Chemical Carnot cycles, Landauer's Principle, and the Thermodynamics of Natural Selection

Eric Smith
Santa Fe Institute

Complex Systems Summer School
Santa Fe, June 2008

Outline

- The subtle task of asking sensible questions about information in the biosphere
- The chemical Carnot construction
- The relation to computation

Big and little questions

- (Big) how does energy *flow* limit the informational *state* of the biosphere?

  Requires theory of biological decay

- (Little) how does energy *flow* limit the *change in information* in the biosphere?

  Can get from equilibrium thermodynamics

(Similar questions can be asked about individuals, species, etc., as about the whole biosphere)

The obvious (little) answer

$$dW = dQ = -k_B T dS \equiv k_B T d\mathcal{I}$$

- Follows from dimensional analysis and the definition of temperature
- Information gain should be entropy loss
- Heat is entropy carried by energy
- Work is an entropy-less energy source

In what senses is such an answer
useful?
wrong?
irrelevant?

I. The complex problem of thinking about information in the biosphere

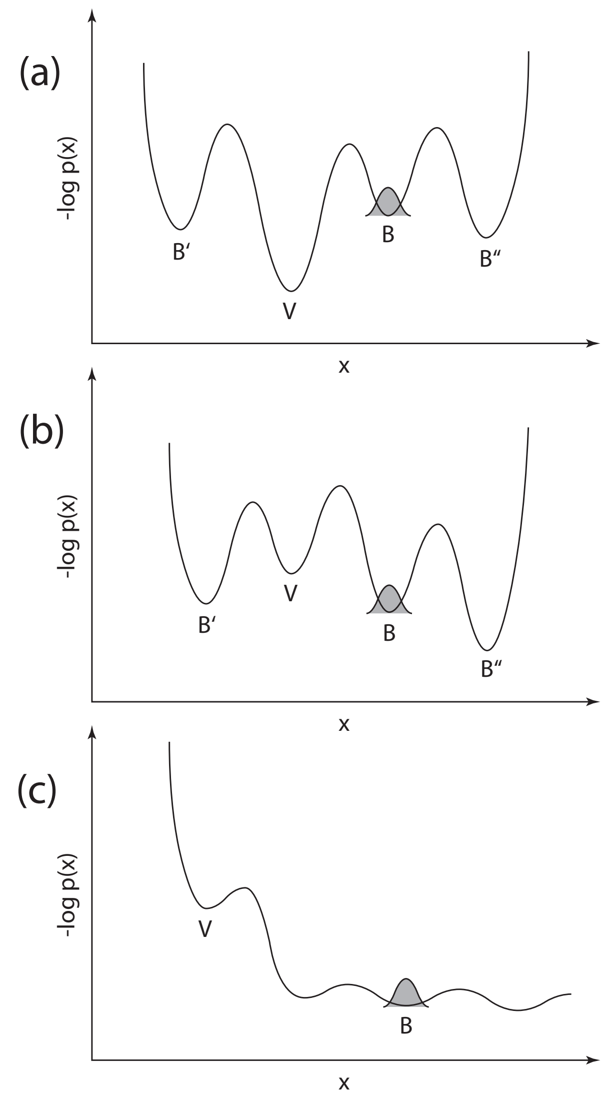
- Many levels, separation of timescales, and flow of constraint and control make assembling from the molecules *very* hard
- Which information? Genes? Heats?
- Which building process? Metabolism? Natural selection?
- What level? Individuals? Ecosystems? Biosphere?

The motivation to think about bounds rather than models

- Bounds from reversible processes also constrain irreversible ones
- Reversible-process bounds can be aggregated through *state variables*; irreversible models usually cannot be
- Bounds supersede models, unknown innovations, and ignorance of details

The challenge of using equilibrium information for the biosphere

- Life involves kinetics as well as energetics
- Our biosphere could (?) be a "frozen accident"
- Only if barriers are small enough that energy flow is limiting is information a relevant constraint



But such limits can be suggested in surprising places...

Allometric scaling of growth

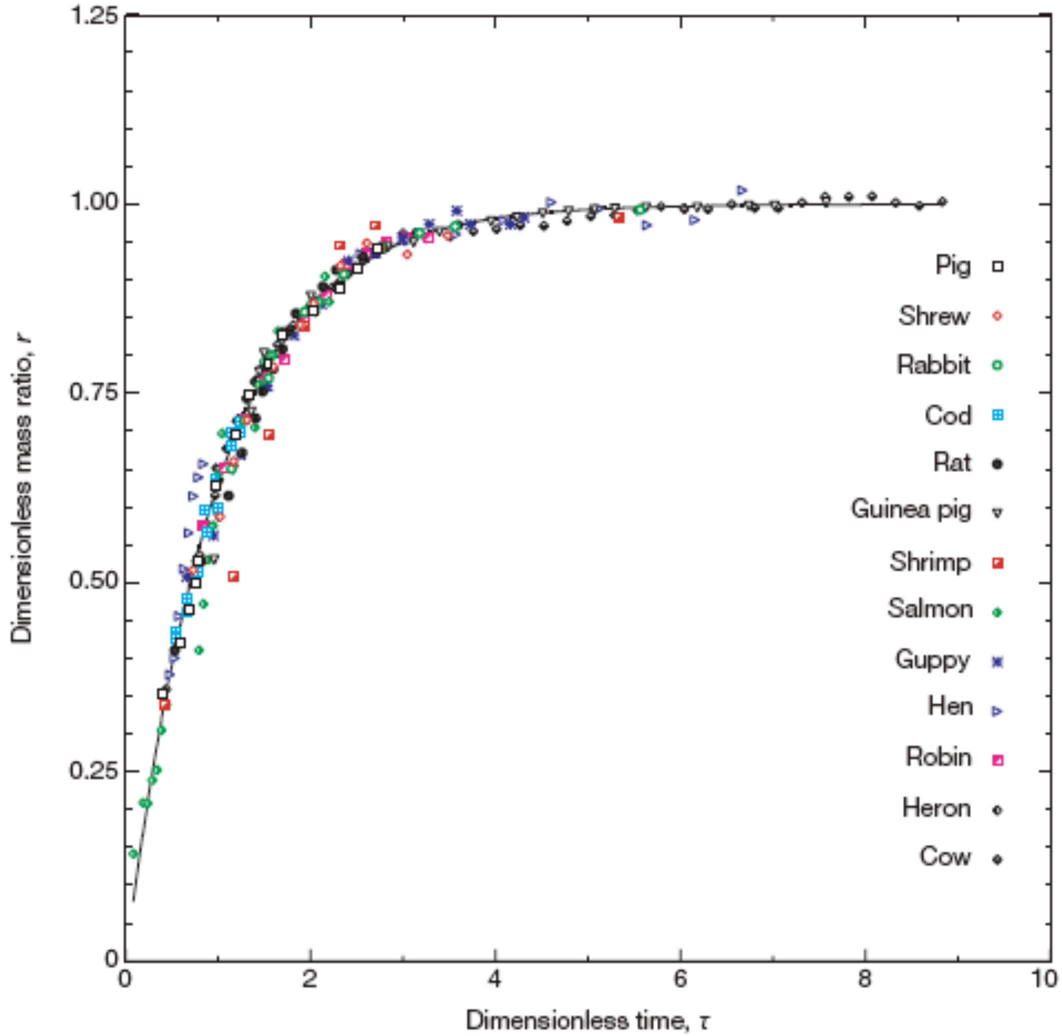
Energy balance in ontogenetic growth

$$B_0 m^{3/4} = \frac{B_c}{m_c} m + \frac{E_c}{m_c}\frac{dm}{dt}$$

Consequence: scale-invariant growth trajectories

$$\frac{d}{d\tau}\left(\frac{m}{M}\right)^{1/4} = 1 - \left(\frac{m}{M}\right)^{1/4}$$

$$\left(\frac{m}{M}\right)^{1/4} = 1 - e^{-\tau}$$



West G.B., Brown J.H. & Enquist B.J. (2001) A general model for otogenetic growth. Nature, 413, 628-631

Informational consequences of allometric scaling

Q: Does life history depend on energy or information?

- Energy/mass used by any stage of life is an invariant

$$\frac{E_{\text{lifetime}}}{M} = \frac{E_c}{m_c} \int_0^{\tau_D} d\tau \left(1 - e^{-\tau}\right)^3$$

- What *minimal energy* would we expect is needed to put "information" into biomass?

$$E_M \sim k_B T \frac{M}{10\text{g}} N_A$$

- Energy/ideal by any life stage is an invariant

$$\frac{E_{\text{lifetime}}}{E_M} = \frac{E_c}{k_B T N_A} \frac{10\text{g}}{m_c} \int_0^{\tau_D} d\tau \left(1 - e^{-\tau}\right)^3$$

- Formation of biomass is clocked by *information*, not directly by energy

$$\frac{E_c}{k_B T N_A} \frac{10\text{g}}{m_c} \approx 30$$

Curious consequences

- No direct evidence from growth that there is a cost to maintaining the living state
- Even decay seems to be created in proportion to growth and repair processes
- Living systems *scale* as if they were on the energy/information bound, even though they deviate from it by an “inefficiency” factor

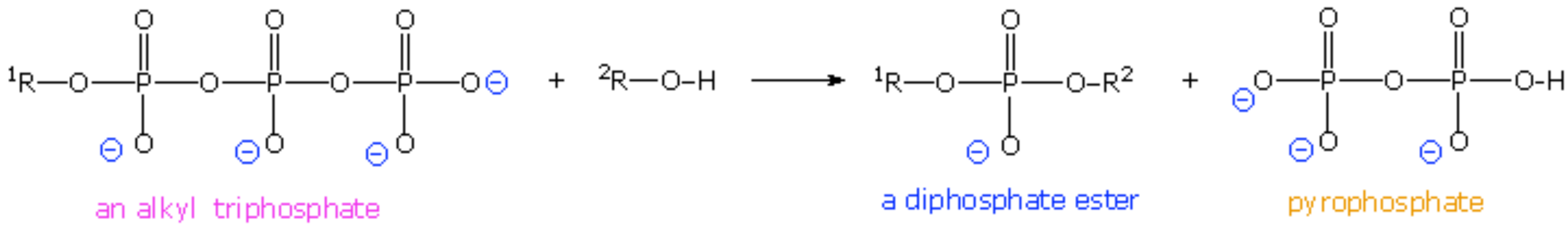
II. Instantiating chemical measures of information

- Would like a model that is equivocally metabolic and evolutionary
- A literal *sub*system is more intuitive than an abstract vision of “life”
- Consider cycles to leverage the Carnot construction from engines

Toy model for metabolism & evolution

Phosphate-driven polymerization

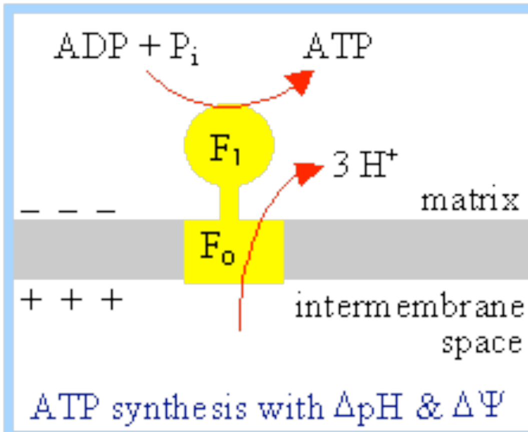
$$N\mathrm{ATP} + \sum_{i=1}^{N+1} \mathrm{M}_{\alpha_i} \rightleftharpoons N\mathrm{AMP} + 2N\mathrm{P}_i + \Pi_{\vec{\alpha}}$$



http://www.cem.msu.edu/~reusch/VirtualText/nucacids.htm

ATP regeneration

$$\mathrm{AMP} + 2\mathrm{P}_i \rightleftharpoons \mathrm{ATP}$$



http://www.rpi.edu/dept/bcbp/molbiochem/MBWeb/mb1/part2/f1fo.htm

(Possibly sequence-dependent) equilibrium relations

$$\left(\frac{[\mathrm{ATP}]}{[\mathrm{AMP}]\,[\mathrm{P}_i]^2}\right)^N = \frac{[\Pi_{\vec{\alpha}}]}{\prod_{z=1}^{Z} [\mathrm{M}_z]^{\nu_z^{\vec{\alpha}}}} K_{\vec{\alpha}}(T)$$

Reactions and chemical work

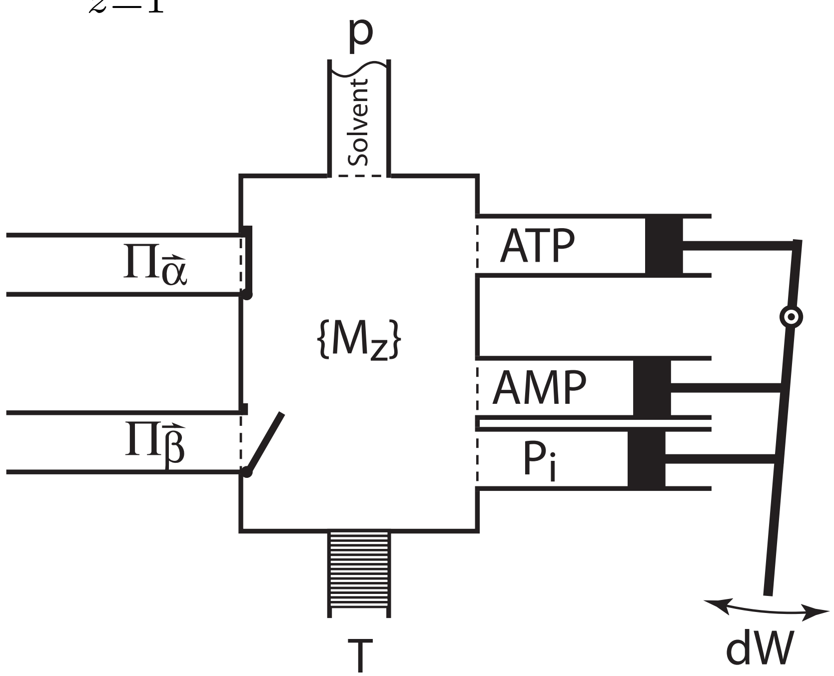
Extensivity $G_X = N_X \mu_X$

Typ. concentration dependence $\mu_X = \bar{\mu}_X + \tau \log\left(\frac{[X]}{[\bar{X}]}\right)$

Partial equilibrium

$$\mu_{\Pi_{\vec{\alpha}}} - \sum_{z=1}^{Z} \nu_z^{\vec{\alpha}} \mu_{M_z} = N \left(\mu_{\mathrm{ATP}} - \mu_{\mathrm{AMP}} - 2\mu_{\mathrm{P}_i}\right)$$

The “van’t Hoff reaction box”

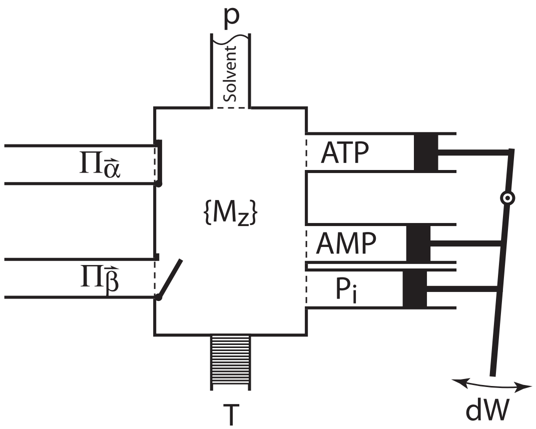


Express chemical work from mechanical work

$$dW \equiv \sum_X dG_X = \sum_X \mu_X dN_X$$

The "chemical Carnot cycle"

$$dW \equiv \sum_X dG_X = \sum_X \mu_X dN_X$$

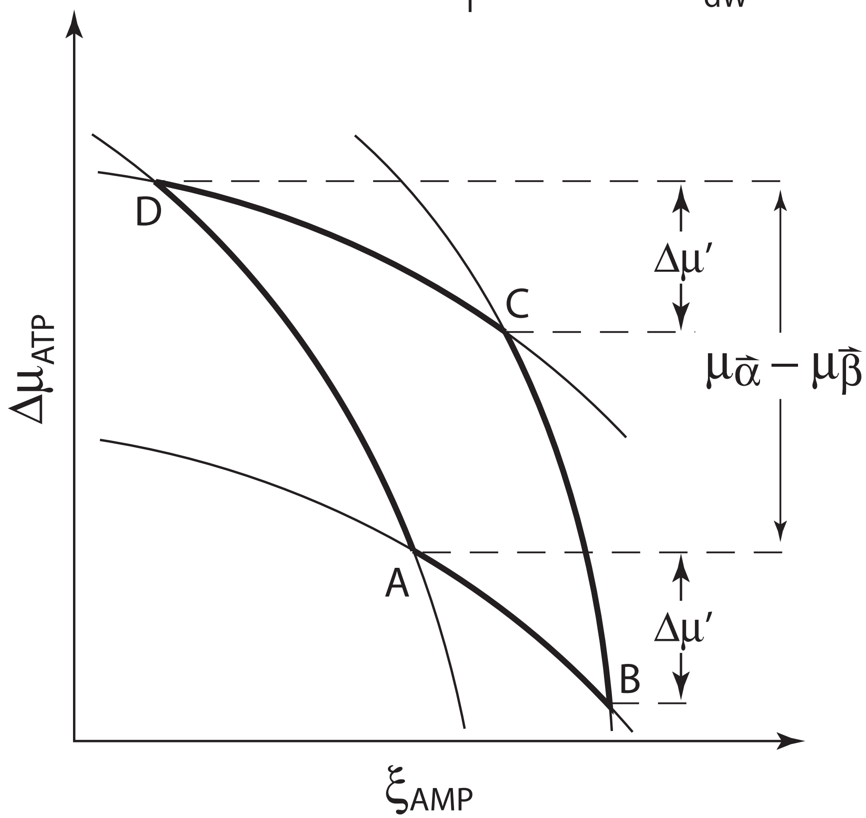


$$\mu_{\Pi_{\vec{\alpha}}} - \sum_{z=1}^{Z} \nu_z^{\vec{\alpha}} \mu_{\mathrm{M}_z} = N\left(\mu_{\mathrm{ATP}} - \mu_{\mathrm{AMP}} - 2\mu_{\mathrm{P}_i}\right)$$

$$\Delta N_{\Pi_{\vec{\alpha}}} = -\Delta N_{\Pi_{\vec{\beta}}}$$

$$\oint dW = \mu_{\Pi_{\vec{\alpha}}} \Delta N_{\Pi_{\vec{\alpha}}} + \mu_{\Pi_{\vec{\beta}}} \Delta N_{\Pi_{\vec{\beta}}} = \Delta G^{CD}_{\Pi_{\vec{\alpha}}} + \Delta G^{AB}_{\Pi_{\vec{\beta}}}$$

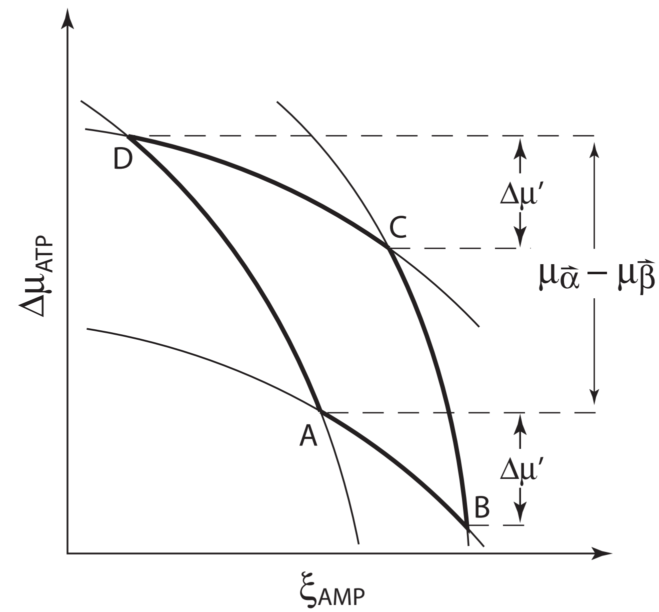
Net work is change in free energies of polymer reservoirs



Chemical "Carnot efficiency"

$$\oint dW = \mu_{\Pi_{\vec\alpha}} \Delta N_{\Pi_{\vec\alpha}} + \mu_{\Pi_{\vec\beta}} \Delta N_{\Pi_{\vec\beta}}$$
$$= \Delta G^{CD}_{\Pi_{\vec\alpha}} + \Delta G^{AB}_{\Pi_{\vec\beta}}$$

$$\Delta N_{\Pi_{\vec\alpha}} = -\Delta N_{\Pi_{\vec\beta}}$$



- Chemical work = area inside the "Carnot" box

- Efficiency relates total work to "capacity" along arc CD

$$\oint dW = \underbrace{\left(1 - \frac{\mu_{\Pi_{\vec\beta}}}{\mu_{\Pi_{\vec\alpha}}}\right)}_{\text{Efficiency}} \Delta G^{CD}_{\Pi_{\vec\alpha}}$$

Bounds between work and entropy

Consider fractions of polymers

$$N_\Pi \equiv \sum_{\vec{\alpha}} N_{\Pi_{\vec{\alpha}}}$$

$$p_{\vec{\alpha}} \equiv \frac{N_{\Pi_{\vec{\alpha}}}}{N_\Pi} = \frac{[\Pi_{\vec{\alpha}}]}{\sum_{\vec{\alpha}} [\Pi_{\vec{\alpha}}]}$$

Dilute-solution chemical potentials

$$dW \equiv \sum_X dG_X = \sum_X \mu_X dN_X$$

$$\mu_X = \bar{\mu}_X + \tau \log \left(\frac{[X]}{[\bar{X}]} \right)$$

- Express cycle work as function of distributions relative to equilibrium

$$\begin{aligned} \oint dW &= N_\Pi \tau \sum_{\vec{\alpha}} \oint dp_{\vec{\alpha}} \log \frac{p_{\vec{\alpha}}}{\pi_{\vec{\alpha}}} \\ &= N_\Pi \tau \oint dD(p \parallel \pi) \end{aligned}$$

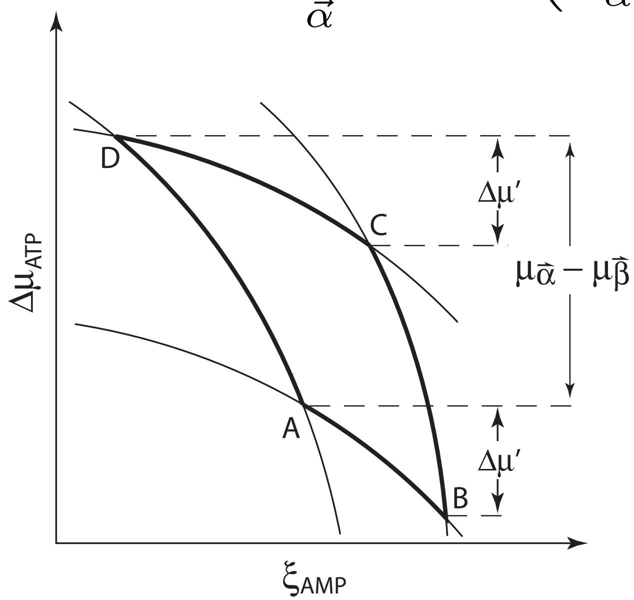
- Kullback-Leibler divergence, or “relative entropy”

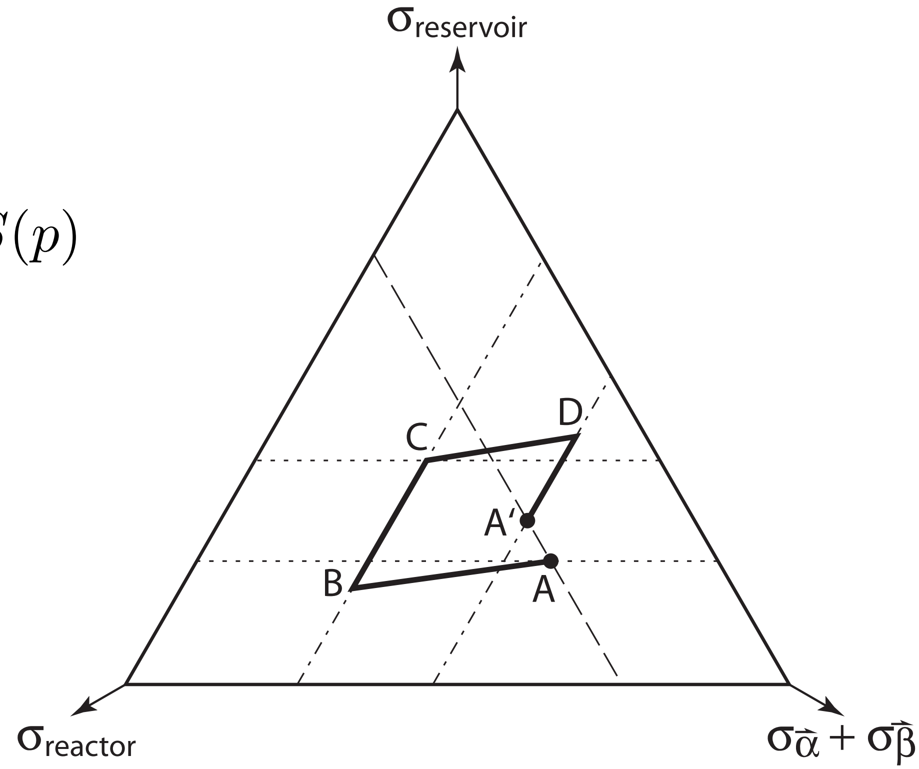
$$D(p \parallel \pi) \equiv \sum_{\vec{\alpha}} p_{\vec{\alpha}} \log \frac{p_{\vec{\alpha}}}{\pi_{\vec{\alpha}}}$$

The energy/entropy representation

$$\oint dW = N_{\Pi} \tau \oint dD(p \parallel \pi)$$

$$D(p \parallel \pi) = \sum_{\vec{\alpha}} p_{\vec{\alpha}} \log\left(\frac{1}{\pi_{\vec{\alpha}}}\right) - S(p)$$





$$\oint dW - \oint dH = -\tau \oint d\sigma.$$

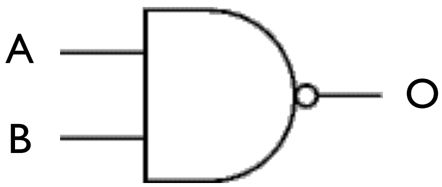
(Chain rule) $$\tau dD(p \parallel \pi) = \sum_{\vec{\alpha}} dp_{\vec{\alpha}} h^{0}_{\Pi_{\vec{\alpha}}} - \tau \left(dS(p) + \sum_{\vec{\alpha}} dp_{\vec{\alpha}} \sigma^{0}_{\Pi_{\vec{\alpha}}} \right)$$

III. The parallel thermodynamics of computation

- Can we attach a minimum energy cost to algorithms, and not merely machines?
- Does the cost aggregate in the same manner as the logic of computation?
- What relation of computation to chemistry?

Attaching energetic costs to algorithms

- All computable functions can be generated from a *finite* list of primitive Boolean operations
- Decompose every such operation into input, logic, output, and erasure
- Recognize that input, logic, and output can be done *reversibly*
- *Erasure alone* converts data entropy to heat entropy
- The cost of a computation is the cost of the erasures it requires

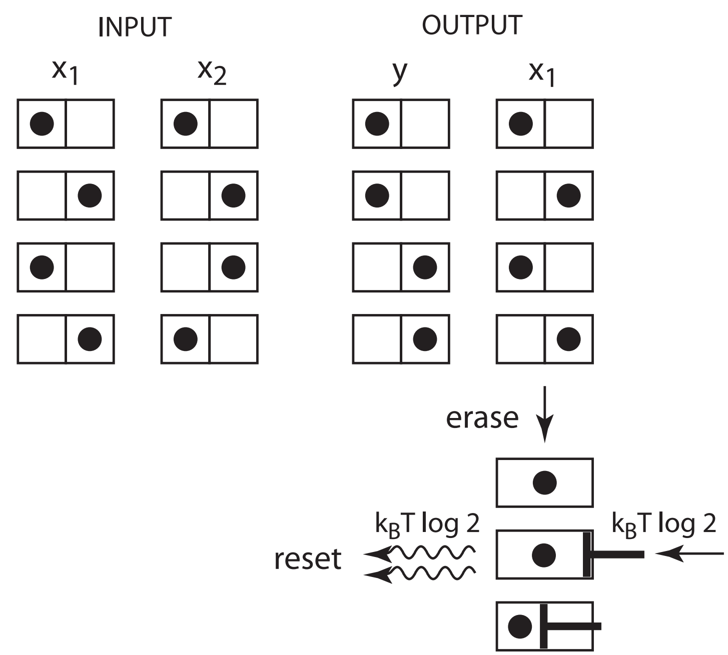


A	B	O
0	0	1
0	1	1
1	0	1
1	1	0

Example: the Szilard single-particle gas

- Consider ideal calculation of XOR
- Input: two IID binary streams
- Output: one IID binary stream
- "Parity"-entropy of output is a component of input entropy
- Sign(x_1)-entropy of input stream is rejected to heat bath

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

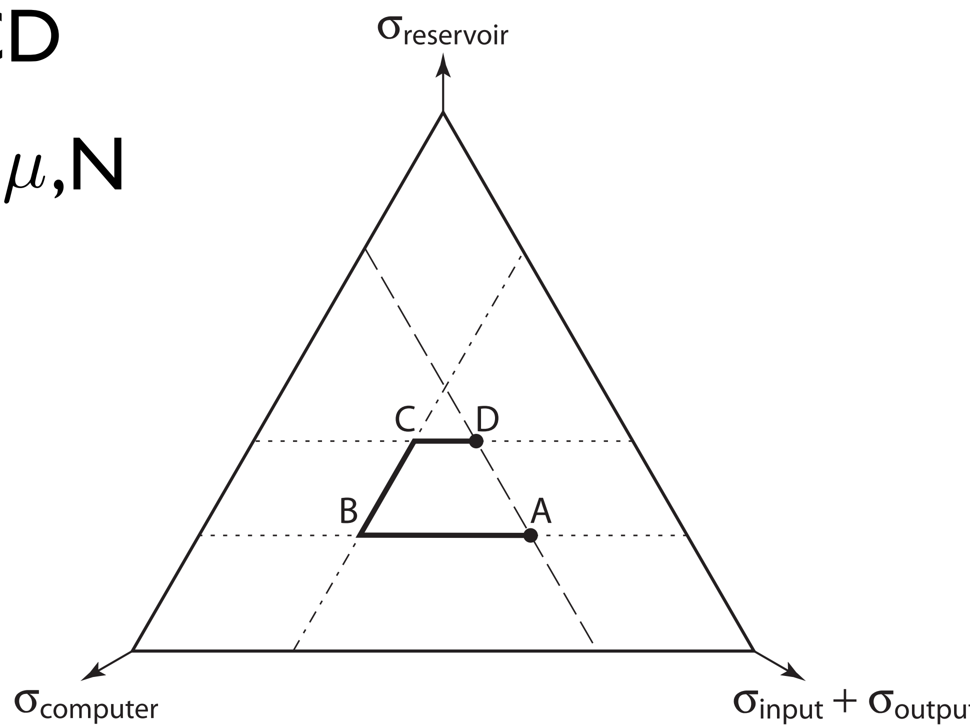


"Landauer's principle"

$$S(X) = S(Y) + Q/T$$

The "Landauer cycle"

- Intake of data bits from high-entropy input stream is arc AB
- Erasure/rejection of heat is BC
- Rejection of data bits to low-entropy output stream is arc CD
- Data take the place of μ,N in chemistry



The Landauer cycle *is* the chemical Carnot cycle

Links of computation to chemistry

- Temperature and entropy are universals for heat engines, chemistry, and computation
- Chemical-number variables are the novelty; correspond to data streams in computation
- Ensemble treatment of data is equivalent to ensemble treatment of molecular arrangement (a new insight for computation from chemistry)

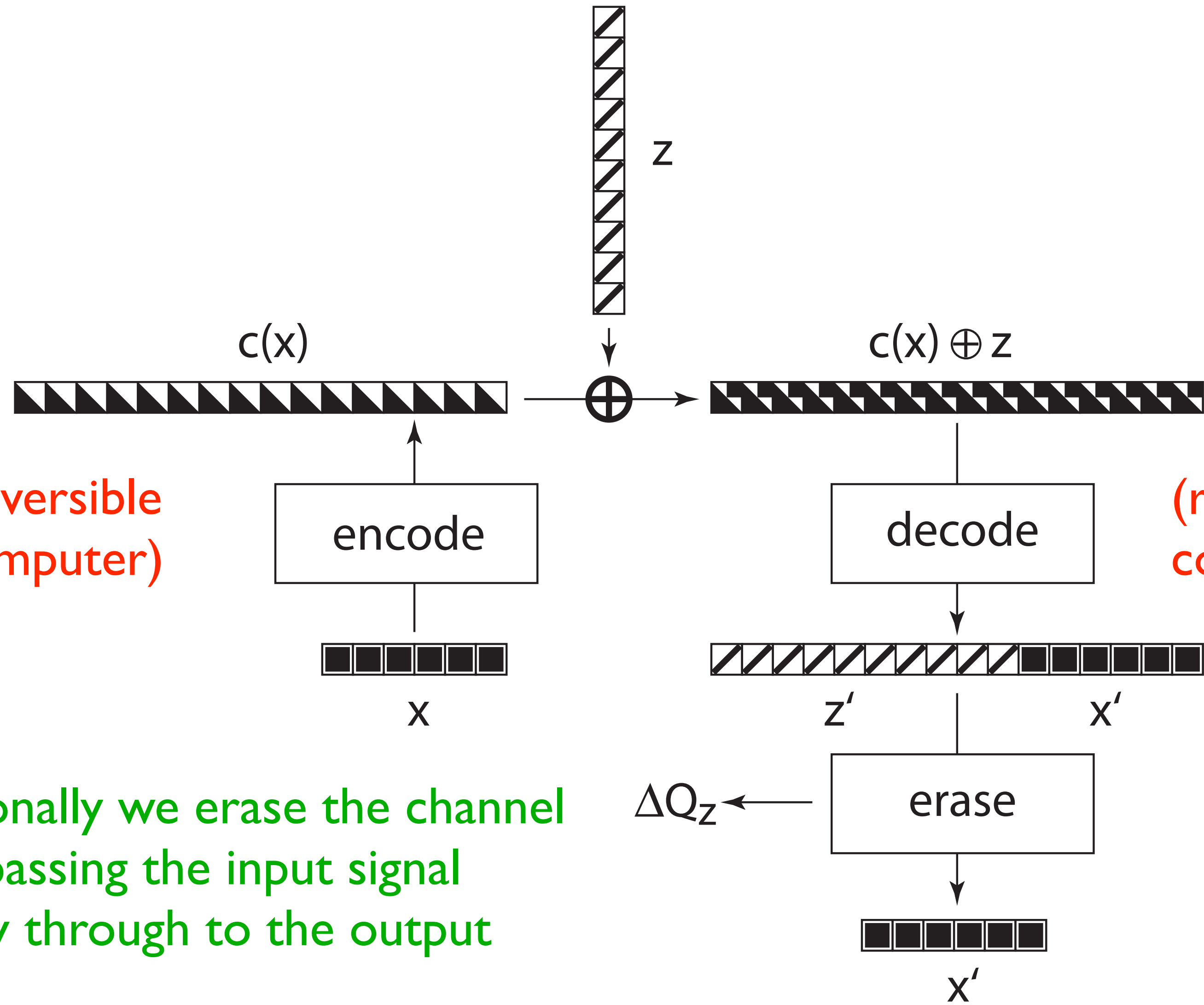
A chemical application of computational theory (Tom Schneider)

http://www-lmmb.ncifcrf.gov/~toms/

- Classic information theory problem: reliable signal communication over noisy channels
- Concept of *error-correcting encoding* can be formulated as a computation problem
- Optimal error correction can be assigned an energetic cost
- Through the Landauer-chemistry map, same ideas can be applied to optimal molecular recognition

Computation in relation to error-correcting encoding

(reversible computer)

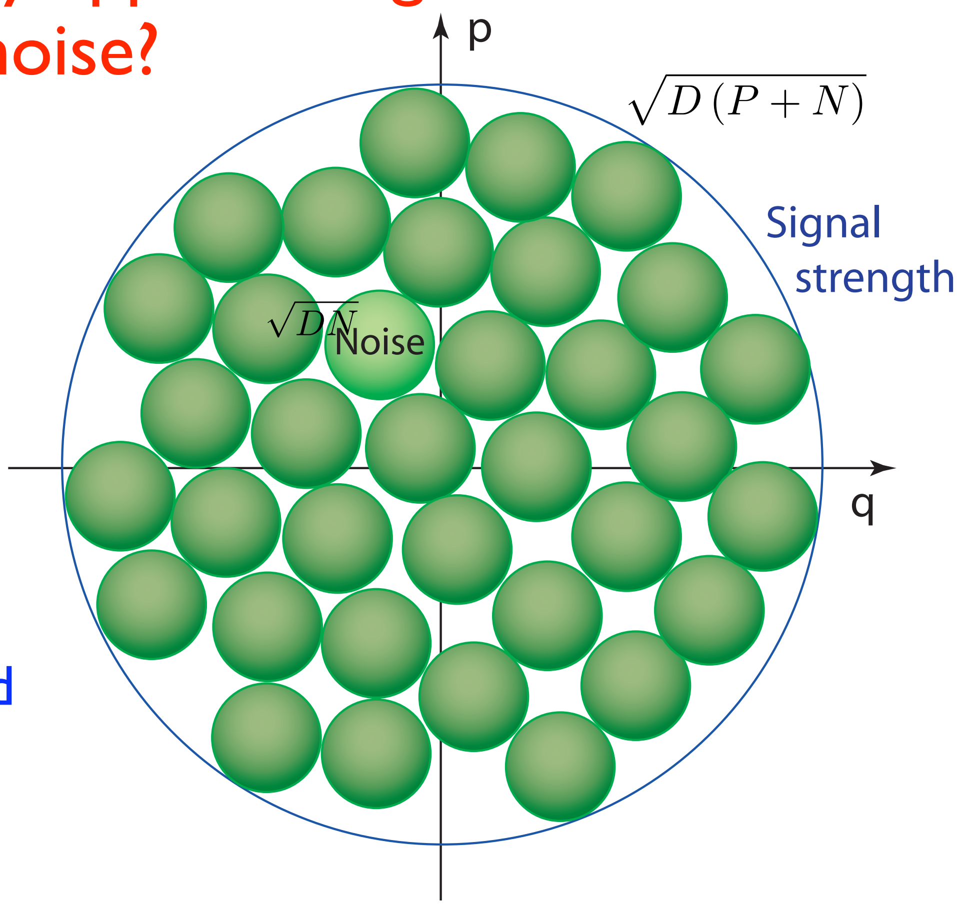


(reversible computer)

Traditionally we erase the channel noise, passing the input signal entropy through to the output

Shannon's theorem for channel capacity (Gaussian channel)

Q: Can we encode messages so that they can be recovered with probability approaching unity, even at finite channel noise?



Fill D-bit code space with maximally distant spheres

$$\frac{[D\,(P+N)]^D}{[DN]^D} \sim \left(\frac{P+N}{N}\right)^D$$

$$= e^{D\log\left(\frac{P+N}{N}\right)}$$

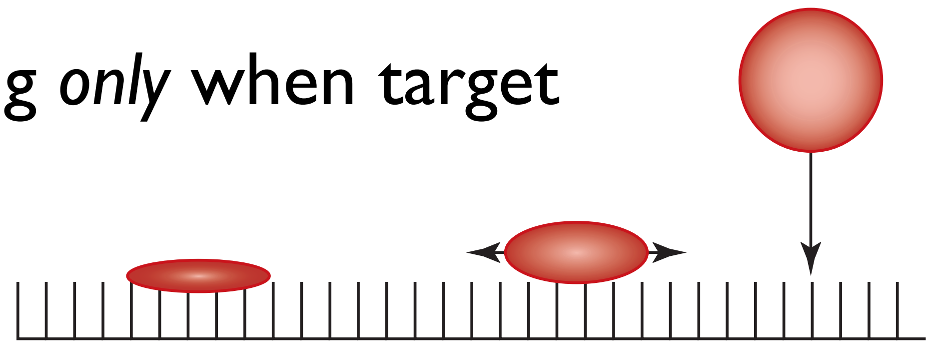
Channel capacity per symbol transmitted

$$C = \frac{1}{2}\log\left(\frac{P+N}{N}\right)$$

Optimal molecular recognition

Q: What is the minimal energy cost to enable a protein to reliably select a single sequence from a suite of random possibilities?

- "Prime" a protein in solution (introduce internal energy to stress its conformation)
- Allow binding to a random site on DNA or RNA
- Allow priming energy to relax as protein migrates along chain, as a function of sequence
- Reliably stop migrating *only* when target sequence is found



Schneider's Shannon theorem for reliable discrimination

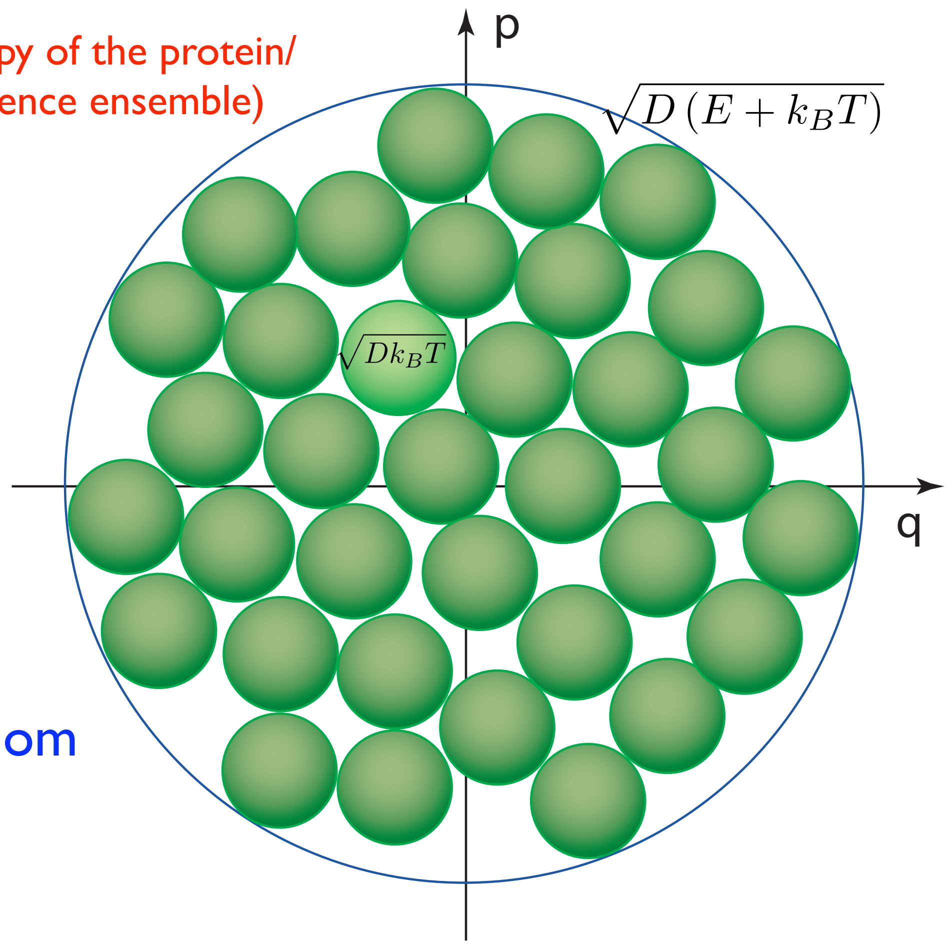
http://www-lmmb.ncifcrf.gov/~toms/

Priming (enthalpy) provides energy for D non-covalent associations

$$\oint dW - \oint dH = -\tau \oint d\sigma$$

(Entropy of the protein/ sequence ensemble)

"$D \times E$"



Coordinate the 2D binding affinities

$$\frac{\left[D\,(E + k_B T)\right]^D}{\left[D k_B T\right]^D} \sim \left(\frac{E + k_B T}{k_B T}\right)^D = e^{D \log\left(\frac{E + k_B T}{k_B T}\right)}$$

"Machine capacity" per degree of freedom

$$C = \frac{1}{2} \log\left(\frac{E + k_B T}{k_B T}\right)$$

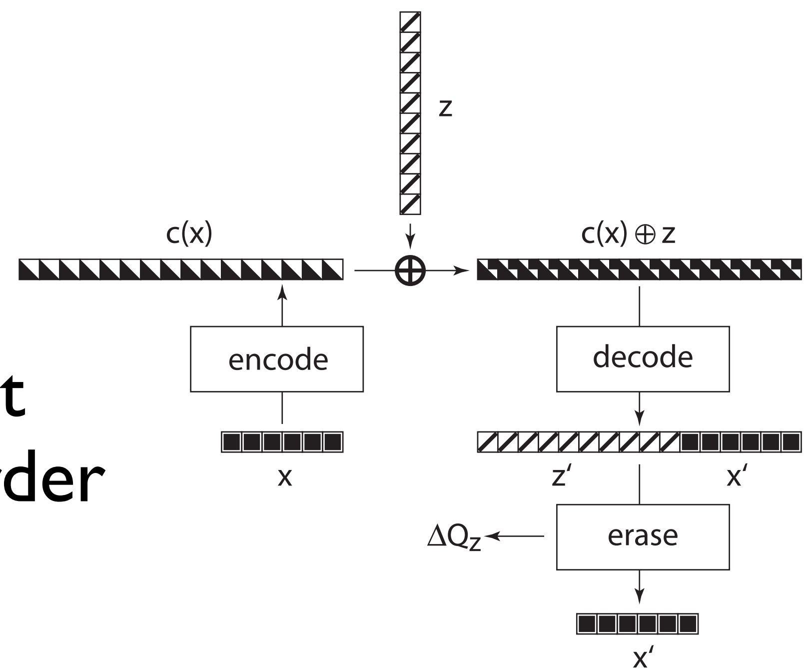
Channel versus molecule problems

- "Priming" energy corresponds to signal power; kT corresponds to channel noise in Shannon bound

$$C = \frac{1}{2} \log \left(\frac{P + N}{N} \right)$$

$$C = \frac{1}{2} \log \left(\frac{E + k_B T}{k_B T} \right)$$

- Shannon erases the noise power; Schneider erases the "signal"



- This use of enthalpy to reject entropy is the math of 1st-order phase transition

Concluding thoughts

- Kinetics of the ensembles of life lend themselves to a machine-like description
- Equilibrium bounds on energy and information work better than they “should”
- Carnot-like decompositions give clarity to both metabolism and evolution
- We have a principled map between chemistry and computation

Some Further Reading

- T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991)
- E. Fermi, *Thermodynamics* (Dover, New York, 1956)
- C. Kittel and H. Kroemer, *Thermal Physics,* (Freeman, New York, 1980)
- E. Smith, Thermodynamics of Natural Selection I - III, *J. Theor. Biol.* http://dx.doi.org/10.1016/j.jtbi.2008.02.010, 008, 013 or SFI preprint #06-03-011