On the stochastic approach to model the double phosphorylation/dephosphorylation cycle

… a word of caution …

A.M. Bersani

SAPIENZA Università di Roma

Consiglio Nazionale delle Ricerche
IASI Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti"

A. Borri F. Carravetta G. Mavelli P. Palumbo

Phosphorylation

A ubiquitous chemical reaction that activates proteins (e.g. enzymes) by adding one (or more) phosphate groups

The basic phosphorylation scheme is that of a basic enzymatic reaction.

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

The enzymes that catalyze a phosphorylation are called **kinases**

A (shorter) alternative notation is

$$S \overset{\overset{E}{\downarrow}}{\longrightarrow} S^p$$

Double phosphorylation:

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

Phosphorylations can be iterated:

$$S \overset{\overset{E}{\downarrow}}{\longrightarrow} S^p \overset{\overset{E}{\downarrow}}{\longrightarrow} S^{pp} \overset{\overset{E}{\downarrow}}{\longrightarrow} S^{ppp}$$

Dephosphorylation

Dephosphorylation is the chemical reaction that “removes” the phosphate group

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p \overset{F \downarrow}{\longrightarrow} S$$

Enzymes that catalyze dephosphorylation are called **phosphatases**

$$S \underset{\uparrow F}{\overset{E \downarrow}{\rightleftarrows}} S^p \underset{\uparrow F}{\overset{E \downarrow}{\rightleftarrows}} S^{pp}$$

Double phospho/dephospho cycle

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

$$S^p + F \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} C_f^p \overset{v_4}{\rightarrow} S + F$$

Double phospho/dephospho cycle

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p + F \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} C_f^p \overset{v_4}{\rightarrow} S + F$$

$$\frac{d[S]}{dt} = -k_1[S][E] + k_{-1}[C_e] + v_4[C_f^p]$$

$$\frac{d[S^p]}{dt} = v_1[C_e] - k_2[S^p][E] + k_{-2}[C_e^p] + v_3[C_f^{pp}] - k_4[S^p][F] + k_{-4}[C_f^p]$$

$$\frac{d[S^{pp}]}{dt} = v_2[C_e^p] - k_3[S^{pp}][F] + k_{-3}[C_f^{pp}]$$

$$\frac{d[E]}{dt} = -k_1[S][E] + (k_{-1} + v_1)[C_e] - k_2[S^p][E] + (k_{-2} + v_2)[C_e^p]$$

$$\frac{d[F]}{dt} = -k_3[S^{pp}][F] + (k_{-3} + v_3)[C_f^{pp}] - k_4[S^p][F] + (k_{-4} + v_4)[C_f^p]$$

Double phospho/dephospho cycle

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p + F \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} C_f^p \overset{v_4}{\rightarrow} S + F$$

$$\frac{d[S]}{dt} = -k_1[S][E] + k_{-1}[C_e] + v_4[C_f^p]$$

$$\frac{d[C_e]}{dt} = k_1[S][E] - (k_{-1} + v_1)[C_e]$$

$$\frac{d[C_f^{pp}]}{dt} = k_3[S^{pp}][F] - (k_{-3} + v_3)[C_f^{pp}]$$

$$\frac{d[C_e^p]}{dt} = k_2[S^p][E] - (k_{-2} + v_2)[C_e^p]$$

$$\frac{d[C_f^p]}{dt} = k_4[S^p][F] - (k_{-4} + v_4)[C_f^p]$$

$$\frac{\ldots}{dt} = -k_1[S][E] + (k_{-1} + v_1)[C_e] - k_2[S^p][E] + (k_{-2} + v_2)[C_e^p]$$

$$\frac{d[F]}{dt} = -k_3[S^{pp}][F] + (k_{-3} + v_3)[C_f^{pp}] - k_4[S^p][F] + (k_{-4} + v_4)[C_f^p]$$

Double phospho/dephospho cycle

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p + F \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} C_f^p \overset{v_4}{\rightarrow} S + F$$

Mass conservation laws simplify the ODE system

- 3 kinds of substrates
- 2 enzymes
- 4 kinds of complexes

$$[E]_{tot} = [E] + [C_e] + [C_e^p]$$

$$[F]_{tot} = [F] + [C_f^{pp}] + [C_f^p]$$

$$[S]_{tot} = [S] + [S^p] + [S^{pp}] + [C_e] + [C_e^p] + [C_f^{pp}] + [C_f^p]$$

Initial conditions: $[S] = [S]_{tot}$ $\quad [S^p]_0 = [S^{pp}]_0 = 0$

$[C_e]_0 = [C_e^p]_0 = [C_f^{pp}]_0 = [C_f^p]_0 = 0$ $\quad [F]_0 = [F]_{tot}$ $\quad [E]_0 = [E]_{tot}$

Steady-state solutions

From the complexes equations set = 0

$$\frac{d[C_e]}{dt} = k_1[S][E] - (k_{-1} + v_1)[C_e]$$
$$\frac{d[C_e^p]}{dt} = k_2[S^p][E] - (k_{-2} + v_2)[C_e^p]$$
$$\frac{d[C_f^{pp}]}{dt} = k_3[S^{pp}][F] - (k_{-3} + v_3)[C_f^{pp}]$$
$$\frac{d[C_f^p]}{dt} = k_4[S^p][F] - (k_{-4} + v_4)[C_f^p]$$

$$K_{M_i} = \frac{k_{-i} + v_i}{k_i} \qquad i = 1, 2, 3, 4$$

$$[C_e] = \frac{[S][E]}{K_{M_1}}$$
$$[C_e^p] = \frac{[S^p][E]}{K_{M_2}}$$
$$[C_f^{pp}] = \frac{[S^{pp}][F]}{K_{M_3}}$$
$$[C_f^p] = \frac{[S^p][F]}{K_{M_4}}$$

Steady-state solutions

From the enzymes conservation laws

$$[E]_{tot} = [E] + [C_e] + [C_e^p]$$

$$[F]_{tot} = [F] + [C_f^{pp}] + [C_f^p]$$

$$[E] = \frac{[E]_{tot}}{1 + \frac{[S]}{K_{M_1}} + \frac{[S^p]}{K_{M_2}}}$$

$$[F] = \frac{[F]_{tot}}{1 + \frac{[S^{pp}]}{K_{M_3}} + \frac{[S^p]}{K_{M_4}}}$$

$$[C_e] = \frac{[S][E]}{K_{M_1}}$$

$$[C_e^p] = \frac{[S^p][E]}{K_{M_2}}$$

$$[C_f^{pp}] = \frac{[S^{pp}][F]}{K_{M_3}}$$

$$[C_f^p] = \frac{[S^p][F]}{K_{M_4}}$$

Steady-state solutions

In summary, 4 complexes and 2 enzymes steady-state solutions can be written in terms of the 3 substrates steady-state solutions

- We still have 3 unknowns (the 3 substrates steady-states) and 3 constraints provided by the 3 substrates algebraic equations

$$\frac{d[S]}{dt} = 0$$
$$\frac{d[S^p]}{dt} = 0$$
$$\frac{d[S^{pp}]}{dt} = 0$$

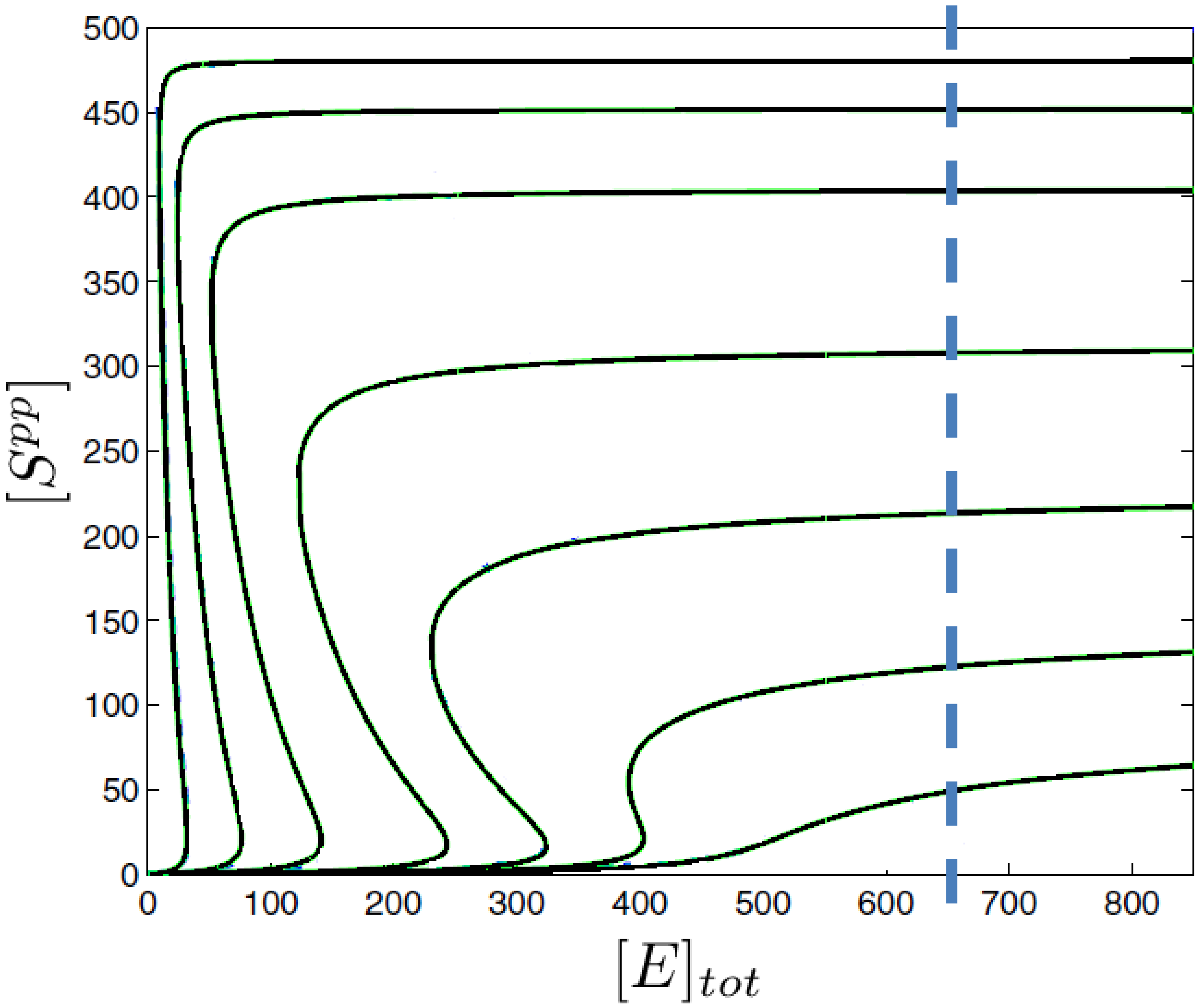
One can be replaced by the substrate conservation law

- We can have 1 or 3 steady-states solutions

Steady-state solutions

Different curves for different values of $[F]_{tot}$

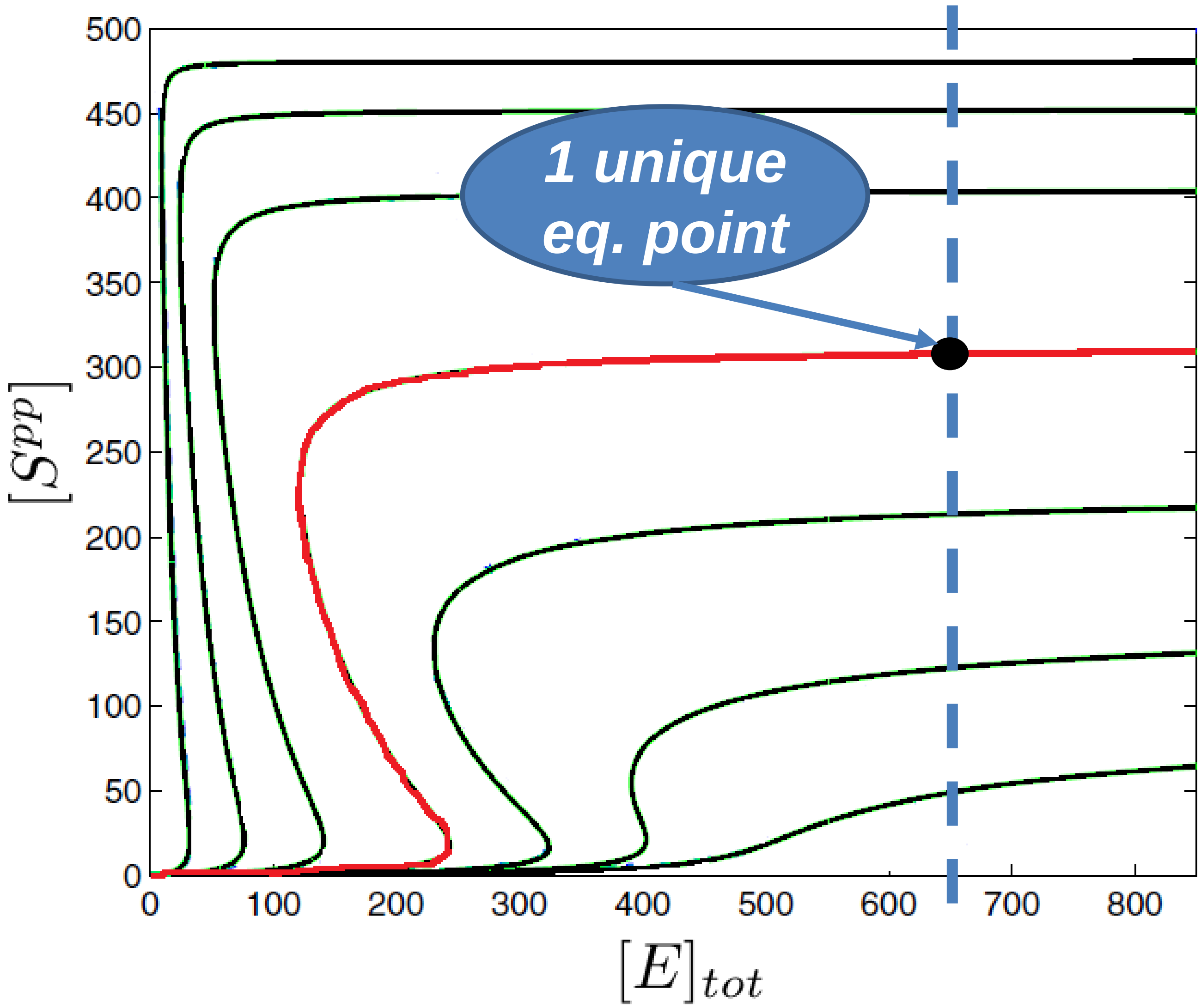
- ✓ for a chosen value of $[S]_{tot}$
- ✓ for a chosen value of $[E]_{tot}$



Steady-state solutions

Different curves for different values of $[F]_{tot}$

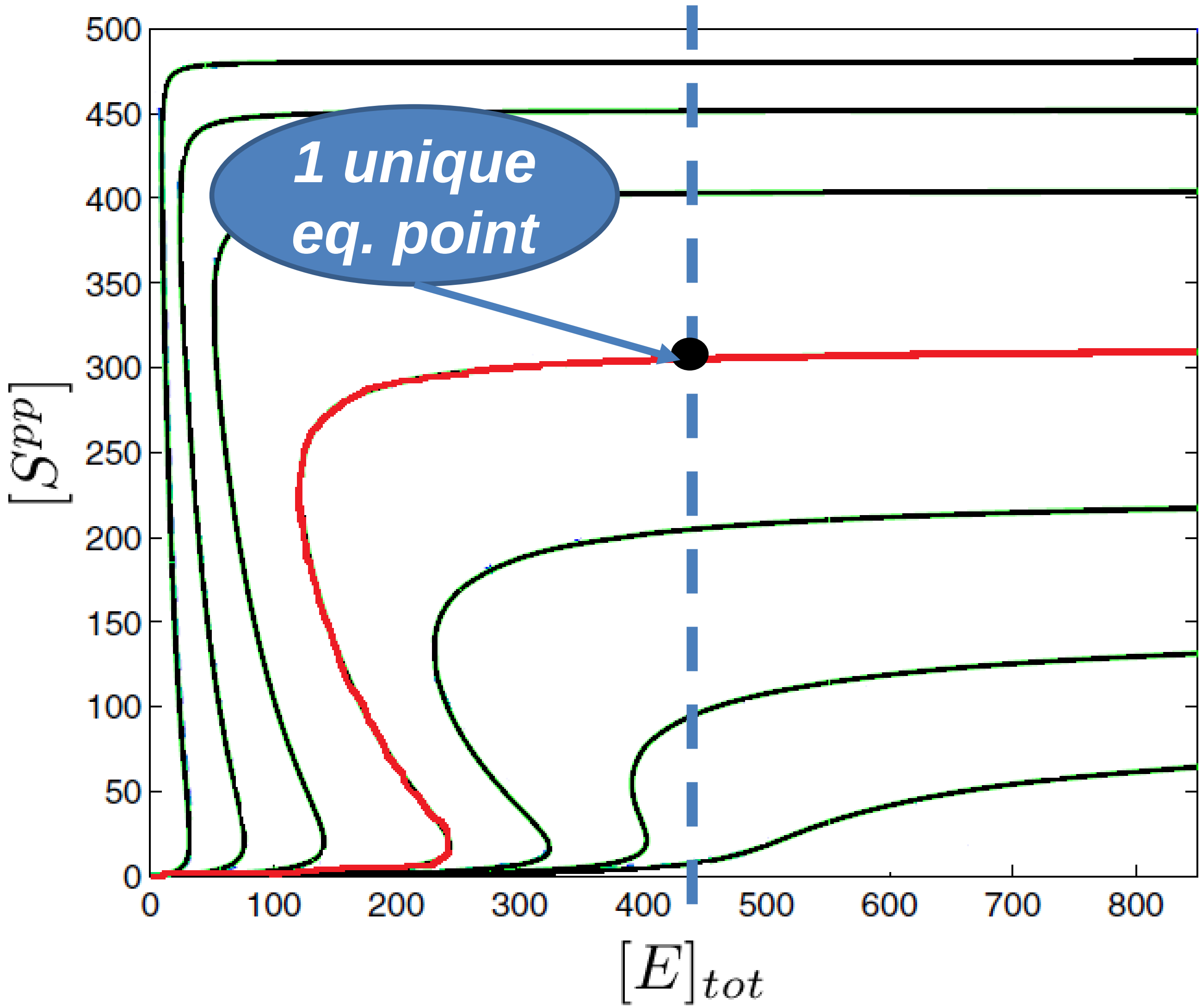
- ✓ for a chosen value of $[S]_{tot}$
- ✓ for a chosen value of $[E]_{tot}$
- ✓ for a chosen value of $[F]_{tot}$



Steady-state solutions

Different curves for different values of **[*F*]tot**

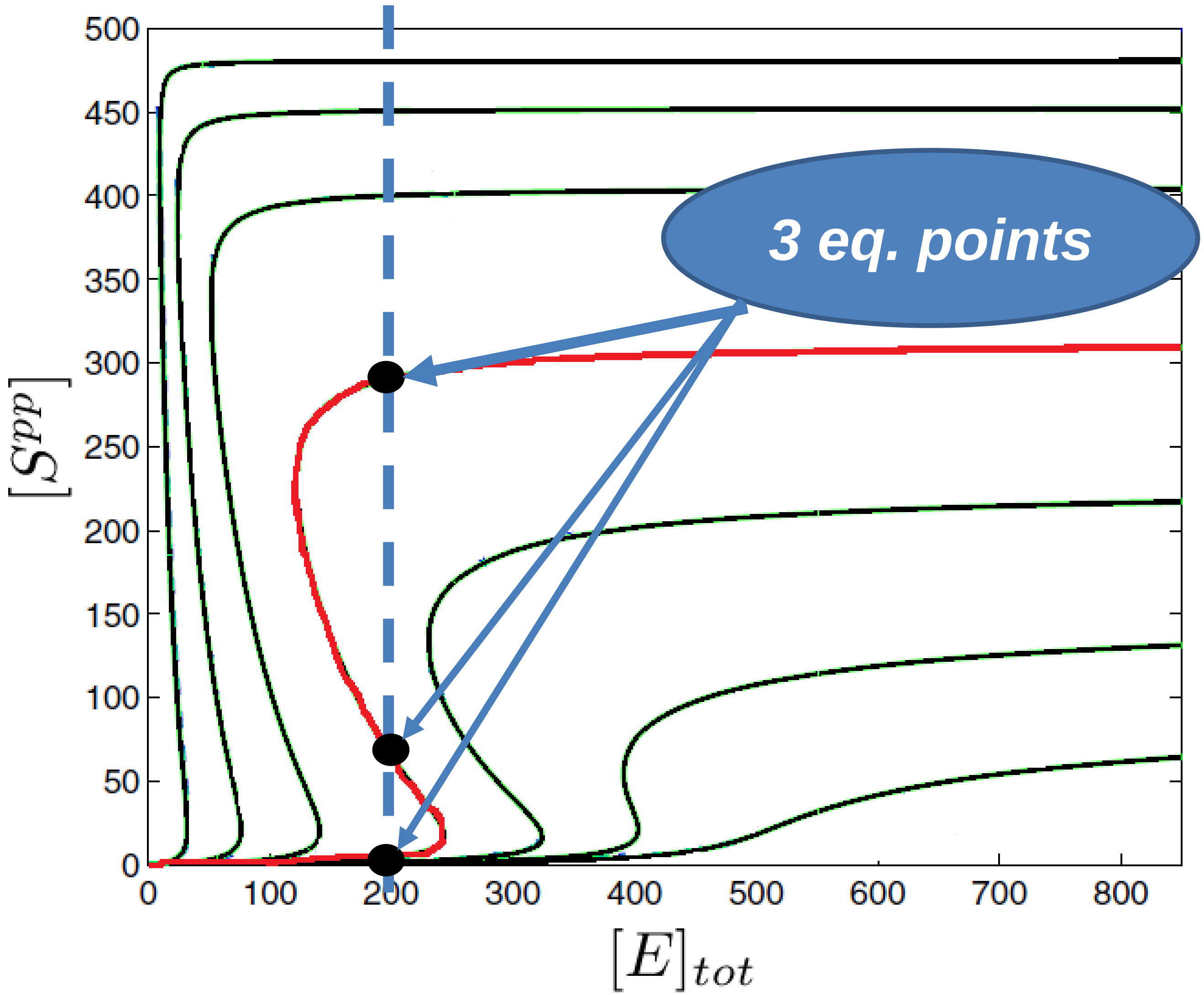
- ✓ for a chosen value of **[*S*]tot**
- ✓ for a chosen value of **[*E*]tot**
- ✓ for a chosen value of **[*F*]tot**



Steady-state solutions

Different curves for different values of $[F]_{tot}$

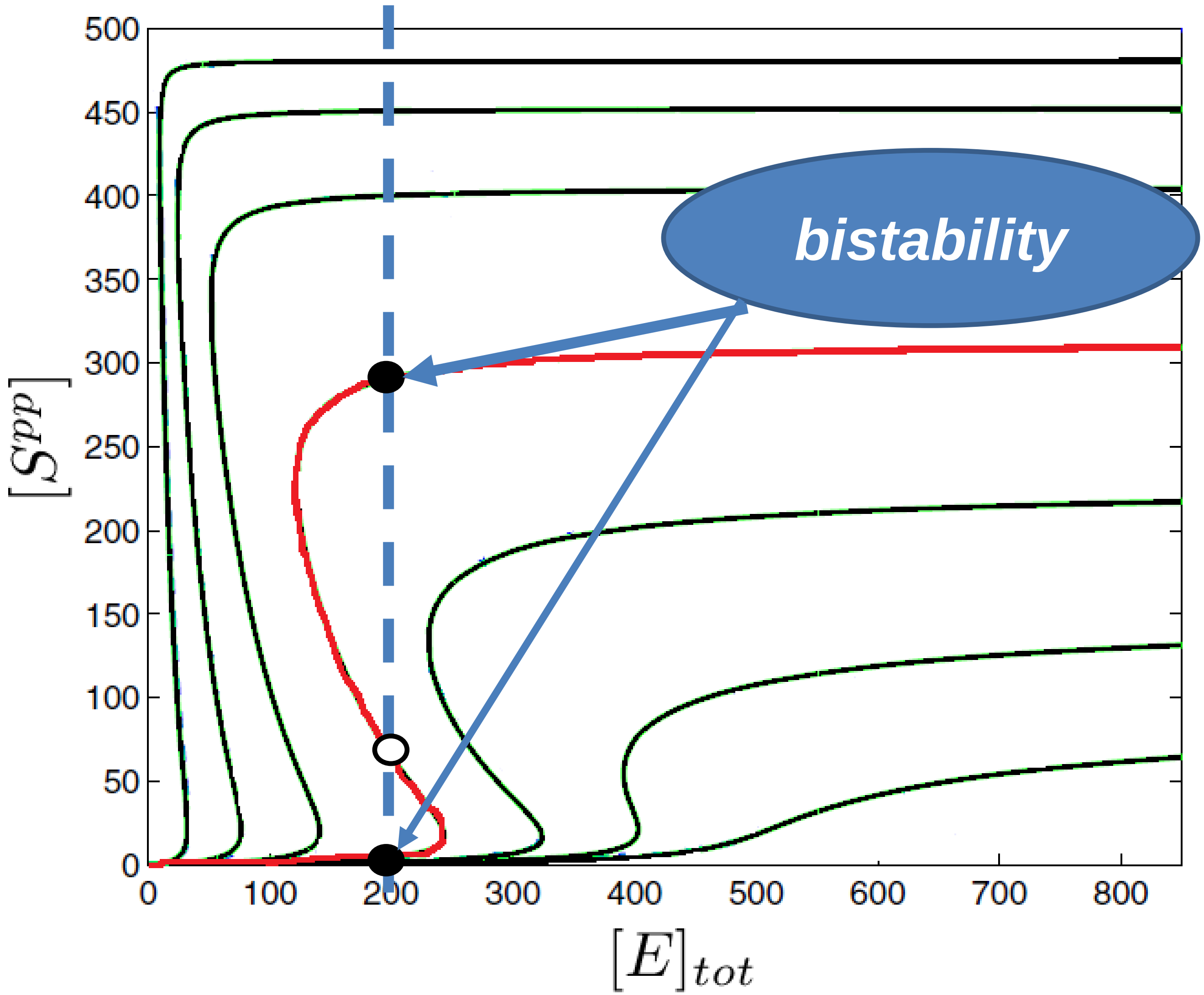
- ✓ for a chosen value of $[S]_{tot}$
- ✓ for a chosen value of $[E]_{tot}$
- ✓ for a chosen value of $[F]_{tot}$



Steady-state solutions

Different curves for different values of **[*F*]tot**

- ✓ for a chosen value of **[*S*]tot**
- ✓ for a chosen value of **[*E*]tot**
- ✓ for a chosen value of **[*F*]tot**



Stochastic approach

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p + F \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} C_f^p \overset{v_4}{\rightarrow} S + F$$

The state of the system is given by the copies of each species:
a **discrete-event system**

$$(n_S, n_{Sp}, n_{Spp}, n_E, n_F, n_{Ce}, n_{Cep}, n_{Cfpp}, n_{Cfp})$$

$$(n_S, n_{Spp}, n_{Ce}, n_{Cep}, n_{Cfpp}, n_{Cfp})$$

Stochastic approach

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p + F \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} C_f^p \overset{v_4}{\rightarrow} S + F$$

Each reaction provides a reset in $(n_S, n_{Spp}, n_{Ce}, n_{Cep}, n_{Cfpp}, n_{Cfp})$

$$S + E \overset{k_1}{\rightarrow} C_e$$

$(n_S, n_{Spp}, n_{Ce}, n_{Cep}, n_{Cfpp}, n_{Cfp})$

$(n_S - 1, n_{Spp}, n_{Ce} + 1, n_{Cep}, n_{Cfpp}, n_{Cfp})$

Stochastic approach

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C_e \overset{v_1}{\rightarrow} S^p + E$$

$$S^p + E \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C_e^p \overset{v_2}{\rightarrow} S^{pp} + E$$

$$S^{pp} + F \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} C_f^{pp} \overset{v_3}{\rightarrow} S^p + F$$

$$S^p + F \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} C_f^p \overset{v_4}{\rightarrow} S + F$$

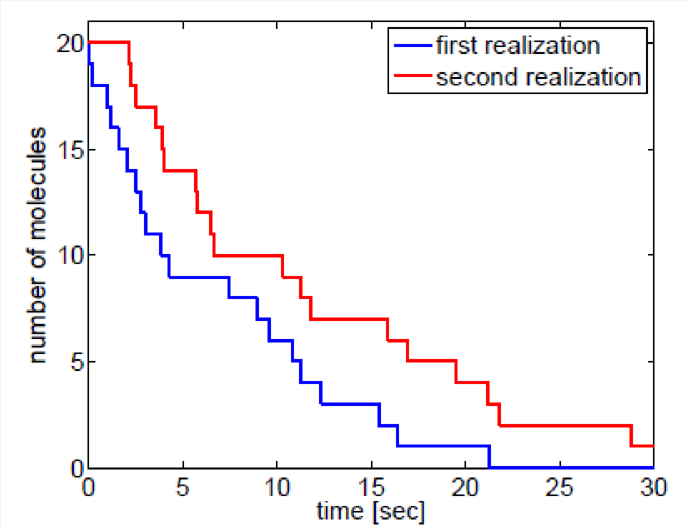
- **Continuous-Time Markov Chains** describe the probabilities to “find” the state on a specific point on a 6-dimensional state space

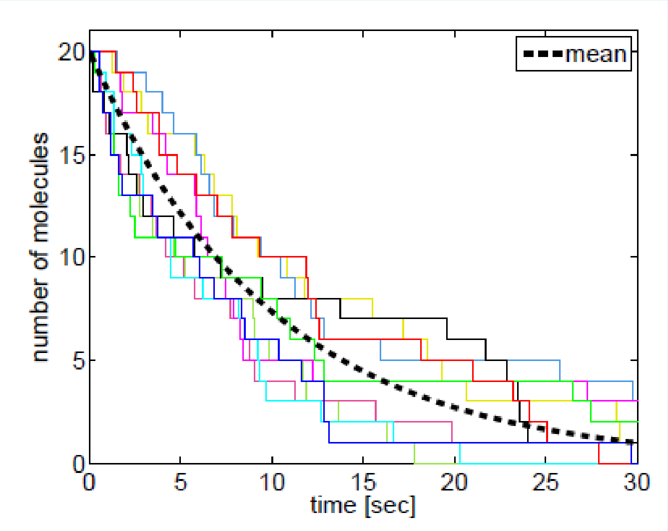
- The dynamics of these probabilities are described by the **Chemical Master Equations (CME)**, with reaction rates providing the “propensities” to jump from one point to another according to a given reaction

- Analytical solutions of CMEs are usually unaffordable (**curse of dimensionality**)

Stochastic approach: Gillespie algorithm

- The **Stochastic Simulation Algorithm** (**SSA**, Gillespie, 1976)
  - ✓ runs a "large" number of random paths
  - ✓ estimates the statistical probability from samples
  - ✓ provides an approximate numerical solution to the CME equation





- The SSA is "exact" in the sense that, for large number of trials, the statistical distribution converges to the exact one
  - ✓ the SSA is usually considered a "golden standard"

Stochastic approach: ergodicity

In case of ergodic CTMC (the case of a unique stationary distribution is ergodic) statistical properties can be deduced from a single, sufficiently long realization (random path) of the process

↓

A unique, large enough random path (instead of many)

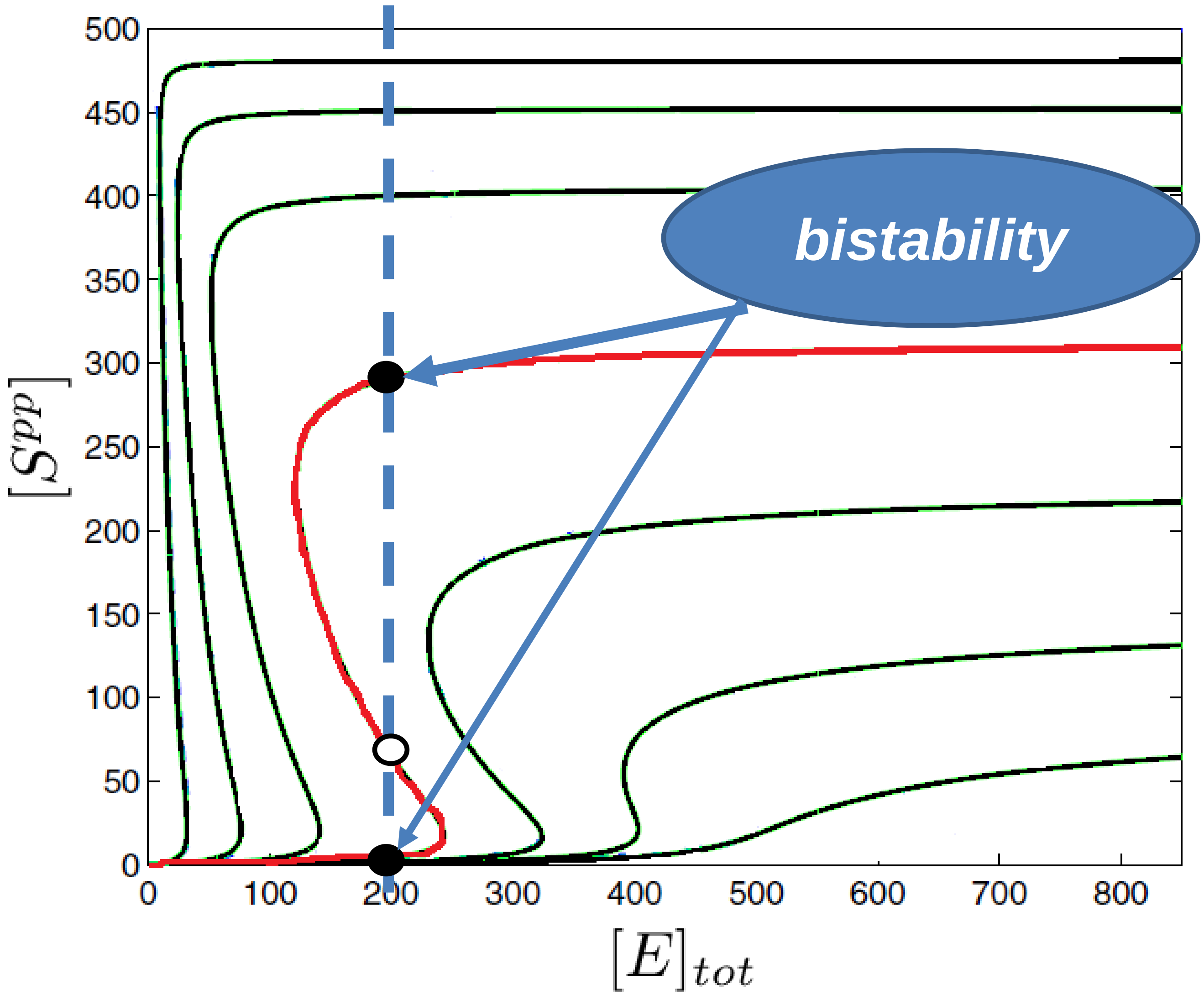
Looking for the stationary distribution, ergodic CTMCs provide it in terms of the temporal distribution of the dwell times over the states of the process

Stochastic approach and double phosphor/dephospho cycle

?

?

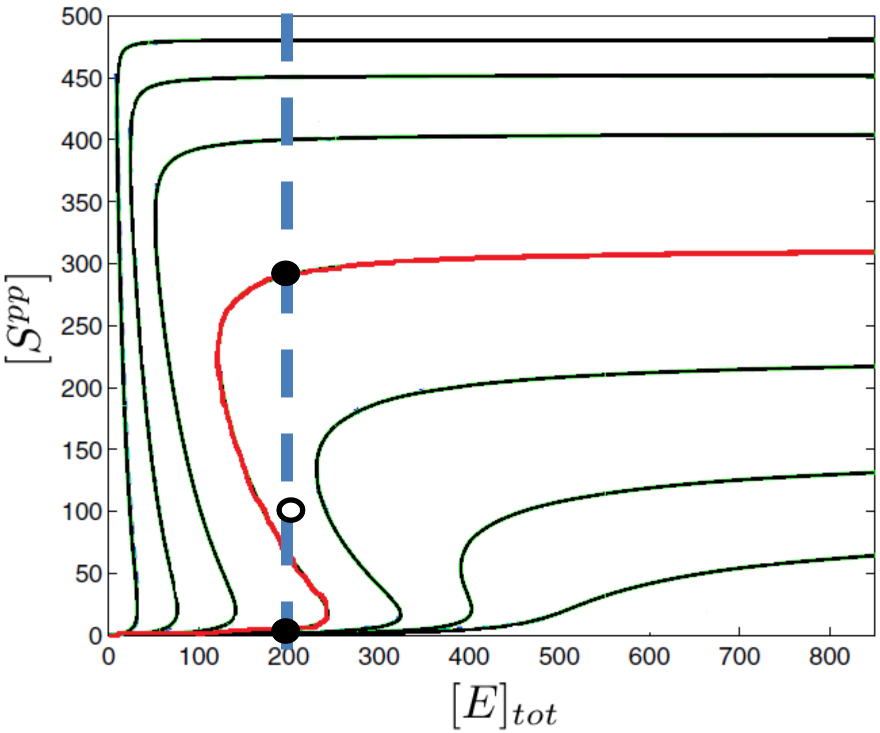
?

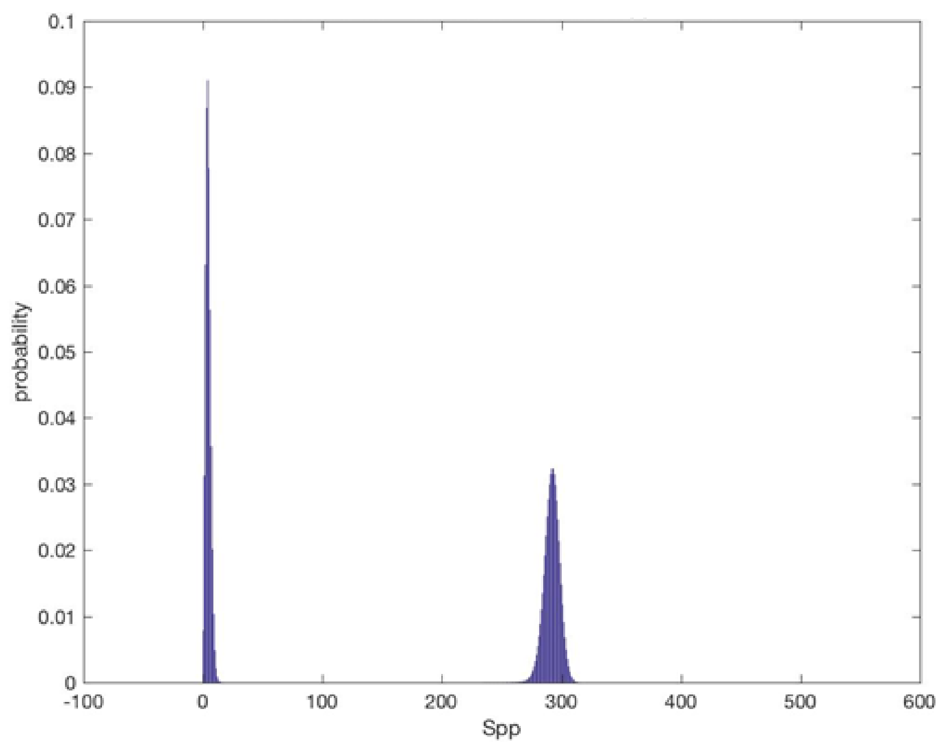


Stochastic approach and double phosphor/dephospho cycle

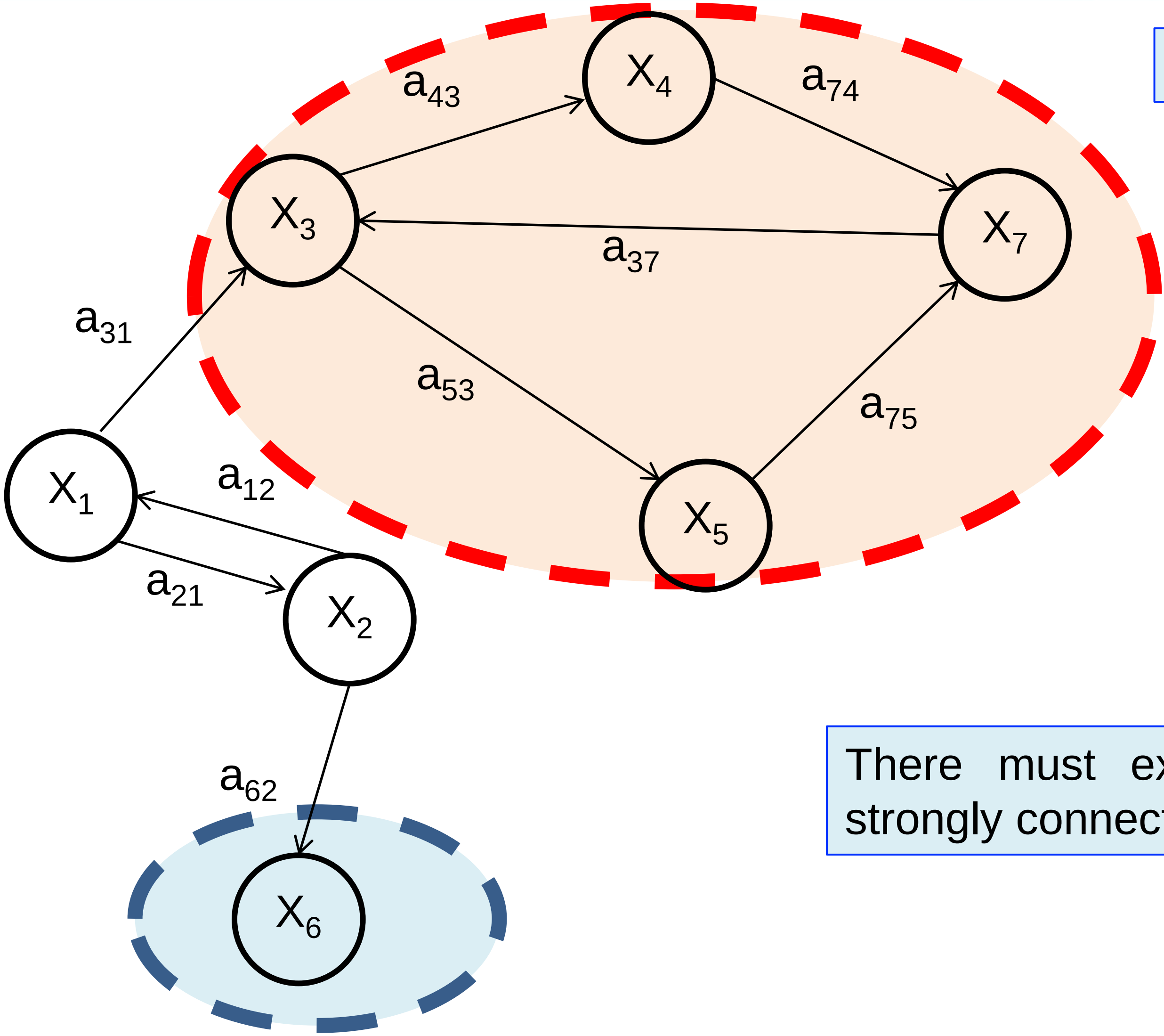
Conjecture (1): ergodicity (that implies a unique stationary probability distribution)

Conjecture (2): deterministic bistability “translates” into stochastic bimodality





Stochastic approach and double phosphor/dephospho cycle



Conjecture (1): ergodicity

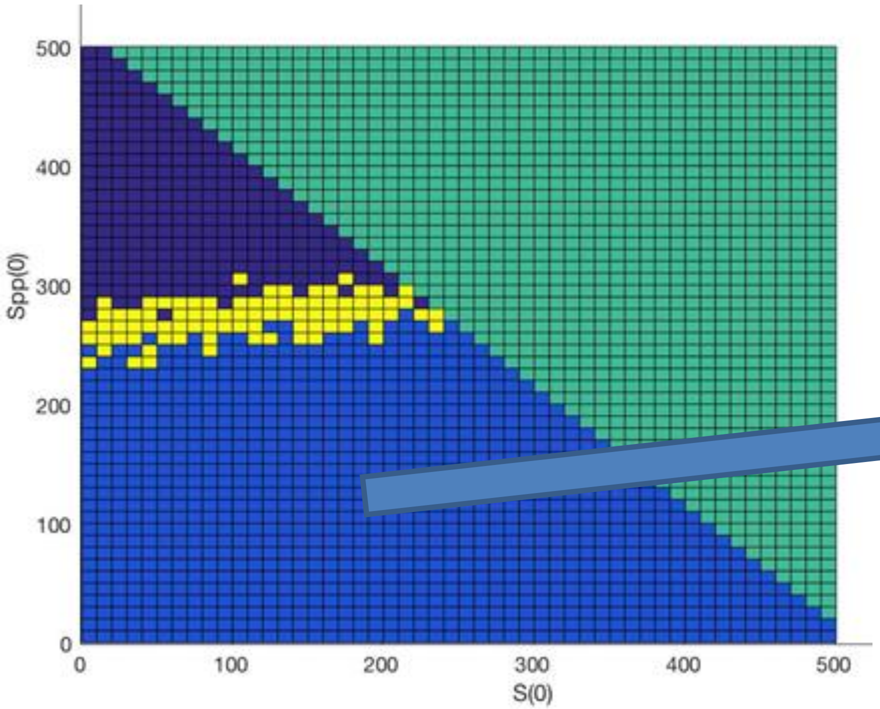
There must exist a unique terminal strongly connected component

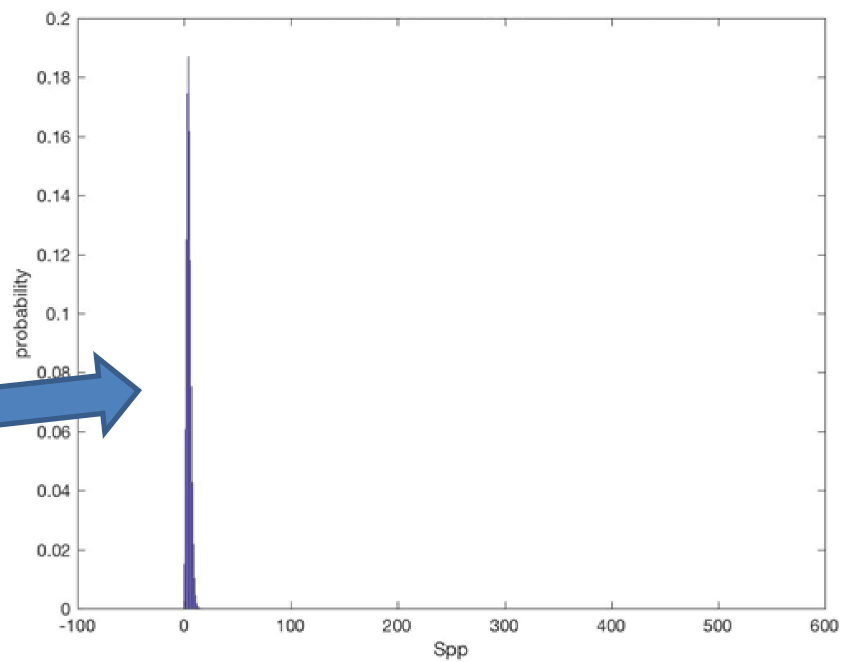
Stochastic approach and double phosphor/dephospho cycle

Assume ergodicity holds. Then:

✓ just 1 (long enough) random path is necessary

✓ stationary distribution is unique whatever is the initial condition



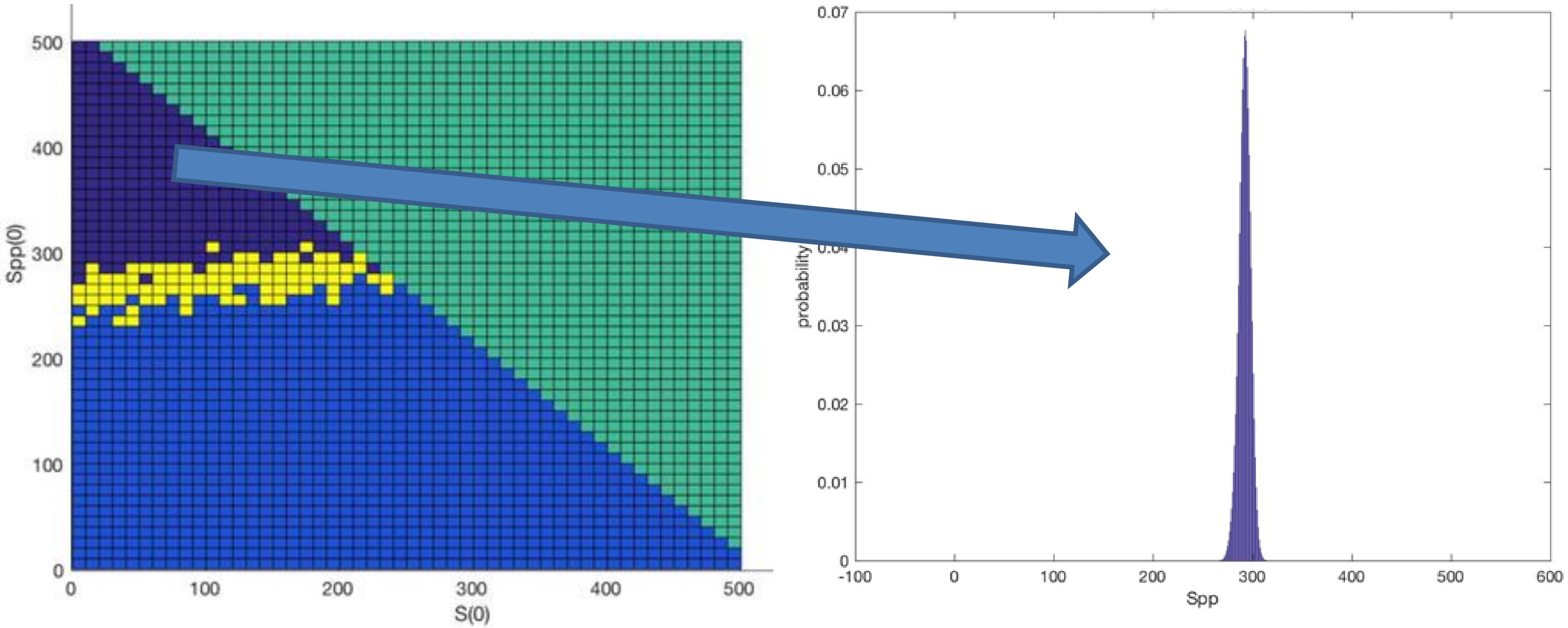


Stochastic approach and double phosphor/dephospho cycle

Assume ergodicity holds. Then:

✓ just 1 (long enough) random path is necessary

✓ stationary distribution is unique whatever is the initial condition



Stochastic approach and double phosphor/dephospho cycle

Problem: the probability of leaving one basin of attraction is so low that it would require a (quite) infinite time to occur. That is why we always see only 1 mode of the bi-modal distribution

Solutions:

- ✓ Exploit the analytical CME solution?
- ✓ Wait for 1 single (long enough) Gillespie simulation?
- ✓ Run a (large enough) number of Gillespie simulations?
- ✓ Run Gillespie simulations starting from the "yellow" region?
- ✓ Run tau-leaping Gillespie simulation?