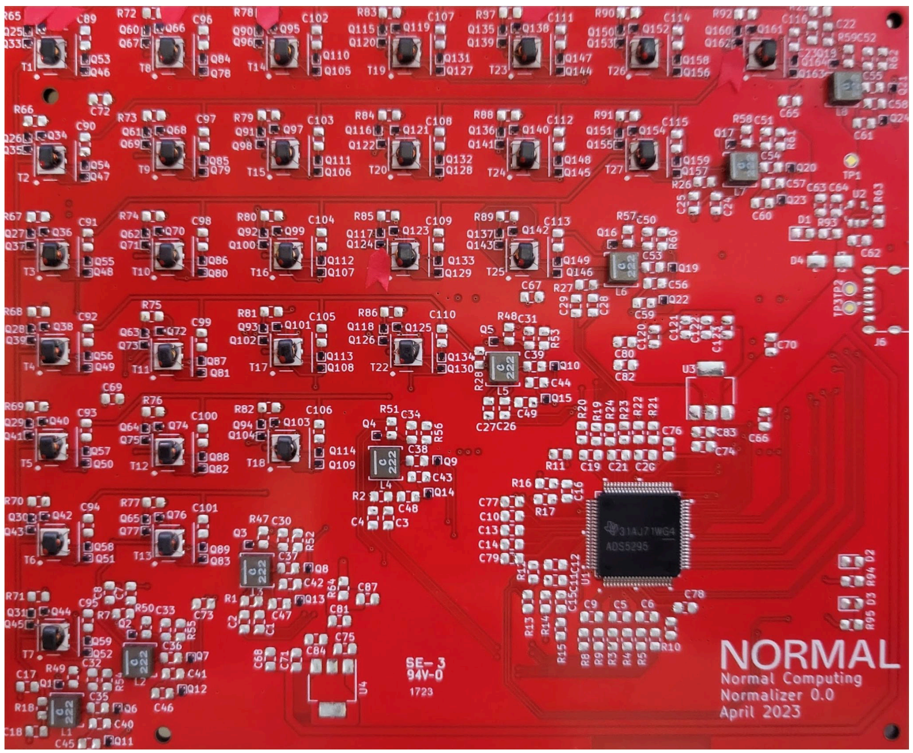
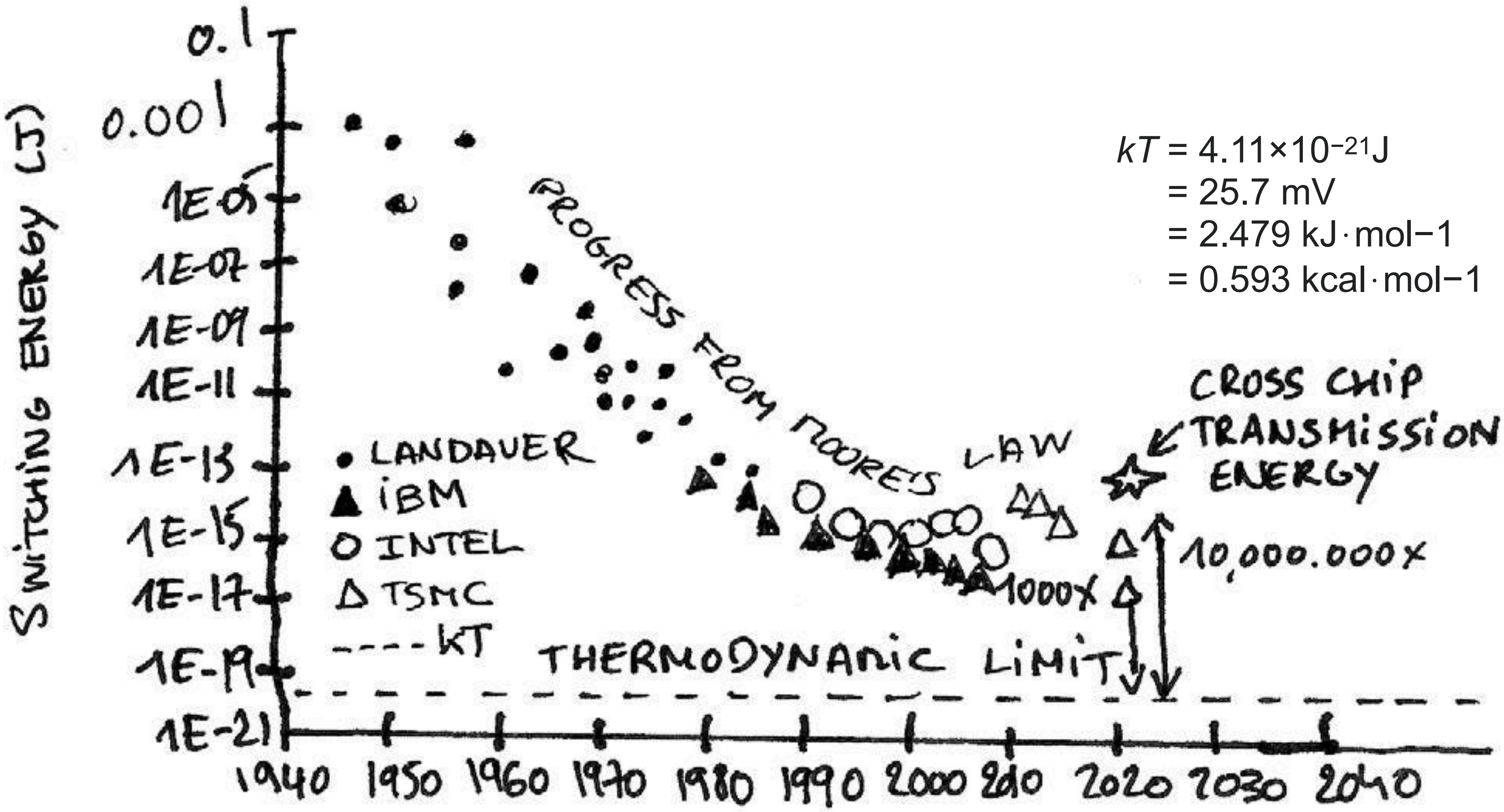
Thermodynamic Linear Algebra

Gavin E. Crooks @ NORMAL Computing



Transistors: Moore's law and Dennard scaling



https://ariaresearch.substack.com/p/spotlight-on-suraj-scaling-compute

Energy Scales

~1,000,000 kT	Floating point operation
~100,000 kT	Neuron spike
~10,000 kT	Cross chip communication
~1,000 kT	Transistor switch (SOTA)
~100 kT	limit of reliable fast information processing?
~10 kT	Molecular biology / ATP hydrolysis
~1 kT	Thermal fluctuations / Landauer's limit

$kT = 4.11 \times 10^{-21}$J
= 25.7 mV
= 2.479 kJ·mol−1
= 0.593 kcal·mol−1

Improving Compute per kT

1) More efficient transistors ?
2) Algorithms
3) Quantum Computing
4) Novel (classical) Hardware

The Hardware Lottery

Sara Hooker

Google Research, Brain Team
shooker@google.com

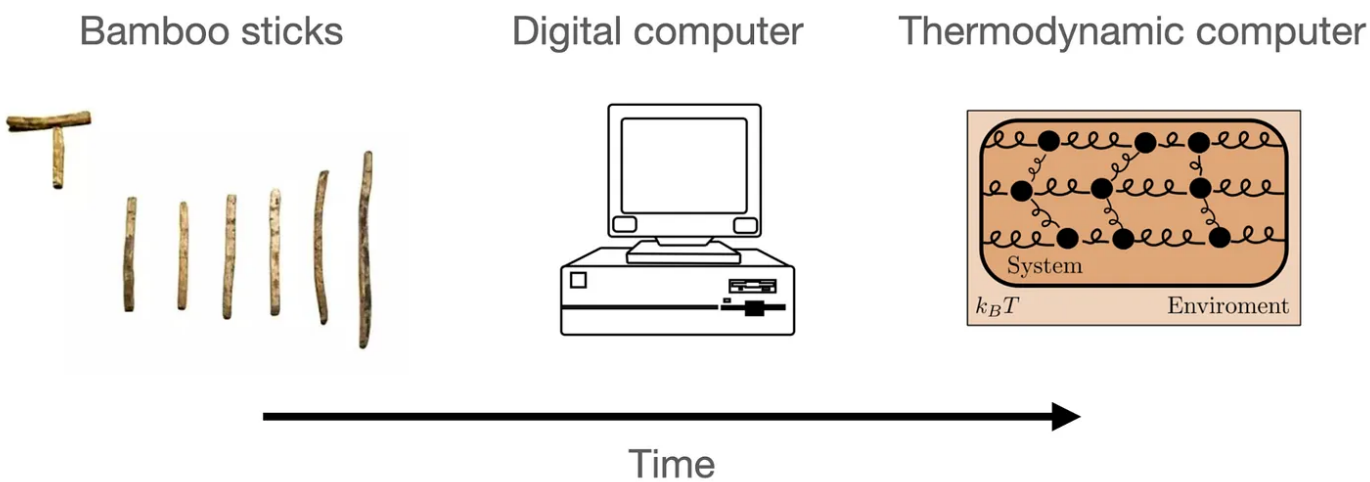
Abstract

Hardware, systems and algorithms research communities have historically had different incentive structures and fluctuating motivation to engage with each other explicitly. This historical treatment is odd given that hardware and software have frequently determined which research ideas succeed (and fail). This essay introduces the term hardware lottery to describe when a research idea wins because it is suited to the available software and hardware and *not* because the idea is superior to alternative research directions.

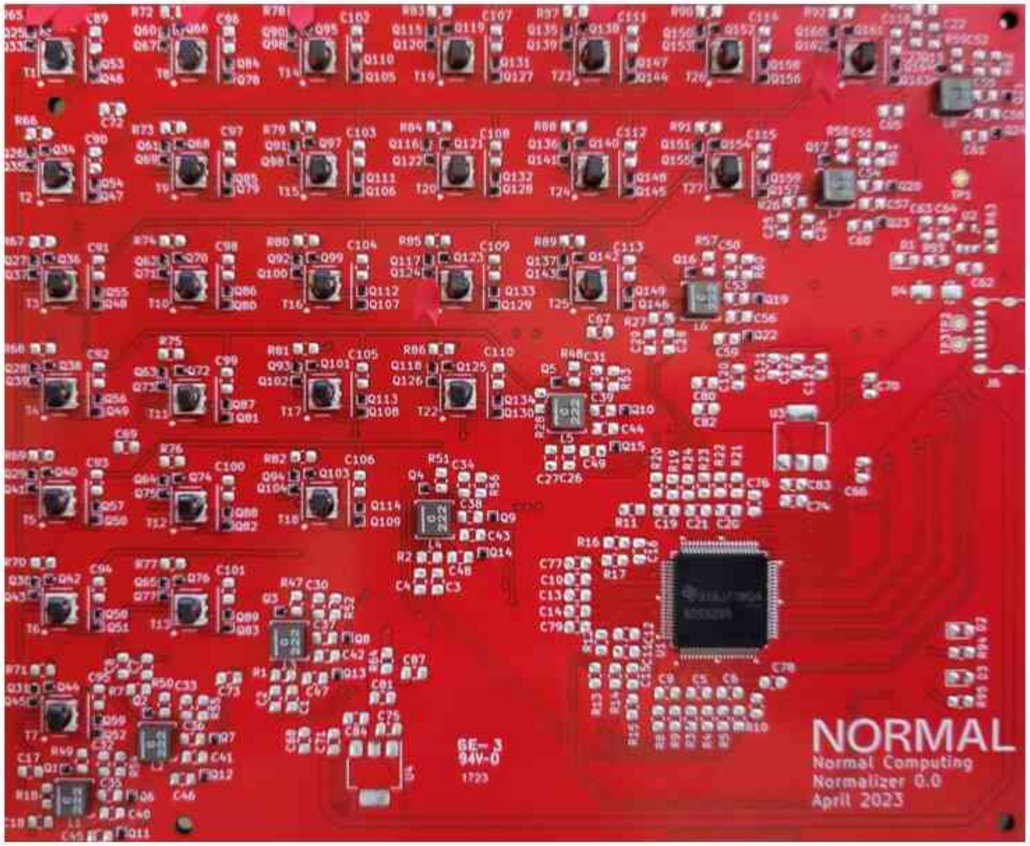
Physic based / Thermodynamic Computing

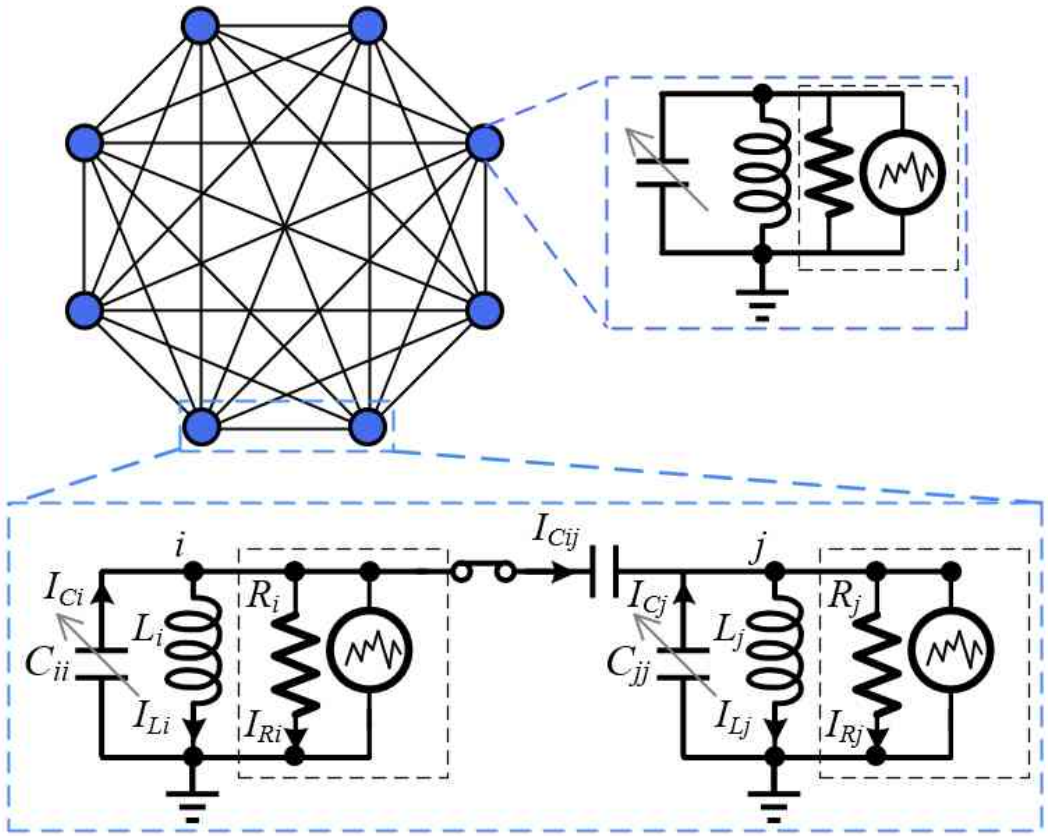
Thermodynamic Computing 1911.01968 Tom Conte et al.

- Harness nature's innate computational capacity
  - Use the underlying physics to compute (compute closer to the hardware)
- Noise is inevitable.
  - Use as a resource, not a curse



Stochastic Processing Unit (SPU)





Stochastic Processing Unit Dynamics

Overdamped or Underdamped Langevin dynamics

Currents

$$\mathrm{d}I = \mathbf{L}^{-1} V \mathrm{d}t$$

$$\mathrm{d}V = -\mathbf{C}^{-1}\mathbf{R}^{-1} V \mathrm{d}t - \mathbf{C}^{-1} I \mathrm{d}t + \sqrt{2\kappa_0}\mathbf{C}^{-1}\mathcal{N}[0, \mathbb{I}\,\mathrm{d}t],$$

Voltages

L: Inductances

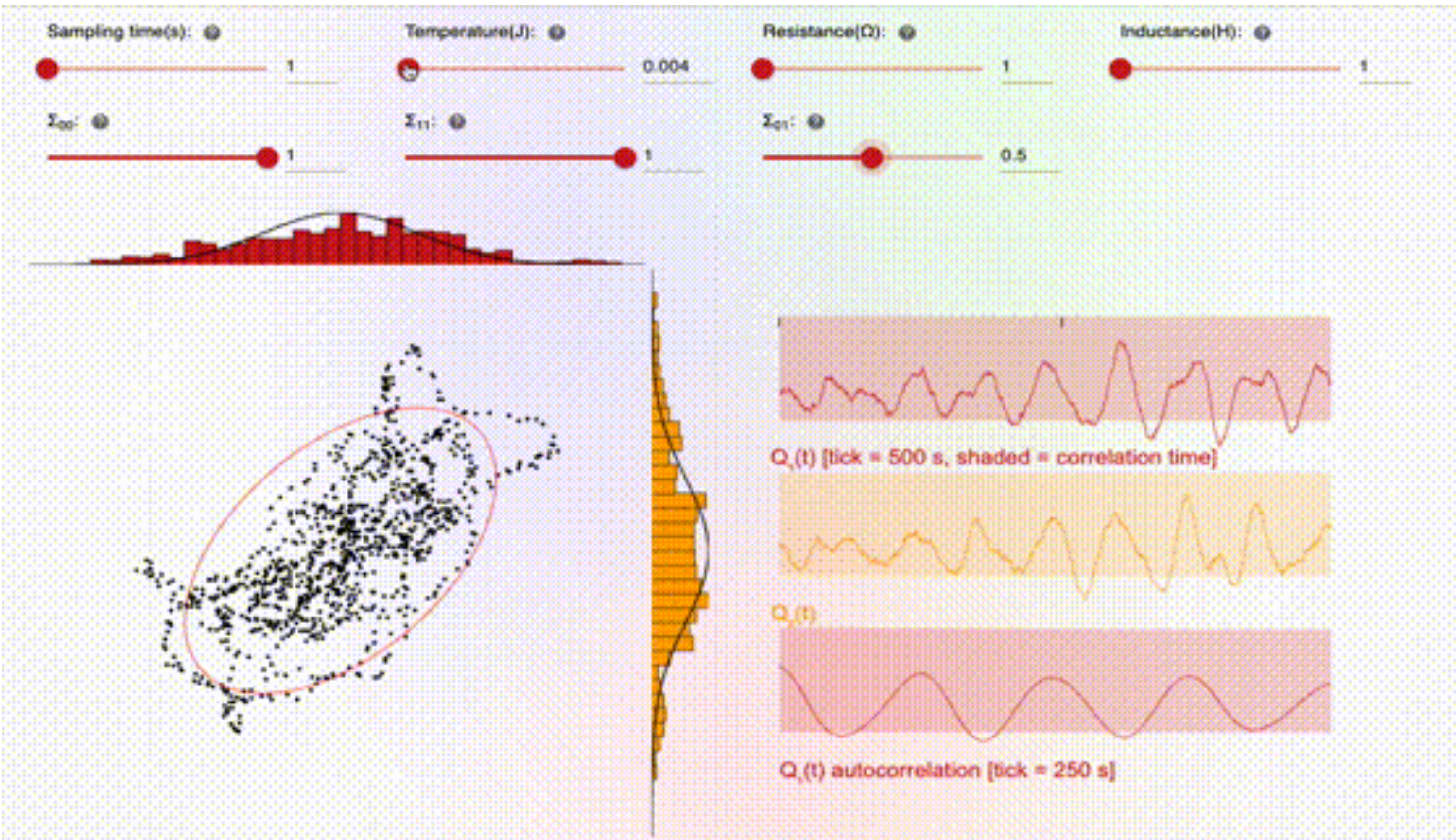
$$\mathcal{H}\left(\vec{I}, \vec{V}\right) = \frac{1}{2}\vec{V}^T \mathbf{C}\vec{V} + \frac{1}{2}\vec{I}^T \mathbf{L}\vec{I},$$

Hamiltonian

C: Maxwell Capacitance Matrix

Thermal Playground

https://app.normalcomputing.ai/composer



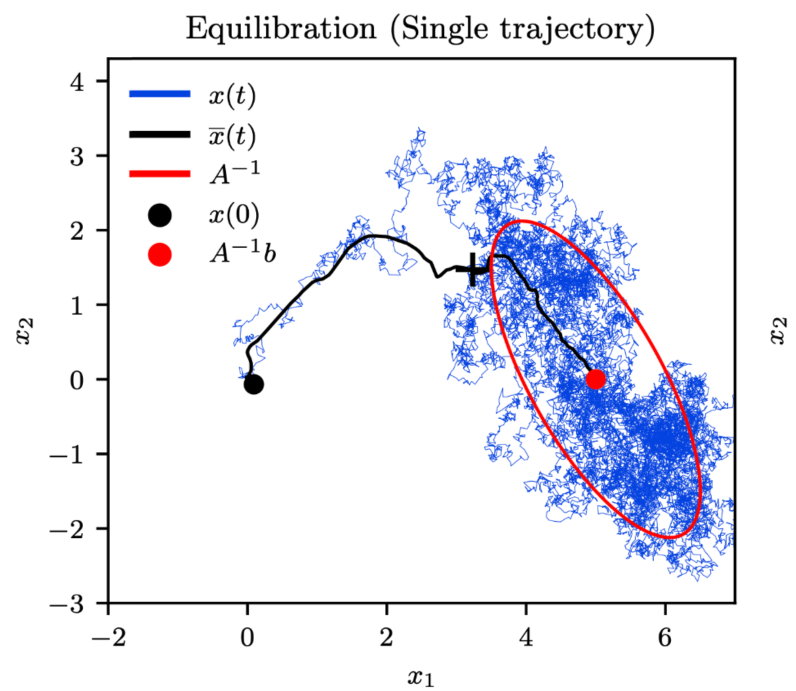
Gaussian Sampling with Stochastic Processing Unit

For harmonic oscillator system, at thermal equilibrium, *x* is Gaussian distributed:

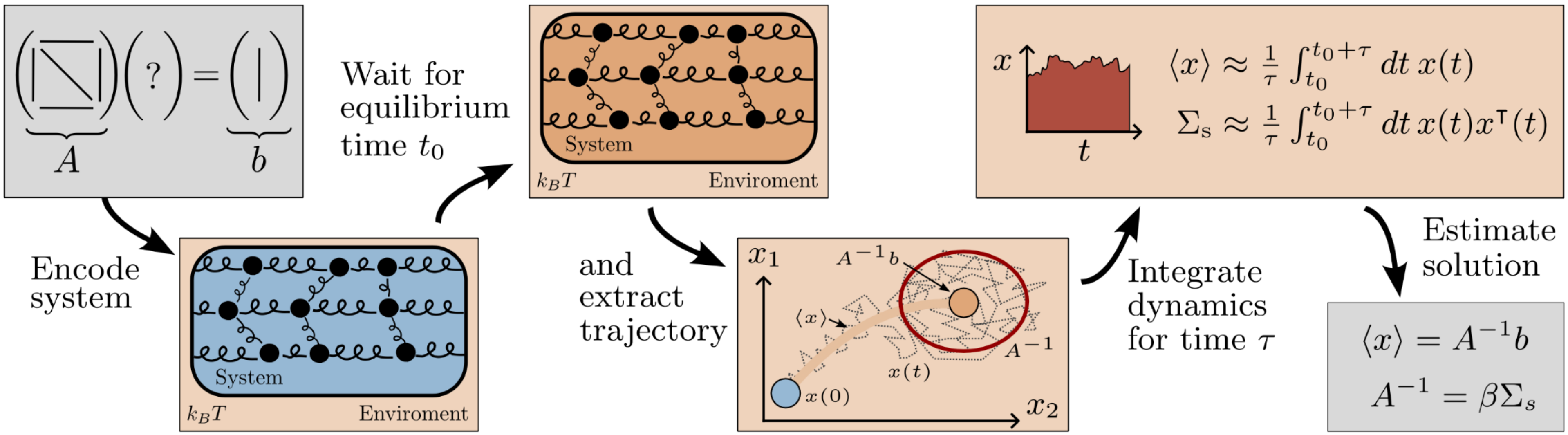
$$\mathcal{N}(\vec{x}|\mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^N|\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}\vec{x}^T\mathbf{\Sigma}^{-1}\vec{x}\right)$$

Maxwell capacitance matrix (C) and covariance matrix are related.

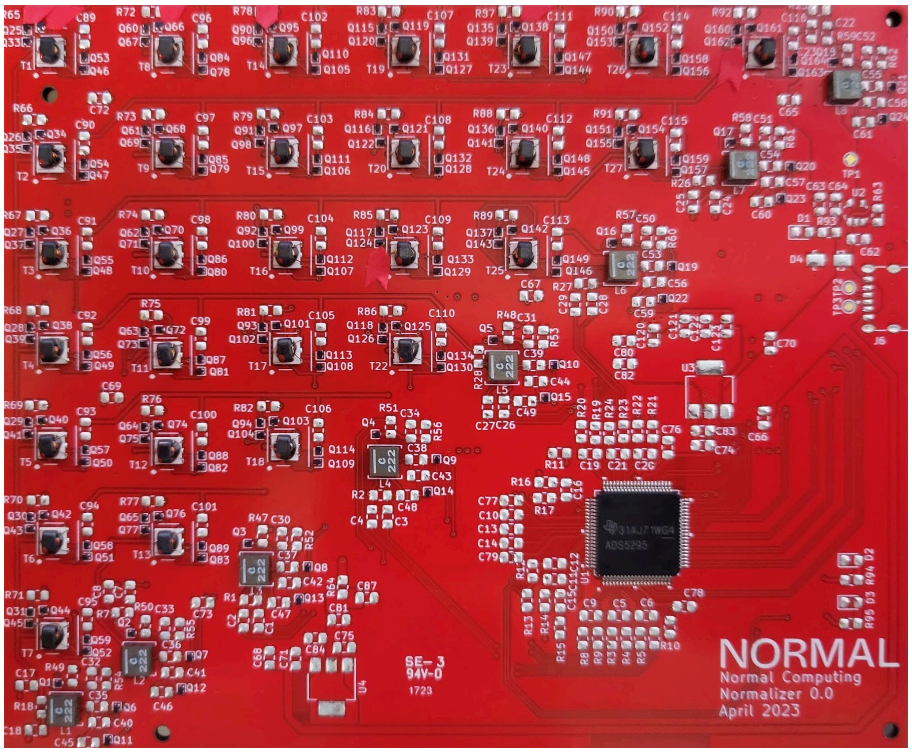
$$C = k_B T \, \Sigma^{-1}$$

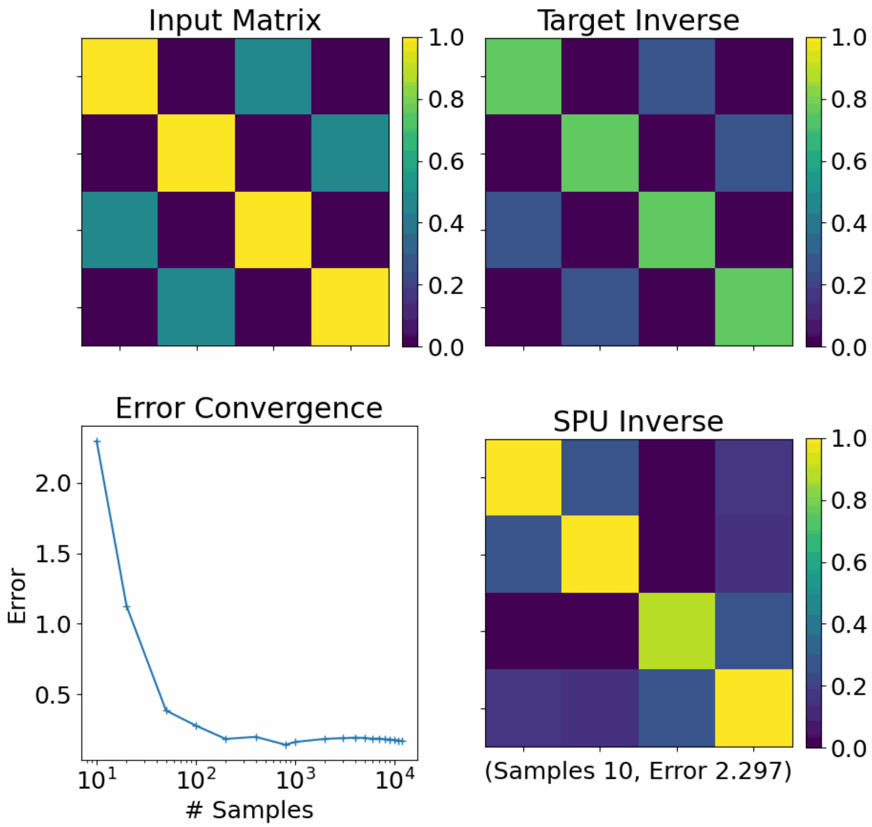


Matrix Inversion with Stochastic Processing Unit



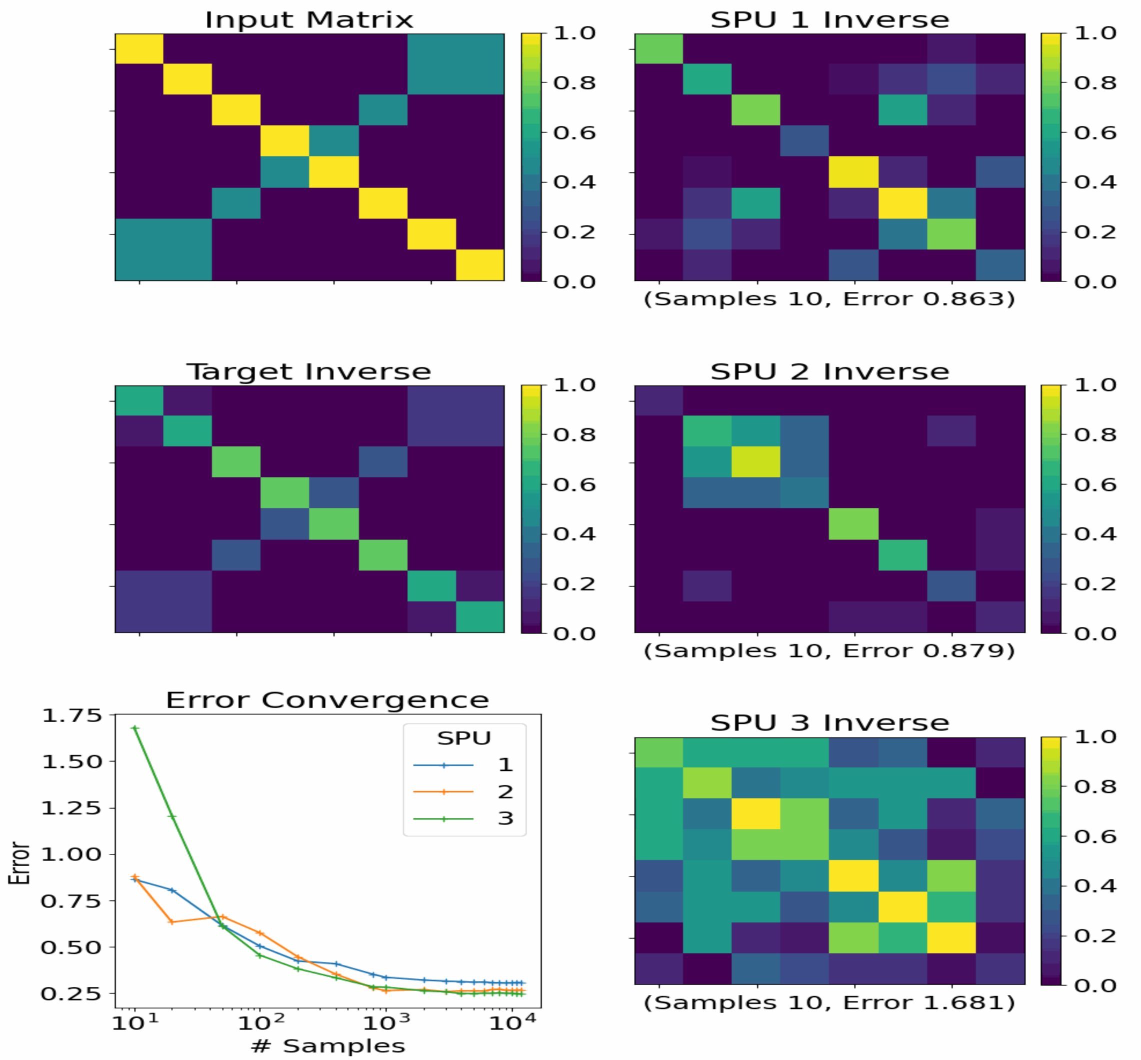
Matrix Inversion with Stochastic Processing Unit





https://blog.normalcomputing.ai/posts/2023-11-09-thermodynamic-inversion/thermo-inversion.html

Matrix Inverse
8x8



Matrix Determinant with Stochastic Processing Unit

$$f_{\mu;\Sigma}(x) = (2\pi)^{-d/2}\,|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}x^\intercal\Sigma^{-1}x\right),$$

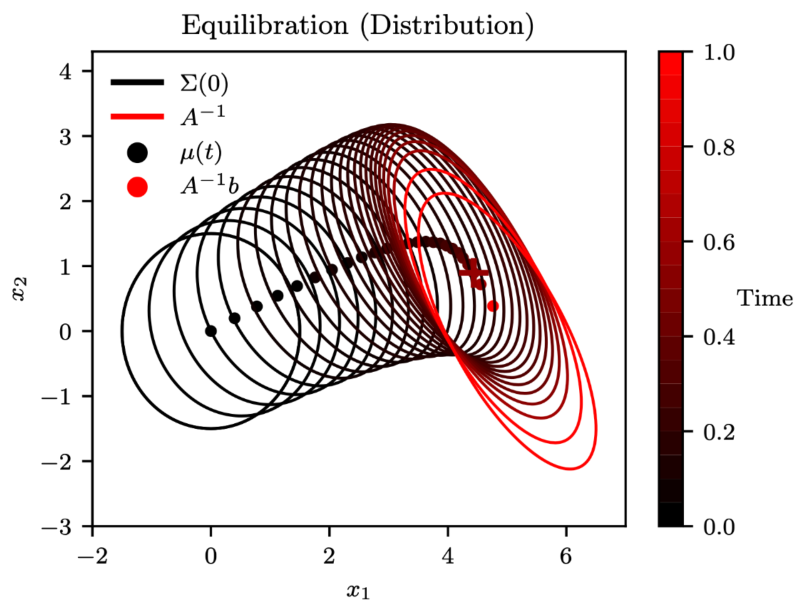
$$S(\Sigma) = \frac{1}{2}\ln|\Sigma| + \frac{d}{2}(1+\ln 2\pi)$$

$$\Delta F = \Delta E - \beta^{-1}\Delta S$$

$$\Delta F = -\beta^{-1}\ln\left(\sqrt{\frac{|A_2^{-1}|}{|A_1^{-1}|}}\right) = -\beta^{-1}\ln\left(\sqrt{\frac{|A_1|}{|A_2|}}\right).$$

Matrix Determinant with Stochastic Processing Unit

$$\Delta F = -\beta^{-1} \ln\left(\sqrt{\frac{|A_2^{-1}|}{|A_1^{-1}|}}\right) = -\beta^{-1} \ln\left(\sqrt{\frac{|A_1|}{|A_2|}}\right)$$



$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle$$

Optimal Control and Thermodynamic Geometry

Riemannian Geometry of finite-time thermodynamic control

free energy — *inverse temperature*

$$p(x|\lambda) = e^{\beta F(\lambda) - \beta E(x,\lambda)}$$

controllable parameters

linear response friction tensor

$$\zeta(\lambda)_{ij} = \beta \int_0^\infty dt \, \langle \delta X_j(0) \delta X_i(t) \rangle_\lambda$$

positive semi-definite symmetric matrix
i.e. thermodynamic metric tenser

correlations of conjugate variables

nonequilibrium excess power

$$\mathcal{P}_\Lambda^{\mathrm{ex}}(t_0) = \left[\frac{\mathrm{d}\boldsymbol{\lambda}^T}{\mathrm{d}t}\right]_{t_0} \cdot \boldsymbol{\zeta}(\boldsymbol{\lambda}(t_0)) \cdot \left[\frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t}\right]_{t_0}$$

Sivak & Crooks PRL (2012)

Geometry of thermodynamics

- Riemannian metric. Minimum dissipation paths are *geodesics*

  Peter Salamon and Steven Berry (1983), Ruppeiner (1979), F. Weinhold (1975)

- Fisher information and information geometry.

  Crooks (2007)

- Finite time thermodynamics with linear response friction tensor

  Sivak & Crooks PRL (2012)

- Steady states, and Drazin inverse derivation

  Mandal & Jarzynski (2015)

Steady state probabilities

$$\xi_{\mu\nu} = -\sum_{i,j} \pi_j \frac{\partial \ln \pi_i}{\partial \lambda_\nu} R^{+}_{ij} \frac{\partial \ln \pi_j}{\partial \lambda_\mu}.$$

Drazin inverse of the rate matrix

- Wasserstein metric

  Chennakesavalu & Rotskoff (2022)

NORMAL Computing

Patrick Coles · Maxwell Aifer · Kaelan Donatella · Denis Melanson · Max Hunter Gordon · Thomas Ahle · Daniel Simpson · Sam Duffield · Gavin Crooks · Antonio Martinez · Faris Sbahi

Error Mitigation for Thermodynamic Computing
arXiv:2401.16231

Thermodynamic Computing System for AI
arXiv:2312.04836

Thermodynamic Matrix Exponentials
arXiv:2311.12759

Thermodynamic Linear Algebra
arXiv:2308.05660

Thermodynamic AI and the fluctuation frontier
arXiv:2302.06584