Non-Stationary Non-Equilibrium Rare Events

Kai Kratzer (*ICP Stuttgart*),
Tanja Schilling (*Uni Luxembourg*),
Axel Arnold (*ICP Stuttgart*) &
Josh Berryman

uni.lu UNIVERSITÉ DU LUXEMBOURG
PHYMS | PHYSICS AND MATERIAL SCIENCES RESEARCH UNIT

Some Definitions:

1) "Full" Equilibrium: detailed balance applies:

$$P(x)P(x \rightarrow y) \quad = \quad P(y)P(y \rightarrow x) \qquad \text{Detailed Balance}$$

2) Stationary (non-)equilibrium: balance only:

$$\Sigma_x P(x)P(x \rightarrow y) \quad = \quad P(y)\Sigma_x P(y \rightarrow x) \qquad \text{Balance}$$

3) Non-stationary non-equilibrium:

$$P(x,t) \quad \neq \quad P(x,t') \qquad \text{No Balance}$$

Directed Rare-Event Sampling:

How do I run a simulation such that I observe some rare event, but also such that I still know what the real, unbiased, distribution is?

Over-sample paths that seem to be going the right way, but record how much you have over-sampled by (FFS, TIS, SPRES, many others…). *"Splitting" methods?*

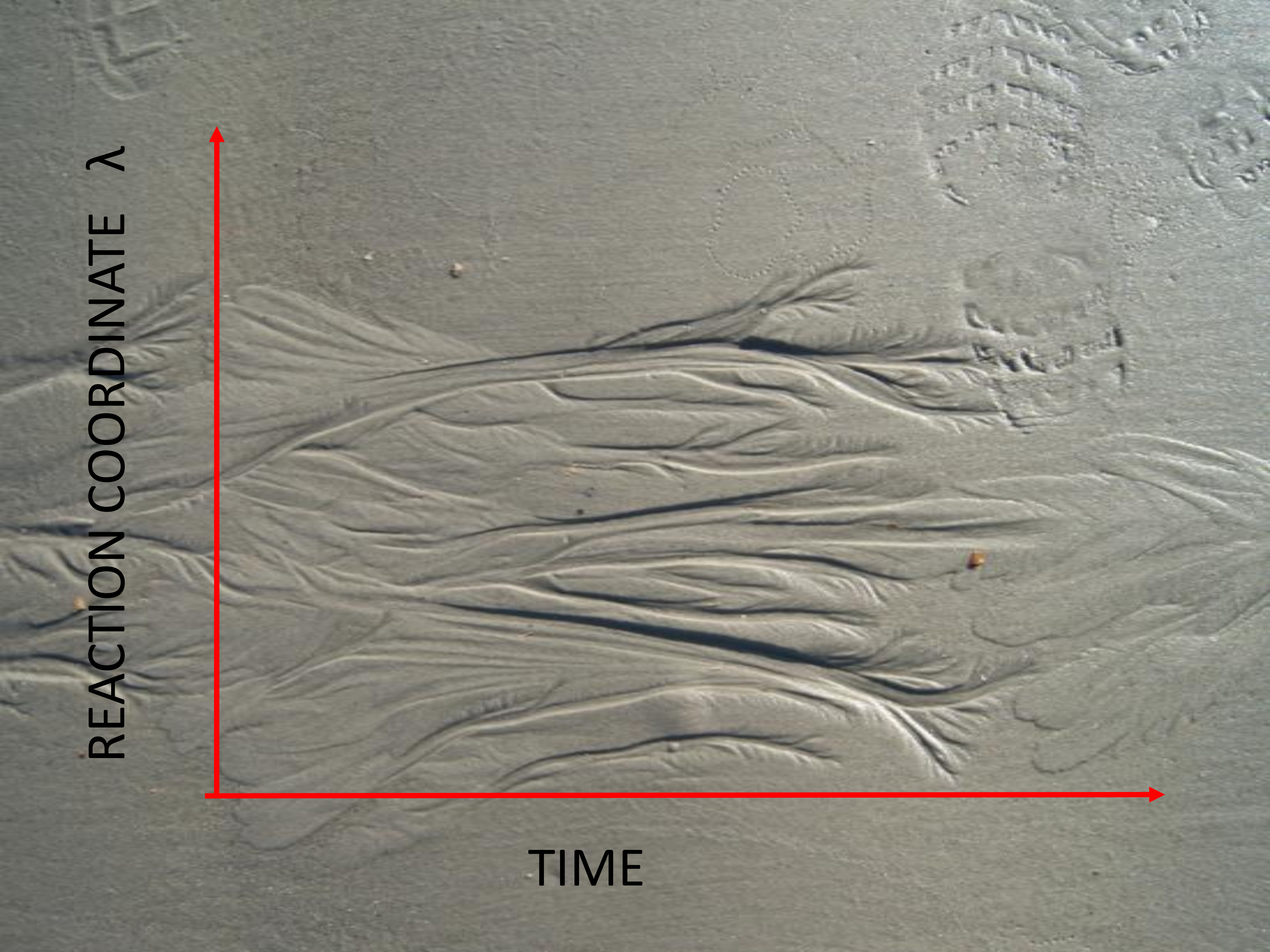
- No requirement for Boltzmann distribution.
- FFS, TIS et al. *still* require/assume Poisson statistics (=>stationary dynamics), therefore we developed SPRES.
- We may be interested in state distributions or in flux across some hypersurface…. slightly different requirements.

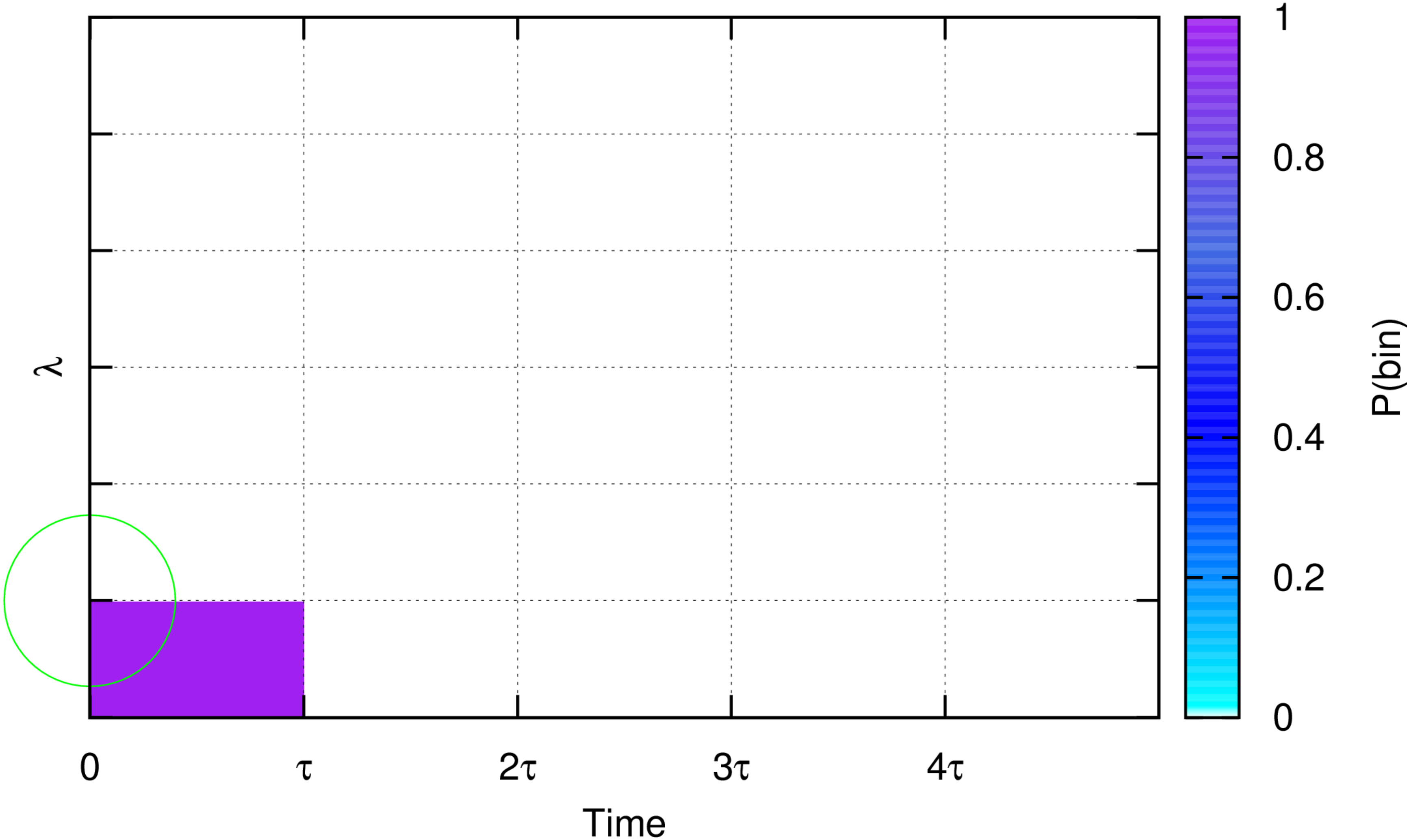
‘Directed’ rare event simulation methods require:

- Some stochastic microscopic dynamics

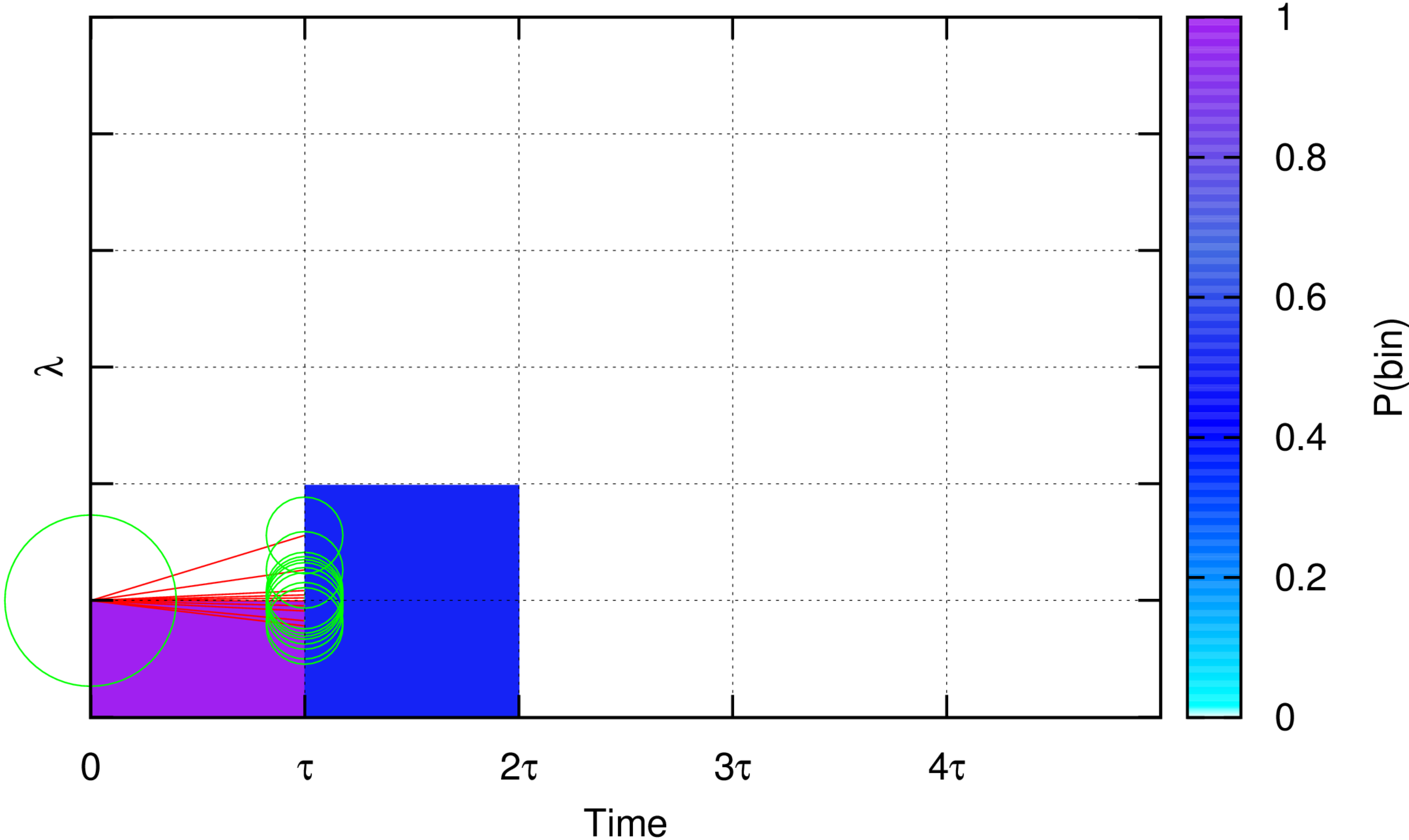
- A macroscopic coordinate λ (e.g. size of a cluster during nucleation, perhaps a fraction of native contacts during protein folding). The coordinate λ defines a subspace of the overall very high-dimensional space of the dynamics.

Before I introduce SPRES (2010).... I should note that it doesn’t seem much different to: RESTART (1991?)

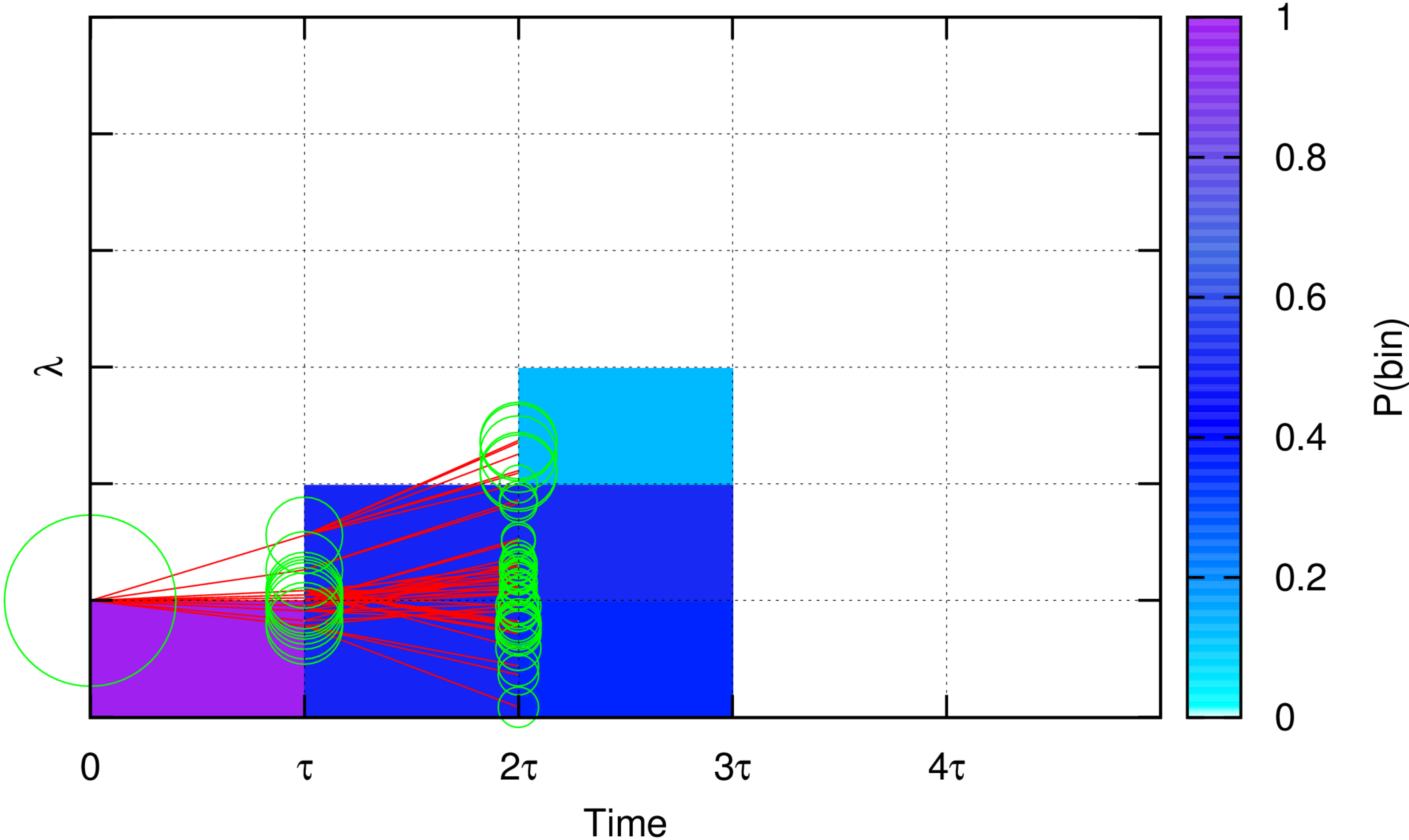

REACTION COORDINATE λ
TIME



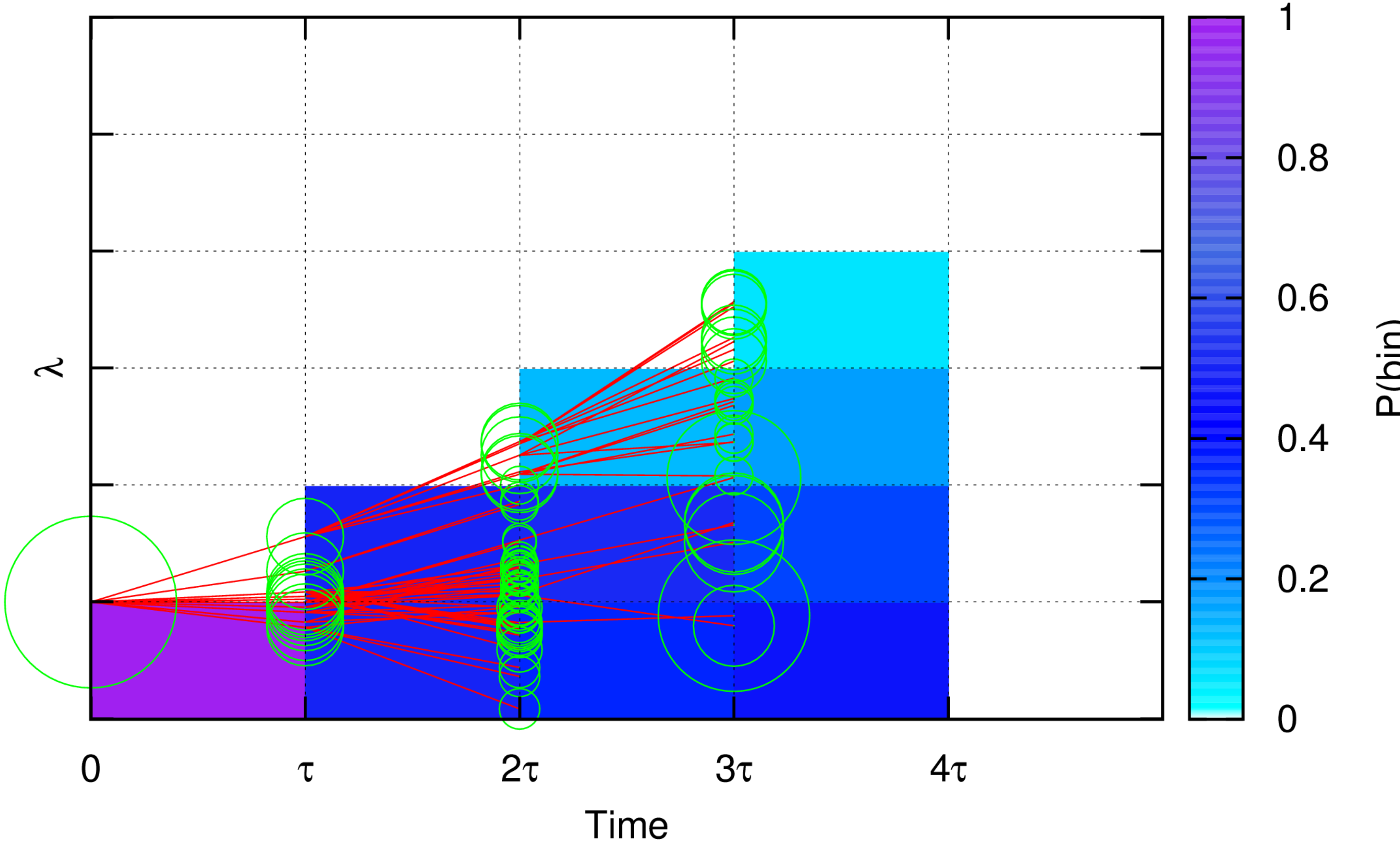
λ
0
τ
2τ
3τ
4τ
Time
1
0.8
0.6
0.4
0.2
0
P(bin)



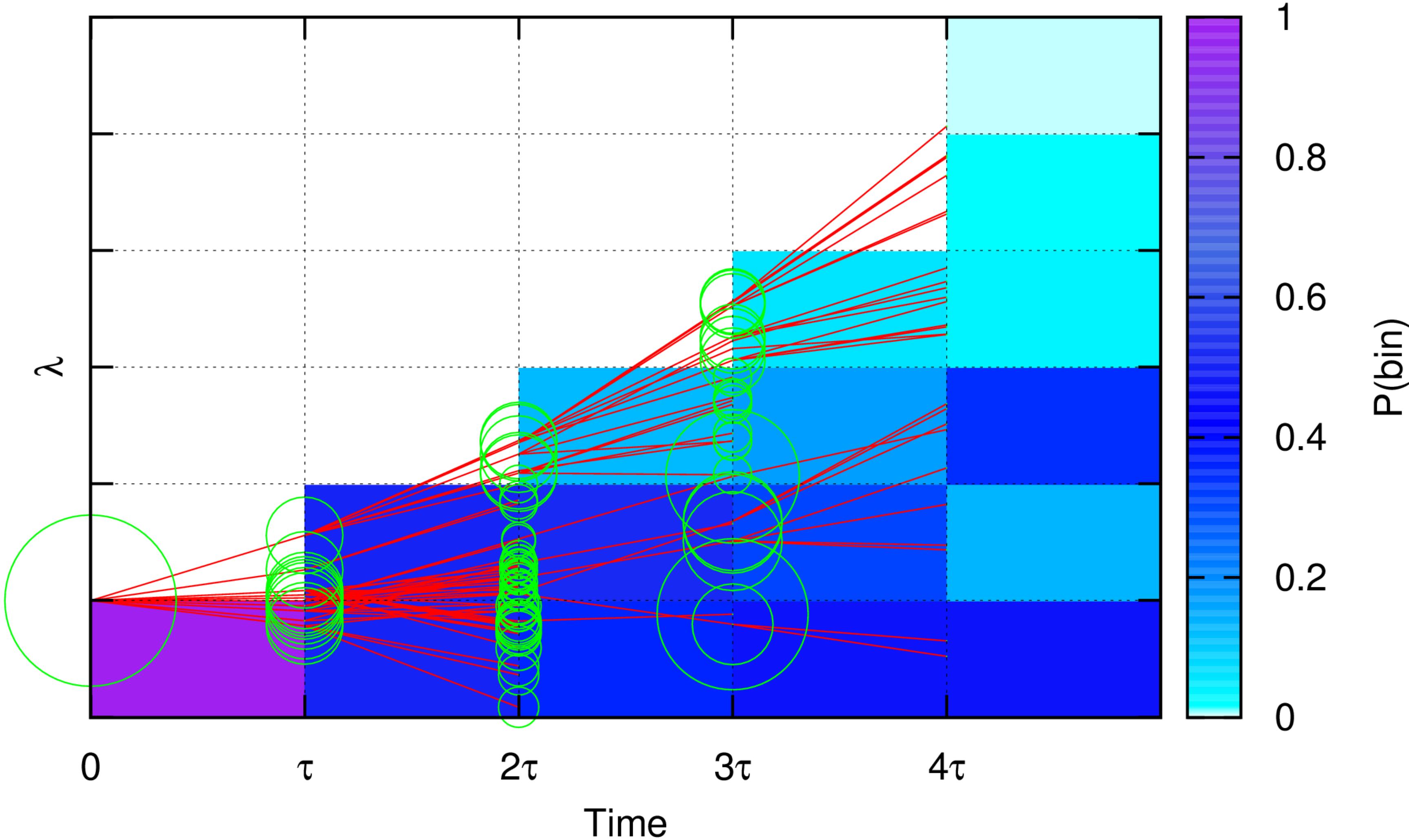
λ
0
τ
2τ
3τ
4τ
Time
1
0.8
0.6
0.4
0.2
0
P(bin)



λ
0
τ
2τ
3τ
4τ
Time
1
0.8
0.6
0.4
0.2
0
P(bin)

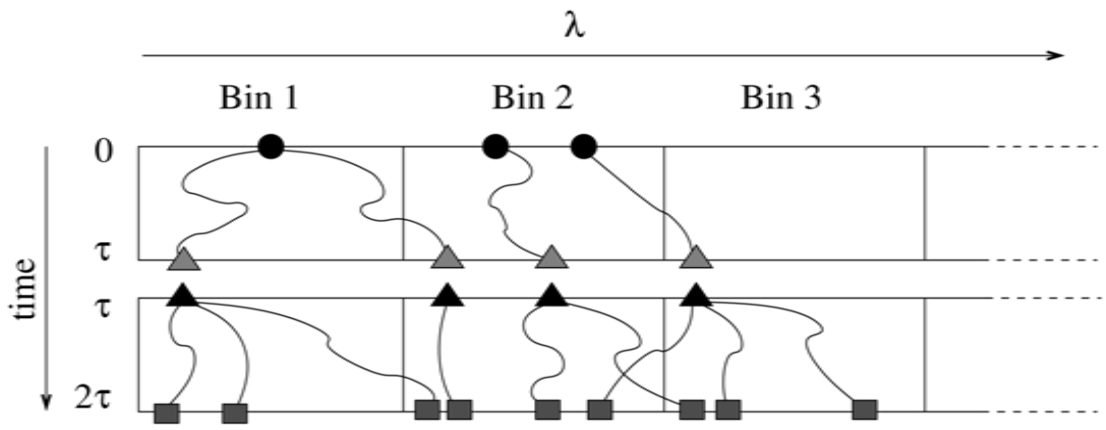


λ
0
τ
2τ
3τ
4τ
Time
1
0.8
0.6
0.4
0.2
0
P(bin)



λ
0
τ
2τ
3τ
4τ
Time
1
0.8
0.6
0.4
0.2
0
P(bin)

SPRES

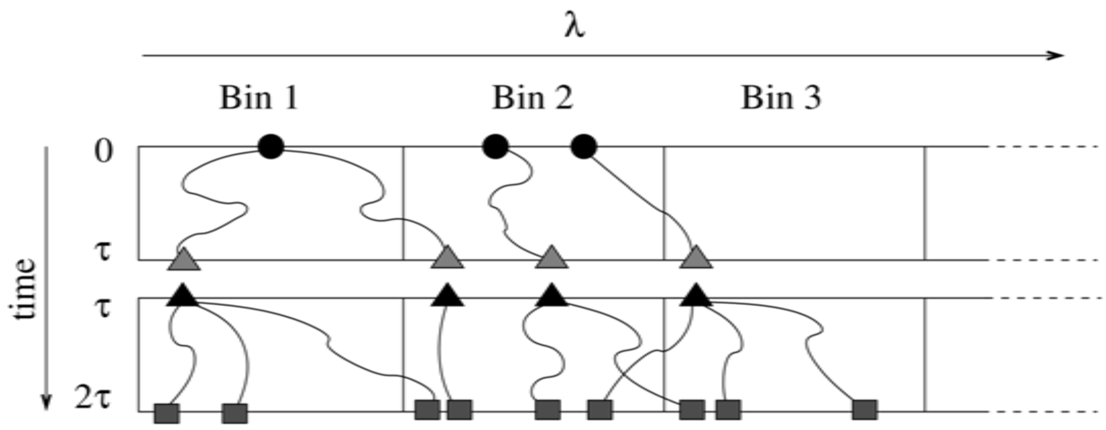


- Use an importance sampling to select the number of shots to run from each bin. R=`number forward'

$$n_i^{t+\tau} = n_i^t + \gamma \left(1 - \frac{R}{N}\right) n_i^t$$

- The softening, gamma, keeps it stable.

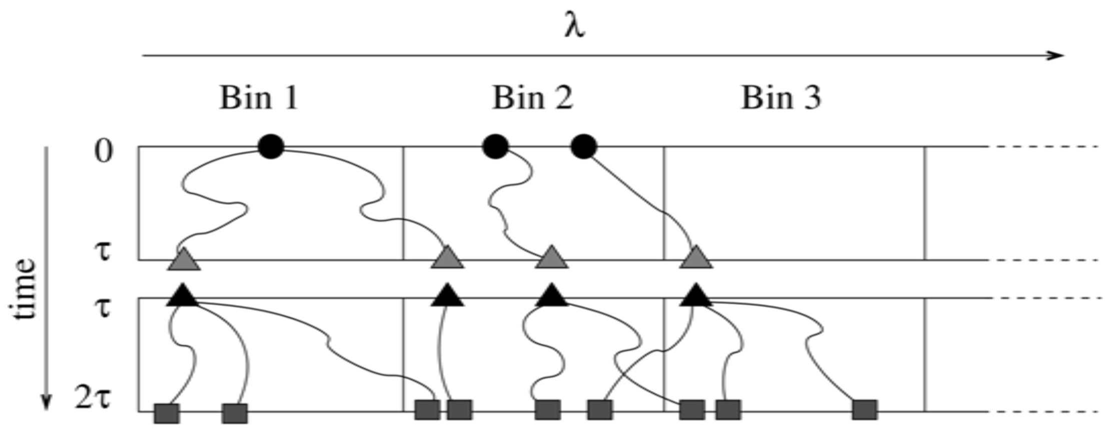
SPRES



A couple of trivial variants of the algorithm:

- `Forward' can mean `out'.
- Forward can mean `decreased Hamming distance' (2+D case).

SPRES



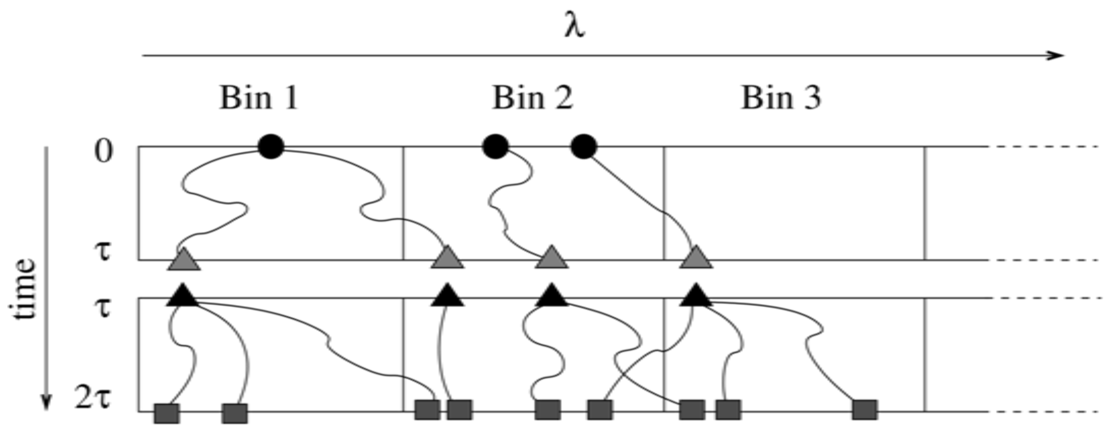
- We know how many shots from each bin... Which configs do we start the shots from?

$$w_s^t = \frac{M_{i,j}^t d_i^t}{N_{i,j}}$$

Weights of individual configs

- Pick the start config with probability proportional to its weight.... Recursion makes everything better.

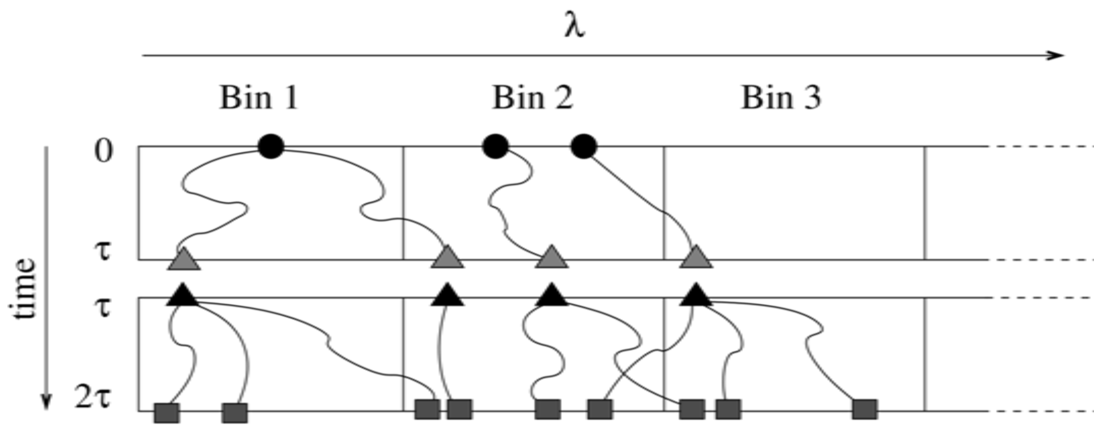
SPRES



- Selection of starting configs by relative weight is rather akin to pruned-enriched Rosenbluth sampling.

- After each round of selection, the weights of paths leaving a given bin are all the same, because high-weight configs had a bigger chance to be chosen, so they were `over-sampled', so their weights have decreased. And vice-versa.

SPRES



$$d_j^t = \sum_{\{s^t\}} w_s^t \delta_j^t,$$

$$M_{i,j}^t = \sum_{\{s^t\}} w_s^t \delta_i^{t-\tau} \delta_j^t / d_i^{t-\tau}$$

Estimated bin weights and conditional transition probabilities for pairs of bins i,j (observe these directly from the simulation)

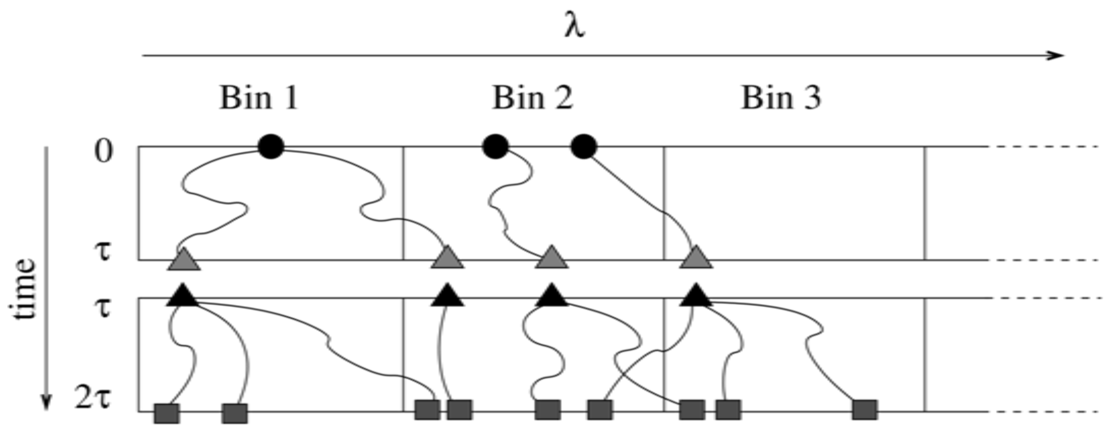
$$w_s^t = \frac{M_{i,j}^t d_i^t}{N_{i,j}}$$

Weights of individual configs

$$\overline{x}^t = \sum_{\{s^t\}} w_s^t x(s^t).$$

Ensemble averages of some observable x

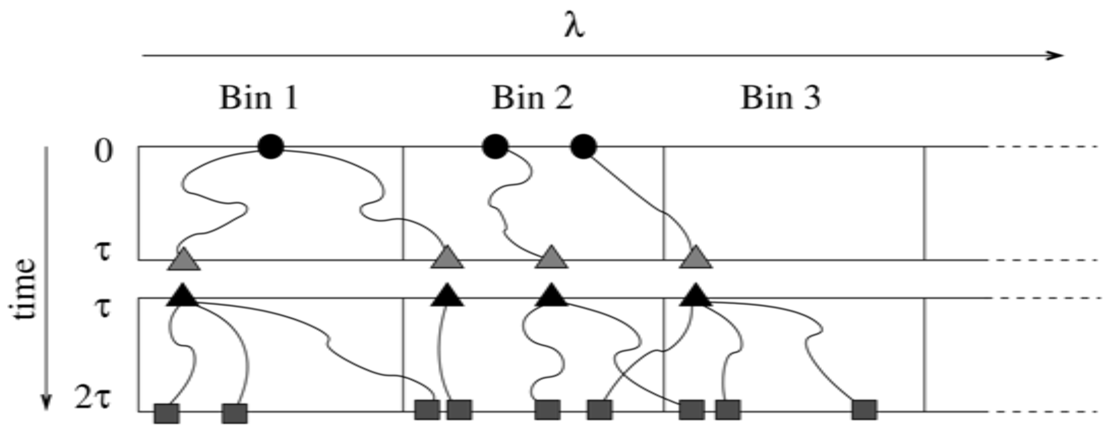
SPRES



- Require *some* continuity on the reaction coordinate w.r.t. configurational space.
- Require a "warm-up" time to populate the bins before rare events are observed.

- Do not require 1D reaction coordinate.
- Prefer, but do not require, the dominant 'least action' reaction paths to be monotonic on the reaction coordinate.

SPRES



- Parameters are τ and the positions of the bin boundaries.

- more bins -> shorter τ .

S-PRES gives:

- Time-dependent dynamics
- An even spread of states/paths over λ is sampled
- A **strictly** unbiased estimate of the probability of each observed state or path is provided, conditioned only on the initial conditions and on the time.

Berryman & Schilling, *J. Chem. Phys.* 2010

Example : Rare fluctuations in a critical system.

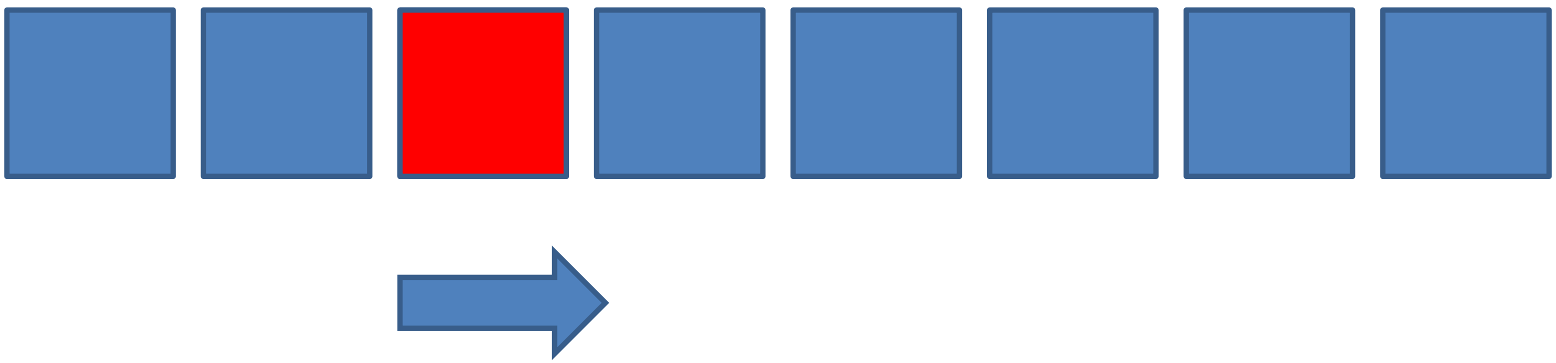
- Asymmetric Exclusion Process – basically a "traffic" model.
- Start with an empty road.
- What is the probability of a total traffic jam versus time?

Theory: Schűtz, *Phys. Rev. E* 1993

Example : Rare fluctuations in a critical system.

On rate = α

Off rate = α



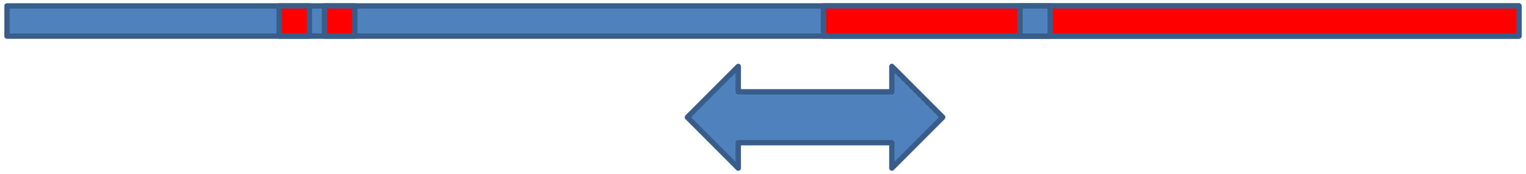
Hop rate = β

Theory: Schűtz, *Phys. Rev. E* 1993

Example: Rare fluctuations in a critical system.

Density = On rate =α

Density = 1 – off rate = 1-α



Interface diffuses randomly

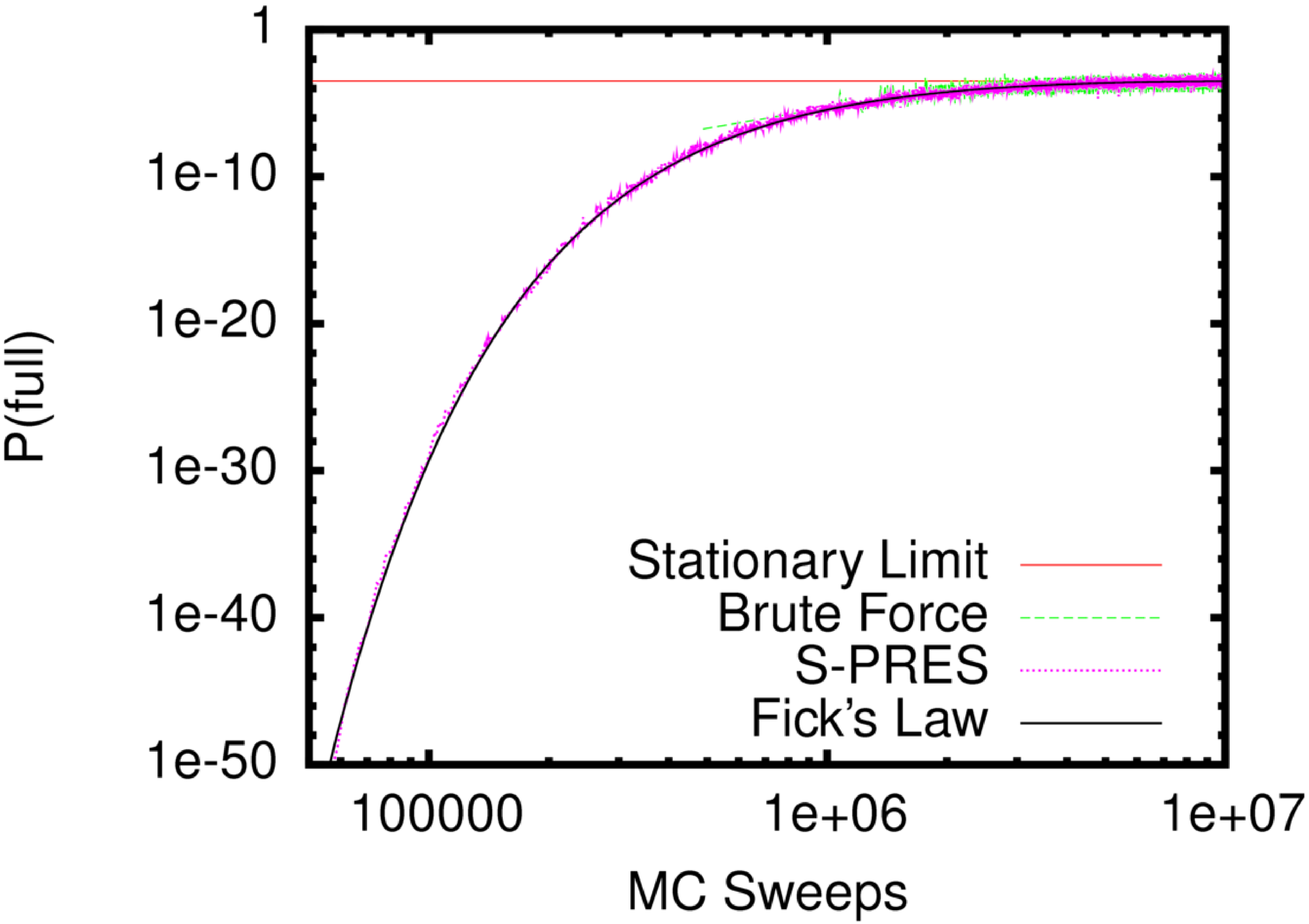
Theory: Schűtz, *Phys. Rev. E* 1993

Example: Rare fluctuations in a critical system.

- Start in the morning with an empty road.

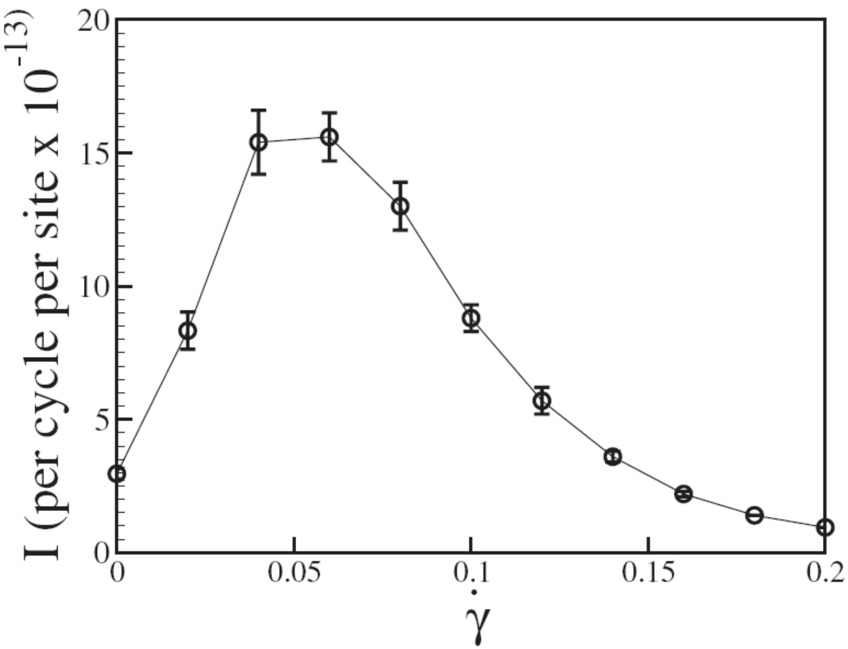
- What is the probability of a total traffic jam versus time (with fixed α)?

Theory: Schűtz, *Phys. Rev. E* 1993



Berryman & Schilling, *J. Chem. Phys.* 2010

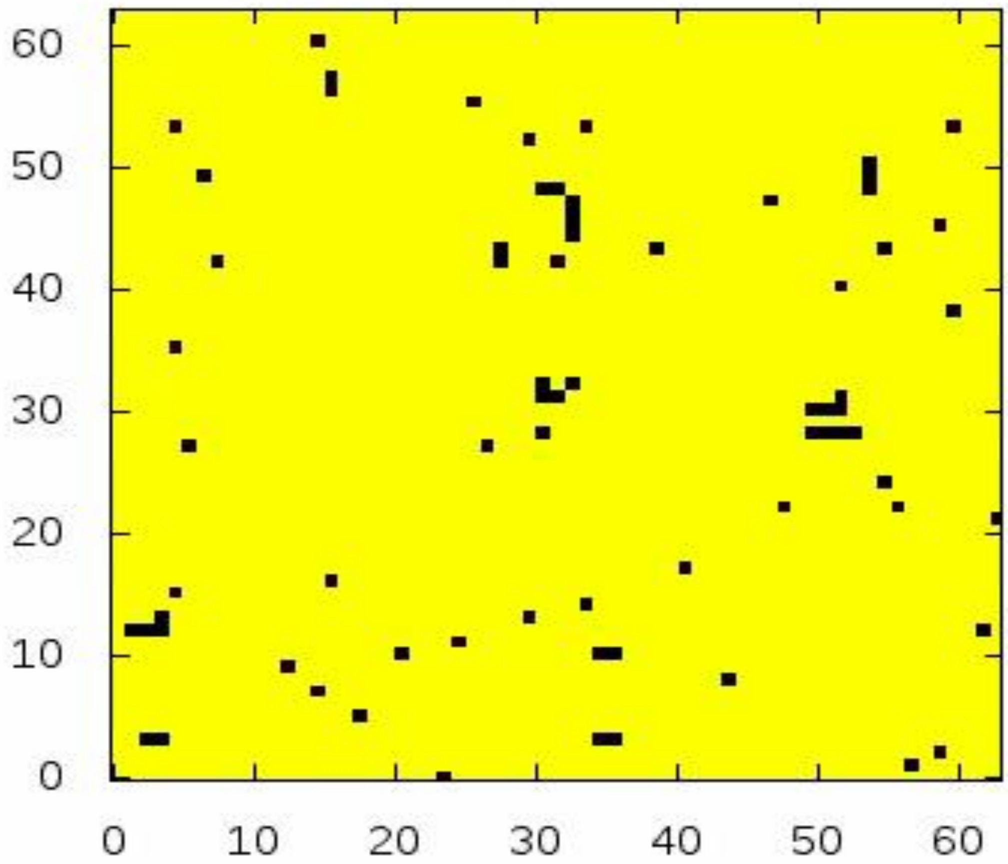
Nucleation under Fixed Shear Flow



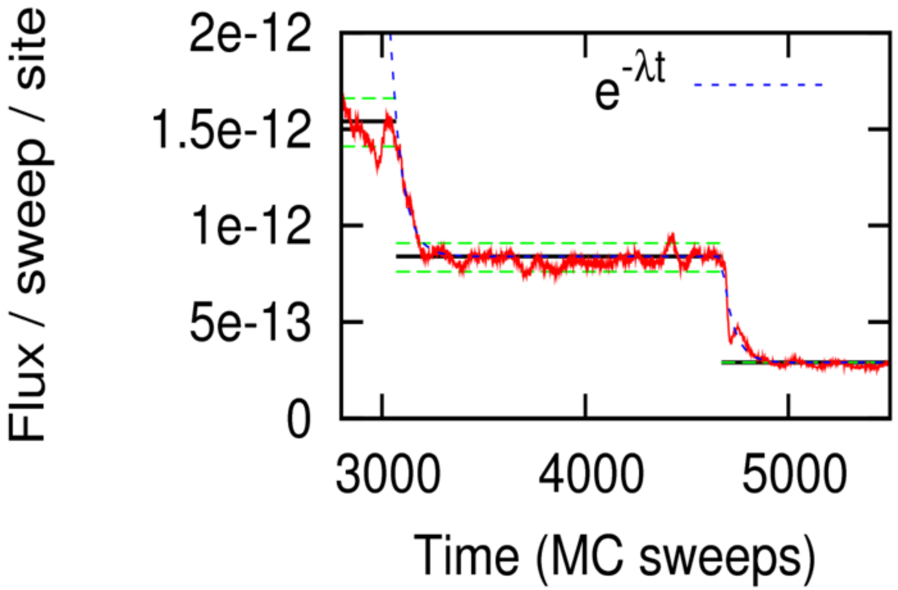
Shear Rate

R. J. Allen, C. Valeriani, S. Tanase-Nicola, P. R. ten Wolde and D. Frenkel.
J. Chem. Phys 2008

Nucleation under Fixed Shear Flow

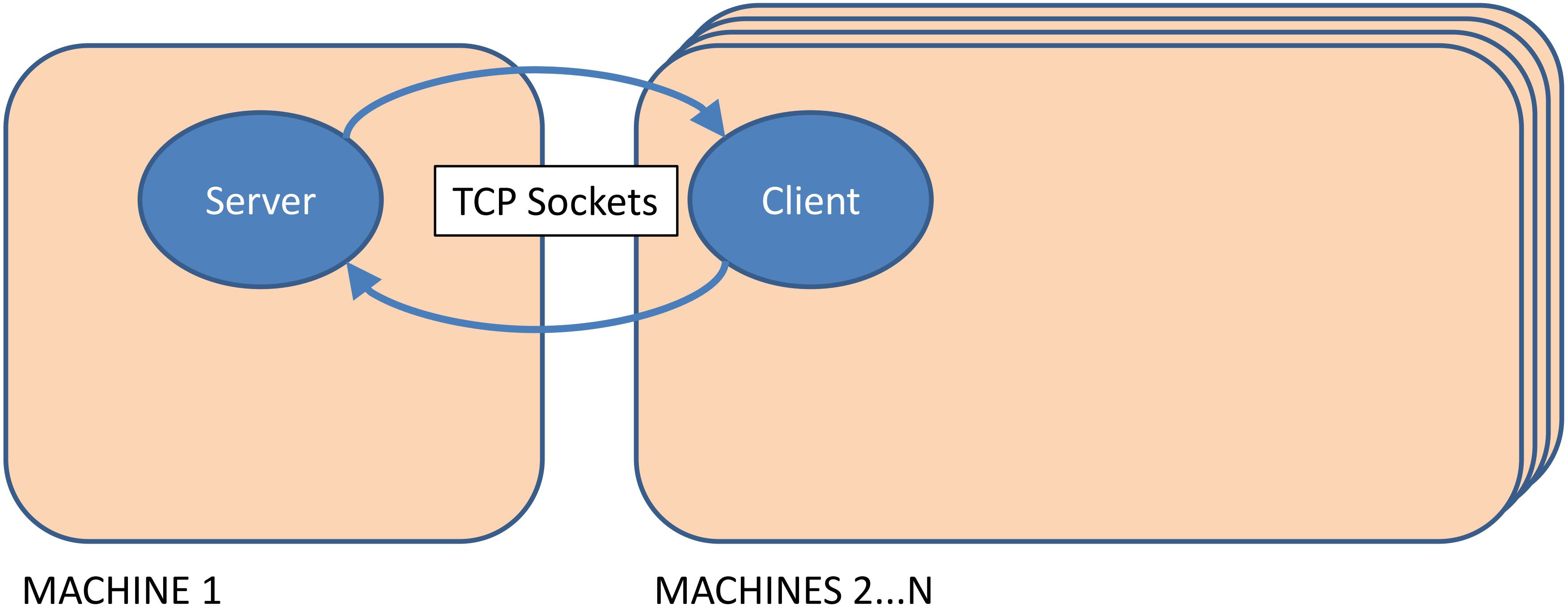


Nucleation under Variable Shear Flow

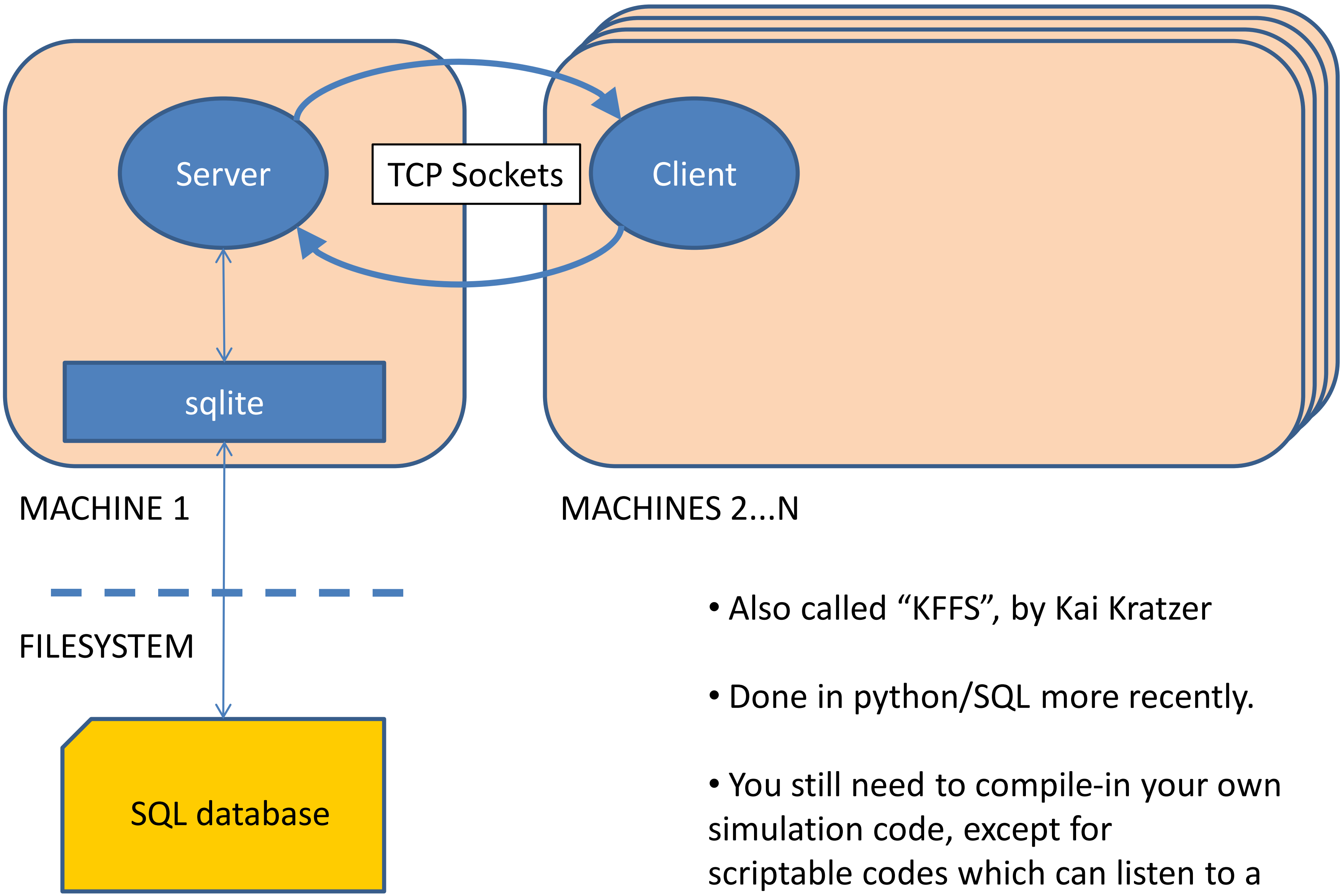


- Quickly approaches the stationary value
- Transients appear to fit an exponential, with same constant in both cases.

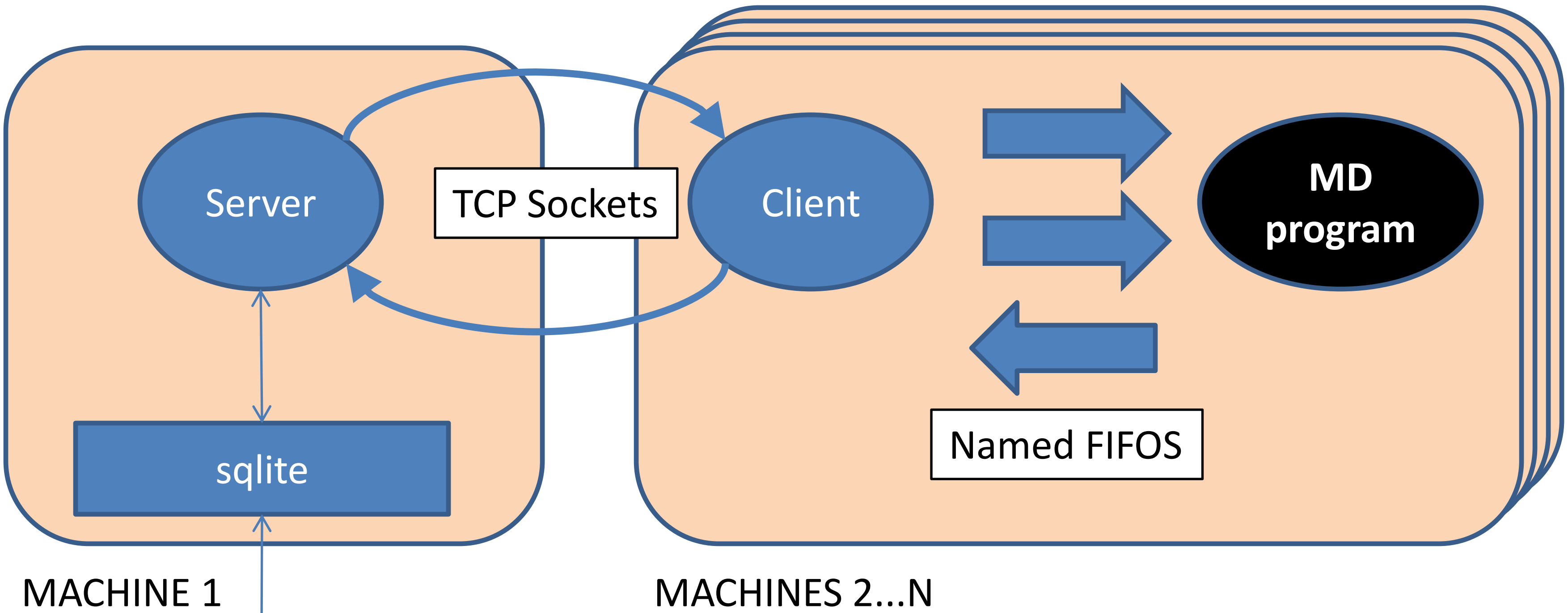
Berryman & Schilling, *J. Chem. Phys.* 2010

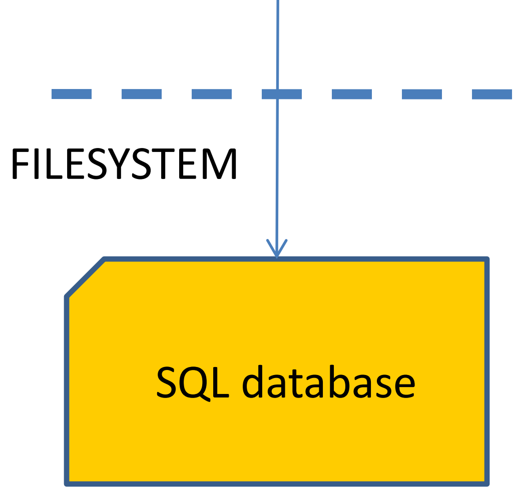


- "KFFS" by Koos van Meel et. al. ,
- Done in C++ a few years ago.
- OO, but you need to compile-in your own simulation code.

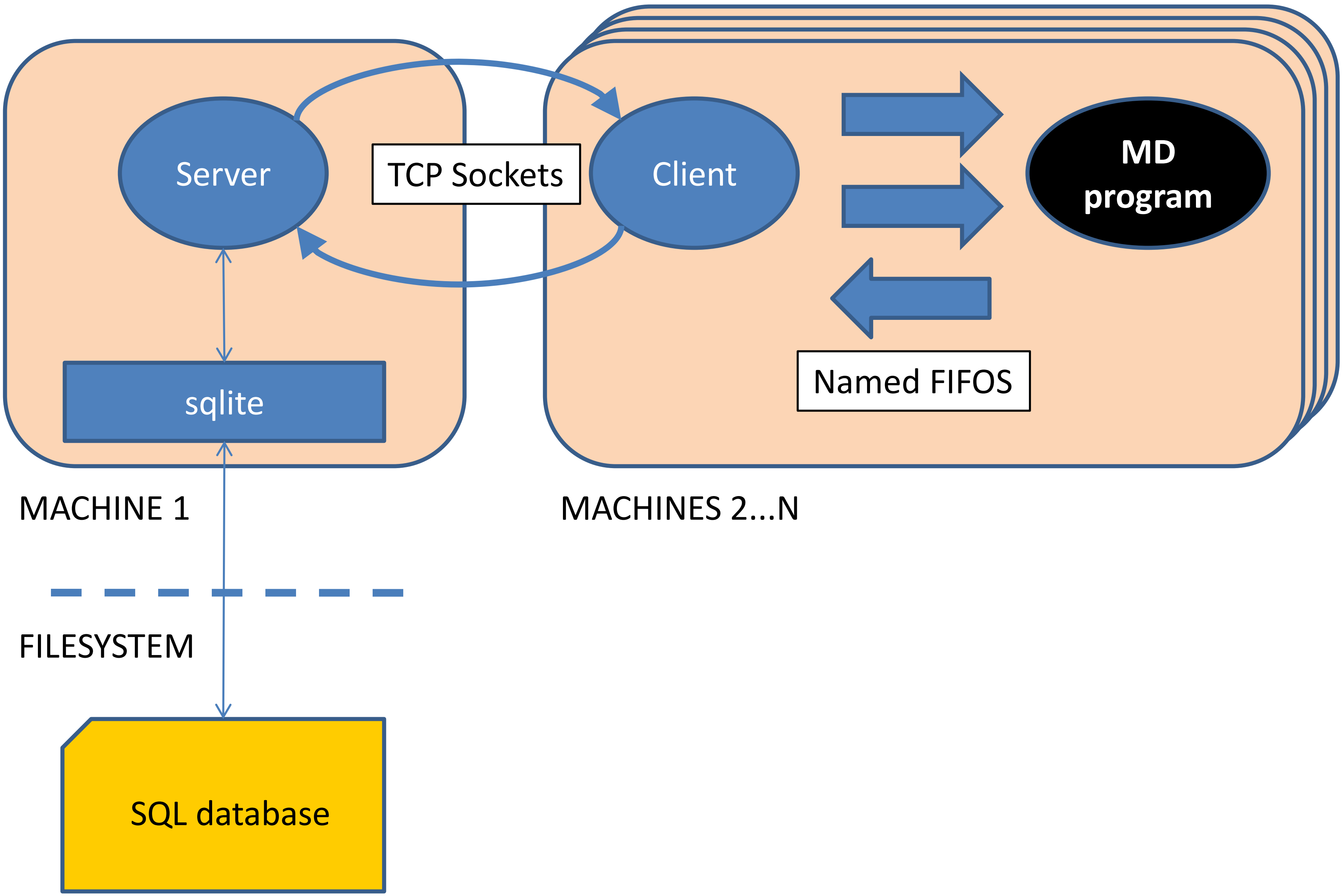


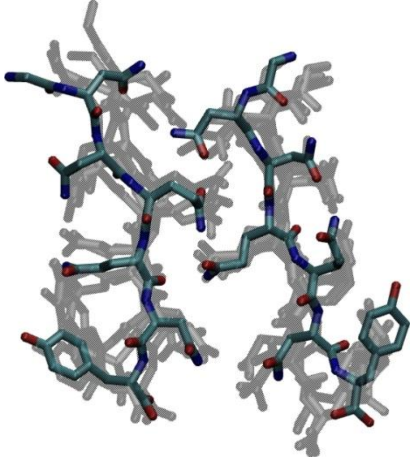
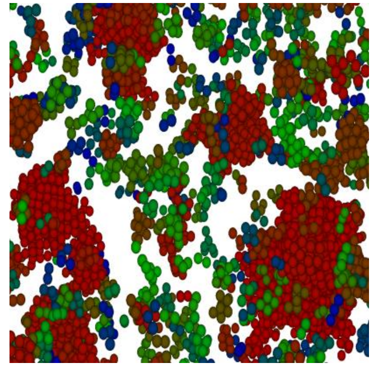
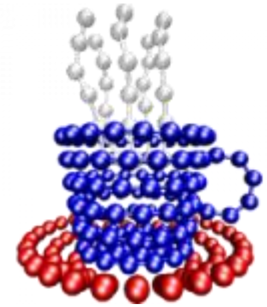
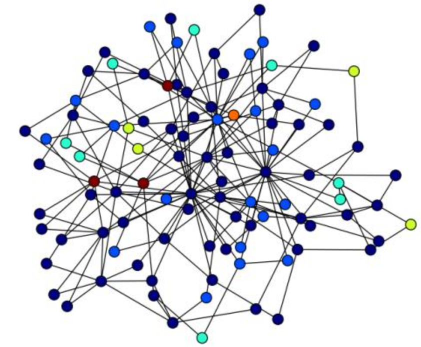
- Also called “KFFS”, by Kai Kratzer
- Done in python/SQL more recently.
- You still need to compile-in your own simulation code, except for scriptable codes which can listen to a TCP socket.





- My extension, renamed FRESHS
- FIFOS: so no need to hammer the filesystem client-side
- Suggest distributing the client + MD prog as a VM.


Server
TCP Sockets
Client
MD program
sqlite
Named FIFOS
MACHINE 1
MACHINES 2…N
FILESYSTEM
SQL database

	GROMACS	ell	espressoMD	networkX
SPRES		Is Fast	Is Fast	Is Fast
FFS		Works	Is Fast	Is Fast

- Open to people wishing to add new algorithms
- Open to people wishing to write wrappers for new simulation codes.
- Developer’s guide on request.

Summary:

- New technique developed for simulating rare events, for example nucleation.

- This works away from steady states; it gives explicit time dependence.

- Tested against some extant steady-state numerical calculations (shear flow) and against theory (T quench & ASEP).

- Trivial parallelisation available; speedup is robust to parameter choice. Harness available.

Afterthought:

- Sometimes people invest effort in mapping a high-level Markov model onto the microscopic dynamics of their system

- SPRES gives you this for free. The matrix of bin-bin transition probabilities (can be) a very neat summary of the important pathways of the dynamics.

Thanks

Tanja & rest of the Binder group (at the time)

Rosalind Allen (inspiration)

Kai Kratzer & Axel Arnold (FRESHS)

Deutsche Forschungsgemeinschaft (funding)

Université du Luxembourg (funding)

Fonds Nationale de Recherche (funding)

NGS UK & Uni Luxembourg HPC & ESMI (CPU)