

# Лекция 3. Молекулярная динамика

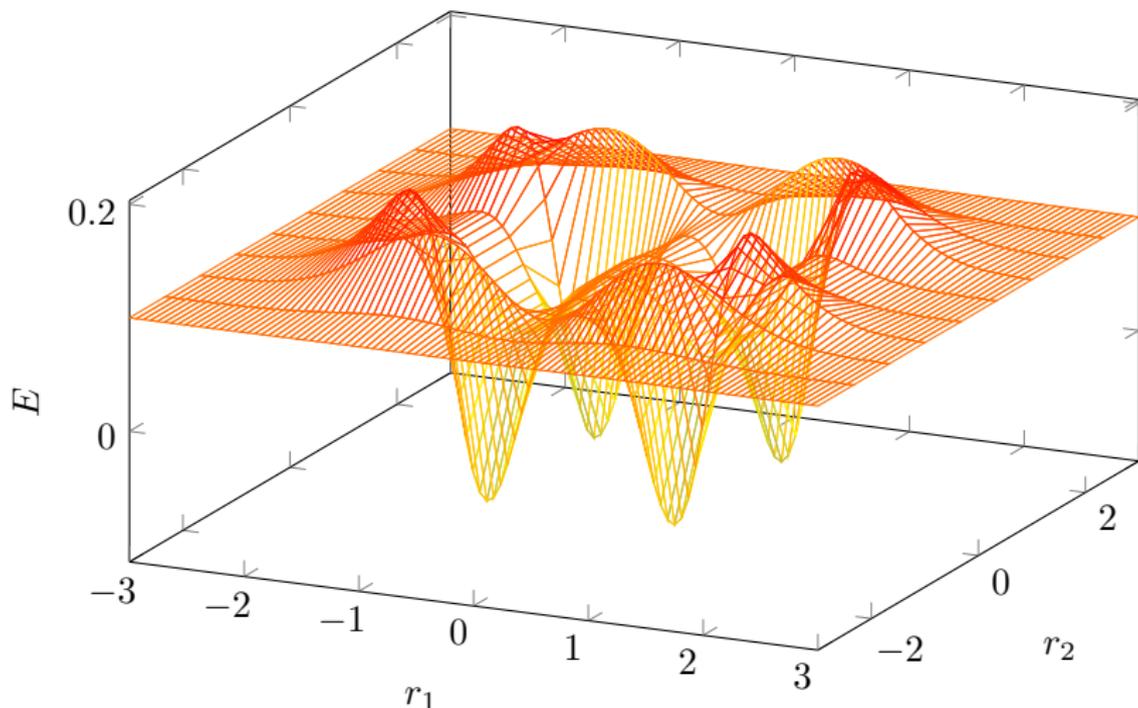
Цикл лекций "Молекулярное моделирование биомолекул и их комплексов" (НТУ Сириус)

Головин А.В.<sup>1</sup>

<sup>1</sup>МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Сириус, 2020

# Минимизация энергии и другие методы исследования поверхности потенциальной энергии



# Минимизация энергии

$$\frac{\partial f}{\partial x_i} = 0; \quad \frac{\partial^2 f}{\partial x_i^2} = 0$$

- Мы можем минимизировать функцию если мы знаем её вид.
- Функция энергии молекулярных систем как правило сложна и мы её не знаем.
- Переменных как минимум 3, это координаты.

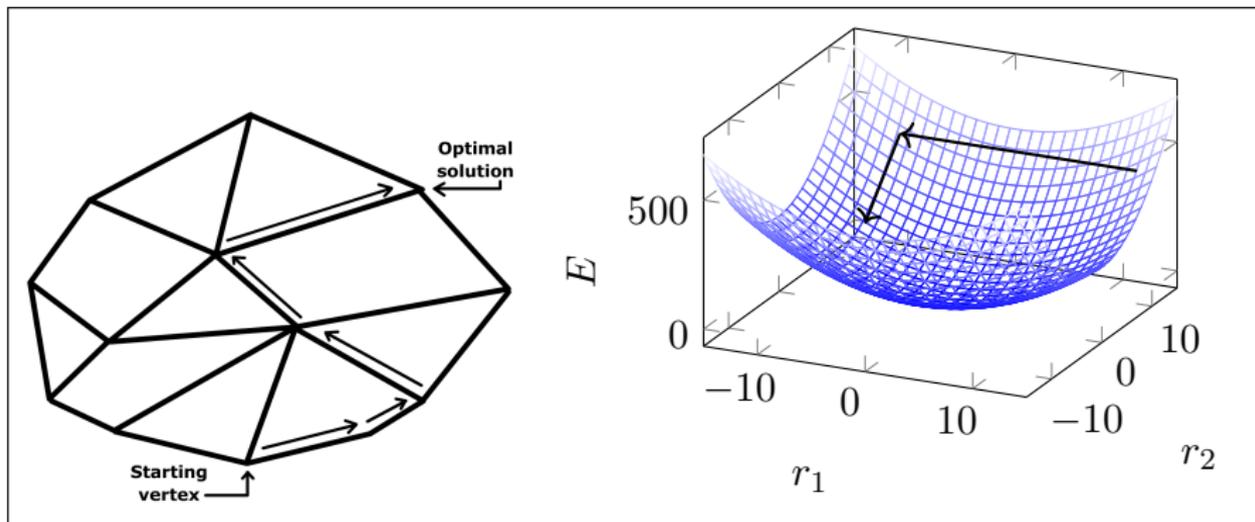
# Минимизация энергии

Два типа алгоритмов:

- Алгоритмы с использованием производных
- Алгоритмы без использования производных
  
- Использование производных может предоставить информацию о форме поверхности.
- Большинство методов минимизации энергии способны двигаться только вниз по поверхности.
  
- Не все методы одинаково эффективны для квантовых и молекулярно-механических систем. Суть разницы состоит в количестве частиц.

# Алгоритмы без использования производных

- Метод Симплекс
- Метод изменения одной переменной



# Алгоритмы с использованием производных

Представим производную как ряд Тейлора:

$$U(x) = U(x_k) + (x-x_k)U'(x_k) + (x-x_k)^2U''(x_k)/2 \dots$$

- где  $x_k$  это матрица векторов текущего состояния системы
- Каждый элемент матрицы  $U'(x_k)$  это первая производная по одной из переменных (x,y,z) и размерность матрицы  $3N \times 1$
- Каждый элемент  $i,j$  матрицы  $U''(x_k)$  это вторая производная по  $\partial x_i \partial x_j$ . Таким образом размерность матрицы  $3N \times 3N$ .
- Эта матрица называется Гессиан или матрица сил

# Алгоритмы с использованием производных первого порядка

## Алгоритм наискорейшего спуска.

Суть: движение вдоль общей силы системы.  $s_k = \frac{F_k}{\max|F_k|}$

Варианты:

### Поиск линии в одном измерении.

Ищем сечение поверхности с наименьшим значением минимуму

### Изменение шага.

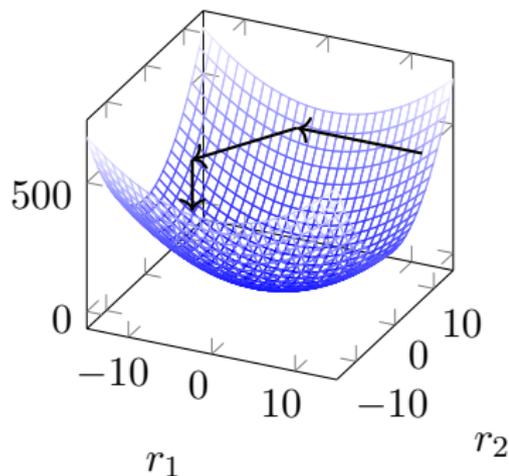
$$x_{k+1} = x_k + l_k * s_k$$

if( $V_{n+1} < V_n$ ) новые координаты приняты и

$$h_{n+1} = 1.2h_n.$$

if( $V_{n+1} \geq V_n$ ) новые координаты не приняты и

$$h_n = 0.2h_n.$$



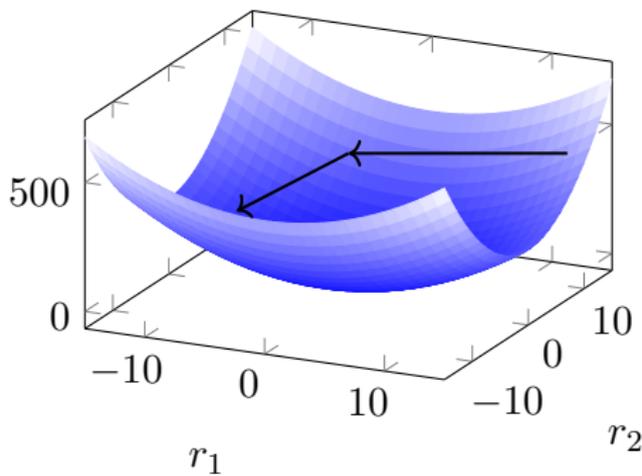
# Алгоритмы с использованием производных первого порядка

## Алгоритм сопряженных градиентов.

Суть: Двигаемся в направлении на основе направления из предыдущего шага.

$$v_k = -g_k - \gamma v_{k-1}$$

$$\gamma_k = \frac{g_k \circ g_k}{g_{k-1} \circ g_{k-1}}$$



# Алгоритмы с использованием производных второго порядка

**Метод Ньютона-Рапсона.** Касательные к функции и её производной:

$$y = U(x_k) + (x-x_k)U'(x_k) + (x-x_k)^2U''(x_k)/2 \dots$$

$$y' = xU'(x_k) + (x-x_k)U''(x_k)$$

где  $x_k$  : текущие координаты системы Если функция квадратичная, то:  $U''(x) = U''(x_k)$  и в минимуме  $U'(m) = 0$   
 $m = x_k - \frac{U'(x_k)}{U''(x_k)}$ ;  $\frac{1}{U''(x_k)}$  инвертированный Гессиан истемы

# Алгоритмы с использованием производных второго порядка

## **Квази Ньютоновские методы:**

Связи со сложностью счёта используют аппроксимации и считают обратный Гессиан только для успешных итераций.

## **Методы:**

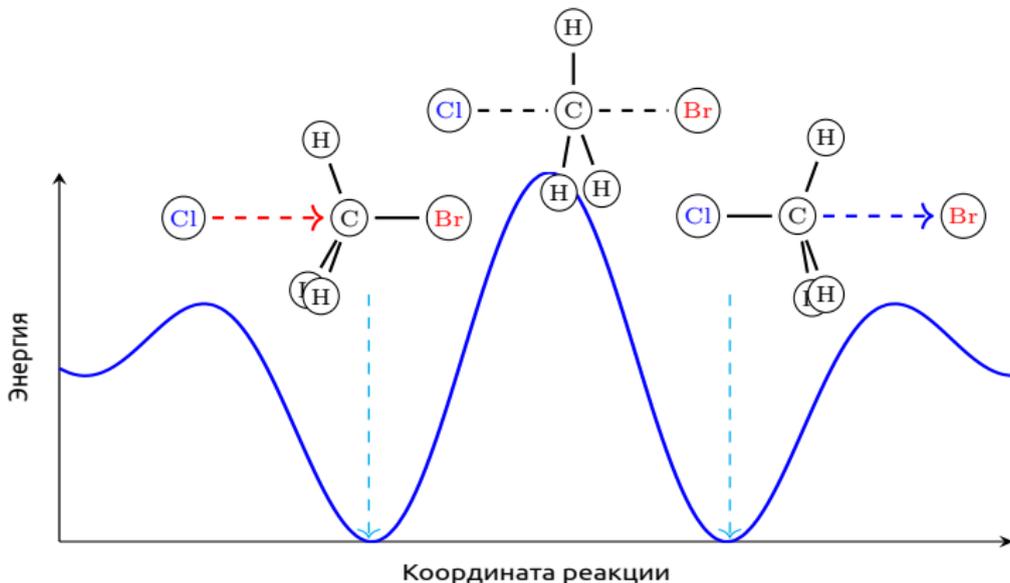
- Девидсно-Флетчер-Пауер (DFP)
- Бройден-Флетчер-Голдфарб-Шано (BFGS)
- Муртуг-Саргент (MS)

# Минимумы, максимумы и стационарные точки

- Мы обсуждаем системы, в которых  $f'(x) = 0$  необязательно может быть максимумом или минимумом, а также точкой перегиба.
- В максимуме все собственные значения Гессиана отрицательные.
- В минимуме собственные значения Гессиана либо 0, либо положительные.
- В стационарной точке не менее одного собственного значения должно быть отрицательным

# Переходные состояния

- Разница в энергии между состояниями определяет направление реакции
- Высота барьера активации определяет скорость реакции

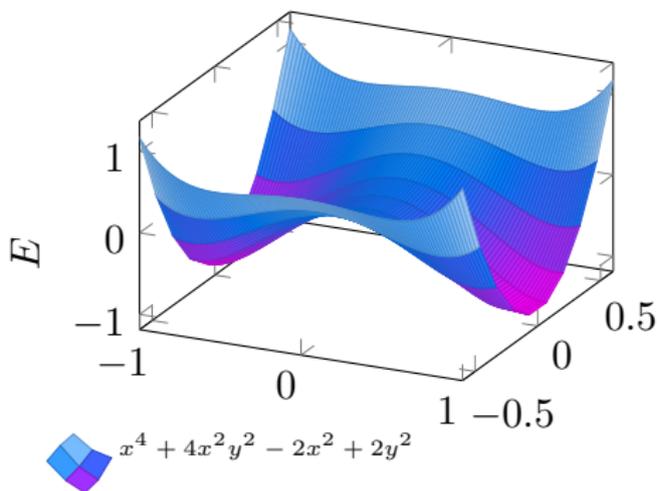


# Квадратичная область переходного состояния

- При приближении к переходному состоянию одно из собственных значений Гессиана становится отрицательным
- Это место называется квадратичной областью переходного состояния
- Большинство алгоритмов поиска переходного состояния нуждаются в структуре, расположенной в квадратичной области.

# Квадратичная область переходного состояния

Давайте представим функцию:



Гессиан:

$$\begin{pmatrix} 12x^2 + 8y^2 - 4 & 16xy \\ 16xy & 8x^2 + 4 \end{pmatrix}$$

И в точке 1,0 а в точке 0,0

$$\begin{pmatrix} 8 & 0 \\ 0 & 4 \end{pmatrix} \quad \begin{pmatrix} 4 & 0 \\ 0 & -4 \end{pmatrix}$$

# Методы поиска

- Метод сканирования поверхности энергии: генерируем координаты в округе стартового состояния и считаем энергии. Работает для малых систем.
- Методы с движением только по одной координате.
- Методы минимизации энергии с использованием первых производных могут принять переходное состояние за минимум.

## Вопросы:

- Является ли одна структура отображением состояния всех молекул вещества?
- Будут ли рассчитанные на основе структуры свойства соответствовать эксперименту?

Скорее всего нет, нам нужен ансамбль конформаций молекул при данной температуре и давлении.

# Уравнение Шредингера

$$\left( -\frac{\hbar^2}{m} \left[ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] + V \right) \Psi(r, t) = i\hbar \frac{\partial \Psi(r, t)}{\partial t}$$

Или:

$$H\Psi = E\Psi; \quad H = \frac{-\hbar^2}{m} \nabla^2 - \frac{Ze^2}{4\pi\epsilon_0 r}$$

В молекулярной механике где аппроксимируем электронную плотность уравнениями классической физики.

$$F = m \frac{\partial^2 r}{\partial t^2}$$

Осталось придумать как следить за эволюцией системы во времени.

# Простое уравнение силового поля (СП)

$$U = \sum_{bonds} \frac{k_i}{2} (l_i - l_0)^2 + \sum_{angles} \frac{k_i}{2} (\phi_i - \phi_0)^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) +$$

$$+ \sum_{i=1}^N \sum_{j=i+1}^N \left( 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)$$

Молекулярная динамика

Монте-Карло

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{F(t)}{m} \Delta t$$

$$acc(o \rightarrow \eta) = \min \left( 1, \exp \left\{ -\beta \left[ U(r^N) - U(r^N) \right] \right\} \right)$$

# Молекулярная динамика

Расчёт силы действующей на атом

$$F_i = \sum_j F_{ij}$$

$\Delta t$

Расчёт новых координат

$$\frac{\partial^2 r_i}{\partial t^2} = \frac{F_i}{m_i}$$
$$\frac{\partial r_i}{\partial t} = v_i, \quad \frac{\partial v_i}{\partial t} = \frac{F_i}{m_i}$$

# Алгоритмы интегратора

В принципе, угадать будущие координаты не просто.

Ряд Тейлора:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \frac{1}{6} \delta t^3 b(t) \dots$$

Алгоритм Верле:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t)$$

$$r(t - \delta t) = r(t) - \delta t v(t) + \frac{1}{2} \delta t^2 a(t)$$

↓

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t)$$

# Интегратор leap-frog

Leap-frog самый быстрый вариант алгоритма Верле:

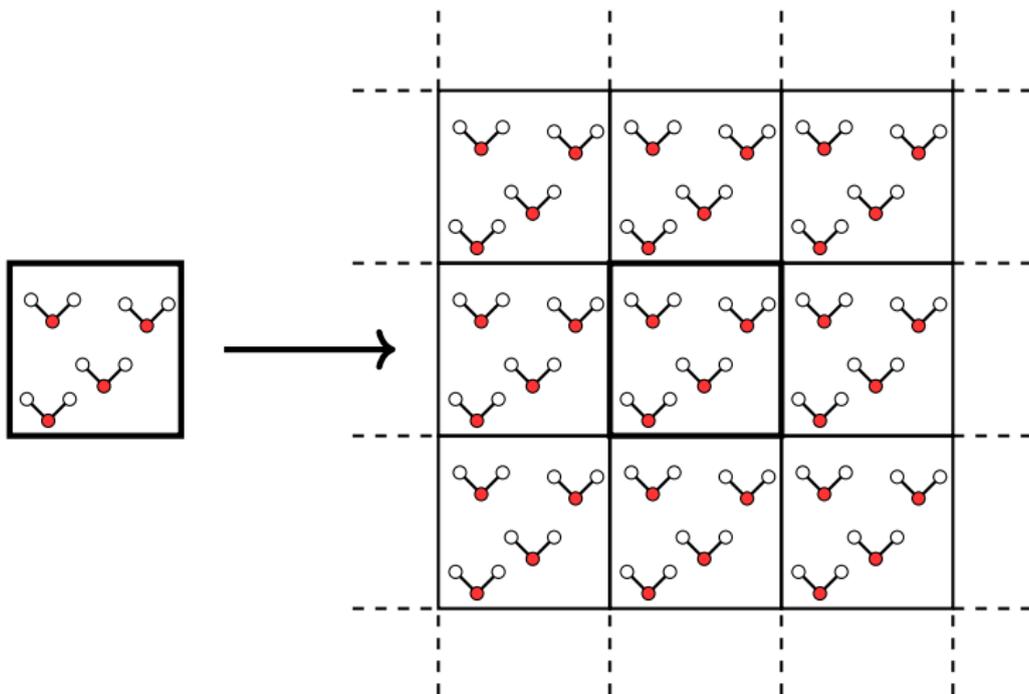
$$r(t + \delta t) = r(t) + \delta t v(t + \frac{1}{2} \delta t)$$

$$v(t + \frac{1}{2} \delta t) = v(t - \frac{1}{2} \delta t) + \delta t a(t)$$

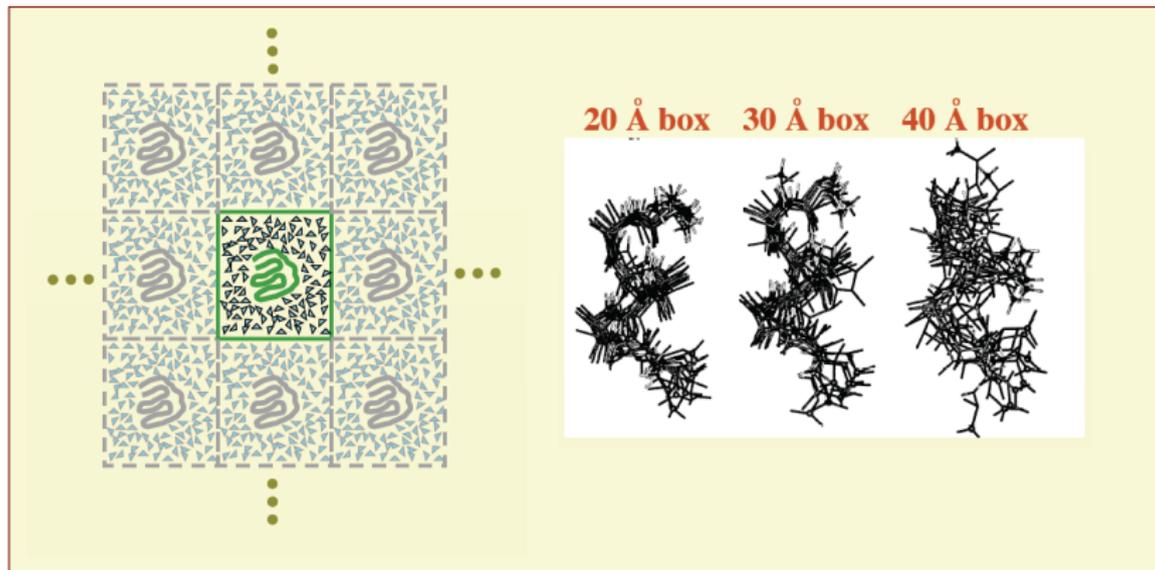
Тогда скорость в момент  $t$ :

$$v(t) = \frac{1}{2} \left[ v(t + \frac{1}{2} \delta t) + v(t - \frac{1}{2} \delta t) \right]$$

# Периодические граничные условия

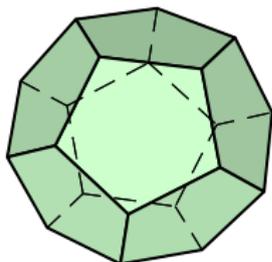
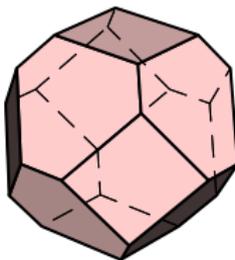
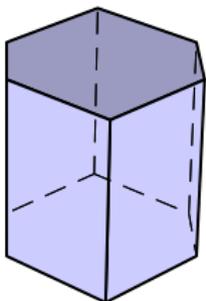
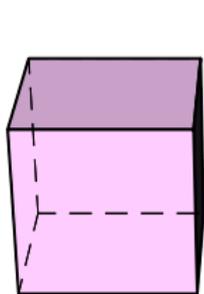
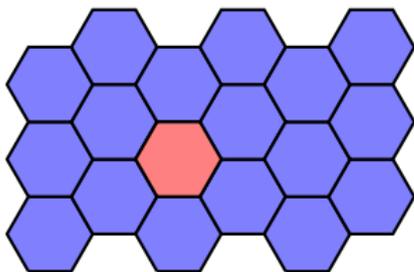
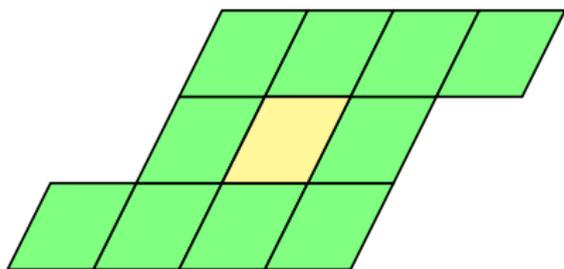


# Периодические граничные условия



МД поли-аланина показала искусственную стабилизацию альфа спирали, при использовании маленькой ячейки.

# РВС и форма ячейки



# Сферические граничные условия

Бывают системы, для которых применение периодических граничных условий неудобно:

- Капли жидкости
- Ван-дер-Ваальсовы кластеры
- Гетерогенные системы при неравновесии
- Моделирование в вакууме

# Список соседей

- Основная тяжесть счёта состоит в вычислении нековалентных взаимодействий.
- Применение обрезания непринципиально меняет скорость счёта, посчитать расстояние — это почти посчитать энергию
- В моделировании жидкостей окружение атома незначительно меняется в течение 10-20 шагов.

# Производительность

Система из 80000 атомов, компьютер Core2Quad:

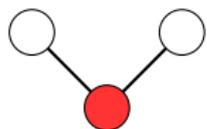
Скорость:	24 шага/сек
Время наблюдения за системой:	1 мкс
Число шагов:	$5 \cdot 10^8$
Длина шага:	2 фс.
Время симуляции:	$5 \cdot 10^8 / 24$ сек 24000 часов 1000 суток около 3 лет

Кластер (96 процессоров) примерно 25 дней, можно до 2000 процессоров.

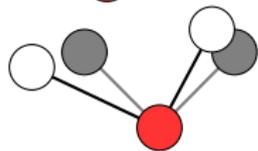
# Ограничения быстрых колебаний

Частота колебаний С-Н, N-H, O-H связей ограничивает временной шаг МД в 1 фс.

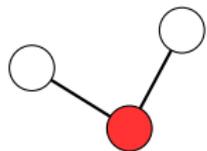
**Shake - алгоритм:**  $SHAKE(r' \rightarrow r''; r)$



Начальные координаты



Координаты после одного шага МД



После применения Shake.

$$\text{Силы} : \frac{\partial}{\partial r_i} \left( V + \sum_{k=1}^K \lambda_k \sigma_k \right)$$

## Увеличение шага интегратора МД

- Можно присвоить атому водорода массу 2 а.е. При этом отняв 1 от тяжелого атома-соседа.
- Использовать специальные конструкции. Dummies.

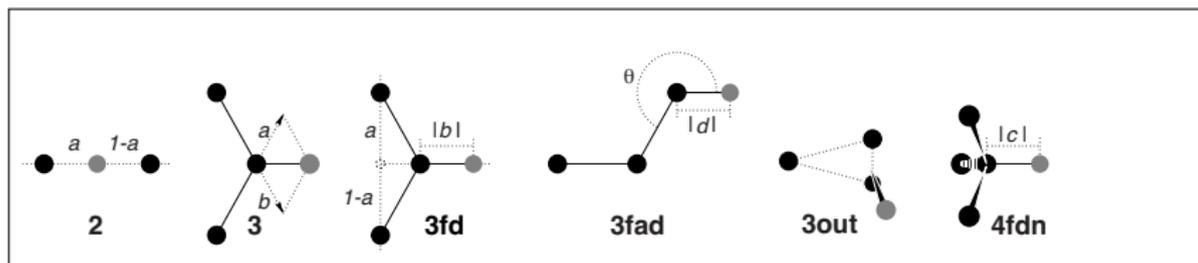
Dummies:

$$V = V(r_d, r_1, \dots, r_n) = V^*(r_1, \dots, r_n)$$

$$\mathbf{F}_i = -\frac{\partial V}{\partial r_i} = -\frac{\partial V}{\partial r_i} - \frac{\partial V}{\partial r_d} \frac{\partial r_d}{\partial r_i} = \mathbf{F}_i^{\text{direct}} + F_i'$$

$$F_i = \begin{bmatrix} \frac{\partial x_s}{\partial x_i} & \frac{\partial y_s}{\partial x_i} & \frac{\partial z_s}{\partial x_i} \\ \frac{\partial x_s}{\partial y_i} & \frac{\partial y_s}{\partial y_i} & \frac{\partial z_s}{\partial y_i} \\ \frac{\partial x_s}{\partial z_i} & \frac{\partial y_s}{\partial z_i} & \frac{\partial z_s}{\partial z_i} \end{bmatrix} F_d$$

# Конструкции атомов-пустышек в GROMACS



- Атомы - пустышки
- Реальные атомы, входящие в конструкцию

Используя атомы-пустышки, можно увеличить шаг до 5-7 фс.

# Температура

При МД часто происходит релаксация структуры и появляется излишек кинетической энергии.

$$E_{NVT} = \frac{3}{2} N k_b T$$

Самый простой способ сохранить температуру — это масштабирование скоростей

$$\frac{\partial T}{\partial t} = \frac{T_0 - T}{\tau} \quad \lambda = \left[ 1 + \frac{n_{TC} \Delta t}{\tau_T} \left\{ \frac{T_0}{T(t - \frac{1}{2} \Delta t)} - 1 \right\} \right]^{1/2}$$

$n_{TC}$  : частота,  $\lambda$  : коэффициент масштабирования

Кроме масштабирующих термостатов, существуют столкновительные термостаты и термостаты с дополнительной степенью свободы.

# Контроль давления в системе

Баростат Берендсена

$$\frac{\partial \mathbf{P}}{\partial t} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_p},$$

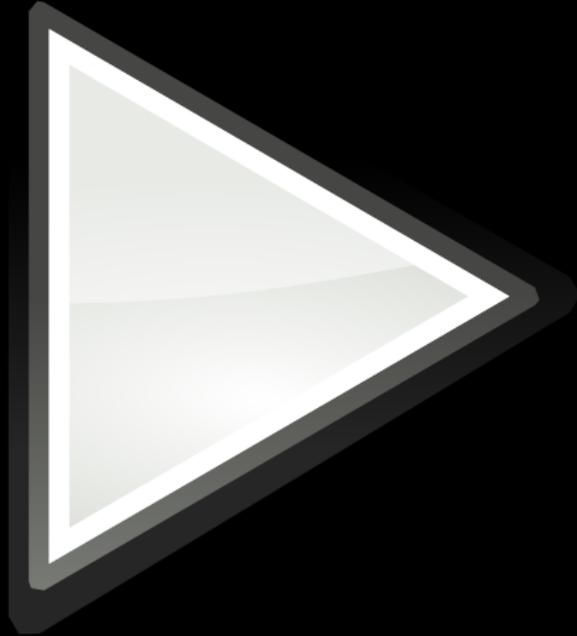
Баростат Паринелло-Рахмана

$$\frac{\partial \mathbf{b}^2}{\partial t^2} = \mathbf{V}\mathbf{W}^{-1}\mathbf{b}'^{-1} (\mathbf{P} - \mathbf{P}_{\text{ref}}).$$

$\mathbf{b}$  - матрица ветров ячейки

$\mathbf{V}$  - объём

$\mathbf{W}$  - матрица параметров определяющих силу сопряжения



# Методология подготовки системы для МД

Построение топологии молекулы на основе координат



Выбор формы и размера ячейки



Минимизация энергии структуры в вакууме методы

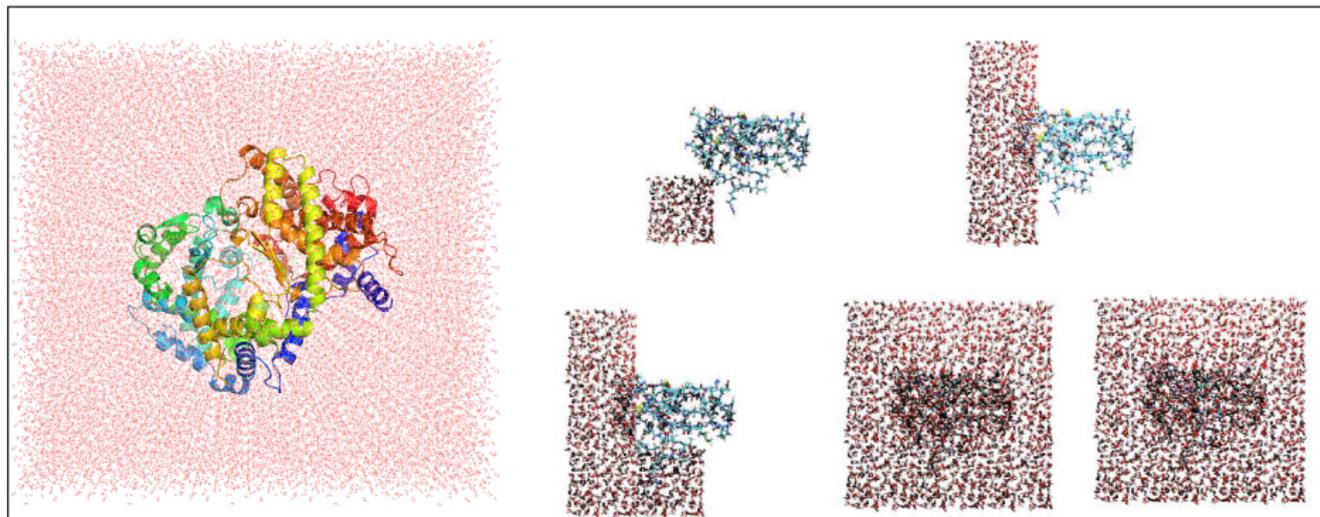


Добавление растворителя и ионов в ячейку



"Утряска" воды и ионов вокруг неподвижной молекулы

# Добавление воды в ячейку



# Молекулярная динамика и неявный растворитель

- Часто явно заданный растворитель, т. е. молекулы воды заменяют потенциалами.
- Методы на основе поверхности доступной растворителю.

$$\Delta G_{solv} = \sum_i \sigma_i ASA_i$$

- Метод Пуассона-Больцмана и его производные

# Молекулярная динамика и неявный растворитель

Метод Пуассона-Больцмана точен, но технически очень сложен для счёта.

$$\vec{\nabla} \cdot \left[ \epsilon(\vec{r}) \vec{\nabla} \Psi(\vec{r}) \right] = -4\pi \rho^f(\vec{r}) - 4\pi \sum_i c_i^\infty z_i q \lambda(\vec{r}) e^{-\frac{z_i q \Psi(\vec{r})}{kT}}$$

Generalized Born и GBSA

$$G_s = \frac{1}{8\pi} \left( \frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \sum_{i,j} \frac{q_i q_j}{f_{GB}} \quad f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2} e^{-D}$$

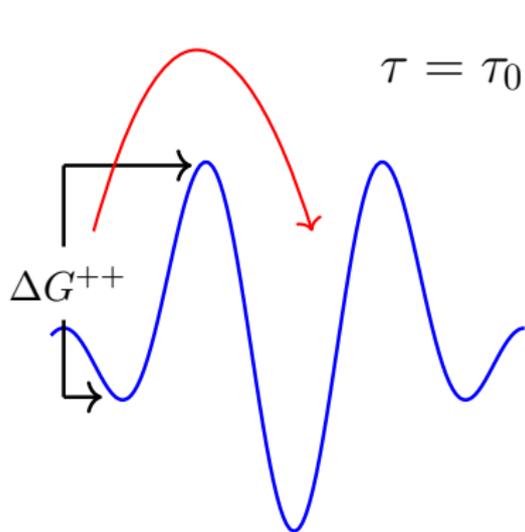
$$D = \left( \frac{r_{ij}}{2a_{ij}} \right)^2, \quad a_{ij} = \sqrt{a_i a_j}$$

# Молекулярная динамика и неявный растворитель

## Основные недостатки неявного растворителя:

- РВ и GB учитывают в основном электростатическую составляющую.
- Гидрофобный эффект не учитывается.
- Вязкость как результат столкновений и скоростей не рассчитывается и не учитывается.
- Водородные связи воды с объектом интереса не могут быть учтены.
- Исчезает возможность учёта водных мостиков.

# Длина траектории МД



$$\tau = \tau_0 e^{\frac{\Delta G^{\ddagger}}{k_b T}}$$

$\Delta G^{\ddagger}, \tau \sim 10^{-12} s$	$t_{MD}$
1 kcal/mol	$\sim 1.2$ ps
5 kcal/mol	$\sim 1.5$ ns
10 kcal/mol	$>1$ ms

# Вопросы?