Polymer simulations and Bayesian data analysis

Daan Frenkel

U.Cambridge

With:

Behnaz Bozorgui

Columbia U/AMOLF

Kostya Shundyak

AMOLF

Outline:

1. Recursive enumeration
   a) Dynamics (simulation)
   b) Polymer statistics (simulation)
2. Molecular Motors (experiments!)

   (well, actually, simulated experiments)

Background:

Velocity auto-correlations functions exhibit algebraic long-time tails.

Always.

On-lattice, off-lattice.

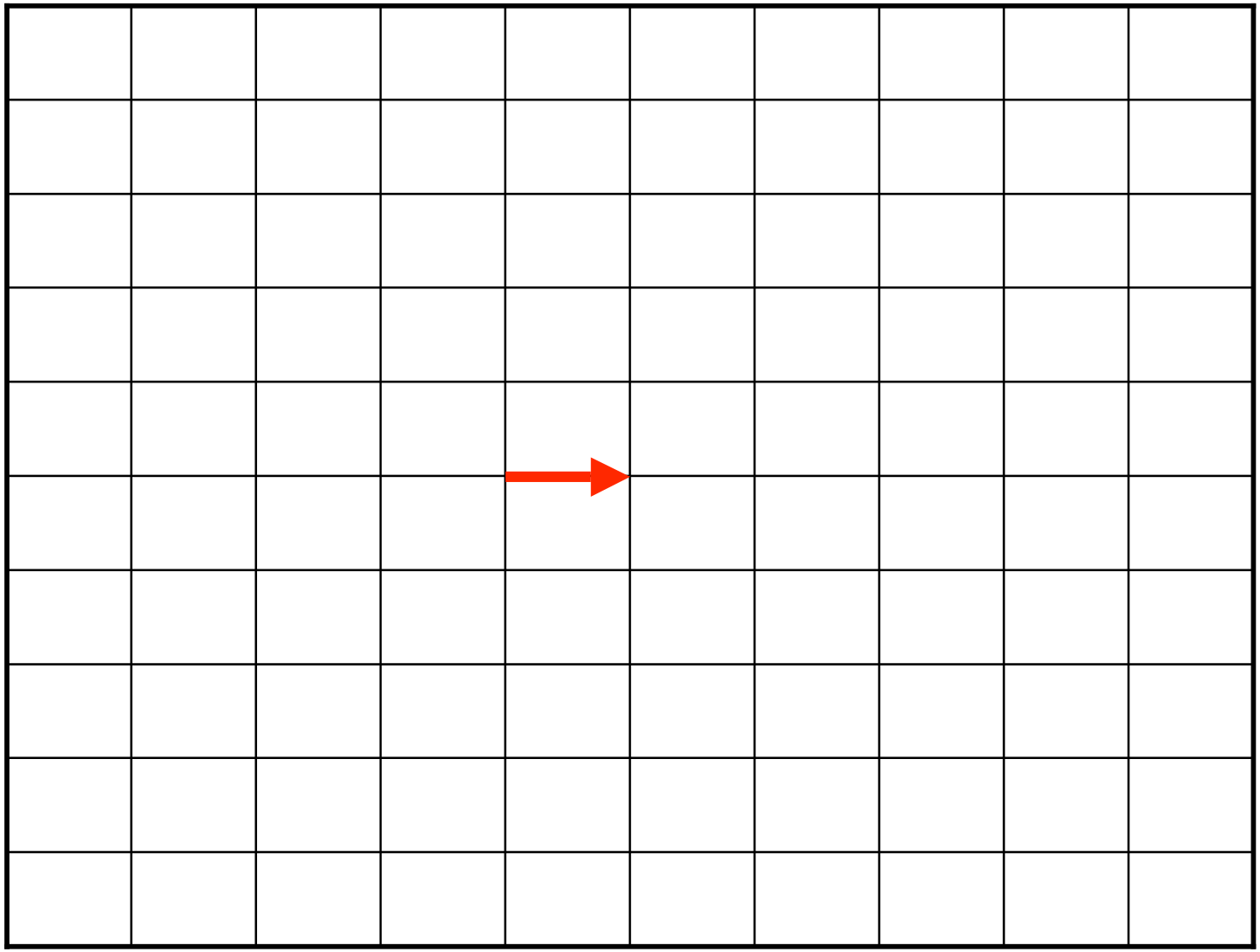
Ballistic (MD), Diffusive (Lorentz-gas)

Background:

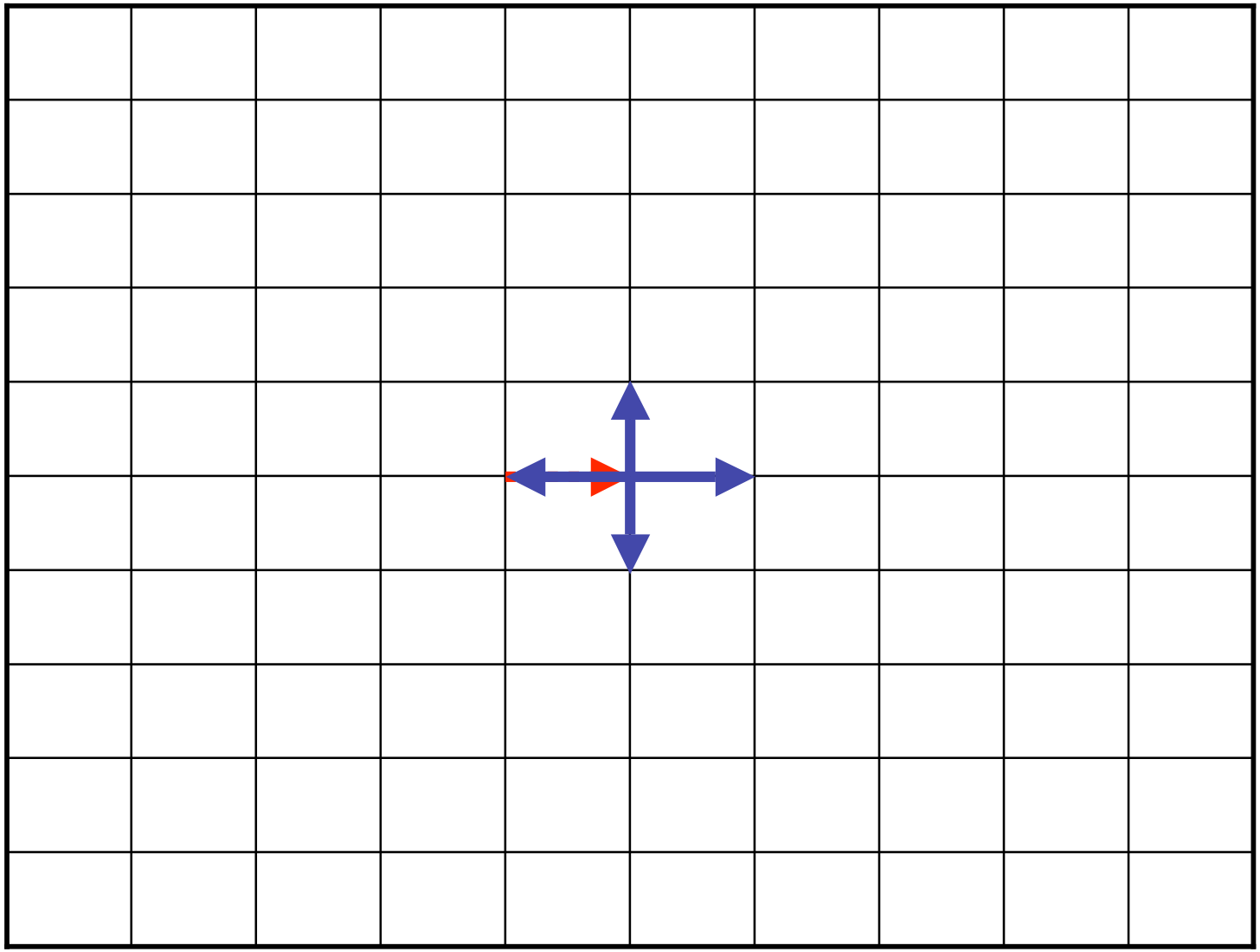
Recursive algorithms allow very accurate computation of <v(0).v(t)> for lattice models.

Example:

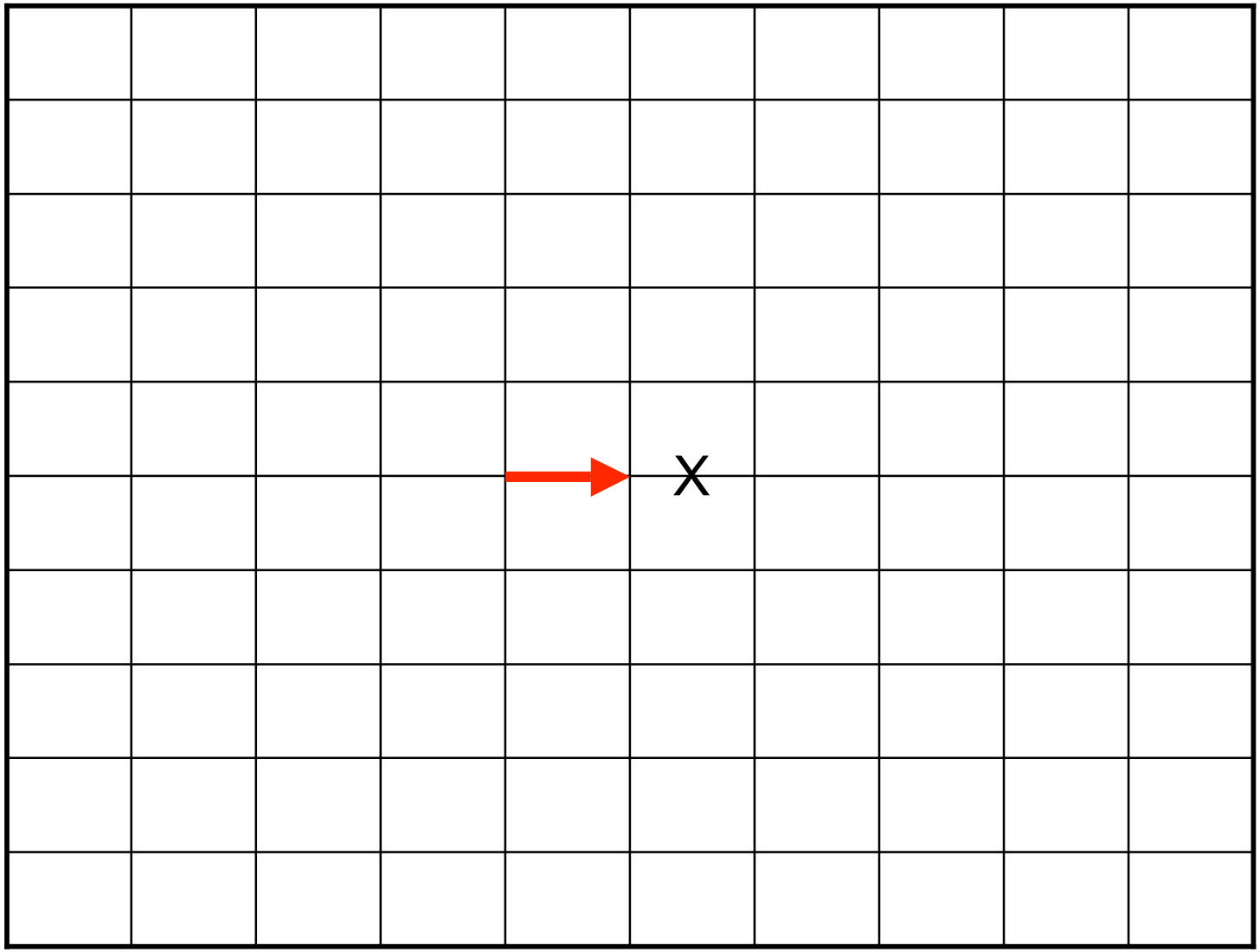
2D Lorentz gas



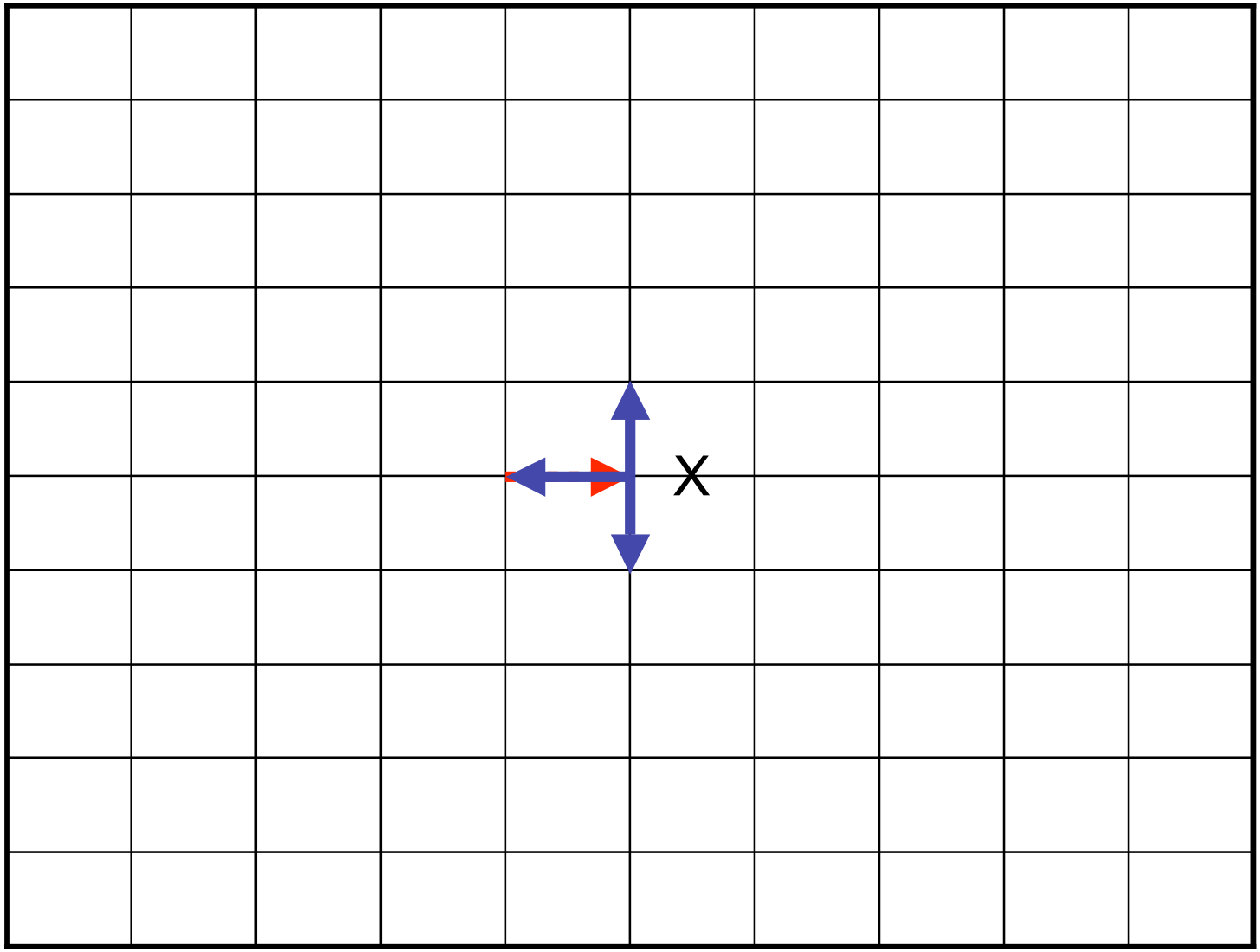
$V_x(0)=1$



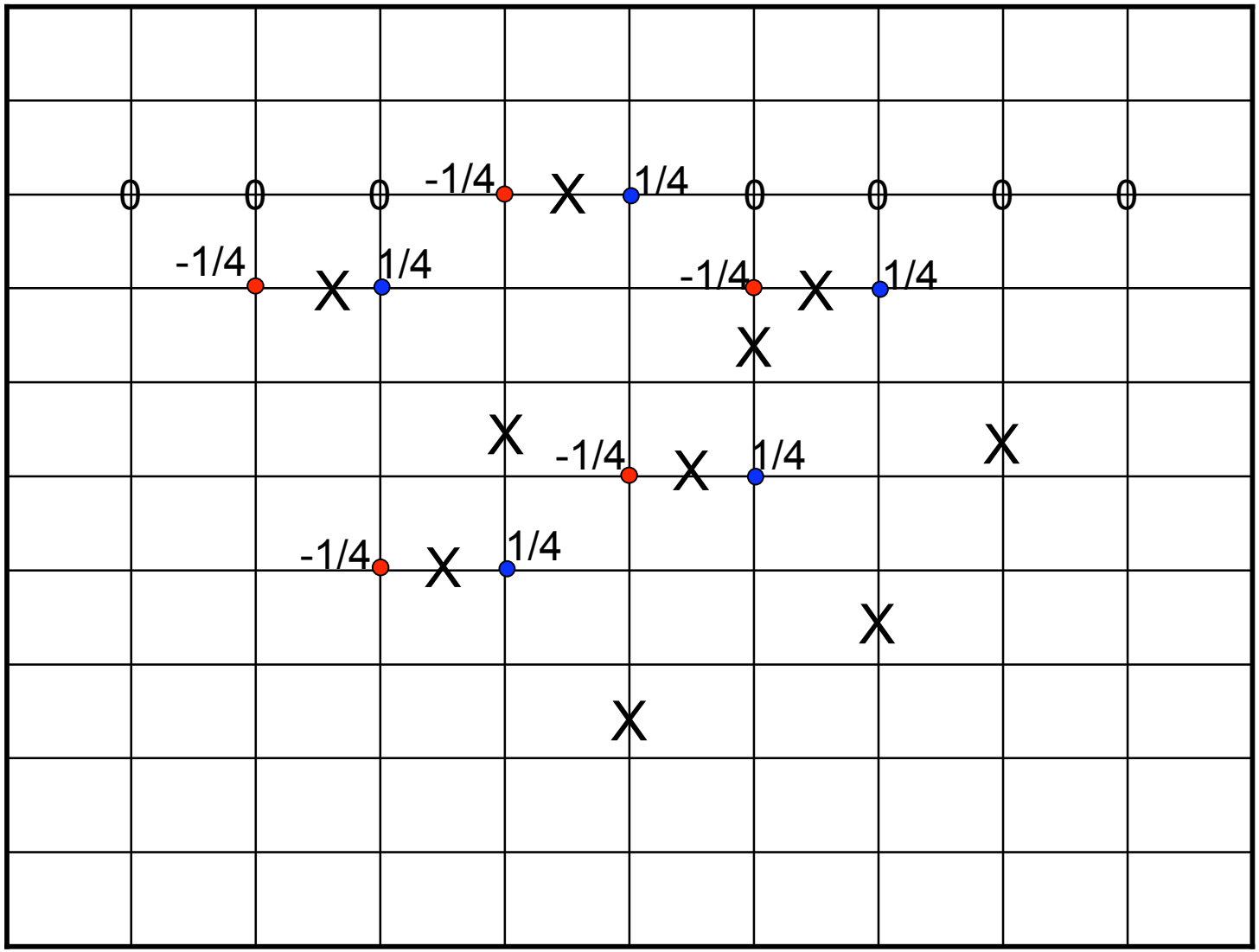
$V_x(0)=1$ $\langle V_x(1)\rangle=0$



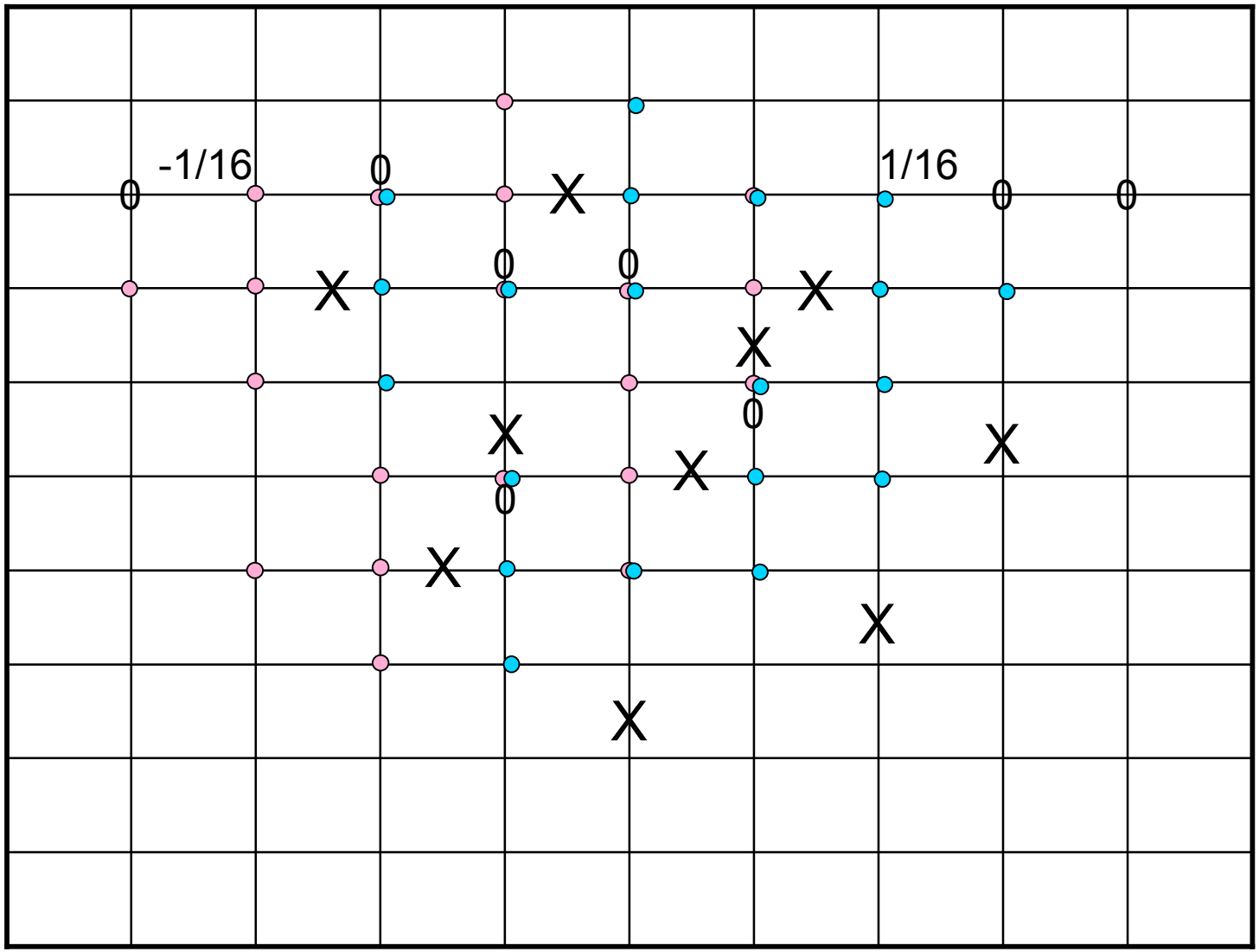
$V_x(0)=1$



$V_x(0)=1$ $\langle V_x(1)\rangle=-1/4$



$\langle V_x(1) \rangle_i$

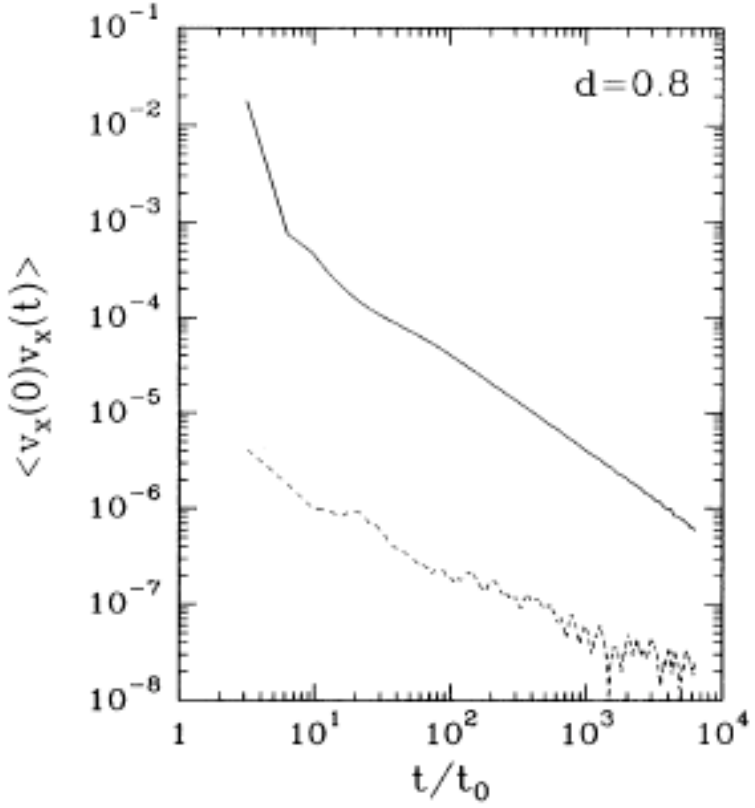


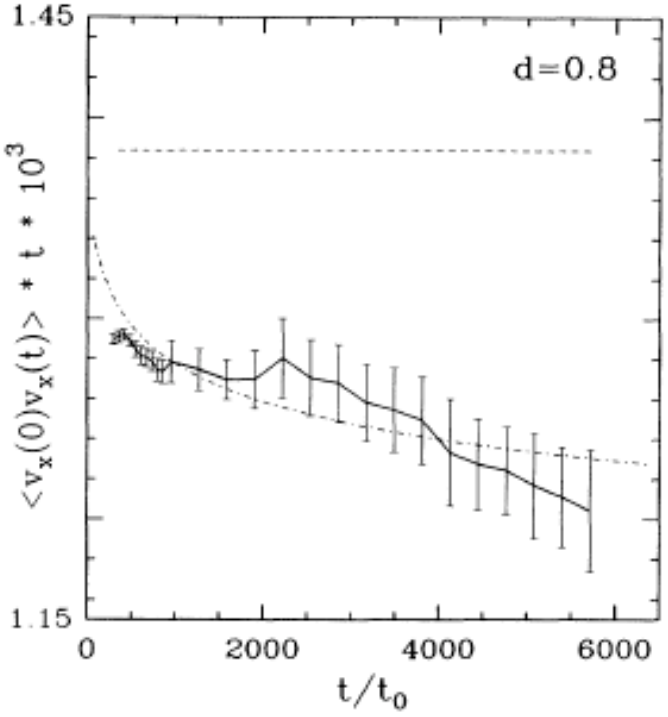
$\langle V_x(2) \rangle_i$

Iterate recursively: ⟹

$\langle V_x(3) \rangle_i$, $\langle V_x(4) \rangle_i$, $\langle V_x(5) \rangle_i$, … , $\langle V_x(n) \rangle_i$,

Method can be generalised to lattice-gas cellular automata ("hydrodynamic long-time tails")





Application to (lattice) polymers:

Consider a lattice (e.g. 2D-square).

At a given point x_i, the potential energy is $U(x_i)$.

The Boltzmann factor for a particle at point x_i is

$\exp(-\beta U(x_i)) \equiv z_i^1$

Consider a lattice (e.g. 2D-square).

At a given point x_i, the potential energy is $U(x_i)$.

The Boltzmann factor for a particle at point x_i is

$\exp(-\beta U(x_i)) \equiv z_i^1$

The partition function for a single point particle is

$Z_1 \equiv \sum_i z_i^1$

Dimers

The Boltzmann factor for a dimer on points x_i and x_{i+1} is

$$\exp(-\beta(U(x_i) + U(x_{i+1})) = z_i^1 \times z_{i+1}^1$$

The Boltzmann factor for all dimers terminating

on point x_i is

$$z_i^{(2)} \equiv z_i^1 \times \sum_{jnni} z_j^1$$

The partition function for a single dimer is

$$Z_2 \equiv \sum_i z_i^{(2)}$$

n-mers

The Boltzmann weight for an n-mer terminating on point x_i is

$$z_i^{(n)} = z_i^1 \times \sum_{jnni} z_j^{(n-1)}$$

and the corresponding partition function is

$$Z^{(n)} = \sum_i z_i^{(n)}$$

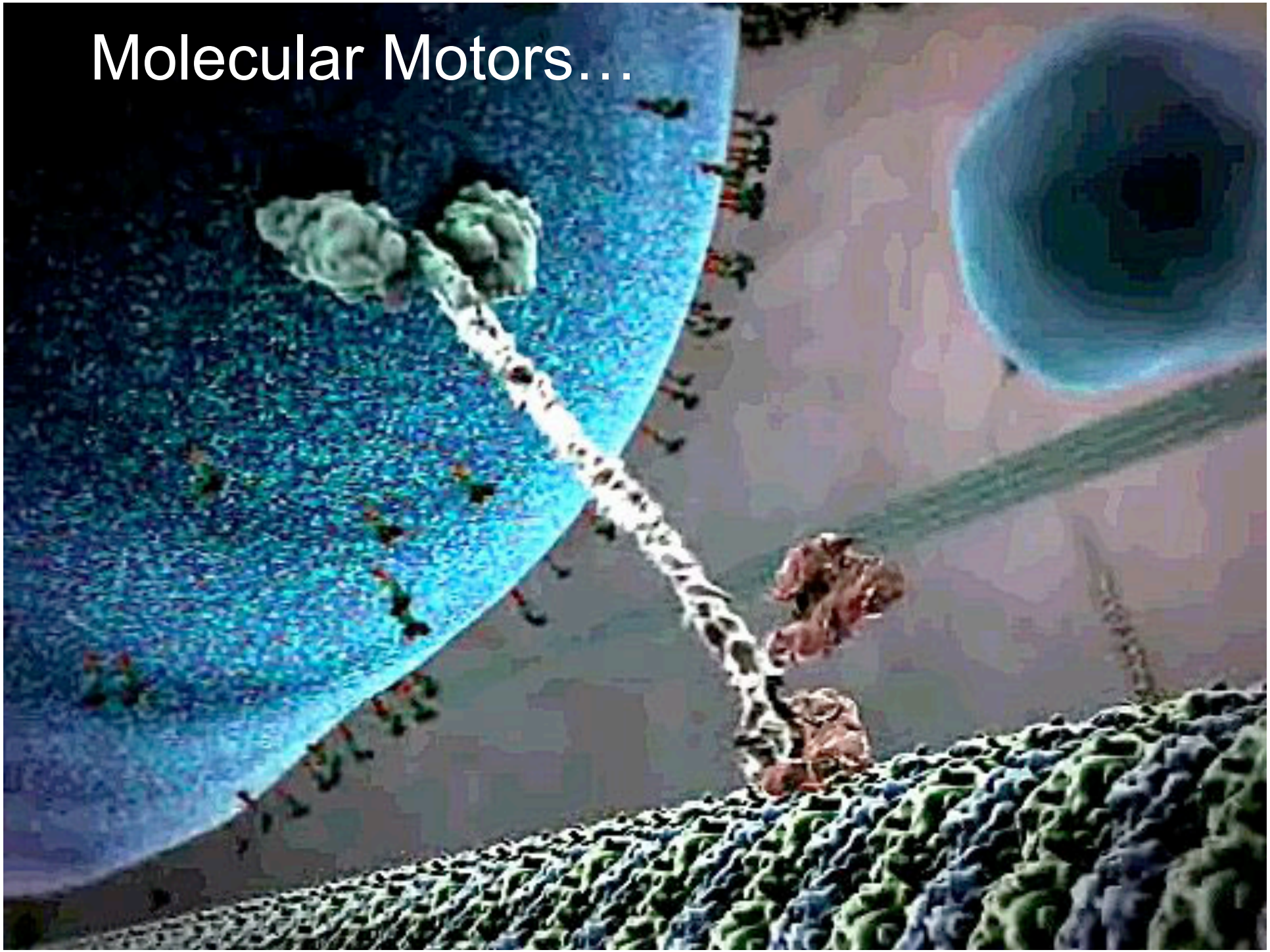
This method is exact for non-self-avoiding, non-interacting lattice polymers.

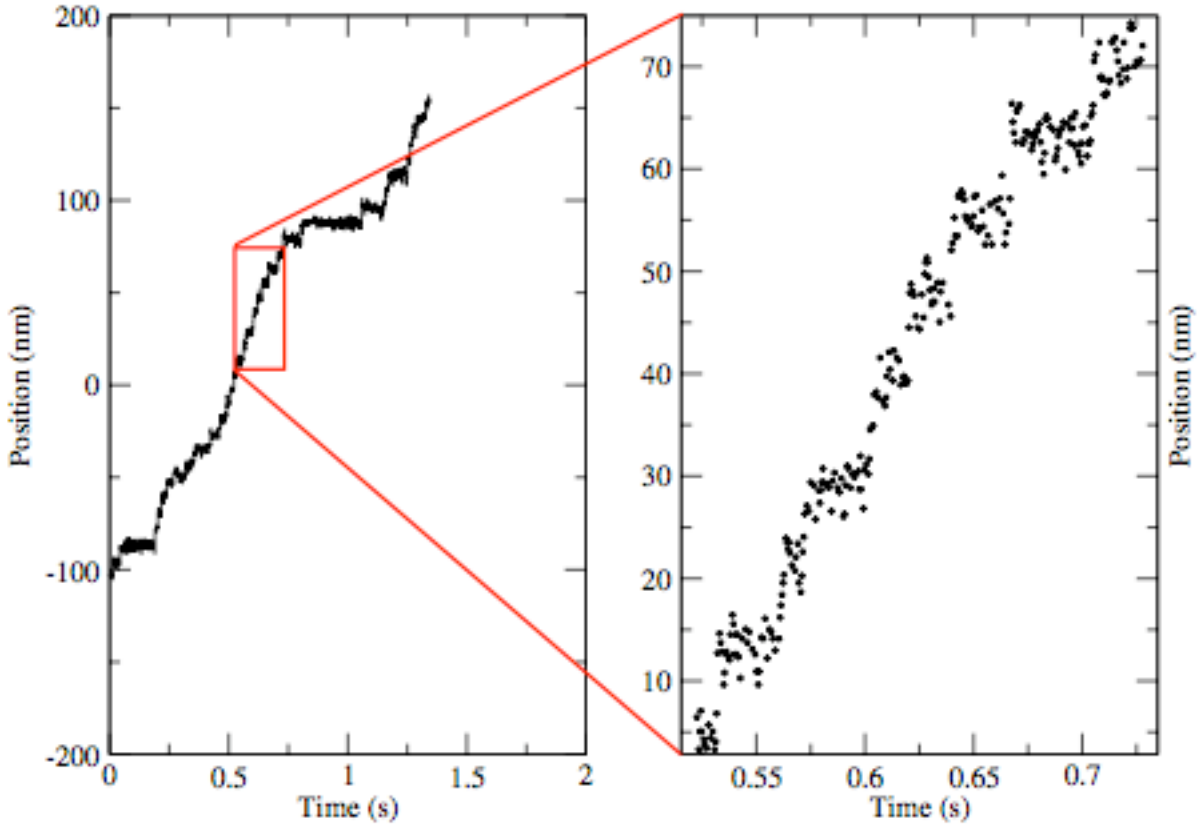
It can be used to speed up MC sampling of (self)interacting polymers

B. Bozorgui and DF, Phys. Rev. E 75, 036708 (2007))

END OF INTRODUCTION

Molecular Motors…

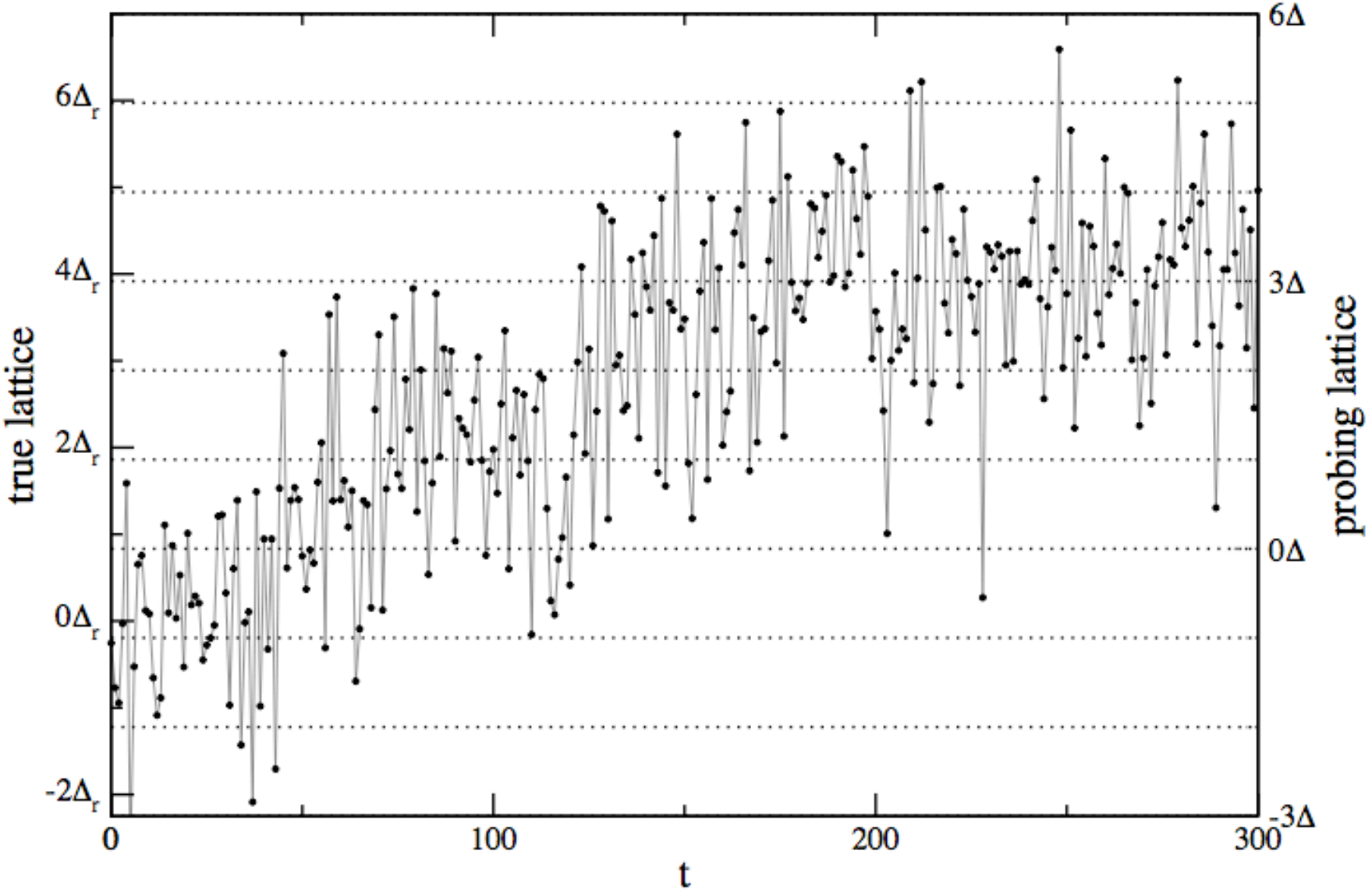




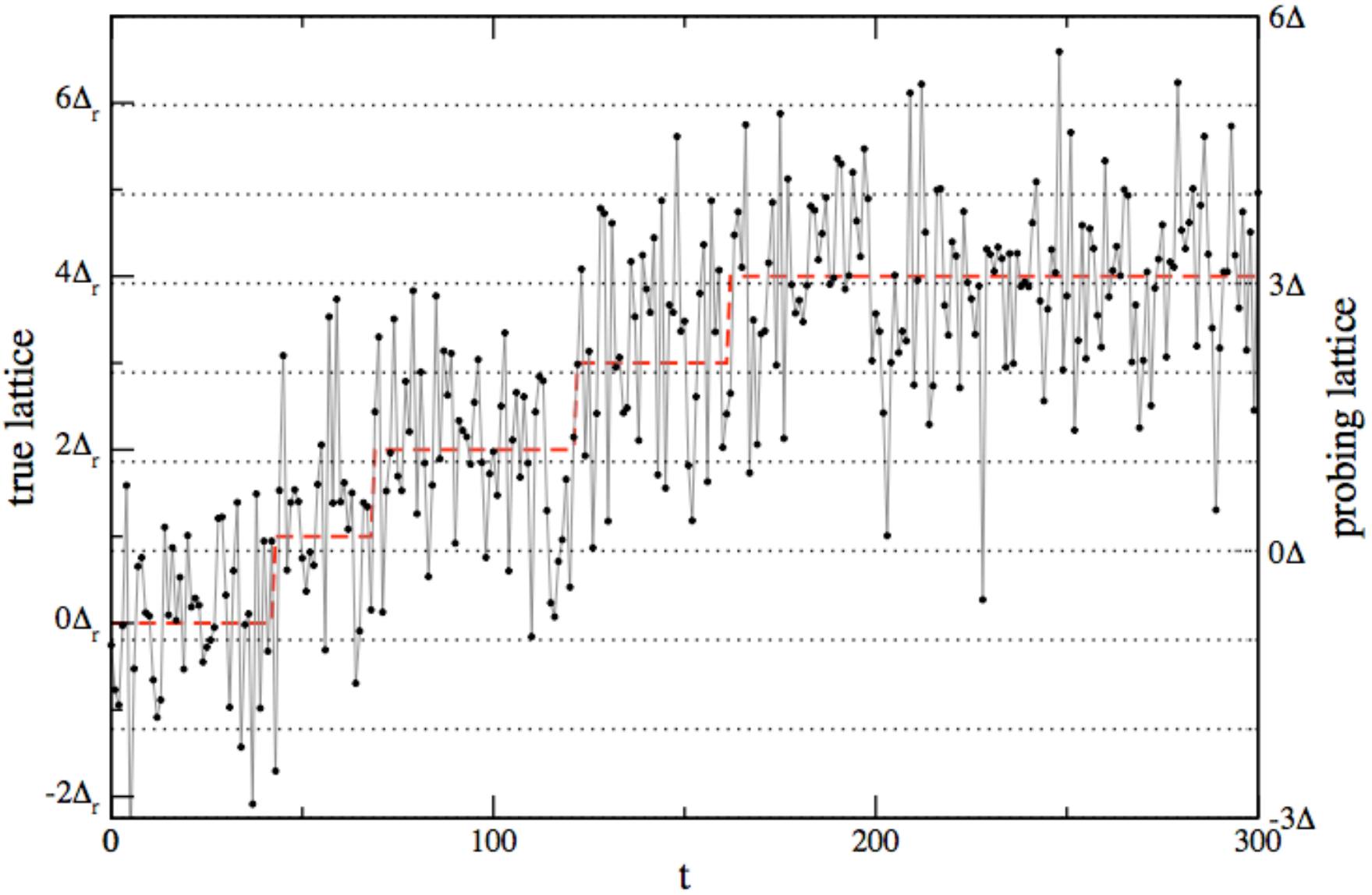
Kinesin motor steps along micro-tubules with a step size of 8nm

Experimentally, the step size is measured by fitting the (noisy) data.

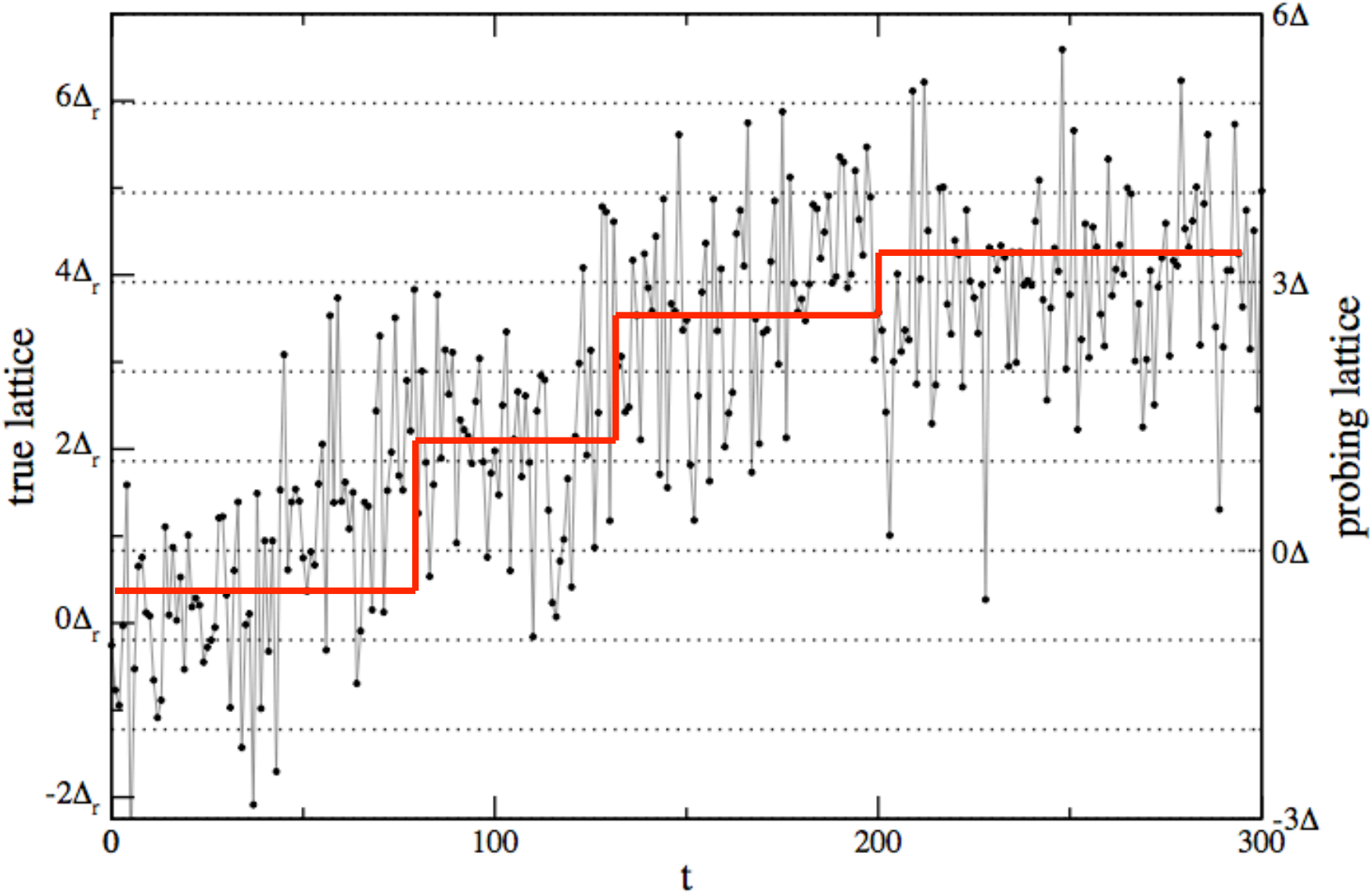
Example: noisy “synthetic data”



Example: noisy “synthetic data”



Best practice: “fit steps to data”



J.W.J. Kerssemakers et al. , Nature 442,709 (2006)

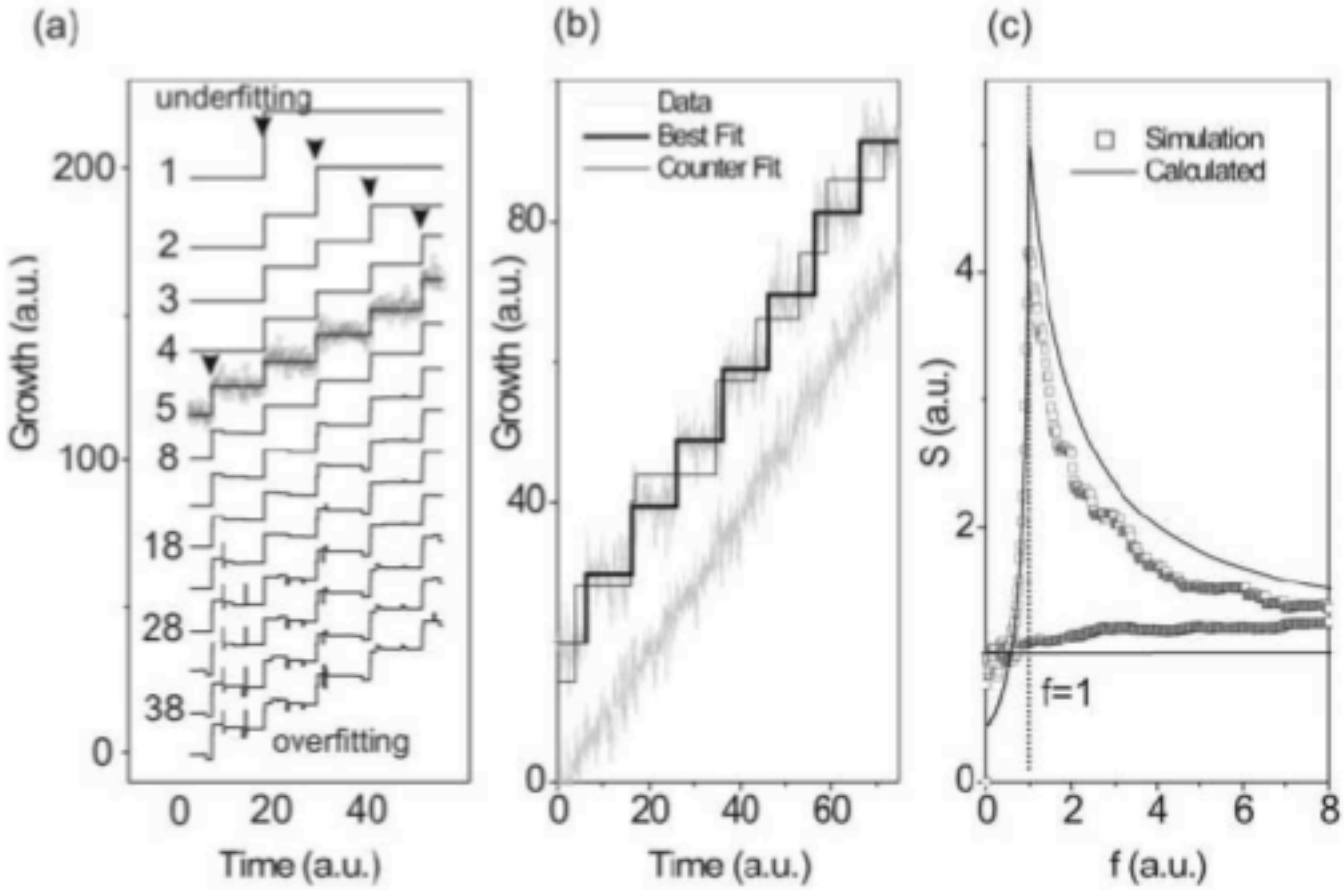
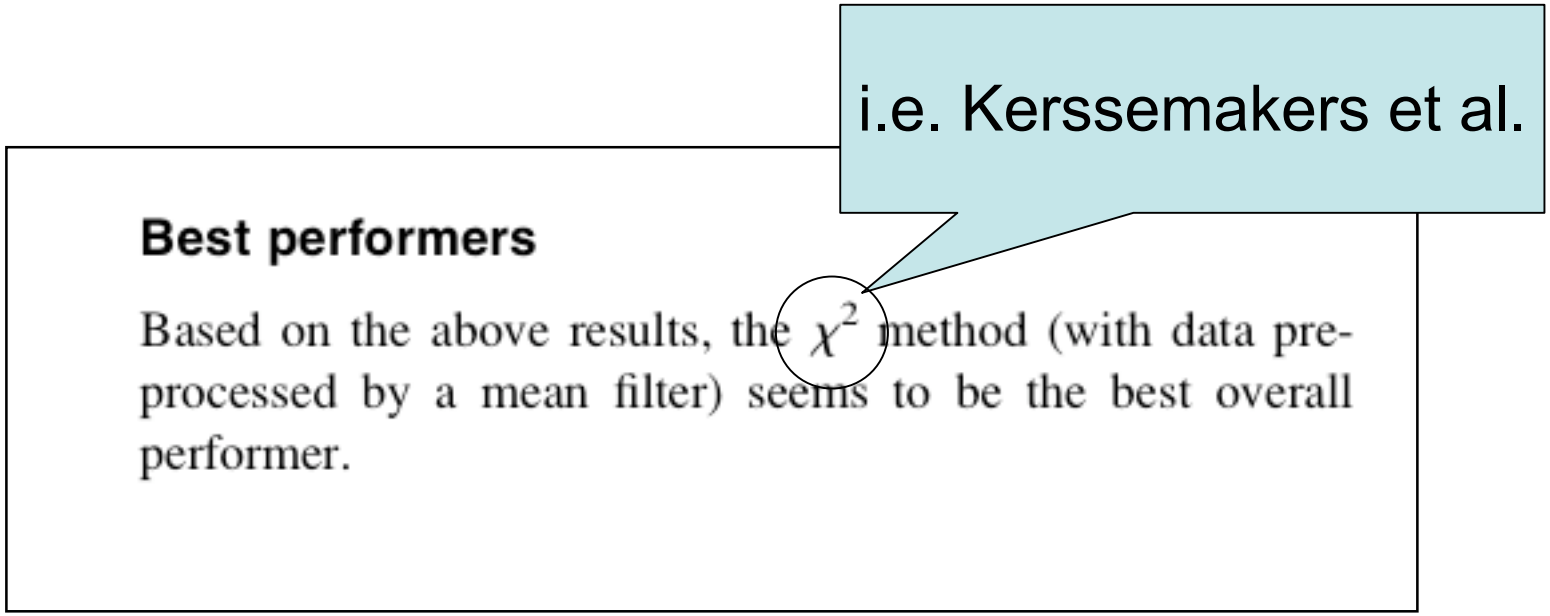


Figure 3.8: Step fitting procedure from Ref. [30]. a) Iterations of the step fitting algorithm. The arrowheads show every new step that is added in each iteration. b) A best-fit together with a counter-fit for a stepping data [30]. The lower curve shows linear non-stepping data c) S parameter versus the relative number of fitted steps from simulation results of Ref. [30]

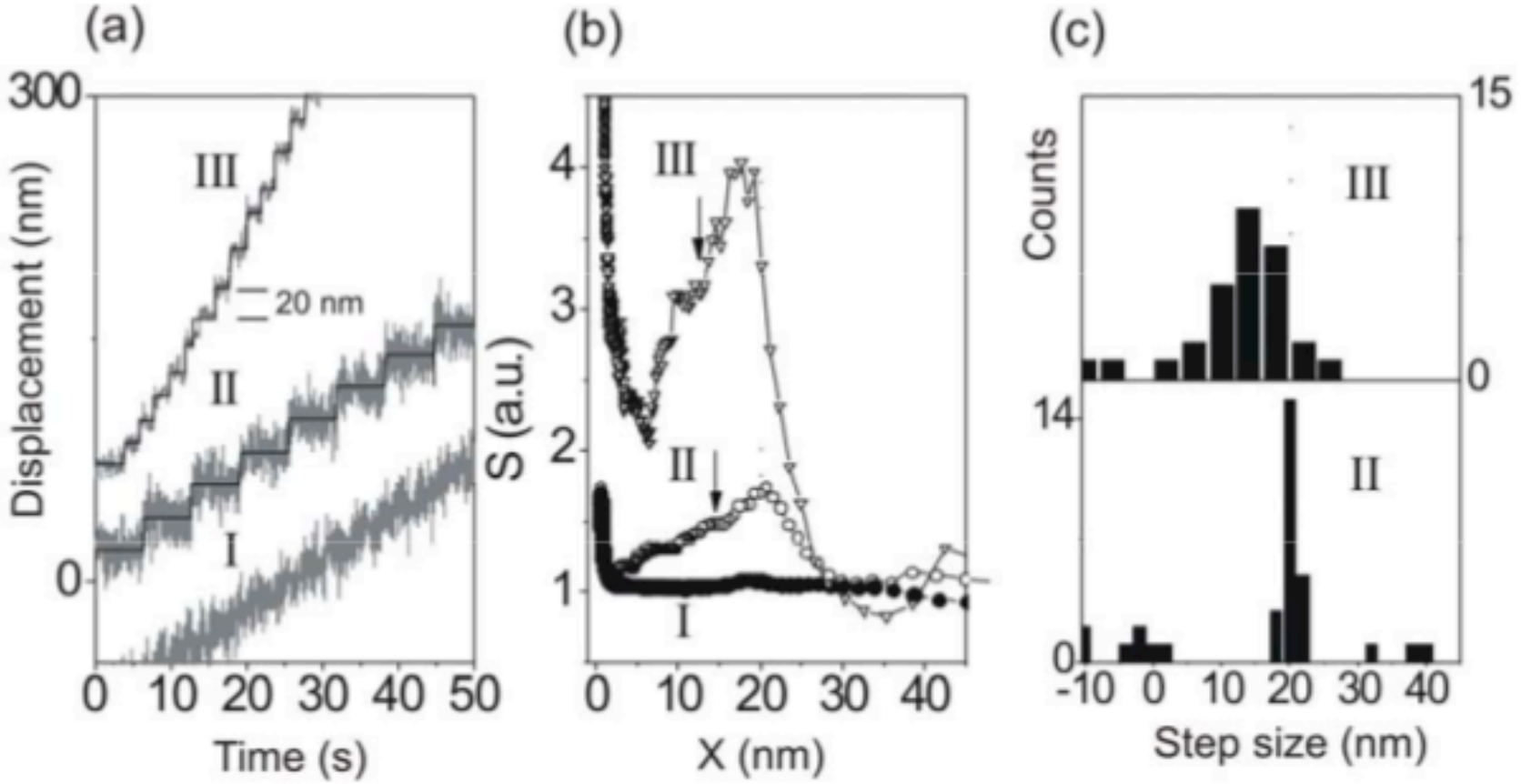
A recent article [Carter et al, Biophys. J. 94, 306(2008)] compared all existing step fitting methods and concluded:



From now on, the χ^2 method is my reference.

How well does it perform?

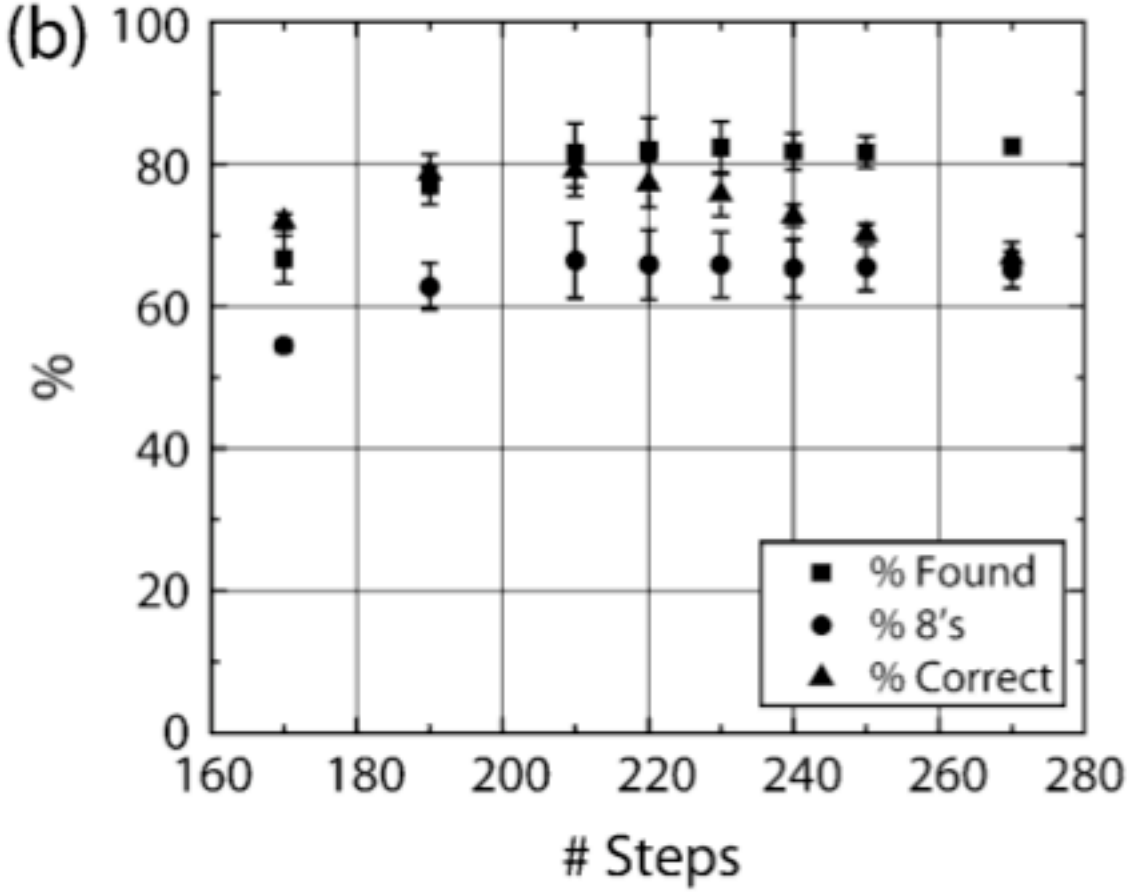
1. It can be used if the noise is less than 60% of the step size.
2. It yields a distribution of step sizes (even if the underlying process has only one step size)



(I): no steps

(II): step = 2 × noise amplitude

(III): step = 4 × noise amplitude



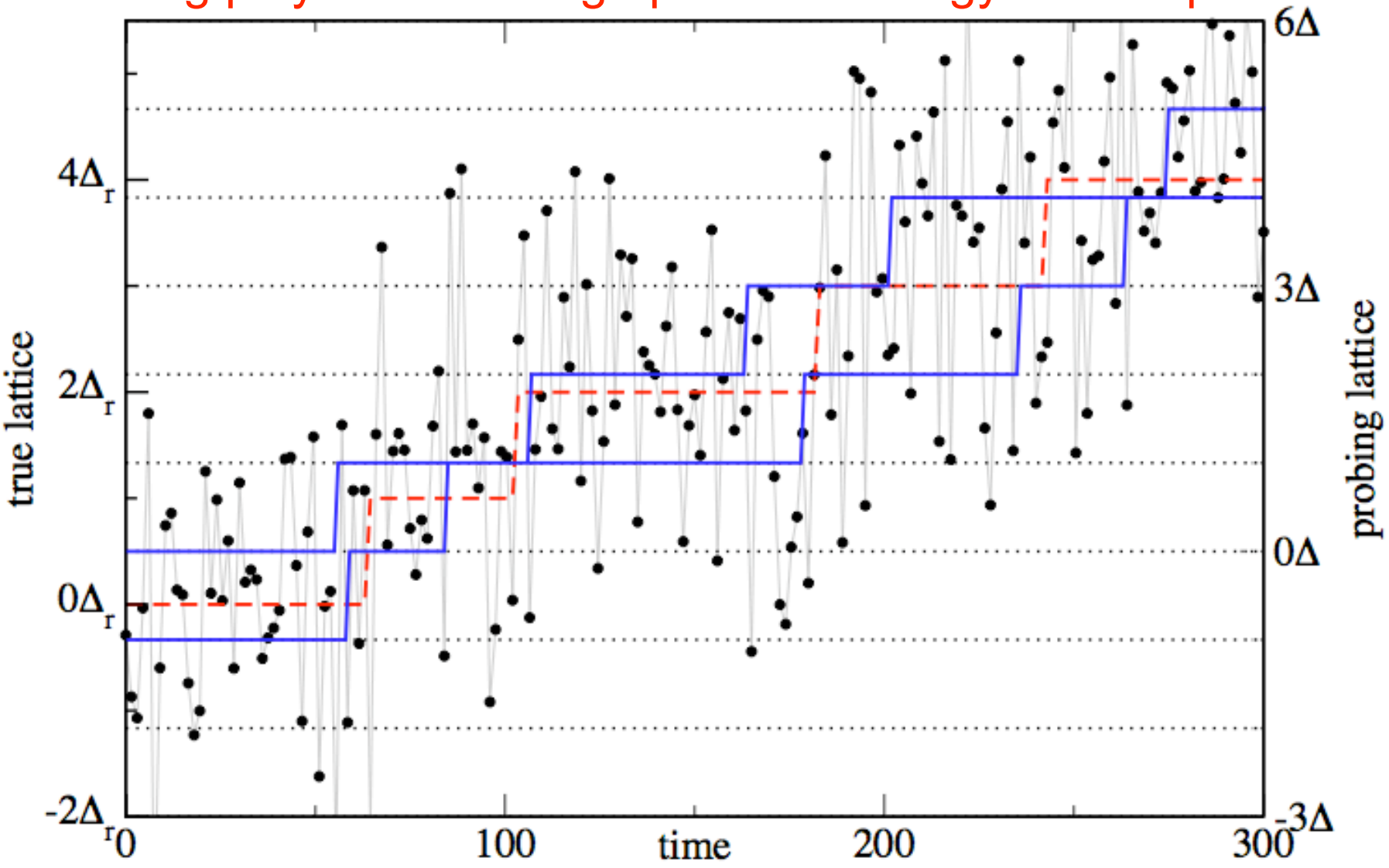
For our benchmark datasets that have only 8-nm steps, we can measure the fraction of correct a posteriori steps whose size is found to be 8 nm (±3 nm). This is reported as the third parameter, ''percent 8's.''

Observation:

We want to know the step size and the step frequency but…

We do not care which trace is the “correct” trace.

Bayesian approach: compute the partition function **Q** of non-reversing polymer in a rough potential energy landscape



"true" trace

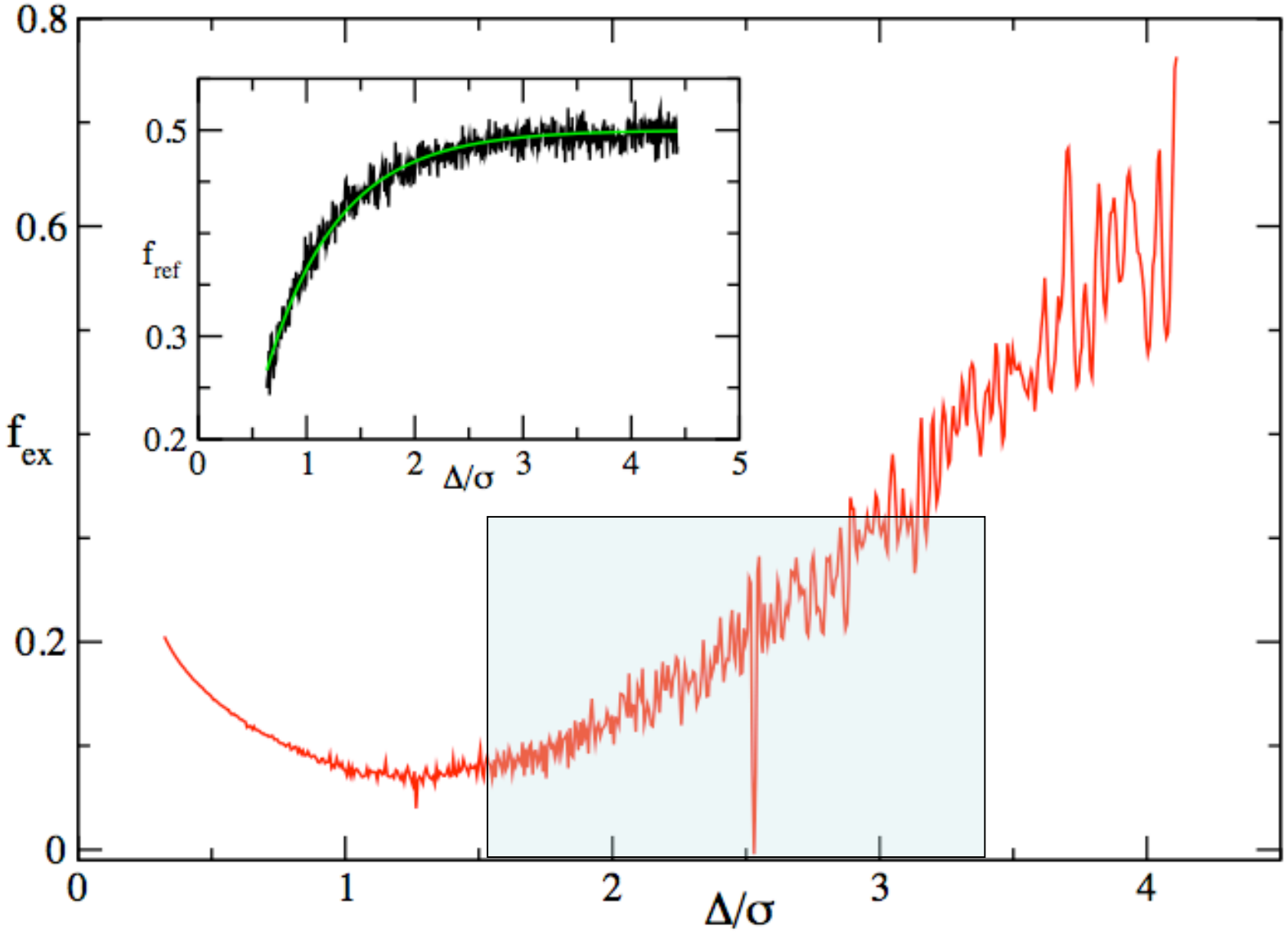
Other directed walks

As shown before: we can enumerate **Q** exactly (and cheaply).

From **Q** we can compute a “free energy”

$$(F = -\ln Q)$$

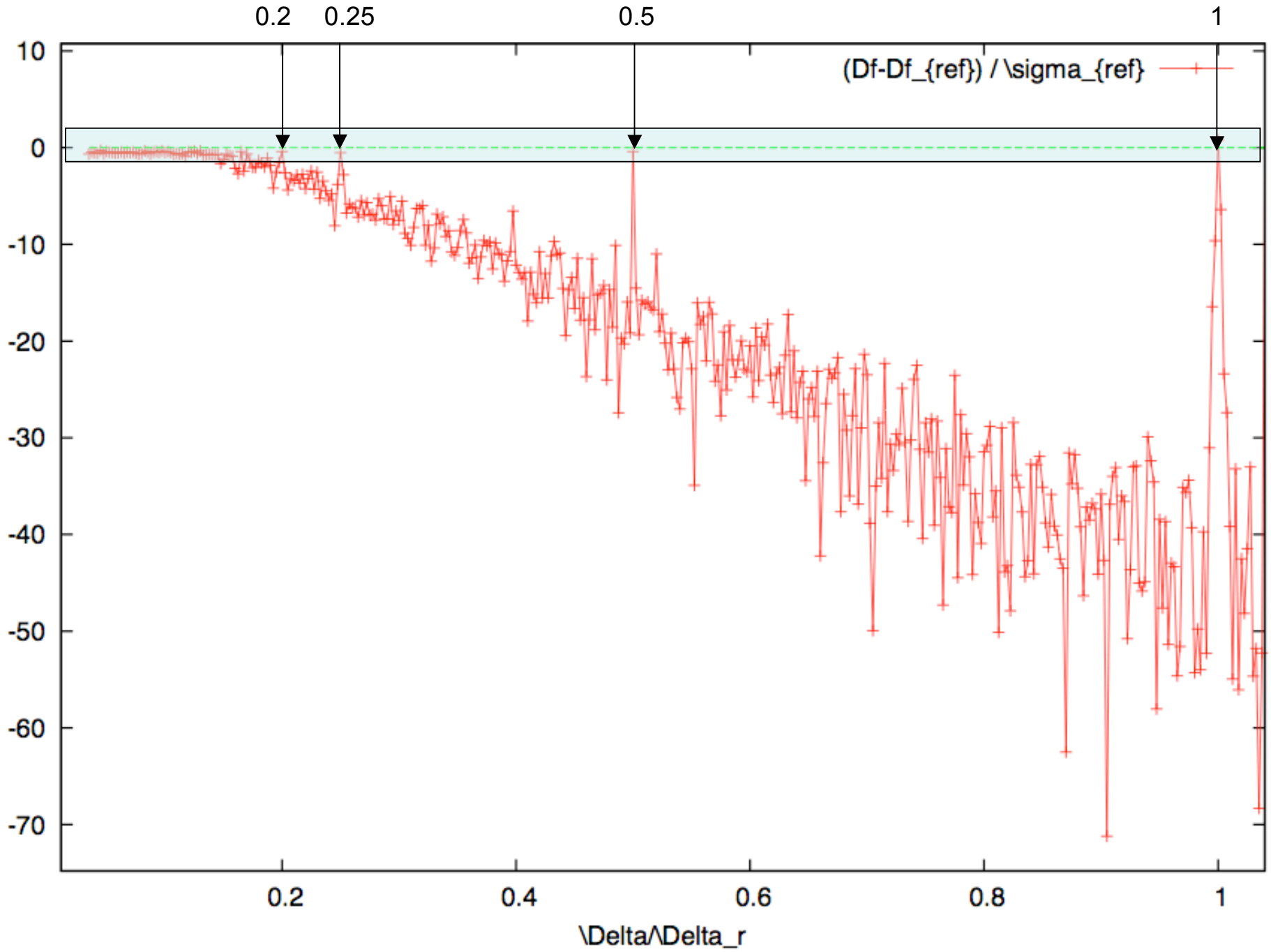
Compute the “excess free energy” with respect to reference data



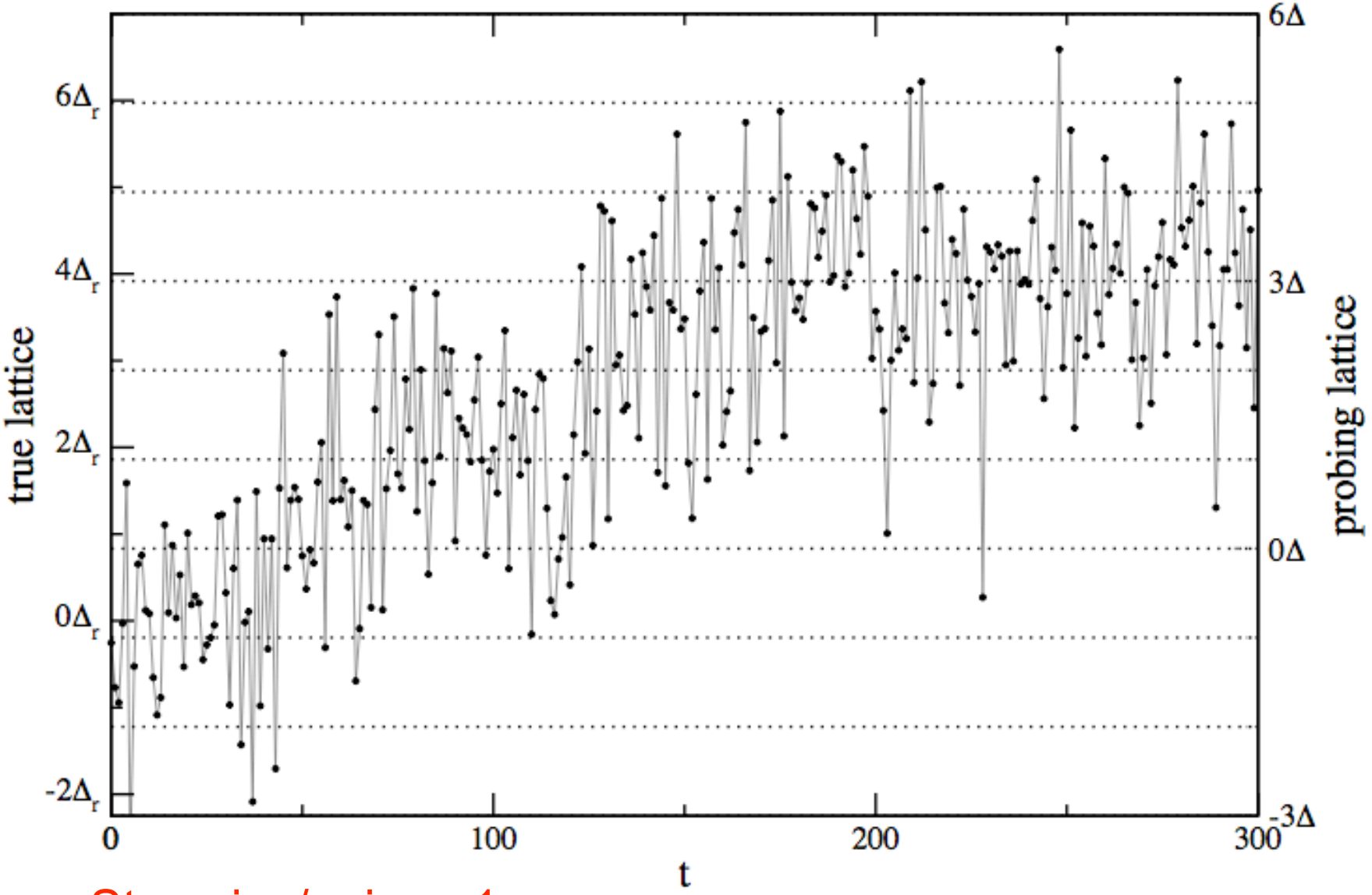
More detailed analysis (including noise estimates etc):

Determine the maximum “modulation amplitude”

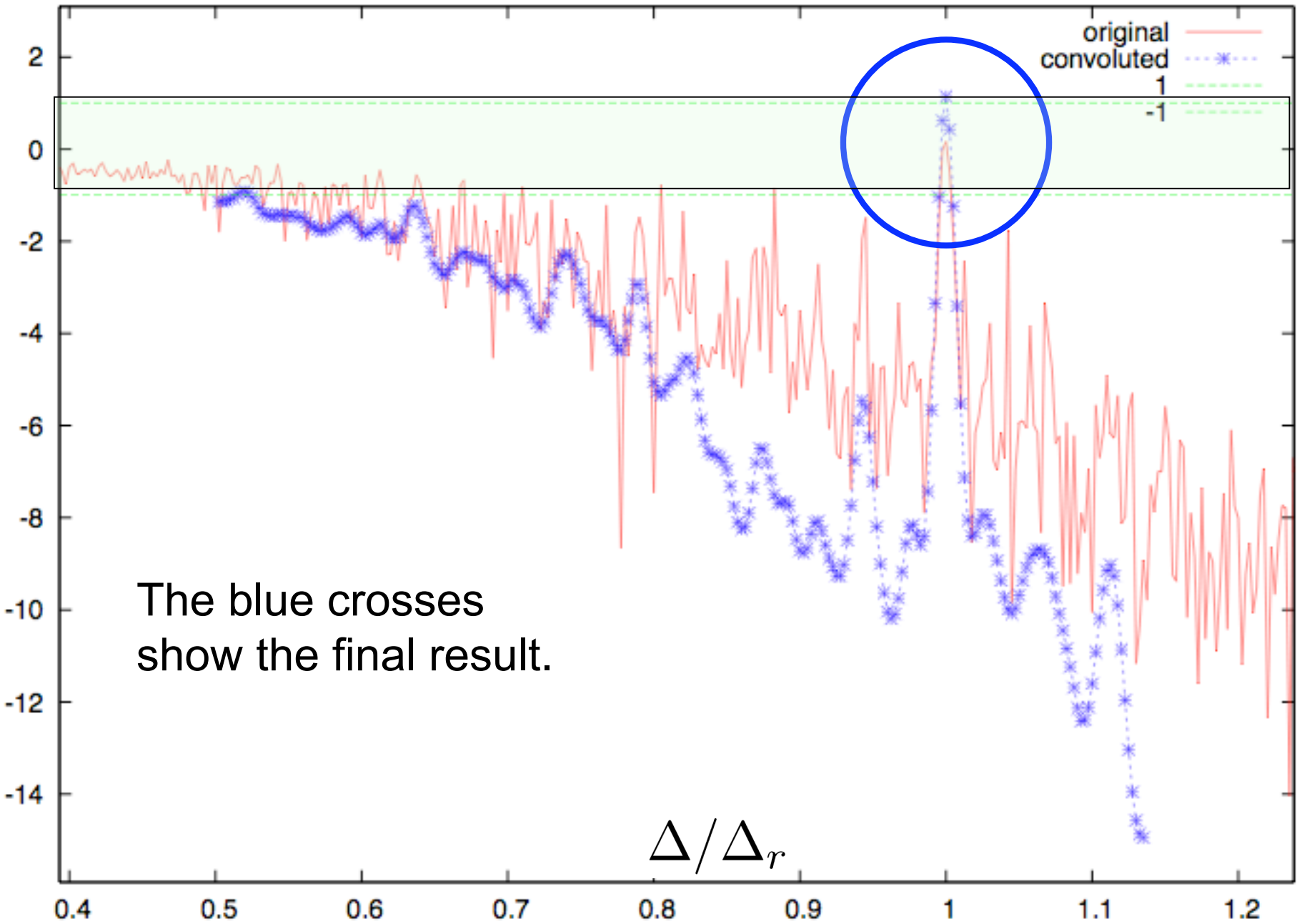
The present method works well, up to twice the noise levels where the best existing methods fail.


0.2
0.25
0.5
1
10
0
-10
-20
-30
-40
-50
-60
-70
(Df-Df_{ref}) / \sigma_{ref}
0.2
0.4
0.6
0.8
1
\Delta/\Delta_r

Now, the analysis of the original “noisy data”

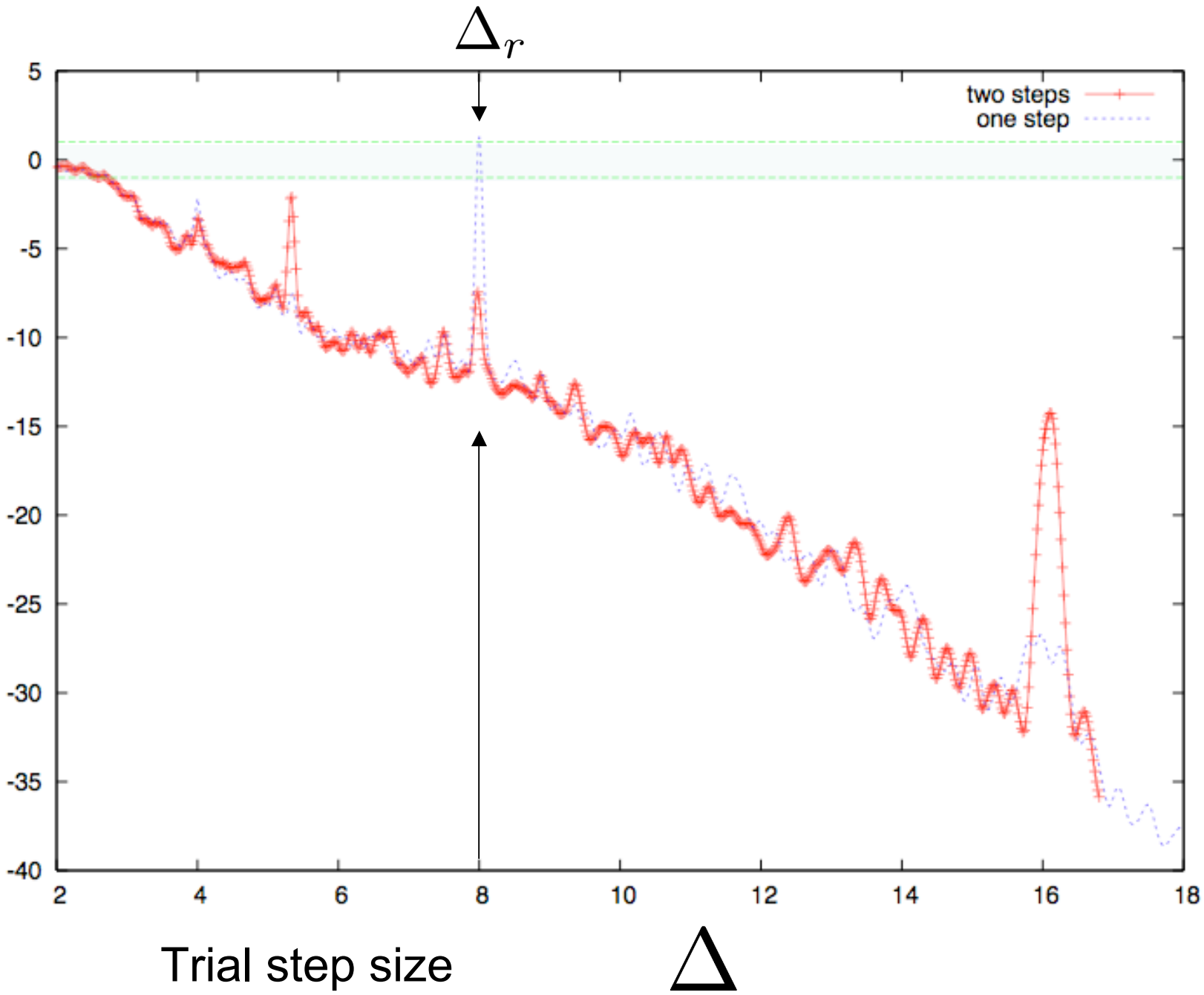


Step size/noise =1


original
convoluted
1
-1
2
0
-2
-4
-6
-8
-10
-12
-14
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
Δ/Δ_r
The blue crosses
show the final result.

Would we be able to detect if a stepping process contains **2** step sizes instead of **1**?

Example: analyse trace with sequence of step sizes:
5-11-5-11-5-11- nm (average: 8nm, as before)


Δ_r
two steps
one step
5
0
-5
-10
-15
-20
-25
-30
-35
-40
2
4
6
8
10
12
14
16
18
Trial step size
Δ

1. Back-stepping and multiple step sizes can be accounted for
2. Non-Poissonian statistics too (…or so I think)
3. Approach is not limited to analysing stepping motors (…this seems a safe bet)