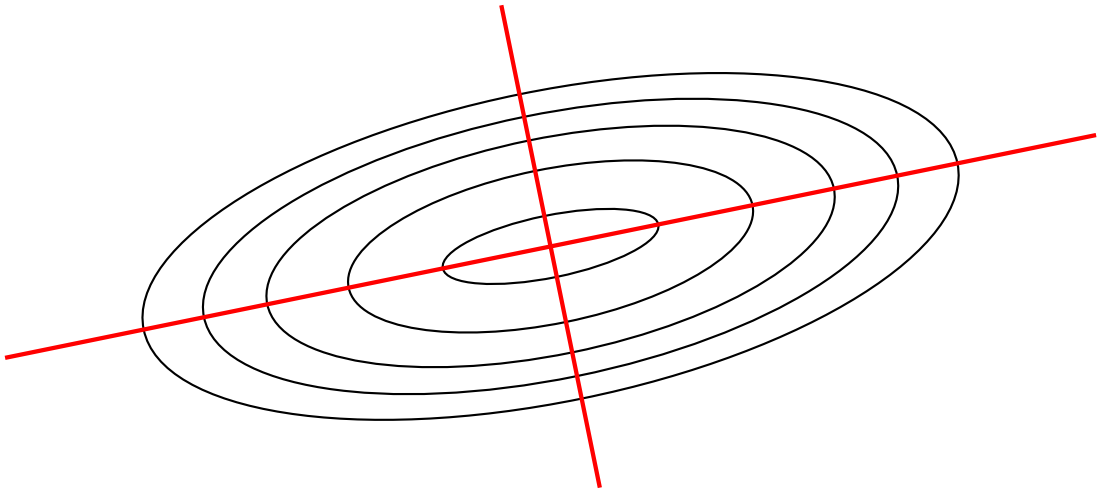
Normal mode analysis: Applications

Konrad Hinsen
Centre de Biophysique Moléculaire (CNRS)
Orléans, France

7 July, 2000

Normal modes

Harmonic potential



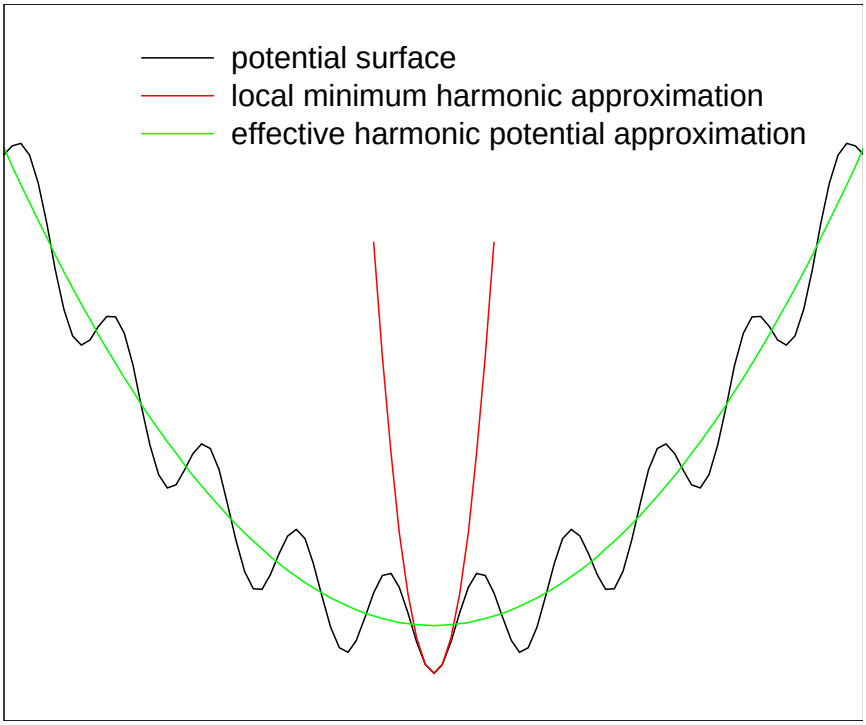
Normal modes = principal axes (mass-weighted)

Two classes of applications:

- Separation of frequencies
- Simple analytic description of the potential

Major limitation: harmonic potential

Harmonic approximations



Local minimum:

- derived from standard all-atom potentials by energy minimization
- vibrational motion
- time scale: $<$ residence time in a minimum
- appropriate for studying fast motions

Effective potential well:

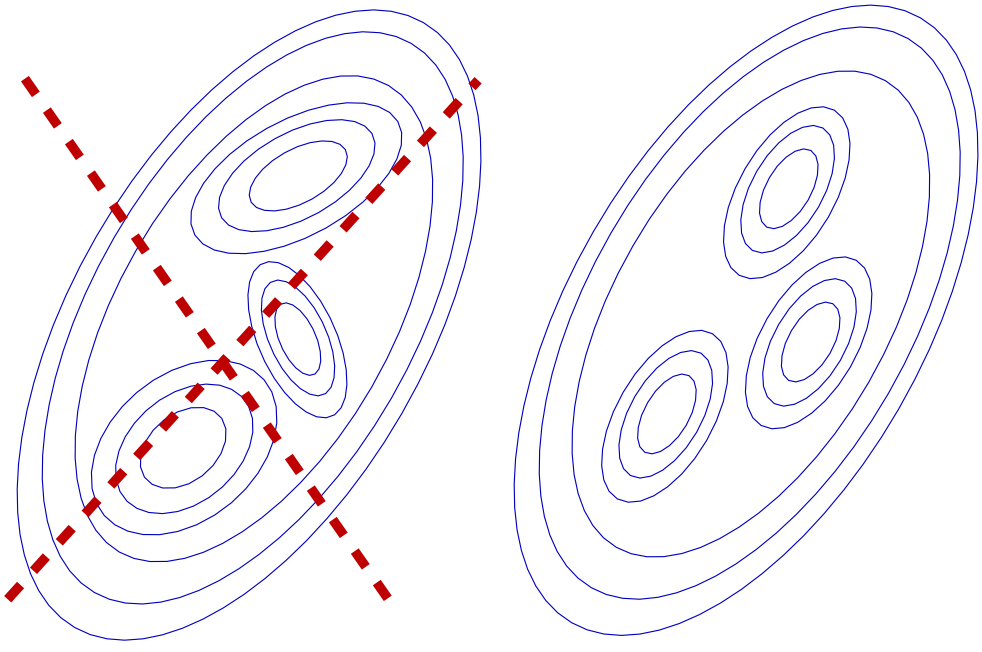
- derived from coarse-grained model
- diffusive motion (jumps between local minima)
- time scale: $>>$ residence time in a minimum
- appropriate for studying slow motions

Harmonic approximations

Standard normal mode analysis:

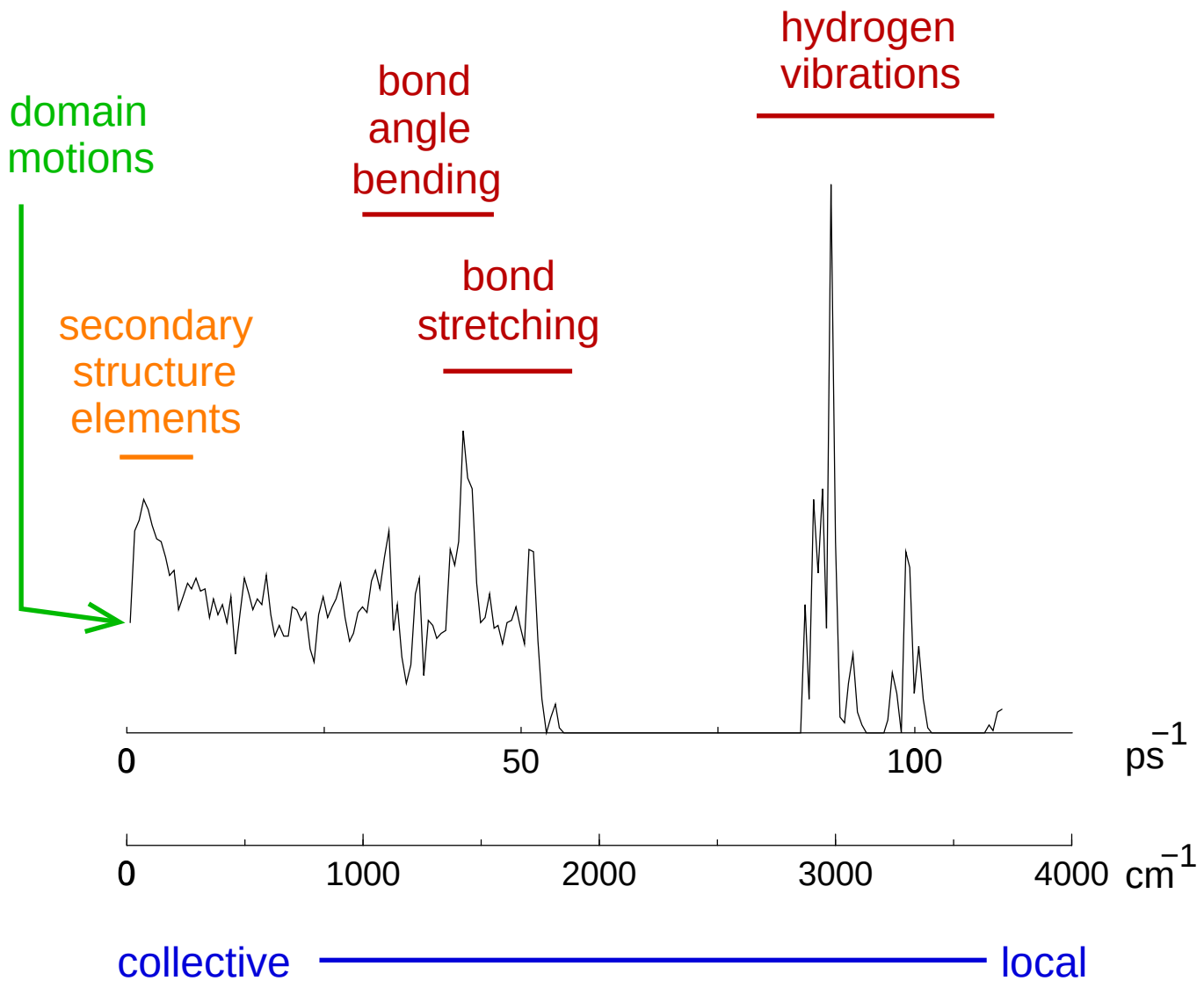
- local minimum
- correctly describes fast vibrations
  ($\longrightarrow$ infrared spectroscopy, Raman, ...)
- correctly identifies the *directions* of slow motions
  ($\longrightarrow$ domain motions, conformational changes)
- overestimates *time scales* and underestimates *amplitudes* of slow motions

Conclusion: the shapes of the local minima and the effective potential well must be similar.



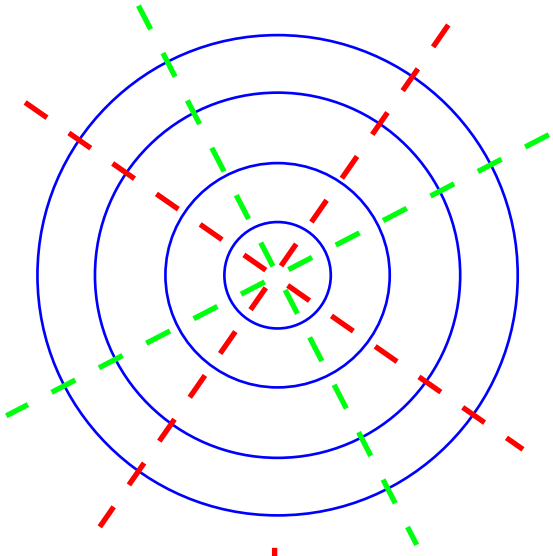
Separating frequencies

Frequency spectrum of a protein:

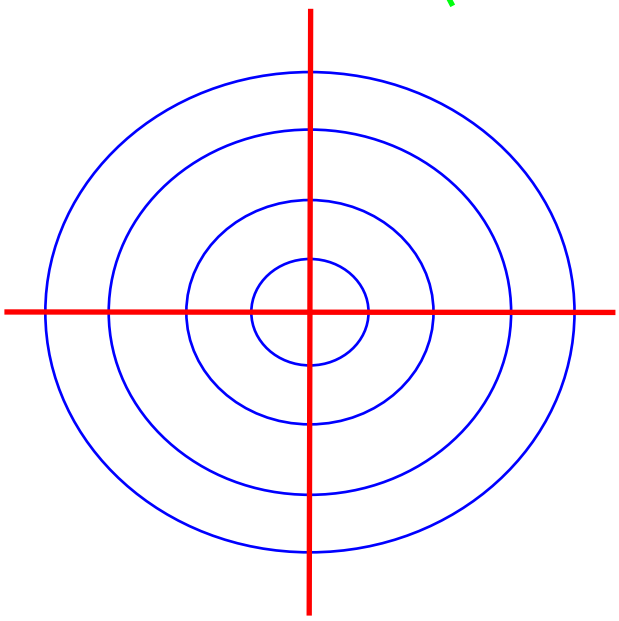


- Most modes describe internal motions of individual residues.
- Only very few modes (about 2%) describe

Separating frequencies



Degenerate modes:
frequencies are equal,
directions are arbitrary



Almost degenerate modes:
frequency difference
is not significant, but
directions are unique

- Always analyze groups of modes with similar frequencies together.
- Do not analyze the differences between modes that are almost degenerate.

Fluctuation amplitudes

Kinetic energy in each mode:

$$\frac{1}{2}\left\langle \dot{x}^2 \right\rangle = \frac{1}{2} k_B T$$

In an harmonic potential:
average potential energy = average kinetic energy

$$\frac{1}{2}\omega^2 \left\langle x^2 \right\rangle = \frac{1}{2}\left\langle \dot{x}^2 \right\rangle$$

It follows:

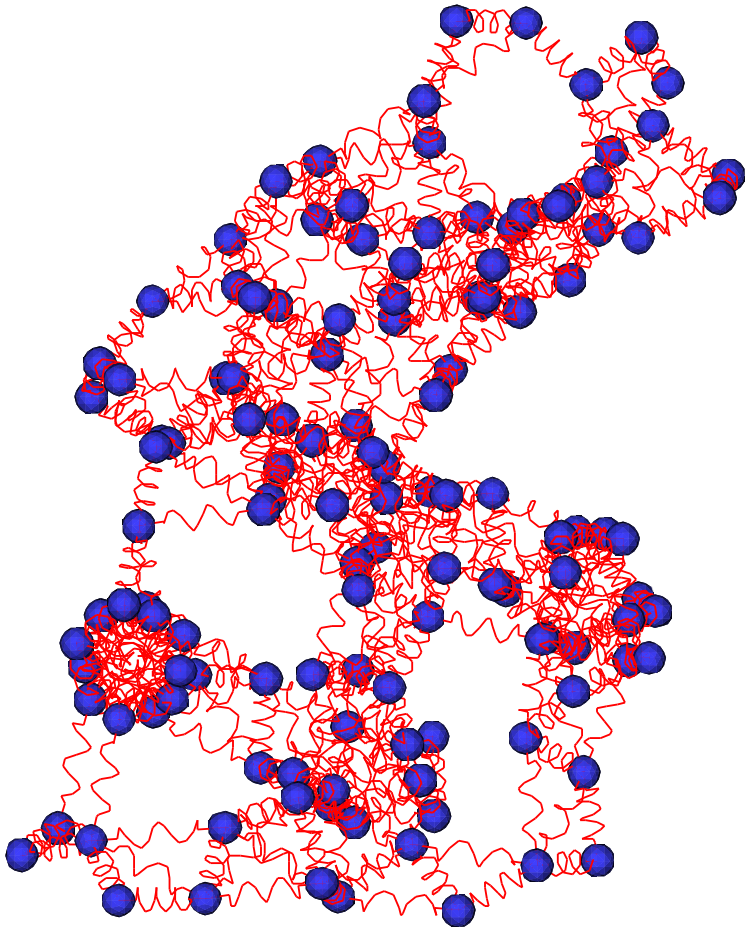
$$\left\langle x^2 \right\rangle = \frac{k_B T}{\omega^2}$$

Slow modes have large amplitudes.

Note: Normal mode analysis in a *local* minimum underestimates the fluctuation amplitudes for slow modes.

Collective motions

- Specific to a protein
- Usually related to its function
- Largest amplitudes
- Can be obtained from simplified models



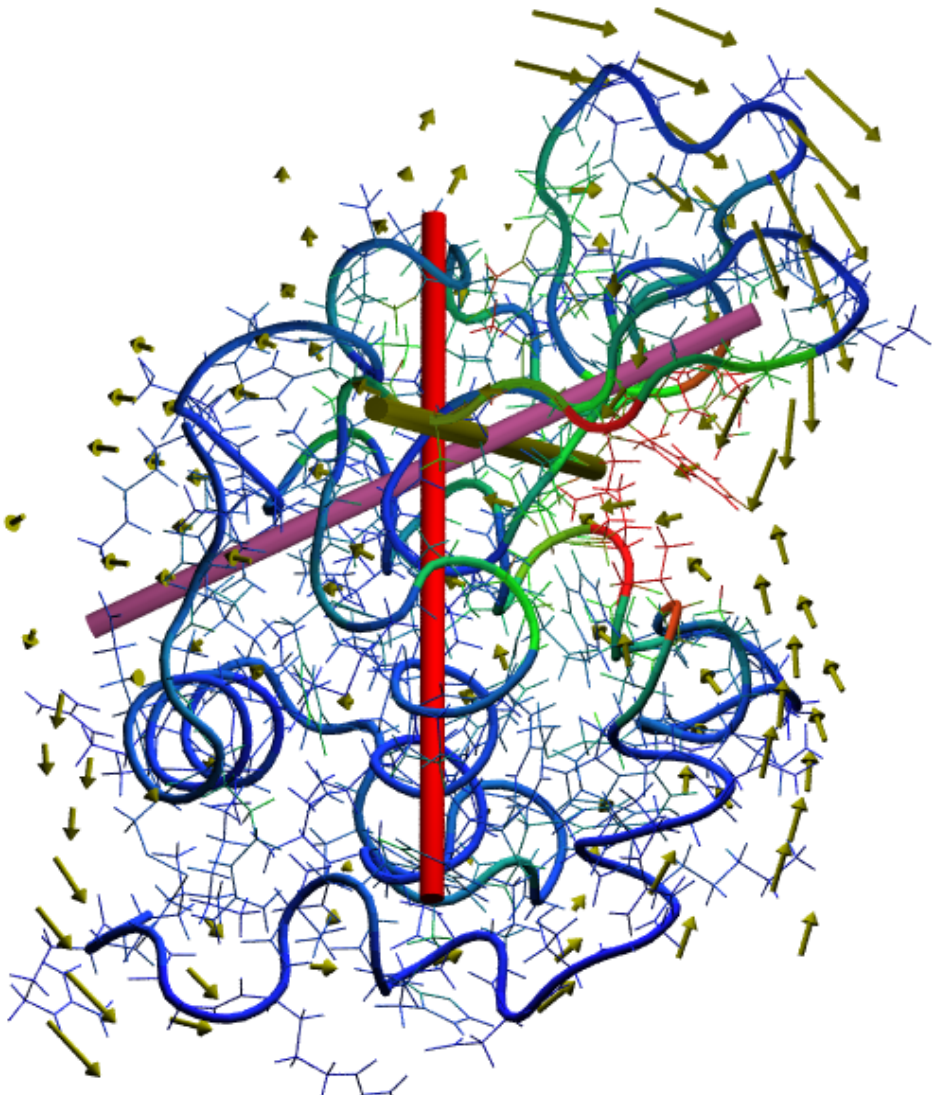
- Residues are point masses
- Springs between nearby points

Domain motions

Domains: quasi-rigid regions in a protein, separated by more flexible inter-domain regions.

Domain motions are

- a useful concept to describe slow dynamics
- not unique – several definitions exist
- an approximation – nothing is really rigid in a protein



Domain motions

1. Identify the most rigid parts.
   Local deformation energy for atom i:

$$E_i = \frac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^{N} k(|\mathbf{R}_i - \mathbf{R}_j|) \times$$

$$\times \left[|\mathbf{R}_i - \mathbf{R}_j + \mathbf{d}_i - \mathbf{d}_j| - |\mathbf{R}_i - \mathbf{R}_j| \right]^2$$

   with

$$k(r) = c \cdot \exp\left(-\frac{r^2}{r_0^2}\right)$$

   Keep atoms for which $E_i < E_{\text{limit}}$ in *all* selected modes.

2. Determine the rigid-body motion in small subregions of the rigid parts for each mode:

$$\mathbf{d}_i = \mathbf{T}_j + \mathbf{D}(\mathbf{n}_j, \phi_j) \cdot \mathbf{R}_i.$$

   $\mathbf{T}_j$: rigid-body translation of subregion j
   $\mathbf{D}(\mathbf{n}, \phi)$: rotation matrix
   $\mathbf{n}_j, \phi_j$: rigid-body rotation (axis and angle) of subregion j
   In total: 6 parameters *per mode*.

Domain motions

3. Group the subregions into clusters of similar rigid-body motion parameters in *all* selected modes.

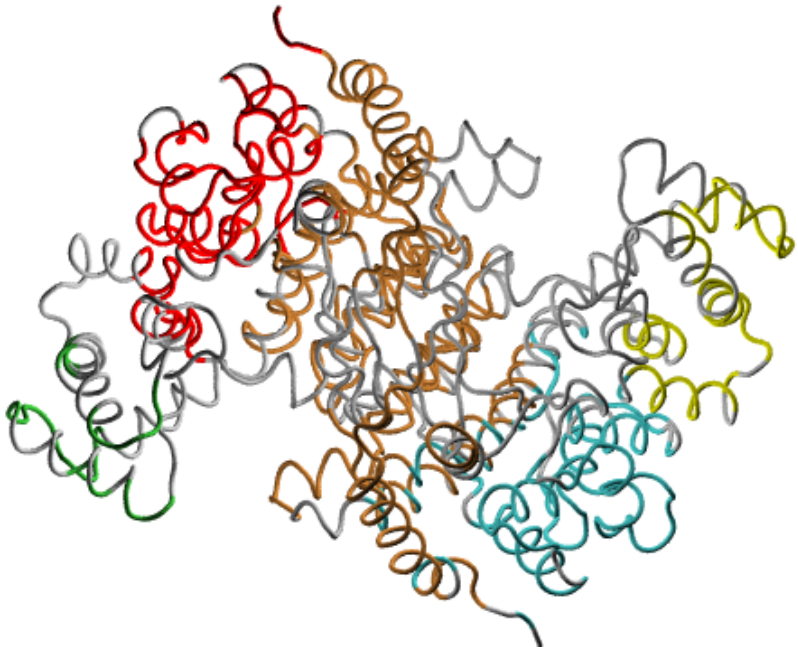
   Similarity measure (empirical):

$$S_{ij} = 3\frac{|\phi_i \mathbf{n}_i + \phi_j \mathbf{n}_j|}{|\phi_i \mathbf{n}_i - \phi_j \mathbf{n}_j|} + \frac{|\mathbf{T}_i + \mathbf{T}_j|}{|\mathbf{T}_i - \mathbf{T}_j|}$$

   Similarity threshold $S_{\min}$ determines cluster size

Result:

- Flexible regions
- Domains
- Semi-flexible inter-domain regions



Monte-Carlo integration

Equilibrium average of a quantity $A(\mathbf{r})$:

$$\langle A \rangle = \frac{1}{Z} \int \cdots \int d^{3N} r A(\mathbf{r}) \exp\left[-\frac{\mathbf{r} \cdot \mathbf{K} \cdot \mathbf{r}}{2 k_B T} \right]$$

Normal mode decomposition:

$$\mathbf{K} = \sum_{i=1}^{3N} \omega_i^2 \mathbf{u}_i \mathbf{u}_i$$

Yields:

$$\langle A \rangle = \frac{1}{Z} \int \cdots \int d^{3N} x A(\sum_i x_i \mathbf{u}_i) \prod_j \exp\left[-\frac{\omega_j^2 x_j^2}{2 k_B T} \right]$$

The normal mode decomposition permits

- sometimes the analytical evaluation
- easy Monte-Carlo integration using a standard Gaussian random number generator
- improving efficiency by eliminating fast small-amplitude modes

Other applications

Prediction of large-scale motions:

- in minimization algorithms
- in sampling algorithms (scaled-variable Monte-Carlo)
- for verifying sampling quality

Modelling atomic fluctuations in structure refinement

Analysing changes in higher-frequency motions due to different environments (analysis of spectroscopic data)

Slow dynamics of proteins using Brownian Modes

System-specific analysis: fluctuations, correlations, ...

Normal modes in comparison

Main advantages:

- no sampling problem
- computational efficiency, especially for coarse-grained models
- simplicity in application

Main limitations:

- single-well potentials, thus no possibility to study conformational changes
- vibrational normal modes do not describe the diffusive motion that dominates slow dynamics ($\rightarrow$ Brownian modes)