$E=mc^2$

计算物理

李文飞

南京大学物理系

E-mail: wfli@nju.edu.cn

2025-05-28

计算机数据处理

时序分析、光滑滤波：

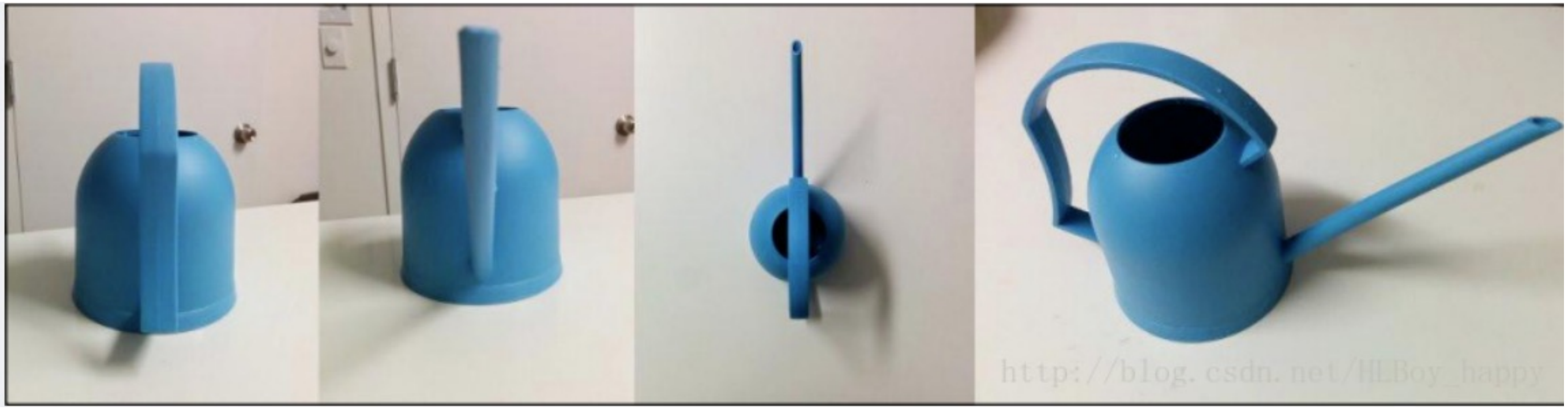
快速傅立叶变换、小波分析

数据拟合与物理建模：

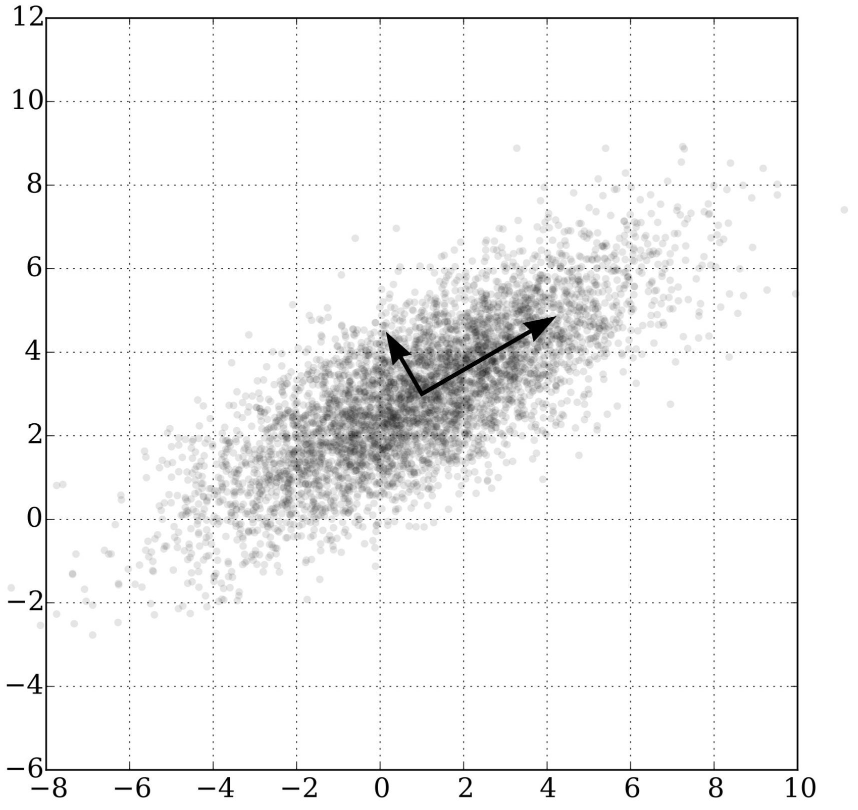
线性与非线性最小二乘拟合、最大似然估计、马尔科夫态模型

数据处理与机器学习：

误差分析、主元分析（PCA）、聚类分析、最大熵原理、神经网络


http://blog.csdn.net/HLBoy_happy

主元分析法 (PCA)



主元分析：
利用正交变换，把描述数据的相互关联的几个变量转换为线性不相关的变量，即主元（正交基矢）。

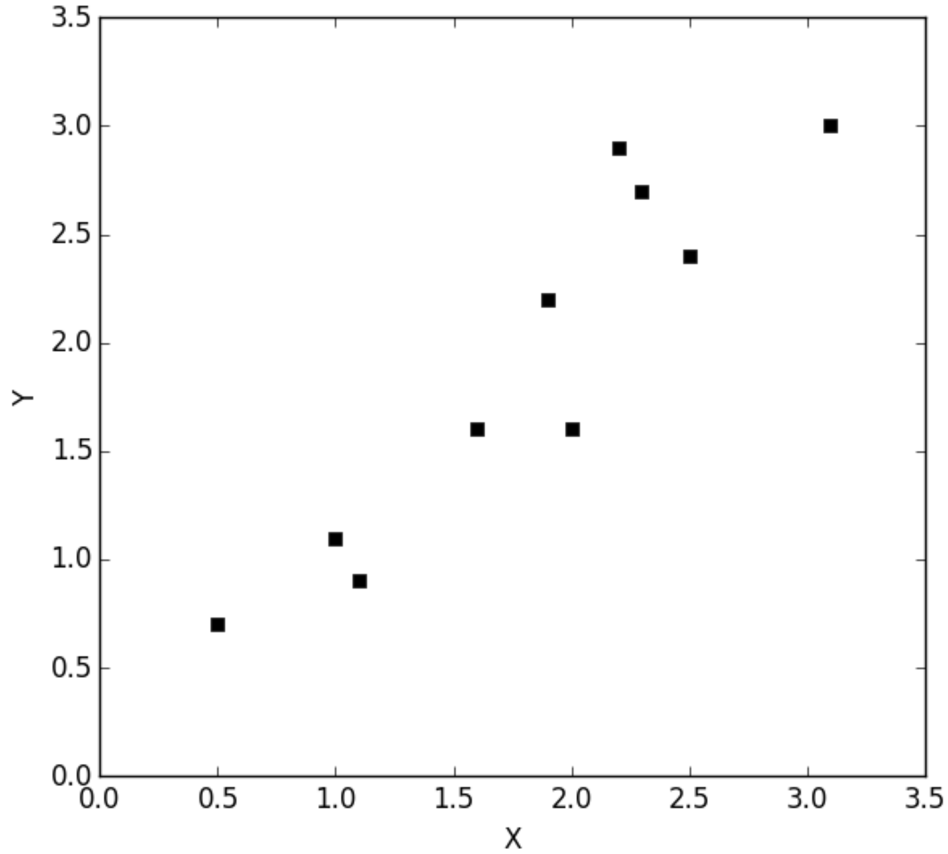
PC1：方差最大
PC2：方差次之
…

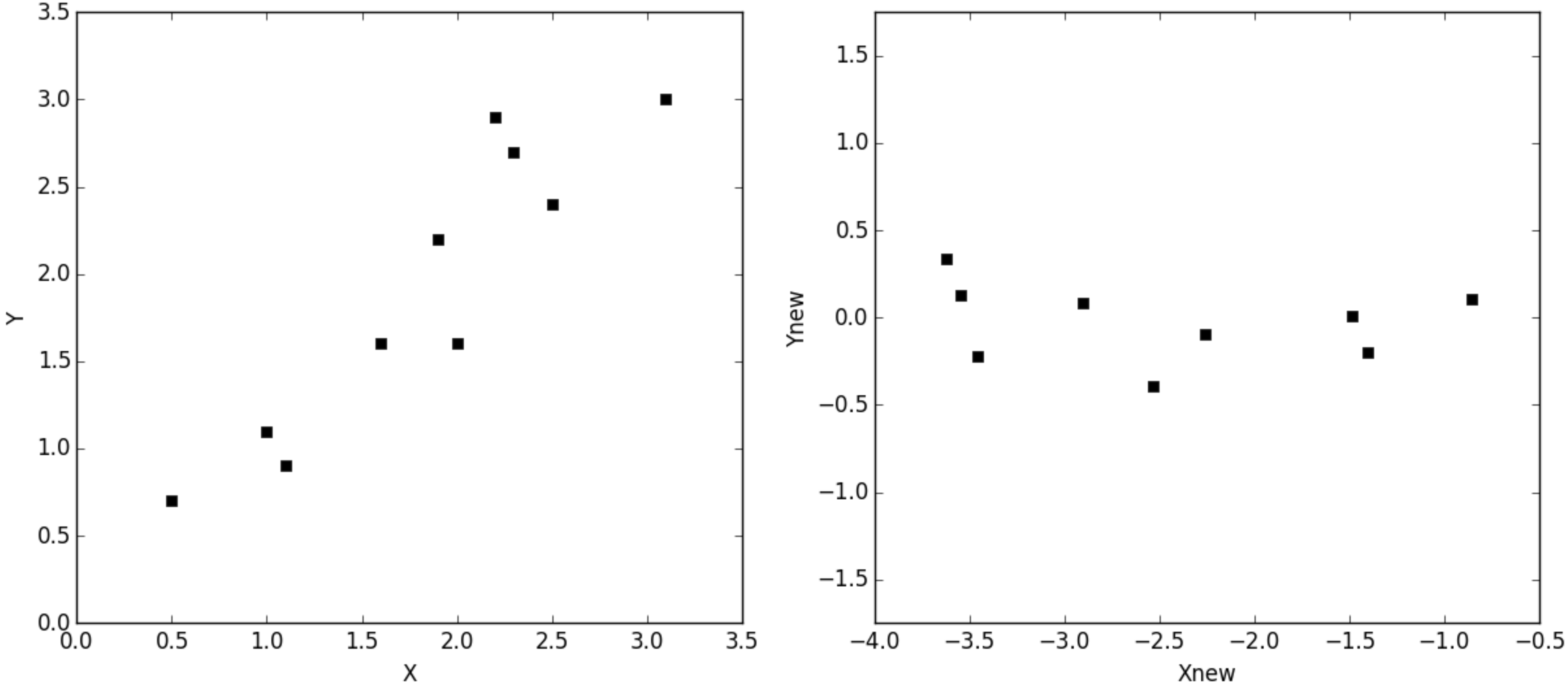
协方差矩阵：

$$X = \begin{bmatrix} \boldsymbol{X1} \\ \boldsymbol{X2} \\ \vdots \\ \boldsymbol{Xn} \end{bmatrix}$$

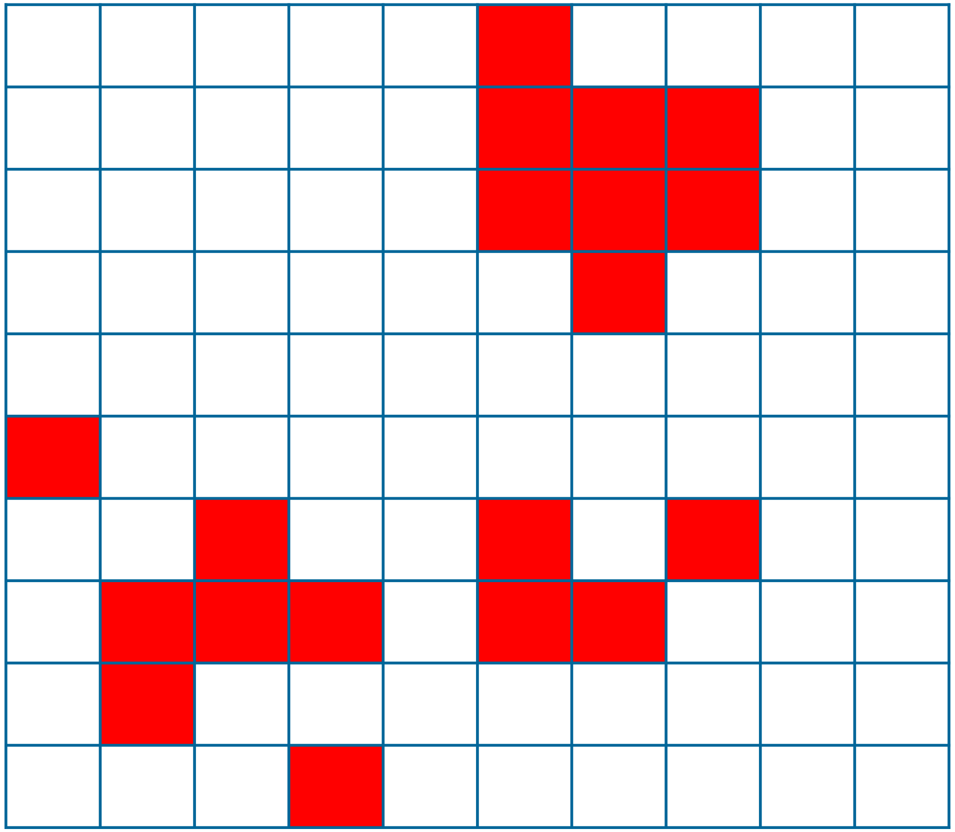
$$\boldsymbol{Q} = \mathrm{C}\boldsymbol{ov}(\boldsymbol{Xi}, \boldsymbol{Xj}) = \boldsymbol{E}((\boldsymbol{Xi} - \boldsymbol{\mu i})(\boldsymbol{Xj} - \boldsymbol{\mu j})) = \boldsymbol{E}(\boldsymbol{XiXj}) - \boldsymbol{E}(\boldsymbol{Xi})\boldsymbol{E}(\boldsymbol{Xj})$$

$$\boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{QW} = \boldsymbol{\Lambda}$$

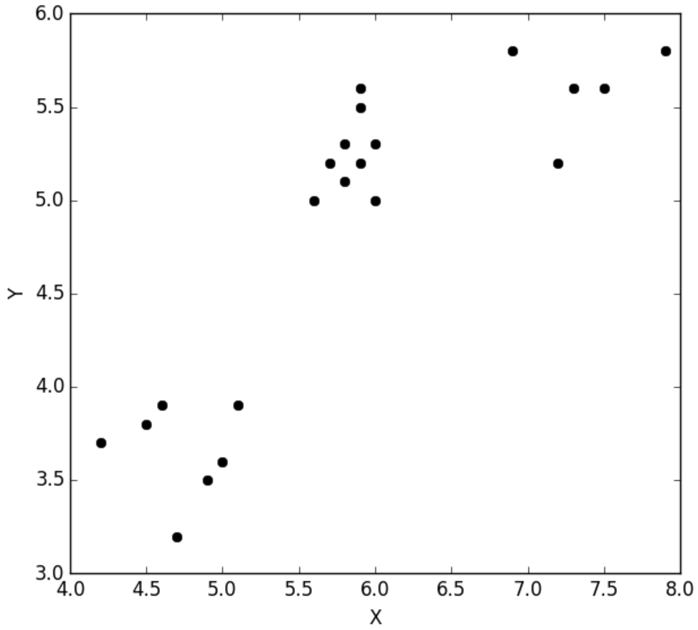

3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Y
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
X


Y
X
Ynew
Xnew
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
−4.0
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

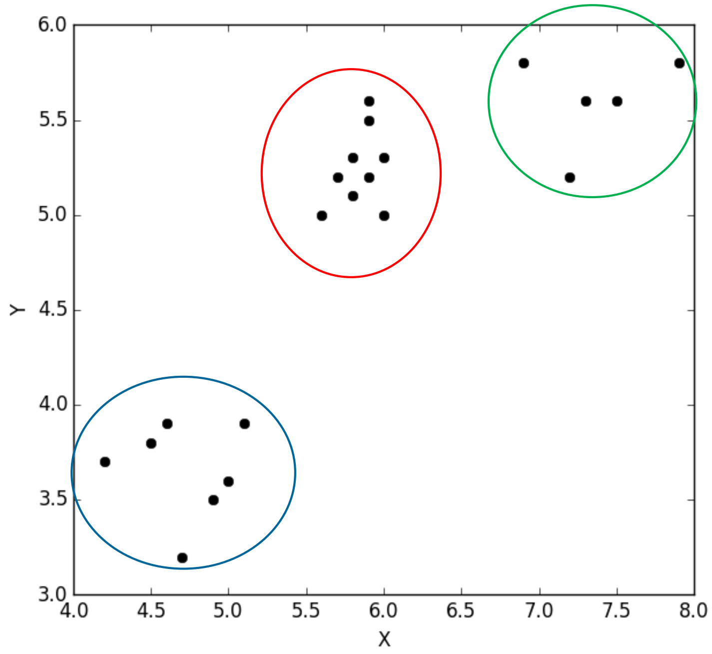
聚类分析(clustering)



k-means clustering
Hierarchical clustering
…



6.0
5.5
5.0
4.5
4.0
3.5
3.0
Y
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
X



6.0
5.5
5.0
4.5
4.0
3.5
3.0
Y
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
X

数据拟合—最小二乘法

目的：

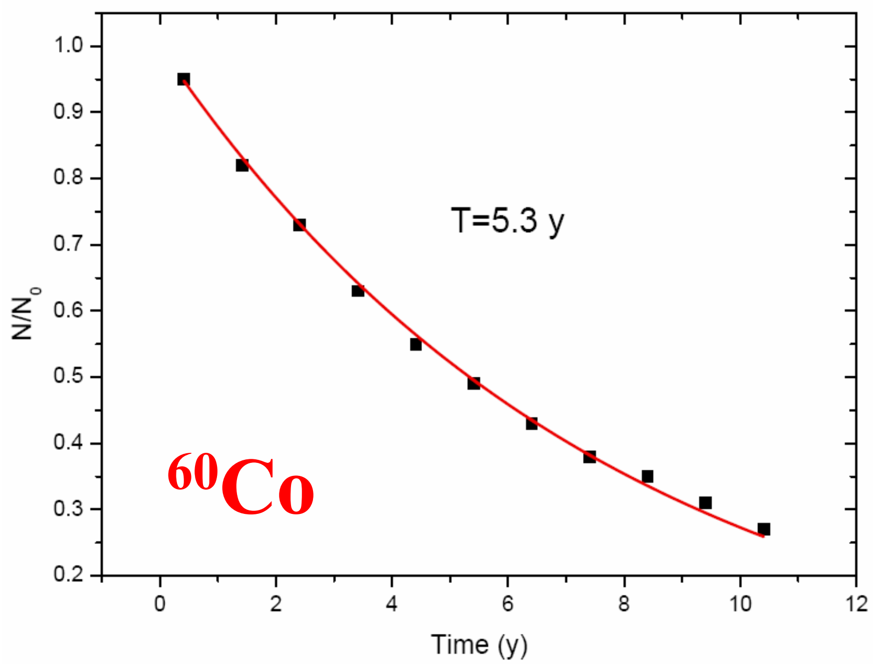
根据实验数据，寻求两个（多个）物理量之间的近似的解析函数关系，或定出函数关系中的参数。

放射性核半衰期测定

Time (y)	N/N_0
0.41	0.94
1.41	0.82
2.41	0.72
3.41	0.63
4.41	0.55
5.40	0.49
6.40	0.43
7.40	0.38
8.40	0.34
9.40	0.30
10.40	0.27

$$N/N_0 = \exp(-\ln 2 * t/T)$$

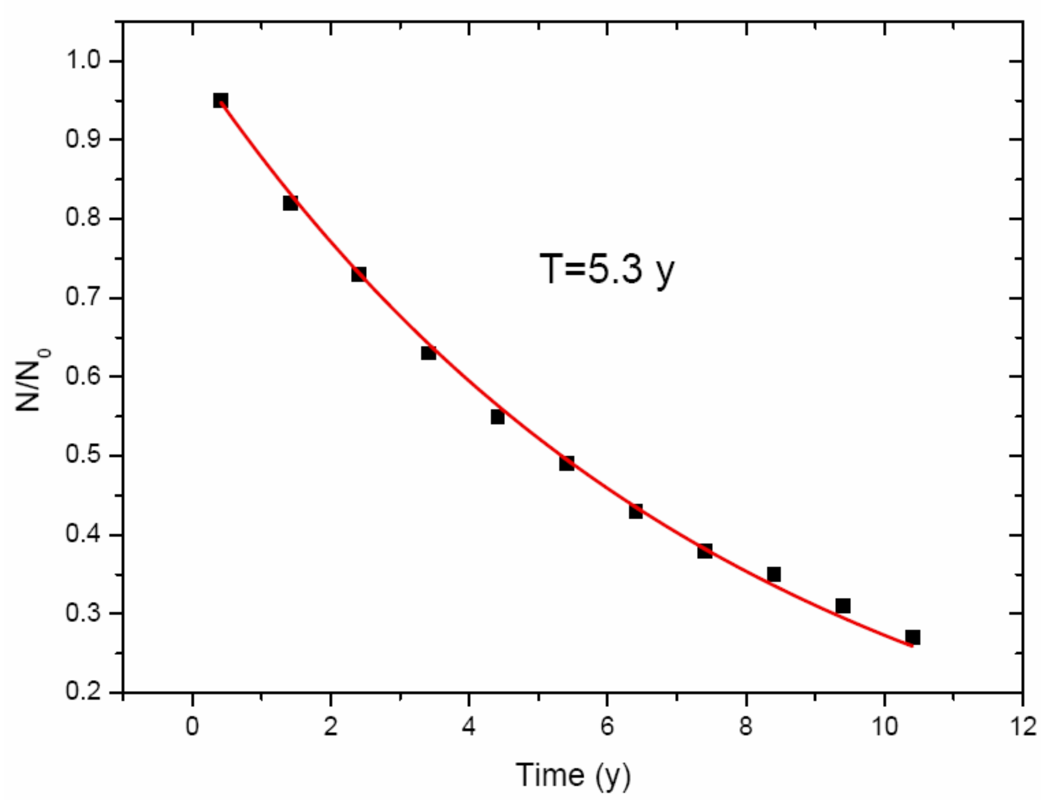
T : 半衰期



最小二乘拟合问题

问题：

如何选择参数T使得拟合曲线与数据点符合最好



最小二乘法

$$N/N_0 = \exp(-\ln 2 * t/T)$$

就是要使偏差

$$N/N_0 = exp(-ln2 * t/T) \qquad ln(N) = ln(N_0) - ln2 * t/T$$

因此可以考虑选取常数T，使得

$$M = \sum_{i=0}^{10}\left[ln(N_i) - ln(N_0) + ln2 * \frac{t_i}{T}\right]^2$$

最小来保证每个偏差的绝对值都很小.

这种根据偏差的平方和为最小的条件来选择常数的方法叫做最小二乘法.

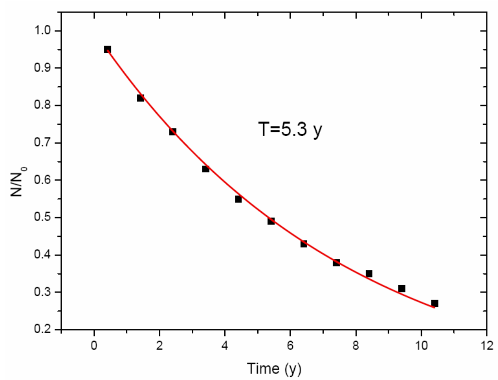
$$M=\sum_{i=0}^{10}\left[ln(N_i)-ln(N_0)+ln2*\frac{t_i}{T}\right]^2$$

把 M 看成自变量 T 的一个一元函数：极值问题

$$\frac{\partial M}{\partial T}=2\sum_{i=0}^{10}\left[ln(N_i)-ln(N_0)+ln2*\frac{t_i}{T}\right]\ \left(-\frac{ln2*t_i}{T^2}\right)=0$$

$$\sum_{i=0}^{10}[ln(N_i)-ln(N_0)]*t_i+\sum_{i=0}^{10}ln2*\frac{t_i^2}{T}\ =0$$

得：T=5.3y



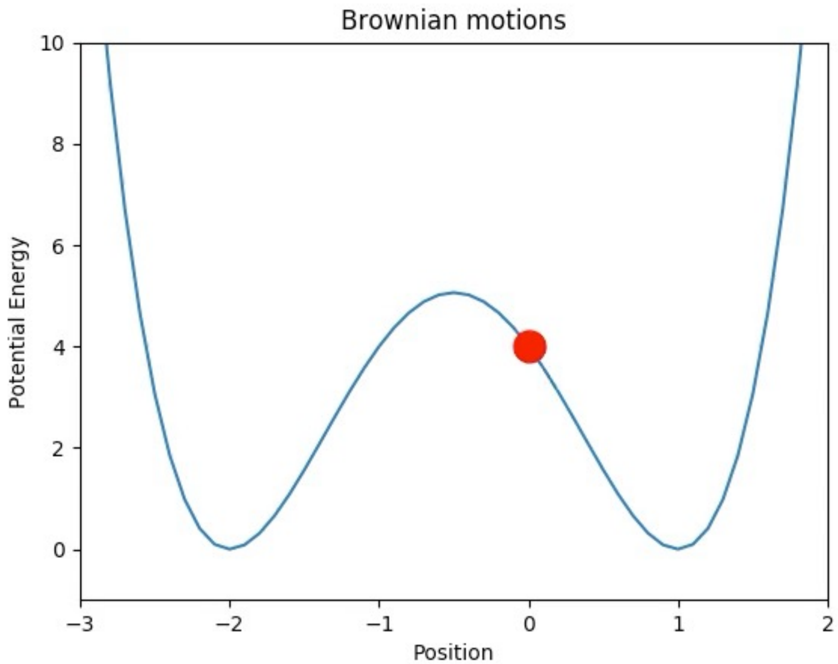
$$N/N_0=exp\left(-ln2*\frac{t}{5.3}\right)$$

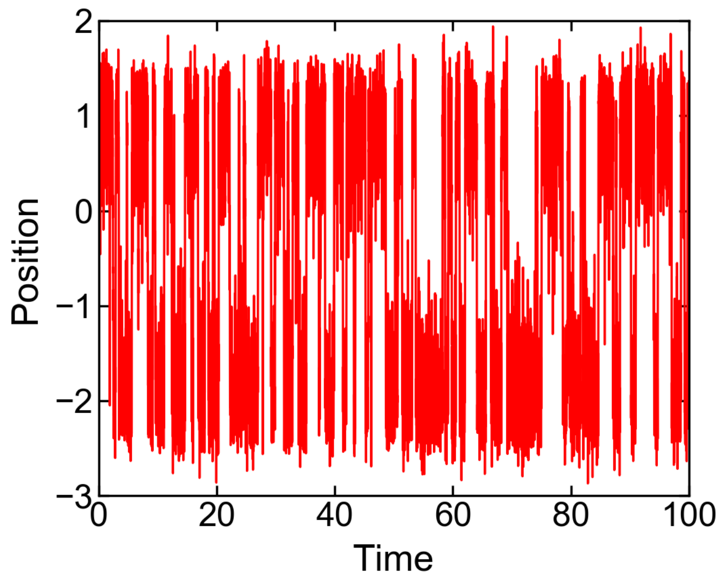
最大熵原理

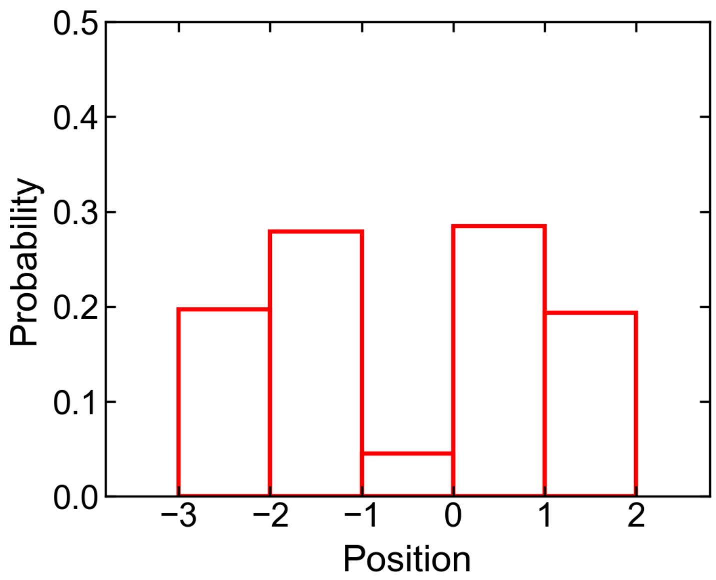
在计算模型中引入实验数据约束

Langevin Equation: Simulating Brownian motions

$$V(x) = (x-1)^2 (x+2)^2$$

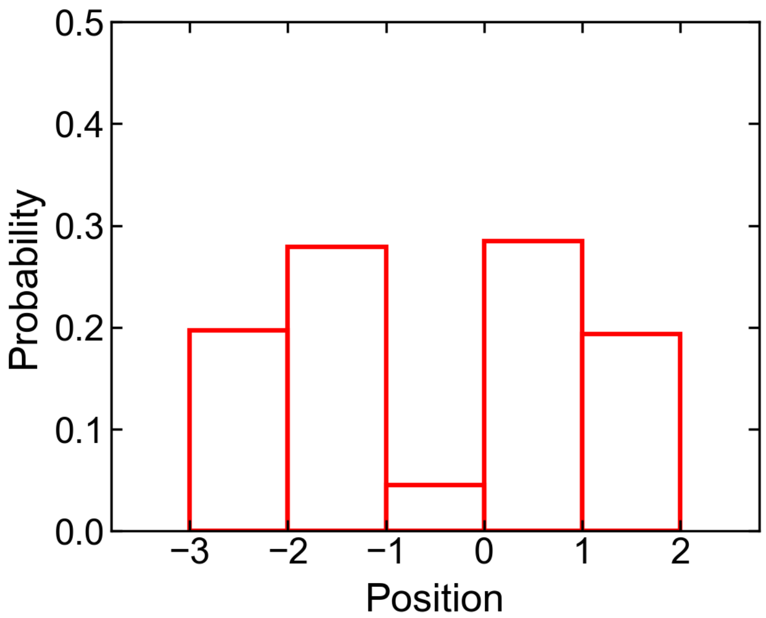




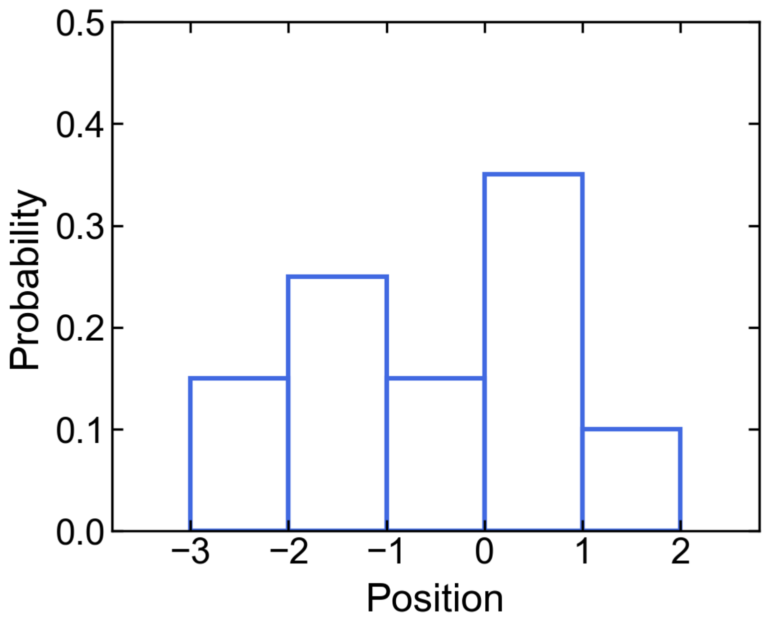


实验数据

Simulation



Experiment



如果实验数据测得的x的分布如右图，如何基于最大熵原理将实验数据约束加入分子模拟中？

最大熵原则（MEP）

$p(x_i)$：Probability distribution.

实验观测值： $<f>= \sum_i^n f(x_i)p(x_i)$

Shannon Entropy: $S(p) = -\sum_i^n p(x_i)\ln p(x_i)$

最大熵原理：
有多种机率分布$p(x_i)$满足实验观测,应该选取熵最大的$p(x_i)$。

约束条件：

$$\sum_i^n f(x_i)p(x_i)=f_{exp}$$

实验数据

$$\sum_i^n p(x_i)=1$$

归一化条件

满足以上约束条件下，-S(p)取极小值，可得：

$$p(x_i) = Z^{-1}\exp[-\beta(E_0(x_i) + \sum_j \alpha_i f_i(x)]$$

$$Z = \sum_i^n \exp[-\beta(E_0(x_i) + \sum_j \alpha_i f_i(x)]$$

正则系综

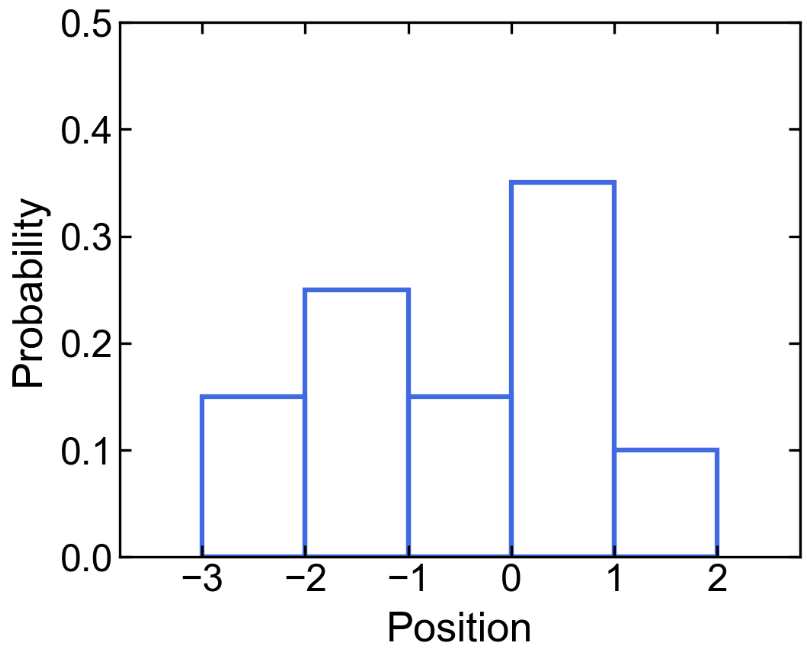
$$p(x_i) = Z^{-1}\exp[-\beta(E_0(x_i)]$$

$$Z = \sum_i^n \exp[-\beta(E_0(x_i)]$$

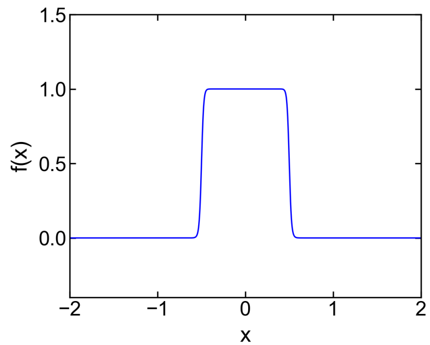
$$E(x) = \underbrace{E_0(x)}_{\text{物理势能函数}} + E_{exp}(x) = E_0(x) + \underbrace{\sum_j \alpha_i f_i(x)}_{\text{实验数据约束}}$$

当实验测量和理论计算不符时，可根据实验测量值对理论加约束修正。

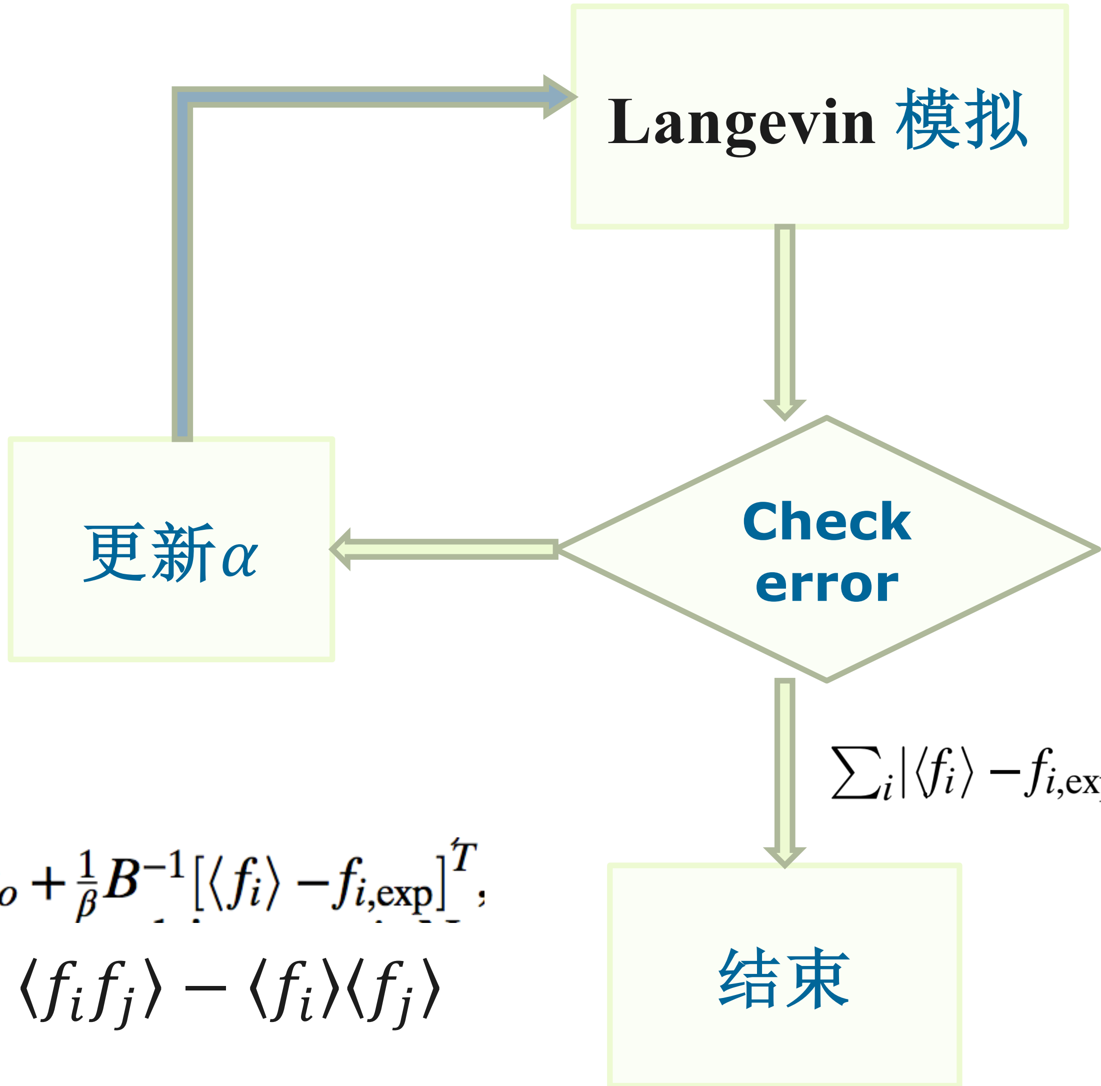
Experiment



$$f_i(x) = \frac{1}{1 + ((x - x_i)/c)^n}$$



拉格朗日乘子α的迭代优化



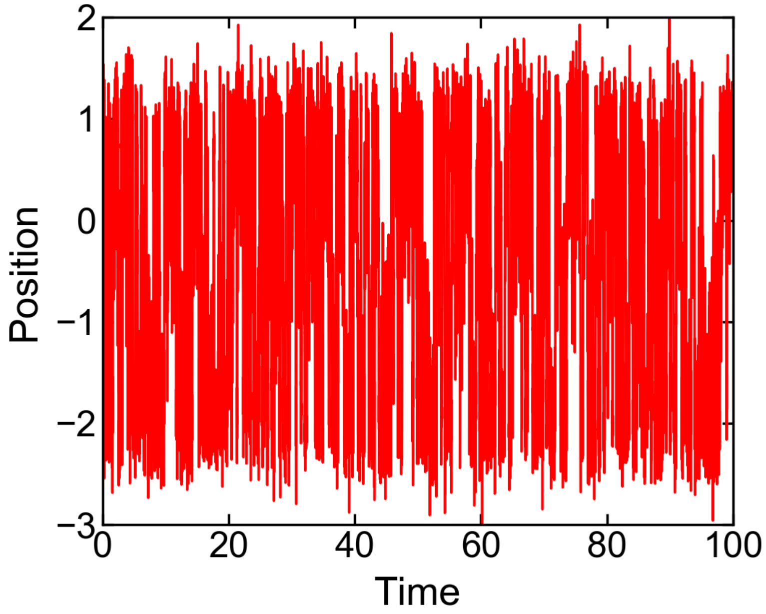
$$V_{ME}(x) = V(x) + \sum_i \alpha_i f_i(x)$$

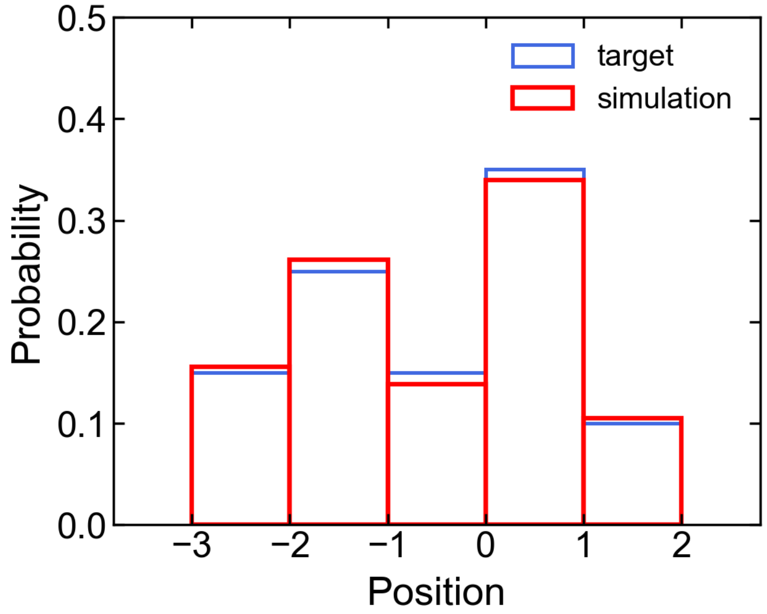
$$\sum_i |\langle f_i \rangle - f_{i,\exp}| / \sum_i f_{i,\exp} < \textbf{tolerance}$$

$$\alpha = \alpha_o + \frac{1}{\beta} B^{-1} [\langle f_i \rangle - f_{i,\exp}]^T;$$

$$B = \langle f_i f_j \rangle - \langle f_i \rangle \langle f_j \rangle$$

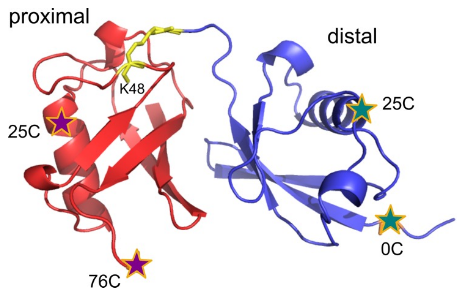
模拟结果

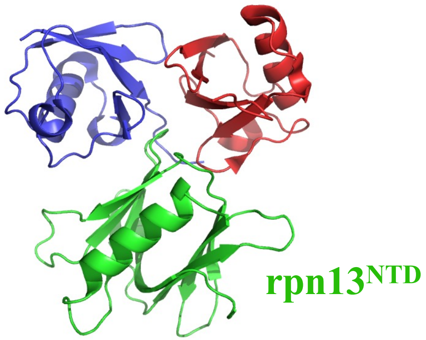

Position
Time
2
1
0
−1
−2
−3
0
20
40
60
80
100

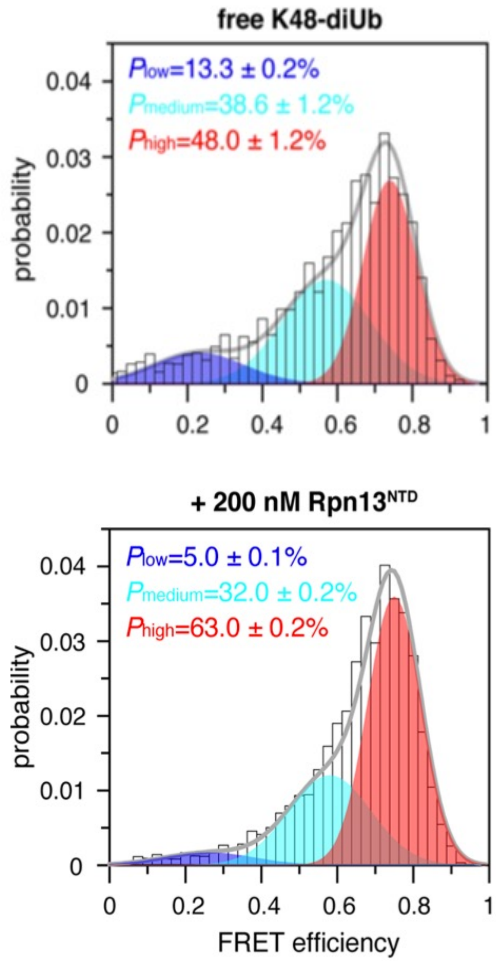

target
simulation
Probability
Position
0.5
0.4
0.3
0.2
0.1
0.0
−3
−2
−1
0
1
2

Coarse-grained simulations with smFRET.

di-ubiquitin dynamics

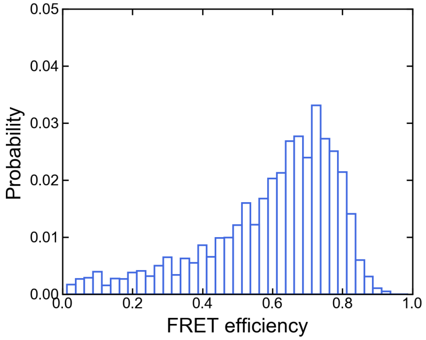


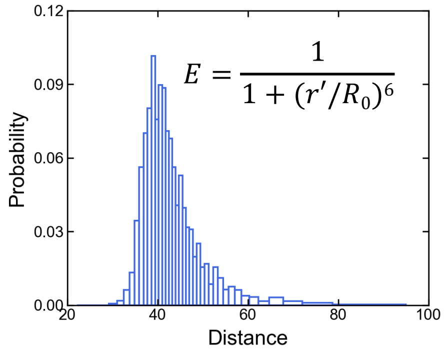


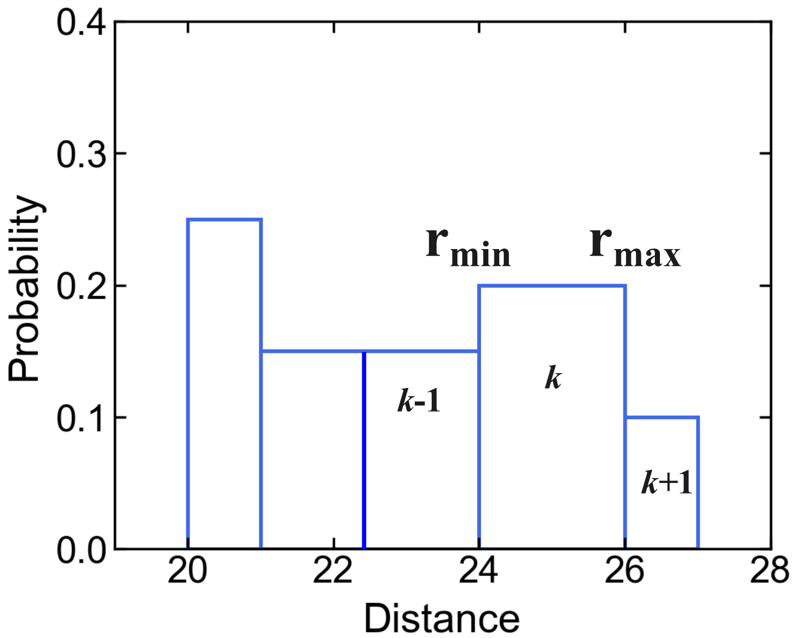


Liu et al. Cell Discovery (2019)5:19

Coarse-grained simulations with smFRET (最大熵)







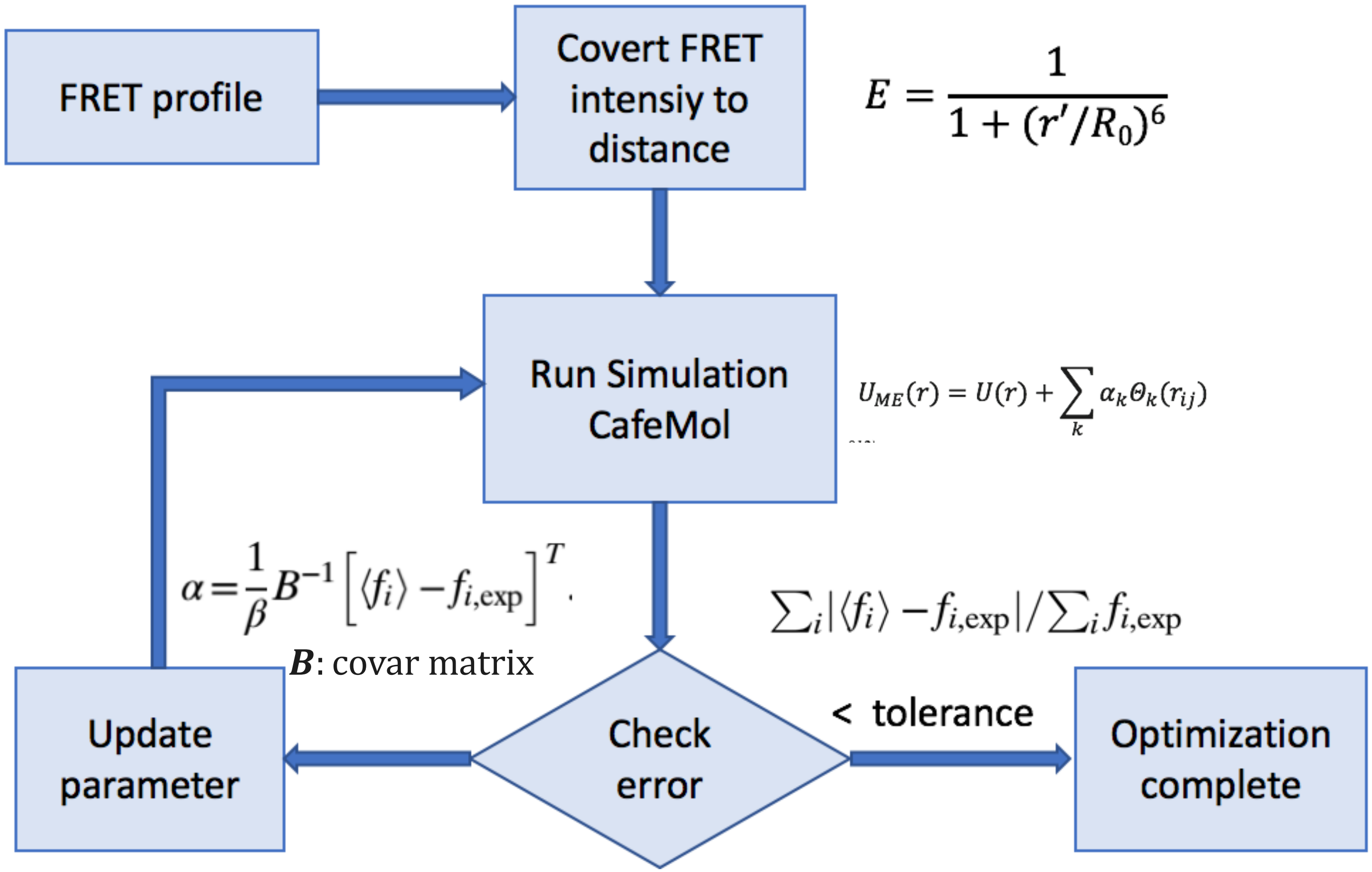
$$\Theta_k(r_{ij}) = \frac{1}{4}\{1 + \tanh[\eta(r_{ij} - r_{\min})]\}\{1 + \tanh[\eta(r_{\max} - r_{ij})]\}$$

最大熵模型： $U_{ME}(r) = U(r) + \sum_k \alpha_k \Theta_k(r_{ij})$

Pirera JW et al., JCTC 8:3445(2012)
Latham AP et al., JPCB 123:1026(2019)

Bin Wen et al., unpublished (preliminary)

Coarse-grained simulations with smFRET (最大熵)

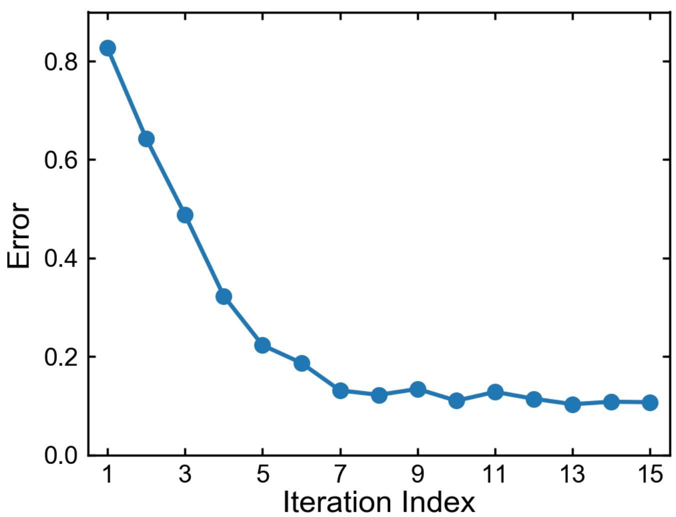


MEP: Zhang B et al, PNAS, 112:6062(2015)

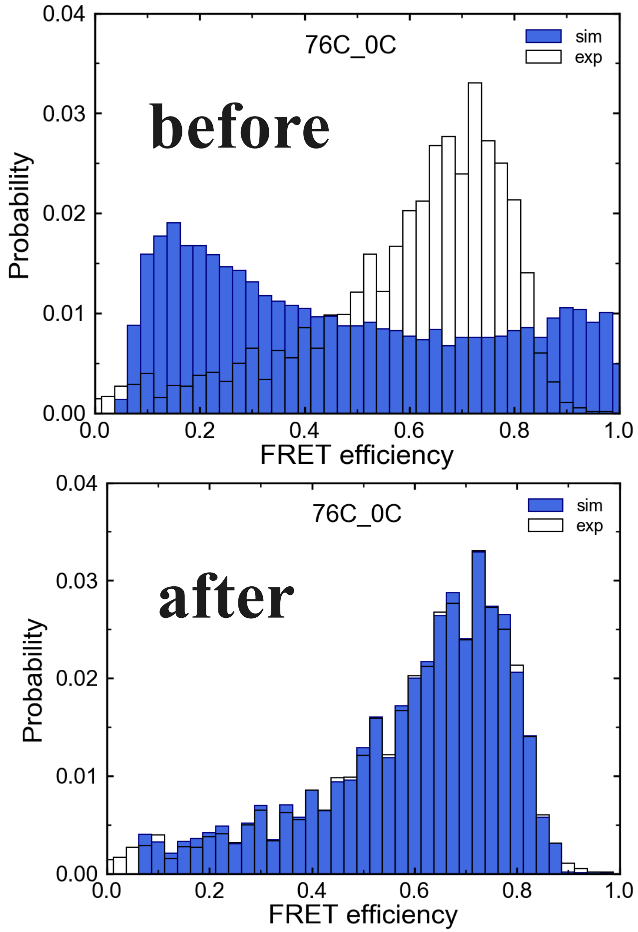
Bin Wen et al., unpublished (preliminary)

optimization process

Coarse-grained simulations with smFRET (最大熵)

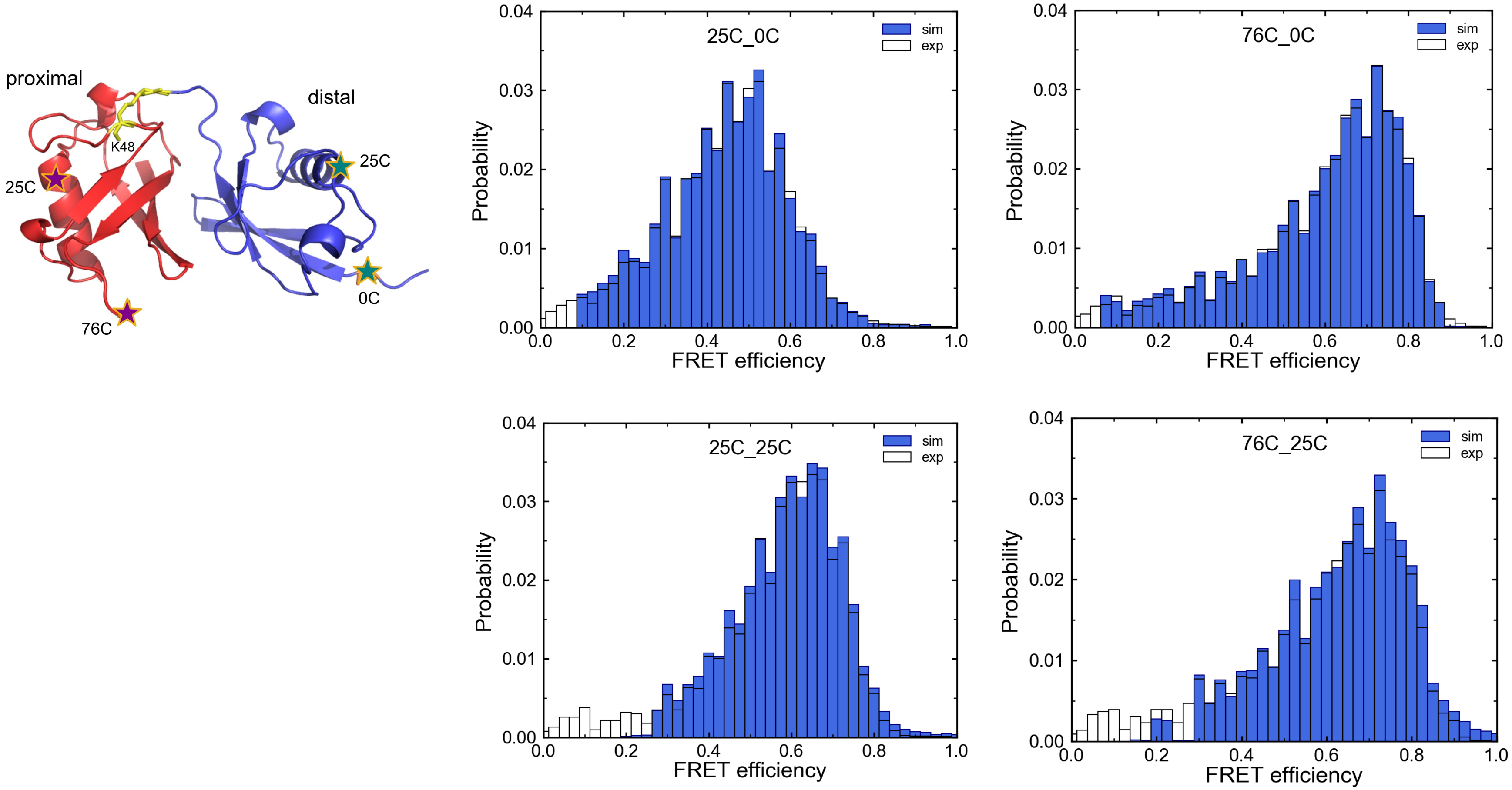


$$\sum_i |\langle f_i \rangle - f_{i,\text{exp}}| / \sum_i f_{i,\text{exp}}$$



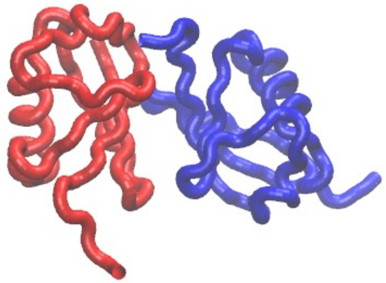
Bin Wen et al., unpublished (preliminary)

Coarse-grained simulations with smFRET (最大熵)

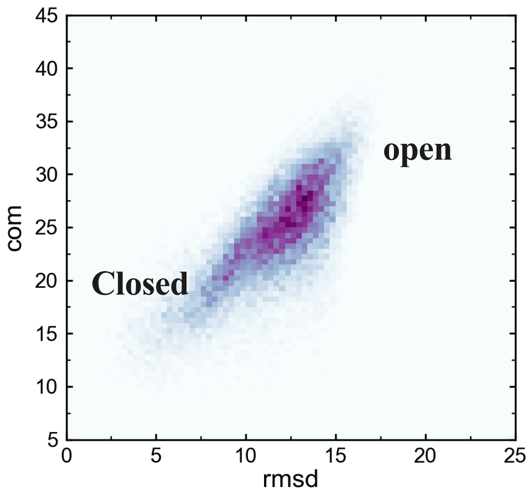


Bin Wen et al., unpublished (preliminary)

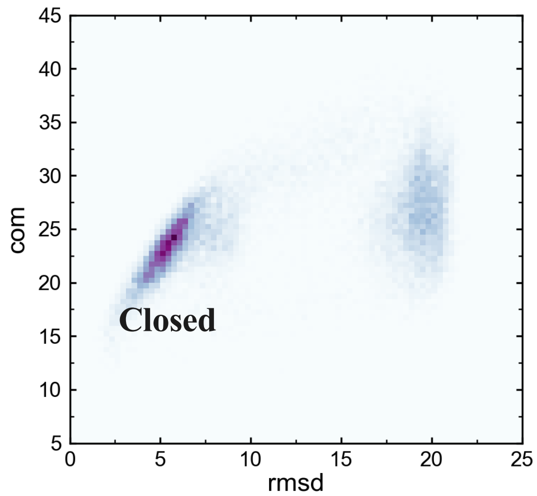
Coarse-grained simulations with smFRET (最大熵)



Conformational selection!



free di-Ub



di-Ub bound with Rpn13$_{NTD}$

Bin Wen et al., unpublished (preliminary)