 1). Let u be its eigenvector, and store $Y \sim \mathcal{N}(0, \sigma_Y^2)$ as $x(0) = W^{-1/2} u\, Y$. Then

$$I_0 = \tfrac{1}{2} \log\Big(1 + \tfrac{\sigma_Y^2}{\sigma^2} |u|^2\Big) < \tfrac{1}{2} \log\Big(1 + \tfrac{\sigma_Y^2}{\sigma^2} \lambda_{\max} |u|^2\Big) = I_t.$$

□

What this means physically

Under the energy metric, a body's information about anything it has stored can change in only two ways: the force does work on it, which can add information, or the dampers dissipate it, which can only remove information. Conservative motion only moves information around, as a rotation in whitened coordinates.
Under any other quadratic noise model, a free oscillation that rotates the state into better-resolved directions creates bits from nothing. A measure of mechanical computation should credit the body for what the force delivered and debit it for what was dissipated, not for coordinate artifacts of free motion. This is the same accounting as in the thermodynamics of computation (Landauer 1961; Bennett 1982): conservative dynamics are reversible and lose nothing, and dissipation is the irreversible step.

Remark 7.2 (Classical versus quantum: why energy and not action). The quantum fluctuation-dissipation theorem (Callen and Welton 1951) gives each mode of frequency ω a fluctuation energy $\frac{\hbar\omega}{2} \coth \frac{\hbar\omega}{2k_B T}$. In whitened coordinates this is the noise variance per mode.

- When $k_B T \ll \hbar\omega$ it tends to $\frac{1}{2}\hbar\omega$: noise energy grows with frequency, which is action weighting (precision $\propto 1/\omega$).

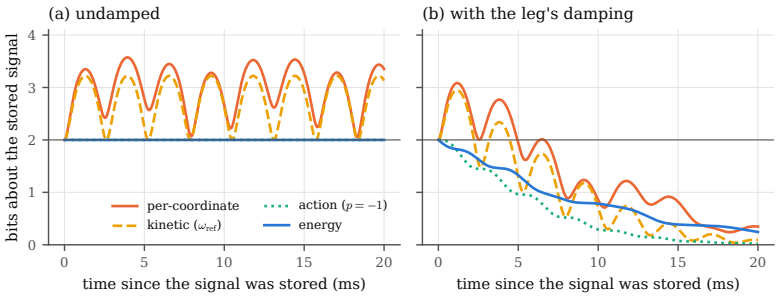


Figure 7.1: Information about a signal stored in the SEA's state ($k_s = 10^4$), under free motion, for four noise models. Each starts from the direction it resolves worst, scaled to 2 bits. (a) Undamped: energy and action weighting stay at exactly 2 bits. Per-coordinate noise reaches 3.58 bits and the kinetic metric 3.22 bits, created by free motion alone. (b) With the leg's damping, energy can only decrease (Proposition 7.1(ii)). Per-coordinate noise still rises to 3.08 bits first (`check_free_motion`).

- When $k_B T \gg \hbar\omega$ it tends to $k_B T$: equal energy per mode, the energy metric.

For a 200 Hz mode at room temperature, $\hbar\omega / k_B T \approx 3 \times 10^{-11}$, so macroscopic mechanisms sit deep in the classical regime. Physics and the locality axiom therefore agree that, for robots, resolution should be measured in energy.

Energy is also the additive conserved quantity associated with time-translation symmetry (Noether), which is why it can be assembled element by element.

7.4 The alternatives compared

Table 7.1 lists the candidate noise models, the requirements each violates, and how it behaves in the two limits of Part I. The frequency-weighted family is

$$W_p = \sum_i \omega_i^p \,(\text{modal energy form } i) = \mathcal{E}^{1/2}\,\mathrm{diag}\big((\Omega\Omega^\top)^{p/2}, (\Omega^\top\Omega)^{p/2}\big)\mathcal{E}^{1/2},$$

with $\Omega = K^{1/2}H^{-1/2}$ as in Theorem 3.1. Here $p = 0$ is the energy metric and $p = -1$ is action weighting. (For numerical comparison each W_p is divided by ω_{ref}^p so that it is measured in joules; this constant cancels in every design gap.)

Table 7.1: Candidate noise models. The stiff-limit column is SEA minus rigid bits at $k_s = 10^{10}$ ($B_w = 100$ Hz). The mesh column is the change from $N = 32$ to $N = 64$ elements on the rod of Theorem 6.1.

Rule	Violates	Stiff limit (bits)	Mesh, $N = 32 \to 64$ (bits)
energy, $W_0 = \mathcal{E}$	none	0.0000	+0.017, converges
action, W_{-1}	U2	0.0000	+0.010, converges
W_{+1}	U2	0.0007, converging slowly	+2.96, keeps growing
W_{+2}	U2	1.489, no limit	
W_{+3}	U2	60.1, diverges	
kinetic, $\mathrm{diag}(\omega_{\mathrm{ref}}^2 H, H)$	U3, U5	0.0000	+0.018, converges
per-coordinate (`"state"`)	U1, U2, U3, U5	8.4716	+33.3, keeps growing
data-driven whitening	not built from the body	erases the body (Proposition 7.2)	
Gramian or balanced coordinates	not built from the body		

Rule	Violates	Stiff limit (bits)	Mesh, $N = 32 \to 64$ (bits)
encoder on the motor (`"motorsensor"`)	answers a different question (the motor)		

The frequency-weighted family. Every W_p is coordinate-free (spread $< 10^{-13}$ over the reparametrizations of Figure 2.1) and flow-invariant (`check_flow_invariance`). Each is non-local, however: it is a matrix function of $H^{-1}K$, so every element influences every entry.

- *Locality, concretely.* Changing the *ground* stiffness k_i from 10^5 to 10^7 changes the action metric's *motor-velocity* entry by a factor of 2.2. The energy metric's entry stays exactly at I_{rot} (`check_locality`).
- *The stiff limit decides $p < 2$.* By Proposition 4.2 the fast mode's energy is $O(k_s^{-1})$, while its frequency, and so its weight, grows like $k_s^{p/2}$. Its weighted contribution therefore behaves like $k_s^{p/2-1}$, which vanishes if and only if $p < 2$. This is the argument of Theorem 4.1 with a weighted norm; we state it as a prediction rather than a theorem, because the slow-mode part is only checked numerically. The numbers confirm it exactly (Figure 7.2(a)):
  - $p \le 0$ converges;
  - $p = 1$ converges at the predicted rate: the gap falls tenfold, from 7.1×10^{-3} to 7×10^{-4}, per factor 100 in k_s;
  - $p = 2$ stalls at 1.49 bits;
  - $p = 3$ diverges.
- *The mesh decides $p \le 0$.* On the rod, $p = -1$ and $p = 0$ converge, and $p = +1$ keeps growing by about 3 bits per doubling (Figure 7.2(b)).

So action weighting is the one serious competitor: it passes both limit tests. What excludes it is locality and, physically,

the classical regime (Remark 7.2).

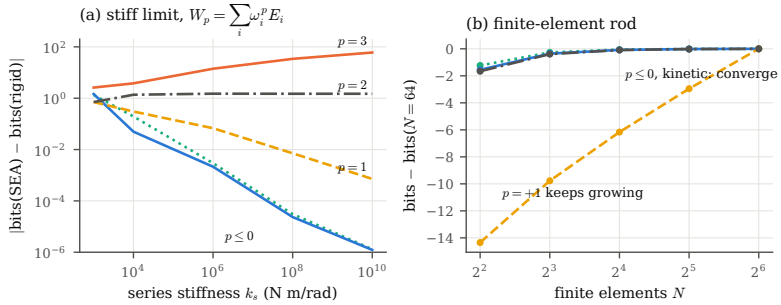


Figure 7.2: (a) Stiff limit of the frequency-weighted metrics W_p: SEA minus rigid bits against k_s. The gap vanishes for $p < 2$, stalls at $p = 2$ and diverges at $p = 3$, as predicted. (b) Mesh refinement of the rod, relative to $N = 64$. Energy, action and the kinetic metric converge; $p = +1$ does not (`check_alternatives_sea`, `check_alternatives_rod`).

Proposition 7.2 (Data-driven whitening erases the body)**.** *Suppose the stacked response $y = Mw$ (M of full column rank d) is whitened by its own covariance, $(\sigma_w^2 MM^\top)^{+1/2}$ on the range of M, and isotropic noise σ_n is then added. Then the information is $\frac{d}{2}\log_2(1+\sigma_n^{-2})$ for* every *body.*

Proof. Write the thin SVD $\sigma_w M = U\Sigma V^\top$. The whitener is $U\Sigma^{-1}U^\top$, and the whitened signal is $U\Sigma^{-1}U^\top U\Sigma V^\top(w/\sigma_w) = UV^\top(w/\sigma_w)$. Its Gram matrix is $\sigma_n^{-2}VU^\top UV^\top = \sigma_n^{-2}I_d$. By Lemma 1.1, the bits are $\frac{1}{2}\log_2\det((1+\sigma_n^{-2})I_d)$. □

What this means physically

Whitening by the data, i.e. z-scoring or ZCA on the observed response, makes every direction equally informative by construction, so the body's contribution vanishes. The check gives 56.59 bits for the rigid leg and for SEAs at 70 and 10^4, equal to the prediction (`check_data_driven_whitening`). A noise model has to be built from the body.

8 A thermodynamically consistent body: Langevin noise

Status: draft · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

The noise of Part I is added to the *readout*, independently at each sample. A body that is physically noisy is driven by thermal forces matched to its own dampers, and its noise is shaped by its dynamics. This chapter builds that model and asks whether the conclusions of Part I survive.

The model. The equation of motion becomes

$$H\ddot{q}+D\dot{q}+Kq = J^\top f(t)+\xi(t), \qquad \mathbb{E}[\xi(t)\xi(s)^\top] = 2\sigma_T^2 D\,\delta(t-s),$$

started from the Gibbs law $x_q(0) \sim \mathcal{N}(0, \sigma_T^2\mathcal{E}^{-1})$ on the clamped bench. By Proposition 2.1 the noise is stationary at covariance $P = \sigma_T^2\mathcal{E}^{-1}$. Sampling with step Δt gives $x_{k+1} = A_d x_k + B_d f_k + w_k$ with w_k i.i.d. $\mathcal{N}(0, Q_d)$ and $Q_d = P - A_d P A_d^\top$; this is `gibbs_pair`, scaled by σ_T^2. We set $\sigma_T^2 = \sigma^2 = 10^{-4}$ J, so a snapshot of the noise has exactly the covariance used in Part I. Only its correlation over time differs.

Proposition 8.1 (Full observation: an exact formula)**.** *If the full state is observed without readout noise, the information about the force profile is*

$$\text{bits} = \tfrac{1}{2}\log_2\det(I_d + \sigma_w^2\, g\,\bar{\Phi}^\top\bar{\Phi}), \qquad g = B_d^\top Q_d^{-1} B_d,$$

where $\bar{\Phi}$ holds the force basis at the samples that drive the state. The design enters only through the scalar g, the per-sample signal-to-noise ratio of a unit force against one thermal kick.

Proof. Step 1: change variables. The map from $(x_0, x_1, \dots, x_N)$ to $(x_0, e_0, \dots, e_{N-1})$, with $e_k = x_{k+1} - A_d x_k$, is invertible. So the two carry the same information (Lemma 2.1(a)).

Step 2: what the new variables contain. x_0 is independent of the force, and $e_k = B_d f_k + w_k$, where the w_k are i.i.d., independent of x_0, and Gaussian with covariance Q_d. This is a Gaussian channel with noise covariance $I \otimes Q_d$.

Step 3: apply the Gaussian formula. Whiten the noise by $Q_d^{-1/2}$ and apply Lemma 1.1 with $f_k = \sum_j w_j \bar{\Phi}_{kj}$. The Gram matrix is $\sigma_w^2 \sum_k \bar{\Phi}_k^\top (B_d^\top Q_d^{-1} B_d) \bar{\Phi}_k$, where $\bar{\Phi}_k$ is row k of $\bar{\Phi}$. Since g is a scalar, this equals $\sigma_w^2 g\, \bar{\Phi}^\top \bar{\Phi}$. □

Proposition 8.2 (Only the dissipation at the contact matters)**.** *Suppose $J^\top$ lies in the range of D. As $\Delta t \to 0$,*

$$g \;\to\; \frac{\Delta t}{2\sigma_T^2}\, J D^+ J^\top.$$

With the full state observed, the information therefore depends only on $JD^+J^\top$: the damping seen at the point where the force acts. For the ground model used here, $JD^+J^\top = 1/b_i$ for the rigid transmission and for every SEA, so every design gives the same bits.

Proof sketch. Continuous time. In the continuous-time limit, observing the state exactly is a white-Gaussian-noise channel for the drift Bf, with noise intensity $GG^\top = \mathrm{diag}(0,\, 2\sigma_T^2 H^{-1} D H^{-1})$. In the low-signal limit, where the causal estimate of the signal is negligible, Duncan's theorem (Duncan 1970) gives the information rate $\frac{1}{2}\mathbb{E}[(Bf)^\top (GG^\top)^+ (Bf)]$ nats per second.

The rate. Substituting $B = [0; H^{-1}J^\top]$ gives

$$\tfrac{1}{2}\mathbb{E}[f^2]\, J H^{-1} \big(2\sigma_T^2 H^{-1} D H^{-1}\big)^+ H^{-1} J^\top = \frac{\mathbb{E}[f^2]}{4\sigma_T^2}\, J D^+ J^\top.$$

Matching to the discrete channel. Over one sample, the discrete channel carries $\frac{1}{2}\log(1 + g\,\mathbb{E}[f^2]) \approx \frac{1}{2} g\,\mathbb{E}[f^2]$ nats about a force sample. The continuous rate gives $\Delta t\, \mathbb{E}[f^2]\, JD^+J^\top/(4\sigma_T^2)$ over

the same interval. Equating the two gives the stated limit for g, which does not depend on the signal amplitude.

The design-independence.

- For the rigid leg, $D = b_i r^2$ and $J = -r$, so $JD^+J^\top = 1/b_i$.
- For the SEA, $D = \begin{bmatrix} b_s + b_i r^2 & -b_s \\ -b_s & b_s \end{bmatrix}$ has $\det D = b_s b_i r^2$, so $(D^{-1})_{11} = b_s/(b_s b_i r^2) = 1/(b_i r^2)$. With $J = (-r, 0)$, this gives $JD^{-1}J^\top = 1/b_i$, independently of k_s and b_s.

What is not proved here. We do not prove that the discrete position-block terms leave the constant unchanged. The numbers below confirm it. □

Numerical check

- $2\sigma_T^2 g/(\Delta t\, JD^+J^\top)$ is 1.0000 at 4, 16 and 64 kHz for the rigid leg and for SEAs at $k_s = 70$, 10^4 and 10^6 (shown as $g\sigma_T^2 f_s/JD^+J^\top = 0.5000$ by `check_langevin_dt_limit`). It deviates only for coarse sampling (0.19 at 100 Hz).
- The exact formula of Proposition 8.1 and a direct computation with the full stacked noise covariance agree, 37.74216 bits (`check_langevin_innovations`).
- At the settings of Section 1.2, *every* design gives 37.742 bits with the full state observed.

What this means physically

This is the thermal limit of force sensing. The force must be read against the fluctuating force of the damper at the contact, whose strength is set by the fluctuation-dissipation theorem (the mechanical analogue of Johnson-Nyquist noise (Nyquist 1928)). If the whole state is observed, the body's inertia and stiffness are irrelevant: no morphology can beat the contact-damper limit.
The body's design matters only when the observation is *partial*, which is always the case for a real robot.

Partial observation: what morphology does. With noise injected through the dampers, observing coordinates C with readout noise ε gives the Gram matrix $\sigma_w^2 M_C^\top (C\Sigma C^\top + \varepsilon^2 I)^{-1} M_C$. Here Σ is the stacked noise covariance, with blocks $\mathrm{Cov}(x_k, x_l) = A_d^{k-l} P$ for $k \geq l$ (`langevin_bits`). By data processing, observing a subset can never exceed the full-state value. So 37.742 bits is a hard ceiling for every design and every sensor.

Table 8.1: Bits about the force profile ($B_w = 100$ Hz) under the Langevin body for several readouts, and under the readout-noise models of Part I (`check_langevin_designs`). Efficiency is motor bits divided by full-state bits.

Design	Langevin: full state	Langevin: motor	Efficiency	Langevin: motor + encoder	Readout noise: energy	Readout noise: motor
rigid	37.742	37.695	0.999	35.707	71.723	66.836
SEA, $k_s = 70$	37.742	2.266	0.060	1.637	69.280	24.379
SEA, 10^3	37.742	18.628	0.494	17.925	70.267	65.892
SEA, 3×10^3	37.742	29.038	0.769	28.194	71.084	72.780
SEA, 10^4	37.742	35.067	0.929	33.820	71.773	69.414
SEA, 10^5	37.742	37.242	0.987	35.631	71.742	67.104
SEA, 10^6	37.742	37.537	0.995	35.704	71.725	66.863

Three findings (Table 8.1; Figure 8.1):

1. *The rigid transmission routes almost everything to its motor encoder.* Through the motor angle alone it captures 99.9% of the thermal ceiling. The SEA's efficiency rises monotonically with stiffness: 6% at the prototype's

$k_s = 70$, 49% at 10^3, 93% at 10^4, 99.5% at 10^6. This is the argument of Wensing et al. (2017) for proprioceptive actuation, derived here from first principles: the rigid transmission delivers contact information to the actuator's own sensor without loss.

2. *The readout-noise model can rank designs wrongly.* With fixed encoder noise added to the readout, the SEA at $k_s = 3 \times 10^3$ beats the rigid leg by 5.9 bits through the motor encoder, and at 10^4 by 2.6 bits. Under Langevin noise this cannot happen: the rigid leg is already within 0.05 bits of the ceiling, so no design can beat it by more.

   The difference is resonance. A passive resonance amplifies the force signal at the motor, and with readout-only noise that amplification looks like free SNR. Under fluctuation-dissipation the same resonance amplifies the thermal force noise by exactly the same factor, so it gains nothing. This is the duplicate-readout artifact again, now in the frequency domain.

3. *With full-state readout, the two models agree qualitatively.* Langevin noise plus energy readout noise gives 36.2 to 36.4 bits for every design. The energy-whitened readout model of Part I gives 69.3 to 71.8. Both say the full state hardly separates the designs, whereas the motor encoder separates them strongly.

Caveats.

- σ_T^2 is an effective temperature, chosen to match Part I.
- The ground damping b_i is a modeling choice ($\zeta_{\text{contact}} = 0.15$), identical across designs by construction (the analysis library checks this in `test_one_ground_for_every_design`). The common ceiling is therefore a property of the ground model.
- Damping is modeled as constant and viscous. The model needs a positive-definite $\mathcal{E}$, so it applies to the clamped bench only.

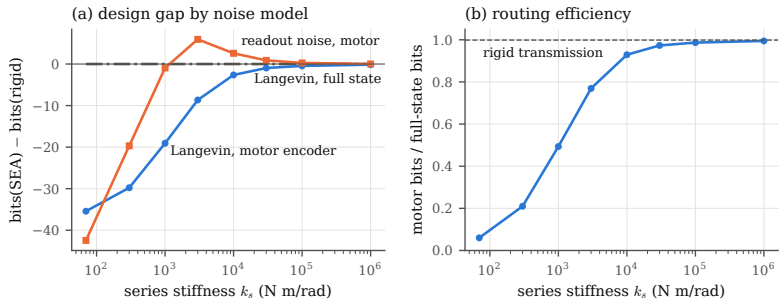


Figure 8.1: (a) SEA minus rigid bits through different noise models. Langevin noise with the full state observed gives zero, to within 10^{-5} bits, for every design. Through the motor encoder, Langevin noise gives a monotone climb to the rigid leg, while readout-only noise lets the SEA overtake the rigid leg near 3×10^3. (b) Routing efficiency under Langevin noise: the fraction of the thermal ceiling that reaches the motor encoder.

9 Open problems

Status: draft · **Updated:** 2026-11-21 · **Checked by:** `tests/test_energy_whitening.py`

This chapter collects what the energy metric settles, what it leaves open, and where the current analysis is limited. The research roadmap turns these into threads with concrete next steps, and research notes record progress on them.

9.1 What energy whitening does and does not capture

1. *Retention is settled.* For the question "how much does the body itself hold about the world," energy is the unique admissible resolution (Theorem 7.1). Every information quantity built on it is coordinate-free, composes across components, converges in constraint and refinement limits, and changes only through work and dissipation (Proposition 7.1).
2. *Routing is what morphology does.* In a thermodynamically consistent body the total is capped by the dissipation at the contact (Proposition 8.2). Designs differ in how much of that total reaches the coordinates a controller can read, at a latency it can use. The efficiency in Table 8.1 is a candidate scalar measure of mechanical intelligence in this sense, and the temporal and frequency decompositions of Part I apply to it unchanged.
3. *Relevance is not captured.* Neither retention nor routing asks whether the information is about what the task needs. The conditional information $I(W; X \mid A)$ of 1 measures relevance to the action, and task-specific variables,

such as the force component that matters for balance, would sharpen it.

4. *Nonlinear and multistable bodies.* The framework extends directly: the noise law becomes the Gibbs law $\exp(-V/T)$ on a non-quadratic potential. A mechanical memory then stores a reliable bit when its energy barrier is large compared with the noise energy, the classical criterion behind Landauer's bound (Landauer 1961). This connects the analysis to mechanical metamaterials and latching mechanisms, though none of it is tested yet.
5. *Validation must be predictive.* A definition is checked by its consequences. The consistency checks pass: invariance, the limits, and agreement with the bandwidth argument of Wensing et al. (2017). The empirical test is whether retention, routing or relevance best predicts closed-loop performance across designs; Table I of 1 gives a first hint.

9.2 Limitations

1. *Body versus sensor.* The energy noise model quantifies what the body retains. Real encoders have per-coordinate noise, and an encoder readout answers a different question. The two should be reported as distinct quantities; Chapter 8 shows how to combine them consistently.
2. *The environment lives in the metric.* K includes the ground spring, so $\mathcal{E}$ differs between stance and flight, and the ground stiffness can flip the sign of the SEA-rigid comparison (the analysis library shows this in `test_the_ground_decides_whether_series_compliance_helps_or_hurts`). That is physically right, since the ground does store energy, but any result should say so.
3. *Dissipation is not stored energy.* D is absent from $\mathcal{E}$, as it should be. Materials with internal variables, such as viscoelastic Maxwell elements, add states whose stored energy has to be defined first.
4. *Nonlinearity and contact.* $H(q)$ varies, contact switches K, and gravity preloads the stiffness. Remark 2.2 outlines a trajectory-wise extension. Hybrid contact needs a

metric per contact mode and a rule at impacts.

5. *The noise scale.* σ is a resolution, not a temperature. Comparisons do not depend on it only when the designs excite the same number of modes (Proposition 4.3(c)), so ranks should be reported alongside bits.
6. *Two metrics, two questions.* $I(W;X)$ measures transmission and MC_W measures retention (Proposition 5.2), so the two should not be expected to agree.

Part III

Research log

Research notes

Dated notes record results as they are found, newest first. Each note says what was found, what checks it, and where it now lives in the chapters, if anywhere.

Date	Note	Status	Now in
2026-11-21	Energy is the right noise model, and morphology routes information	promoted	Chapter 7, Chapter 8

2026-11-21: Energy is the right noise model, and morphology routes information

Status: promoted to Chapter 7 and Chapter 8 · **Checked by:** `tests/test_energy_whitening.py`

Question. Is the energy metric the best way to model a body's resolution for measuring mechanical intelligence, or only a reasonable one?

Findings.

1. Maximum entropy fixes the shape of the noise once energy is chosen as the constrained quantity, but it does not choose energy. Constraining mean action instead gives the quantum zero-point shape.
2. Energy is the only noise metric that is coordinate-free, assembled element by element, free of outside scales, symmetric in time and invariant under free motion (Theorem 7.1).
3. Invariance under free motion is equivalent to "free motion never creates information" (Proposition 7.1). Under per-coordinate noise, free motion raised the stored information from 2 to 3.58 bits with no input at all.
4. Of the alternatives, only action weighting passes both the stiff-spring limit and the mesh-refinement limit. Locality and the classical regime exclude it. The frequency-weighted family converges in the stiff limit exactly when $p < 2$, as predicted (Table 7.1).
5. Whitening by the data gives every body the same bits (Proposition 7.2).
6. In a Langevin body, with noise injected through the dampers, every design gives the same 37.742 bits when

the full state is observed. The limit is set by the damping at the contact, $JD^{+}J^{\top} = 1/b_i$ (Proposition 8.2). Through the motor encoder, the rigid transmission captures 99.9% of that limit, and the SEA climbs from 6% to 99.5% with stiffness (Table 8.1).

7. Readout-only noise lets a resonant SEA beat the rigid leg through the encoder by up to 5.9 bits. Under the fluctuation-dissipation theorem the resonance amplifies the thermal noise by the same factor, so that advantage is an artifact.

Open from this note. The constant 1/2 in Proposition 8.2 is confirmed numerically but proved only in continuous time (roadmap thread T3). The routing efficiency in point 6 needs a definition and properties of its own (thread T2).

Research roadmap

This page is the working plan for the book. It lists the open threads and the next step on each, and keeps a dated record of findings and decisions. It is updated at the end of every working session. The book itself is not planned in advance; chapters grow out of the findings recorded here.

How we work

- **Session protocol.** Each session starts by reading this page and ends by updating it: thread statuses, findings, decisions, and a changelog entry if anything was published.
- **From finding to page.** A question becomes a thread here. Exploration happens in scratch code. A result that holds up gets a function in `mibook` and a test in `tests/`, then a dated research note (status *exploratory*). The note is folded into a chapter once its result is checked, its proofs are labelled, and it connects to the existing material. The note stays as the dated record and links to the chapter.
- **Checked numbers.** Every number in the text is computed by `mibook` and pinned by a test. Figures are code cells that call `mibook.figures`.
- **Proofs.** An incomplete proof is titled "Proof sketch" and states what it assumes.
- **Citations.** Each citation is marked *full text* or *secondary* in the sources appendix.
- **Prose.** All text (chapters, notes, this page, the changelog and commit messages) is written and edited with the humanizer skill: no staged contrasts, closers or inflated claims, no em dashes in prose, plain verbs.

- **Releases.** Dated tags (`v0.1`, `v0.2`, …) with a changelog entry. Before a result goes to a venue, check that venue's preprint policy.
- **Branches.** Book content goes straight to `main`, which deploys the site. Changes to other repositories, such as the lab website, go through pull requests. This is a default and can change.

Threads

Order is chosen together at the start of a session. Status is one of *active*, *queued*, *parked* or *done*.

T1. Foundations of the energy metric. *Done*, except for one open item.

- Where: Chapter 2, Chapter 7.
- Open: confirm the citations marked *secondary* against the sources. The PDFs that could not be fetched automatically are Li, Li and Fritzen (2007); Moore (1981); Kubo (1966); Lee, Wensing and Park (2020); Zahedi and Ay (2013); Cover and Thomas, chapters 9 and 12; Kammer (1991); Geveci (1984), or Baker (1976) or Dupont (1973).

T2. Retained versus routed information. *Queued.*

- Question: what is the right definition of routing efficiency (sensor bits divided by full-state bits under Langevin noise), and what are its properties? Candidates to prove: bounds, invariance under relabeling the sensor, behaviour under composition of mechanisms.
- Next step: state the definition, prove the bound $0 \leq \eta \leq 1$ and the invariances, and compute η for the joint encoder as well as the motor encoder.

T3. The Langevin body as the default sensor model. *Queued.*

- Question: can the discrete-time constant in Proposition 8.2 be proved, not only checked?

- Next step: expand $B_d^\top Q_d^{-1} B_d$ in Δt, including the position-block terms, and show the leading coefficient is $JD^+J^\top/(2\sigma_T^2)$.

T4. Predictive validation on the Go2 designs. *Parked: needs data.*

- Question: which of retention, routing efficiency and $I(W; X \mid A)$ best predicts closed-loop tracking error across designs and seeds?
- Depends on: the RL training pipeline and trained policies from the quadruped experiments, which are not in either repository.

T5. The Langevin body in the MuJoCo leg. *Queued.*

- Question: does the analytic Langevin result (flat full-state bits, motor bits climbing with stiffness) hold in the nonlinear simulated leg?
- Next step: inject damping-matched noise forces in the simulation and estimate with InfoNCE. Needs the `sim` extra (torch and warp), ideally on a GPU.

T6. Floating bases. *Queued.*

- Question: how should base positions that store no energy be treated?
- Options: keep velocities only; regularize with $\omega_{\text{ref}}^2 H$ and sweep ω_{ref}; use gravity and contact stiffness at the operating point.

T7. Nonlinear and multistable bodies. *Queued.*

- Question: what does the framework say about mechanical memory, where the noise law is the Gibbs law on a non-quadratic potential?
- Next step: a bistable element with an energy barrier; bits stored against barrier height and noise energy.

T8. Latency. *Queued.*

- Question: when does routed information arrive at the sensor, and how does that depend on the transmission? Extends the temporal Shapley analysis.

T9. Relevance. *Queued.*

- Question: which measures capture information relevant to the controller or the task, such as $I(W; X \mid A)$ or control information?

T10. Literature. *Queued.*

- Confirm the *secondary* citations (see T1), then survey morphological computation, physical reservoir computing and the thermodynamics of computation.

Findings log

Newest first. Each entry gives the finding, the test that pins it, and where it is written up.

- **2026-11-21.** Energy is the unique noise metric under five requirements, and invariance under free motion is equivalent to "free motion never creates information" (`test_uniqueness_step_four_identity`, `test_free_motion_creates_information_only_without_energy_noise`; Chapter 7).
- **2026-11-21.** The frequency-weighted family W_p converges in the stiff limit exactly when $p < 2$; only action weighting ($p = -1$) also passes mesh refinement (`test_frequency_weighted_stiff_limit_threshold_is_p_equals_two`, `test_frequency_weighted_mesh_behaviour`; Section 7.4).
- **2026-11-21.** Data-driven whitening gives every body the same bits (`test_data_driven_whitening_erases_the_body`; Proposition 7.2).
- **2026-11-21.** Langevin body: with the full state observed, every design gives the same bits, set by the contact damping $JD^{+}J^{\top} = 1/b_i$. The rigid transmission routes 99.9% of that to its motor encoder; the SEA climbs from 6% to 99.5% with stiffness. Readout-only noise misranks resonant SEAs by up to 5.9 bits (`test_langevin_full_state_information_depends_only_on_contact_damping`, `test_langevin_routing_reproduces_the_proprioceptive_argument`; Chapter 8).

- **2026-11-20.** Coordinate invariance, the port-Hamiltonian structure, the stiff-SEA limit with explicit rates, the 8.4716-bit duplicate-readout artifact, SNR asymptotics, consistent Shapley noise, the closed form for MC_W, mesh independence and the floating-base condition (the remaining tests in `tests/test_energy_whitening.py`; Part I).

Decisions log

- **2026-11-21.** The work becomes a live textbook on GitHub Pages, published as results are checked; important results go to venues later. The book is not planned in advance.
- **2026-11-21.** Quarto for authoring (one source for the website and the PDF; numbered theorems with cross-references; figures from code).
- **2026-11-21.** A dedicated repository, `CyPhiLab/mechintelligence`, owned by the lab. It depends on the analysis library in `zpatty/quadruped-mech-info`, pinned to commit `2b25bca`, so the book's numbers stay reproducible.
- **2026-11-21.** The book is served at `mechintelligence.cyphilab.com` and linked from the lab website's navigation bar (CyPhiLab/CyPhiLab.github.io pull request 3).
- **2026-11-21.** Two-tier publishing: dated research notes, and chapters that absorb them once stable.
- **2026-11-21.** All prose goes through the humanizer skill.
- **2026-11-21.** Branch workflow: direct pushes to `main` for book content; pull requests for other repositories.
- **2026-11-21.** The book is live at `mechintelligence.cyphilab.com`, tagged `v0.1`. The Publish workflow installs Quarto and TinyTeX with `quarto-dev/quarto-actions/setup`, because `quarto install tinytex` hit GitHub's anonymous API rate limit on the shared runners.

Paper candidates

- **The energy metric is unique, and free motion cannot create information under it.** Theorem 7.1 and Proposition 7.1, with the comparison of alternatives. Possible venues: a robotics venue (RSS, ICRA, IEEE Transactions on Robotics) or a physics or information-theory venue.
- **Morphology routes information; the contact sets the ceiling.** Proposition 8.2 and the routing results, ideally after threads T2, T3 and T5.

Changelog

Newest first. Each release is tagged in the repository.

v0.1 (2026-11-21)

First version of the book.

- Part I, *Information in mechanisms*: the energy metric, its invariance and noise law, the port-Hamiltonian structure of the whitened dynamics, bodies with different numbers of coordinates (the stiff-SEA limit and the duplicate-readout artifact), attribution and morphological computation, and continua, gears and floating bases.
- Part II, *Is energy the right choice?*: the uniqueness theorem, the proof that free motion cannot create information under energy noise, the alternatives compared, and the Langevin body with its routing results.
- Open problems, the first research note, the roadmap, and the sources appendix.
- 27 tests pin every number in the text.

A Status of the literature claims

Status: stable · **Updated:** 2026-11-21

Each citation in the book is marked by how it was checked. *Full text* means the primary source was read, and only these citations carry equation or section numbers. *Secondary* means the claim rests on abstracts or citing works and should be confirmed against the source. No mathematical statement in the book depends on a citation; each is proved in the text. This table is updated as sources are read in full (thread T10 of the roadmap).

Source	Access	What is used
Jaynes (1957)	full text	Eqs. (2-4) and (2-10); Sec. 3 (Boltzmann distribution); footnote 6 (discrete case only)
Kessy et al. (2018)	full text	Eq. (4), rotational freedom; Eq. (8), ZCA; Eqs. (13) and (14), optimality
Diggavi and Cover (2001)	full text	Lemma II.2 (Ihara): Gaussian noise is worst at fixed covariance

Source	Access	What is used
Schaft and Jeltsema (2014)	full text	Eq. (4.10), input-state-output port-Hamiltonian form
Bui-Thanh et al. (2013)	full text	Sec. 3.3, mass-weighted inner product
Cotter et al. (2013)	full text	robustness under mesh refinement (abstract and Sec. 1)
Kubo (1966)	secondary	the FDT links noise intensity to dissipation; equipartition
Moore (1981)	secondary	Gramians, coordinate dependence, balancing
Li et al. (2007)	secondary	the EfI and MKE connection, via iteration and re-orthonormalization
Lee et al. (2020)	secondary	affine-invariant metric on inertial parameters (parameter space, not state space)
Zahedi and Ay (2013)	secondary	discrete $\mathrm{MC}_W = I(W'; W \mid A)$ on empirical distributions
1	own work	the leg models, settings and quadruped experiments that the book starts from

Source	Access	What is used
Wensing et al. (2017)	secondary	model parameters, as reproduced in 1 [Sec. III]
Cover and Thomas (2006)	secondary	Ch. 12, maximum entropy; colored-noise channels
Kammer (1991); Papadimitriou (2004)	secondary	EfI as D-optimal design; entropy-based sensor placement
Willems (1972); Stuart (2010)	secondary	storage functions; the function-space viewpoint
Dupont (1973); Baker (1976); Geveci (1984)	secondary	semi-discrete finite-element convergence (an assumption of Theorem 6.1)
Landauer (1961); Bennett (1982)	secondary	reversibility of conservative dynamics; dissipation as erasure
Callen and Welton (1951)	secondary	quantum FDT, $\frac{\hbar\omega}{2}\coth\frac{\hbar\omega}{2k_BT}$
Nyquist (1928)	secondary	thermal noise set by dissipation
Duncan (1970)	secondary	mutual information for signals in white Gaussian noise

References

Baker, Garth A. 1976. "Error Estimates for Finite Element Methods for Second Order Hyperbolic Equations." *SIAM Journal on Numerical Analysis* 13 (4): 564–76. https://doi.org/10.1137/0713048.

Bennett, Charles H. 1982. "The Thermodynamics of Computation—a Review." *International Journal of Theoretical Physics* 21 (12): 905–40. https://doi.org/10.1007/BF02084158.

Bui-Thanh, Tan, Omar Ghattas, James Martin, and Georg Stadler. 2013. "A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems. Part I: The Linearized Case, with Application to Global Seismic Inversion." *SIAM Journal on Scientific Computing* 35 (6): A2494–523. https://doi.org/10.1137/12089586X.

Callen, Herbert B., and Theodore A. Welton. 1951. "Irreversibility and Generalized Noise." *Physical Review* 83 (1): 34–40. https://doi.org/10.1103/PhysRev.83.34.

Cotter, Simon L., Gareth O. Roberts, Andrew M. Stuart, and David White. 2013. "MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster." *Statistical Science* 28 (3): 424–46. https://doi.org/10.1214/13-STS421.

Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory*. 2nd ed. Wiley-Interscience.

Diggavi, Suhas N., and Thomas M. Cover. 2001. "The Worst Additive Noise Under a Covariance Constraint." *IEEE Transactions on Information Theory* 47 (7): 3072–81. https://doi.org/10.1109/18.959289.

Duncan, Tyrone E. 1970. "On the Calculation of Mutual Information." *SIAM Journal on Applied Mathematics* 19 (1): 215–20. https://doi.org/10.1137/0119020.

Dupont, Todd. 1973. "L^2-Estimates for Galerkin Methods for Second Order Hyperbolic Equations." *SIAM Journal on Numerical Analysis* 10 (5): 880–89. https://doi.org/10.1137/0710073.

Geveci, Tunc. 1984. "On the Convergence of Galerkin Approximation Schemes for Second-Order Hyperbolic Equations in Energy and Negative Norms." *Mathematics of Computation* 42 (166): 393–415. https://doi.org/10.1090/S0025-5718-1984-0736443-5.

Jaynes, Edwin T. 1957. "Information Theory and Statistical Mechanics." *Physical Review* 106 (4): 620–30. https://doi.org/10.1103/PhysRev.106.620.

Kammer, Daniel C. 1991. "Sensor Placement for on-Orbit Modal Identification and Correlation of Large Space Structures." *Journal of Guidance, Control, and Dynamics* 14 (2): 251–59. https://doi.org/10.2514/3.20635.

Kessy, Agnan, Alex Lewin, and Korbinian Strimmer. 2018. "Optimal Whitening and Decorrelation." *The American Statistician* 72 (4): 309–14. https://doi.org/10.1080/00031305.2016.1277159.

Kubo, R. 1966. "The Fluctuation-Dissipation Theorem." *Reports on Progress in Physics* 29 (1): 255–84. https://doi.org/10.1088/0034-4885/29/1/306.

Landauer, Rolf. 1961. "Irreversibility and Heat Generation in the Computing Process." *IBM Journal of Research and Development* 5 (3): 183–91. https://doi.org/10.1147/rd.53.0183.

Lee, Taeyoon, Patrick M. Wensing, and Frank C. Park. 2020. "Geometric Robot Dynamic Identification: A Convex Programming Approach." *IEEE Transactions on Robotics* 36

(2): 348–65. https://doi.org/10.1109/TRO.2019.2926491.

Li, D. S., H. N. Li, and C. P. Fritzen. 2007. "The Connection Between Effective Independence and Modal Kinetic Energy Methods for Sensor Placement." *Journal of Sound and Vibration* 305 (4–5): 945–55. https://doi.org/10.1016/j.jsv.2007.05.004.

Moore, Bruce C. 1981. "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction." *IEEE Transactions on Automatic Control* 26 (1): 17–32. https://doi.org/10.1109/TAC.1981.1102568.

Nyquist, H. 1928. "Thermal Agitation of Electric Charge in Conductors." *Physical Review* 32 (1): 110–13. https://doi.org/10.1103/PhysRev.32.110.

Papadimitriou, Costas. 2004. "Optimal Sensor Placement Methodology for Parametric Identification of Structural Systems." *Journal of Sound and Vibration* 278 (4–5): 923–47. https://doi.org/10.1016/j.jsv.2003.10.063.

Patterson, Zach J. 2026. *Quantifying Mechanical Intelligence in Legged Robots with Information Theory*. https://arxiv.org/abs/2609.19588.

Schaft, Arjan van der, and Dimitri Jeltsema. 2014. "Port-Hamiltonian Systems Theory: An Introductory Overview." *Foundations and Trends in Systems and Control* 1 (2–3): 173–378. https://doi.org/10.1561/2600000002.

Stuart, Andrew M. 2010. "Inverse Problems: A Bayesian Perspective." *Acta Numerica* 19: 451–559. https://doi.org/10.1017/S0962492910000061.

Wensing, Patrick M., Albert Wang, Sangok Seok, David Otten, Jeffrey Lang, and Sangbae Kim. 2017. "Proprioceptive Actuator Design in the MIT Cheetah: Impact Mitigation and High-Bandwidth Physical Interaction for Dynamic Legged Robots." *IEEE Transactions on Robotics* 33 (3): 509–22. https://doi.org/10.1109/TRO.2016.2640183.

Willems, Jan C. 1972. "Dissipative Dynamical Systems Part I: General Theory." *Archive for Rational Mechanics and Analysis* 45 (5): 321–51. https://doi.org/10.1007/BF00276493.

Zahedi, Keyan, and Nihat Ay. 2013. "Quantifying Morphological Computation." *Entropy* 15 (5): 1887–915. https://doi.org/10.3390/e15051887.