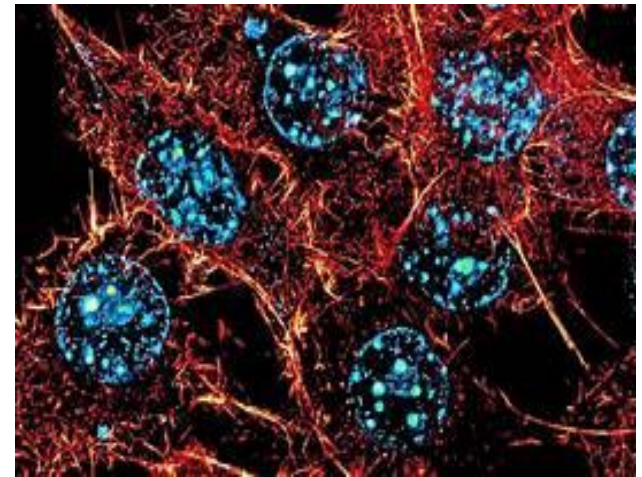
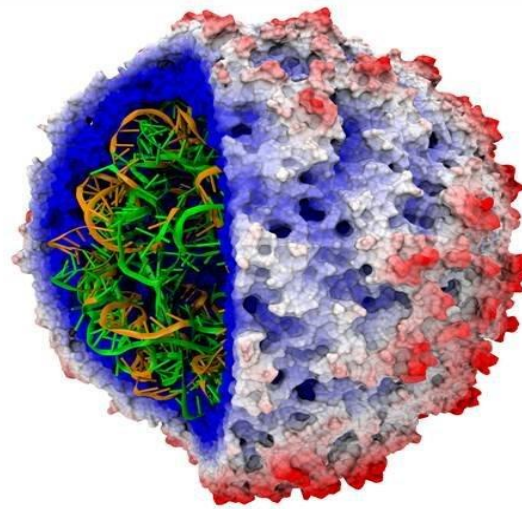
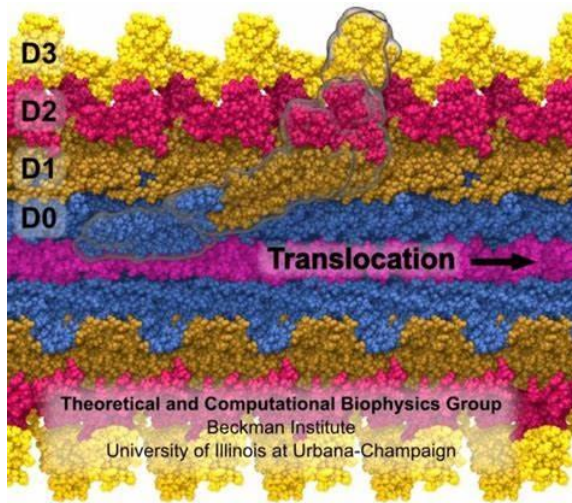


# Vật lý Y sinh

Trần Thị Minh Thu

Email: [ttmthu@hcmus.edu.vn](mailto:ttmthu@hcmus.edu.vn)



# TÀI LIỆU THAM KHẢO

1. Vật lý thống kê, Nguyễn Nhật Khanh, Tủ sách Trường ĐHKHTN, 1999.
2. Mô phỏng trong Vật lý, Võ Văn Hoàng (Chủ biên) NXB Đại học Quốc gia TP HCM, 2016.
3. Understanding Molecular Simulation: From algorithms to application, Daan Frenkel and Berend Smit, Second edition, 2005, Academic press.

# HÌNH THỨC THI VÀ CHẤM ĐIỂM

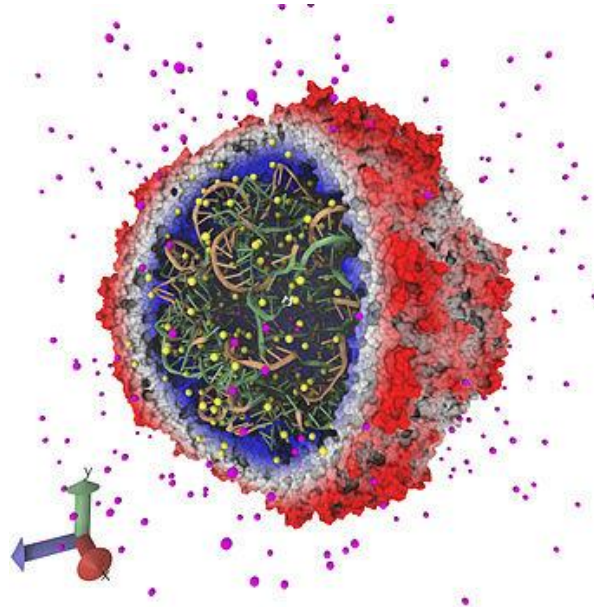
- Bài tập cá nhân và điểm danh tại lớp: 20%
- Báo cáo đồ án: 40%
- Điểm cuối kỳ: 40%



03 TC lý thuyết: 12 buổi; 04 tiết / 01 buổi

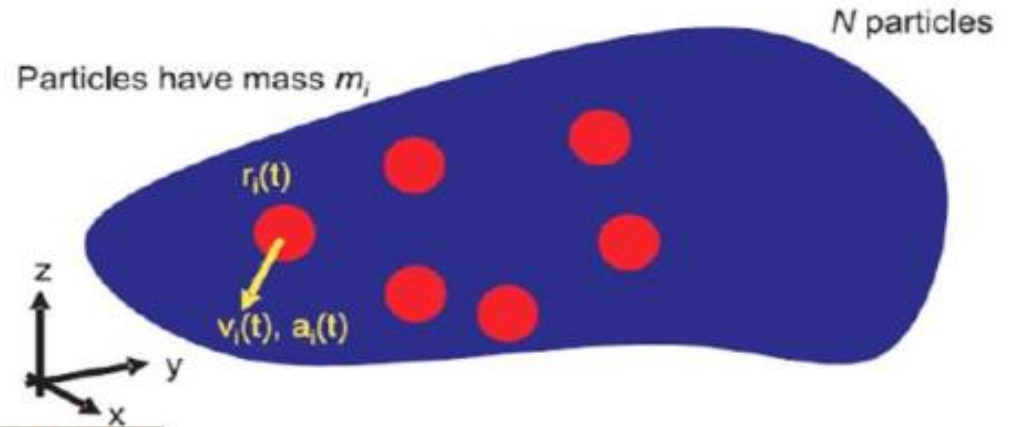
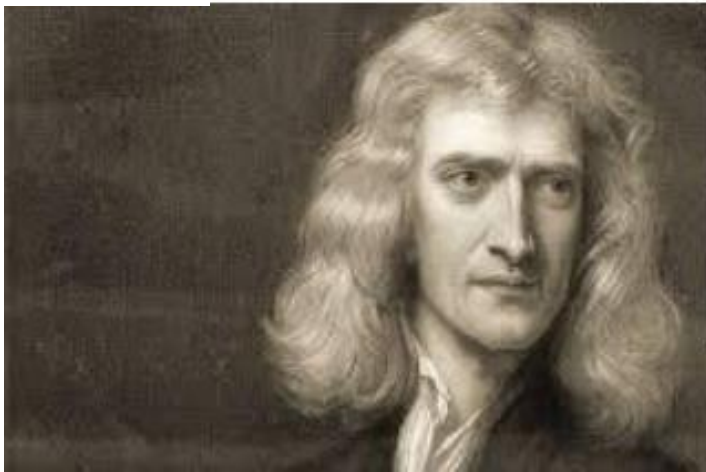
- Buổi 1+2: Giới thiệu về protein và khoa học tính toán
- Buổi 3: Phương pháp mô phỏng MD.
- Buổi 4: Kỹ thuật mô phỏng MD: thuật giải, trường lực, các phần mềm.
- Buổi 5: Kỹ thuật mô phỏng MD: thuật giải, trường lực, các phần mềm (tiếp theo).
- Buổi 6+7+8: Giới thiệu và hướng dẫn sử dụng phần mềm VMD, XMGRACE, GROMACS.
- Buổi 9+10: Giao đồ án môn học và hướng dẫn làm đồ án
- Buổi 11+12: Báo cáo đồ án

# Buổi 4: Kỹ thuật mô phỏng MD: thuật giải, trường lực, các phần mềm





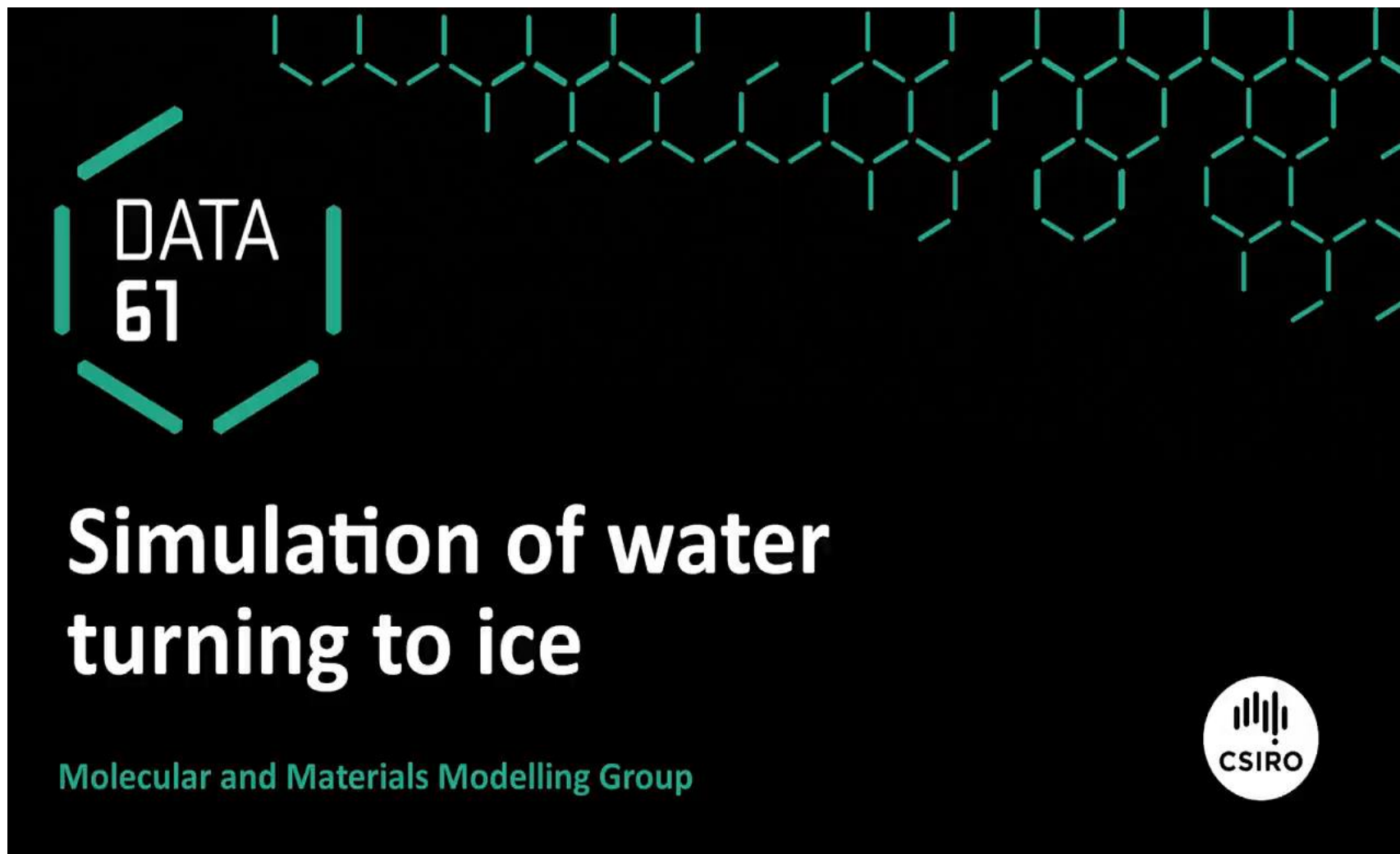
# Mô phỏng động học phân tử cổ điển



Là mô phỏng máy tính các chuyển động vật lý của nguyên tử và phân tử (trong hệ nhiều hạt)

Giải số các phương trình hai Newton: cơ học cổ điển - không có vật lý mới!


# *Conformations – Frames – Snapshots*



DATA  
61

Simulation of water  
turning to ice

Molecular and Materials Modelling Group



# Molecular Modeling

For each atom in every molecule, we need:

- Position ( $r$ )
- Momentum ( $m + v$ )
- Charge ( $q$ )
- Bond information (which atoms, bond angles, etc.)



# From Potential to Movement

To run the simulation,  
we need the force on  
each particle.

$$F_i = m_i a_i$$

We use the gradient of  
the potential energy  
function.

$$F_i = -\nabla_i V$$

Now we can find the  
acceleration.

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

# What is the Potential?

A single atom will be affected by the potential energy functions of every atom in the system:

- Bonded Neighbors
- Non-Bonded Atoms (either other atoms in the same molecule, or atoms from different molecules)

$$V(R) = E_{bonded} + E_{non-bonded}$$

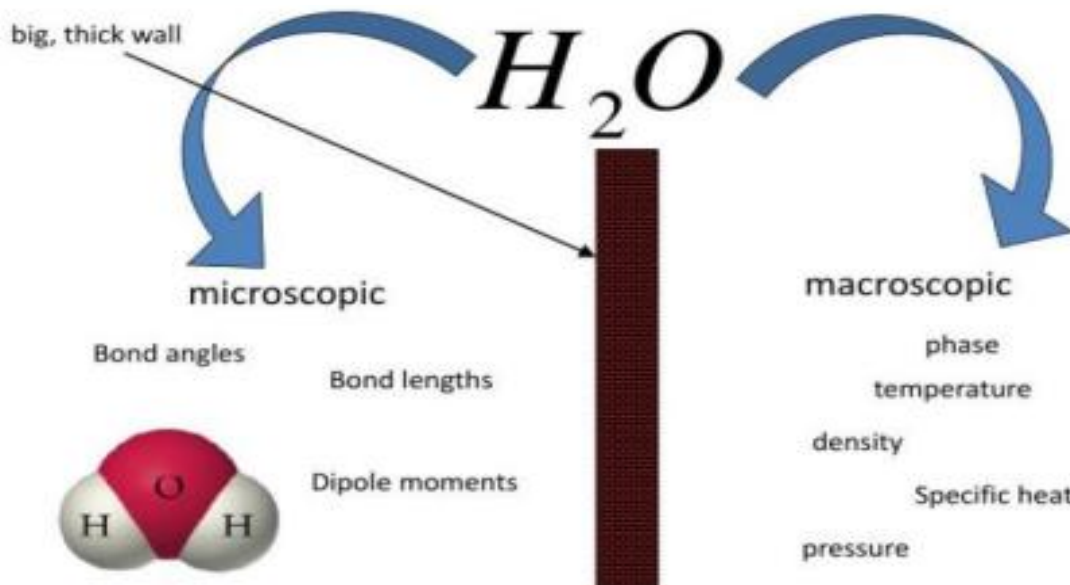
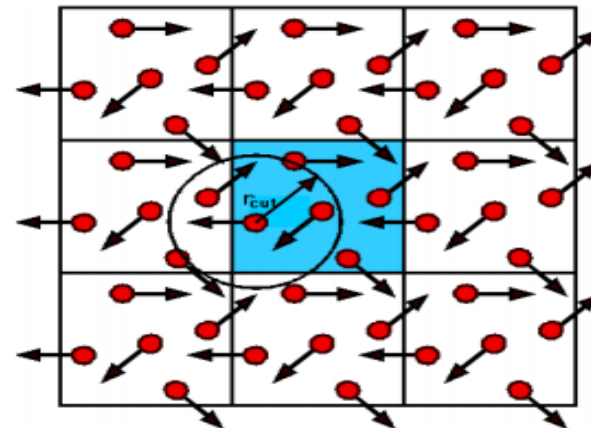
# **Tương quan giữa Cơ học thống kê và Cơ học cổ điển**

# Tại sao phải sử dụng cơ học thống kê?

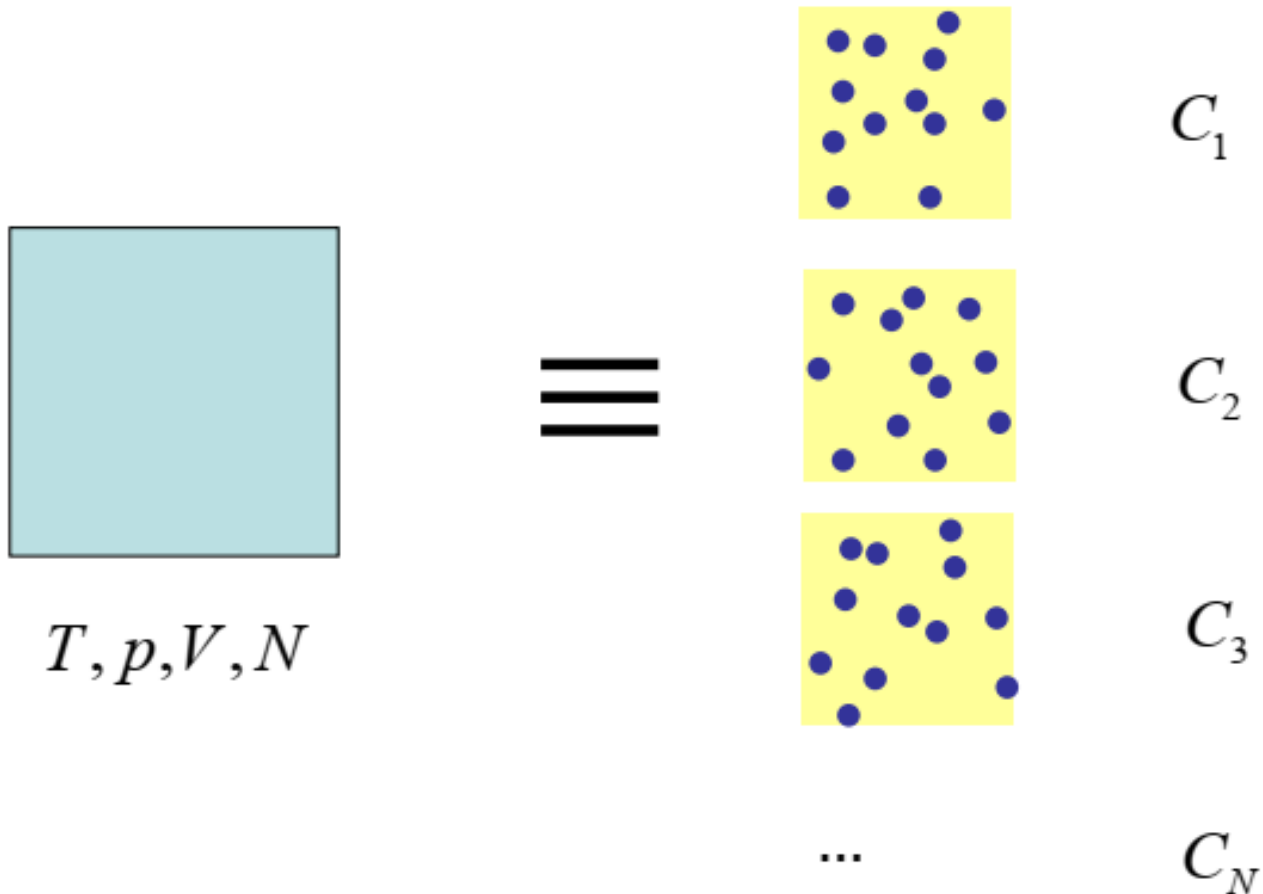
Các đại lượng vĩ mô **quan sát được** (thể tích, nhiệt độ, nồng độ, ...)



Các trạng thái vi mô (toạ độ, vận tốc của các nguyên tử)



# Trạng thái vi mô – Trạng thái vĩ mô



- Cùng một trạng thái vĩ mô, hệ vật liệu có thể được biểu diễn bởi rất nhiều cấu hình  $C_i$  vi mô khác nhau.

# Trạng thái vi mô – Trạng thái vĩ mô

- Cơ học thống kê: Mỗi trạng thái vi mô của hệ được mô tả bằng một điểm trong không gian pha, là tập hợp tất cả các giá trị khả dĩ của tọa độ  $\{q_i\}$  và xung lượng  $\{p_i\}$ , ( $i=1,2,\dots,N$  và  $N$  là tổng số nguyên tử).
  - Theo cơ học thống kê, một đại lượng quan sát được trong trạng thái vĩ mô của hệ có thể được viết dưới dạng giá trị trung bình trên toàn không gian pha.
- Làm sao để tính giá trị trung bình trên toàn không gian pha ???



# **Trạng thái vi mô: Mô tả cơ học hệ nhiều hạt**

# Từ chuyển động Newton đến chuyển động Lagrange

- Newton: chuyển động không ràng buộc

$$m\ddot{\mathbf{r}} = \mathbf{F}$$
$$\Leftrightarrow \begin{cases} m\ddot{x} = F_x \\ m\ddot{y} = F_y \\ m\ddot{z} = F_z \end{cases}$$



Từ những lực được cung cấp, chúng ta có thể tính được 3N biến dựa trên 3N phương trình như trên.

Định luật Newton thường được dùng để phân tích chuyển động của hệ vật. Tuy các định luật của Newton rất dễ nhớ, triển khai chúng trong thực tế là một điều vô cùng khó khăn vì số bậc tự do quá lớn và các đại lượng tính toán là có hướng.

# Cơ học Lagrange

- Cơ học Lagrange hướng tới những đại lượng vô hướng như năng lượng để mô tả chuyển động của một hệ, do đó khiến bài toán trở nên đơn giản hơn rất nhiều.
- Các chất điểm trong hệ thường sẽ có những liên hệ gọi là **ràng buộc**.
- Chúng ta xem xét định luật 2 Newton với hai thành phần lực tác động vào hệ: thành phần lực ràng buộc  $F_c$  và thành phần lực gây ra chuyển động tự do  $F$ .

$$m\ddot{\mathbf{r}} = \mathbf{F} + \mathbf{F}_c$$

# Cơ học Lagrange

- Do lực ràng buộc không tạo công tại một thời điểm nhất định nào, chúng ta sẽ nhân thêm độ dời ảo (vi phân của tọa độ không phụ thuộc thời gian) để triệt tiêu lực ràng buộc. Chúng ta có :

$$\begin{aligned}\sum_{i=1}^N m_i \ddot{\mathbf{r}}_i &= \sum_{i=1}^N (\mathbf{F}_i + \mathbf{F}_i^c) \\ \Leftrightarrow \sum_{i=1}^N (m_i \ddot{\mathbf{r}}_i - \mathbf{F}_i) &= \sum_{i=1}^N \mathbf{F}_i^c \\ \Leftrightarrow \sum_{i=1}^N (m_i \ddot{\mathbf{r}}_i - \mathbf{F}_i) \delta \mathbf{r}_i &= \sum_{i=1}^N \mathbf{F}_i^c \delta \mathbf{r}_i \\ \Leftrightarrow \sum_{i=1}^N (m_i \ddot{\mathbf{r}}_i - \mathbf{F}_i) \delta \mathbf{r}_i &= 0\end{aligned}$$

Phương trình trên gọi là nguyên lý D'Alembert

# Cơ học Lagrange - Các tọa độ tổng quát

- Chuyển hệ tọa độ Decartes thông thường thành một hệ tọa độ tổng quát  $\mathbf{q}$ , trong đó các thành phần  $q_1, q_2, \dots, q_f$  đặc trưng cho  $f$  bậc tự do của hệ.
  - Hệ có 2 bậc tự do  $\mathbf{q}$  có hai thành phần, hệ có 6 bậc tự do,  $\mathbf{q}$  có 6 thành phần.
- > Viết lại hệ tọa độ thông thường  $\mathbf{r}$  dưới dạng một hàm của các biến mới  $\mathbf{q}$ :

$$\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, \dots, q_f, t)$$

$$\mathbf{r}_i = \mathbf{r}_i(q_j, t)$$

$f$ : bậc tự do của hệ.

# Cơ học Lagrange - Các tọa độ tổng quát

- Từ đây, chúng ta có thể suy ra độ dời ảo và đạo hàm của  $\mathbf{r}_i$ :

$$\delta \mathbf{r}_i = \sum_{j=1}^f \frac{\partial \mathbf{r}_i}{\partial q_j} \delta q_j$$

$$\dot{\mathbf{r}}_i = \sum_{j=1}^f \frac{\partial \mathbf{r}_i}{\partial q_j} \dot{q}_j + \frac{\partial \mathbf{r}_i}{\partial t}$$



# Cơ học Lagrange - Các tọa độ tổng quát

- Vận tốc của một phương theo tọa độ tổng quát là độc lập với phương khác (do các phương đặc trưng cho bậc tự do riêng của chúng), nếu chúng ta lấy đạo hàm từng phần  $\dot{\mathbf{r}}_i$  theo  $\dot{q}_j$

$$\frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_j} = \frac{\partial \mathbf{r}_i}{\partial q_j}$$

- Thay phương trình độ dời ảo vào phương trình nguyên lý D'Alembert:

$$\begin{aligned} & \sum_{i=1}^N (m_i \ddot{\mathbf{r}}_i - \mathbf{F}_i) \delta \mathbf{r}_i = 0 \\ \Leftrightarrow & \sum_{i=1}^N (m_i \ddot{\mathbf{r}}_i - \mathbf{F}_i) \sum_{j=1}^f \frac{\partial \mathbf{r}_i}{\partial q_j} \delta q_j = 0 \\ \Leftrightarrow & \sum_{j=1}^f \left( \sum_{i=1}^N m_i \ddot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial q_j} - \sum_{i=1}^N \mathbf{F}_i \frac{\partial \mathbf{r}_i}{\partial q_j} \right) \delta q_j = 0 \end{aligned}$$

# Cơ học Lagrange - Các tọa độ tổng quát

- Sử dụng quy tắc nhân trong đạo hàm, chúng ta có:

$$\frac{d}{dt} \left( \dot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial q_j} \right) = \ddot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial q_j} + \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial q_j}$$

Áp dụng vào phương trình trên:

$$\sum_{j=1}^f \left( \sum_{i=1}^N m_i \frac{d}{dt} \left( \dot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial q_j} \right) - \sum_{i=1}^N m_i \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial q_j} - \sum_{i=1}^N \mathbf{F}_i \frac{\partial \mathbf{r}_i}{\partial q_j} \right) \delta q_j = 0 \quad (*)$$

# Cơ học Lagrange - Các đại lượng vô hướng

- Lagrange nghĩ rằng các đại lượng vô hướng có thể đặc trưng cho chuyển động của toàn hệ, ví dụ như động năng. Động năng của hệ được định nghĩa như

$$T = \sum_{i=1}^N \frac{1}{2} m_i \dot{\mathbf{r}}_i^2$$

Đạo hàm từng phần của động năng theo các tọa độ tổng quát là:

$$\frac{\partial T}{\partial q_j} = \sum_{i=1}^N m_i \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial q_j} \quad (**)$$

$$\frac{\partial T}{\partial \dot{q}_j} = \sum_{i=1}^N m_i \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_j} = \sum_{i=1}^N m_i \dot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial q_j} \quad (***)$$

$Q_j$  là lực tổng quát

$$Q_j = \sum_{i=1}^N \mathbf{F}_i \frac{\partial \mathbf{r}_i}{\partial q_j}$$

Thay  $Q$ ,  $(**)$  và  $(***)$  vào  $(*)$ , ta được:

$$\sum_{j=1}^f \left( \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} - Q_j \right) \delta q_j = 0$$

# Cơ học Lagrange - Lực bảo toàn và Thế năng

- Phương trình chuyển động tổng quát là:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} = Q_j$$

Chúng ta có thể đưa thế năng vào phương trình này để giảm bớt đi số lực cần khảo sát. Đầu tiên chúng ta phân tích thành phần lực thành các lực bảo toàn (lực thế) và các lực khác.

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} = \sum_{i=1}^N \mathbf{F}_i^p \frac{\partial \mathbf{r}_i}{\partial q_j} + Q_j$$

Mà các lực thế lại chính là trừ gradient của thế năng:

$$\mathbf{F}_i^p = -\nabla V$$

Do đó:

$$\sum_{i=1}^N \mathbf{F}_i^p \frac{\partial \mathbf{r}_i}{\partial q_j} = -\frac{\partial V}{\partial q_j}$$

# Cơ học Lagrange - Lực bảo toàn và Thế năng

- Và chúng ta có thể viết lại phương trình chuyển động tổng quát như sau:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} + \frac{\partial V}{\partial q_j} = Q_j$$

Đây là phương trình mô tả chuyển động của một hệ với thế năng. Dạng này được sử dụng rộng rãi nhất trong phân tích chuyển động của hệ vật do nó tuy tổng quát nhưng vẫn tính đến những đại lượng vô hướng có thể dễ dàng tính được như động năng và thế năng.

# Đa thức Lagrange và Phương trình Euler-Lagrange

- Bởi vì thế năng chỉ phụ thuộc vào vị trí của vật, do đó đạo hàm theo vận tốc của nó bằng 0:

$$\frac{d}{dt} \frac{\partial V}{\partial \dot{q}_i} = 0$$

Thêm biểu thức trên vào PT chuyển động tổng quát, ta có:

$$\frac{d}{dt} \frac{\partial (T - V)}{\partial \dot{q}_j} - \frac{\partial (T - V)}{\partial q_j} = Q_j$$

Chúng ta định nghĩa đa thức Lagrange như sau:

$$L = T - V$$

Nếu hệ là bảo toàn (cơ năng không bị thất thoát hay được bơm thêm, tức các thành phần lực tổng quát bằng 0) thì chúng ta có phương trình sau:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0$$



# Mô tả cơ học hệ nhiều hạt

- Xét một hệ có  $f$  bậc tự do, nếu hệ cổ điển  $N$  hạt thì  $f=3N$ .
- Theo hình thức luận Lagrange, trạng thái cơ học (trạng thái vi mô) của hệ được xác định bởi  $f$  tọa độ suy rộng  $q_1, q_2, \dots, q_f$  và  $f$  vận tốc suy rộng  $\dot{q}_1, \dot{q}_2, \dots, \dot{q}_f$ , trong đó:

$$\dot{q}_i = \frac{dq_i}{dt} \quad (i=1, 2 \dots f)$$

- Các phương trình chuyển động là các phương trình Lagrange:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \quad (i=1, 2, \dots, f) \quad (1.1)$$

- Đối với hệ bảo toàn:

$$L_{(q, \dot{q})} = T - V, \quad (1.2)$$

- $T$ : động năng,  $V$ : thế năng

# Mô tả cơ học hệ nhiều hạt

- Vật lý thống kê: Hình thức luận Hamilton, trong đó trạng thái của hệ được cho bởi  $f$  tọa độ suy rộng  $q_1, q_2, \dots, q_f$  và  $f$  xung lượng suy rộng  $p_1, p_2, \dots, p_f$ , hay bởi tập các giá trị:

$$(q_1, \dots, q_f, p_1, \dots, p_f) \equiv (q, p) \quad (1.3)$$

Theo định nghĩa:

$$p_i \equiv \frac{\partial L}{\partial \dot{q}_i} \quad (1.4)$$

# Mô tả cơ học hệ nhiều hạt

- Các phương trình chuyển động là các phương trình Hamilton:

$$\frac{\partial H}{\partial p_i} = \dot{q}_i \quad ; \quad \frac{\partial H}{\partial q_i} = -\dot{p}_i \quad (i = 1, 2, \dots, f) \quad (1.5)$$

Trong đó hàm Hamilton được xác định bằng:

$$H(p, q) = \sum_{i=1}^f p_i \dot{q}_i - L(q, \dot{q})$$

# Mô tả cơ học hệ nhiều hạt

Giải (1.5) với các điều kiện ban đầu cho trước

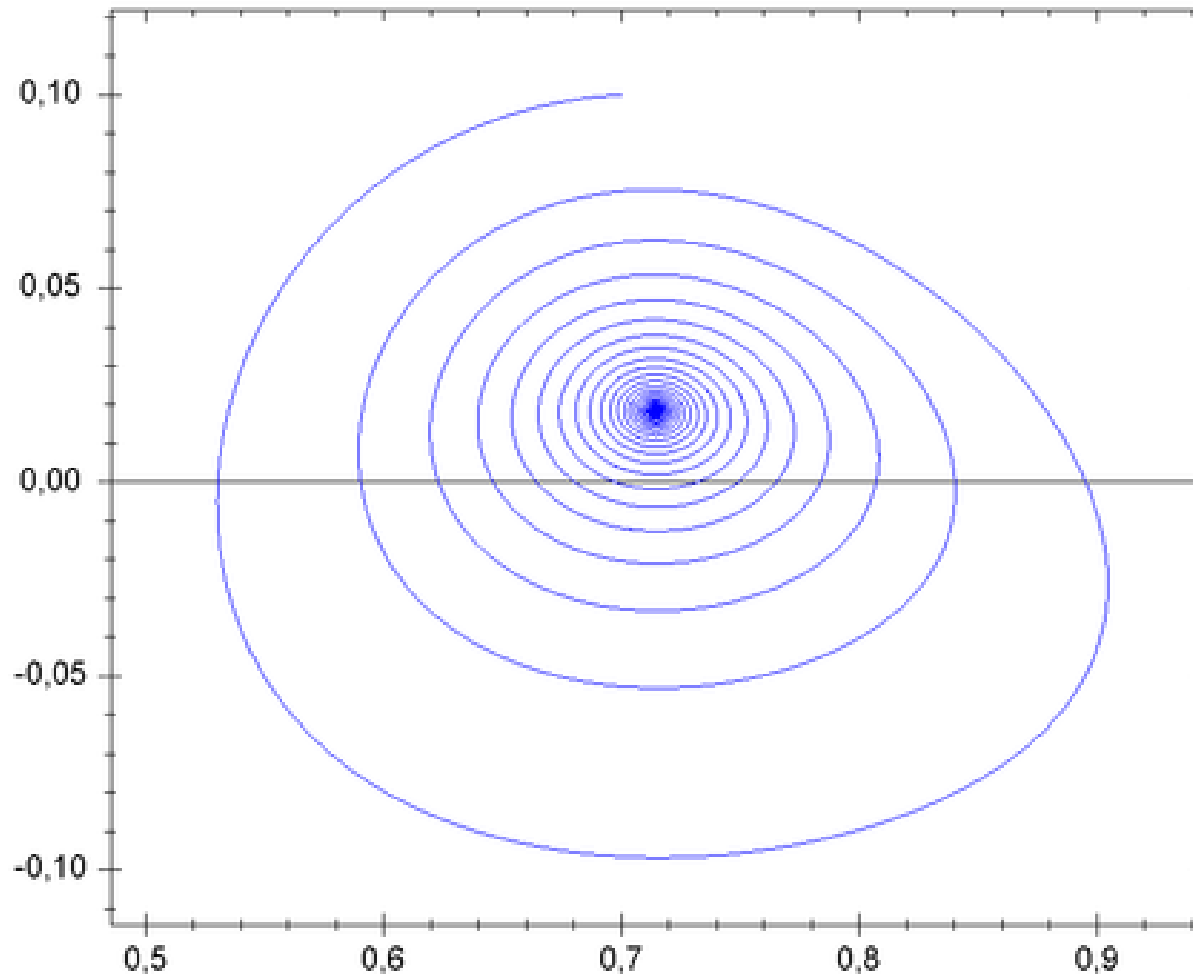
$$\left. \begin{aligned} p_i(t=0) &= p_i^0 \\ q_i(t=0) &= q_i^0 \end{aligned} \right\} \quad (1.6)$$

Về nguyên tắc có thể biết được trạng thái vi mô của hệ ở từng thời điểm. Nhưng trên thực tế bài toán này không thể giải được vì số hạt quá lớn hoặc ngay cả khi biết được tọa độ và xung lượng của từng hạt tại thời điểm ban đầu cũng không đủ thời gian để thể chúng vào nghiệm tổng quát của phương trình chuyển động

# Không gian pha

- Trạng thái vi mô của hệ nhiều hạt  $f$  bậc tự do có thể xác định bằng  $f$  tọa độ  $q_i$  và  $f$  xung lượng  $p_i$  nên có thể được biểu diễn bằng không gian  $2f$  chiều, không gian này được gọi là không gian pha (không gian  $\Gamma$ - là một khái niệm toán học).
- Không gian pha:  $6N$  biến,  $\Gamma = (r,p)$  hoặc  $(q,p)$
- Điểm biểu diễn trạng thái vi mô của hệ được gọi là điểm pha.
- Theo thời gian, trạng thái vi mô của hệ thay đổi dẫn đến điểm pha chuyển động vạch một đường trong không gian pha, gọi là quỹ đạo pha.

# Quỹ đạo pha



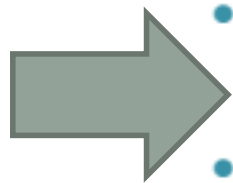
Không gian pha của một hệ động lực với tâm không ổn định, biểu diễn một quỹ đạo pha.

[https://vi.wikipedia.org/wiki/Kh%C3%B4ng\\_gian\\_phạ](https://vi.wikipedia.org/wiki/Kh%C3%B4ng_gian_phạ)



# Nguyên lý thống kê

- Do số hạt quá lớn và chuyển động không ngừng, chúng tương tác với nhau và với môi trường xung quanh một cách phức tạp, ngẫu nhiên. Khi đó hệ vĩ mô luôn chuyển từ trạng thái vi mô này sang trạng thái vi mô khác. Nói cách khác, hệ vĩ mô không thể nằm lâu trong bất kỳ trạng thái vi mô khả dĩ nào.
- Các đại lượng ở trạng thái vĩ mô chỉ có thể tính được dựa trên giá trị trung bình của các đại lượng vật lý theo trạng thái vi mô.
- Các kết luận và tiên đoán về trạng thái của một hệ vĩ mô mang tính xác suất.
- Vật lý thống kê là lý thuyết thống kê về hệ nhiều hạt.



# Nguyên lý tiến tới trạng thái cân bằng

- Các giá trị đo được trong các thí nghiệm của các thông số vĩ mô được cho là giá trị trung bình của chúng.
- Ở trạng thái vĩ mô nào mà mức độ hỗn loạn trong phân bố và chuyển động của các hạt cấu thành hệ càng cao thì số trạng thái vi mô càng nhiều, xác suất tồn tại trạng thái vĩ mô đó càng lớn.
- Trạng thái vĩ mô có mức độ hỗn loạn cao nhất gọi là trạng thái cân bằng thống kê (hoặc cân bằng nhiệt động, hoặc cân bằng nhiệt).
- Trong tự nhiên, mọi hệ kín đều liên tục chuyển từ trạng thái vĩ mô có xác suất nhỏ sang trạng thái vĩ mô có xác suất lớn hơn, cho đến khi đạt trạng thái cân bằng. Đây là nguyên lý tiến tới trạng thái cân bằng của vật lý thống kê.

# Tập hợp (ensembles)

- Một số lượng lớn các bản sao của hệ trong cùng một điều kiện vĩ mô.
- Mỗi bản sao thể hiện một trạng thái vi mô khả dĩ của hệ vĩ mô trong cùng một điều kiện nhiệt động ( $T, p, V, N \dots$ )

*Gibbs, 1878*

# Tập hợp

Result of thermodynamical constraints, e.g. temperature, pressure...

Microcanonical

$NVE$

Canonical

$NVT$

Isobaric-isothermal

$NpT$

Grand canonical

$TV\mu$

$\mu$

chemical potential (e.g. concentration)

# Tập hợp vi chính tắc

- Xét một hệ cô lập, do không trao đổi năng lượng với bên ngoài nên  $E = \text{const}$ ,
- $N, V, E = \text{constants}$
- Năng lượng nằm trong khoảng:

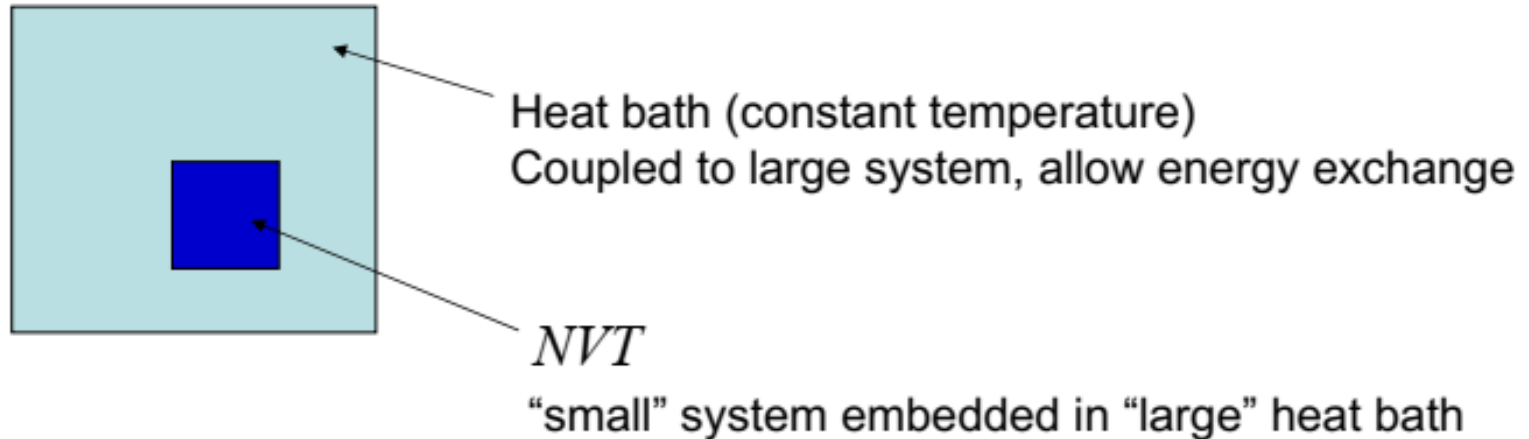
$$E \in [E_0, E_0 + \delta E]$$

- Hàm số phân bố  $\rho(q, p)$  được biểu diễn:

$$\rho(q, p) = \begin{cases} C & \text{nếu } E(q, p) \in [E_0, E_0 + \delta E] \\ 0 & \text{nếu } E(q, p) \notin [E_0, E_0 + \delta E] \end{cases}$$

- Phương pháp động lực học phân tử (MD): tạo ra tập vi chính tắc ( $E=\text{constant}$ ), đồng thời bảo toàn xung lượng tổng cộng.

# Tập hợp chính tắc



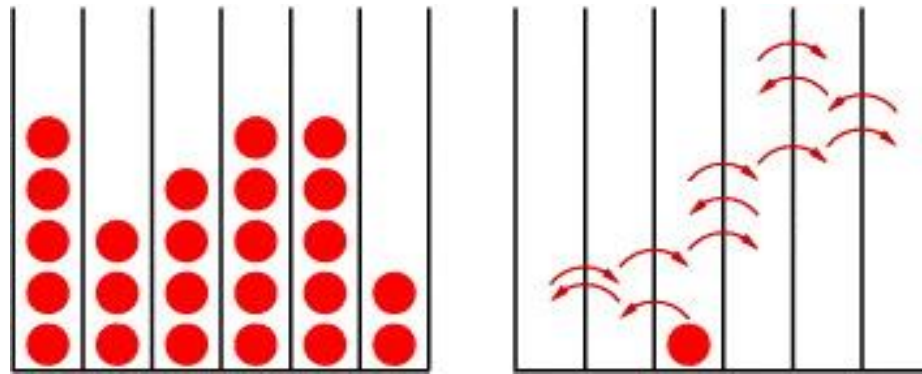
Constant number of particles =  $N$

Constant volume =  $V$

Constant temperature =  $T$

# Nguyên lý ERGODIC

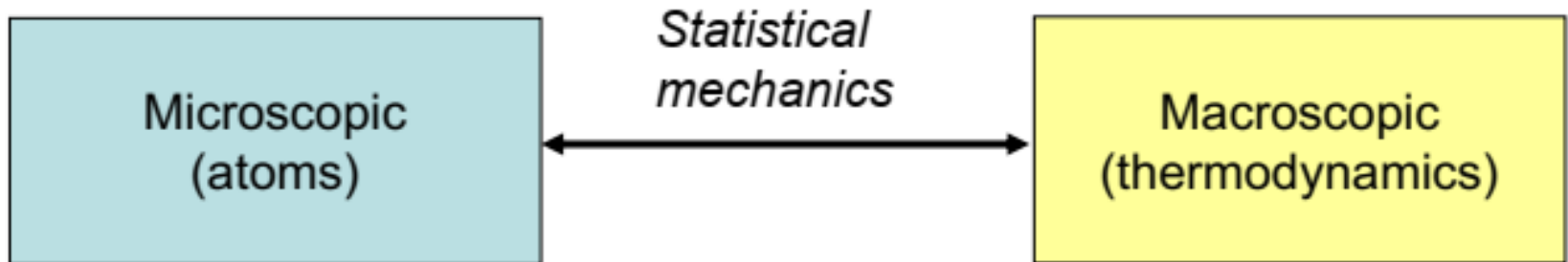
- Ý tưởng: Thay vì khảo sát một hệ ở nhiều thời điểm, tiến hành khảo sát một tập hợp nhiều hệ tương tự tại cùng một thời điểm



- Một đại lượng nhiệt động mà ta đo đạc được trong thế giới vĩ mô là giá trị trung bình trên tất cả các trạng thái vi n

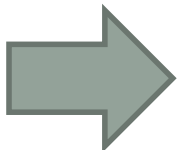
$$\langle A \rangle_{Ens} = \langle A \rangle_{Time}$$

# Mối tương quan giữa cơ học thống kê và các tính chất nhiệt động



- **Hệ vĩ mô được xác định bởi những tính chất:** Ví dụ  $(N, V, E) = NVE$  ensemble
- **Điều kiện trong trạng thái vĩ mô** (Không đổi về thể tích, nhiệt độ, số hạt trong hệ) được chuyển đổi sang trạng thái vi mô.

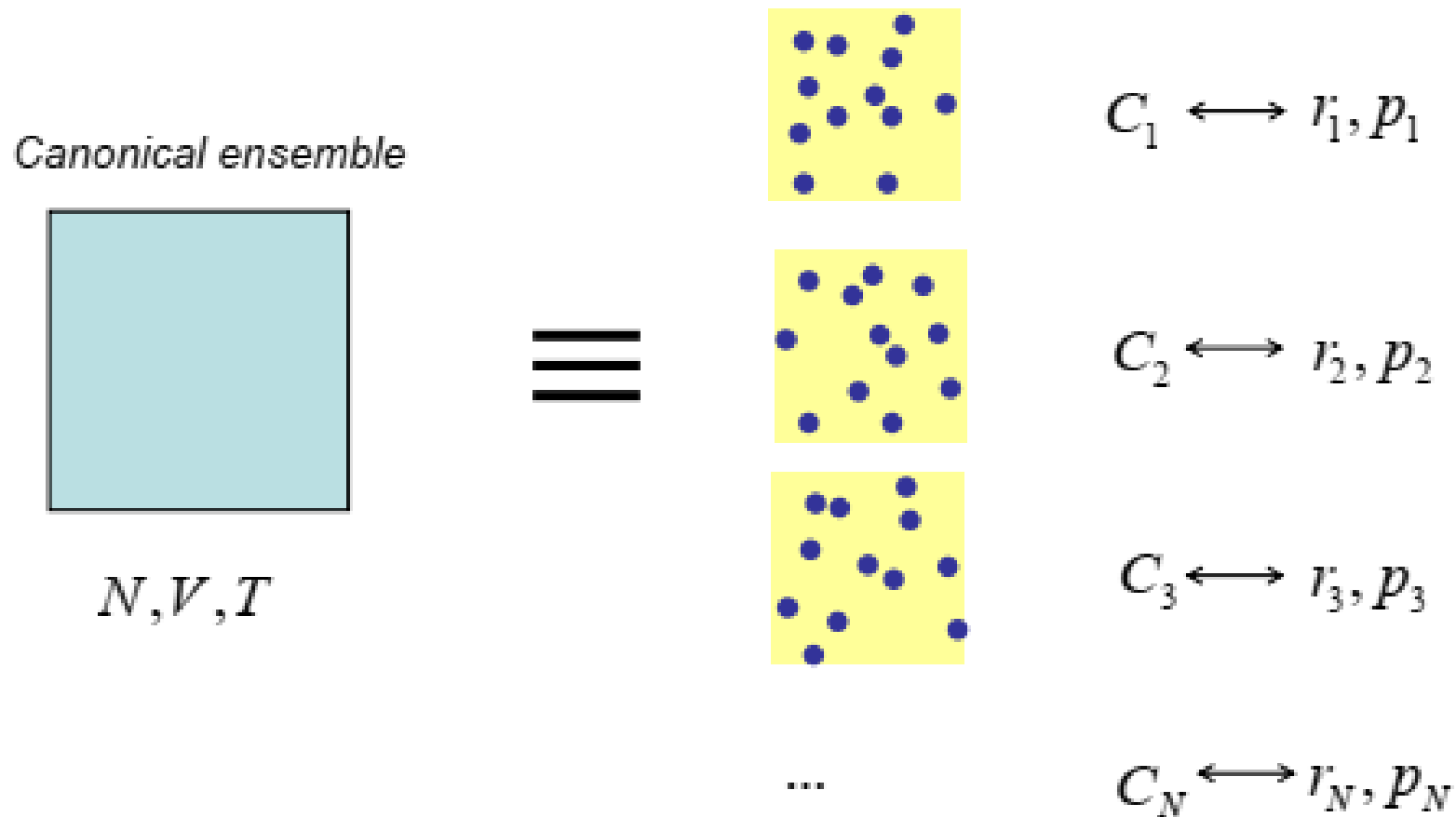
“Behavior” trong trạng thái vi mô có quan hệ ràng buộc với điều kiện vĩ mô, hay nói cách khác, phân bố của các trạng thái vi mô ảnh hưởng đến tính chất của hệ vĩ mô



Để tính một tính chất trong trạng thái vĩ mô, cần biết **phân bố (distribution)** của hệ trong trạng thái vi mô.



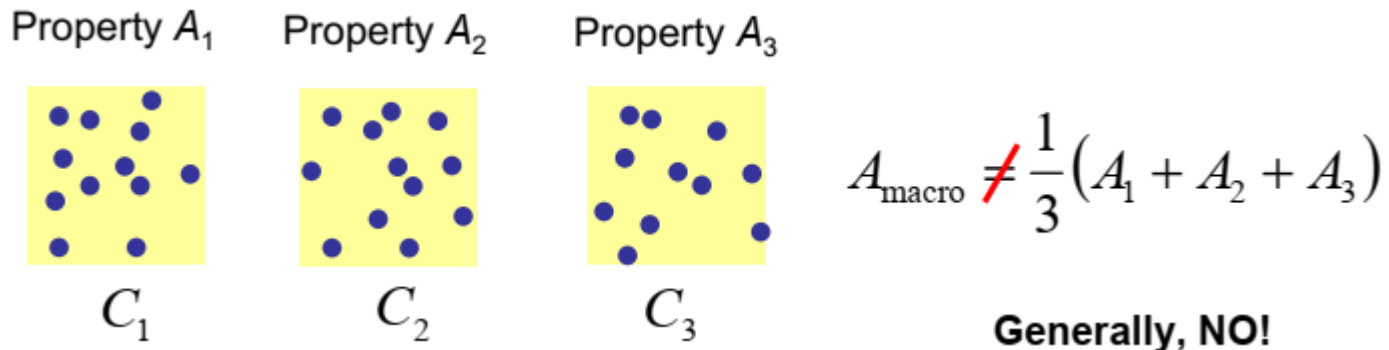
# Trạng thái vĩ mô - trạng thái vi mô



Cùng một trạng thái vĩ mô, hệ vật liệu có thể được biểu diễn bởi rất nhiều cầu hình  $C_i$  vi mô khác nhau và có một số cầu hình  $C_i$  có xu hướng giống nhau

# Trung bình trên tập hợp

- Rather than taking single measurement, need to average over “all” microscopic states that represent the corresponding macroscopic condition
- This averaging needs to be done in a suitable fashion, that is, we need to **consider the specific distribution of microscopic states** (e.g. some microscopic states may be more likely than others)



## Trung bình trên tập hợp

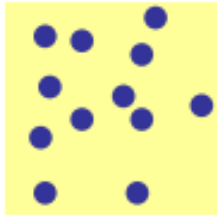
Để tính một tính chất trong trạng thái vĩ mô, cần biết **phân bố (distribution)** số trạng thái vi mô để tìm ra **mật độ xác suất**.

**Do vậy:**

Không thể chỉ dùng dữ liệu trên một cấu hình vi mô đơn lẻ để tính toán tính chất trong trạng thái vĩ mô mà cần tính trung bình trên toàn tập hợp.

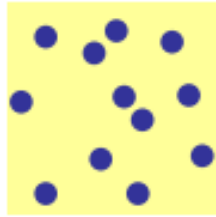
# Trung bình trên tập hợp

Property  $A_1$



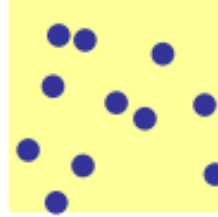
$C_1$

Property  $A_2$



$C_2$

Property  $A_3$



$C_3$

$$A_{\text{macro}} \neq \frac{1}{3}(A_1 + A_2 + A_3)$$

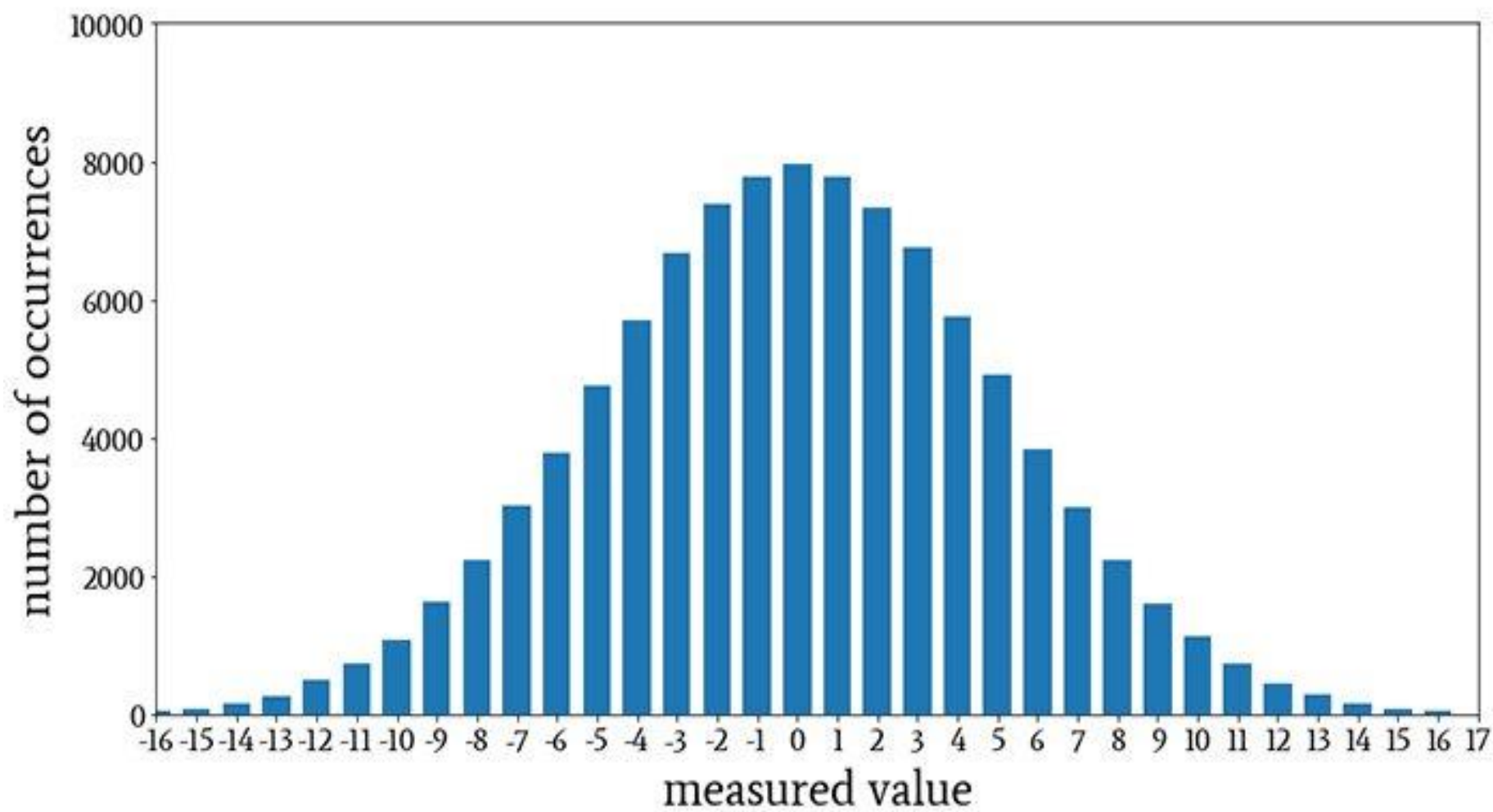
- Không thể tính trung bình theo cách thông thường, mà cần nhân các trọng số tương ứng với từng trạng thái

$$A_{\text{macro}} = \rho_1 A_1 + \rho_2 A_2 + \rho_3 A_3 =$$

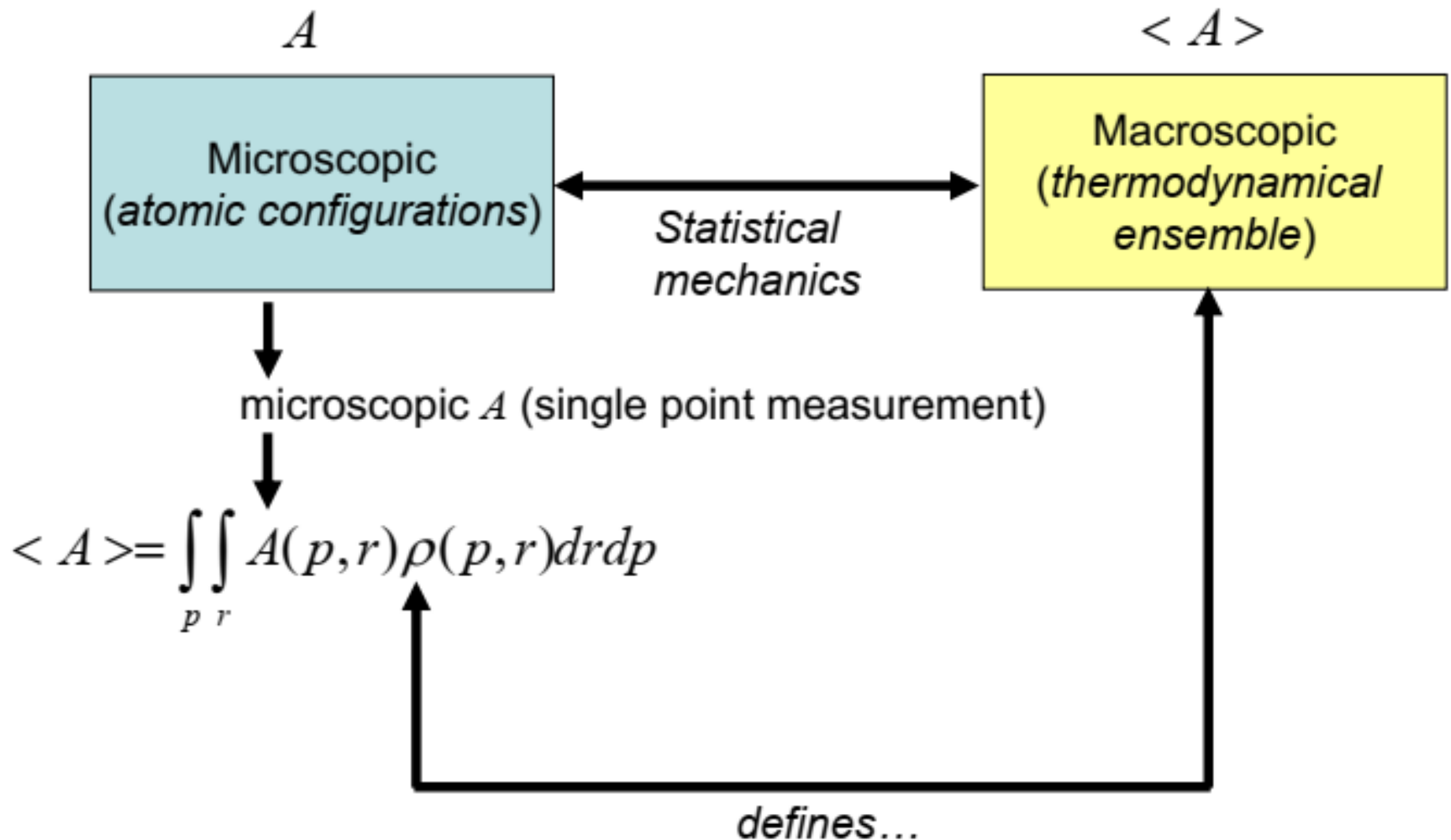
$$\rho_1(r_1, p_1) A_1(r_1, p_1) + \rho_2(r_2, p_2) A_2(r_2, p_2) + \rho_3(r_3, p_3) A_3(r_3, p_3)$$



Mật độ xác suất có trạng thái  $C_1$



# Quan hệ giữa trạng thái vi mô và trạng thái vĩ mô



# Trị trung bình trên toàn không gian pha

Đại lượng quan sát được trong một trạng thái vĩ mô của hệ **A** có thể được viết dưới dạng giá trị trung bình trên toàn không gian pha

$$\langle A \rangle = \int \int_{p \ r} A(p, r) \rho(p, r) dr dp$$

- $\langle A \rangle$  là trung bình thống kê của đại lượng cần quan sát trong trạng thái vĩ mô và được biểu diễn như là hàm của tọa độ  $r$  và xung lượng  $p$ .
- $\rho(p, r)$  là hàm mật độ xác suất của các trạng thái vi mô.
- $d\vec{r} = d\vec{r}_1 d\vec{r}_2 \dots d\vec{r}_N$
- $d\vec{p} = d\vec{p}_1 d\vec{p}_2 \dots d\vec{p}_N$
- $N$  là tổng số hạt của hệ.

# Trị trung bình trên toàn không gian pha

$$\langle A \rangle = \int \int_{p, r} A(p, r) \rho(p, r) dr dp$$

Trong điều kiện hệ không thay đổi số hạt, thể tích và nhiệt độ (điều kiện NVT), tất cả các trạng thái vi mô khả dĩ của hệ tuân theo phân bố của tập hợp thống kê chính tắc, hàm mật độ xác suất của các trạng thái vi mô có dạng:

$$\rho_{NVT} = \frac{\exp \left[ -\frac{H(\{\vec{r}_i\}, \{\vec{p}_i\})}{k_B T} \right]}{Z_{NVT}}$$

Trong đó  $k_B$  là hằng số Boltzmann,  $T$  là nhiệt độ,  $H$  là Hamiltonian cổ điển của hệ và  $Z$  là tổng thống kê:

$$Z_{NVT} = \int \exp \left[ -\frac{H}{k_B T} \right] d\vec{p} d\vec{r}$$



## How to relate microscopic states to macroscopic variables?

$A(r, p)$  Property due to specific microstate

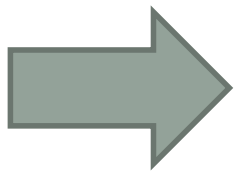
*To measure an observable quantity from MD simulation we must express this observable as a function of the positions and linear momenta of the particles in the system, that is,  $r, p$*

Recall, microscopic states characterized by  $r, p$

$$r = \{\vec{x}_i\}, p = \{m_i \dot{\vec{x}}_i\} \quad i = 1..N$$
$$\parallel$$
$$p_i$$

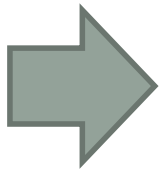
## Trị trung bình trên toàn không gian pha, trung bình trên tập hợp

$$\langle A \rangle = \int \int_{p, r} A(p, r) \rho(p, r) dr dp$$



Theo cơ học thống kê, để tính các đại lượng vĩ mô từ các đại lượng vi mô, ta cần phải tính tích phân trên toàn bộ không gian pha.

**Điều này rất phức tạp và trên thực tế là không thể thực hiện được.**



**Do đó, ta chỉ có thể ước lượng đại lượng nhiệt động A một cách gần đúng bằng lời giải số và phương pháp mô phỏng động lực học phân tử là một trong những phương pháp hiệu quả nhất.**

# Rời rạc quỹ đạo pha để tính giá trị bằng lời giải số

- Lagrange (variational formulation of classical mechanics):
  - Consider a system that is at a point  $r_0$  at time 0 and at point  $r_t$  at time  $t$ , then the system follows a trajectory  $r(t)$  such that:

$$S = \int_{t_b}^{t_e} dt [\mathcal{K} - \mathcal{U}]$$

is an extremum.



Quỹ đạo rời rạc



Quỹ đạo thực

# Rời rạc quỹ đạo pha để tính giá trị trung bình bằng lời giải số

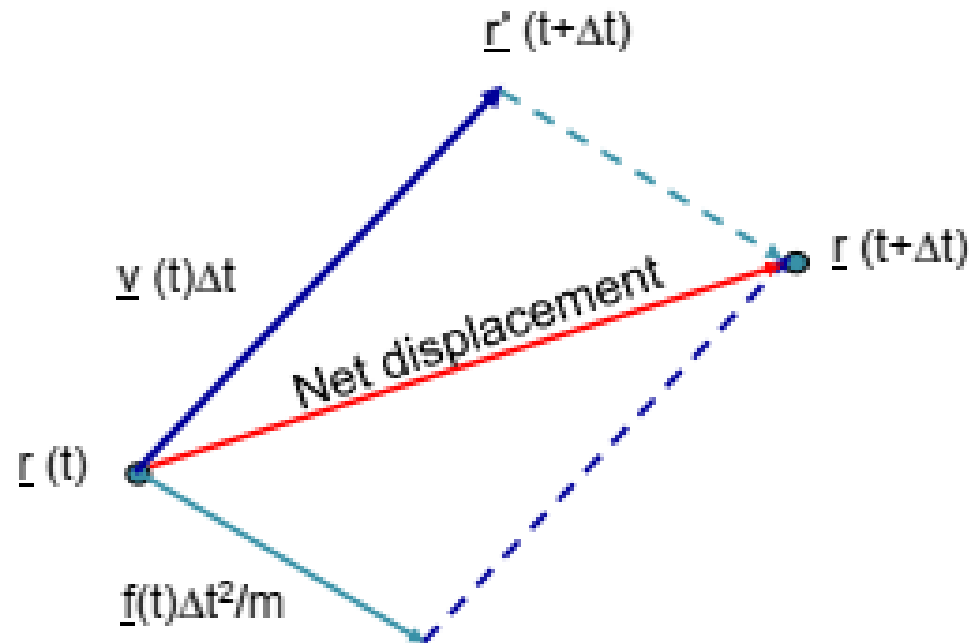
Trung bình thống kê trên tập hợp được tính bởi phương trình:

$$\langle A \rangle = \int \int_{p, r} A(p, r) \rho(p, r) dr dp$$

được ước lượng gần đúng khi rời rạc hóa quỹ đạo pha trong trạng thái vi mô của hệ:

$$\langle A \rangle; \bar{A} = \frac{1}{N} \sum_{i=1}^N A(\{r_i(i\delta t)\})$$

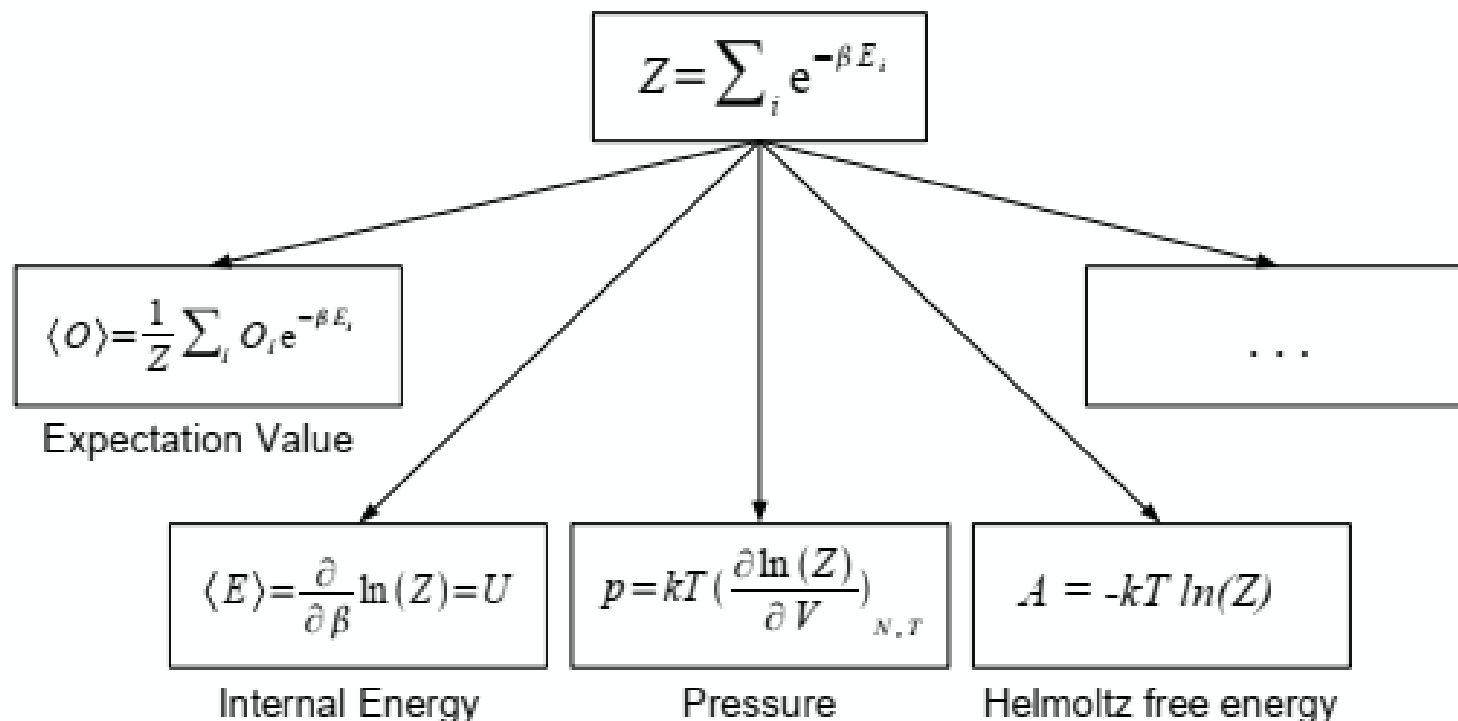
# Tích phân PT chuyển động: Verlet, leapfrog



$$[\underline{r}(t), \underline{v}(t), \underline{f}(t)] \longrightarrow [\underline{r}(t+\Delta t), \underline{v}(t+\Delta t), \underline{f}(t+\Delta t)]$$

# Vai trò của tổng thống kê

- Để đo được các đại lượng vĩ mô của một hệ, ta chỉ cần xác định được tổng thống kê  $Z$  của hệ đó (dưới góc nhìn của vật lý thống kê!)



## Xác định tổng thống kê?

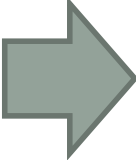
- Việc tính toán hàm tổng thống kê thường rất phức tạp, và trong hầu hết các trường hợp, chúng ta sử dụng các gần đúng

$$Z = \underbrace{\sum_i}_{2)} e^{-\beta \underbrace{E_i}_{1)}} \quad 1)$$

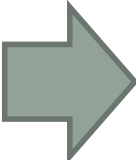
- 1) Tính năng lượng của hệ cho mỗi trạng thái vi mô  $i$  (trong mô phỏng MD, thường sử dụng các hàm thế - trường lực cho trước: CHARMM, Amber, GROMOS, ...)
- 2) Cần phương pháp để “tìm được” tất cả các trạng thái vi mô của hệ (sampling method) ứng với một phân bố cho trước của hệ vĩ mô:
  - Phân bố/tập hợp vi chính tắc ( $N, V, E = \text{const}$ )
  - Phân bố /tập hợp chính tắc ( $N, V, T = \text{const}$ )
  - Phân bố /tập hợp đẳng nhiệt - đẳng áp ( $N, P, T = \text{const}$ )
  - ...

# Rời rạc quỹ đạo pha bằng phương pháp mô phỏng MD

- **Ý tưởng:** giải số phương trình chuyển động của hạt để tạo ra một loạt các cấu hình của hệ tại các thời điểm rời rạc  $\{\vec{r}_i(\delta t)\}, \{\vec{r}_i(2\delta t)\}, \dots, \{\vec{r}_i(n\delta t)\}$ .
- Giá trị trung bình thống kê có thể tính một cách gần đúng khi lấy trung bình trên **tập hợp các cấu hình** thu được từ mô phỏng MD



Mô phỏng MD là một phương pháp tạo mẫu (sampling) và ta có thể ước lượng gần đúng các đại lượng nhiệt động mà không cần phải tích phân trên toàn bộ không gian pha.



Số cấu hình của hệ càng nhiều thì độ chính xác càng cao. Tăng mẫu, tăng độ chính xác.



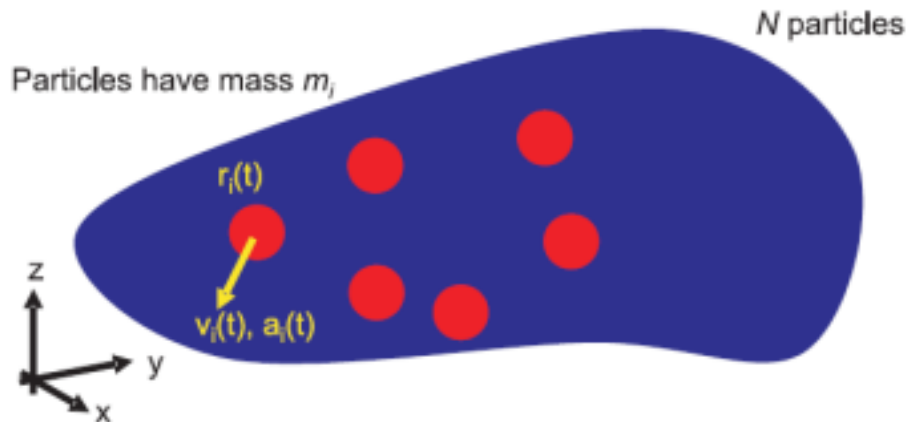
# Rời rạc quỹ đạo pha bằng phương pháp mô phỏng MD

- During integration of equations of motion – must impose thermodynamical constraints
- For example, *Verlet central difference method* leads to a microcanonical ensemble (*NVE*)
- Other integration methods exist to generate *NVT*, *NpT* ensembles etc.

$$r_i(t_0 + \Delta t) = \underbrace{-r_i(t_0 - \Delta t)}_{\text{Positions at } t_0 - \Delta t} + \underbrace{2r_i(t_0)\Delta t}_{\text{Positions at } t_0} + \underbrace{a_i(t_0)(\Delta t)^2}_{\text{Accelerations at } t_0} + \dots$$

$$a_i = f_i / m$$

# Rời rạc quỹ đạo pha bằng phương pháp mô phỏng MD



Follow trajectories of atoms  
(*classical mechanics*,  
*Newton's laws*)

*"Verlet central difference method"*

$$r_i(t_0 + \Delta t) = \underbrace{-r_i(t_0 - \Delta t)}_{\text{Positions at } t_0 - \Delta t} + \underbrace{2r_i(t_0)\Delta t}_{\text{Positions at } t_0} + \underbrace{a_i(t_0)(\Delta t)^2}_{\text{Accelerations at } t_0} + \dots$$

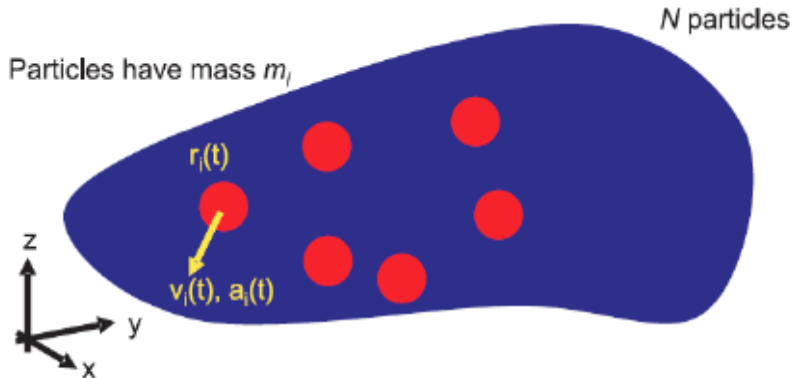
Positions  
at  $t_0 - \Delta t$

Positions  
at  $t_0$

Accelerations  
at  $t_0$

$$a_i = f_i / m$$

# Rời rạc quỹ đạo pha bằng phương pháp mô phỏng MD



Follow trajectories of atoms  
(*classical mechanics,*  
*Newton's laws*)

*"Verlet central difference method"*

$$r_i(t_0 + \Delta t) = \underbrace{-r_i(t_0 - \Delta t)}_{\text{Positions at } t_0 - \Delta t} + \underbrace{2r_i(t_0)\Delta t}_{\text{Positions at } t_0} + \underbrace{a_i(t_0)(\Delta t)^2}_{\text{Accelerations at } t_0} + \dots \quad a_i = f_i / m$$

***It's sufficient to simply  
average over all MD steps...***

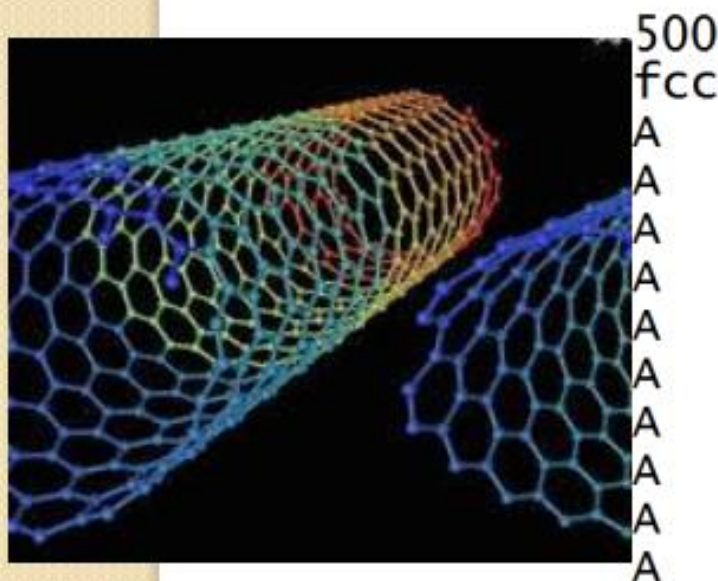
$$\langle A \rangle_{Time} = \frac{1}{N_t} \sum_{i=1..N_t} A(i)$$

Protein: Từ năng lượng đến cấu hình

# Tọa độ nguyên tử

Với hệ có N nguyên tử, vị trí của các nguyên tử được biểu diễn bởi vector ***x chứa các tọa độ*** x, y, and z của mỗi nguyên tử.

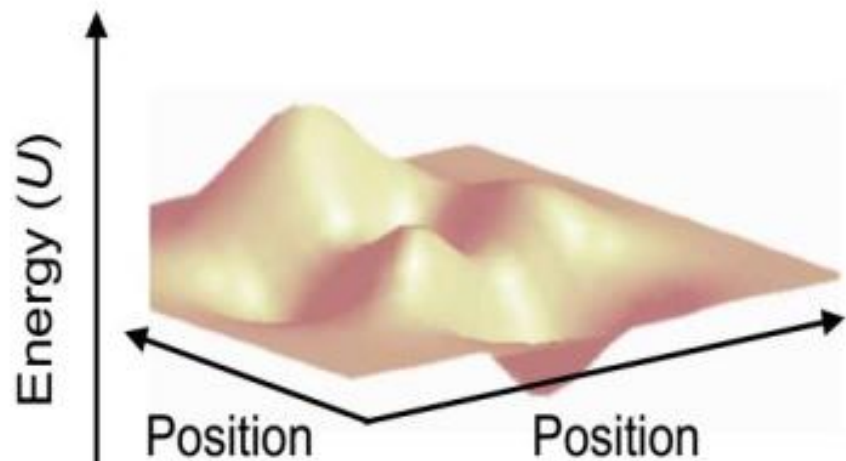
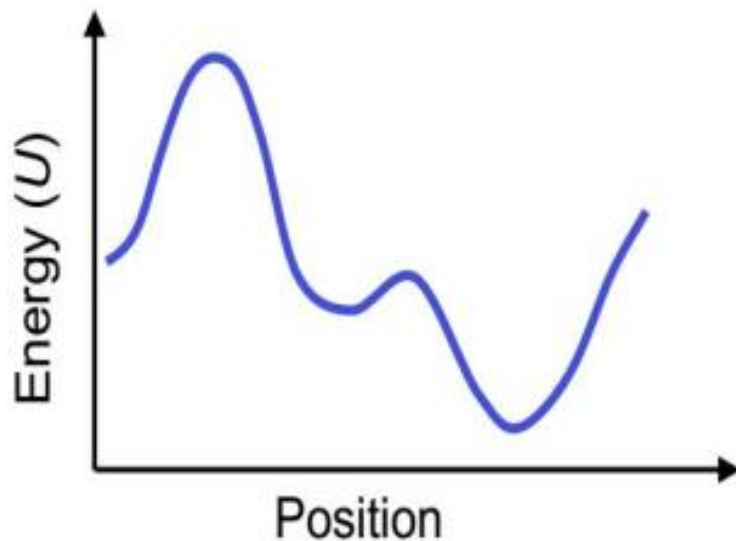
$$\mathbf{x} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ x_2 \\ y_2 \\ z_2 \\ \vdots \\ x_N \\ y_N \\ z_N \end{pmatrix}$$



A	7.215	8.064	1.484
A	6.159	5.683	2.910
A	7.306	2.762	3.461
A	8.016	1.468	4.693
A	5.436	5.978	2.149
A	7.973	5.408	7.343
A	5.907	3.338	.731
A	2.836	.866	7.721
A	6.808	4.176	7.311
A	7.882	3.194	.949

# Năng lượng

- Hàm thế  $U(\mathbf{x})$  biểu diễn thế năng tổng cho hệ các nguyên tử theo tọa độ chuyển động ( $\mathbf{x}$ ) của nguyên tử
- Hệ bao gồm protein và dung môi.

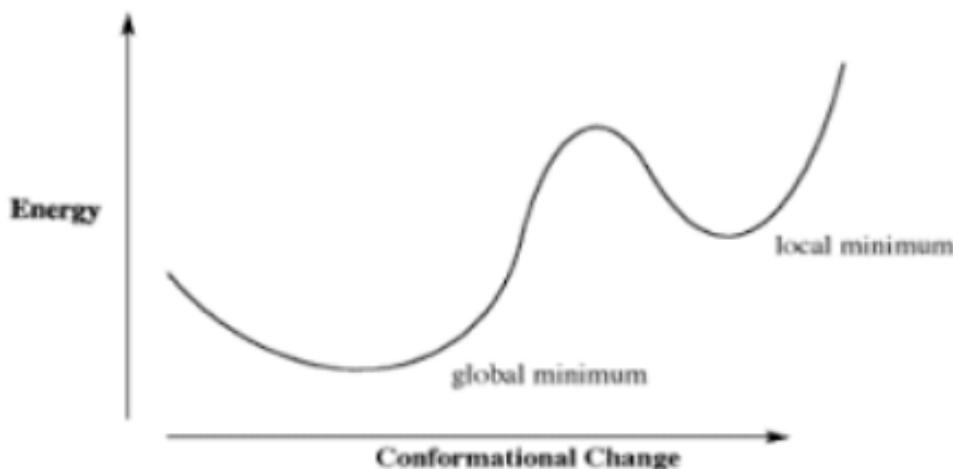


# Liên hệ giữa năng lượng và lực

- Lực tác dụng lên mỗi nguyên tử là đạo hàm của thế năng tại tọa độ nguyên tử đó:

$$\mathbf{F}(\mathbf{x}) = -\nabla U(\mathbf{x})$$

- Tại cực tiểu địa phương, lực bằng không.
- Thế năng  $U$  được gọi là trường lực (force field)



# Phân bố năng lượng của protein

- Năng lượng cho trước tương ứng với hệ tọa độ cụ thể của các nguyên tử (vị trí xác định của nguyên tử),

Vậy tọa độ của nguyên tử có theo một quy luật phân bố nào hay không???

Với:

- Hệ có nhiệt độ cố định (nguyên tử dao động xung quanh một vị trí cố định)
- Hệ đã đạt trạng thái cân bằng.

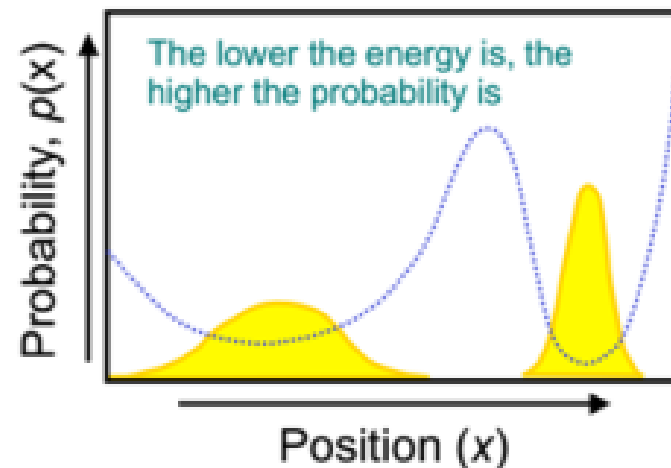
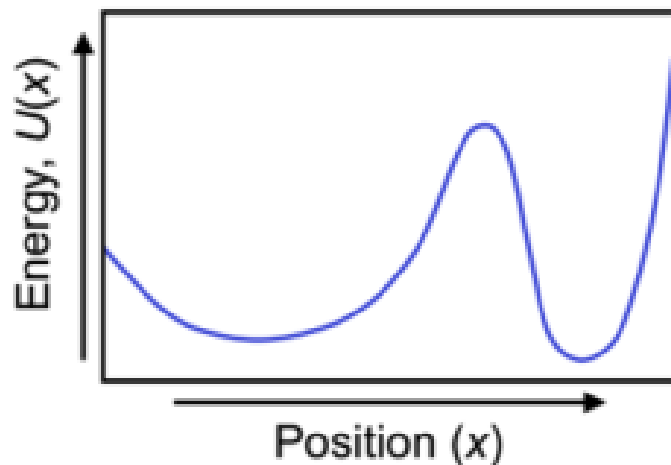


# Phân bố Boltzman

Xác suất tồn tại năng lượng của hệ theo phân bố Boltzman

$$p(\mathbf{x}) \propto \exp\left(\frac{-U(\mathbf{x})}{k_B T}\right)$$

T: nhiệt độ,  $K_B$ : hằng số Boltzman

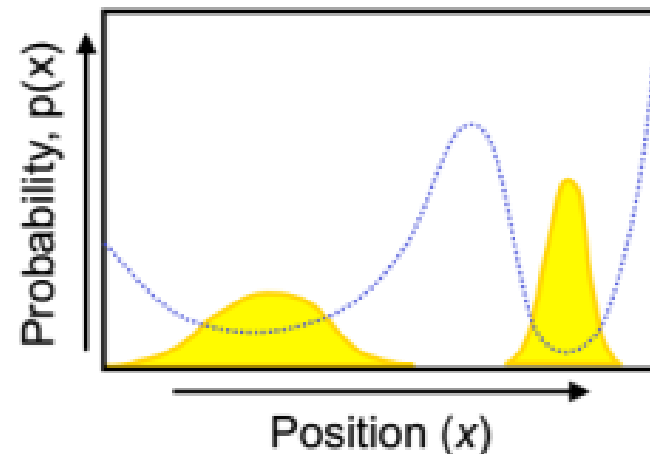
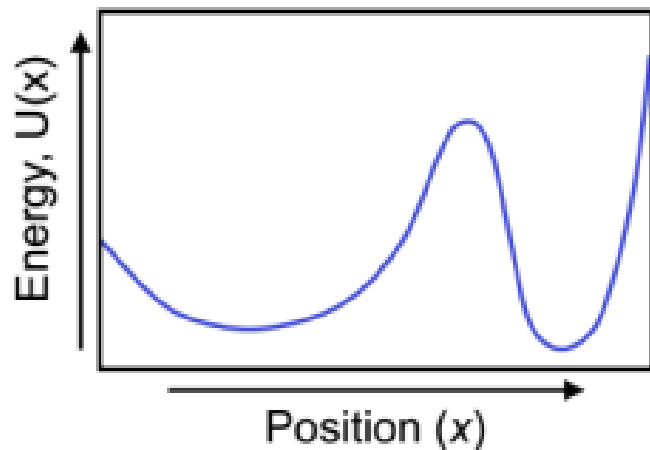


# Phân bố Boltzman

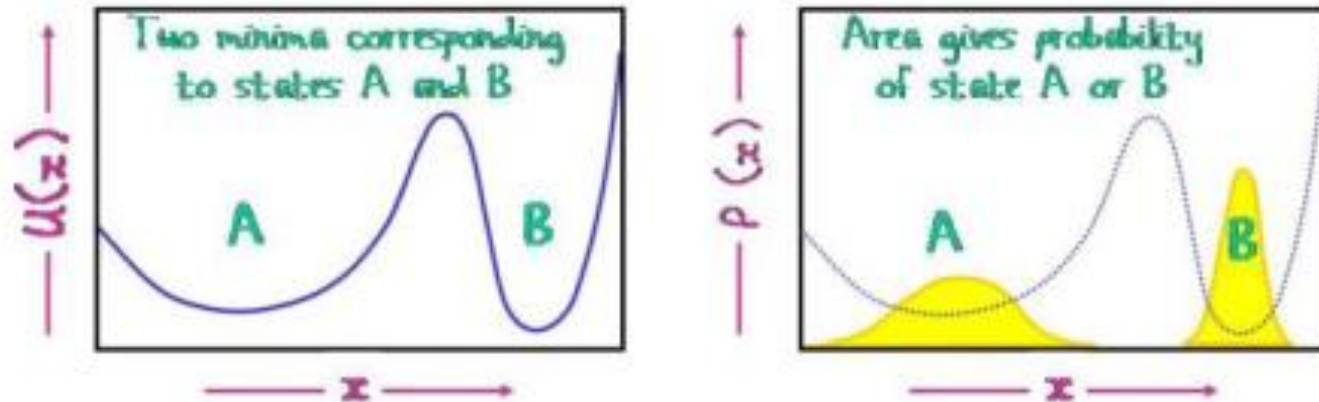
Tính chất cơ bản của phân bố Boltzman:

- Năng lượng lớn cho xác suất thấp
- Biểu diễn dưới dạng hàm mũ
- Ở nhiệt độ cao, năng lượng sẽ rất cao.

$$p(\mathbf{x}) \propto \exp\left(-U(\mathbf{x}) / k_B T\right)$$



# Phân bố Boltzman



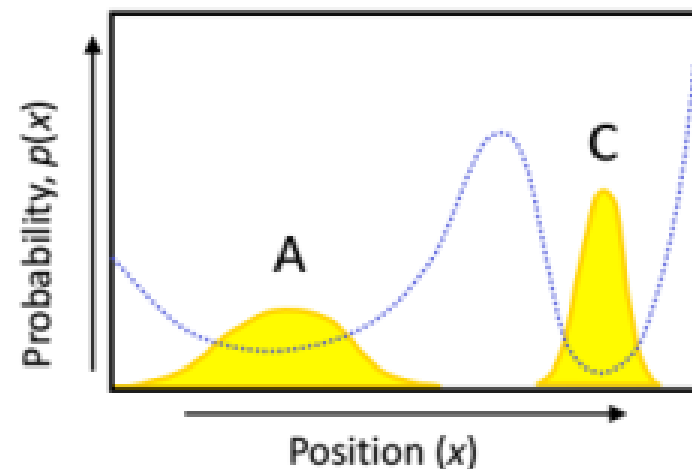
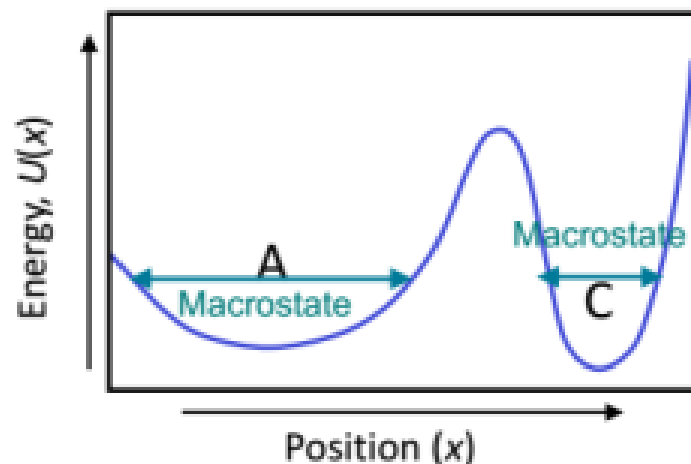
- Probability of system being at position  $x$  is  
$$P(x) = \exp(-U(x)/kT) / Q.$$
$$U(x) \text{ is Potential Energy at position } x.$$
- Find  $Q$ , the "Partition Function", so total probability is 1.  
$$Q = \sum \exp(-U(x)/kT)$$

# Cấu trúc protein

- Điều gì cần quan tâm?
  - Xác suất của các nguyên tử protein sẽ sắp xếp như thế nào trong dung môi.
  - Xác suất tương đồng trong tọa độ của các hệ nguyên tử để tạo nên các cấu hình protein giống nhau.

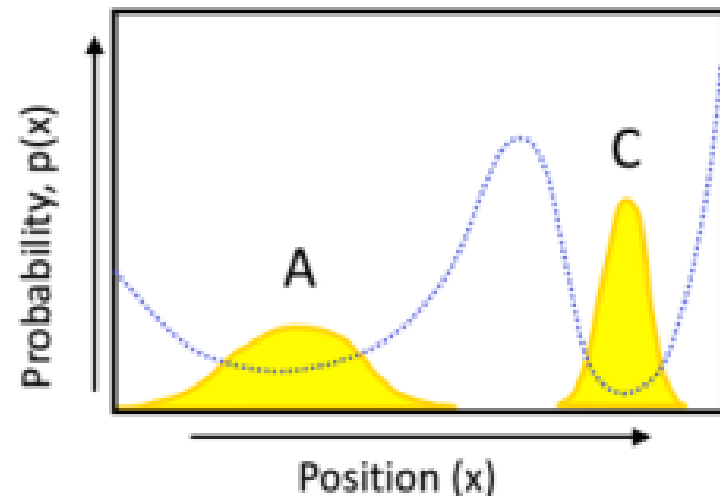
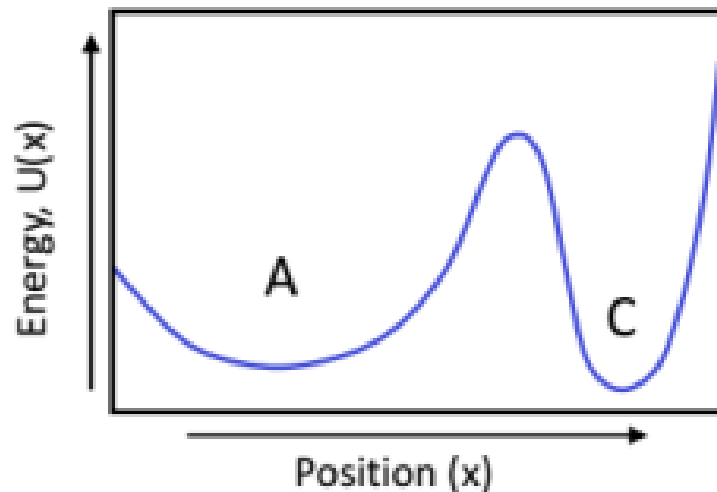
# Cấu trúc protein

- Điều chúng ta muốn là so sánh xác suất tạo nên các tập hợp của các cấu hình phổ biến của protein
- Gọi các tập hợp vĩ mô này là (A, C). Mỗi tập hợp vĩ mô gồm nhiều trạng thái vi mô.
  - Trạng thái vĩ mô đại diện cho cấu hình phổ biến của protein và có bề mặt năng lượng tự do xác định.



# Phân bố của hệ vĩ mô

- C : giếng thế sâu, nguyên tử có xu hướng “arrangement” giống nhau
- A: giếng thế rộng hơn, nguyên tử “arrangement” khác nhau.

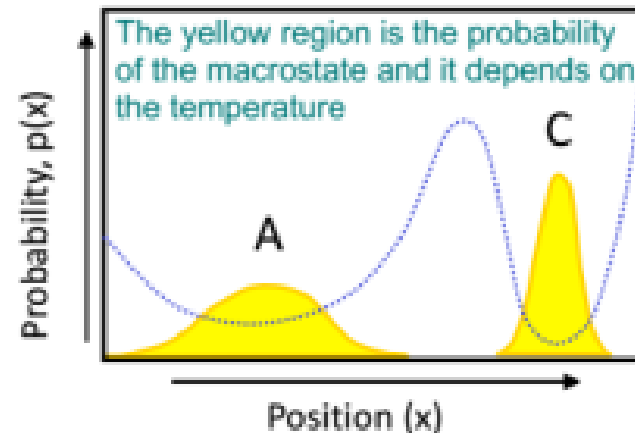
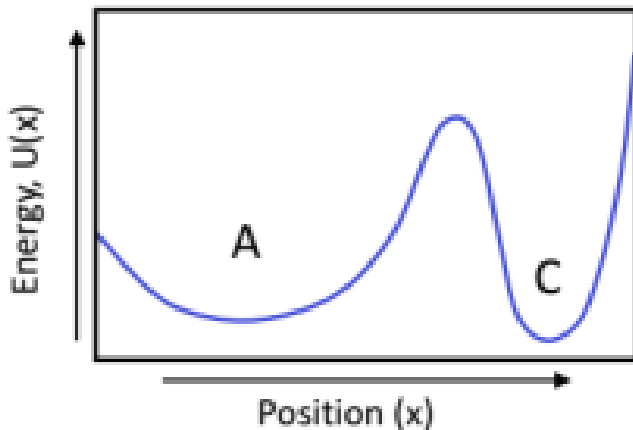


# Phân bố của hệ vĩ mô

- A hay C có xác suất cao hơn?
- Xác suất phân bố của hệ vĩ mô

$$P(A) = \int_{x \in A} P(x) \propto \int_{x \in A} \exp\left(\frac{-U(x)}{k_B T}\right) dx$$

- Ở nhiệt độ thấp,  $P(C) > P(A)$
- Ở nhiệt độ cao,  $P(A) > P(C)$



# Năng lượng tự do của hệ vĩ mô

- Năng lượng tự do  $G_A$  của hệ vĩ mô A được định nghĩa:

$$P(A) = \exp\left(-G_A / k_B T\right)$$

- Tương tự phân bố Boltzmann:

$$p(\mathbf{x}) \propto \exp\left(-U(\mathbf{x}) / k_B T\right)$$



# Năng lượng tự do của hệ vĩ mô

Năng lượng tự do  $G_A$  của hệ vĩ mô A được định nghĩa:

$$P(A) = \exp\left(-G_A / k_B T\right)$$

$$G_A = -k_B T \log_e(P(A))$$

Năng lượng tự do theo thế năng và H và entropy S)

$$G_A = H_A - TS_A$$

# So sánh cấu trúc protein

*RMSD (root mean squared distance/deviation):*

$$\begin{aligned}\text{RMSD}(\mathbf{v}, \mathbf{w}) &= \sqrt{\frac{1}{n} \sum_{i=1}^n \|v_i - w_i\|^2} \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)}\end{aligned}$$

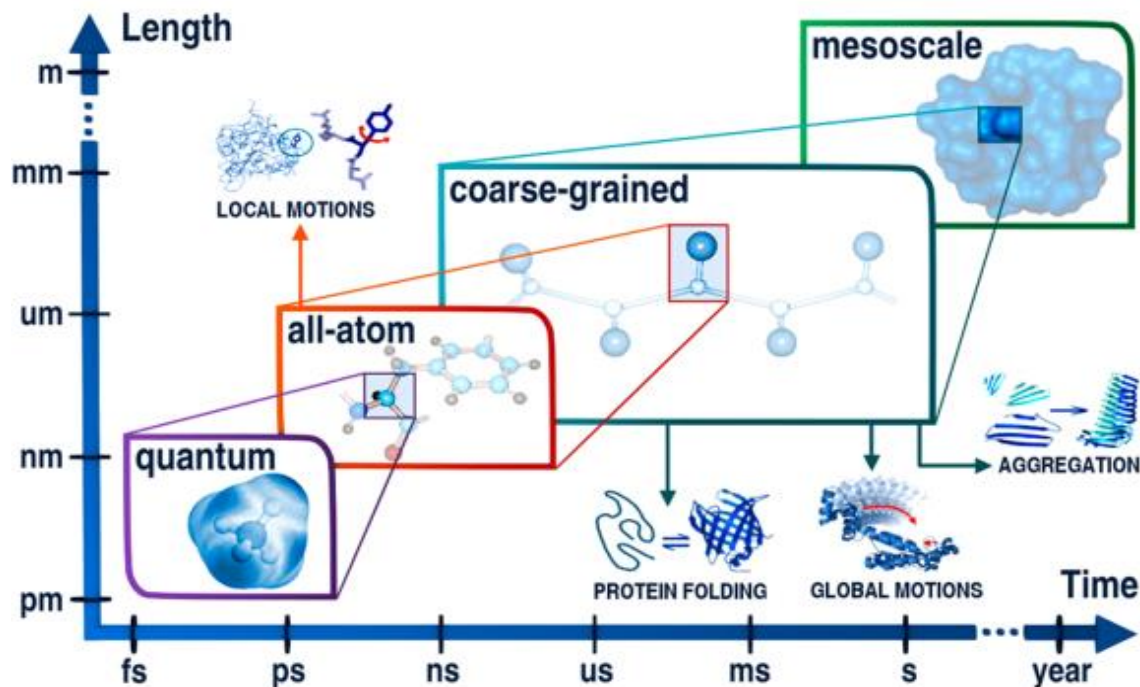
*v: tọa độ của cấu trúc ban đầu*

*w: tọa độ cấu trúc thay đổi theo thời gian*

# Thách thức cho mô phỏng hệ sinh học

- Biophysical processes involve hundreds of **thousands of atoms** often in **intricate interactions** that are difficult to **simply**.
- They span a wide range of **time scales**: enzymatic and regulatory process take ms, and structural reorganizations may exceed seconds.
- The **small driving forces** that cause molecular changes result from large, opposing energetic effect. This required careful fine-tuning of the force fields that describe interatomic interactions.

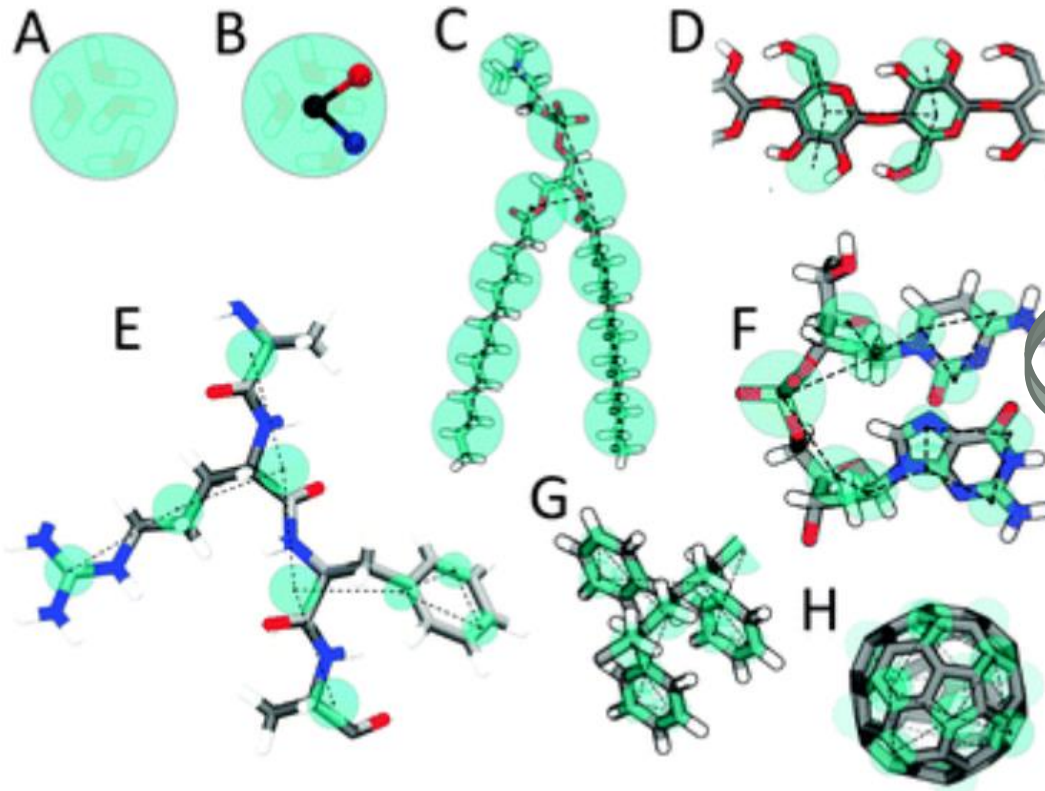
# Application ranges for molecular modeling at different resolutions



- All-atom model remains inaccessible for many protein systems.
- Lowering the level of protein representation extends the modeling opportunities to much larger simulation timescales and system sizes.

*Approximate ranges of time scales and system sizes in different resolutions*

# Mô hình thô coarse-grained



- The model uses a 4:1 mapping.
- Four main types of bead: polar, intermediate, apolar, and charged.

Each particle type has a number of subtypes, which allow for an accurate representation of the chemical nature of the underlying atomistic structure.