
©Denis Hawkins

THE AUSTRALIAN NATIONAL UNIVERSITY

# Maximum entropy & maximum entropy production in biological systems: survival of the likeliest?

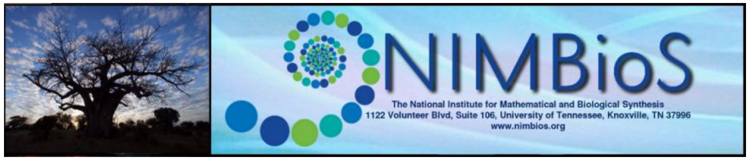
*Roderick Dewar*

*Research School of Biology*

*The Australian National University, Canberra*

***Information and Entropy in Biological Systems***

*8–10 April 2015, NIMBIOS*



# Maximum Entropy

# (MaxEnt)

# Maximum (relative) entropy:

Maximise $H = -\sum_i p_i \ln \frac{p_i}{q_i}$ w.r.t. $p_i$ subject to

$$\begin{cases} \sum_i p_i x_i = X \\ \sum_i p_i = 1 \end{cases}$$

↑

constraints (*C*)

$$\therefore \; p_{i|C} = \frac{q_i e^{-\beta x_i}}{\sum_i q_i e^{-\beta x_i}}$$

## What is the rationale for MaxEnt ?

- information theory (least biased $p_{i|C}$ )
- combinatorics of sample frequencies (most likely $p_{i|C}$)

# MaxEnt: the combinatorial rationale

$N$ independent observations

$M$ possible outcomes for each observation, $i$ = 1 ….M

$q_i$ = prior probability of outcome $i$

Outcome $i$ observed $n_i$ times → frequency distribution $p_i = \frac{n_i}{N}$

$$\Pr(\{n_i\}) = N! \prod_{i=1}^{M} \frac{q_i^{\,n_i}}{n_i!}$$

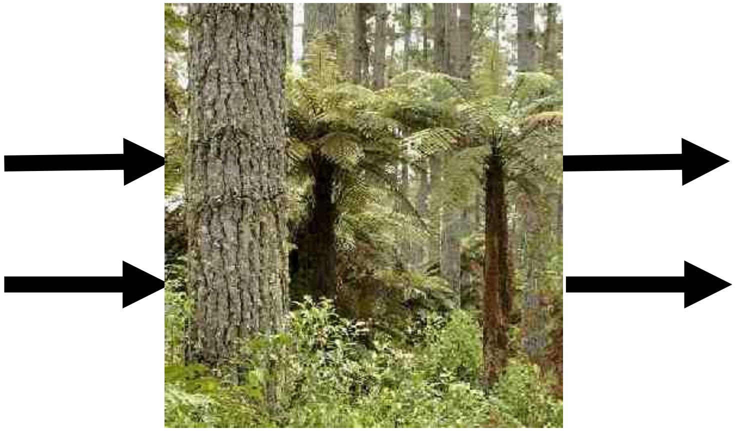
$$\lim_{N \to \infty} \frac{1}{N} \ln \Pr(\{n_i\}) = -\sum_{i=1}^{M} p_i \ln \frac{p_i}{q_i}$$ = relative entropy of $p_i$ and $q_i$

∴ The MaxEnt distribution $p_{i|C}$ is **by far the most likely** long-term frequency distribution of outcomes, among all those distributions consistent with given constraints $C$

(transparent connection to observations)

# The prediction challenge in biology:

biological systems are complex, open,  non-equilibrium



energy in

matter in

energy out

matter out

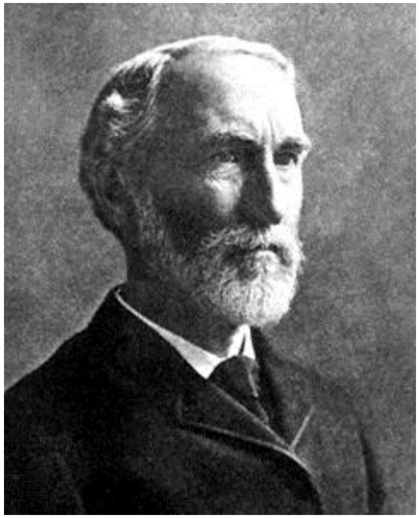
*many internal degrees of freedom*

# Statistical mechanics:

Some (many?) details of the underlying dynamics are irrelevant for making predictions at larger scales

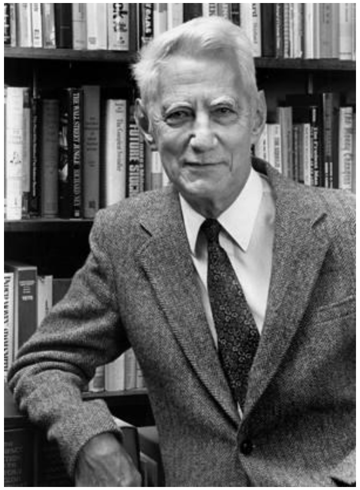
# statistical mechanics

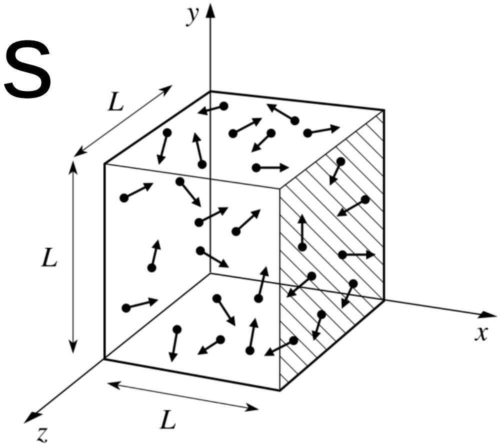
*Boltzmann*



*Gibbs*

*Shannon*





maximum (relative) entropy

↓

*Jaynes*

~~detailed dynamics~~ x($t$)

most likely

$p(x \mid \text{key dynamical constraints})$

# Maximum (relative) entropy:

Maximise $H = -\sum_i p_i \ln \frac{p_i}{q_i}$ w.r.t. $p_i$ subject to $\begin{cases} \sum_i p_i x_i = X \\ \sum_i p_i = 1 \end{cases}$

$$\therefore \; p_{i|C} = \frac{q_i e^{-\beta x_i}}{\sum_i q_i e^{-\beta x_i}}$$

↑

constraints (*C*)

**What do the constraints represent ?**

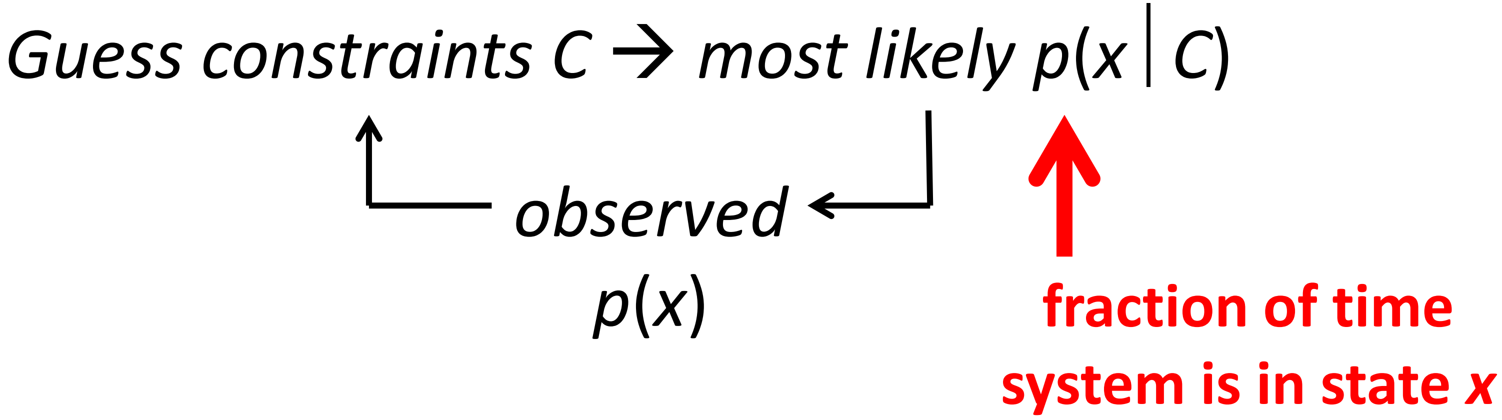
- information theory: *C* = what we know
- statistical mechanics: *C* = the relevant dynamics (key resource constraints, steady-state balance …)

# The role of MaxEnt in statistical mechanics

- **MaxEnt as a statistical selection principle**

*Known constraints* $C \rightarrow$ *most likely* $p(x \mid C)$

- **MaxEnt as a tool for identifying the relevant dynamics (*C*)**



*Guess constraints* $C \rightarrow$ *most likely* $p(x \mid C)$

*observed*
$p(x)$

**fraction of time system is in state *x***

# Combining mechanism and drift* in ecology

*Bertram & Dewar (2015)*

*Relevant dynamics* $C$ + *MaxEnt* → *most likely* $p(x \mid C)$

**mechanism**

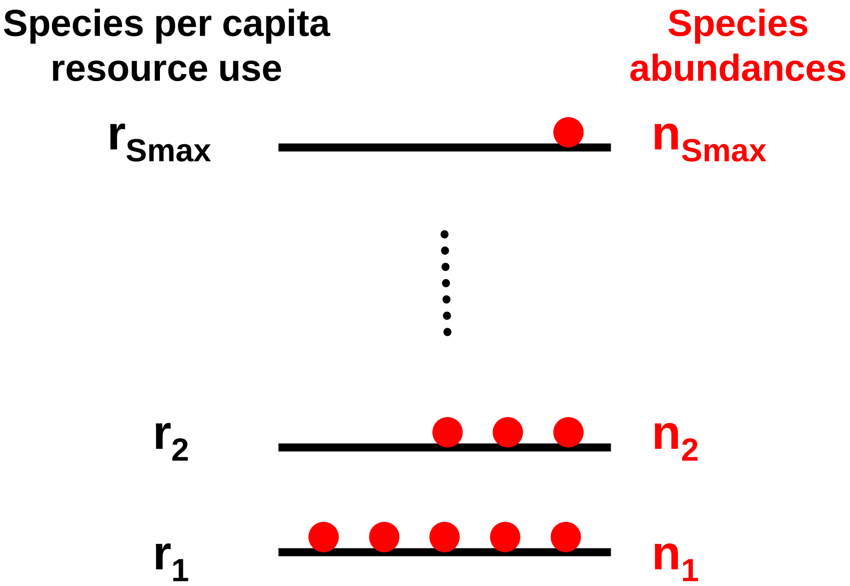
**drift***

*** in the biological sense!**

# MaxEnt: a non-neutral, resource-based approach

*Dewar & Porté (2008), Bertram & Dewar (2013, 2015)*

## *mechanism*



**Mean annual resource balance :**
**(mean use = mean supply)**

$$\sum_{i=1}^{Smax} \overline{n}_i r_i = \overline{R}$$

**Area constraint (hard) :**

$$\sum_{i=1}^{Smax} n_i a_i = A$$

## *drift*

**MaxEnt**

↓

$$p(n_1, n_2 \ldots. n_{Smax}|R)$$

**(≈ Bose-Einstein)**

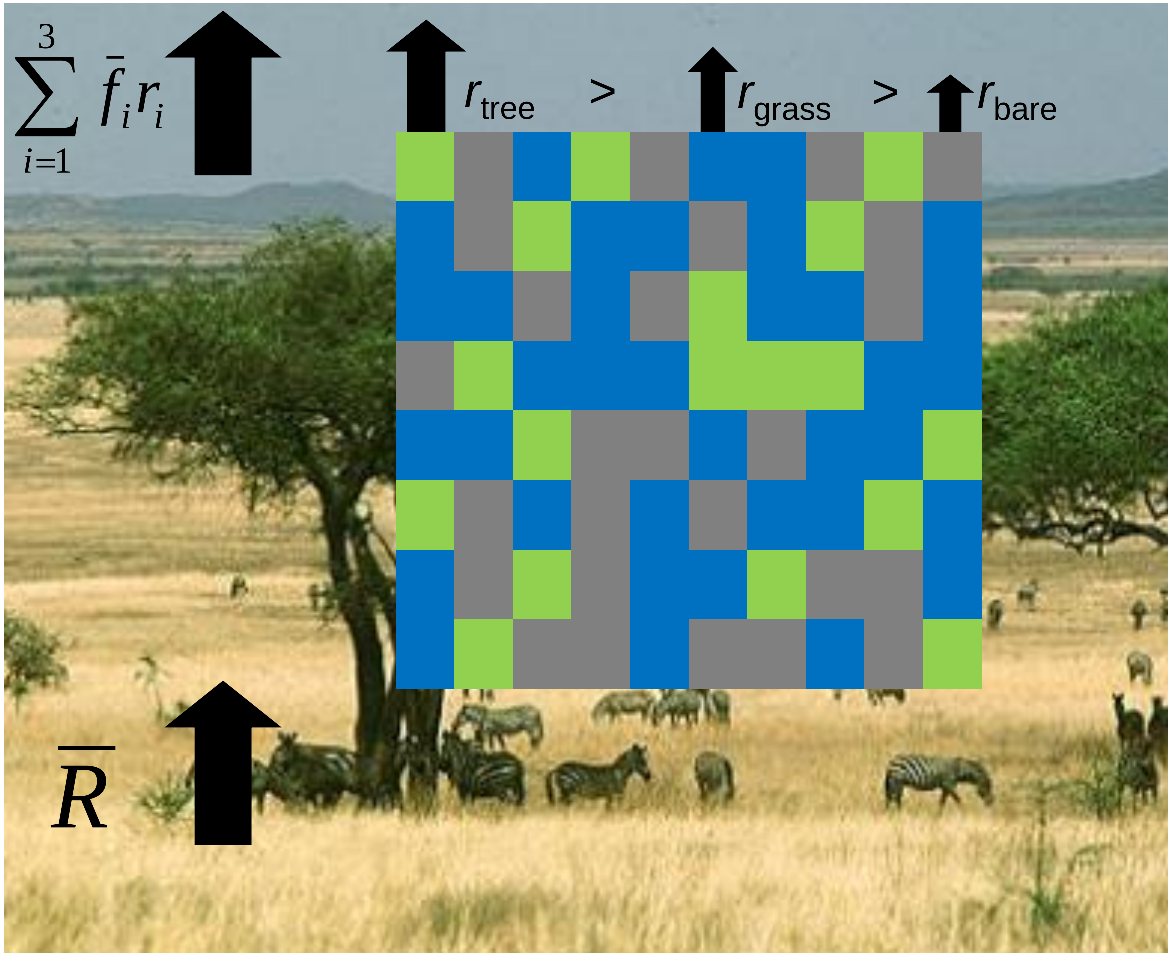
↓

**unifies different macroecological patterns:**

***SAD's***
***energy equivalence***
***diversity-productivity***
***diversity-stability***

# The statistical mechanics of savannas

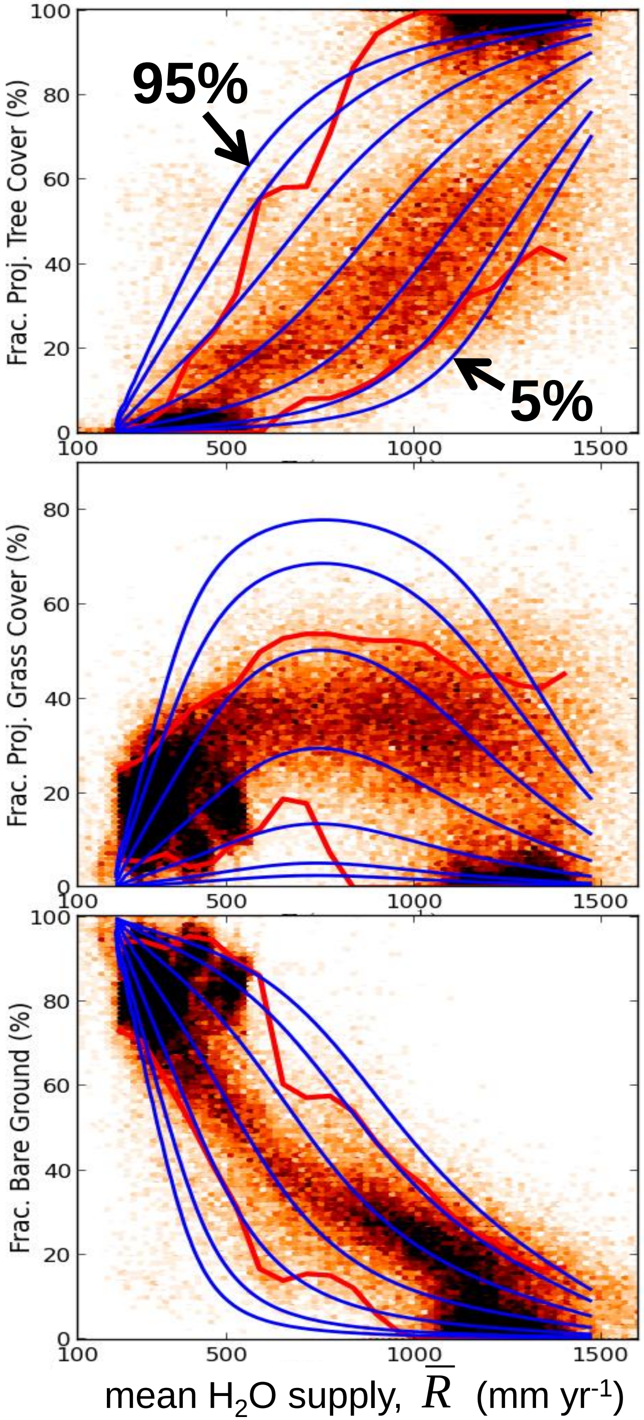
*Bertram & Dewar (2013)*



$\sum_{i=1}^{3} \bar{f}_i r_i$

$r_{\text{tree}} > r_{\text{grass}} > r_{\text{bare}}$

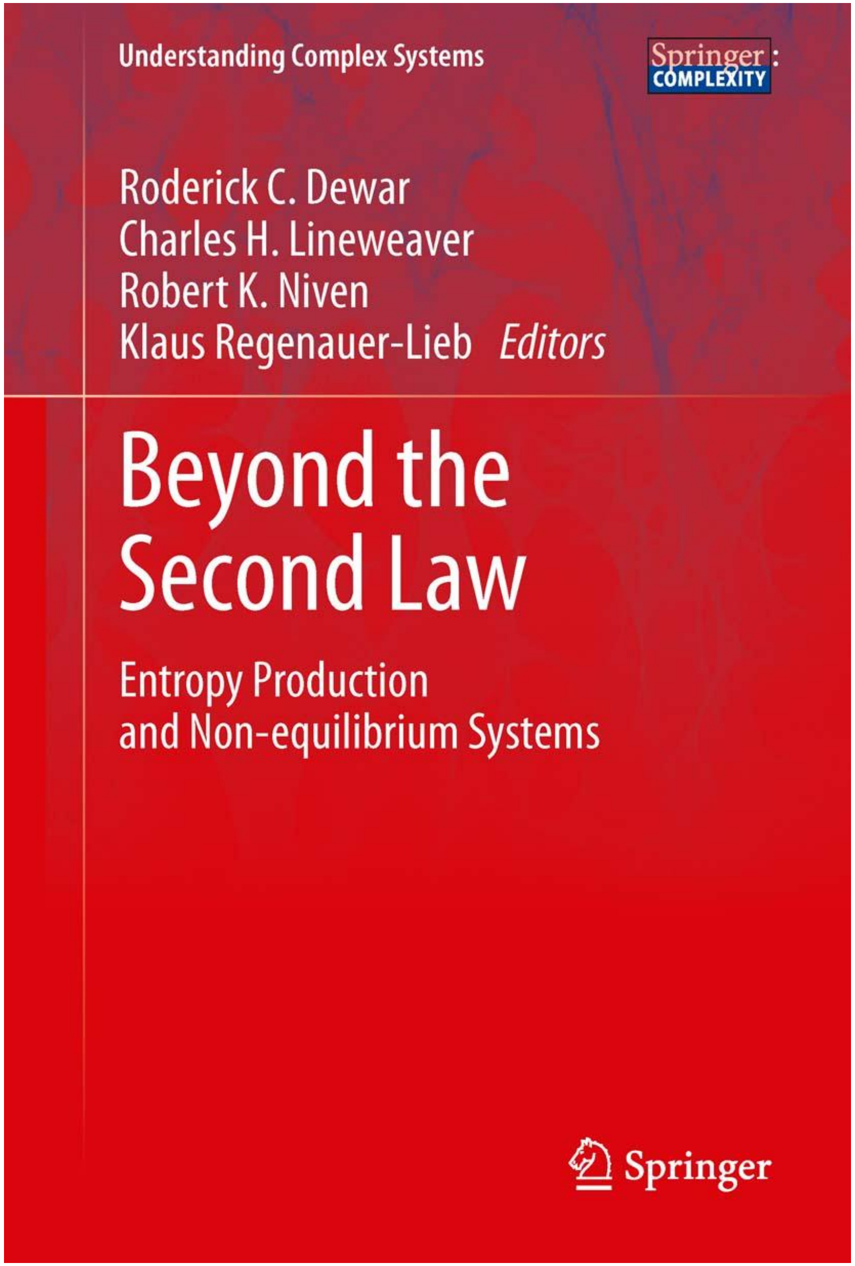
$\bar{R}$



For given $\bar{R}$ what is most likely $p(f_{\text{tree}}, f_{\text{grass}})$?

# Maximum Entropy Production

# (MEP)



Understanding Complex Systems

Springer COMPLEXITY

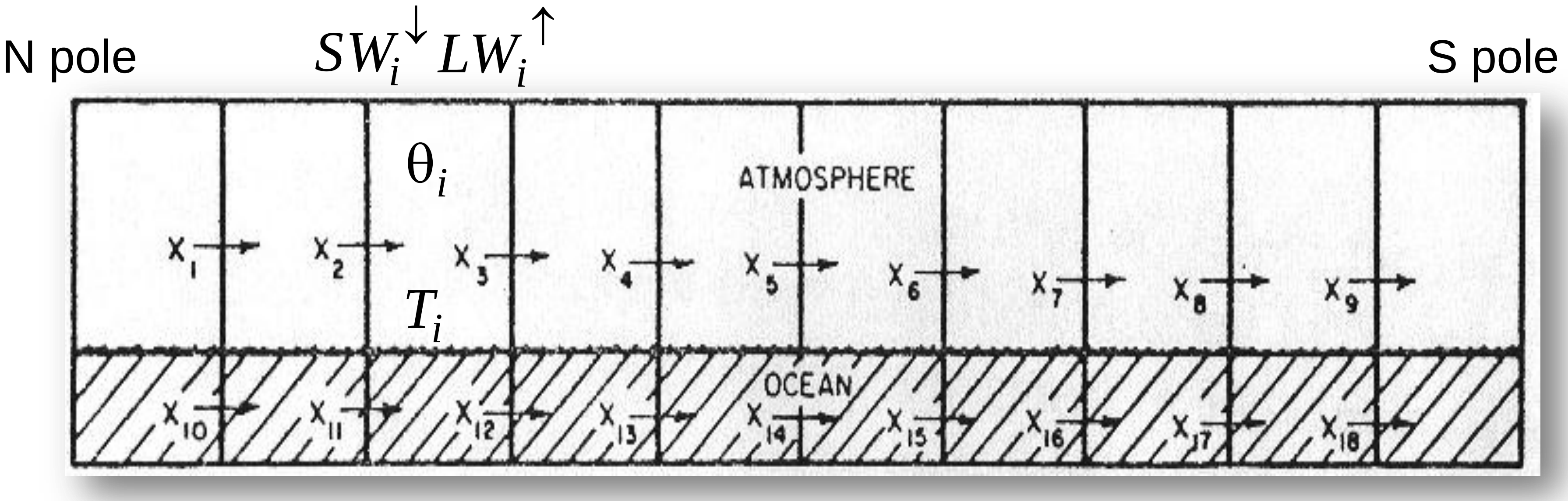
Roderick C. Dewar
Charles H. Lineweaver
Robert K. Niven
Klaus Regenauer-Lieb *Editors*

# Beyond the Second Law

## Entropy Production and Non-equilibrium Systems

Springer

# Paltridge (1978): 10-zone energy balance model



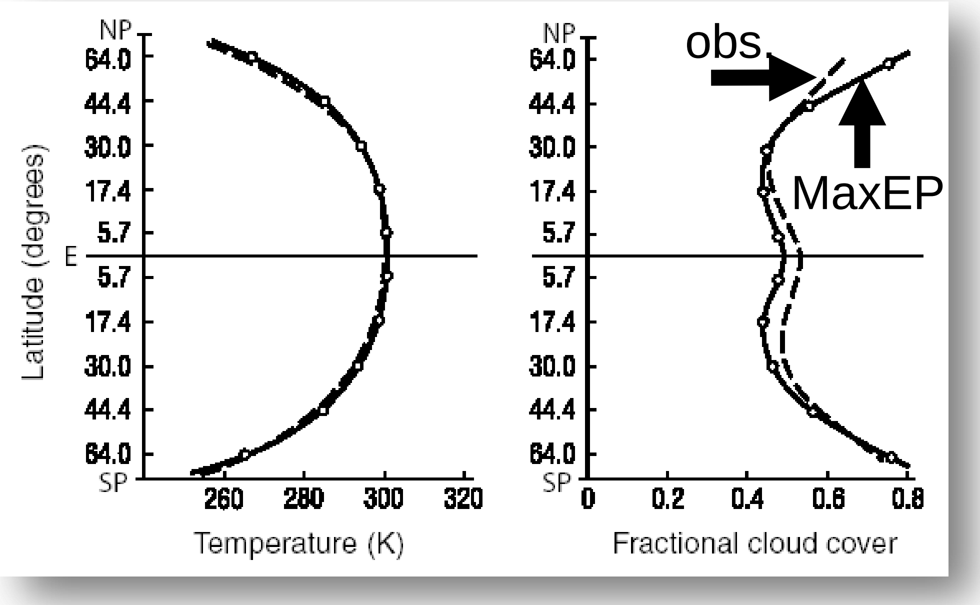
Maximize

$$\boxed{EP \equiv \sum_{i=1}^{10} \frac{LW_i^{\uparrow} - SW_i^{\downarrow}}{T_i} = \sum_{i=1}^{9} X_i \left( \frac{1}{T_{i+1}} - \frac{1}{T_i} \right)}$$

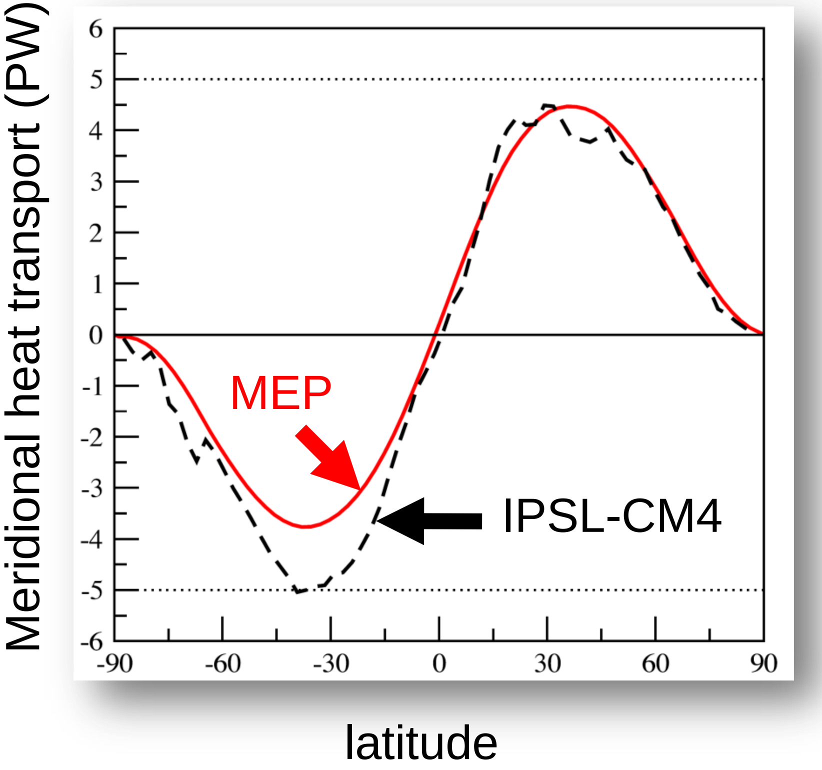
‘entropy production’
(it’s not!)

subject only to steady-state energy balance

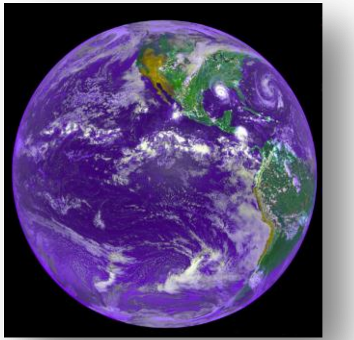
*Paltridge (1978): 1D model*



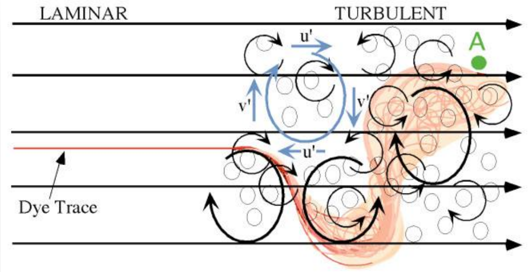
*Herbert & Paillard (2013): 2D model*



# MEP applications across physics & biology

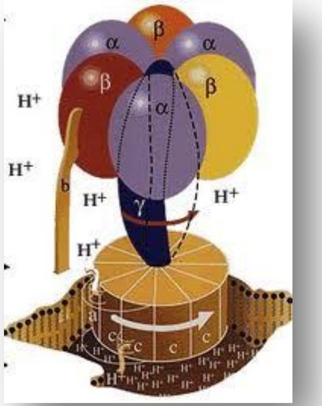


*Paltridge (1978) …*



*Malkus (2003) …*

*Main & Naylor (2008)*

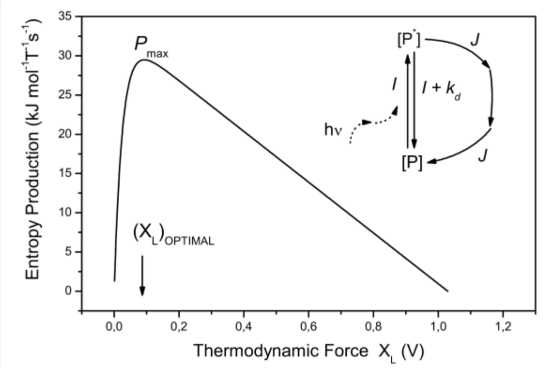


*Dewar et al (2006)*

some intriguing successes,
but what does it mean?

*Martyushev et al (2000)*

*Juretić et al (2003)*



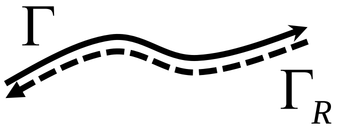
*Dewar (2010)*

# Some potentially misleading statements about Maximum Entropy Production

- MEP is a corollary to the second law ($\mathrm{d}S_{\mathrm{universe}}/\mathrm{d}t \geq 0$) which states that $S_{\mathrm{universe}}$ increases as fast as possible

- MEP means that, over time, EP approaches a maximum in the steady-state

# MEP as a statistical stability criterion for non-equilibrium stationary states



$$D = \sum_{\Gamma} p_{\Gamma} \underbrace{\ln \frac{p_{\Gamma}}{p_{\Gamma_R}}}_{d_{\Gamma}} \geq 0$$ with equality iff $p_{\Gamma_R} = p_{\Gamma}$

- *D* is a generic measure of irreversibility / time-reversal symmetry breaking / distance from equilibrium

- Fluctuation theorem (follows from definition of $d_{\Gamma}$): $$\frac{p(d)}{p(-d)} = e^{d}$$

- If $p(d) = N(D,\sigma^2)$: $$\sigma^2 = 2D \quad \therefore \mathrm{CV} = \frac{\sigma}{D} \propto \frac{1}{\sqrt{D}}$$ is minimal when *D* is maximal

- Use MaxEnt to construct $p_{\Gamma}$ : *D* depends on the constraints *C*

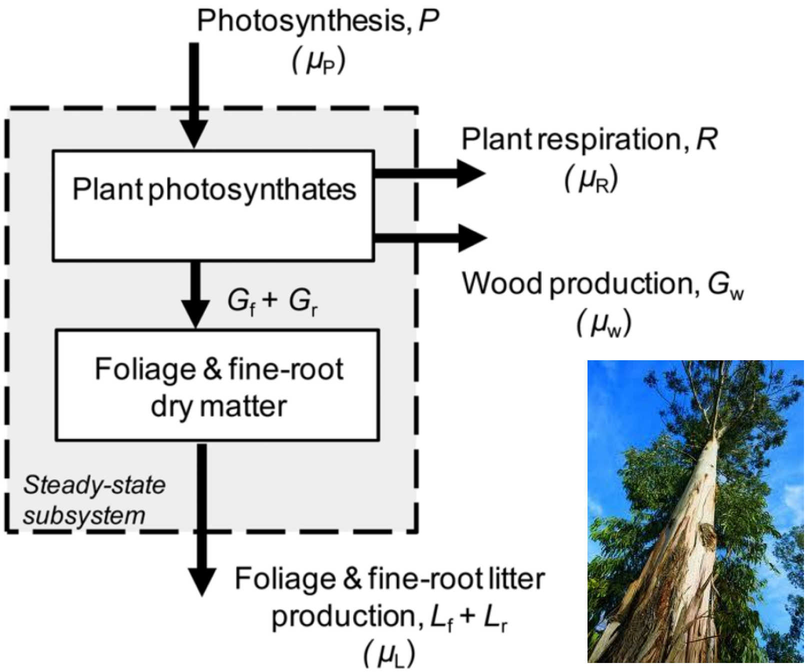
Invited review: Part of an invited issue on carbon allocation

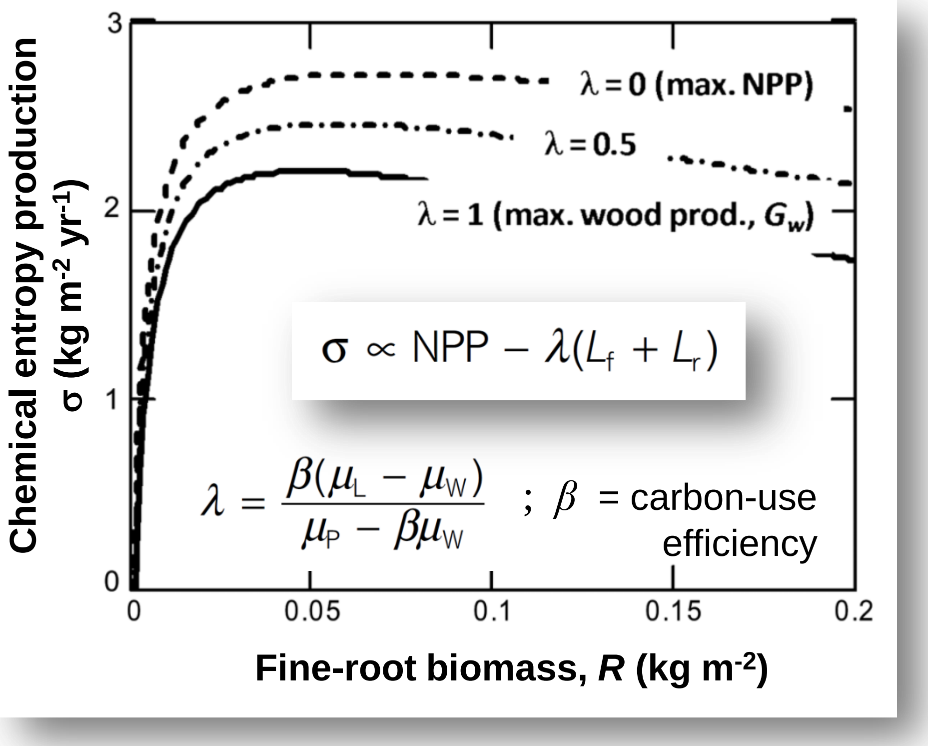
## Modeling carbon allocation in trees: a search for principles

Oskar Franklin[1,7], Jacob Johansson[1,2], Roderick C. Dewar[3], Ulf Dieckmann[1], Ross E. McMurtrie[4], Åke Brännström[1,6] and Ray Dybzinski[5]

*Tree Physiol.* **32**, 648-666 (2012)

# MEP mimics traditional maximum fitness models

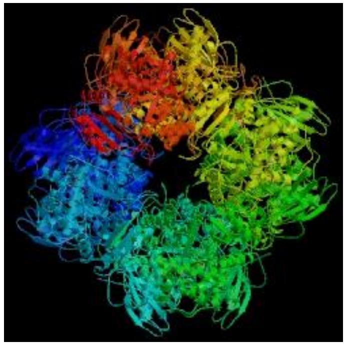




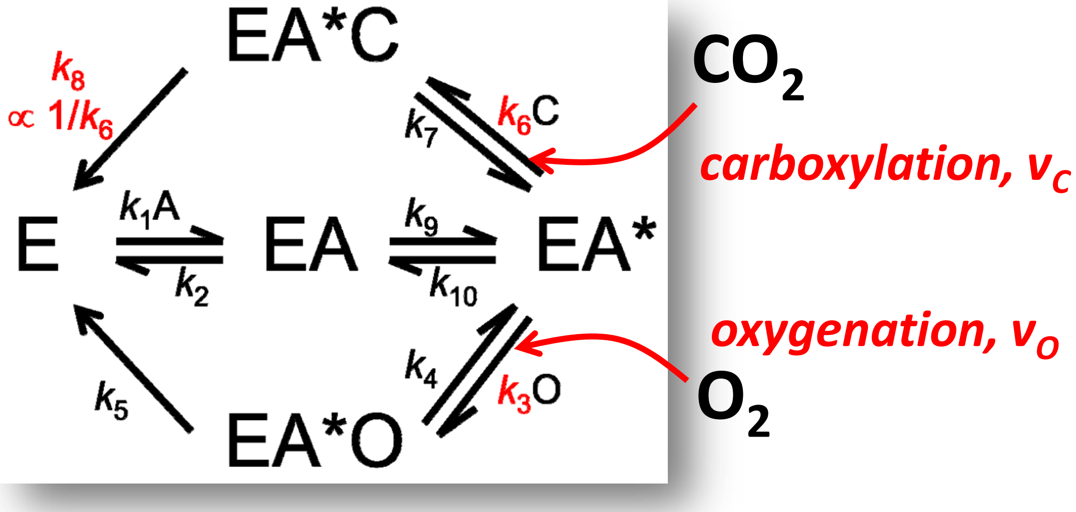
$$\sigma \propto \mathrm{NPP} - \lambda(L_f + L_r)$$

$$\lambda = \frac{\beta(\mu_L - \mu_W)}{\mu_P - \beta\mu_W}$$ ; $\beta$ = carbon-use efficiency

# Evolutionary optimisation of Rubisco: Earth's most abundant protein



### *Enzyme kinetics*



Enzyme state i

Kinetic design ***k*** = ($k_1$, $k_2$ … $k_{10}$)

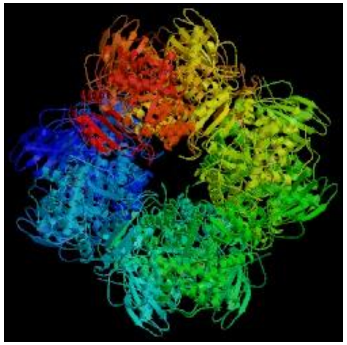
Apply MaxEnt to $p(i,\mathbf{k}) = p(i|\mathbf{k})\,p(\mathbf{k})$

$$\therefore p(\mathbf{k}) \propto e^{H(\mathbf{k})}$$

$$H(\mathbf{k}) = -\sum_i p(i|\mathbf{k}) \ln p(i|\mathbf{k})$$

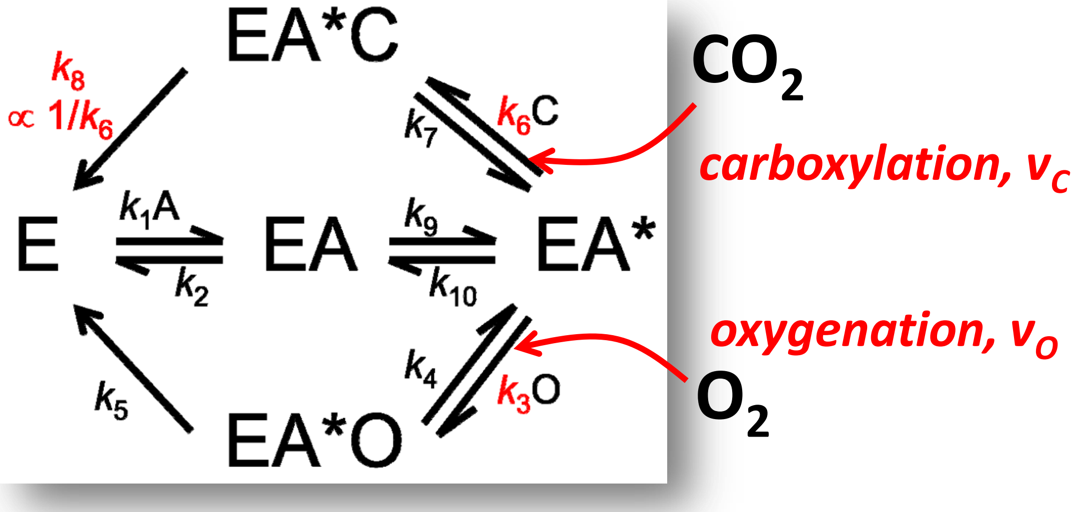
$$EP(\mathbf{k}) = v_C \ln \frac{v_C^+}{v_C^-} + v_O \ln \frac{v_O^+}{v_O^-}$$

Maximise simultaneously
w.r.t. $k_6$ & $k_3$



# Evolutionary optimisation of Rubisco: Earth’s most abundant protein

## *Enzyme kinetics*



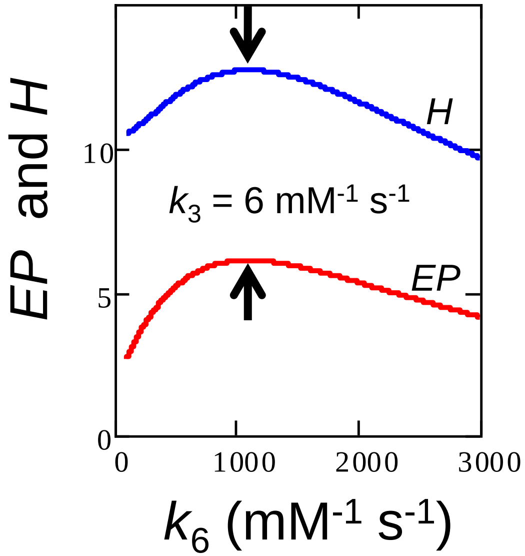
**$CO_2$**

***carboxylation, $v_C$***

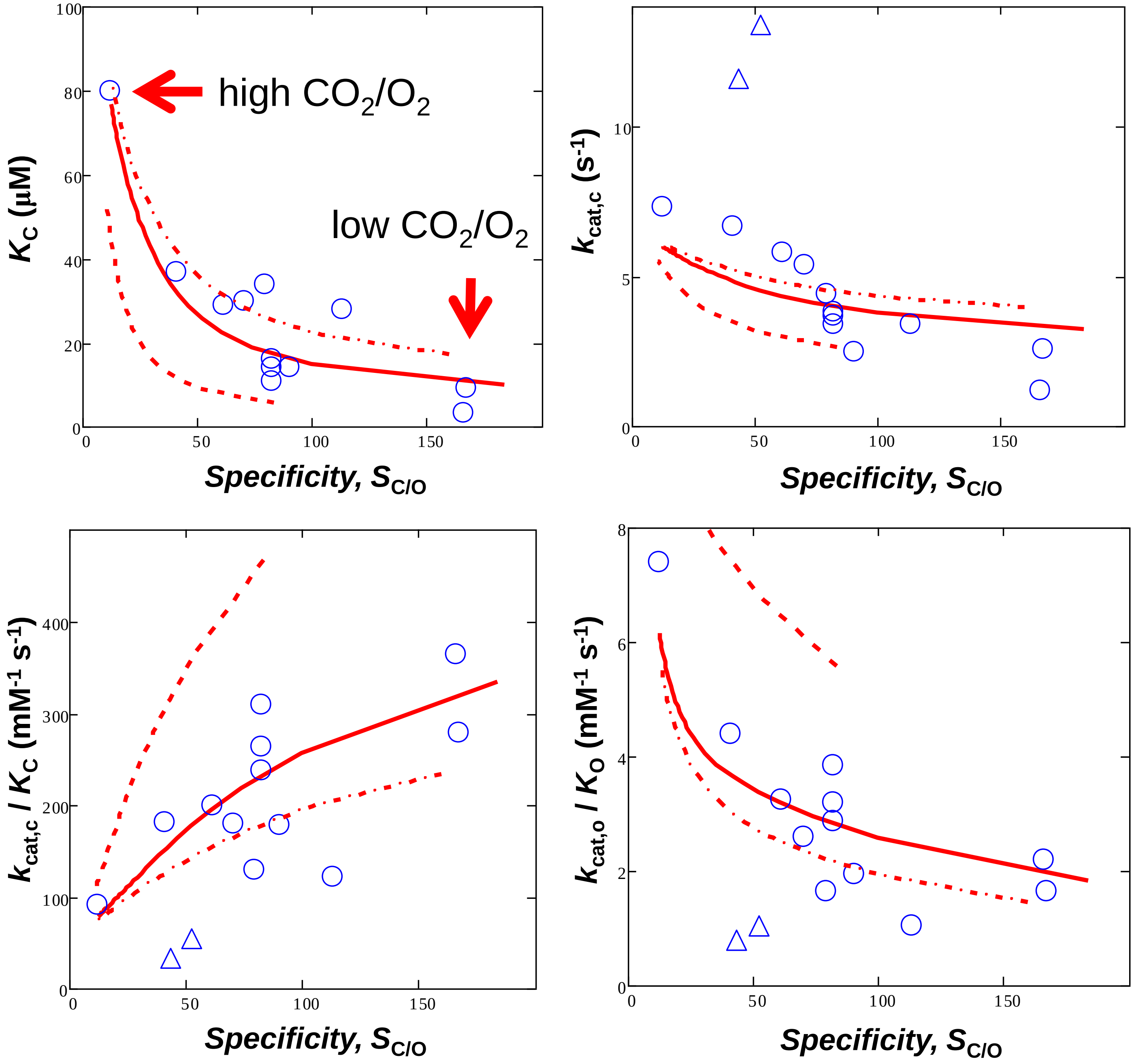
***oxygenation, $v_O$***

**$O_2$**

## *MaxEnt & MEP*



# Rubisco adaptation to different $CO_2/O_2$ environments



MaxEnt & MEP predictions

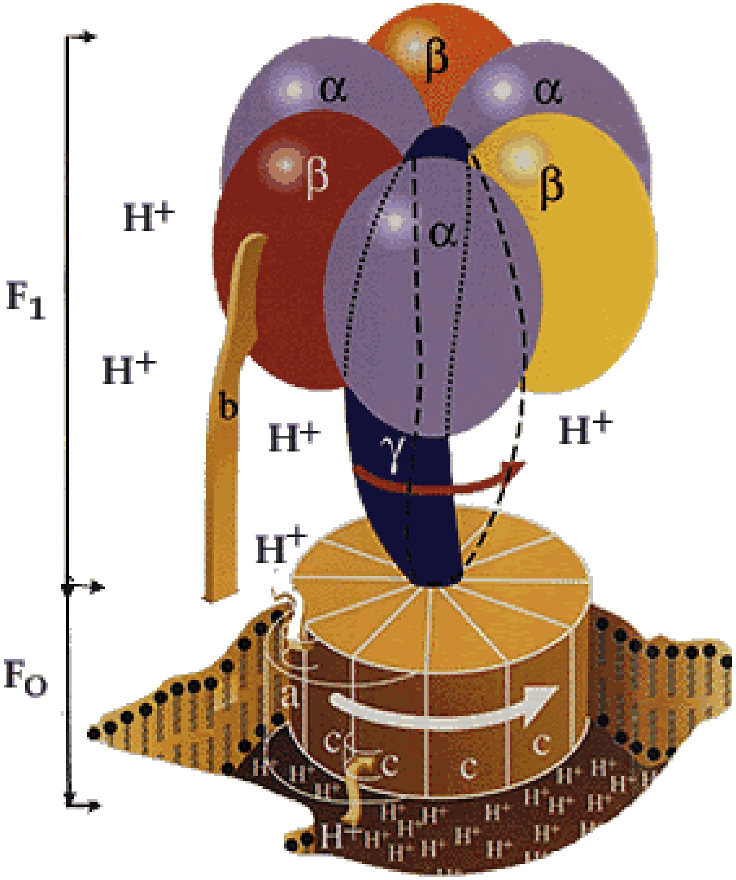
— default parameters

- - - $k_7$ × 1/10

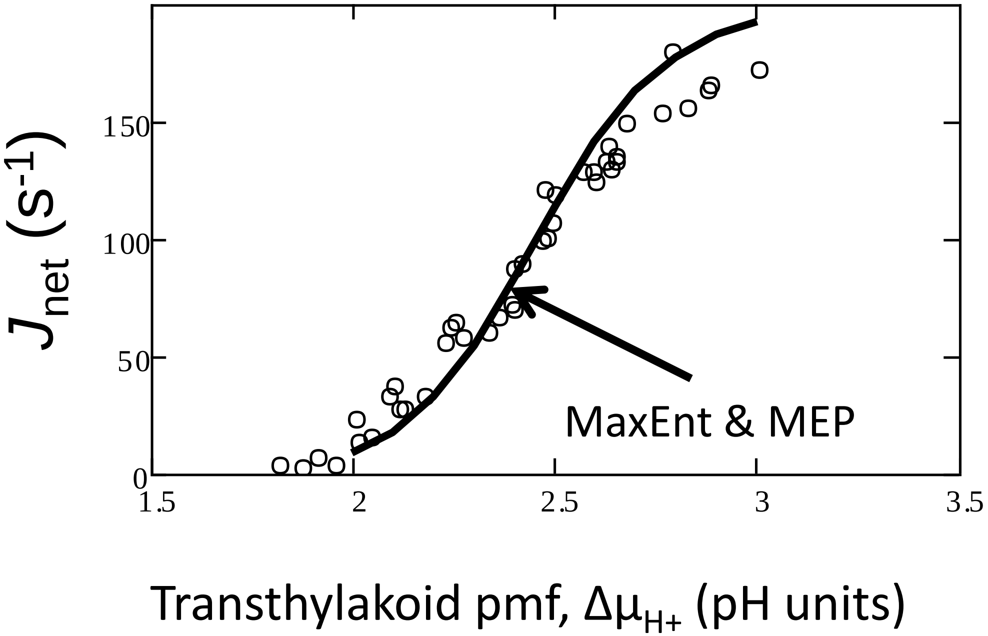
-·- $k_4$ × 1/10

*Dewar et al.*
*(in prep.)*

# $F_0F_1$-ATP synthase : Nature’s smallest rotary motor

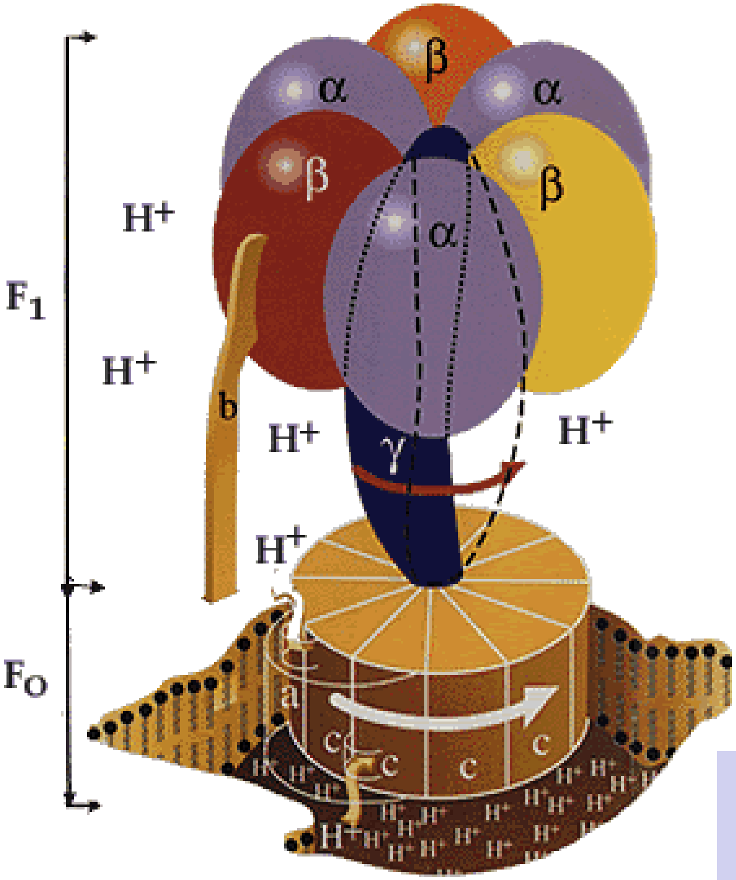


Net ATP synthesis vs. proton driving force



*Dewar, Juretic & Zupanovic (2006)*

# Evolution of ATP-synthase kinetics

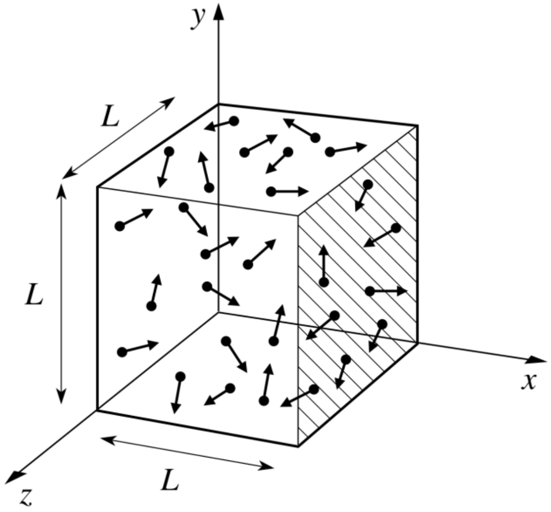


MaxEnt & MEP predict …

- optimal angular position for ATP synthesis close to observed (0.6)
- optimal gearing ratio (H$^+$/ATP) $\propto$ 1 / pmf
- $J_{net}$ maximally sensitive to pmf
- high free-energy conversion efficiency (69%) within the experimental range (50 – 80%)

→ observed kinetic design of ATP synthase consistent with most likely design

# Optimal trait adaptation to changing constraints ($C_1 \rightarrow C_2$): survival of the likeliest



**MaxEnt:**

$C_1$

*most likely*

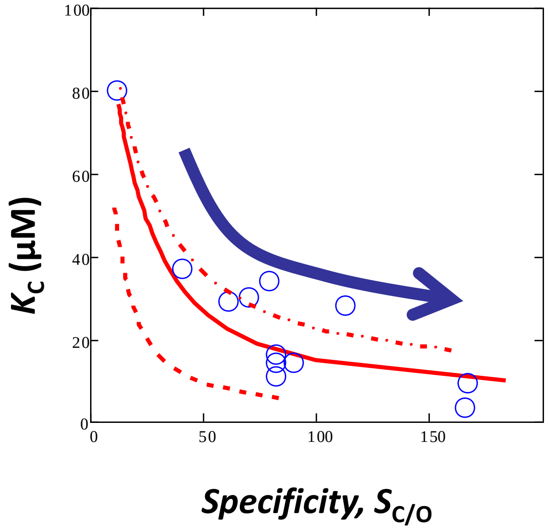
$p(x \mid C_1)$

$C_2$

*most likely*

$p(x \mid C_2)$

**MEP: max stability ?**



(cf. *P-V-T gas laws*)

# Where next?

***Theory***

- Basis of MEP
- Non-stationary behaviour

***Applications***

- Ecological interactions: $r_i \rightarrow r_{ij}$
- Food webs
- Plant adaptations (e.g. leaf stomatal responses)