

Введение в биологическую кристаллографию

Лекция 2

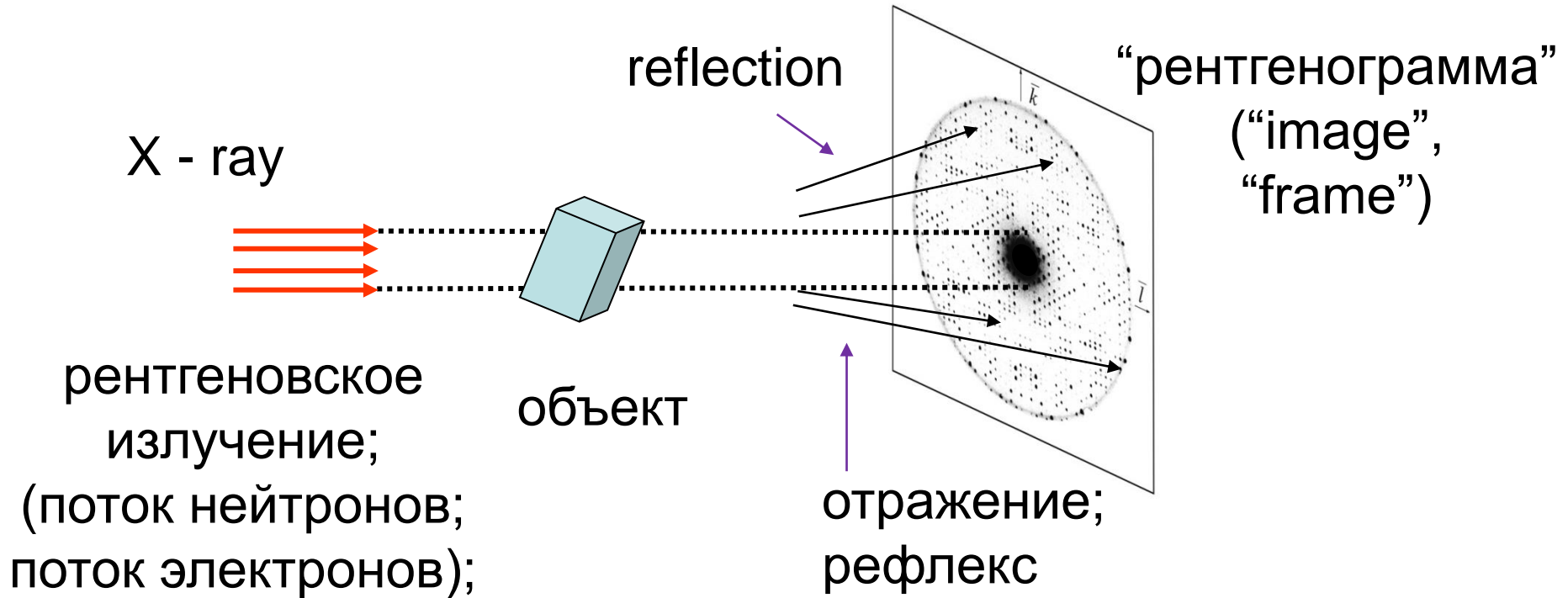
Владимир Юрьевич ЛУНИН
(lunin@impb.ru)

Институт Математических Проблем Биологии РАН

(филиал Института Прикладной Математики
им. М.В. Келдыша РАН)

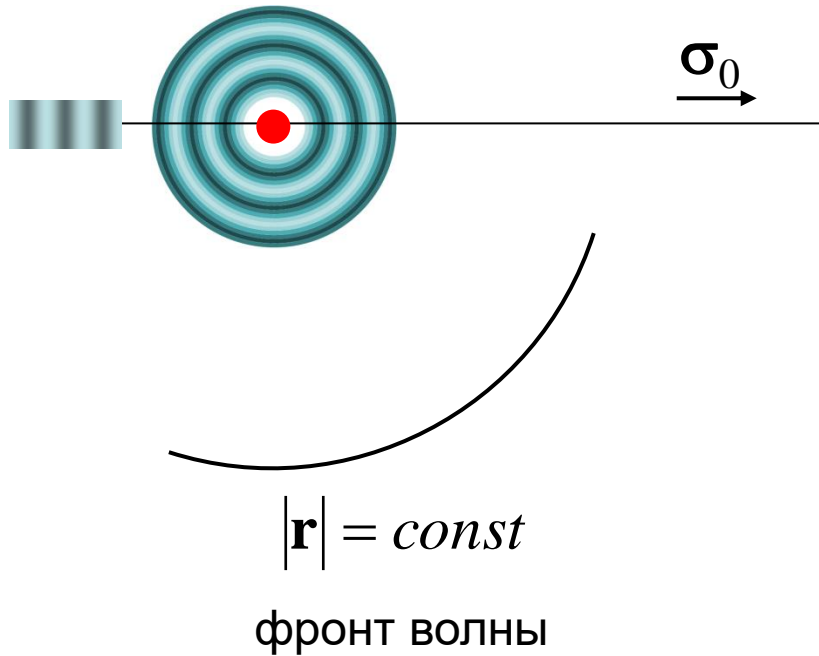
Пущино

Дифракционный эксперимент



- изучаем «отраженные» от объекта лучи;
- нет фокусирующих линз;
- вращение объекта – трехмерный набор данных;
- необходимо «восстановить» структуру объекта

Откуда берутся «отраженные» лучи?



падающая волна

$$E(\mathbf{r}, t) = E_0 \sin \left[2\pi \left(\frac{(\boldsymbol{\sigma}_0, \mathbf{r})}{\lambda} - vt + \delta \right) \right]$$



уравнения Ньютона,

осцилляция электрона



уравнения Максвелла

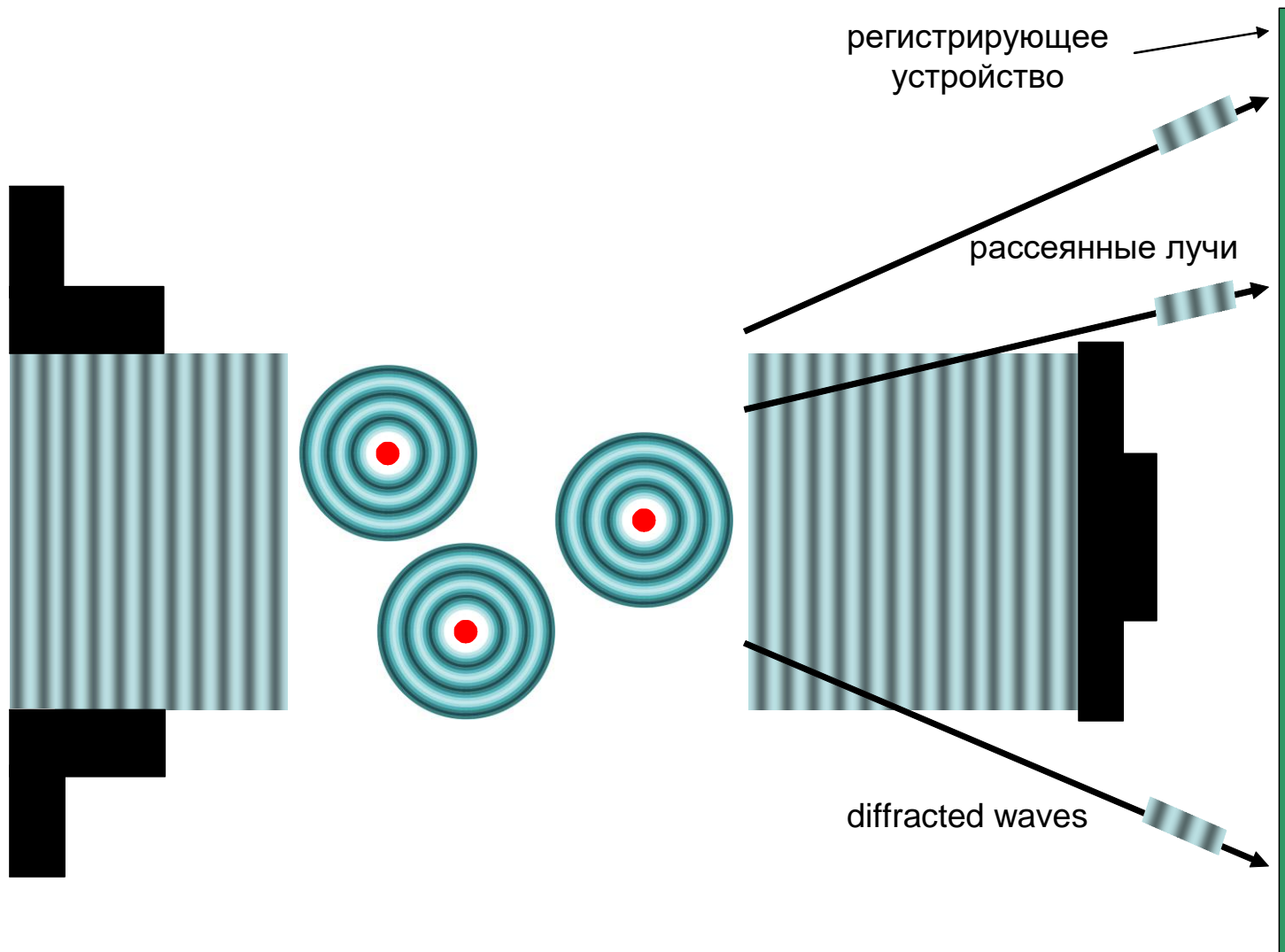
рассеянная волна

$$E(\mathbf{r}, t) = \frac{\varepsilon}{|\mathbf{r}|} E_0 \sin \left[2\pi \left(\frac{|\mathbf{r}|}{\lambda} - vt + \delta \right) \right]$$

$$\frac{\varepsilon}{|\mathbf{r}|} \approx 10^{-12}$$



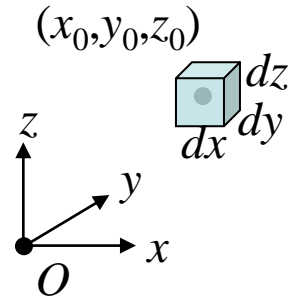
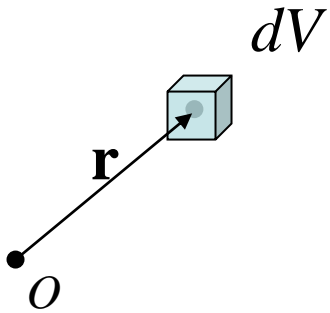
Дифракционный эксперимент



Картина дифракции определяется распределением электронов в объекте

Функция распределения электронной плотности

$\rho(\mathbf{r})dV$ (или $\rho(x_0, y_0, z_0)dxdydz$) -
 средний (по времени) заряд
 в объеме dV

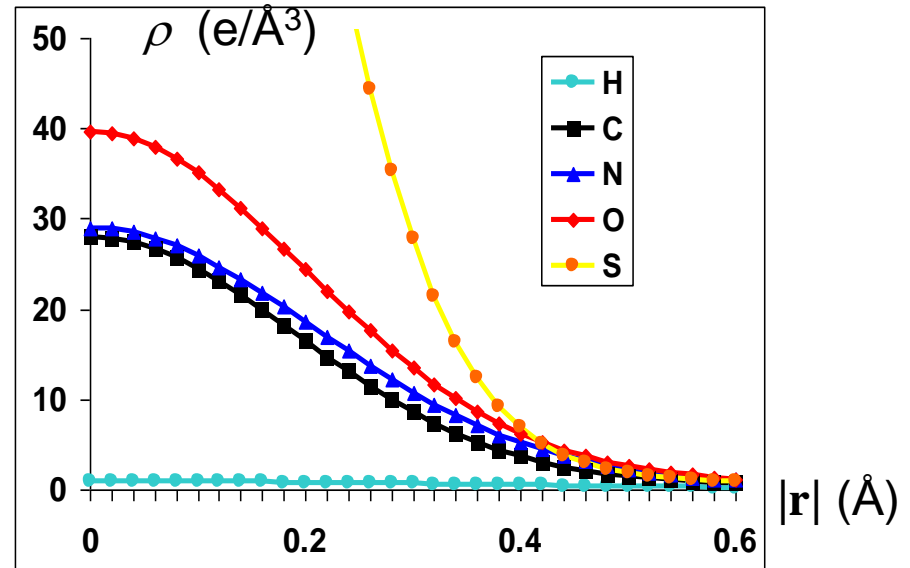


Для простых объектов можно
 найти из решения уравнения
 Шредингера

**Распределение электронной
 плотности в атоме**
 (гауссово приближение)

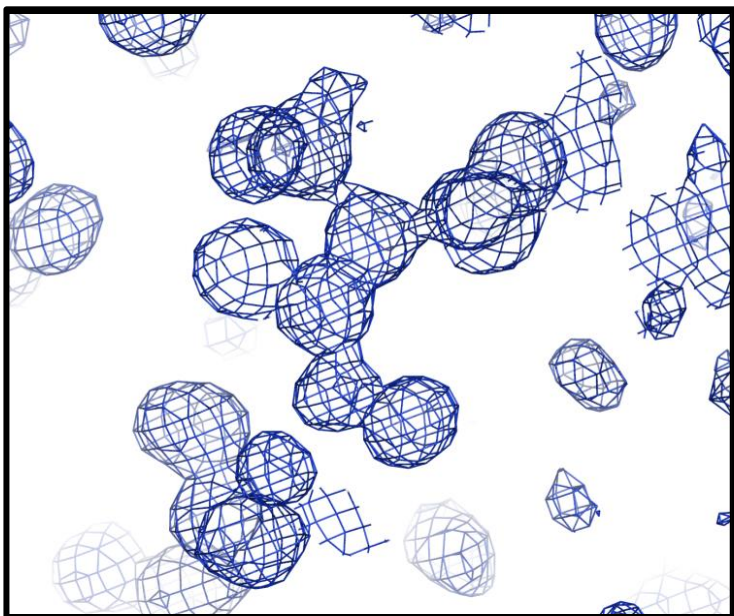
$$\rho(\mathbf{r}) \approx \alpha \exp\left(-\frac{|\mathbf{r}|^2}{\beta}\right)$$

Таблицы:
 5-гауссовое приближение



Известны для всех химических типов атомов

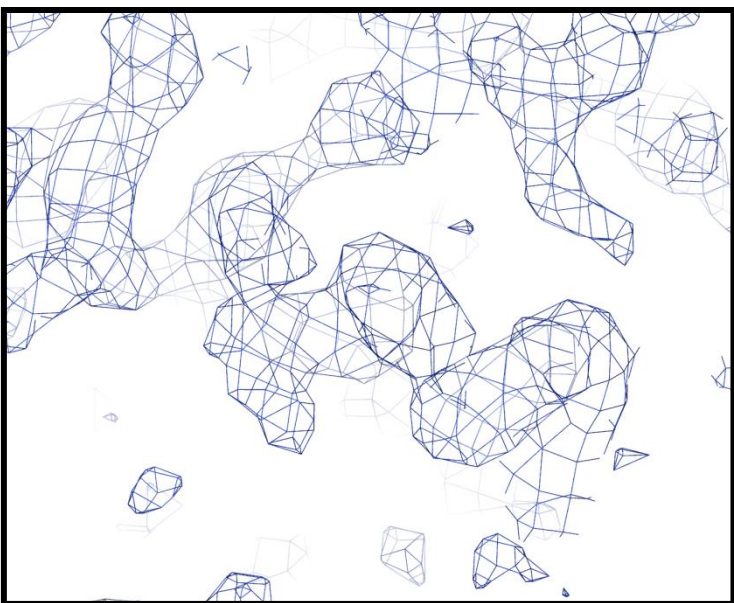
Электронная плотность в молекуле



$\{\mathbf{r}_j\}$ – координаты атомов

расчет

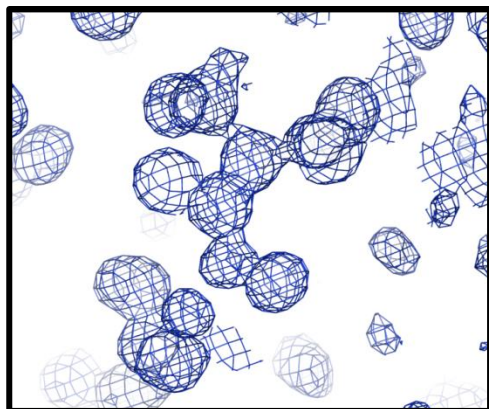
$$\leftarrow \rho(\mathbf{r}) = \sum_{j=1}^{N_{atoms}} \rho_j^{atom}(\mathbf{r} - \mathbf{r}_j)$$



интерпретация

$$\rightarrow \rho(\mathbf{r}) = \sum_{j=1}^{N_{atoms}} \rho_j^{atom}(\mathbf{r} - \mathbf{r}_j)$$

Электронная плотность в молекуле

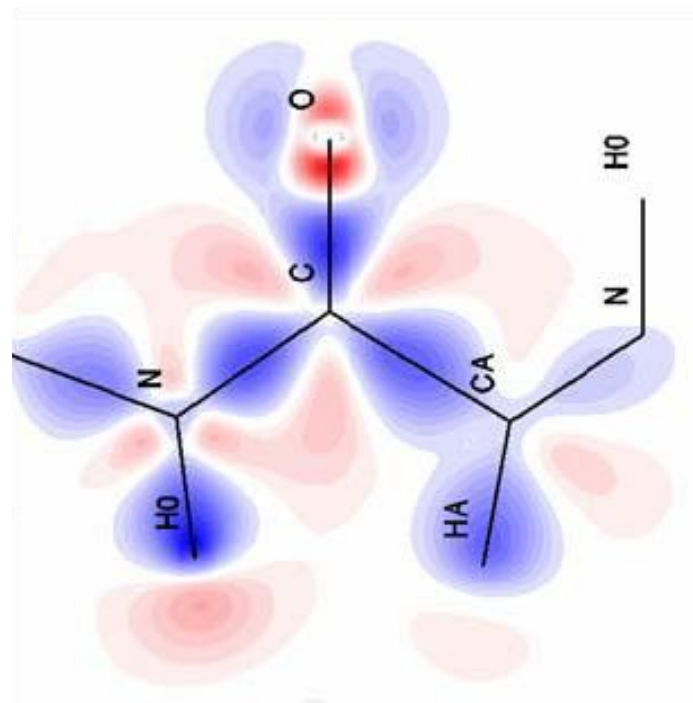
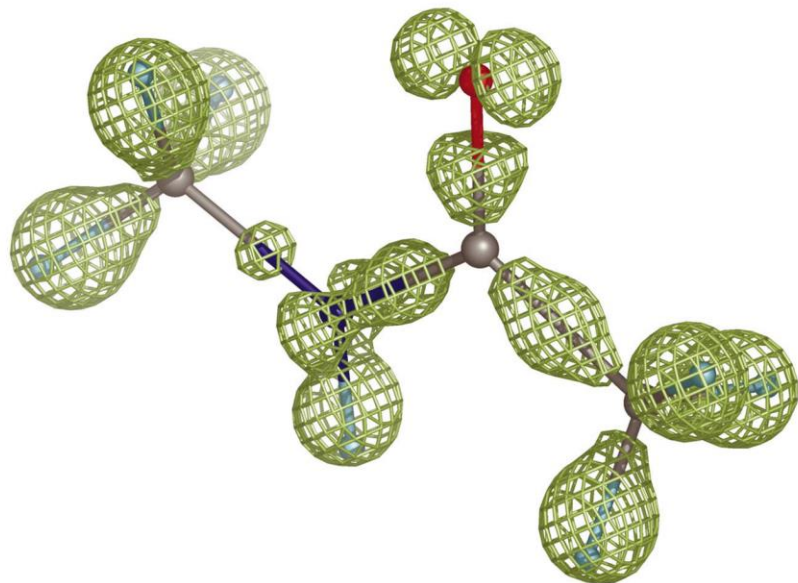


расчет («сферические,
независимые атомы»)

$$\leftarrow \rho_{sph}(\mathbf{r}) = \sum_{j=1}^{N_{atoms}} \rho_j^{atom}(\mathbf{r} - \mathbf{r}_j)$$

деформационная плотность

$$\rho_{def}(\mathbf{r}) = \rho(\mathbf{r}) - \rho_{sph}(\mathbf{r})$$



Прямая задача теории рассеяния

Знаем:

распределение электронной плотности в объекте

Хотим:

предсказать картину рассеяния (энергии волн, распространяющихся в разных направлениях)

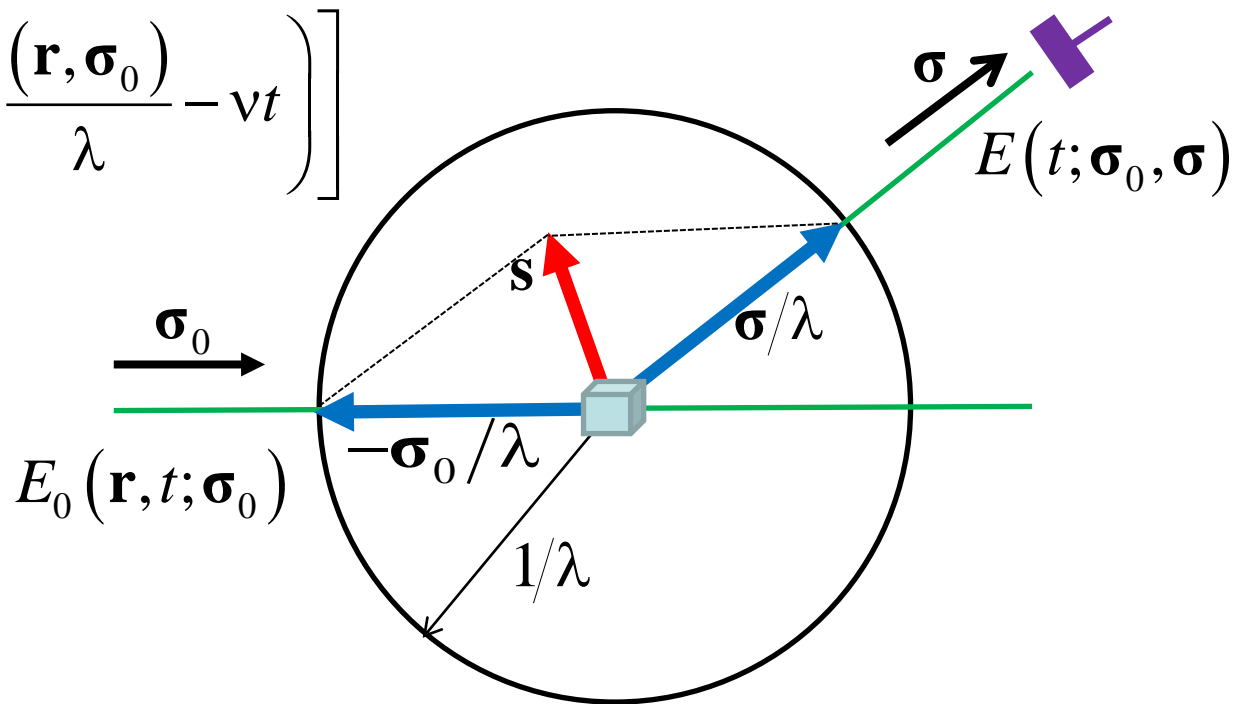
$$E_0(\mathbf{r}, t; \boldsymbol{\sigma}_0) = E_0 \sin \left[2\pi \left(\frac{(\mathbf{r}, \boldsymbol{\sigma}_0)}{\lambda} - vt \right) \right]$$

$$|\boldsymbol{\sigma}_0| = |\boldsymbol{\sigma}| = 1$$

λ - длина волны

R_{s-d} - расстояние
объект-детектор

$$\varepsilon \approx 10^{-12}$$



$$E(t; \boldsymbol{\sigma}_0, \boldsymbol{\sigma}) = E(t; \mathbf{s}) = \frac{\varepsilon}{R_{s-d}} E_0 F(\mathbf{s}) \sin \left[2\pi \left(\frac{R_{s-d}}{\lambda} - vt \right) - \varphi(\mathbf{s}) \right]$$

$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda} \quad \text{- вектор рассеяния}$$

Вся информация об объекте – в структурном факторе

$$F(\mathbf{s}), \varphi(\mathbf{s})$$

Волна на детекторе:

$$E(t; \boldsymbol{\sigma}_0, \boldsymbol{\sigma}) = E(t; \mathbf{s}) = \frac{\varepsilon}{R_{s-d}} E_0 F(\mathbf{s}) \sin \left[2\pi \left(\frac{R_{s-d}}{\lambda} - \mathbf{v}t \right) - \varphi(\mathbf{s}) \right]$$

Зависит от комбинации $\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda}$ - вектор рассеяния

Вся информация об объекте – в структурном факторе

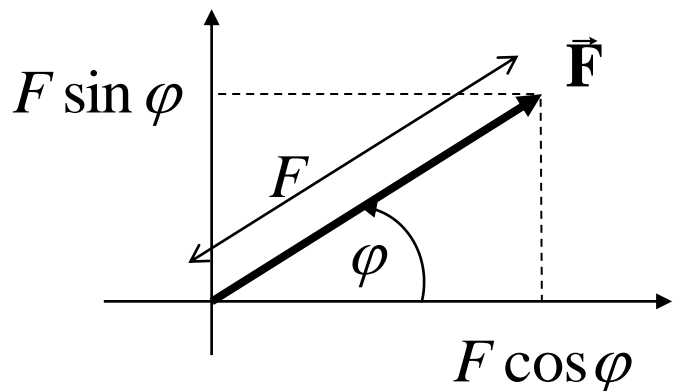
$$F(\mathbf{s}), \varphi(\mathbf{s})$$

$$A = F(\mathbf{s}) \cos \varphi(\mathbf{s}) =$$

$$B = F(\mathbf{s}) \sin \varphi(\mathbf{s}) =$$

$$\int \rho(\mathbf{r}) \cos 2\pi(\mathbf{s}, \mathbf{r}) dV_{\mathbf{r}}$$

$$\int \rho(\mathbf{r}) \sin 2\pi(\mathbf{s}, \mathbf{r}) dV_{\mathbf{r}}$$



$$F(\mathbf{s}) = \sqrt{A^2 + B^2}$$

$$\varphi(\mathbf{s}) = \operatorname{arctg} \frac{B}{A}$$

решение
прямой
задачи

Эксперимент

Волна на детекторе:

$$E(t; \boldsymbol{\sigma}_0, \boldsymbol{\sigma}) = E(t; \mathbf{s}) = \frac{\varepsilon}{R_{s-d}} E_0 F(\mathbf{s}) \sin \left[2\pi \left(\frac{R_{s-d}}{\lambda} - \mathbf{v}t \right) - \varphi(\mathbf{s}) \right]$$

Измеряется интенсивность (энергия) рассеянных волн

$$I(\mathbf{s}) = \left(\frac{\varepsilon E_0}{R_{s-d}} \right)^2 F^2(\mathbf{s}) \Delta t \quad I(\mathbf{s}) \propto F^2(\mathbf{s})$$

Эксперимент дает набор величин модулей структурных факторов $F(\mathbf{s})$ для разных векторов рассеяния \mathbf{s} .

Информация о фазах $\varphi(\mathbf{s})$ в «стандартном» эксперименте теряется.

Прямая задача теории рассеяния:

Знаем: $\rho(\mathbf{r})$. **Предсказываем:** $I(\mathbf{s}) \propto F^2(\mathbf{s})$ для разных \mathbf{s} .

Обратная задача теории рассеяния:

Знаем: $F^{obs}(\mathbf{s})$ для разных \mathbf{s} . **Хотим найти** $\rho(\mathbf{r})$.

Метод проб и ошибок

$$\mathbf{r}_j, j = 1, \dots, N$$

координаты атомов

$$\rho(\mathbf{r}) = \sum_j \rho_{type_j}(\mathbf{r} - \mathbf{r}_j)$$

электронная плотность

$$F^{calc}(\mathbf{s}_k), k = 1, \dots, M$$

структурные факторы

$$Q(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_k \left(F^{calc}(\mathbf{s}_k; \mathbf{r}_1, \dots, \mathbf{r}_N) - F^{obs}(\mathbf{s}_k) \right)^2 \quad \text{критерий}$$

Уточнение структуры

$\mathbf{r}_j, j = 1, \dots, N$ приближенная модель; координаты атомов

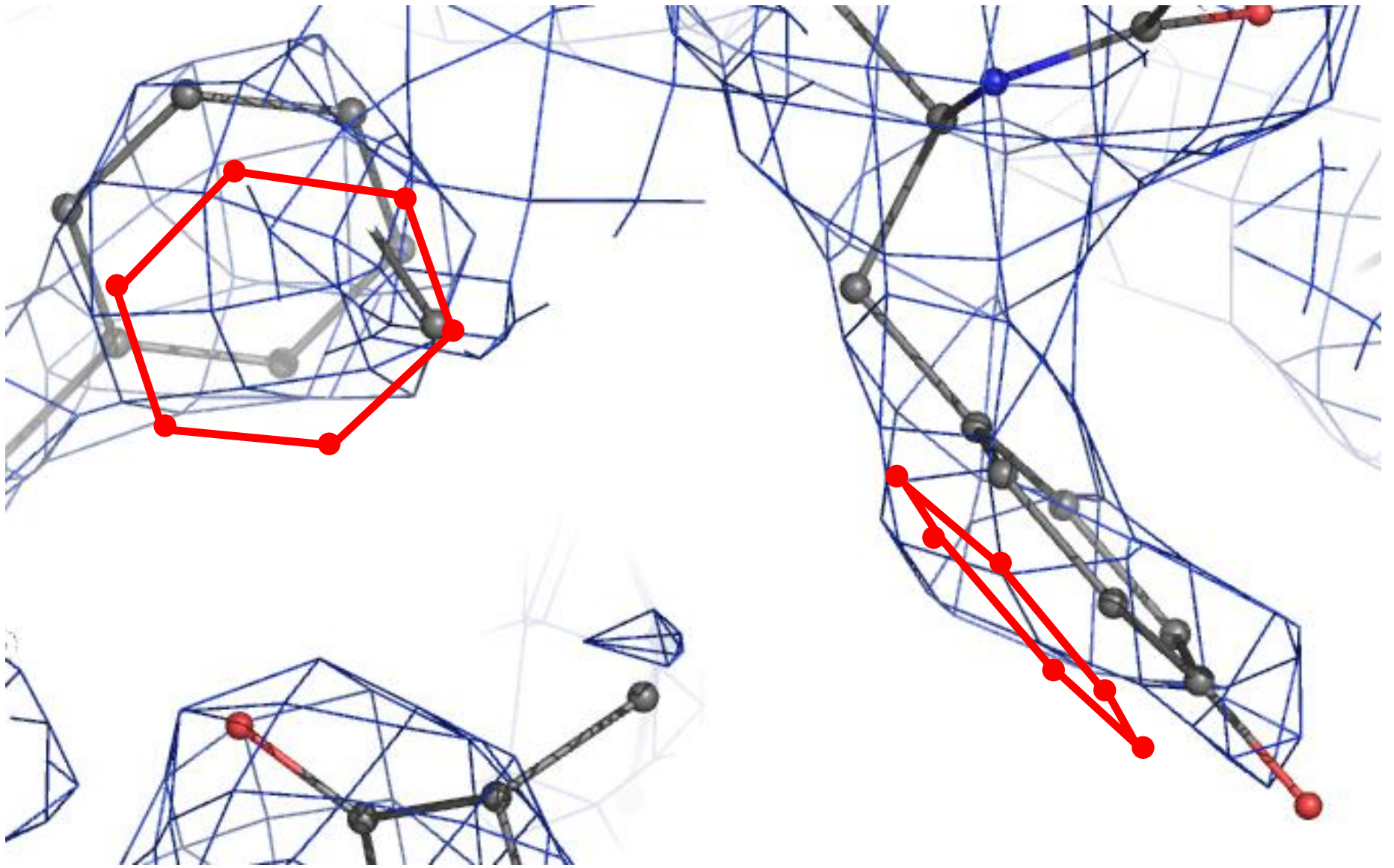
ищем **небольшие** поправки $\Delta_j, j = 1, \dots, N$ такие, что

$$Q(\Delta_1, \dots, \Delta_N) = \sum_k \left(F^{calc}(\mathbf{s}_k; \mathbf{r}_1 + \Delta_1, \dots, \mathbf{r}_N + \Delta_N) - F^{obs}(\mathbf{s}_k) \right)^2 \Rightarrow \min$$

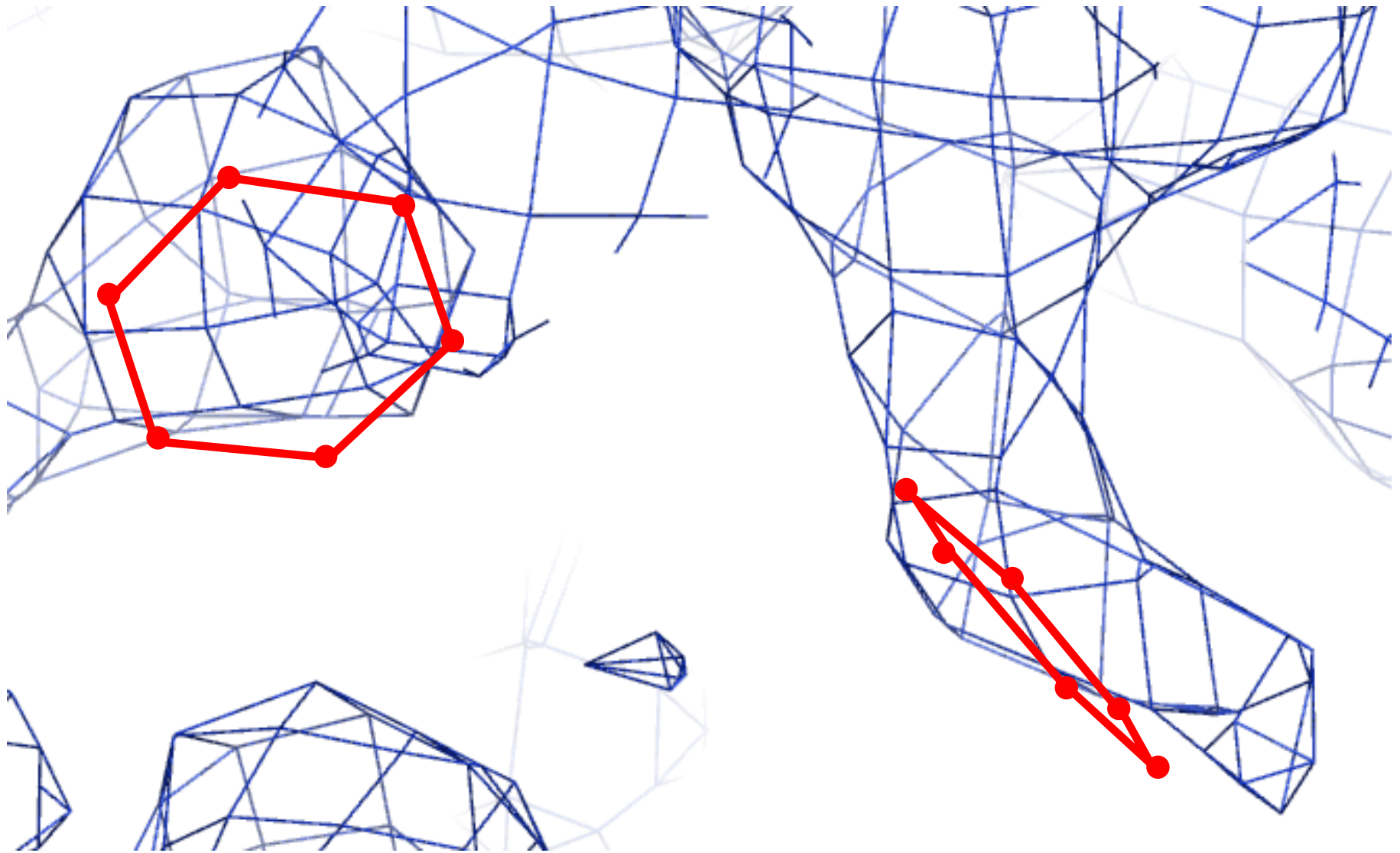
$F^{calc}(\mathbf{s}_k), k = 1, \dots, M$ структурные факторы

$\rho(\mathbf{r}) = \sum_j \rho_{type_j}(\mathbf{r} - (\mathbf{r}_j + \Delta_j))$ электронная плотность

Современные компьютеры умеют решать такую задачу, если ошибки Δ_j в координатах атомов находятся в пределах 0.7 Å.



RNase sa, 2.5Å, MIR



RNase sa, 2.5Å, MIR

$$I(\mathbf{s}) = \left(\frac{\varepsilon E_0}{R_{s-d}} \right)^2 F^2(\mathbf{s}) \Delta t \quad \varepsilon^2 \approx 10^{-24} \quad !!!$$

Проблема

Интенсивность рассеяния отдельной молекулой слишком мала для регистрации.

Возможное решение:

- увеличение мощности источника излучения;
- повышение чувствительности регистрирующего устройства;
- увеличение времени экспозиции;
- рассеяние большим числом идентичных молекул.

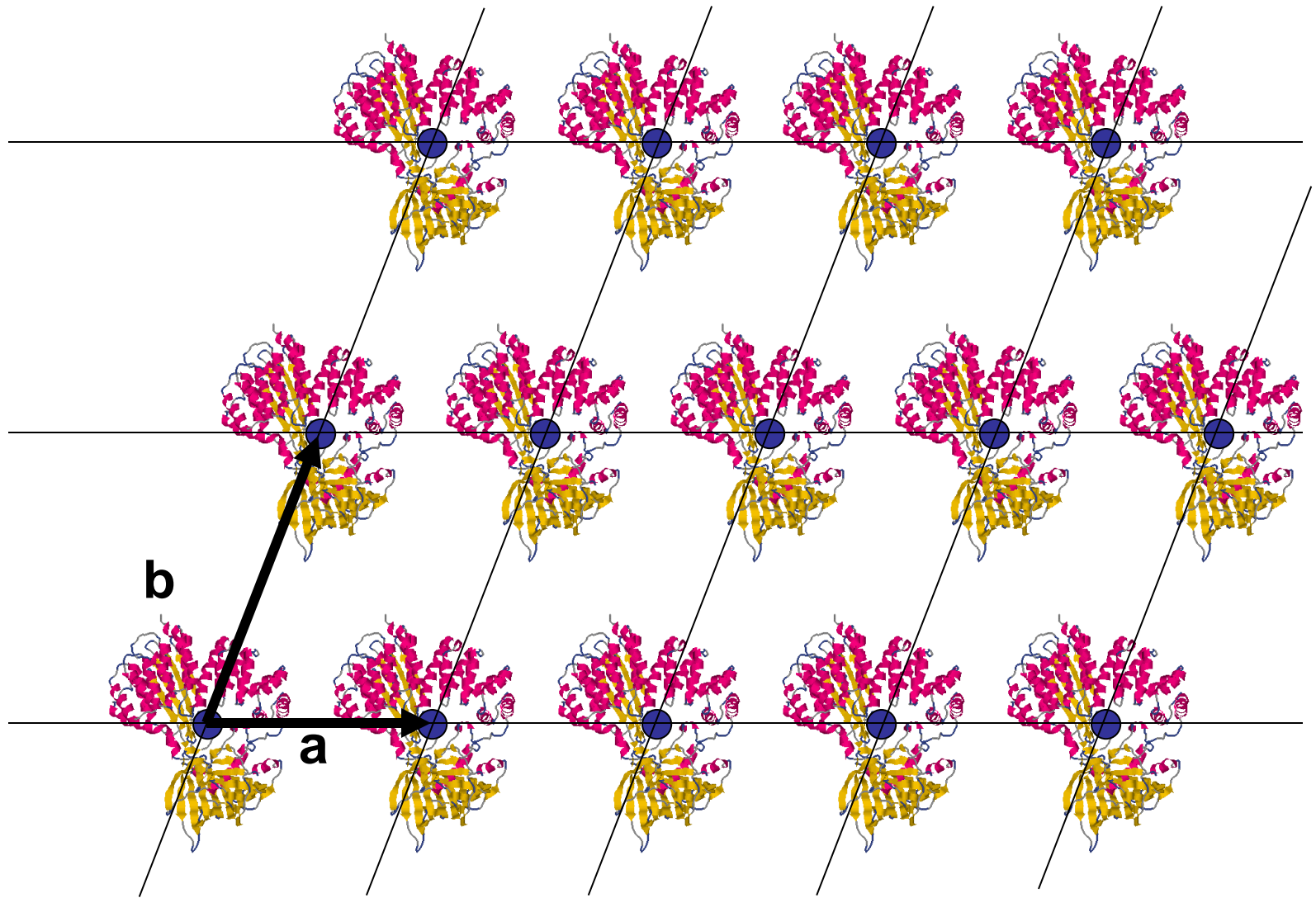
Много молекул:

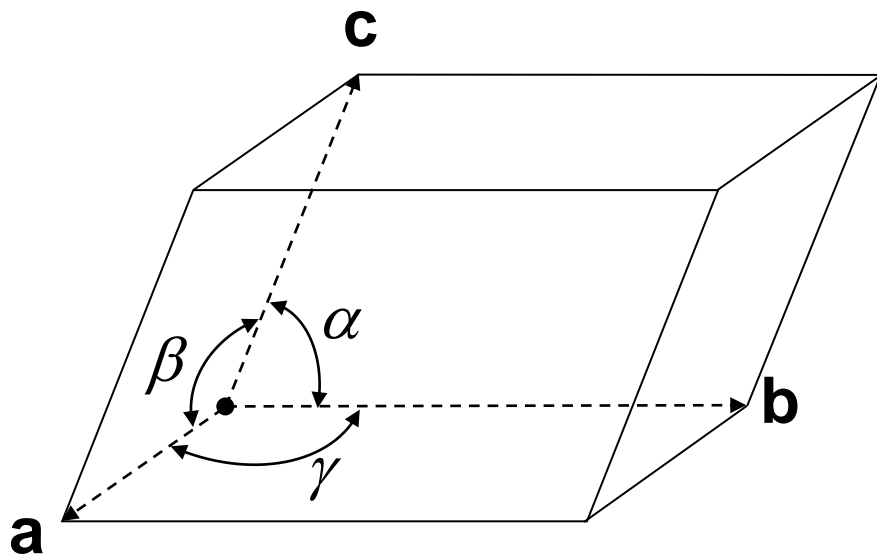
- растворы (малоугловое рентгеновское рассеяние, МУР);
- газы;
- порошки;
- **кристаллы.**

Рассеяние кристаллами



Buried a thousand feet (300 meters) below Naica mountain in the Chihuahuan Desert, the cave was discovered by two miners excavating a new tunnel for the Industrias Peñoles company in 2000.





элементарная ячейка
(unit cell)

$a, b, c, \alpha, \beta, \gamma$
параметры элементарной
ячейки
(unit cell parameters)

содержимое элементарной ячейки полностью
определяет «внутреннюю» структуру кристалла

$$\rho(\mathbf{r}) = \rho(\mathbf{r} + \mathbf{a}) = \rho(\mathbf{r} + \mathbf{b}) = \rho(\mathbf{r} + \mathbf{c})$$

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

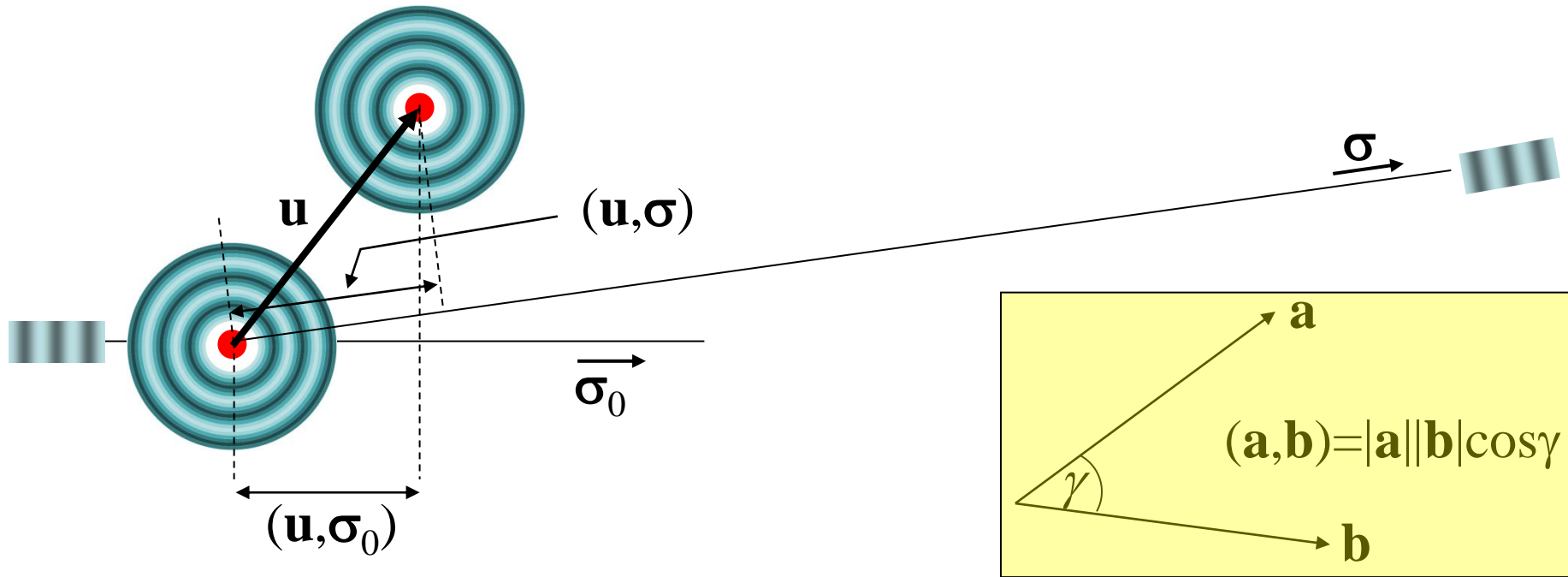
x, y, z – относительные
(кристаллографические)
координаты

$$\tilde{\rho}(x, y, z) = \rho(x\mathbf{a} + y\mathbf{b} + z\mathbf{c})$$

элементарная ячейка – куб
 $0 \leq x, y, z \leq 1$

$\tilde{\rho}(x, y, z)$ имеет период 1. по всем переменным.

Рассеяние рентгеновских лучей двумя электронами



- В «точке» детектора складываются два электрических поля
- Можно измерить амплитуду суммарной волны

$$E(\mathbf{r}, t) = E_0 \sin \left[2\pi \left(\frac{(\mathbf{r}, \boldsymbol{\sigma})}{\lambda} - vt + \delta \right) \right] \quad \Delta_1 = \frac{(\mathbf{u}, \boldsymbol{\sigma}_0)}{\lambda} \quad \Delta_2 = \frac{(\mathbf{u}, \boldsymbol{\sigma})}{\lambda}$$

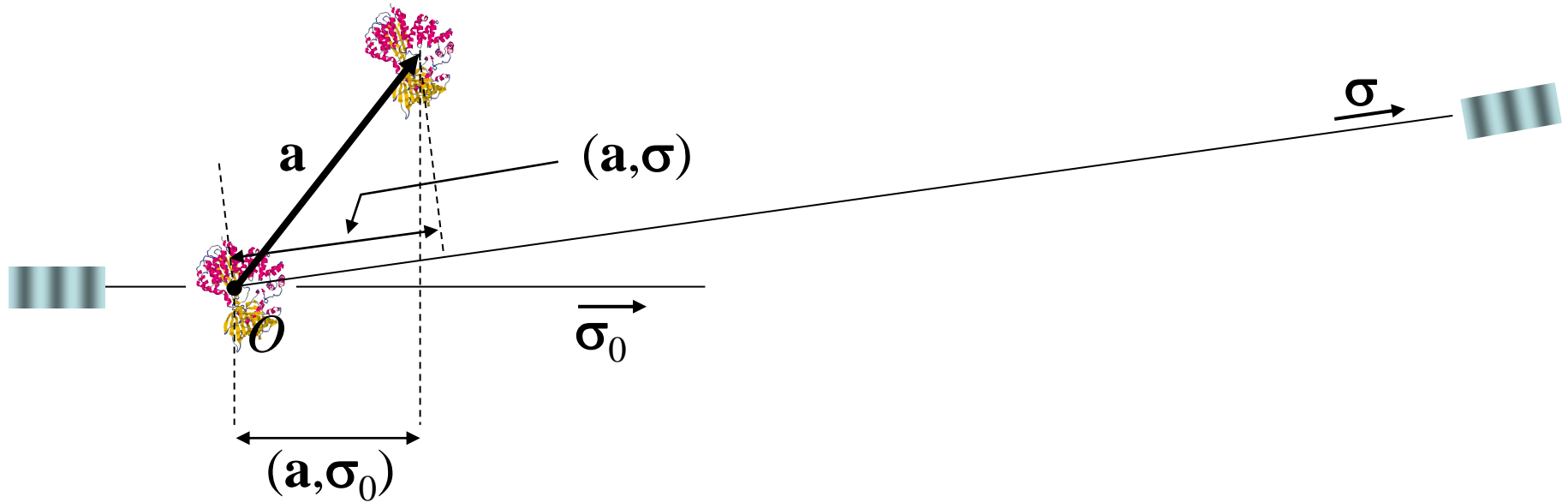
$$\Delta = \Delta_2 - \Delta_1 = (\mathbf{u}, \mathbf{s})$$

$$E_1(t) \propto E_0 \sin[2\pi(-vt)]$$

$$E_2(t) \propto E_0 \sin[2\pi(-vt - \Delta)]$$

$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda} \quad \text{вектор рассеяния}$$

Рассеяние рентгеновских лучей двумя молекулами



- В «точке» детектора складываются два электрических поля
- Можно измерить амплитуду суммарной волны

$$E_1(t) \propto F(\mathbf{s}) E_0 \sin[\omega]$$

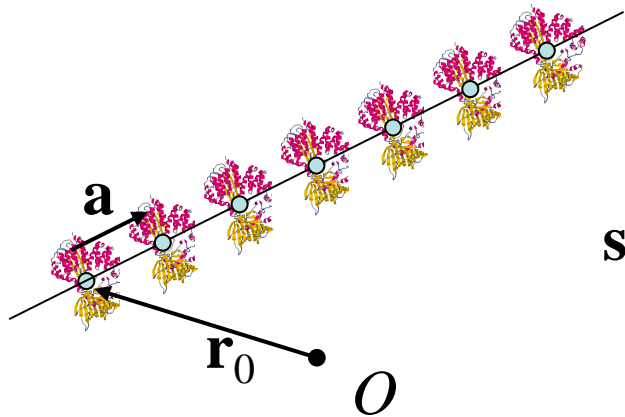
$$E_2(t) \propto F(\mathbf{s}) E_0 \sin[\omega - 2\pi\Delta]$$

$$\omega = 2\pi \left(\frac{R_{s-d}}{\lambda} - vt \right) - \varphi(\mathbf{s})$$

$F(\mathbf{s})$ -модуль структурного фактора молекулы

$$\Delta = (\mathbf{a}, \mathbf{s})$$

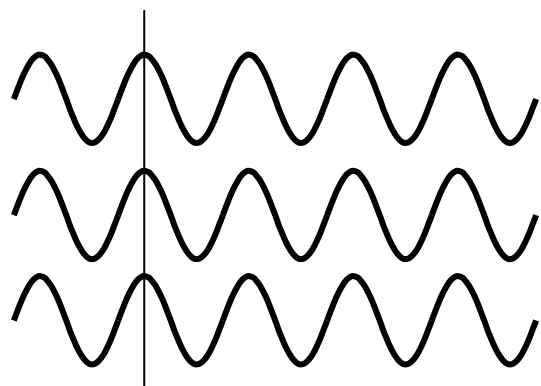
$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda} \quad \text{вектор рассеяния}$$



$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda}$$

$$\Delta = (\mathbf{s}, \mathbf{a})$$

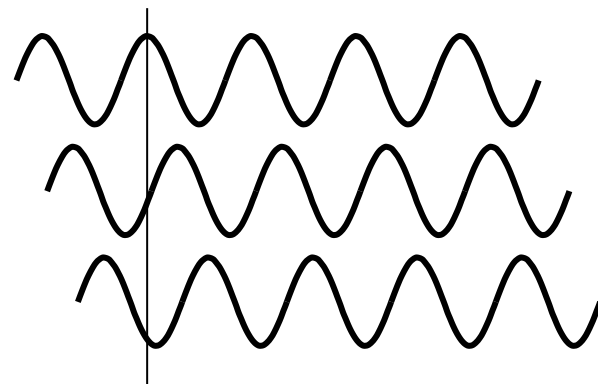
$$F(\mathbf{s}) E_0 \left\{ \sin[\omega] + \sin[\omega + 2\pi(\Delta)] + \sin[\omega + 2\pi(2\Delta)] + \sin[\omega + 2\pi(3\Delta)] \dots \right\}$$



$(\mathbf{s}, \mathbf{a}) = h$ - целое

волны усиливаются

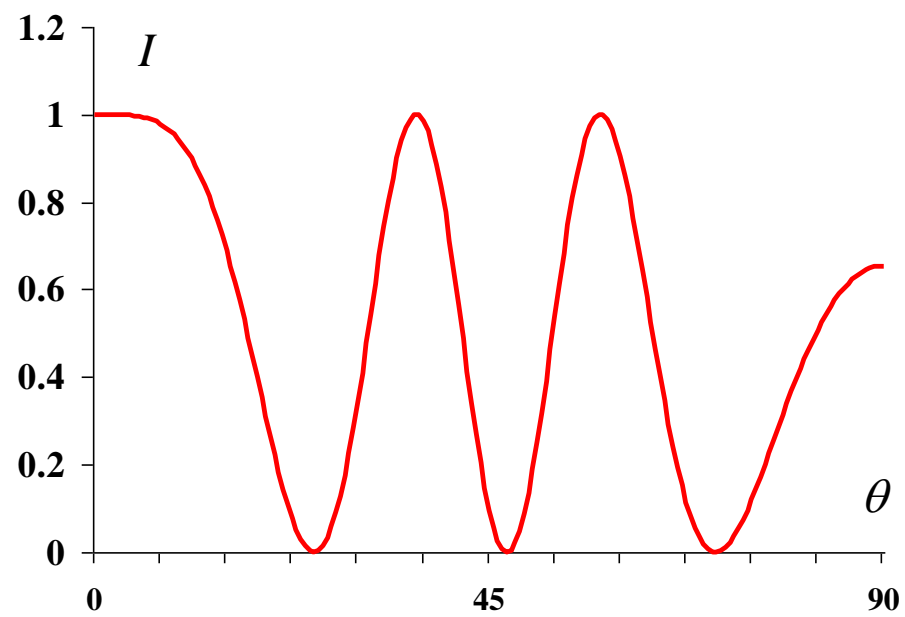
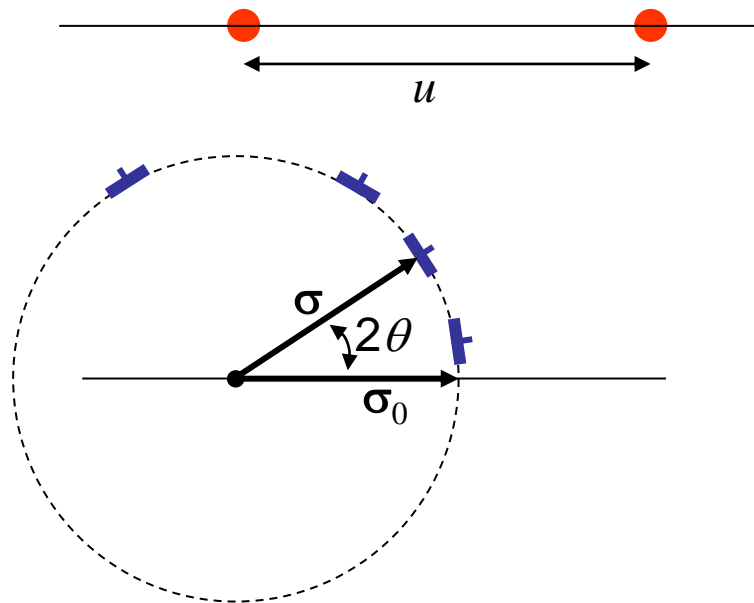
амплитуда волны $F(\mathbf{s}) E_0 N_{mol}$

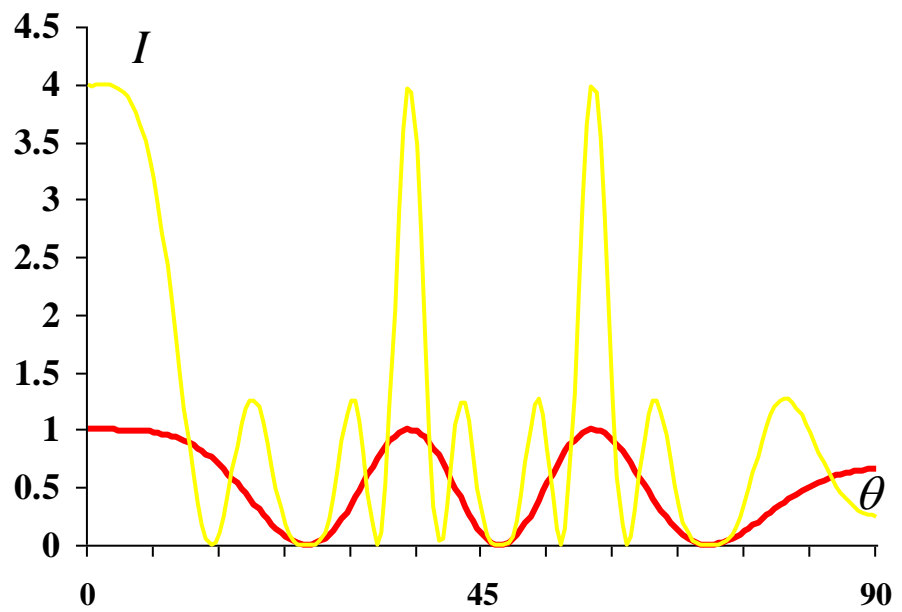
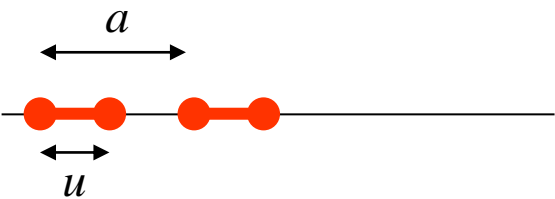
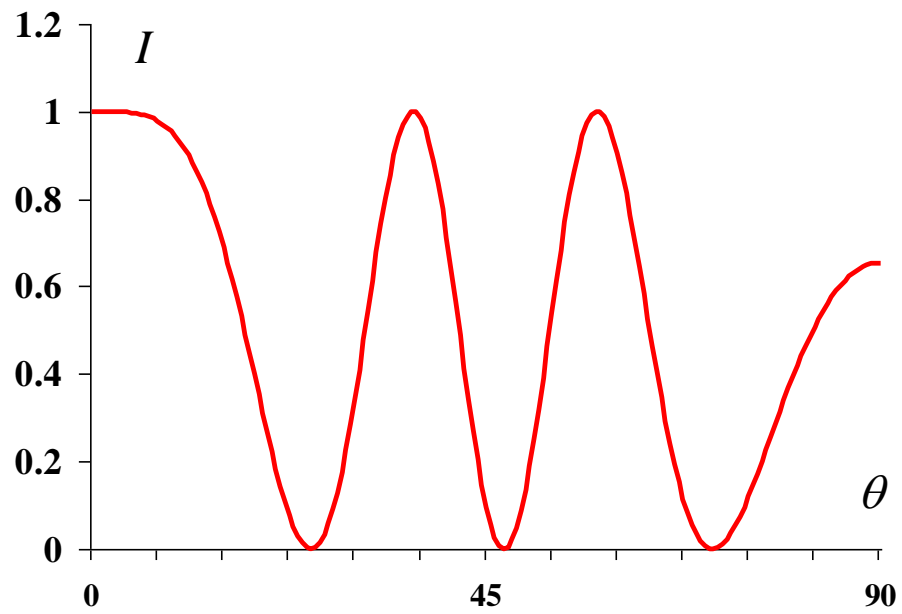
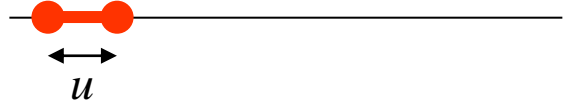


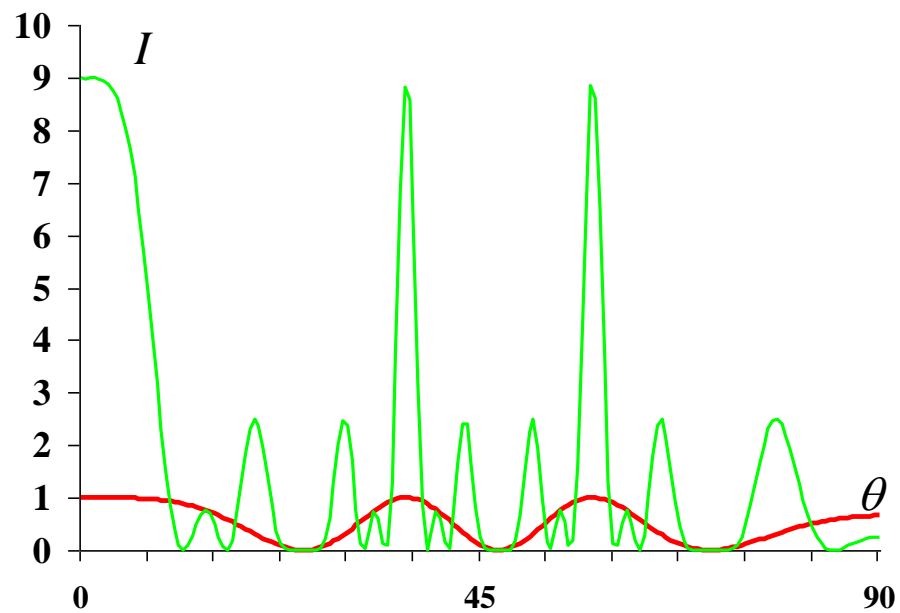
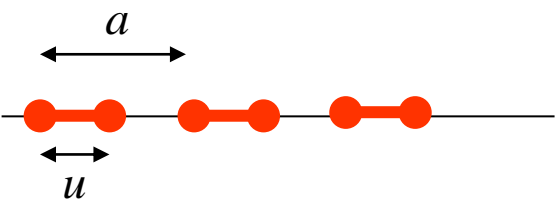
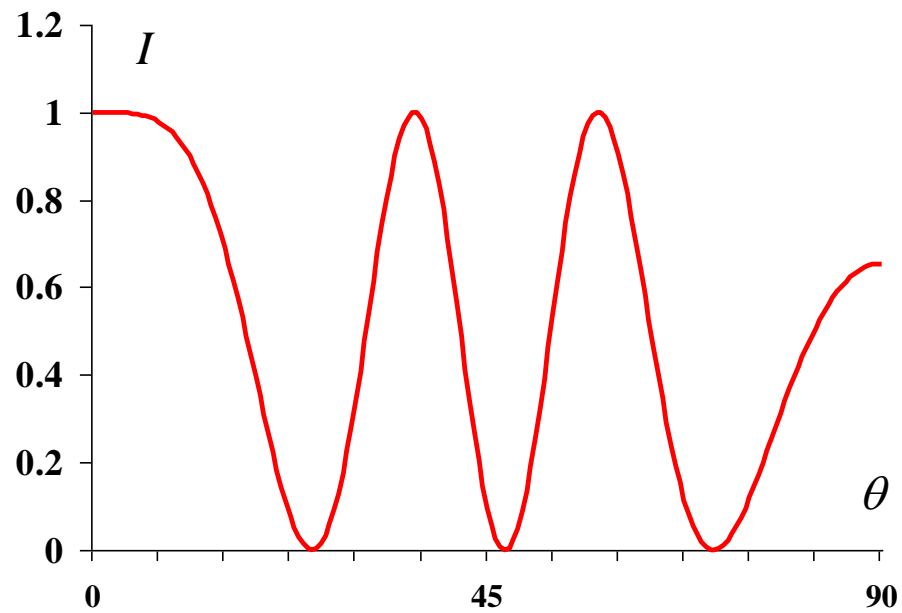
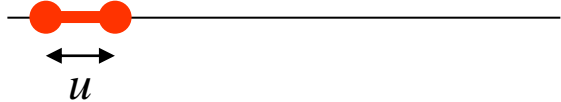
$(\mathbf{s}, \mathbf{a}) = h$ - нецелое

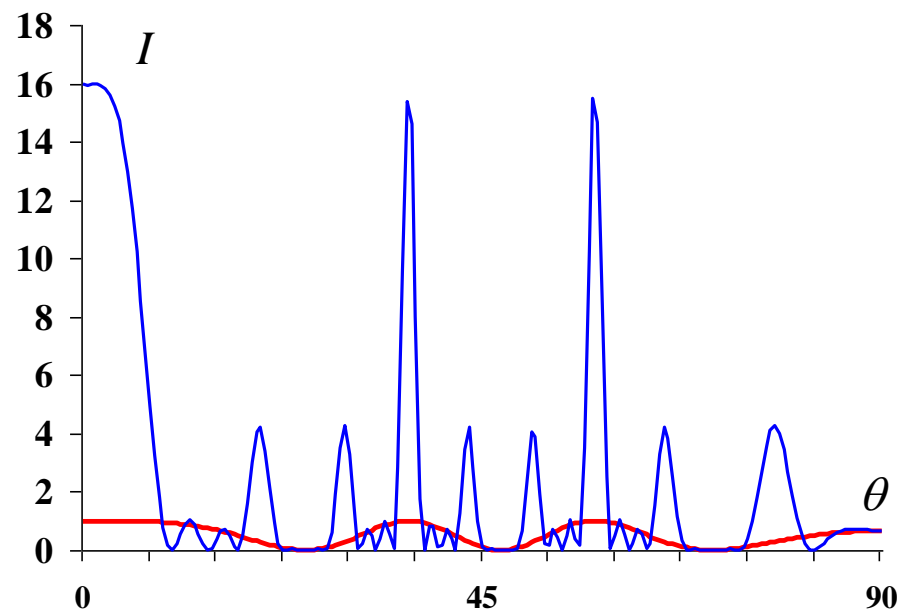
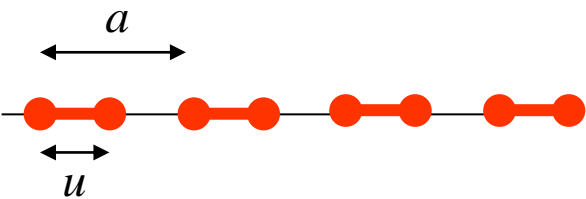
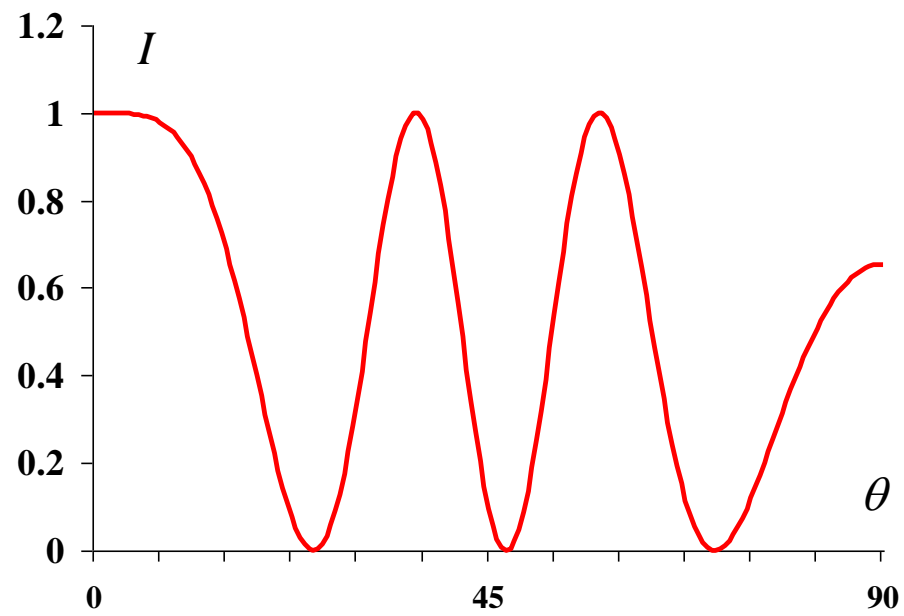
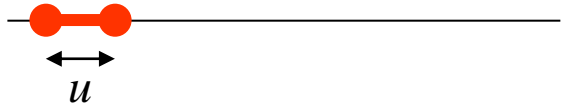
усиления нет

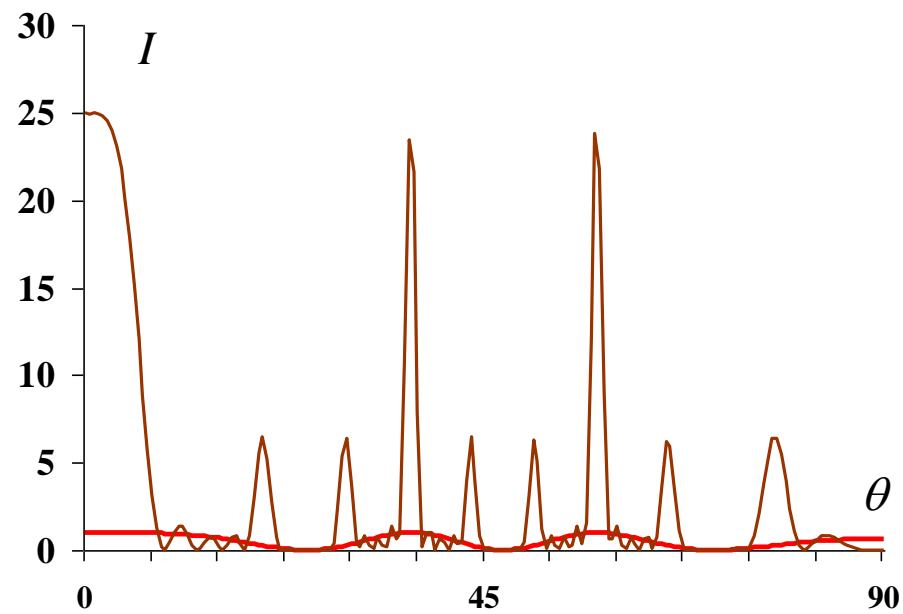
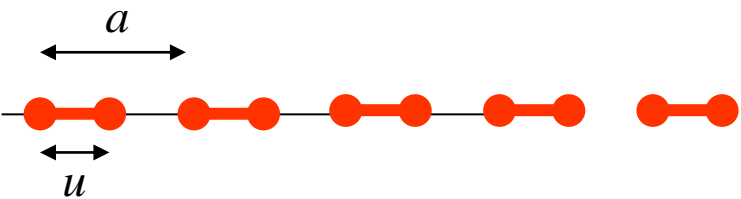
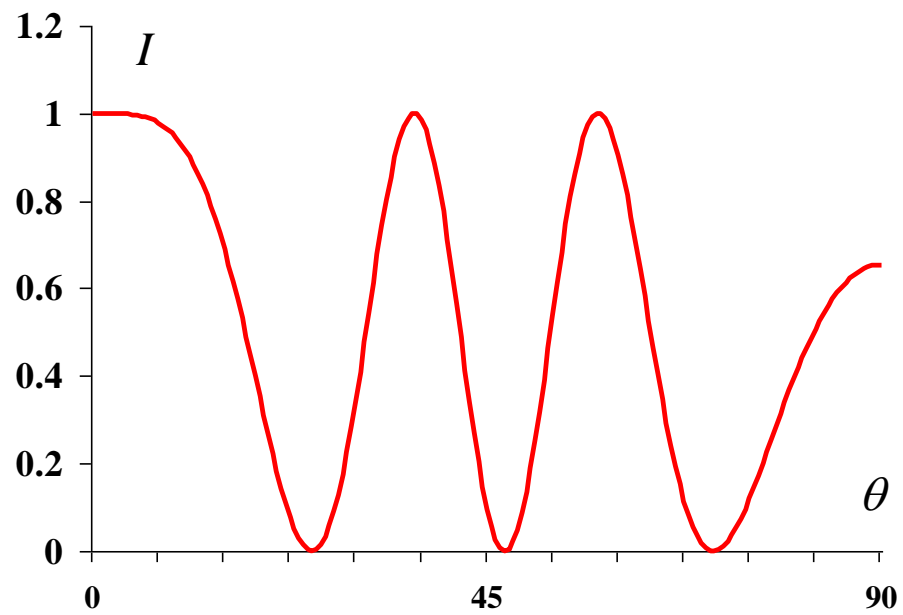
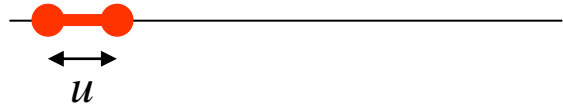
амплитуда волны ≈ 0

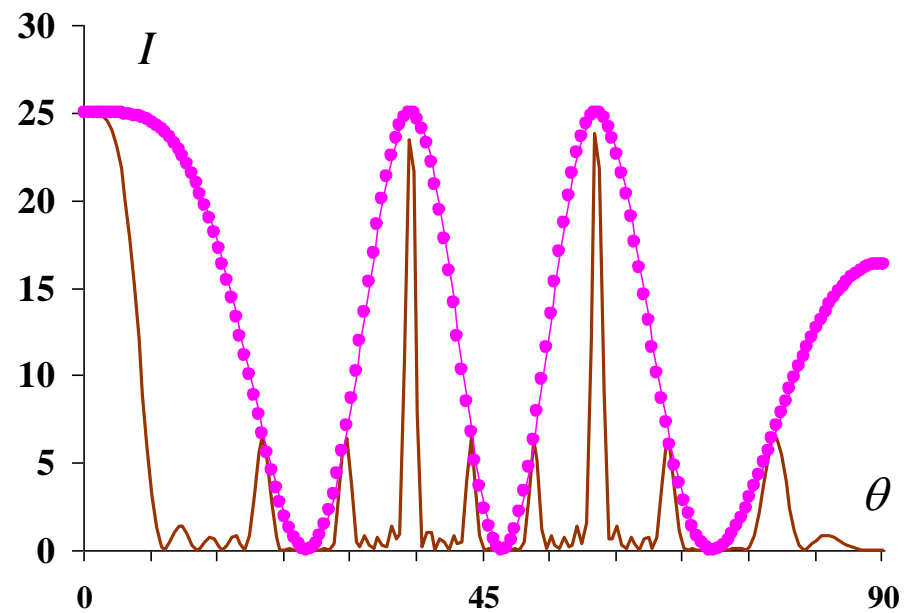
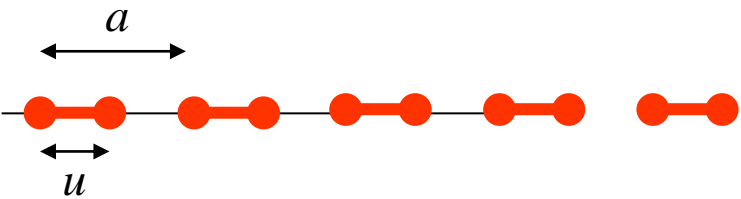
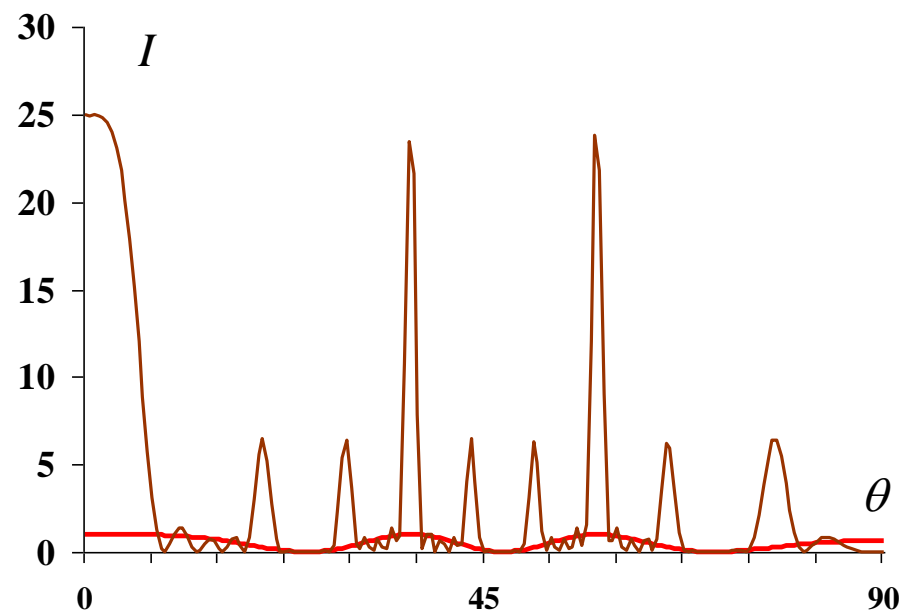
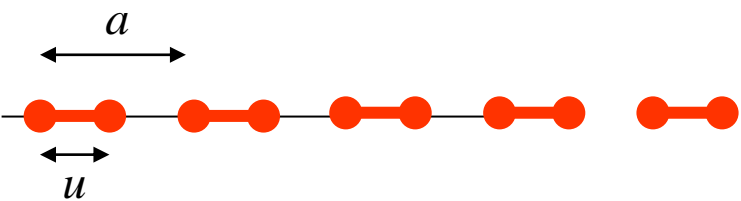


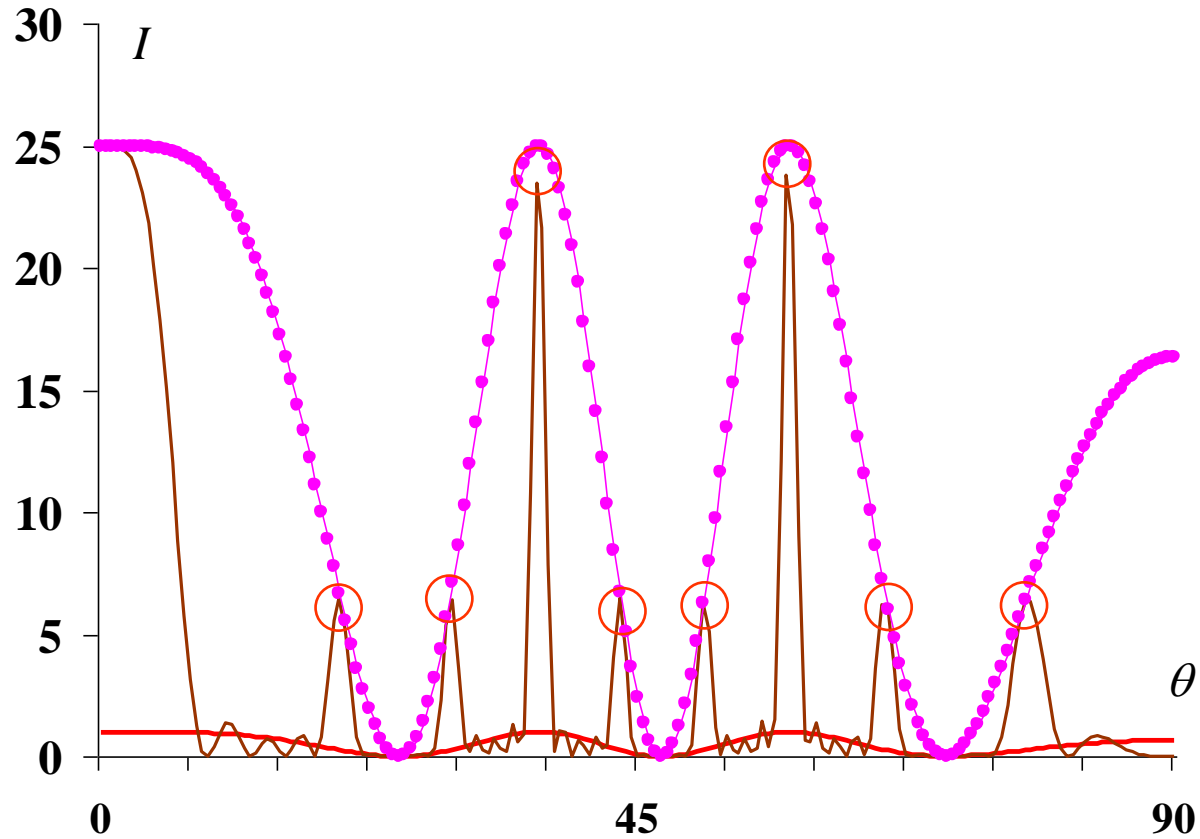
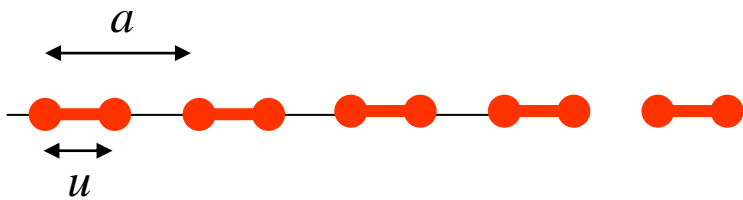






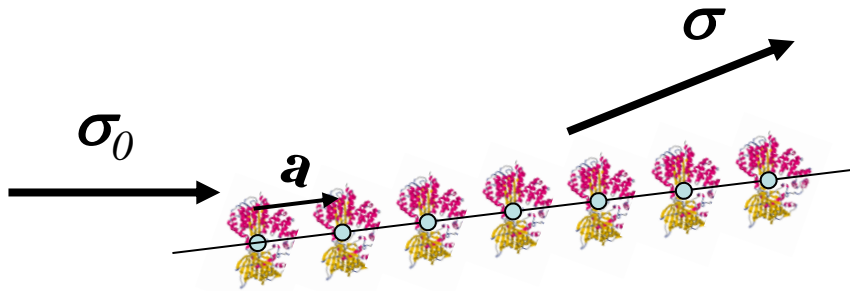




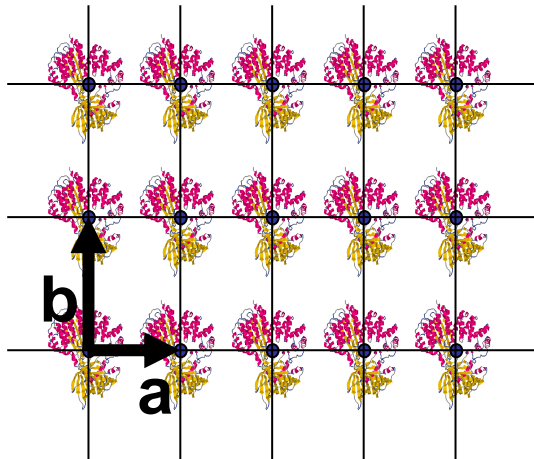


- усиление сигнала в 25 раз для некоторых углов рассеяния
- потеря информации для других направлений

Условия Лауэ



$$\Delta = \left(\frac{\sigma - \sigma_0}{\lambda}, \mathbf{a} \right) = h \quad \text{integer}$$



$$\left(\frac{\sigma - \sigma_0}{\lambda}, \mathbf{b} \right) = k \quad \text{integer}$$

трехмерный кристалл

$$\left(\frac{\sigma - \sigma_0}{\lambda}, \mathbf{c} \right) = l \quad \text{integer}$$

h, k, l – индексы (Miller indices) рефлекса

линейные размеры кристалла $0.1\text{мм}=10^6\text{\AA}$
линейные размеры элементарной ячейки 100\AA

количество копий молекулы в кристалле $(10^4)^3=10^{12}$

Кристалл усиливает интенсивность
в 10^{24} раз !!!

Условия дифракции

(Лауэ):

$$(\mathbf{s}, \mathbf{a}) = h$$

$$(\mathbf{s}, \mathbf{b}) = k$$

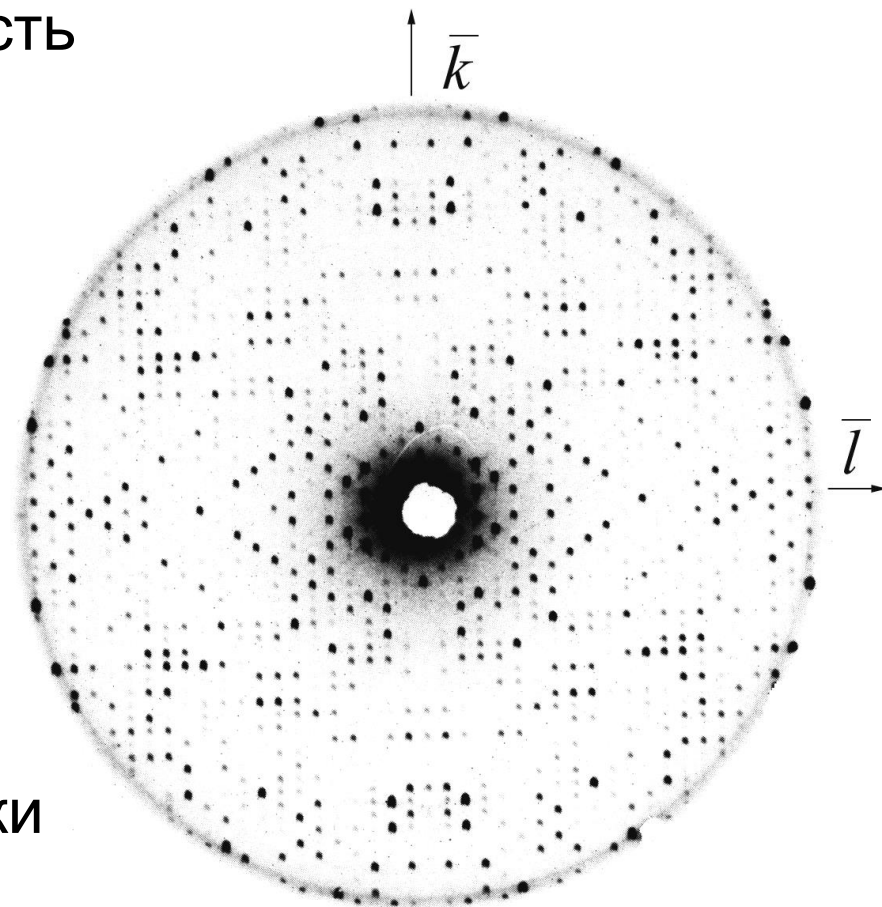
$$(\mathbf{s}, \mathbf{c}) = l$$

h, k, l - целые числа




(индексы рефлекса)

$\mathbf{a}, \mathbf{b}, \mathbf{c}$ - ребра элементарной ячейки

$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda} \text{ - вектор рассеяния}$$



Кристалл

-  Позволяет радикально увеличить амплитуду рассеянной волны для некоторых («Брэгговских») рефлексов.
-  Теряется информация о модулях структурных факторов для остальных рефлексов.
-  Проблема получения кристаллов.

Метод проб и ошибок

Гипотетическая
структура; координаты
атомов $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$



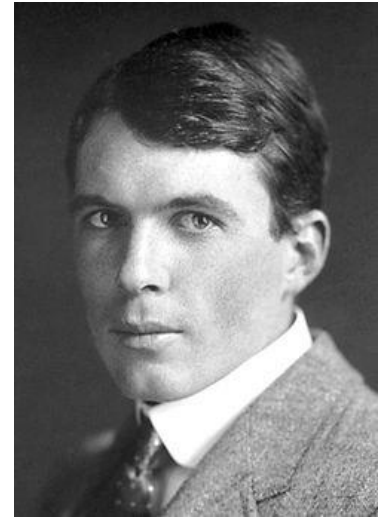
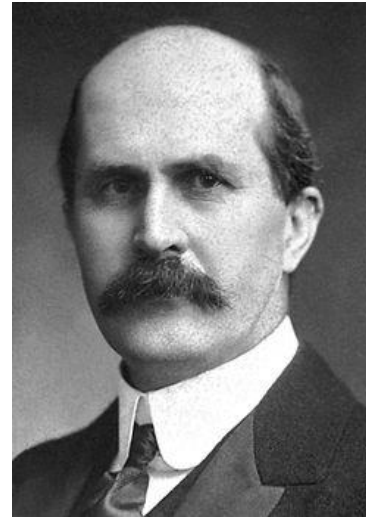
Расчет соответствующей
картины дифракции

$$F(\mathbf{s}_1), F(\mathbf{s}_2), \dots, F(\mathbf{s}_K)$$



Сравнение с
экспериментом

$$R = \sum_{k=1}^K \left(F^{obs}(\mathbf{s}_k) - F^{calc}(\mathbf{s}_k) \right)^2$$



Sir William Henry Bragg
Sir William Lawrence Bragg

Нобелевская премия 1915 г.

Bernal J.D., Crowfoot D.C., Nature, 143, 794 (1934)
первые рентгенограммы пепсина

Perutz M.F. 1937 тема диссертации – определение кристаллической структуры гемоглобина

1950 – 1960

- появление ЭВМ;
- “двухшаговое” решение структуры;

Perutz M.F., Kendrew J.C. Нобелевская премия 1962 г.
структуры гемоглобина и миоглобина

1965 – лизоцим

1967 – рибонуклеаза, химотрипсин, карбоксипептидаза

Метод проб и ошибок

гипотеза о структуре
(координаты атомов)



расчет ожидаемой картины
дифракции



сравнение с экспериментом



ВЫВОДЫ

Новый подход

распределение
электронов в объекте
(приблизенно)



интерпретация
(расстановка атомов
модели)

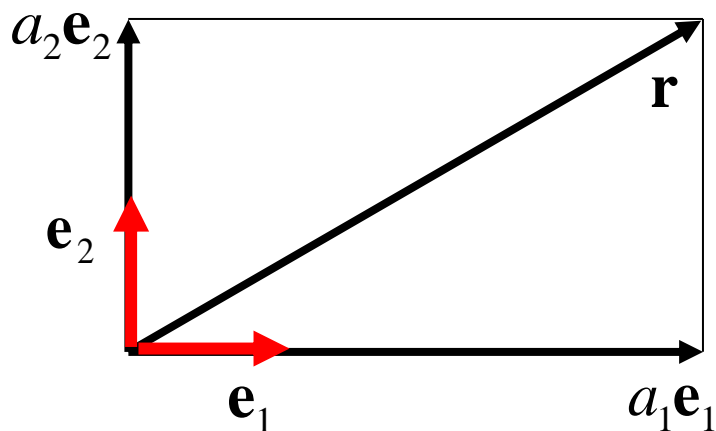


уточнение модели

Ряды Фурье



Жан Батист Жозеф Фурье
1768 - 1830



$$\mathbf{r} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2$$

$\mathbf{e}_1, \mathbf{e}_2$ - базисные вектора

a_1, a_2 - коэффициенты разложения

$$\mathbf{r} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$$

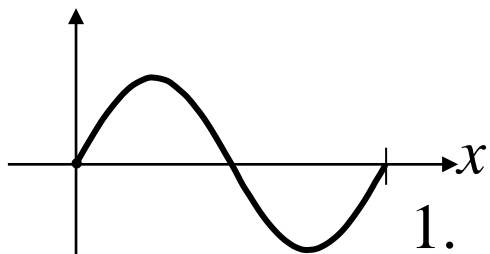
$$\rho(x) = a_1e_1(x) + a_2e_2(x) + a_3e_3(x) + \dots$$

$e_1(x), e_2(x), e_3(x), \dots$ - базисные функции (1, x, x², x³, ...)

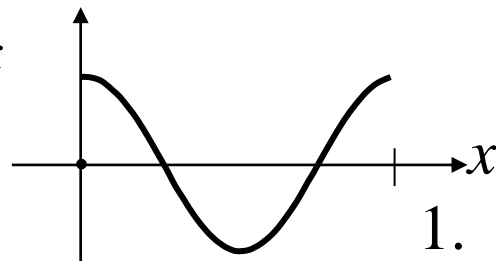
a_1, a_2, a_3, \dots - коэффициенты разложения

базисы из степенных функций – Грегори, Ньютон XVII век

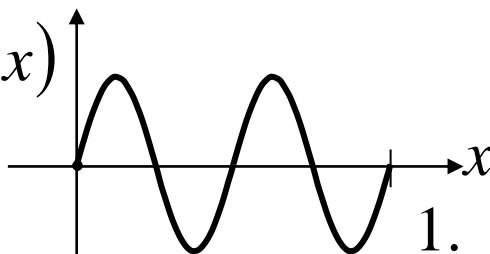
$$\sin 2\pi x$$



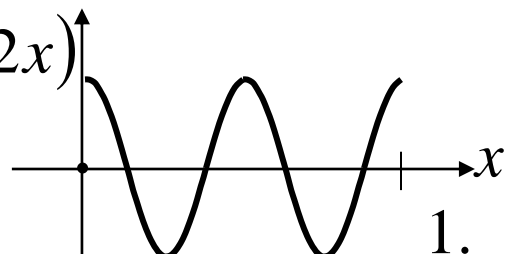
$$\cos 2\pi x$$



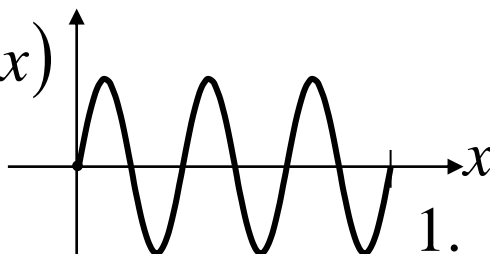
$$\sin 2\pi(2x)$$



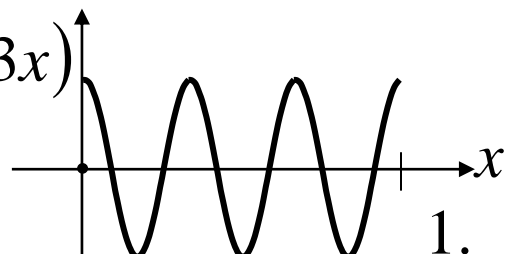
$$\cos 2\pi(2x)$$



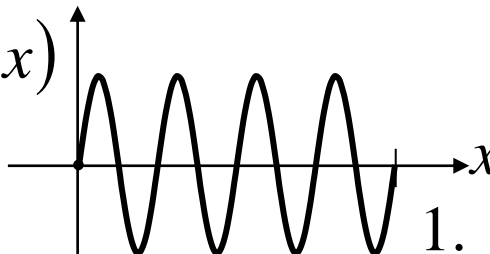
$$\sin 2\pi(3x)$$



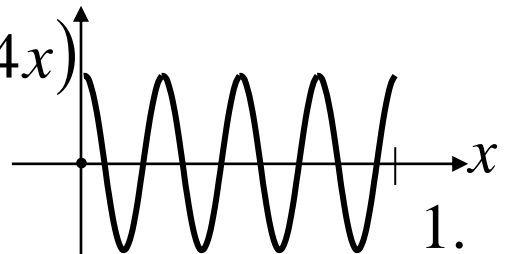
$$\cos 2\pi(3x)$$



$$\sin 2\pi(4x)$$



$$\cos 2\pi(4x)$$



$$\sin 2\pi(hx)$$

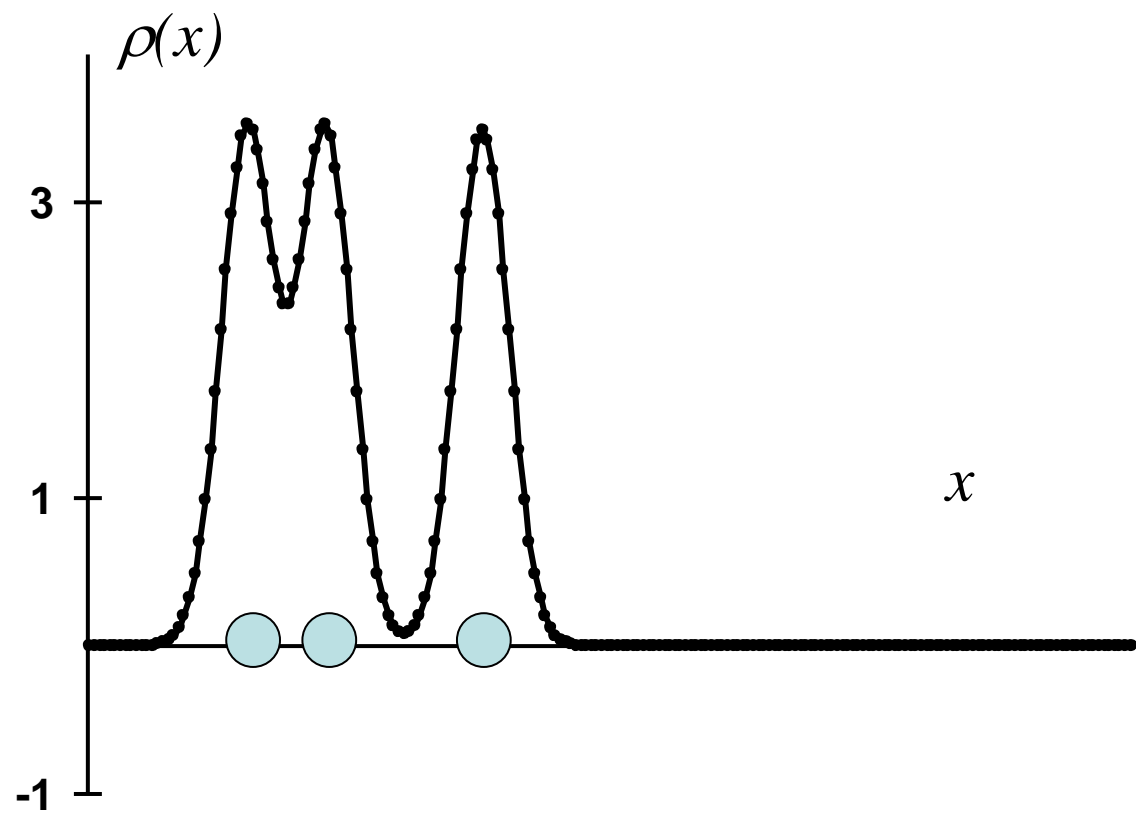
$$\cos 2\pi(hx)$$

"Классический" ("тригонометрический") ряд Фурье

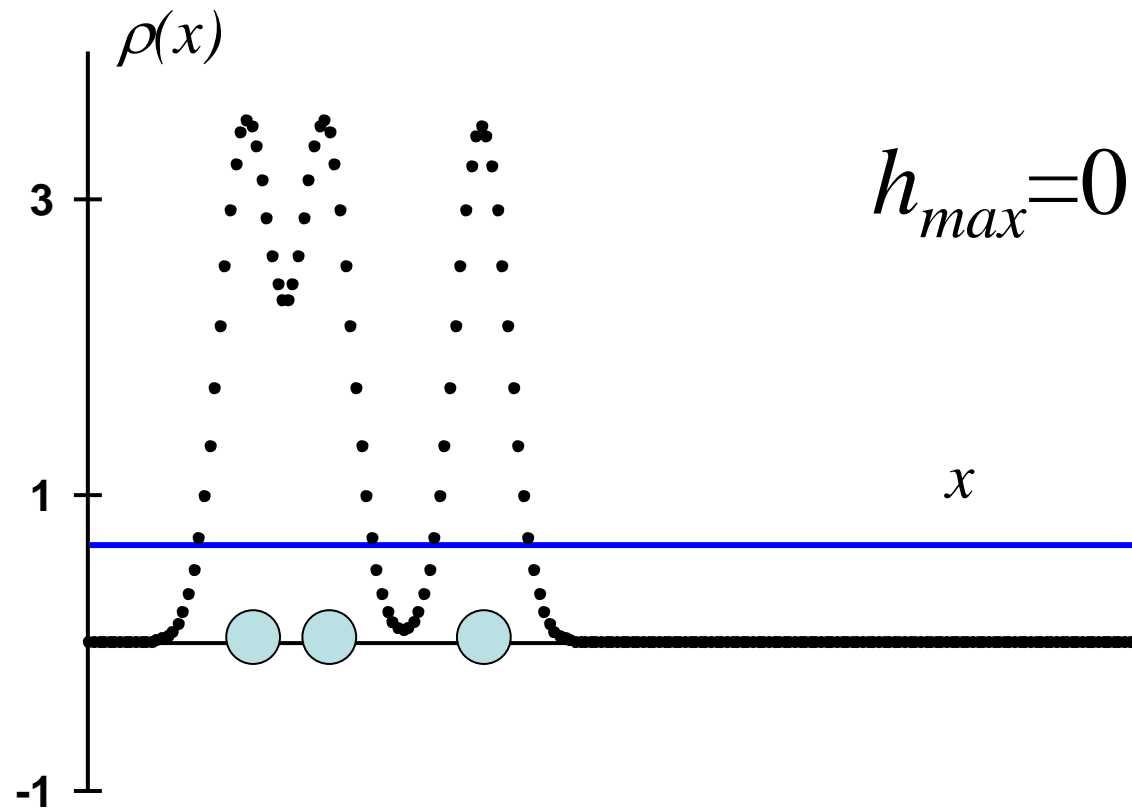
Любая функция может быть приближена на отрезке $[0,1]$ комбинацией "синусов и косинусов".

$$\begin{aligned}\rho(x) &= A_0 \\ &+ A_1 \cos 2\pi x + B_1 \sin 2\pi x \\ &+ A_2 \cos 2\pi(2x) + B_2 \sin 2\pi(2x) \\ &+ A_3 \cos 2\pi(3x) + B_3 \sin 2\pi(3x) \\ &+ \dots\end{aligned}$$

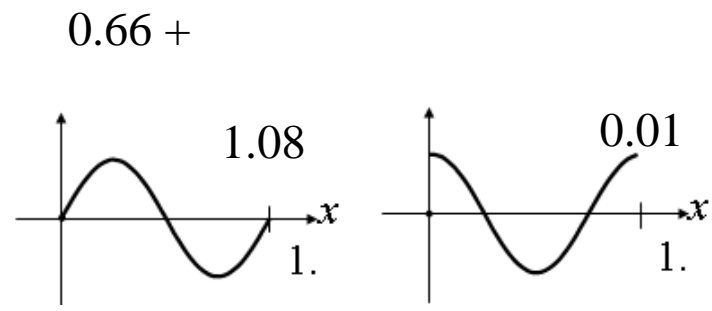
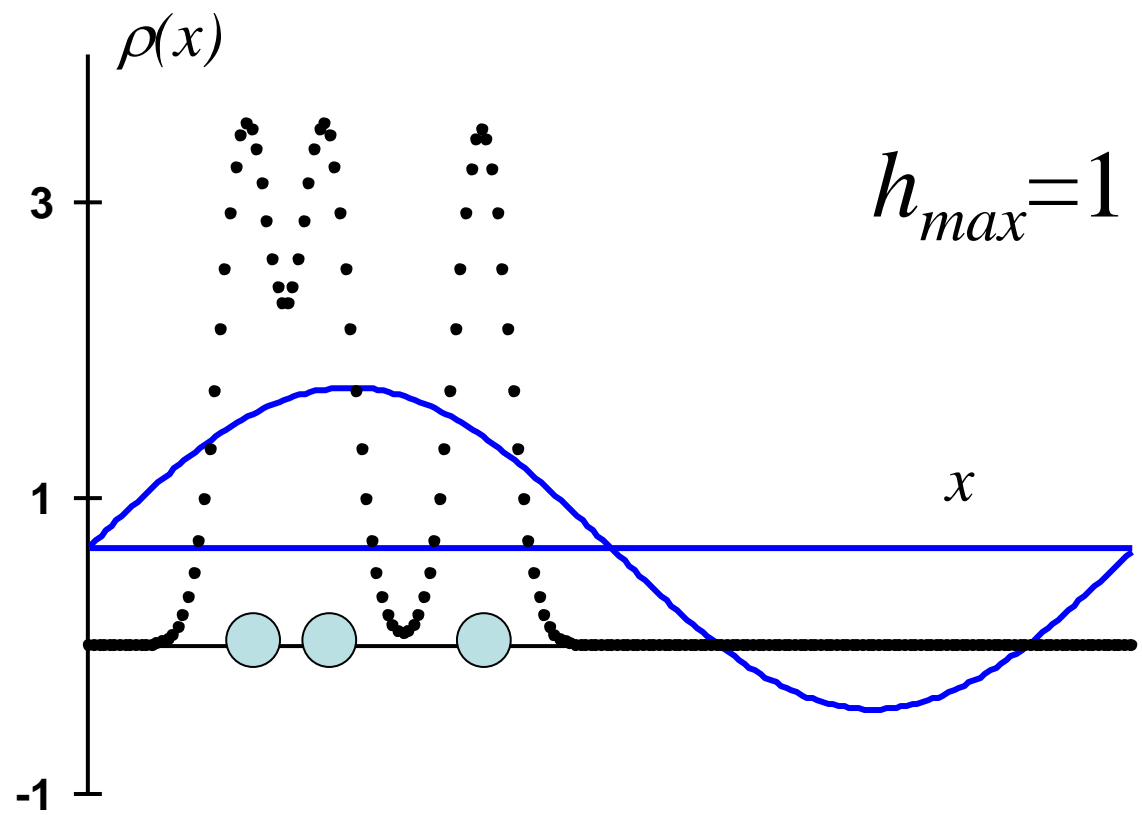
$$\rho(x) \approx A_0 + \sum_h \left\{ A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx) \right\}$$



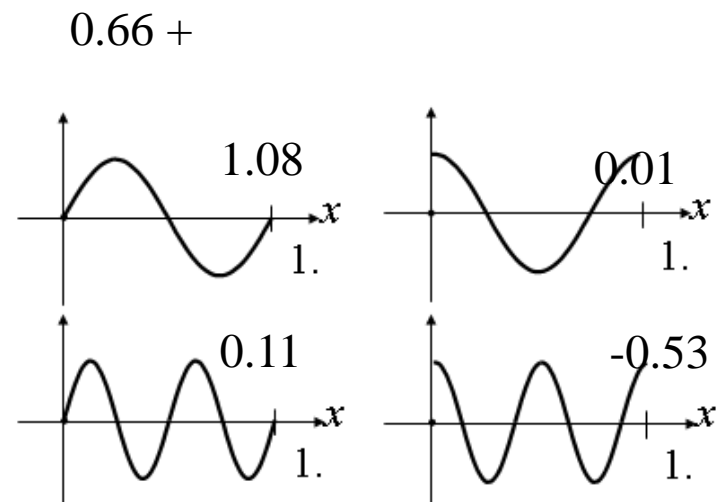
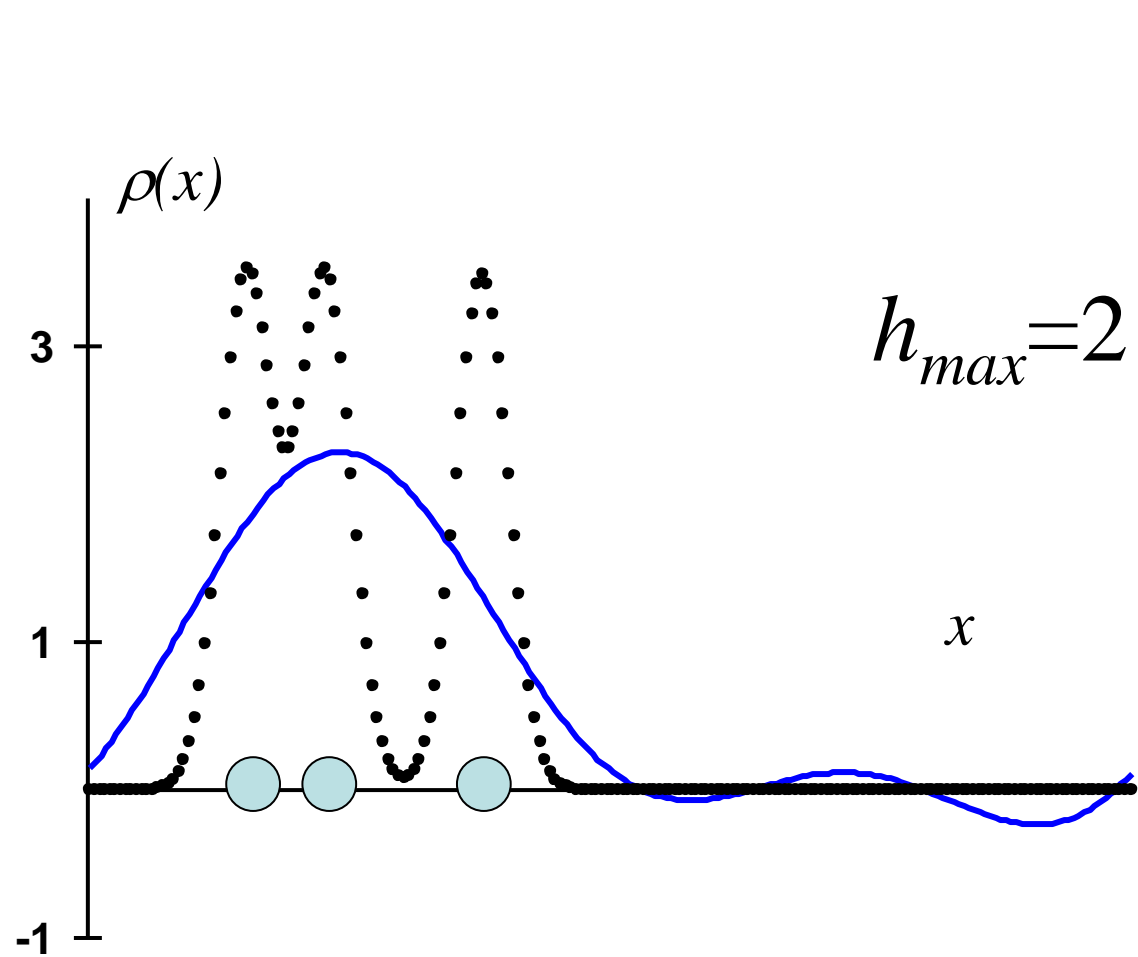
0.66 +



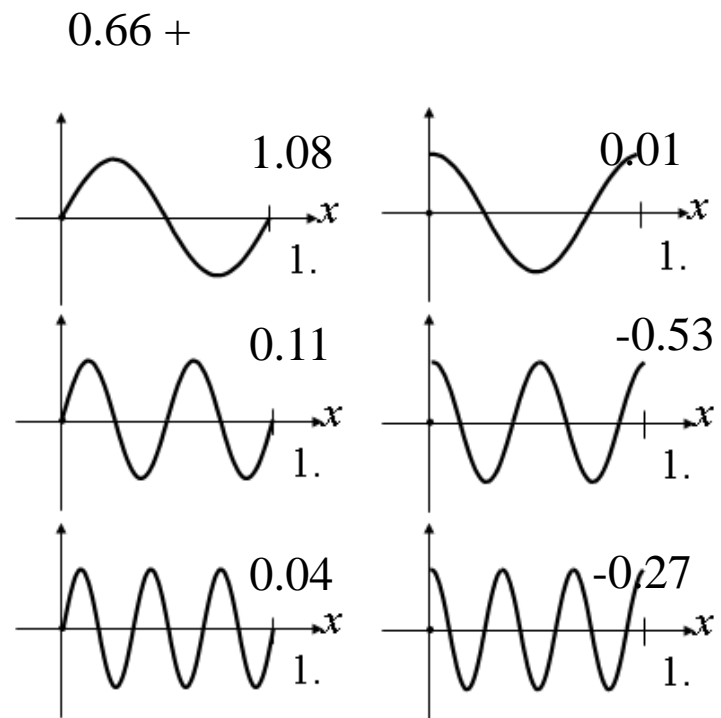
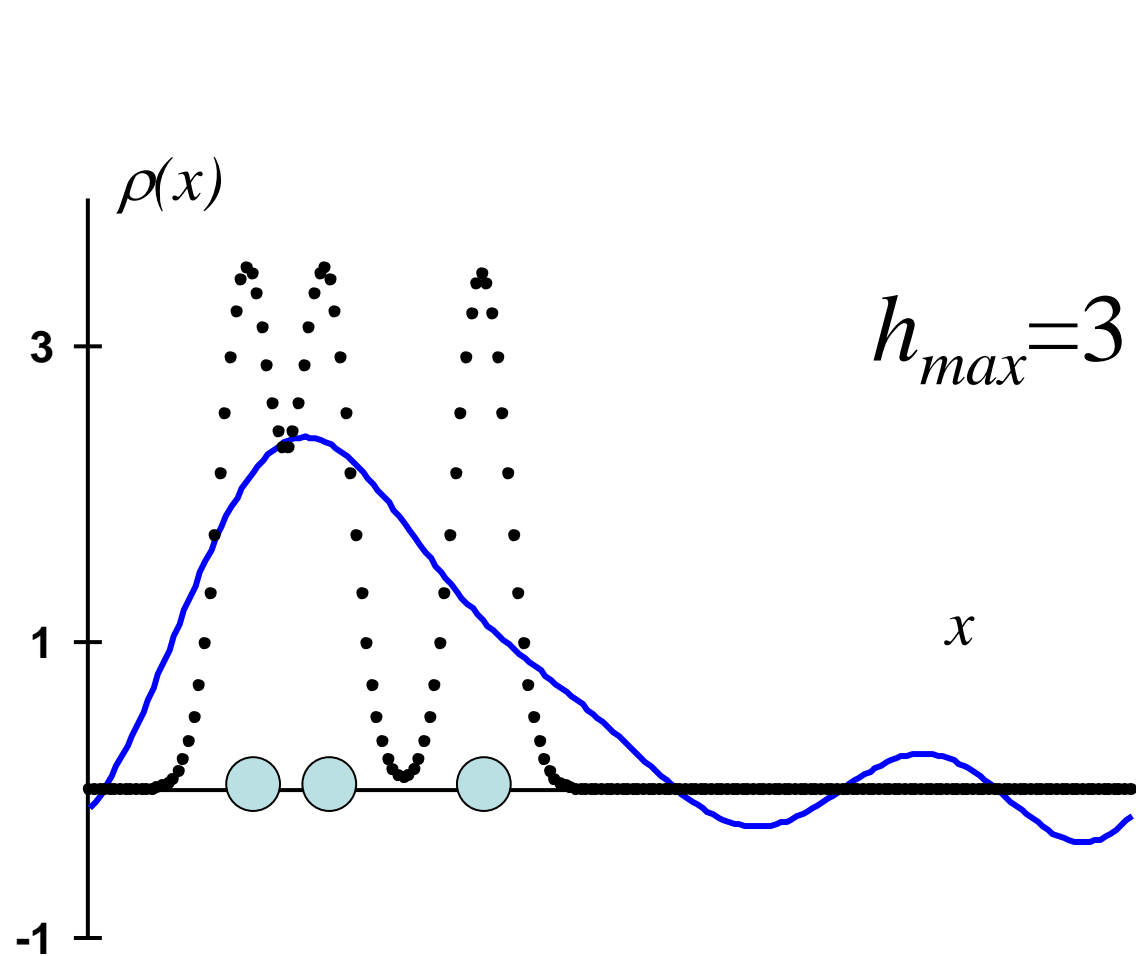
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



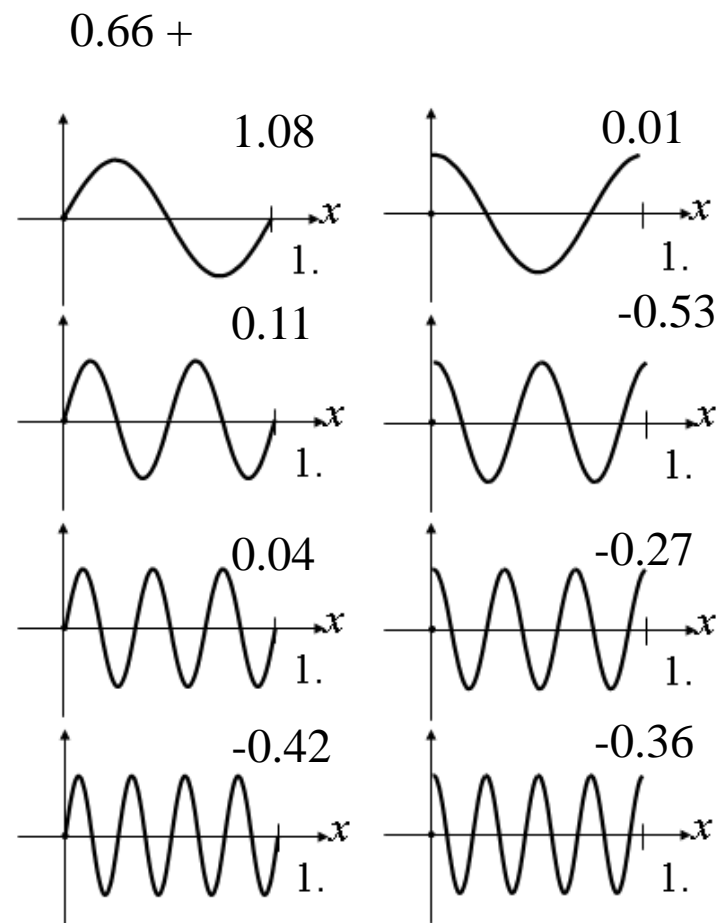
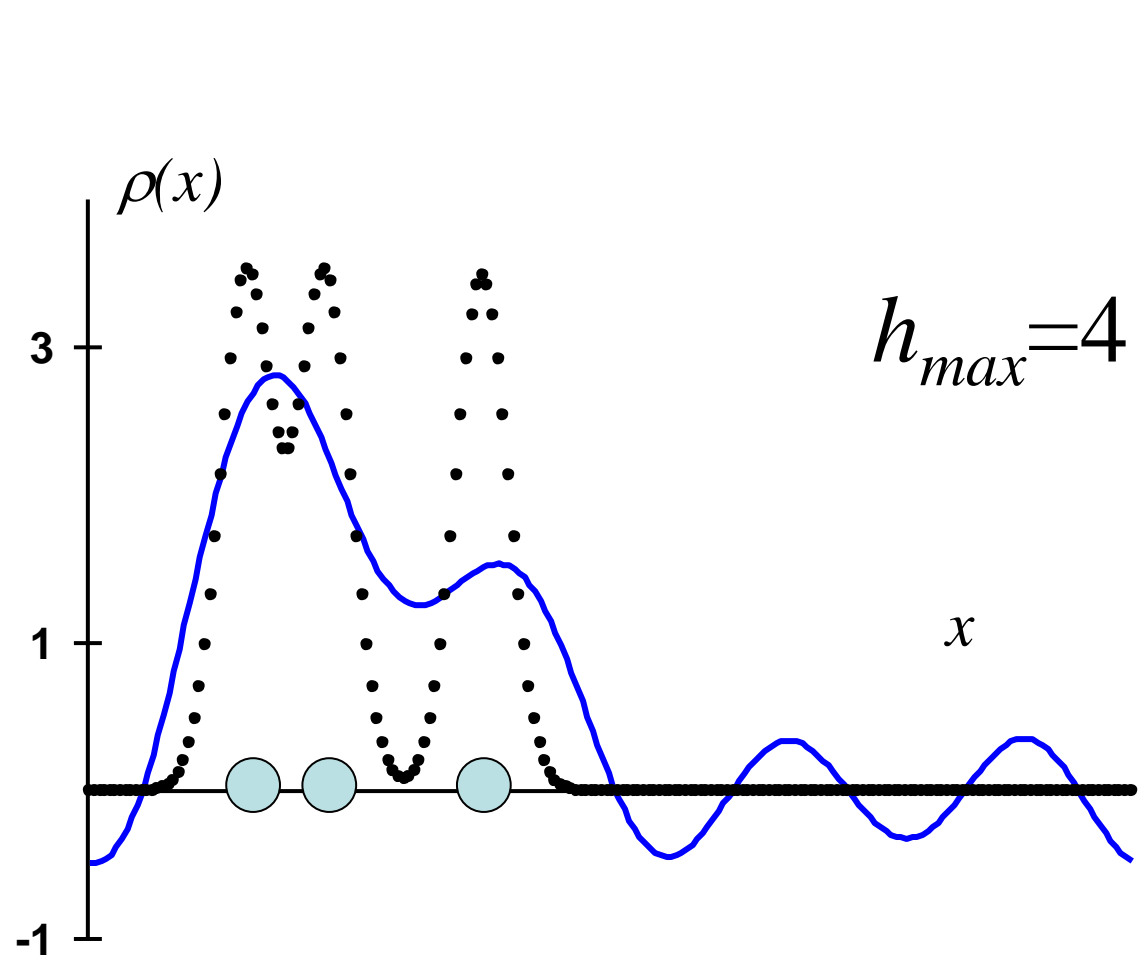
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



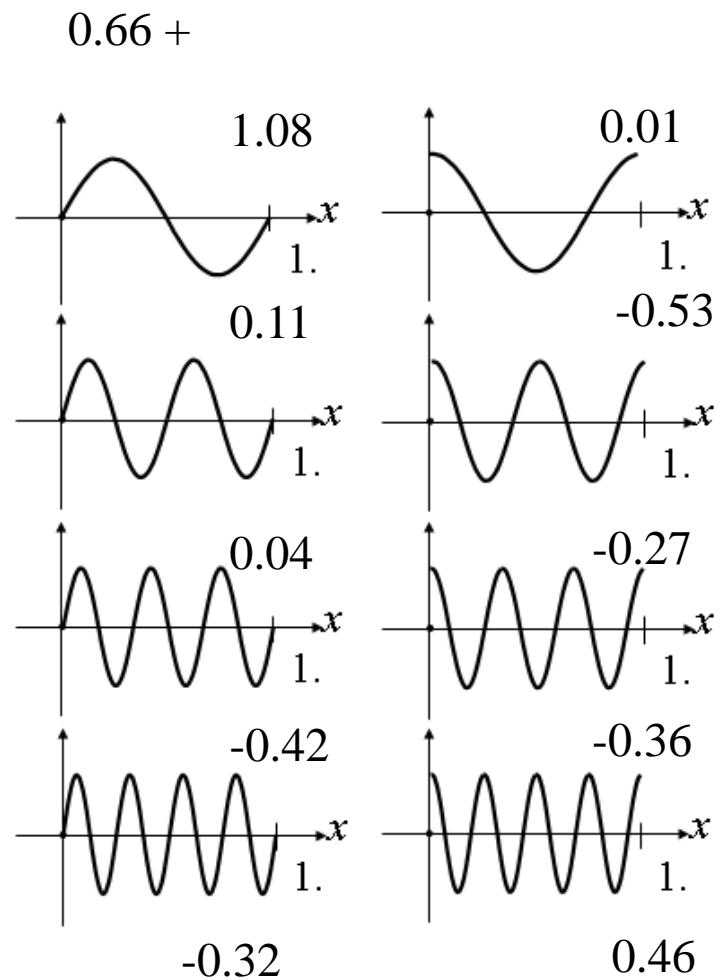
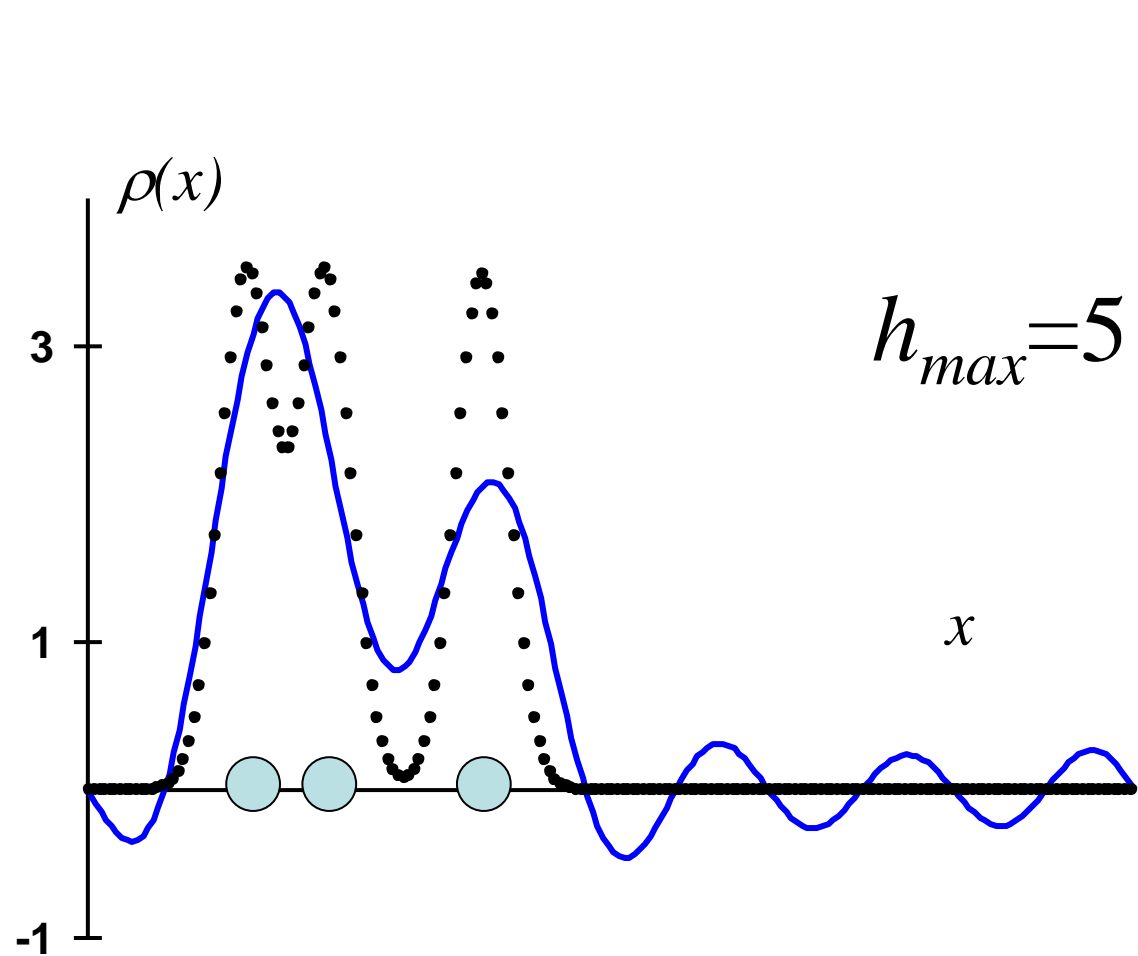
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



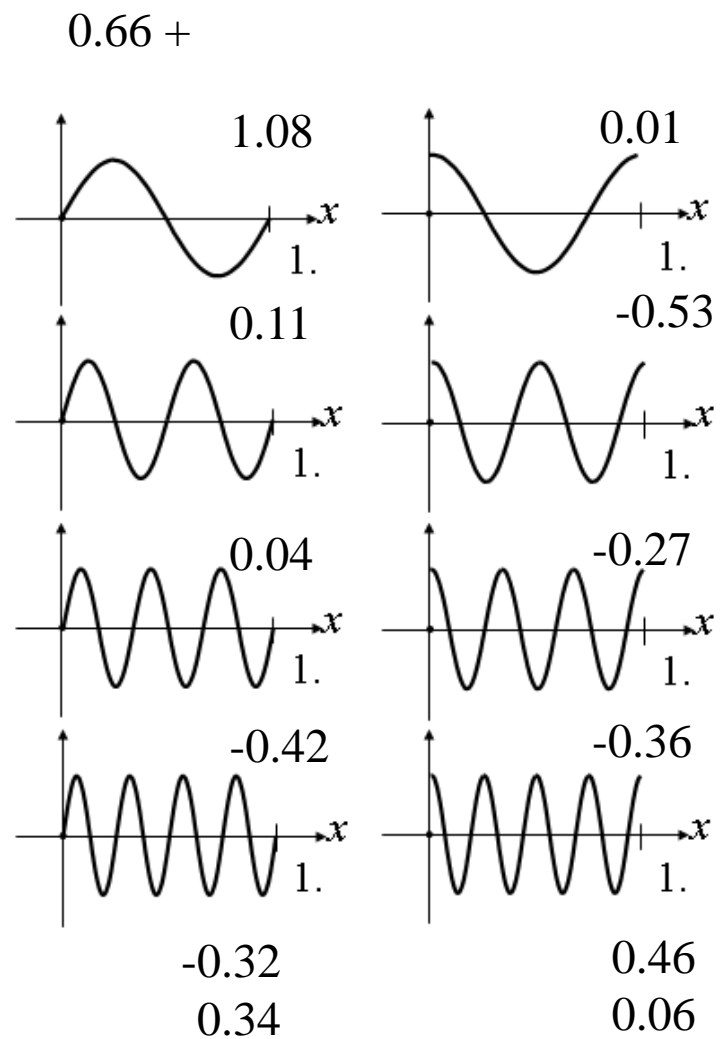
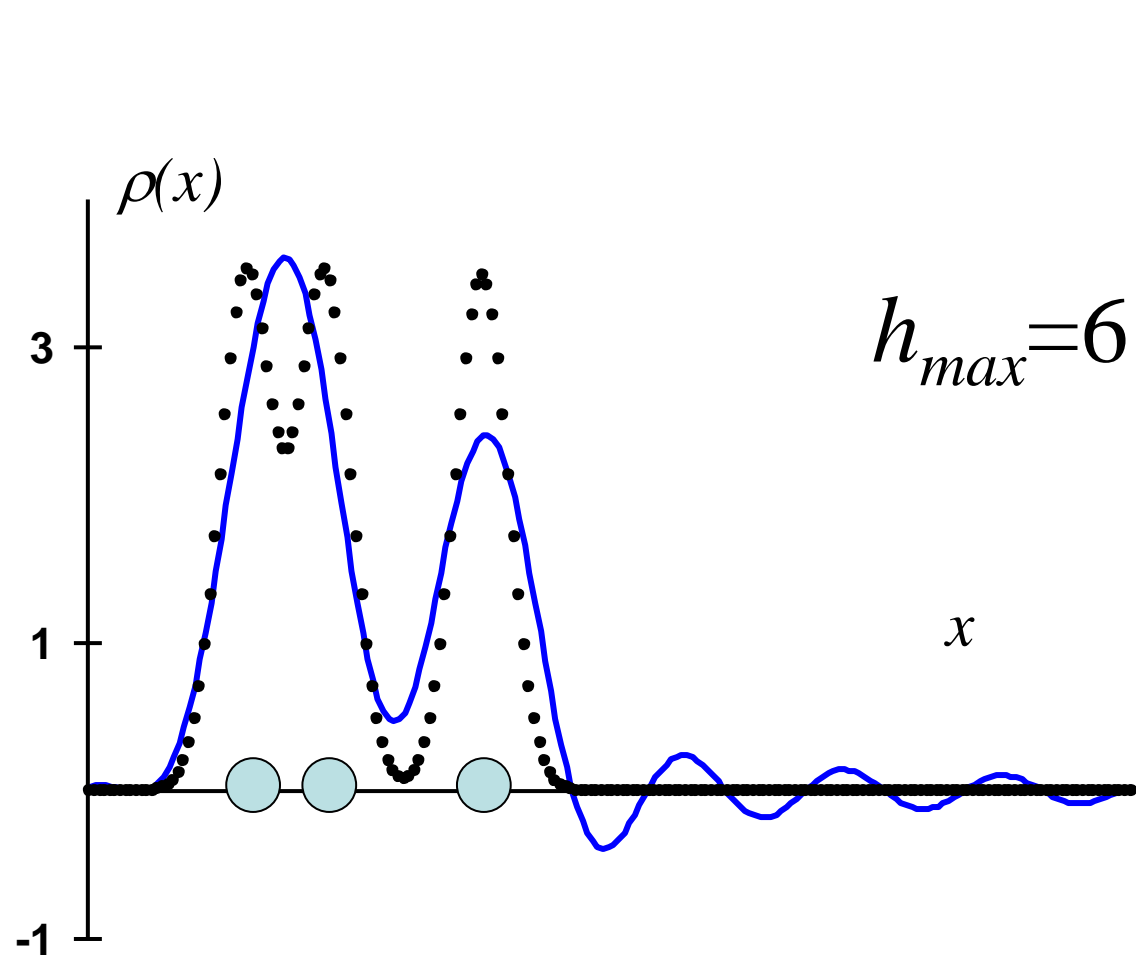
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



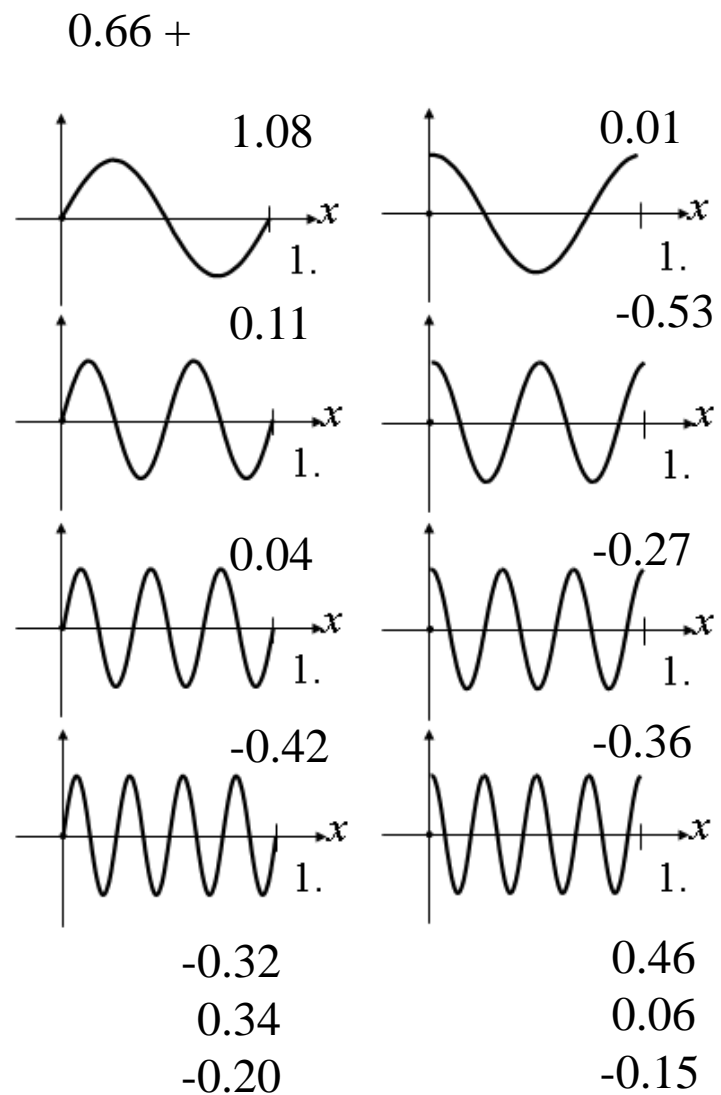
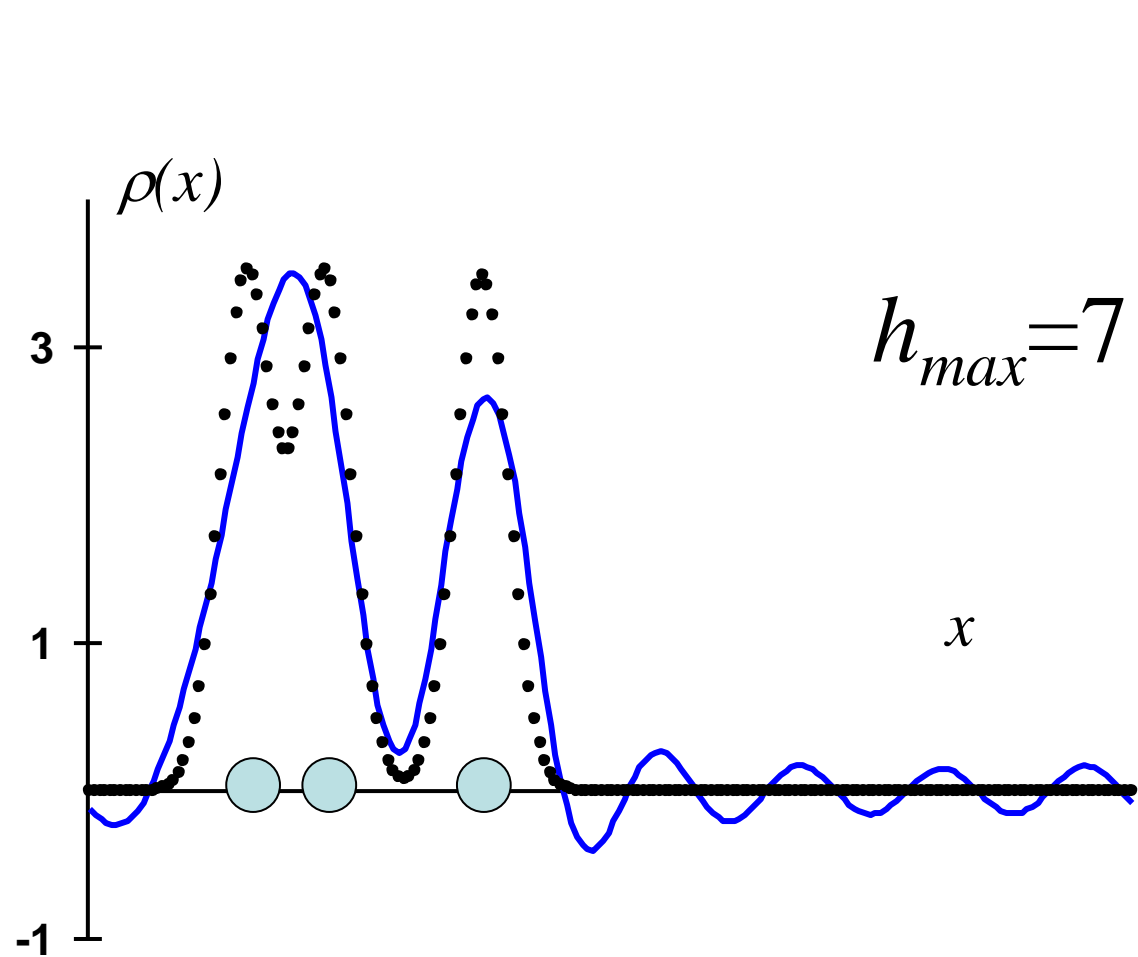
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



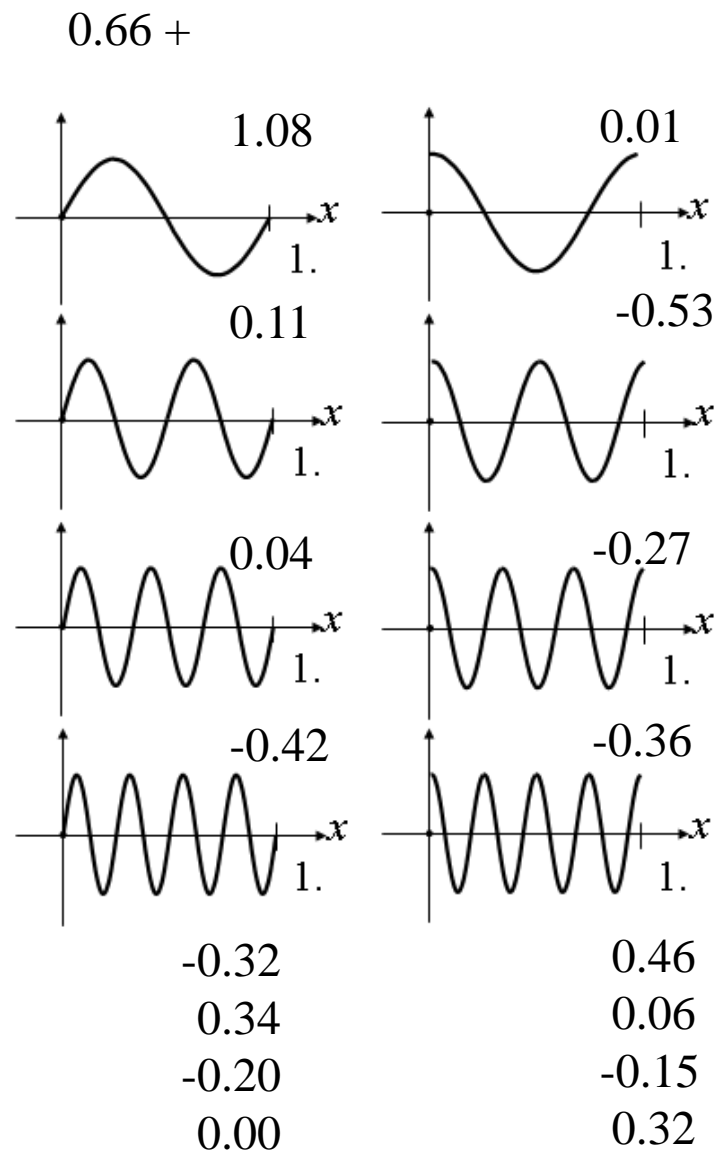
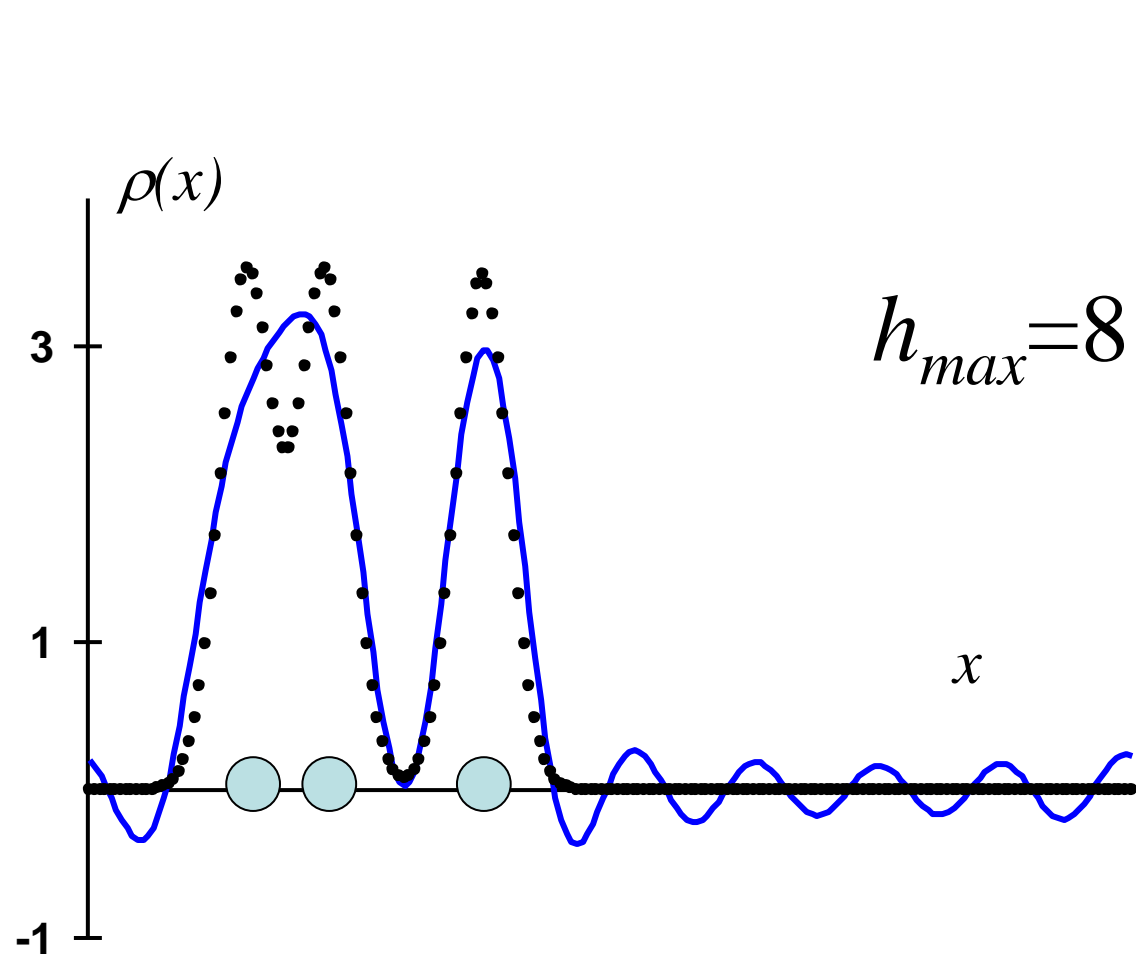
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



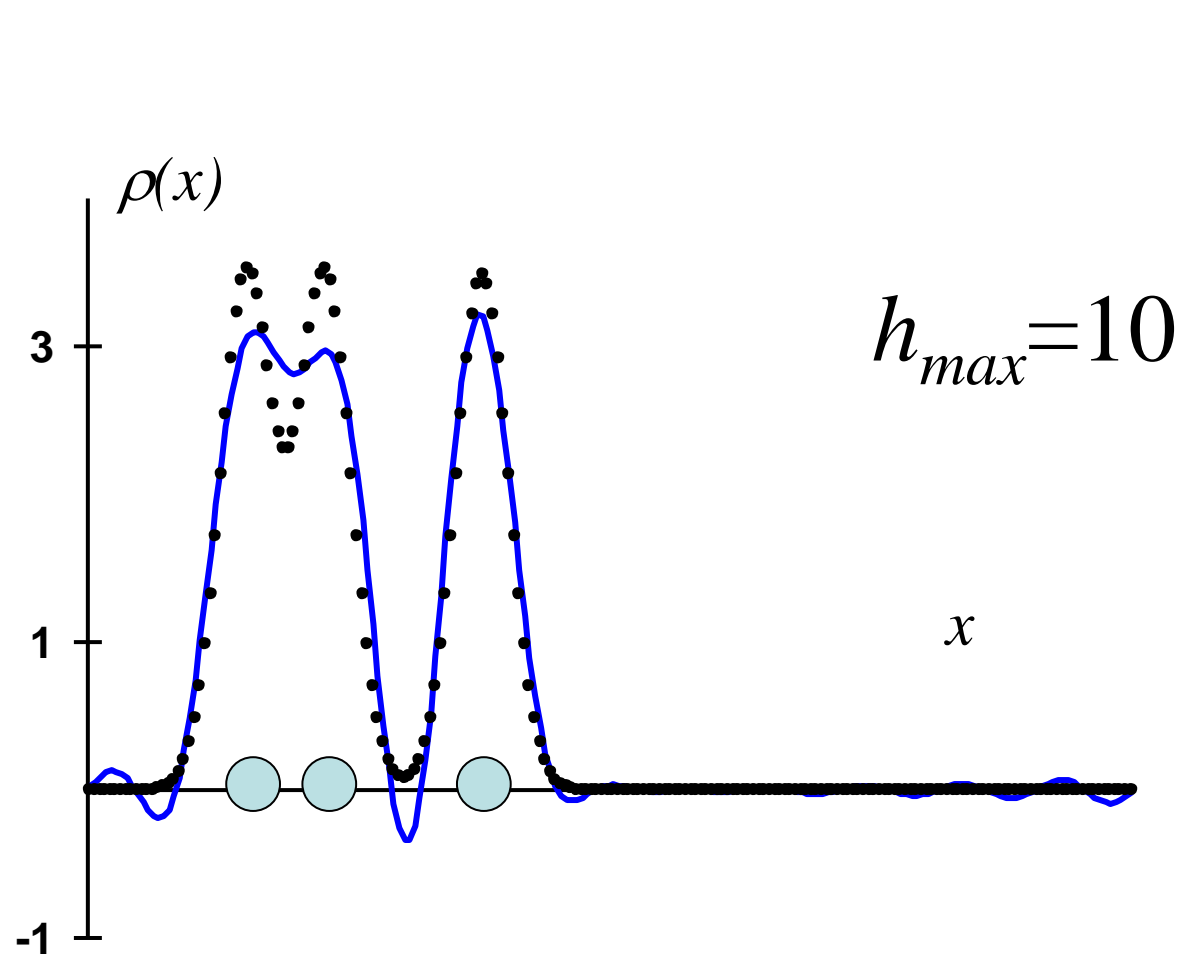
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



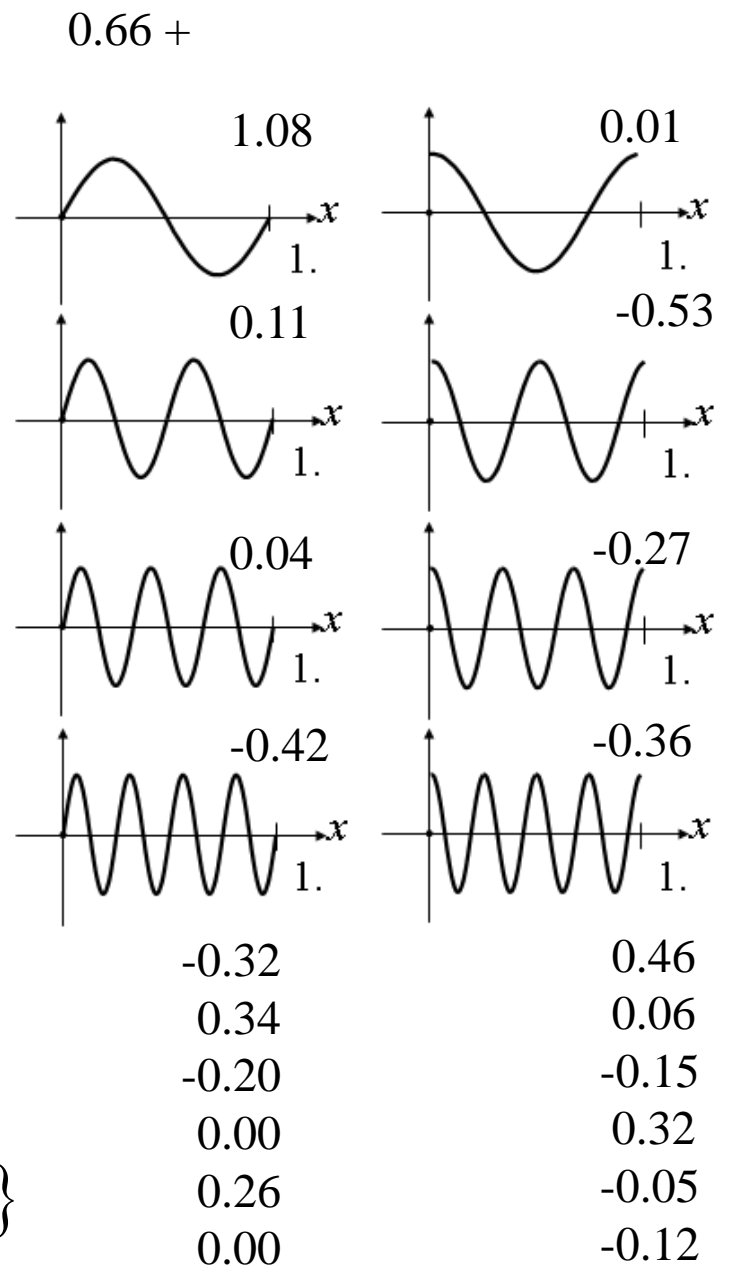
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$

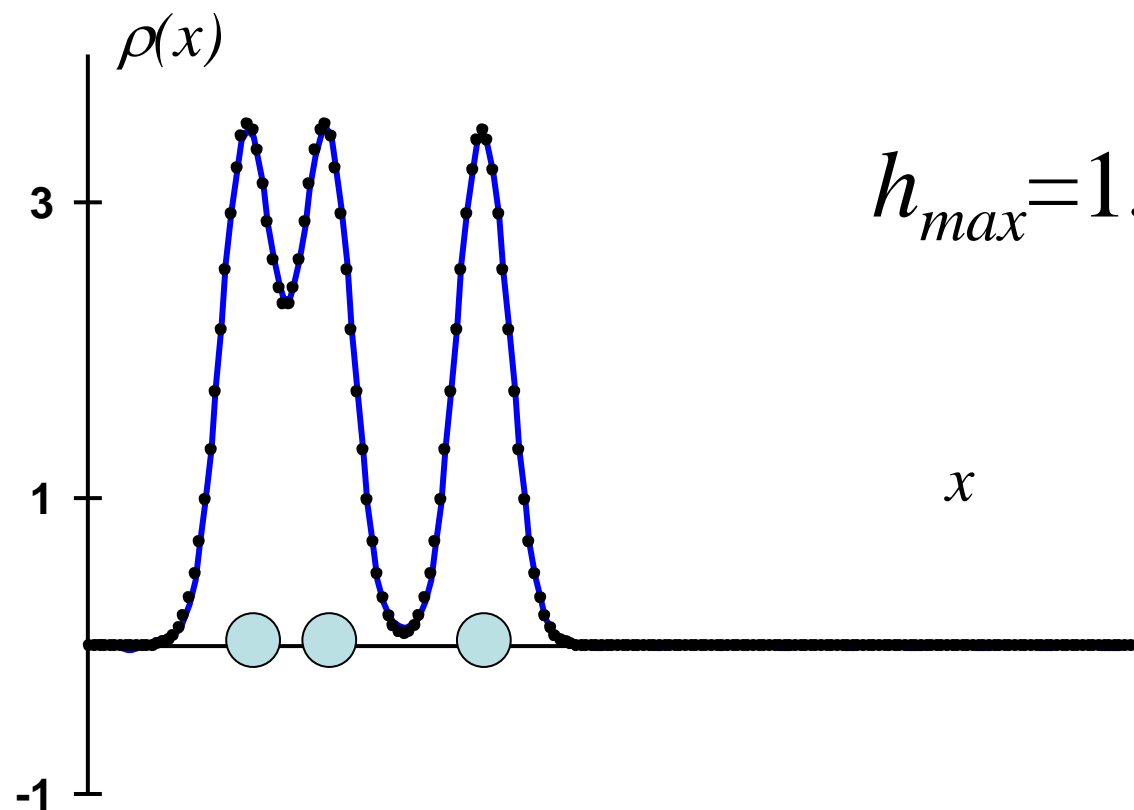


$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



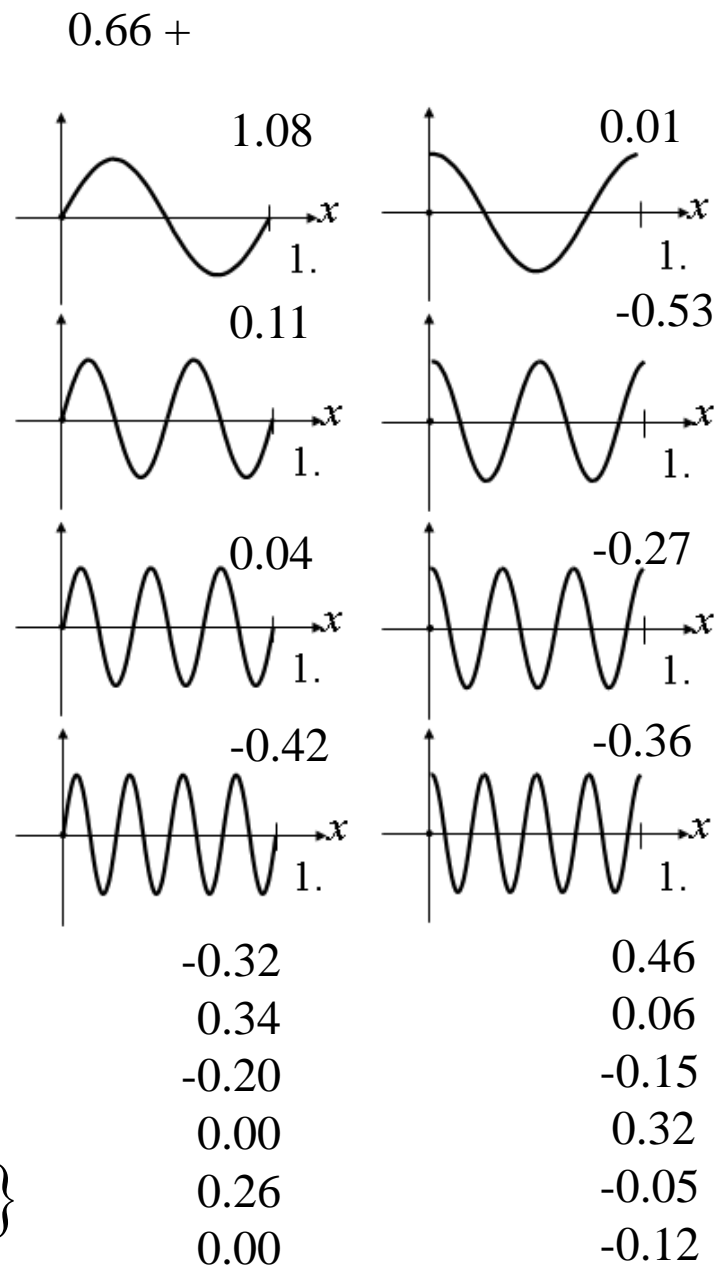
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



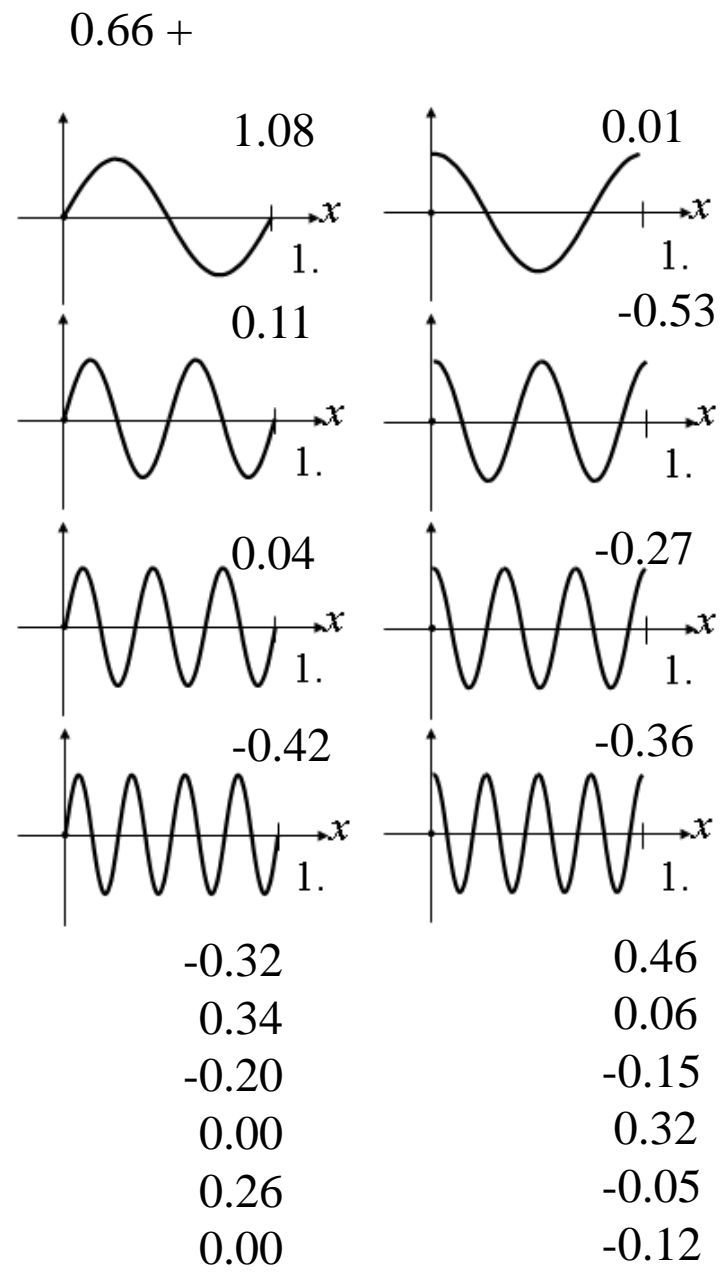
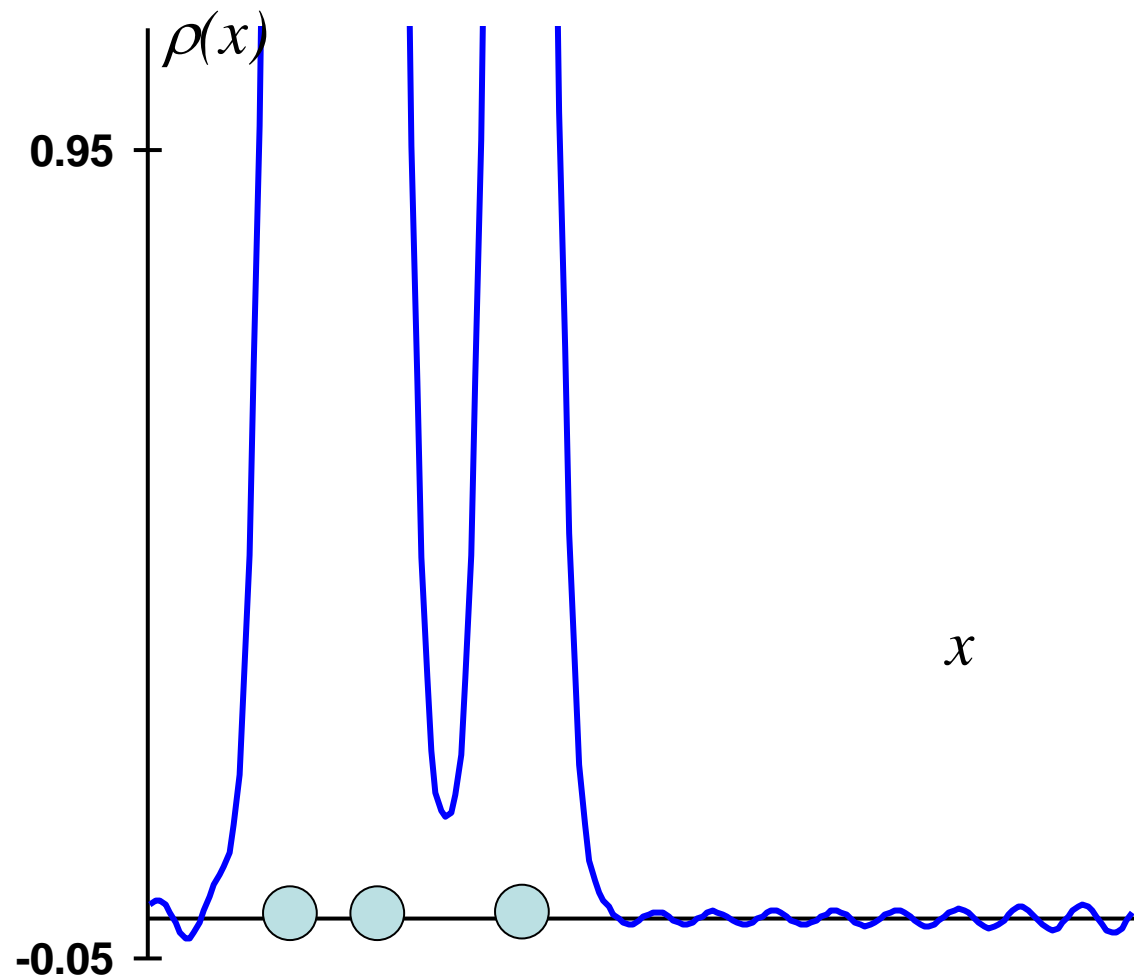


$$h_{max} = 15$$

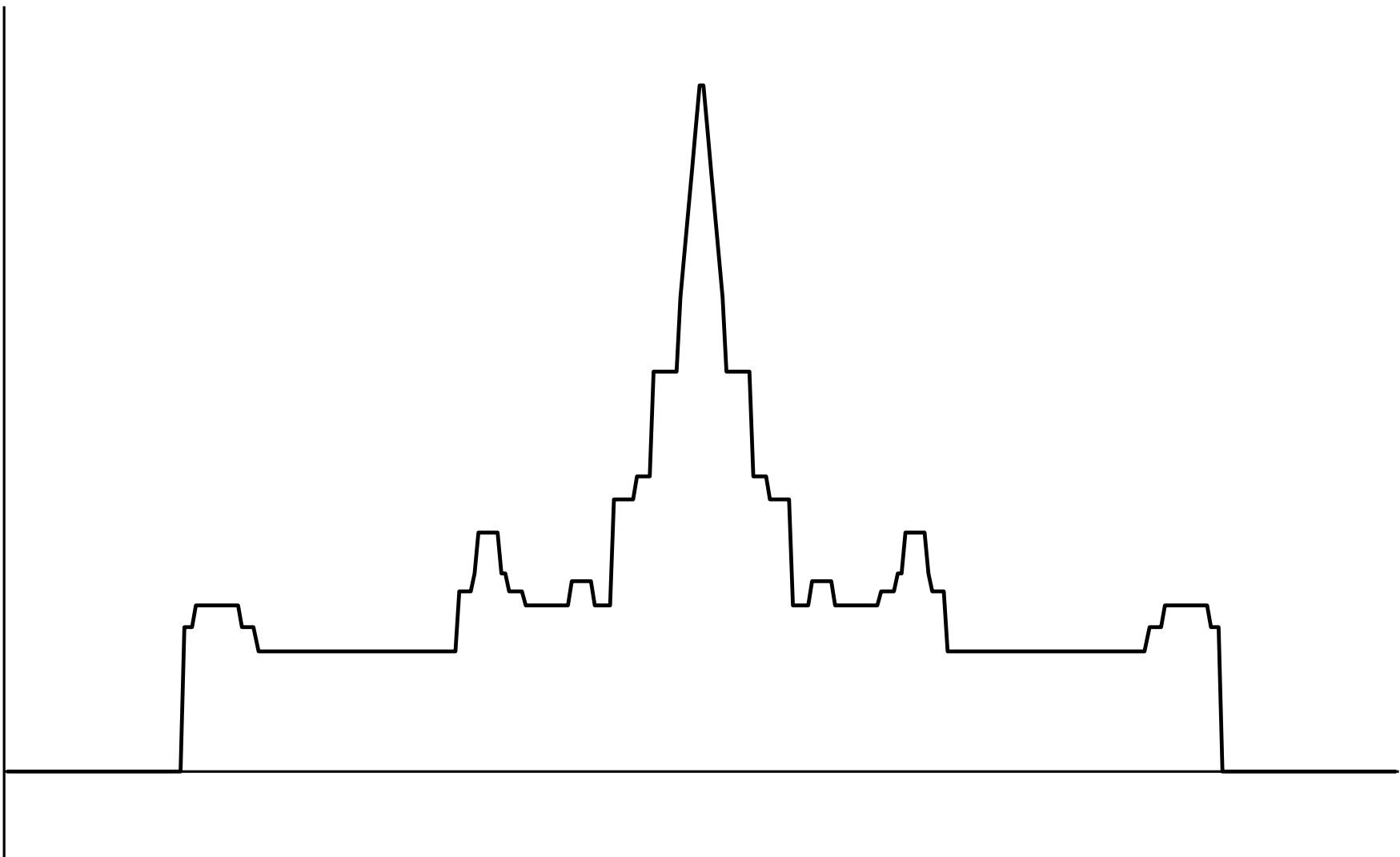
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$

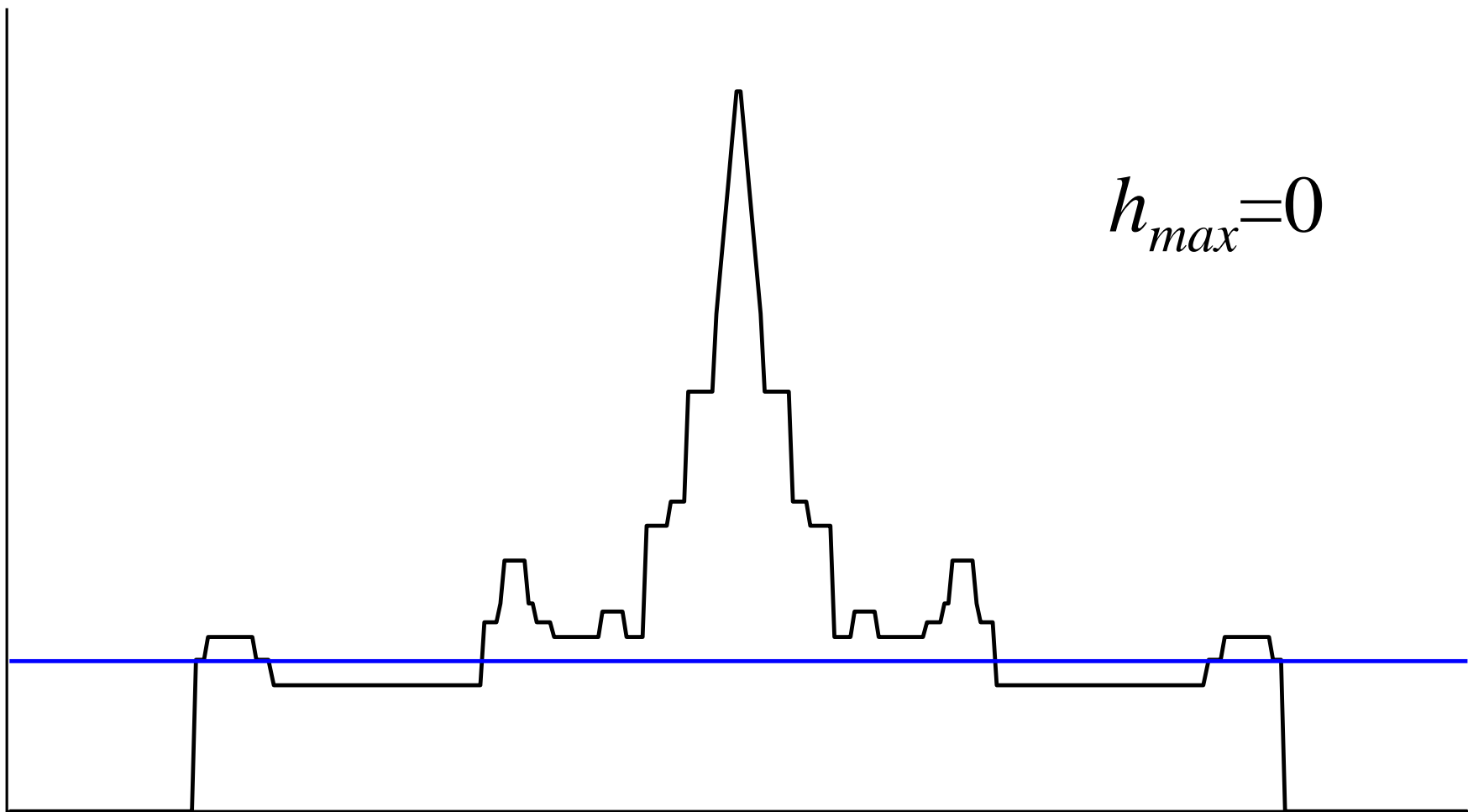


+ ...



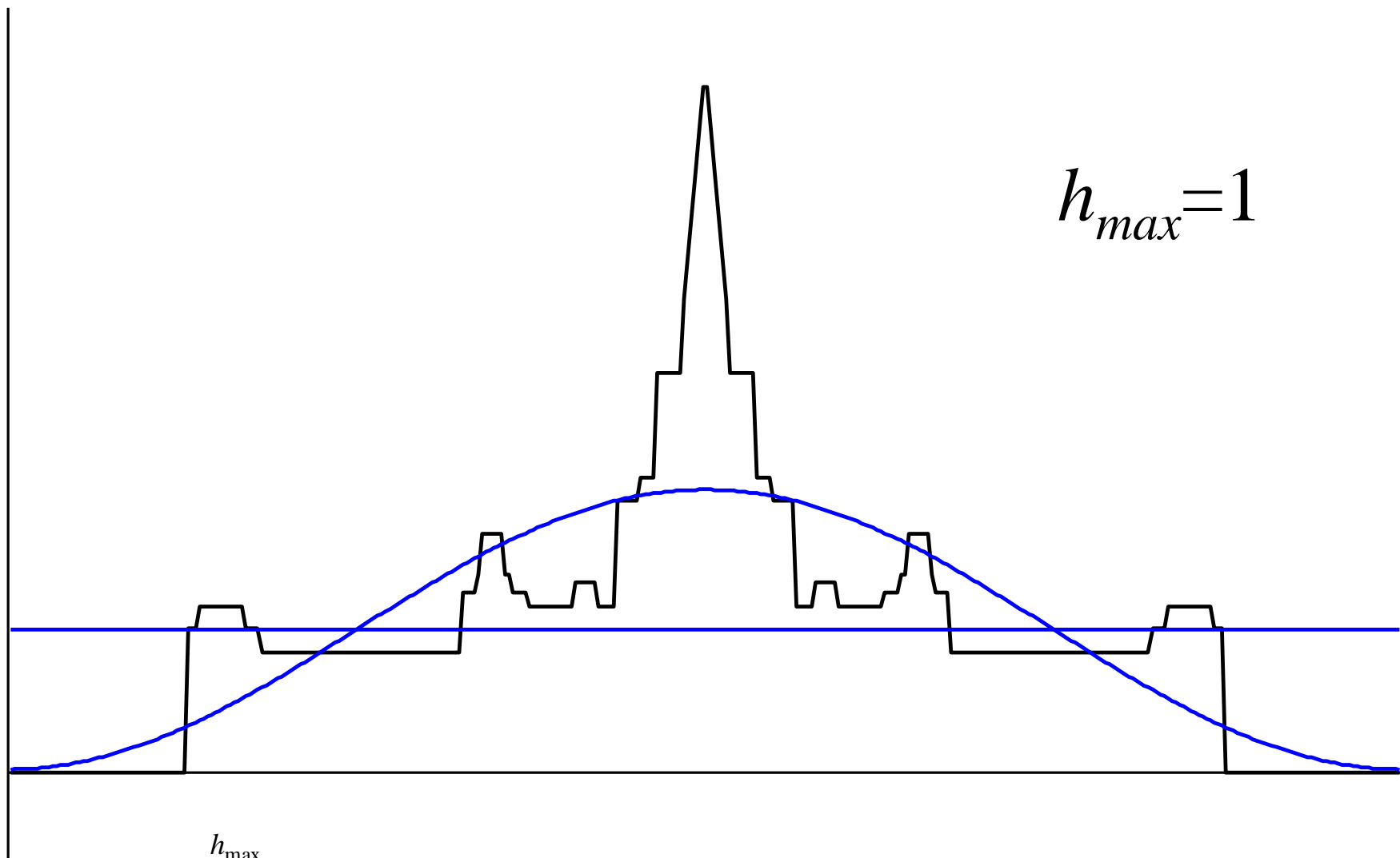
+ ...



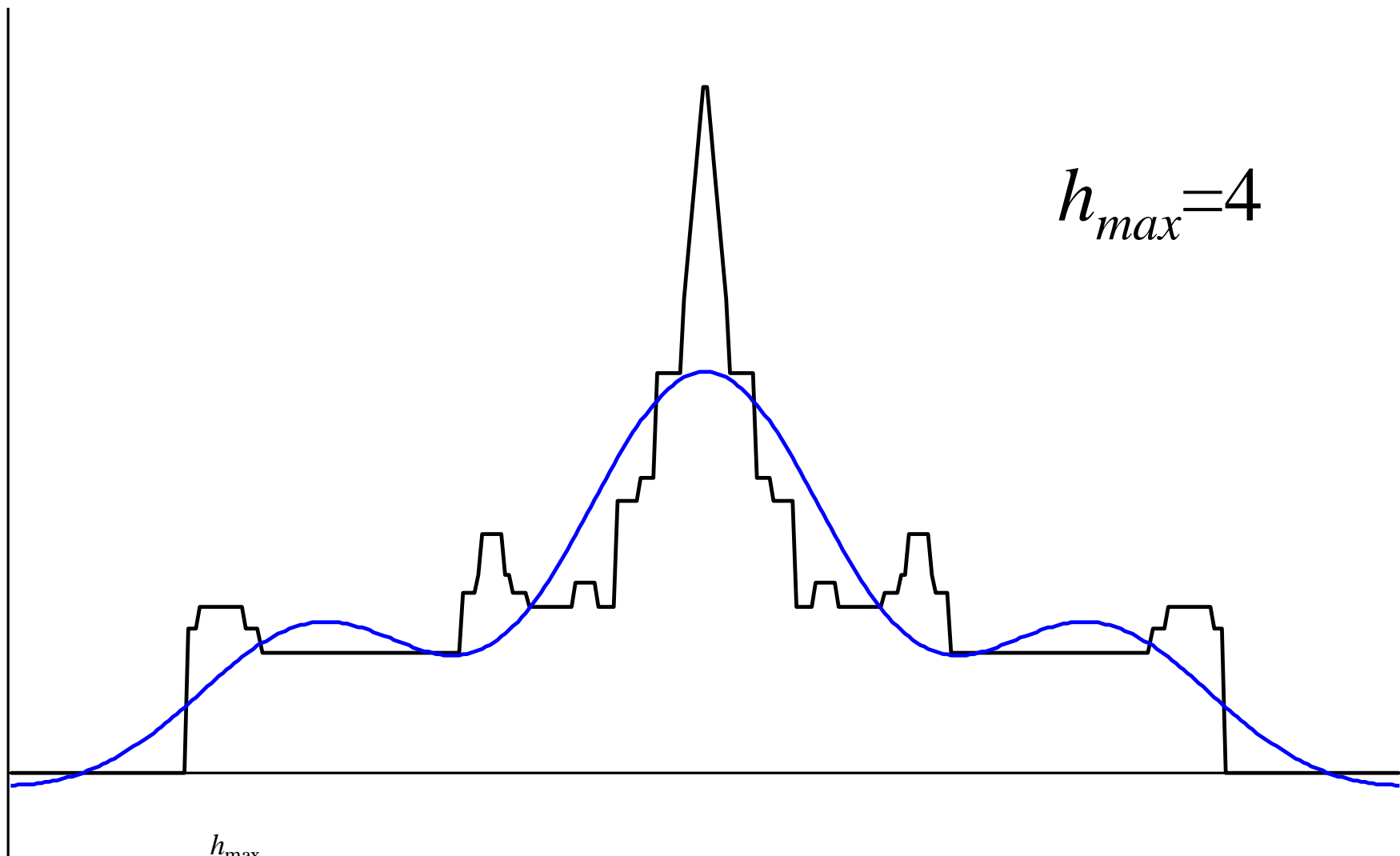


$h_{max} = 0$

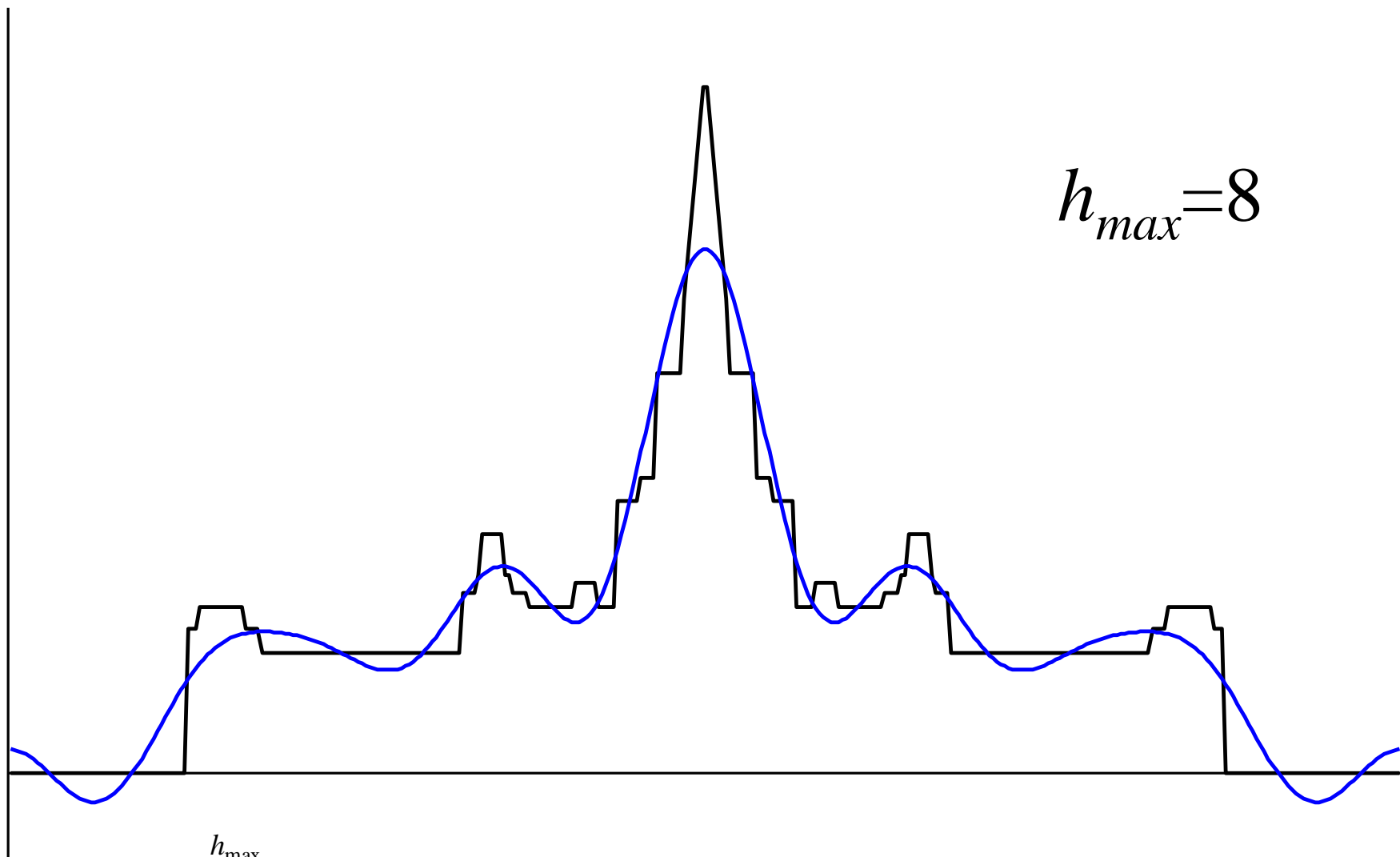
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



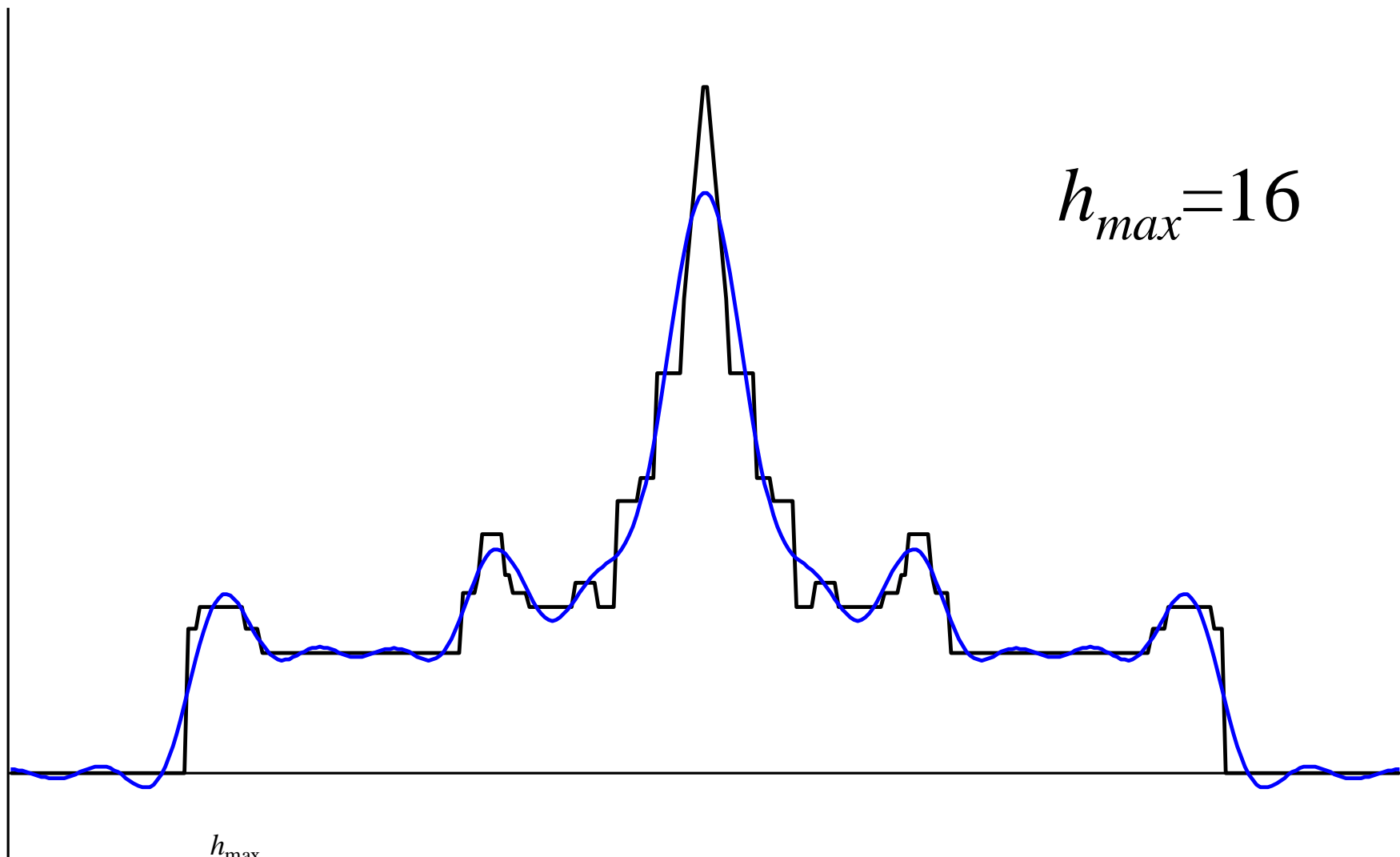
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{\max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



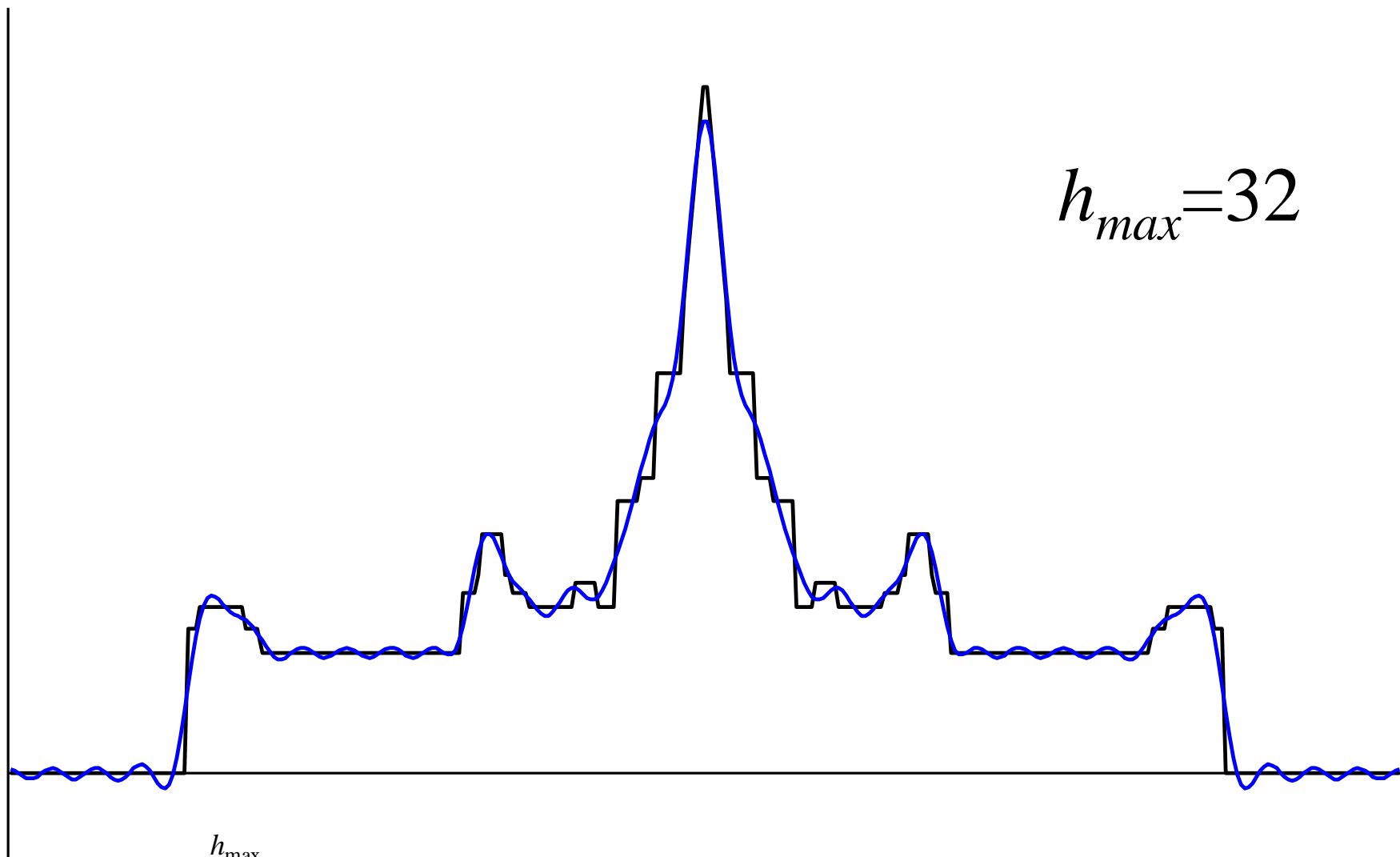
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{\max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{\max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



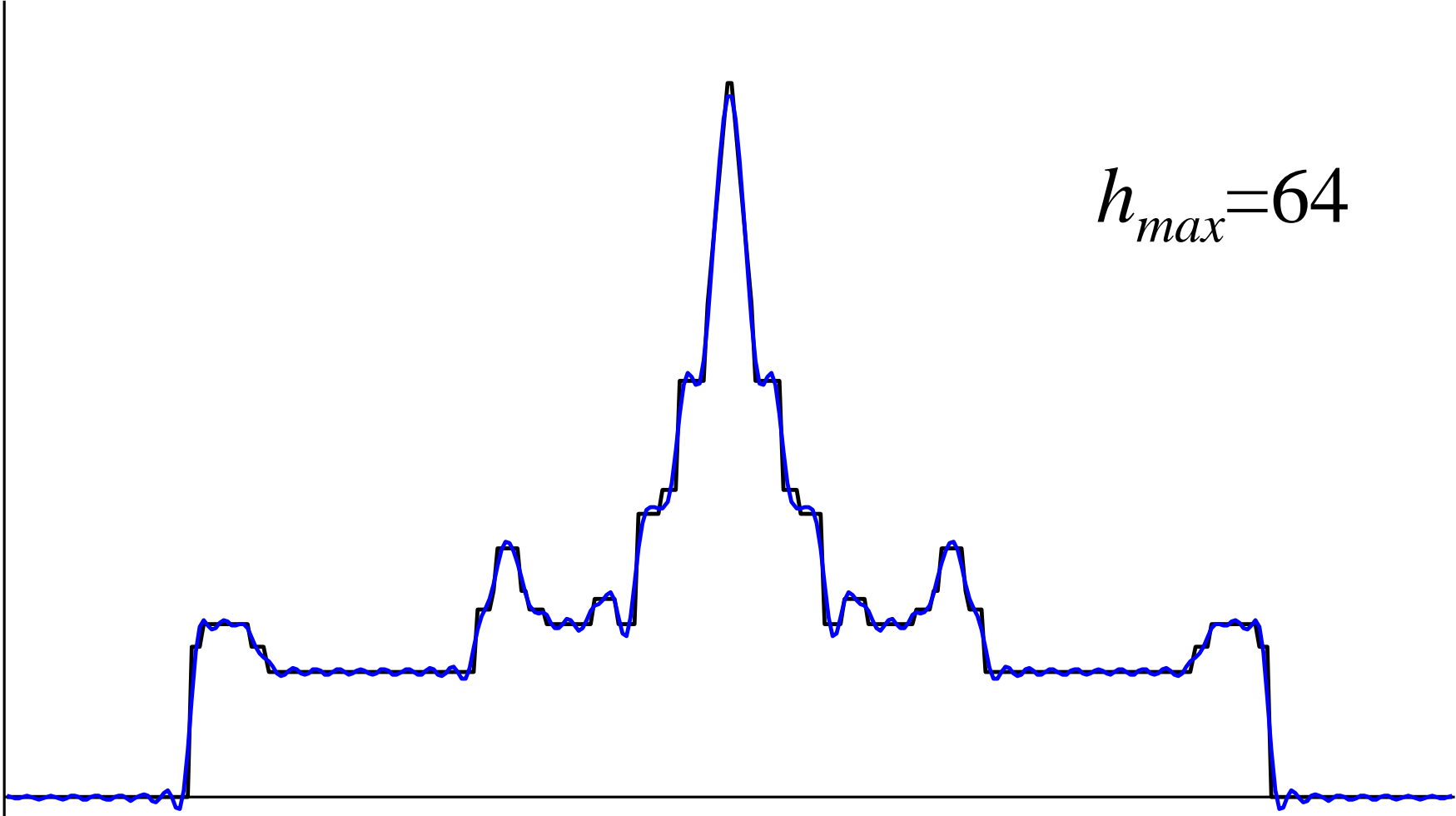
$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{\max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$



$h_{max} = 32$

$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$

$h_{max}=64$


$$\rho(x) \approx A_0 + \sum_{h=1}^{h_{max}} \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$

Вычисление коэффициентов ряда Фурье

Любая функция на отрезке $[0,1]$ может быть разложена, при этом единственным образом, в ряд Фурье.

Фурье – синтез.

$$\rho(x) = \sum_h \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$

Коэффициенты разложения **ВЫЧИСЛЯЮТСЯ** по формулам

$$A_h = \int_0^1 \rho(x) \cos[2\pi(hx)] dx \quad B_h = \int_0^1 \rho(x) \sin[2\pi(hx)] dx$$

Коэффициенты не надо угадывать, их можно вычислить ! Фурье – анализ.

Модуль и фаза коэффициента Фурье

$$\rho(x) = \sum_h \{A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx)\}$$

$$A_h \cos 2\pi(hx) + B_h \sin 2\pi(hx) = F_h \cos[2\pi(hx) - \varphi_h]$$

$$A_h = F_h \cos \varphi_h = \int_0^1 \rho(x) \cos [2\pi(hx)] dx$$

$$B_h = F_h \sin \varphi_h = \int_0^1 \rho(x) \sin [2\pi(hx)] dx$$

$$F_h = \sqrt{A_h^2 + B_h^2} \quad \text{tg } \varphi_h = \frac{B_h}{A_h}$$

$$\rho(x) = \sum_h F_h \cos[2\pi(hx) - \varphi_h]$$

F_h, φ_h - модуль и фаза коэффициента Фурье, отвечающего частоте h

Вся информация об объекте – в структурном факторе

$$F(\mathbf{s}), \varphi(\mathbf{s})$$

$$A = F(\mathbf{s}) \cos \varphi(\mathbf{s}) =$$

$$\int \rho(\mathbf{r}) \cos 2\pi(\mathbf{s}, \mathbf{r}) dV_{\mathbf{r}}$$

$$B = F(\mathbf{s}) \sin \varphi(\mathbf{s}) =$$

$$\int \rho(\mathbf{r}) \sin 2\pi(\mathbf{s}, \mathbf{r}) dV_{\mathbf{r}}$$

$$\rho(x) = \sum_h F_h \cos[2\pi(hx) - \varphi_h]$$

$$A_h = F_h \cos \varphi_h = \int_0^1 \rho(x) \cos[2\pi(hx)] dx$$

$$B_h = F_h \sin \varphi_h = \int_0^1 \rho(x) \sin[2\pi(hx)] dx$$

Структурные факторы – совпадают с коэффициентами разложения функции распределения электронной плотности в ряд Фурье.

Если модули и фазы структурных факторов известны, то распределение электронной плотности может быть рассчитано по формуле

$$\rho(x) = \sum_h F_h \cos[2\pi(hx) - \varphi_h]$$

Фазовая проблема: значения фаз в эксперименте теряются, надо их как-то восстановить.

Рентгеноструктурный анализ

- Выделение и очистка белка
- Кристаллизация
- Дифракционный эксперимент
- Решение фазовой проблемы
- Расчет и интерпретация синтеза Фурье; построение предварительной атомной модели
- Уточнение атомной модели
- Проверка правильности модели

Ряды Фурье для функций нескольких переменных

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

$$\tilde{\rho}(x, y, z) = \rho(x\mathbf{a} + y\mathbf{b} + z\mathbf{c})$$

x, y, z – относительные
(кристаллографические)
координаты

элементарная ячейка – куб
 $0 \leq x, y, z \leq 1$

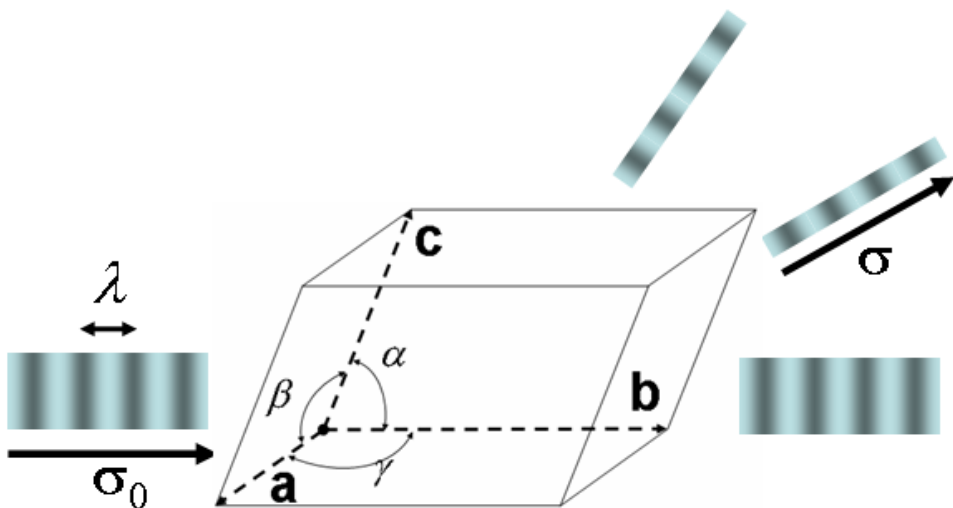
$\tilde{\rho}(x, y, z)$ имеет период 1. по всем переменным.

$$\tilde{\rho}(x, y, z) \approx F_{000} + \sum_{hkl} F_{hkl} \cos \left[2\pi (hx + ky + lz) - \varphi_{hkl} \right]$$

$$F_{hkl} = \sqrt{A_{hkl}^2 + B_{hkl}^2} \quad \operatorname{tg} \varphi_{hkl} = \frac{B_{hkl}}{A_{hkl}}$$

$$A_{hkl} = \int_{0 \leq x, y, z \leq 1} \tilde{\rho}(x, y, z) \cos \left[2\pi (hx + ky + lz) \right] dx dy dz$$

$$B_{hkl} = \int_{0 \leq x, y, z \leq 1} \tilde{\rho}(x, y, z) \sin \left[2\pi (hx + ky + lz) \right] dx dy dz$$



Рентгеновский эксперимент с монокристаллом позволяет измерить интенсивность волн, рассеянных в направлениях, определяемых условиями:

h, k, l - целые

$$h = (\mathbf{s}, \mathbf{a}), \quad k = (\mathbf{s}, \mathbf{b}), \quad l = (\mathbf{s}, \mathbf{c})$$

$$\mathbf{s} = (\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}) / \lambda$$

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

Амплитуда и фаза рассеянной волны зависят от распределения электронной плотности в элементарной ячейке и могут быть рассчитаны по формулам

$$F_{hkl} = \sqrt{A_{hkl}^2 + B_{hkl}^2} \quad \text{tg } \varphi_{hkl} = \frac{B_{hkl}}{A_{hkl}}$$

$$A_{hkl} = \int_V \rho(x, y, z) \cos[2\pi(hx + ky + lz)] dx dy dz$$

$$B_{hkl} = \int_V \rho(x, y, z) \sin[2\pi(hx + ky + lz)] dx dy dz$$

Структурные факторы, отвечающие Брэгговским рефлексам совпадают с коэффициентами в разложении в ряд Фурье функции распределения электронной плотности.

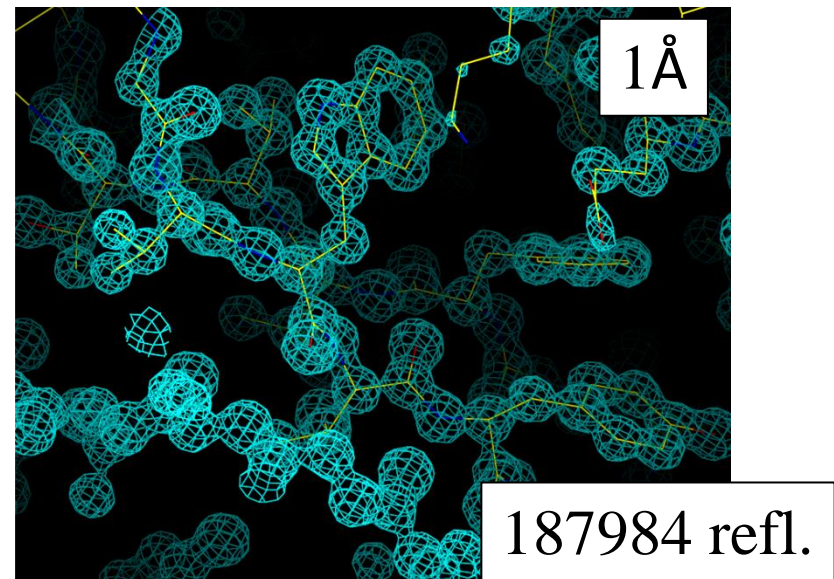
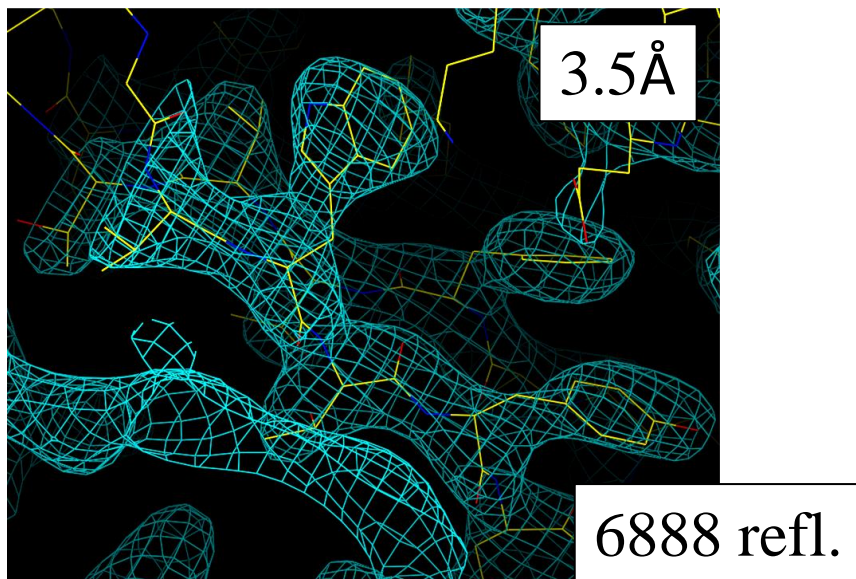
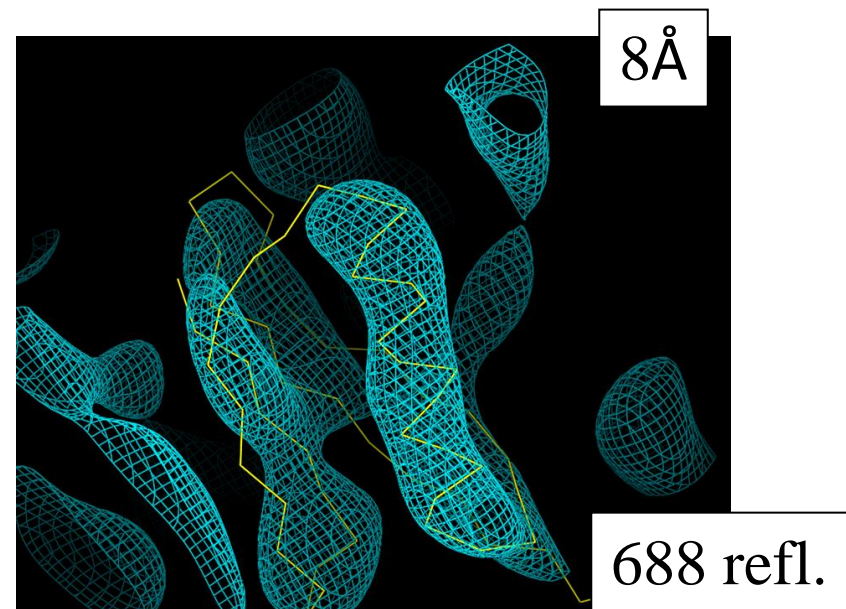
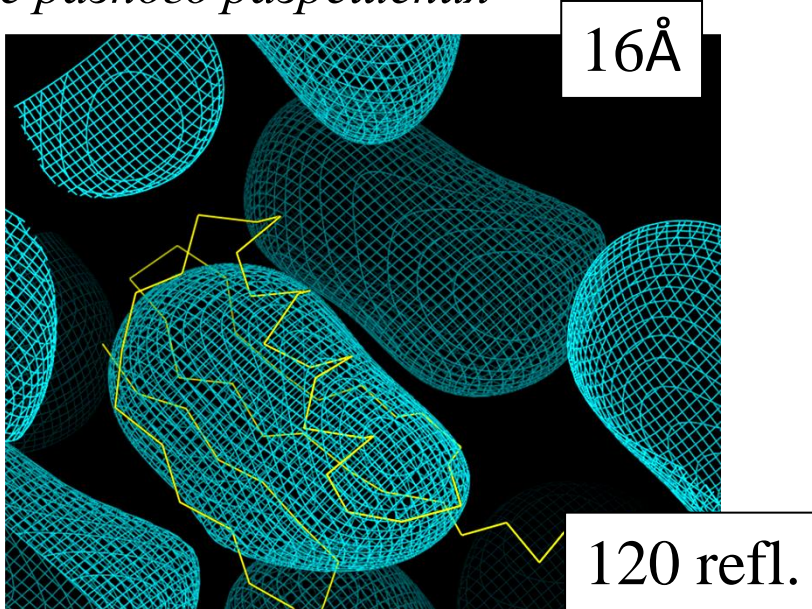
$$\rho(x, y, z) = \sum_{\substack{hkl \\ \text{целые}}} F_{hkl} \cos[2\pi(hx + ky + lz) - \varphi_{hkl}]$$

Знание значений модулей и фаз структурных факторов дает возможность рассчитать "синтез Фурье электронной плотности"

$$\tilde{\rho}_S(x, y, z) \approx \sum_{(hkl) \in S} F_{hkl} \cos[2\pi(hx + ky + lz) - \varphi_{hkl}]$$

Синтез Фурье рассчитывается по набору рефлексов S , для которых известны и модули и фазы с.ф.

Изображение областей высокой плотности для белка Protein G, полученные на основе синтезов Фурье разного разрешения

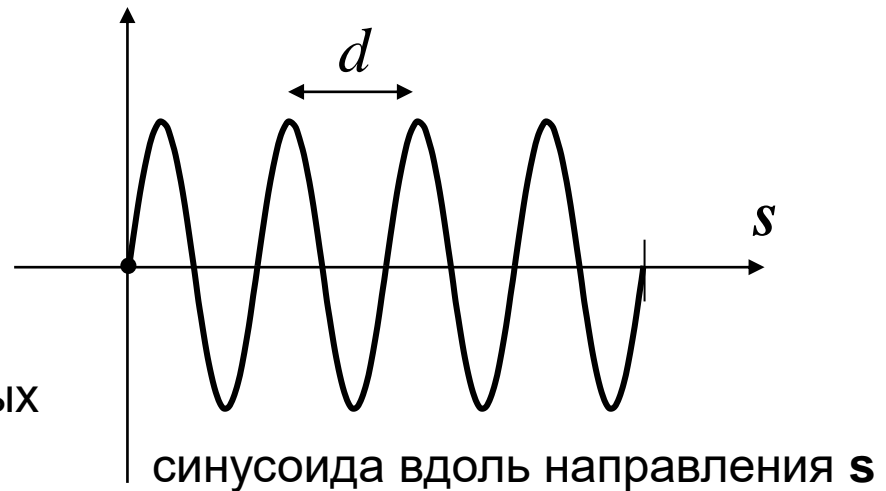
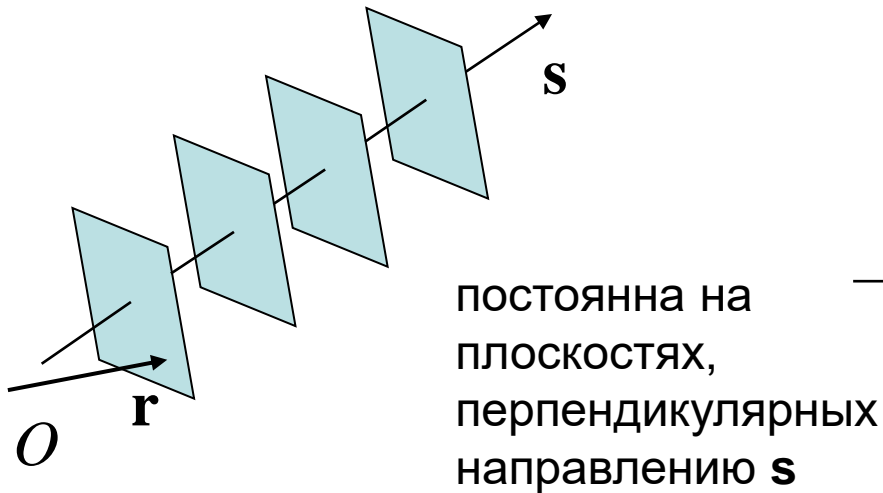


Гармоника Фурье

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

$$H_s(\mathbf{r}) = \cos[2\pi(hx + ky + lz)] = \cos[2\pi(\mathbf{s}, \mathbf{r})]$$

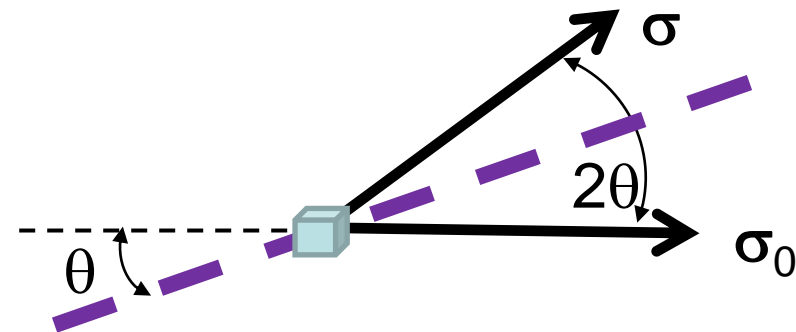
$$h = (\mathbf{s}, \mathbf{a}), k = (\mathbf{s}, \mathbf{b}), l = (\mathbf{s}, \mathbf{c})$$



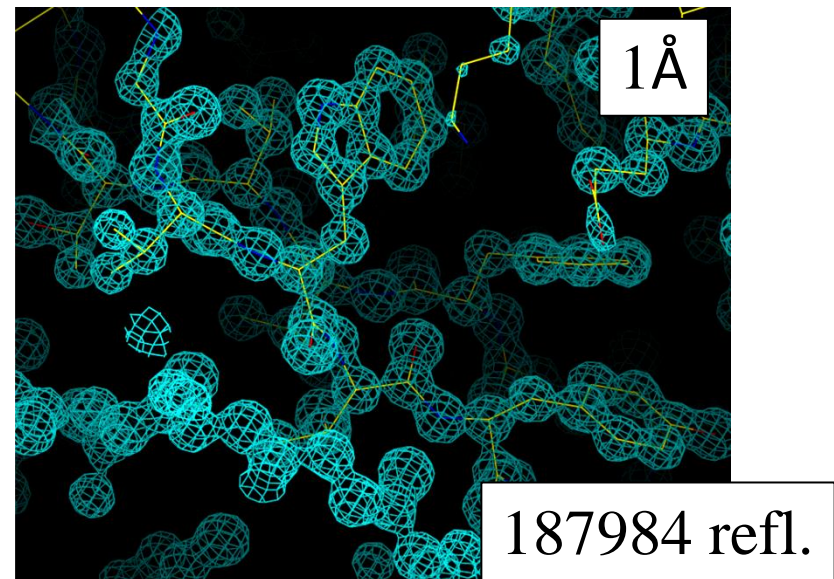
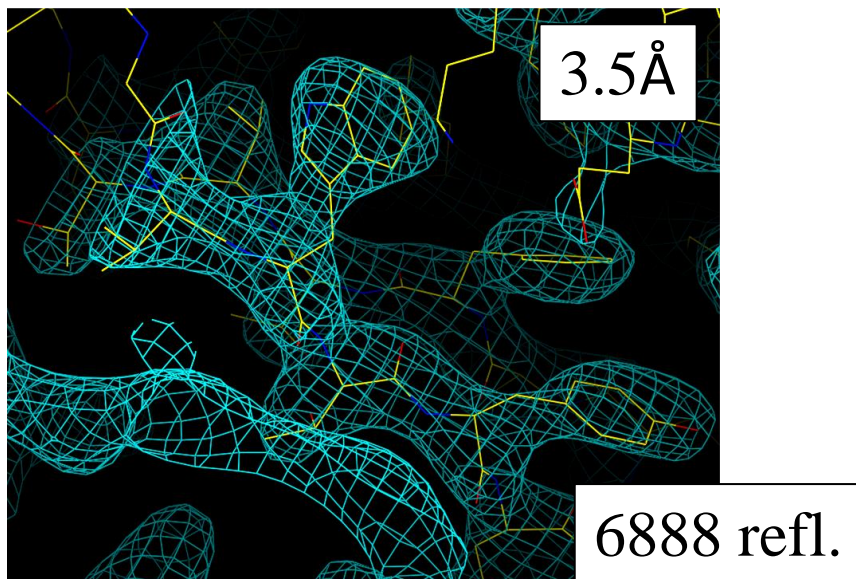
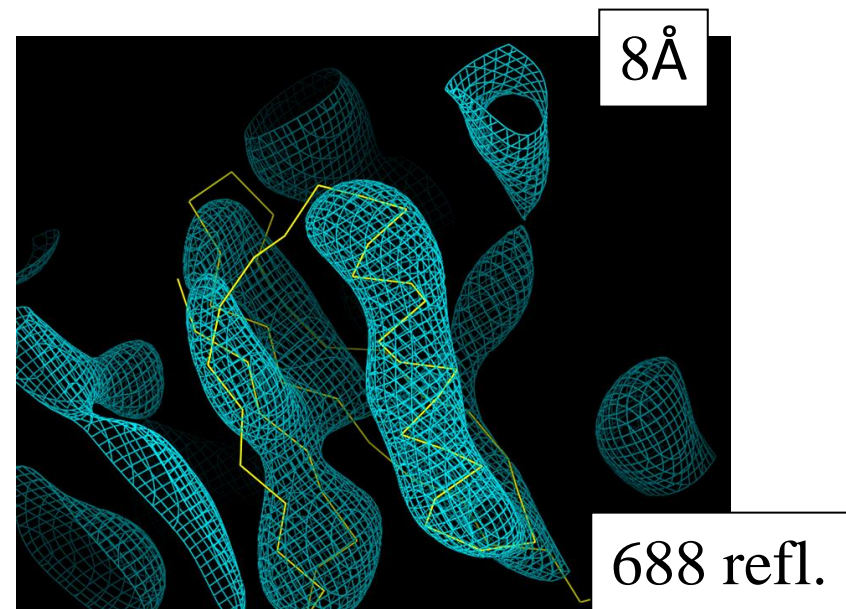
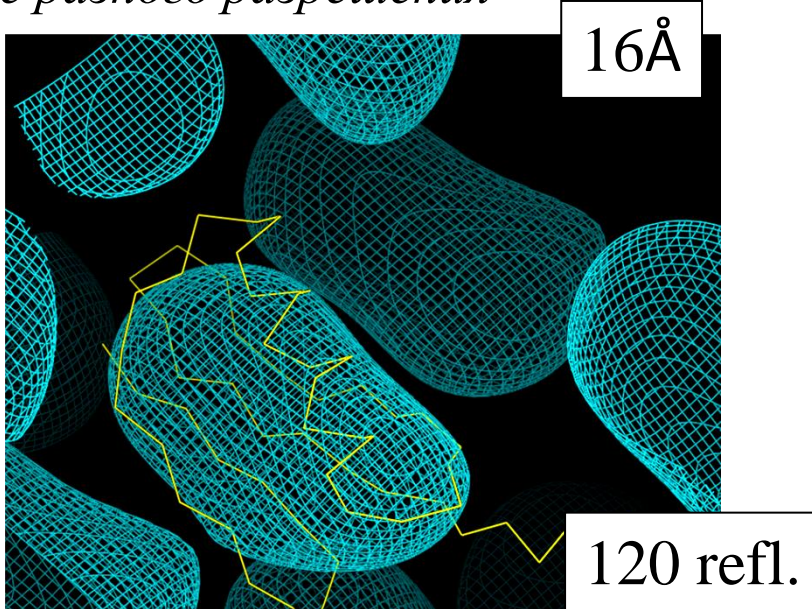
Разрешение, соответствующее гармонике Фурье

$d = \frac{1}{|\mathbf{s}|}$ - расстояние между соседними максимумами в направлении \mathbf{s} ;

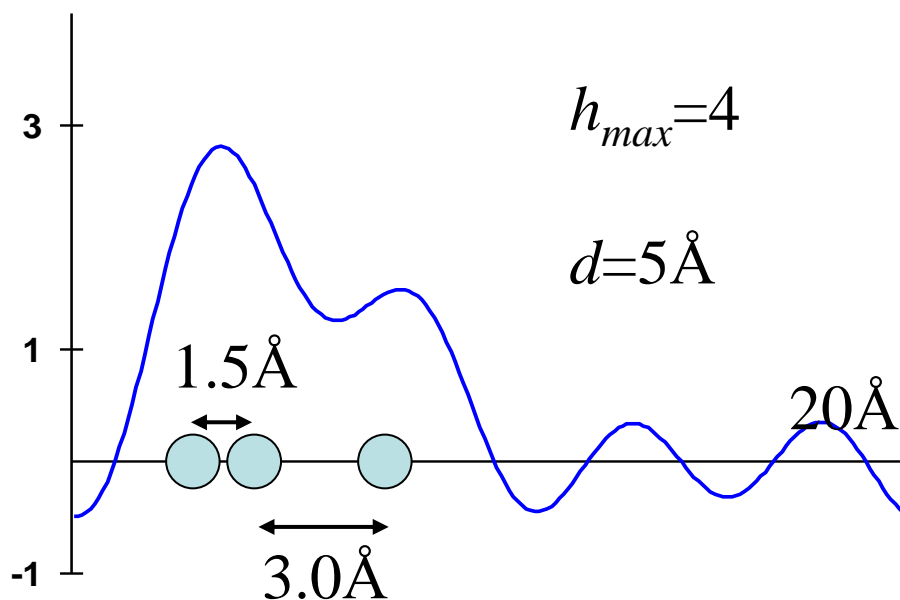
$$|\mathbf{s}| = \frac{2 \sin \theta}{\lambda} \quad d = \frac{\lambda}{2 \sin \theta}$$



Изображение областей высокой плотности для белка Protein G, полученные на основе синтезов Фурье разного разрешения



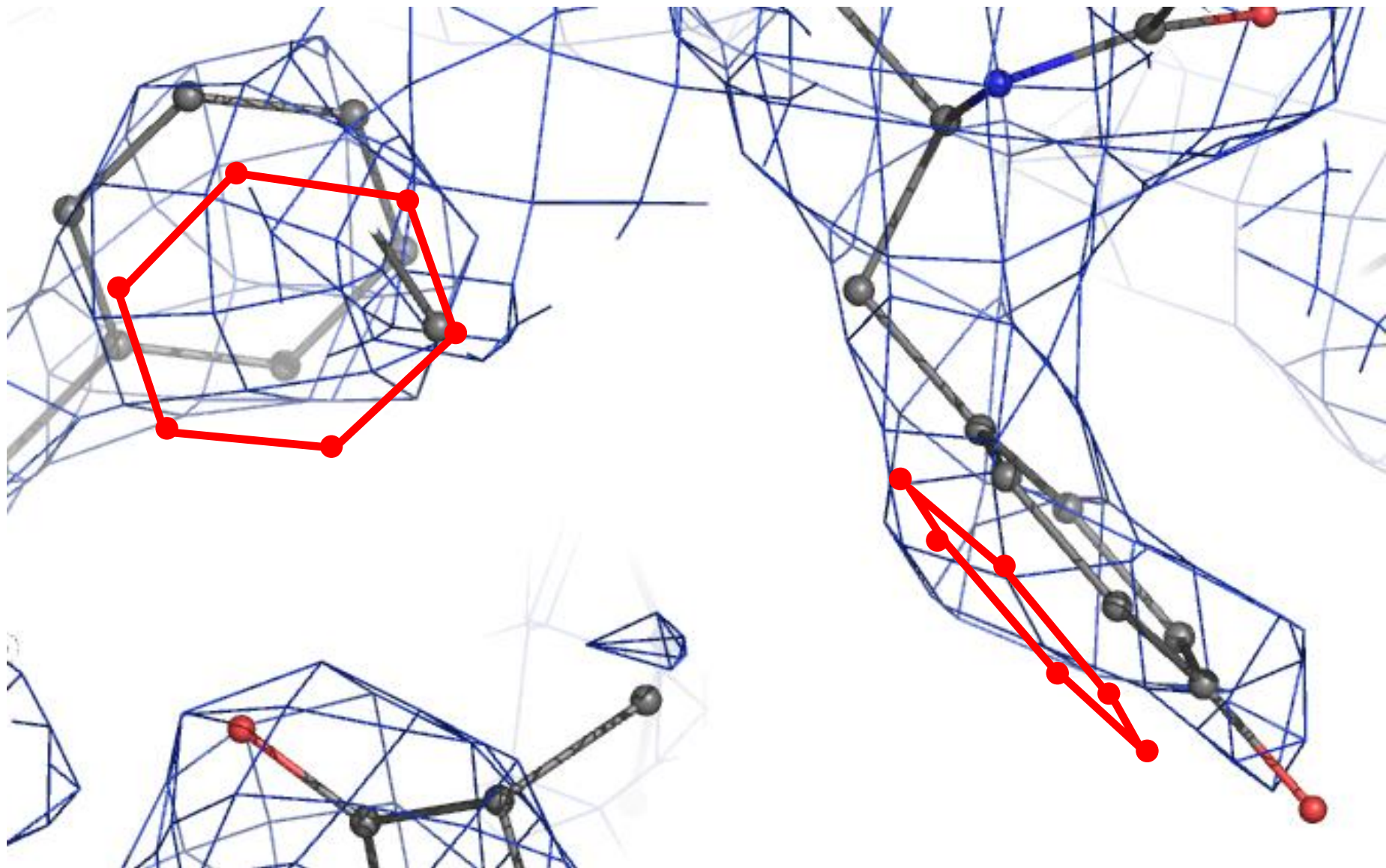
Разрешение набора структурных факторов характеризует размер деталей, «видимых» на картах распределения электронной плотности. Оно не связано прямо с точностью определения координат атомов методом РСА!



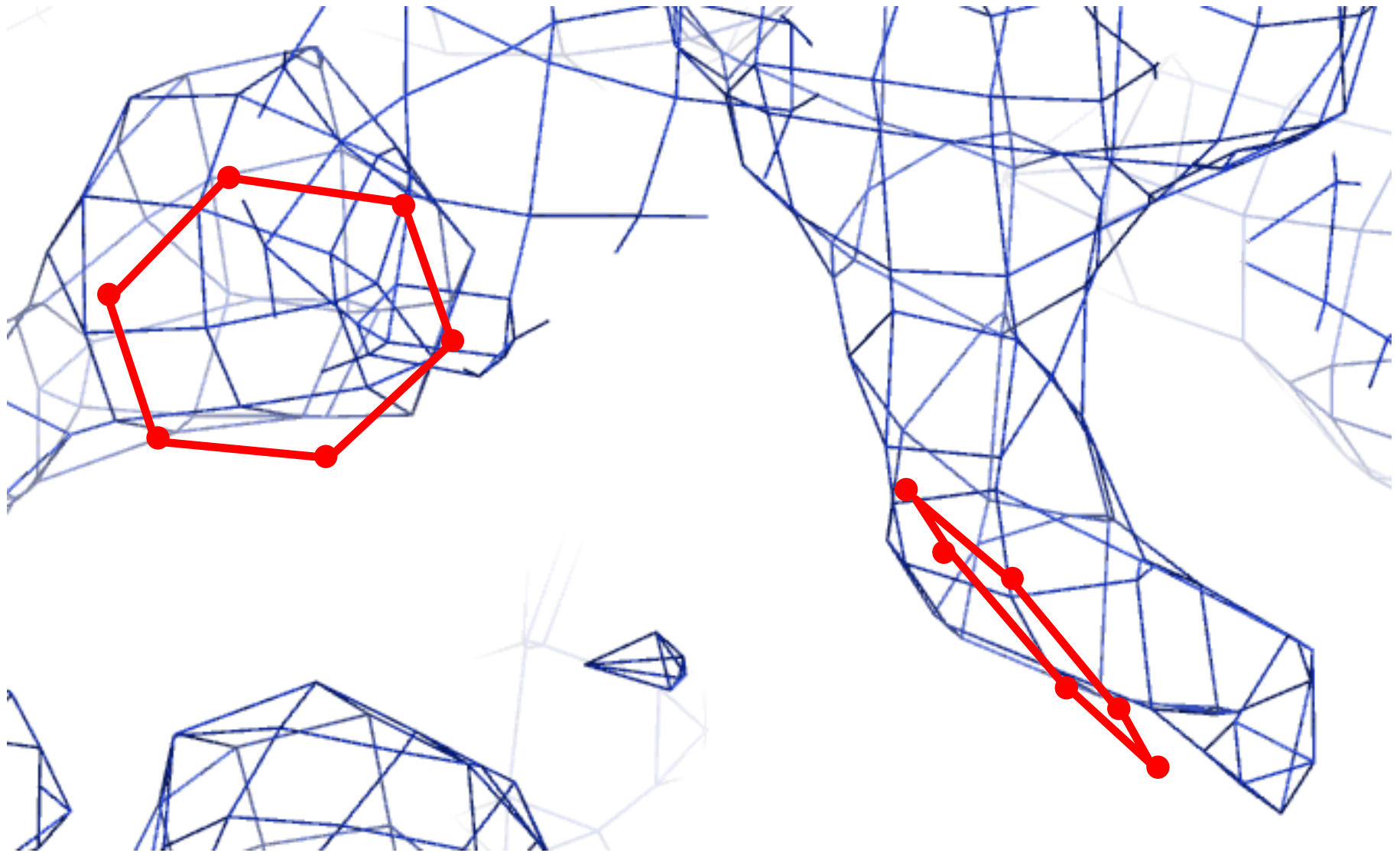
- электронная плотность является суммой вкладов атомов;
- форма вклада атома известна; неизвестны - координаты центра.



точность определения координат выше, нежели номинальное разрешение синтеза Фурье



RNase sa, 2.5Å, MIR



RNase sa, 2.5Å, MIR

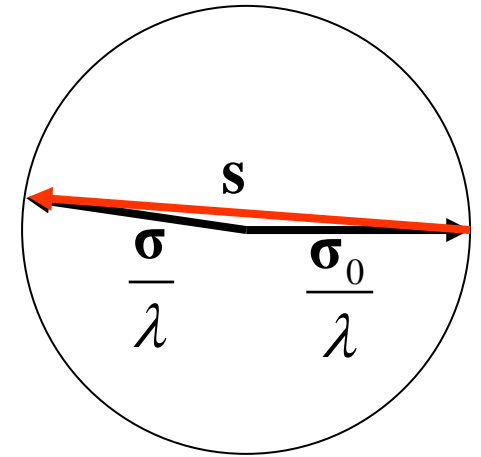
$$|\mathbf{s}|_{\max} = \frac{2}{\lambda}$$

$$d_{\min}^{\text{theoretical}} = \frac{\lambda}{2}$$

$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda}$$

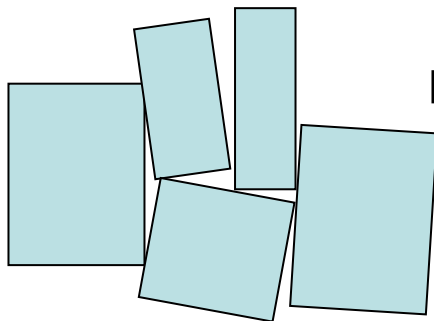
$$|\boldsymbol{\sigma}| = |\boldsymbol{\sigma}_0| = 1$$

$$d = \frac{1}{|\mathbf{s}|}$$

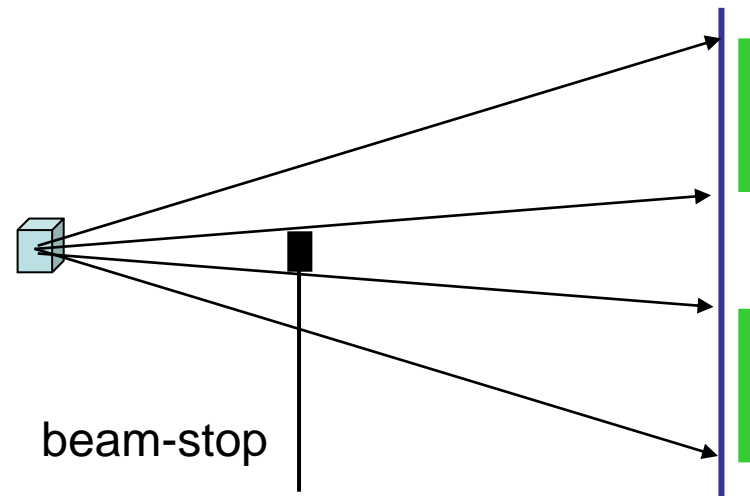


Разрешение:

- теоретический предел $\lambda/2$;
- недостаток чувствительности детектора – ухудшает разрешение;
- дефекты кристалла – ухудшают разрешение;
- “beam-stop” мешает регистрировать низкоугловые рефлексы.



“МОЗАИЧНОСТЬ
кристалла”



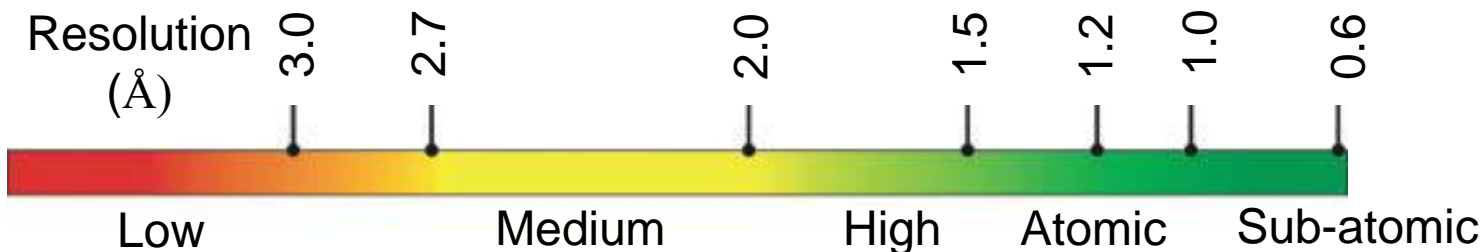
Разрешение набора структурных факторов, измеряемых в реальном эксперименте, определяется качеством кристалла.

crambin, 0.54 Å (Teeter et al., 1993); 0.48 Å (2011)

antifreeze protein, 0.62 Å (Ko et al., 2003);

aldose reductase, 0.66 Å (Howard et al., 2004)].

```
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.4
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 30.0
REMARK 3 DATA CUTOFF (SIGMA (F)) : NONE
REMARK 3 COMPLETENESS FOR RANGE (%) : 97.0
REMARK 3 NUMBER OF REFLECTIONS : 53449
REMARK 3
```



Фазовая проблема



Подходы к решению фазовой проблемы

SIR - Single Isomorphous Replacement
Однократное изоморфное замещение;

MIR - Multiple Isomorphous Replacement
Множественное изоморфное замещение;

AD - Anomalous Dispersion
Аномальное рассеяние;

MAD - Multiwavelength Anomalous Diffraction
Многоволновое аномальное рассеяние;

MR - Molecular Replacement
Молекулярное замещение;

Direct Methods (*ab initio* methods)
"Прямые" методы.

Молекулярное замещение

Molecular replacement

Метод молекулярного замещения

1. Постулаты:

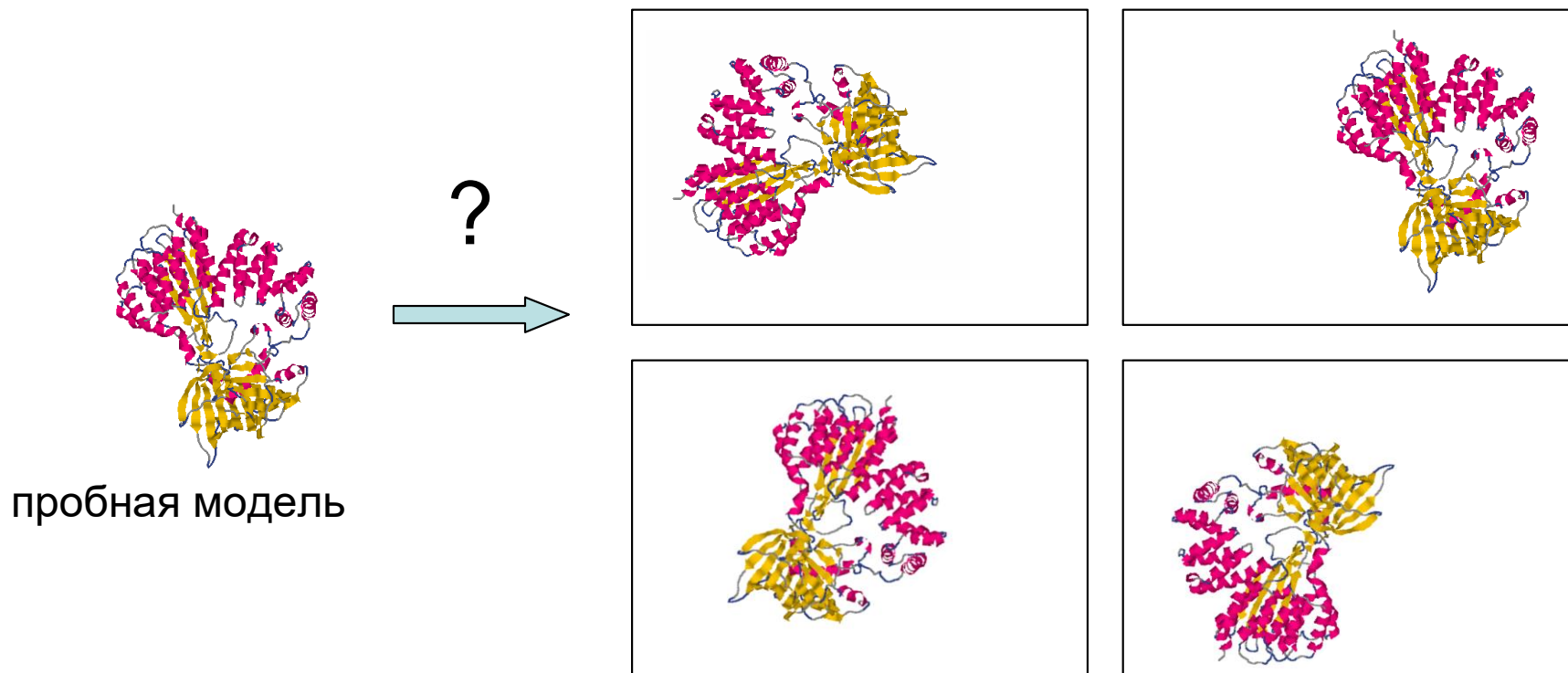
- белок – цепочка разнообразных аминокислотных остатков, свернутая в пространстве;
- последовательность типов аминокислот в цепочке определяет пространственную структуру белка;
- цепочки с похожей (гомологичной) последовательностью имеют похожую пространственную структуру.
- значения фаз, рассчитанные по модели гомологичного белка могут служить приближением к искомым значениям

Метод молекулярного замещения

2. Программа действий:

- найти в банке структур белок с похожей последовательностью и уже известной структурой;
- известную структуру использовать в качестве первого приближения к искомой структуре; рассчитать фазы; построить синтезы Фурье;
- внести корректировки в модель в соответствии с рассчитанными синтезами.

MR - Molecular Replacement Молекулярное замещение;



В общем случае положение твердого тела определяется шестью параметрами $(\alpha, \beta, \gamma, t_x, t_y, t_z)$.

↑↑↑
углы вращения

↑↑↑
вектор трансляции

MR - Molecular Replacement
Молекулярное замещение;

Оптимальное размещение пробной модели

Для каждого допустимого набора параметров $(\alpha, \beta, \gamma, t_x, t_y, t_z)$ можно рассчитать соответствующие значения модулей структурных факторов $F_{hkl}^{calc}(\alpha, \beta, \gamma, t_x, t_y, t_z)$

Хотим иметь наилучшее совпадение с экспериментально определенными модулями F_{hkl}^{obs}

$$\sum_{hkl} \left(F_{hkl}^{calc}(\alpha, \beta, \gamma, t_x, t_y, t_z) - F_{hkl}^{obs} \right)^2 \Rightarrow \min$$

Задачу можно разделить на два этапа: поиск углов вращения ("функция вращения") и поиск трансляции ("функция трансляции").

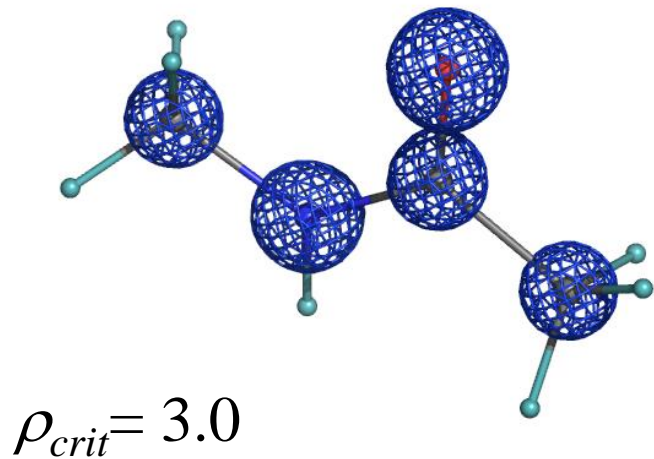
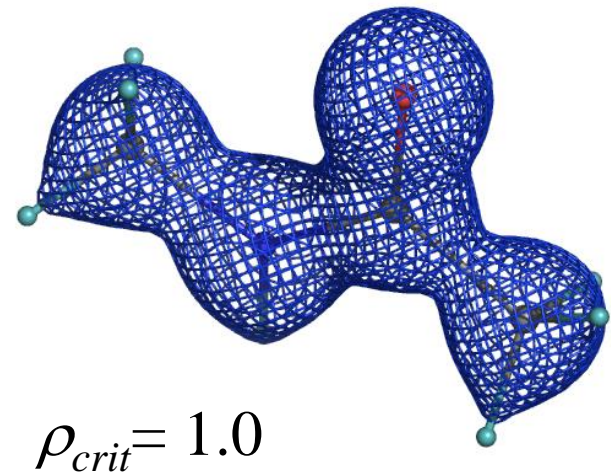
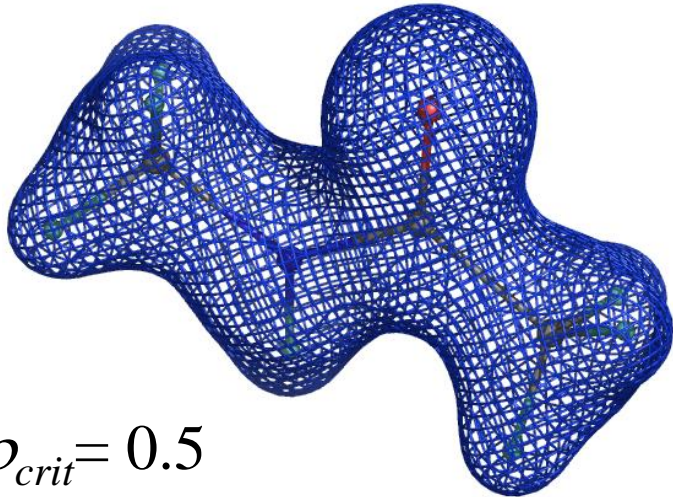
MR - Molecular Replacement
Молекулярное замещение;

С нахождением оптимальной ориентации и положения пробной модели работа не кончается!

Надо откорректировать модель.

Как "показать" распределение электронной плотности?

$\rho(x, y, z) = \rho_{crit}$ - поверхность



σ – шкала (σ -scale, z-score)

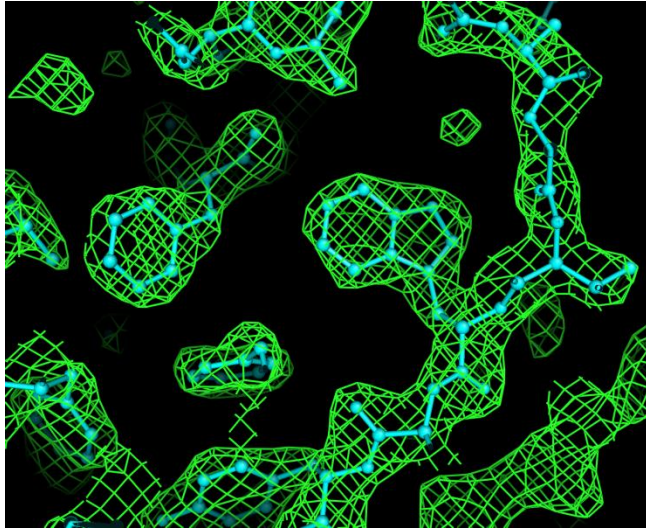
$$\langle \rho \rangle = \frac{1}{|V|} \int_V \rho(\mathbf{r}) dV_{\mathbf{r}} \quad - \text{средняя электронная плотность в объекте}$$

$$\sigma^2 = \frac{1}{|V|} \int_V (\rho(\mathbf{r}) - \langle \rho \rangle)^2 dV_{\mathbf{r}} \quad \sigma - \text{среднеквадратичное отклонение плотности (от среднего)}$$

$$z_{crit} = \frac{\rho_{crit} - \langle \rho \rangle}{\sigma} \quad - \text{уровень срезки в "сигмах"}$$

Комбинированные синтезы Фурье

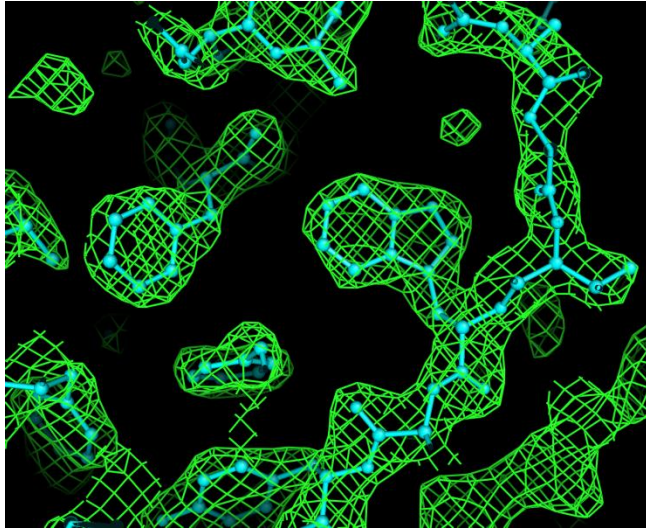
Комбинированные синтезы Фурье



$$F_{full}, \varphi_{full}$$

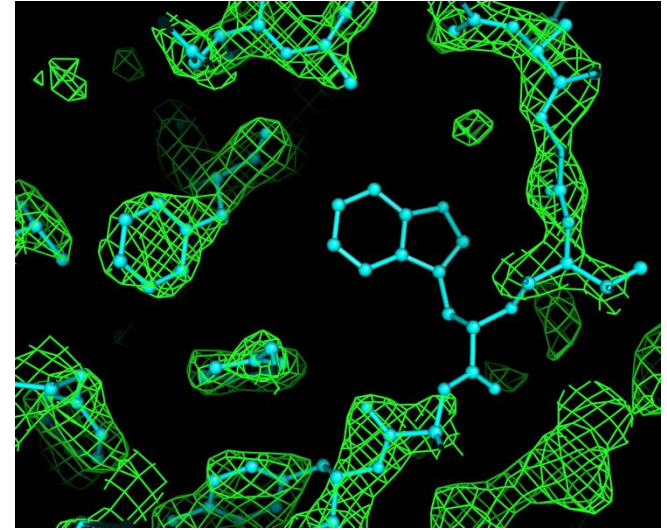
$$F_{full} = F_{obs}$$

Комбинированные синтезы Фурье



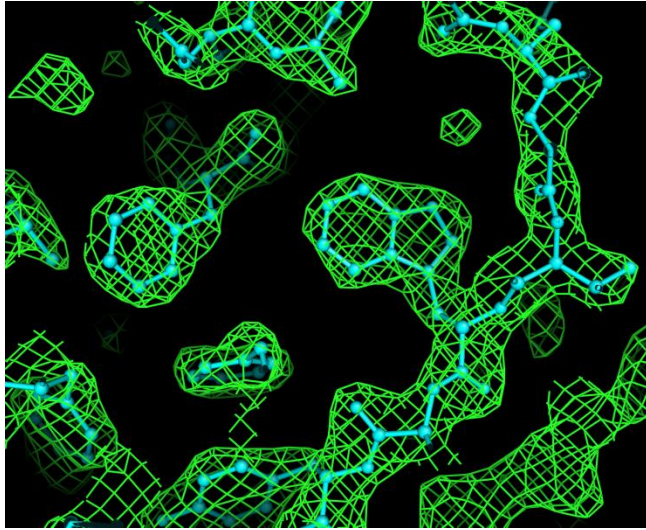
$$F_{full}, \varphi_{full}$$

$$F_{full} = F_{obs}$$



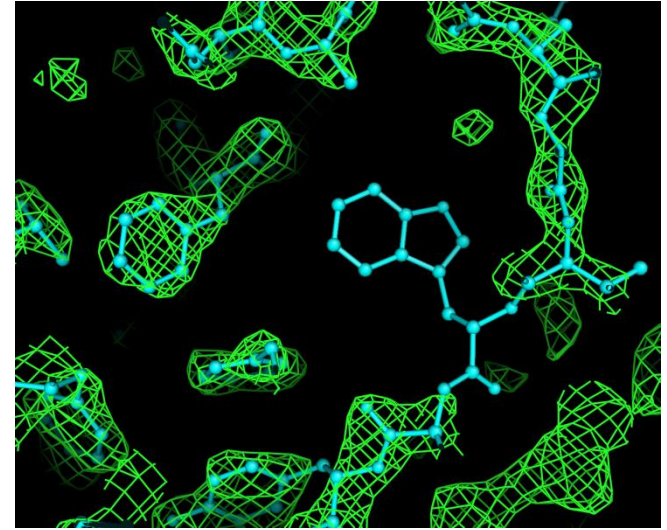
$$F_{part}, \varphi_{part}$$

Комбинированные синтезы Фурье



$$F_{full}, \varphi_{full}$$

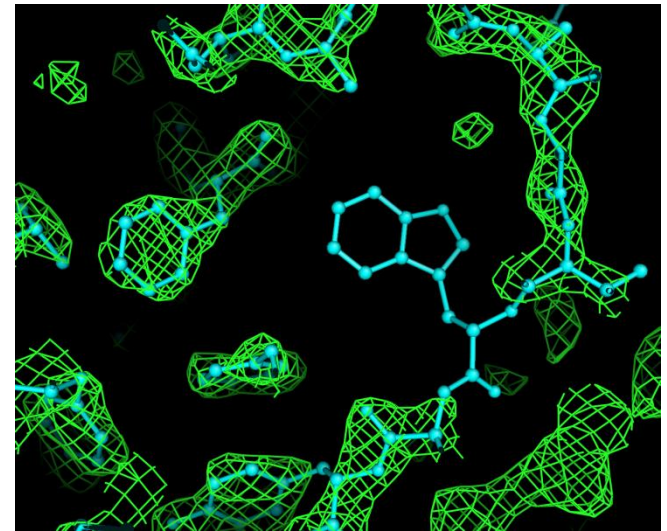
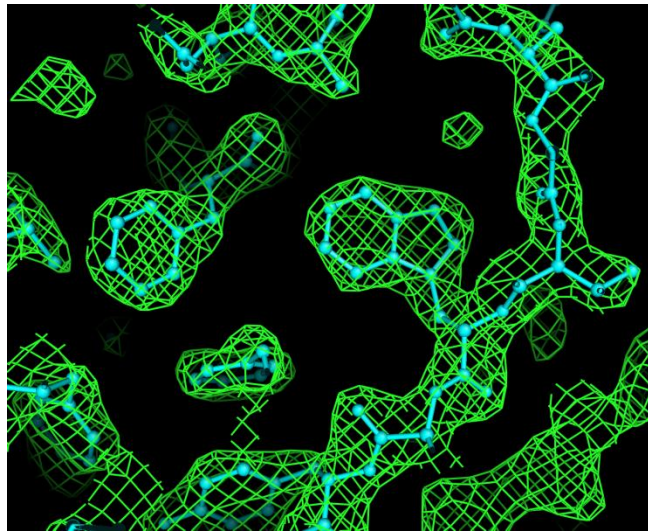
$$F_{full} = F_{obs}$$



$$F_{part}, \varphi_{part}$$

$$F_{full}, \varphi_{part}$$

Комбинированные синтезы Фурье

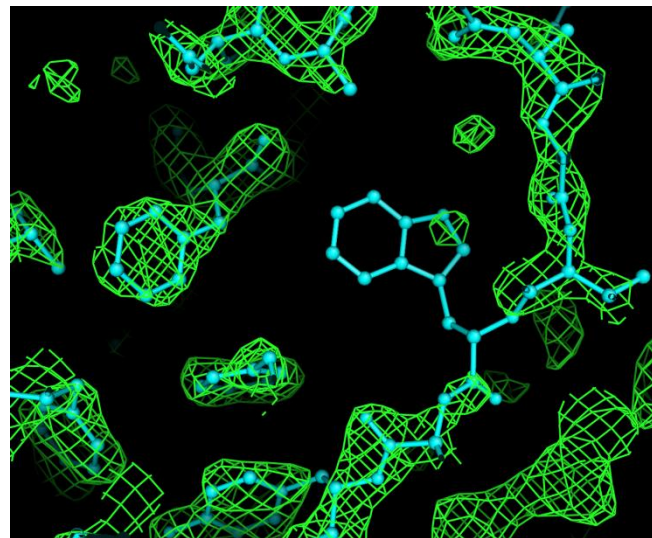


1.5σ

F_{full}, φ_{full}

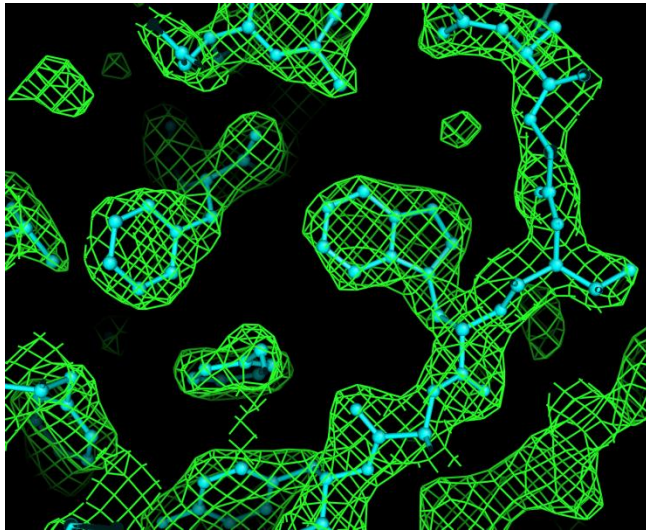
$F_{full} = F_{obs}$

F_{part}, φ_{part}

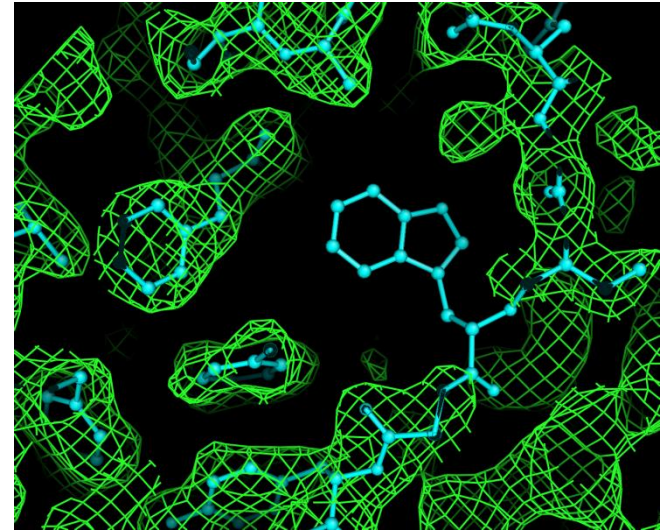


F_{full}, φ_{part}

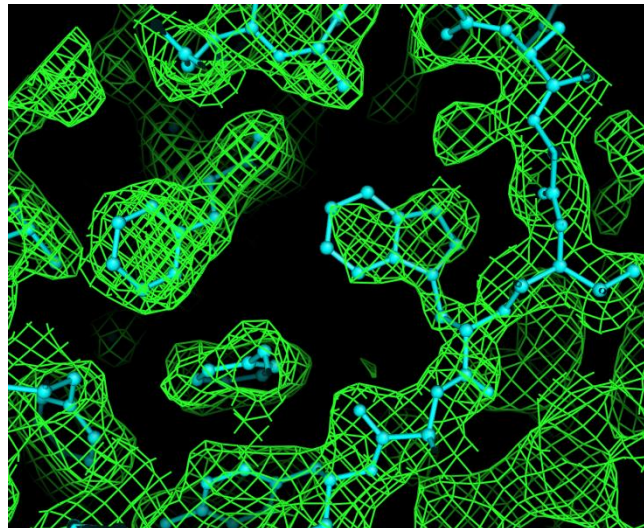
Комбинированные синтезы Фурье



F_{full}, φ_{full}
 1.5σ



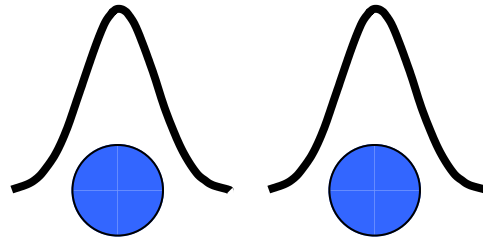
F_{part}, φ_{part}
 0.5σ



$F_{full}, \varphi_{part}, 0.5 \sigma$

«Комбинированные» синтезы Фурье. Корректировка атомной модели.

«точная»
структура



F^{obs}

φ^{true}

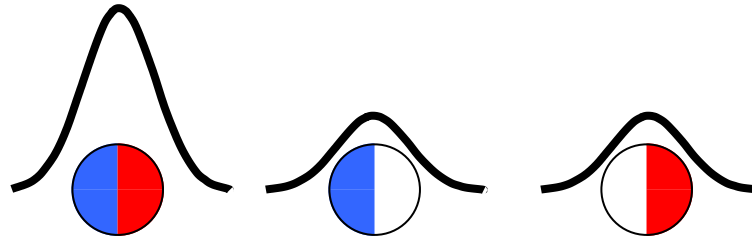
модель



F^{calc}

φ^{calc}

комбинированный
синтез Фурье



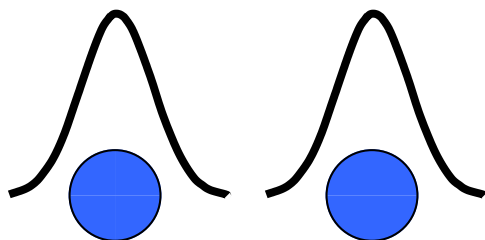
F^{obs}

φ^{calc}

- можем добавлять в модель «потерянные» атомы
- можем удалять из модели «неправильные» атомы

«Комбинированные» синтезы Фурье.

«точная»
структура



F^{obs}

φ^{true}

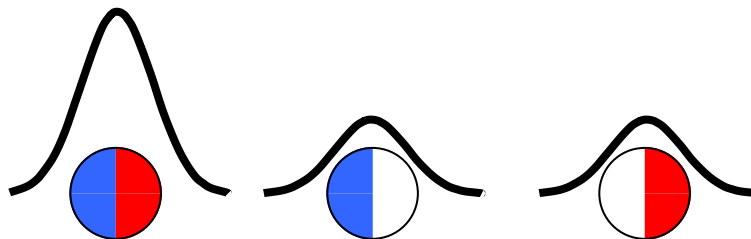
модель



F^{calc}

φ^{calc}

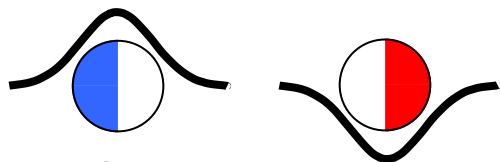
комбинированный
синтез Фурье



F^{obs}

φ^{calc}

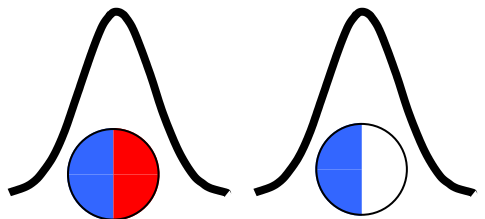
разностный
синтез Фурье



$F^{obs} - F^{calc}$

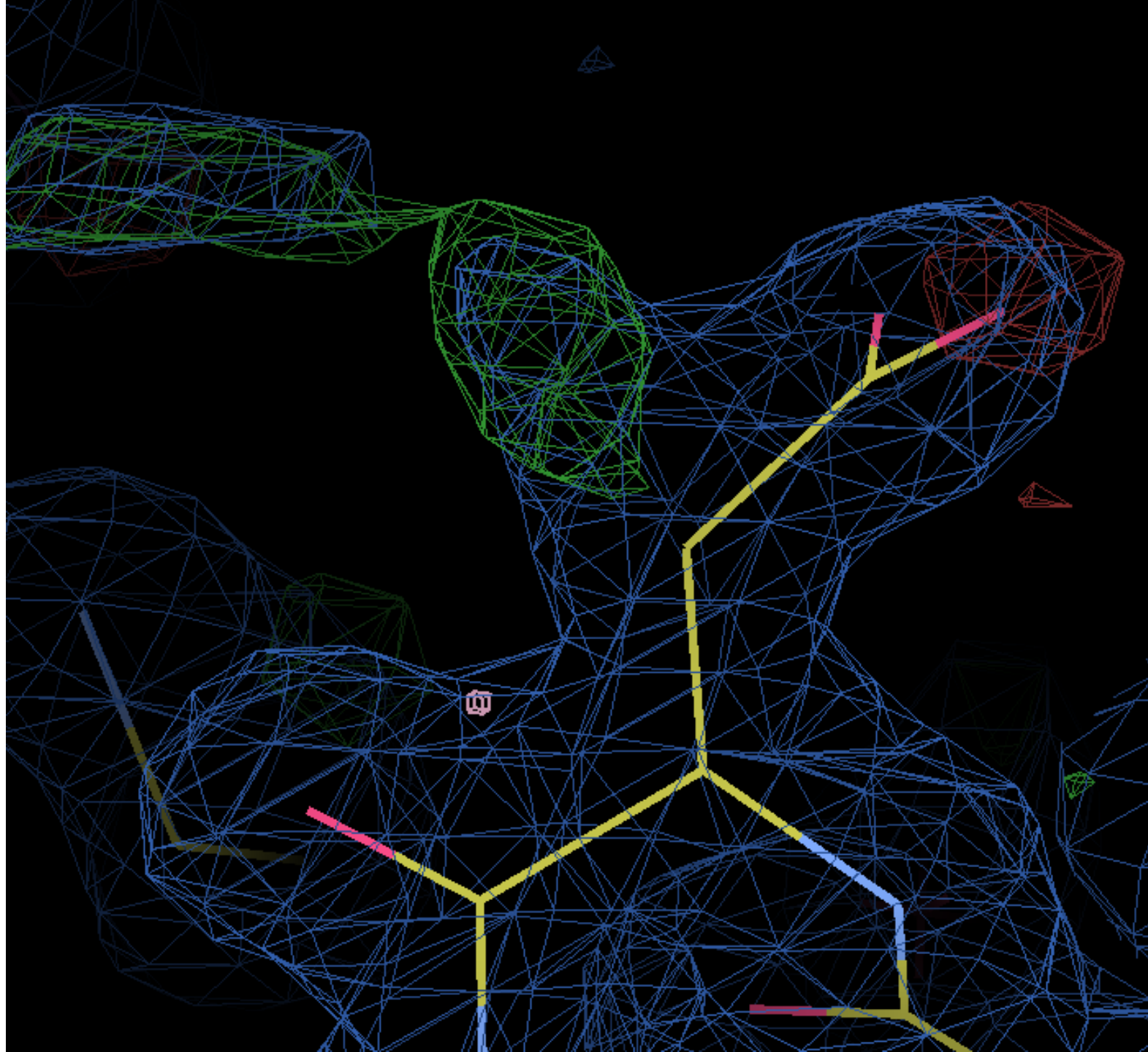
φ^{calc}

«2-1»
 $2F^o - F^c$
синтез Фурье



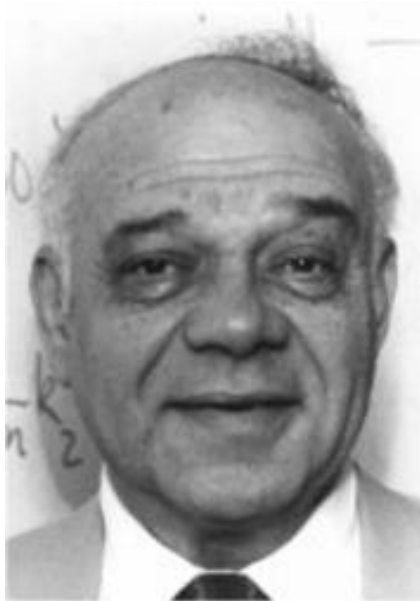
$2F^{obs} - F^{calc}$

φ^{calc}



Jerome Karle Gerbert Hauptman

$\mathcal{F}^{\text{Karle}},$
 φ^{Karle}



$\mathcal{F}^{\text{Hauptman}},$
 $\varphi^{\text{Hauptman}}$

**Что важнее для
изображения:
модули или фазы?**

$\mathcal{F}^{\text{Karle}},$
 $\varphi^{\text{Hauptman}}$



$\mathcal{F}^{\text{Hauptman}},$
 φ^{Karle}

Взвешенные синтезы Фурье

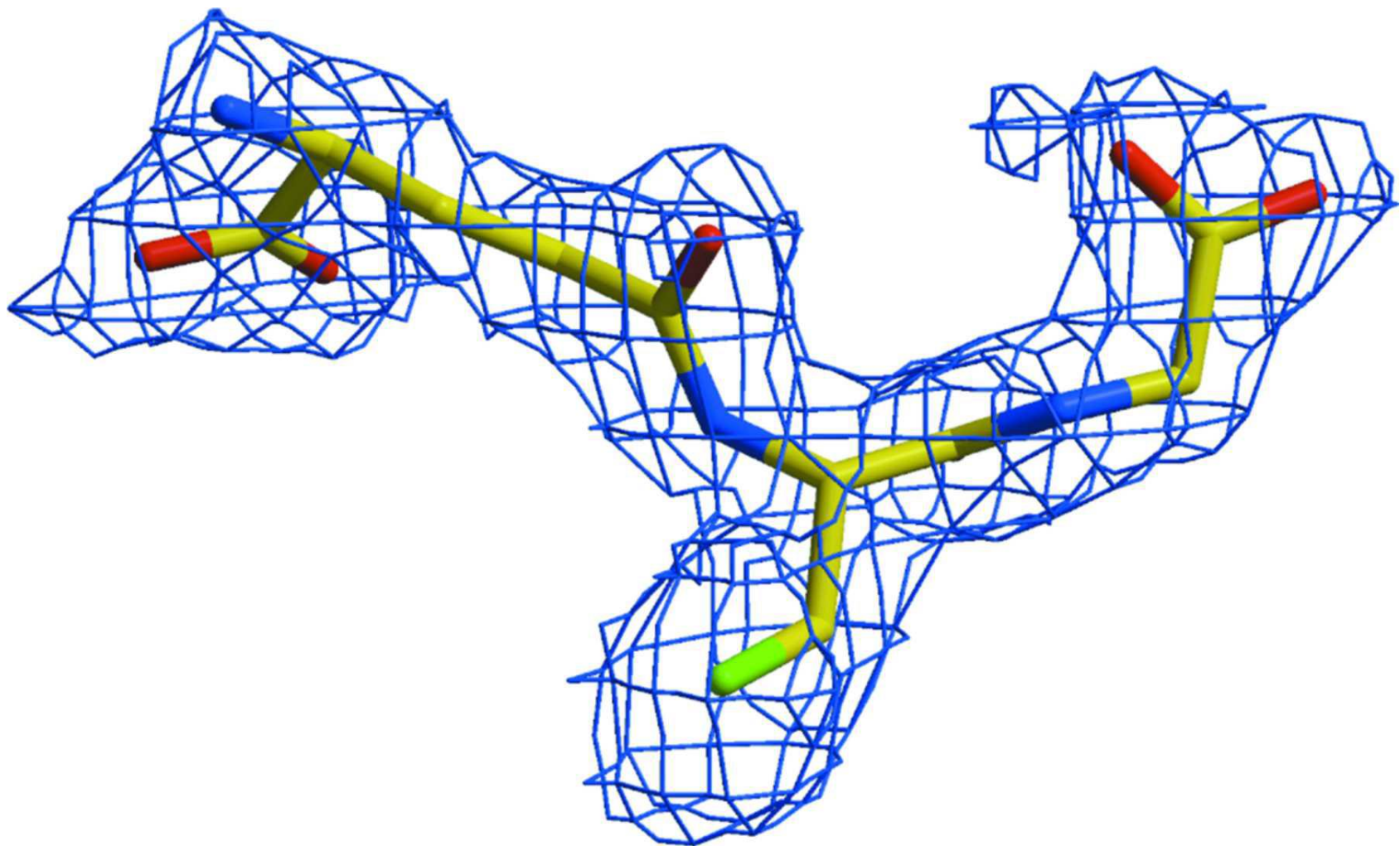
$F_{hkl}^{obs}, \varphi_{hkl}^{calc}$ - синтез Фурье

$w_{hkl} F_{hkl}^{obs}, \varphi_{hkl}^{calc}$ - взвешенный синтез Фурье (чем надежнее определена фаза, тем больше вес w_{hkl})

$2F_{hkl}^{obs} - F^{calc}, \varphi_{hkl}^{calc}$ - комбинированный синтез Фурье

$2mF_{hkl}^{obs} - DF^{calc}, \varphi_{hkl}^{calc}$ - взвешенный комбинированный синтез Фурье
(σ_A -weighted $2F^o - F^c$ synthesis)

Final σ_A -weighted $2|F_o| - |F_c|$ electron density (Read, 1986) for the GSH molecule bound to subunit A of DmGSTe6 contoured at 1σ .



MR - Molecular Replacement
Молекулярное замещение;

- 😊 • один из наиболее используемых методов решения фазовой проблемы;
- 😊 • чем больше белков исследовано, тем шире возможности метода;
- 😞 • необходимо наличие гомологичной пробной модели;
- 😞 • результат зависит от обоснованности гипотезы о гомологии между пробной моделью и исследуемой структурой.

Phaser (R.Read *et al.*)

MolRep, BALBES (A.Vagin *et al.*)

AMoRe (J.Navaza *et al.*)

MR - Molecular Replacement
Молекулярное замещение;

Где взять гомологичную пробную модель?

- среди структур, определенных РСА;
- среди структур, определенных ЯМР;
- электронная микроскопия;
- теоретическое предсказание модели.

Rosetta (D.Baker *et al.*)

Gilski *et al* (2011).Acta Cryst D67, 907-914.

Synopsis: The crystal structure of Mason-Pfizer Monkey Virus protease in monomeric fold has been solved by Molecular Replacement using a model generated by players of the online game Foldit.

Метод изоморфного замещения

"нативный" белок

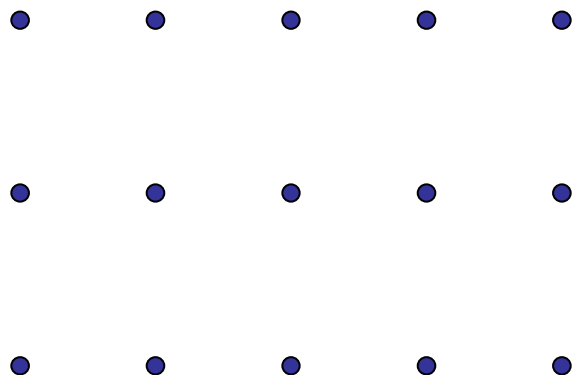


$\{F^P\}$

изоморфное производное



подструктура тяжелых атомов



$\{F^{PH}\}$

$$F_{hkl}^H \approx \left| F_{hkl}^{PH} - F_{hkl}^P \right|$$

Шаг 1. Определение подструктуры «тяжелых атомов»

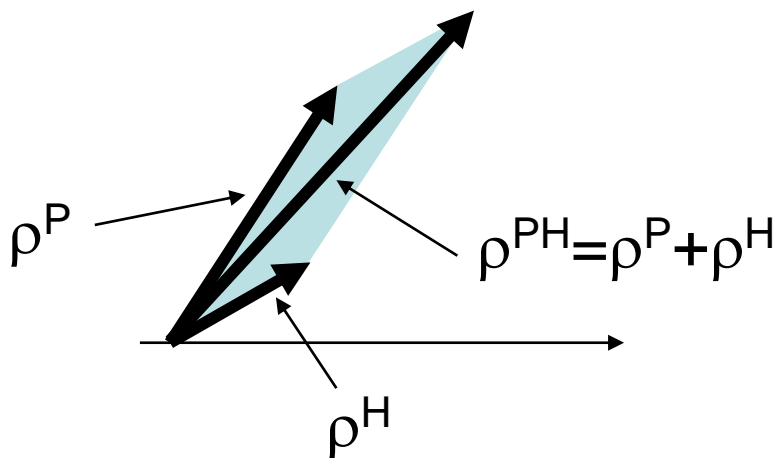
- небольшое число атомов: $\{\mathbf{r}_j^H\}, j = 1, \dots, m$

$$F_{hkl}^H \approx |F_{hkl}^{PH} - F_{hkl}^P|$$

- расчет модулей и фаз $F_{hkl}^H, \varphi_{hkl}^H$

Шаг 2. Расчет значения фазы для каждого коэффициента Фурье

$$(F^{PH})^2 = (F^P)^2 + (F^H)^2 + 2F^P F^H \cos(\varphi^P - \varphi^H)$$



F^P, F^{PH} - получили из эксперимента

F^H, φ^H - определили на Шаге 1

φ^P - находим из уравнения; для каждого коэффициента

нативный белок F_P, φ_P

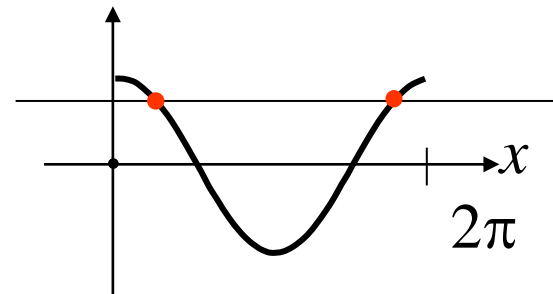
изоморфное производное F_{PH}, φ_{PH}




подструктура тяжелых атомов F_H, φ_H

уравнение $F_P^2 + 2F_P F_H \cos(\varphi_P - \varphi_H) + F_H^2 = F_{PH}^2$

$$\cos(\varphi^P - \varphi^H) = \frac{(F^{PH})^2 - (F^P)^2 - (F^H)^2}{2F^P F^H}$$

решение не единственно
(проблема снимается при
наличии двух производных)



-  метод позволяет решать фазовую проблему;
-  наличие каналов в кристаллах белка позволяет получать производные;
-  изоморфизм имеет место лишь приближенно;
-  проблема определения мест присоединения тяжелых меток;
-  проблема создания тяжелых меток для больших макромолекулярных комплексов
-  фазы определяются неоднозначно; необходимо несколько производных.

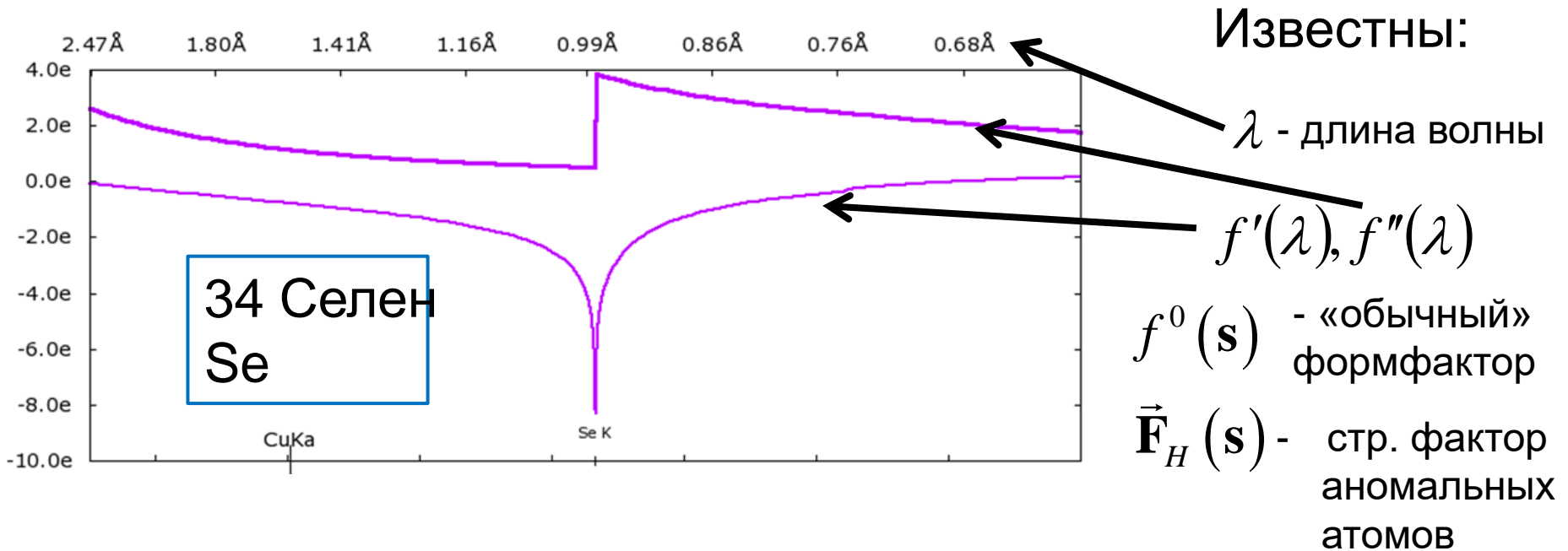
MIR - Multiple Isomorphous Replacement
Множественное изоморфное замещение;

AD - Anomalous Dispersion / Аномальное рассеяние

Кинематическая теория рассеяния: $I^{obs}(\mathbf{s}) = |\vec{\mathbf{F}}_P(\mathbf{s})|^2$

Наличие аномальных рассеивающих атомов:

$$I^{obs}(\mathbf{s}; \lambda) = \left| \vec{\mathbf{F}}^P(\mathbf{s}) + \frac{f'(\lambda)}{f^0(\mathbf{s})} \vec{\mathbf{F}}^H(\mathbf{s}) + \frac{f''(\lambda)}{f^0(\mathbf{s})} \vec{\mathbf{F}}^H(\mathbf{s}) \right|_{rot\ 90^\circ}^2$$



$F_P(\mathbf{s}), \varphi_P(\mathbf{s})$ - НЕИЗВЕСТНЫ

AD - Anomalous Dispersion Аномальное рассеяние



- метод позволяет решать фазовую проблему;



- в белках встречаются аномально рассеивающие атомы; можно использовать производные с тяжелыми атомами;



- проблема определения мест нахождения аномально рассеивающих атомов;



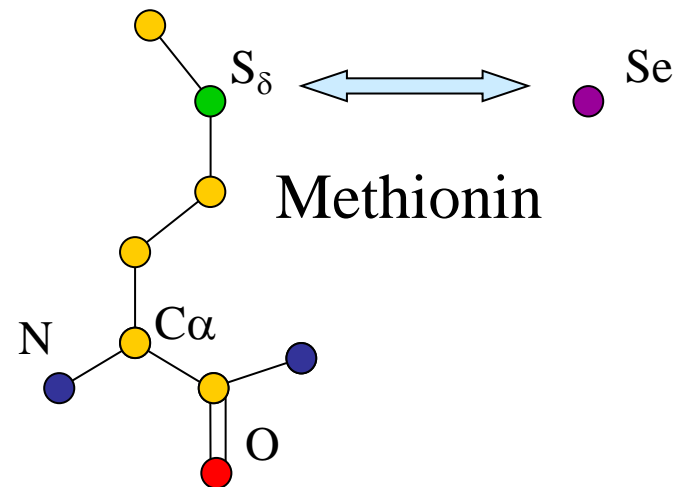
- фазы определяются неоднозначно;

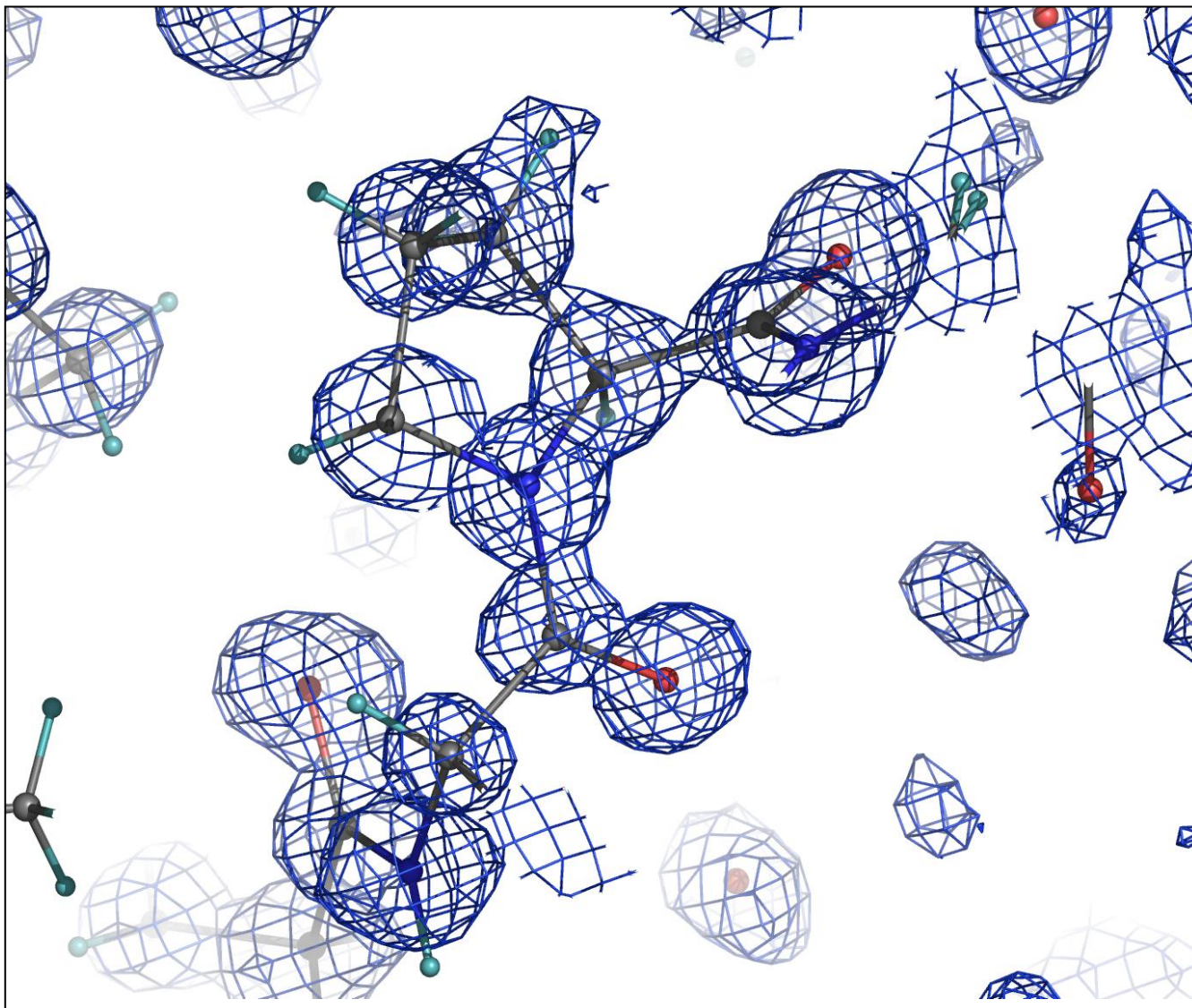
MAD - Multiwavelength Anomalous Diffraction

Многоволновое аномальное рассеяние;

В основе подхода лежит изменение интенсивности аномального рассеяния при изменении длины волны.

- ☺ • один из активно используемых методов решения фазовой проблемы;
- ☺ • возможность использования селен-метиониновых производных;
- ☹ • проблема определения мест нахождения аномально рассеивающих атомов;
- ☹ • не всегда получается.





Синтез Фурье электронной плотности для белка альдоз-редуктазы. Разрешение 0.9Å. Фазы определены методом MAD.