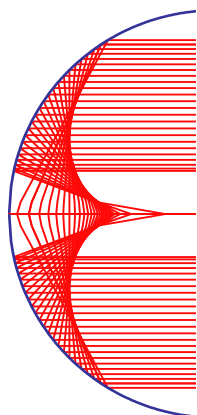


Решение обратной задачи светорассеяния для определения морфологических характеристик биологических частиц с помощью сканирующего проточного цитометра

Максим А. Юркин



Институт Химической Кинетики и Горения СО РАН,
Лаборатория Цитометрии и Биокинетики
<http://cyto.kinetics.nsc.ru/>
тел: +7(383)3333240
yurkin@gmail.com

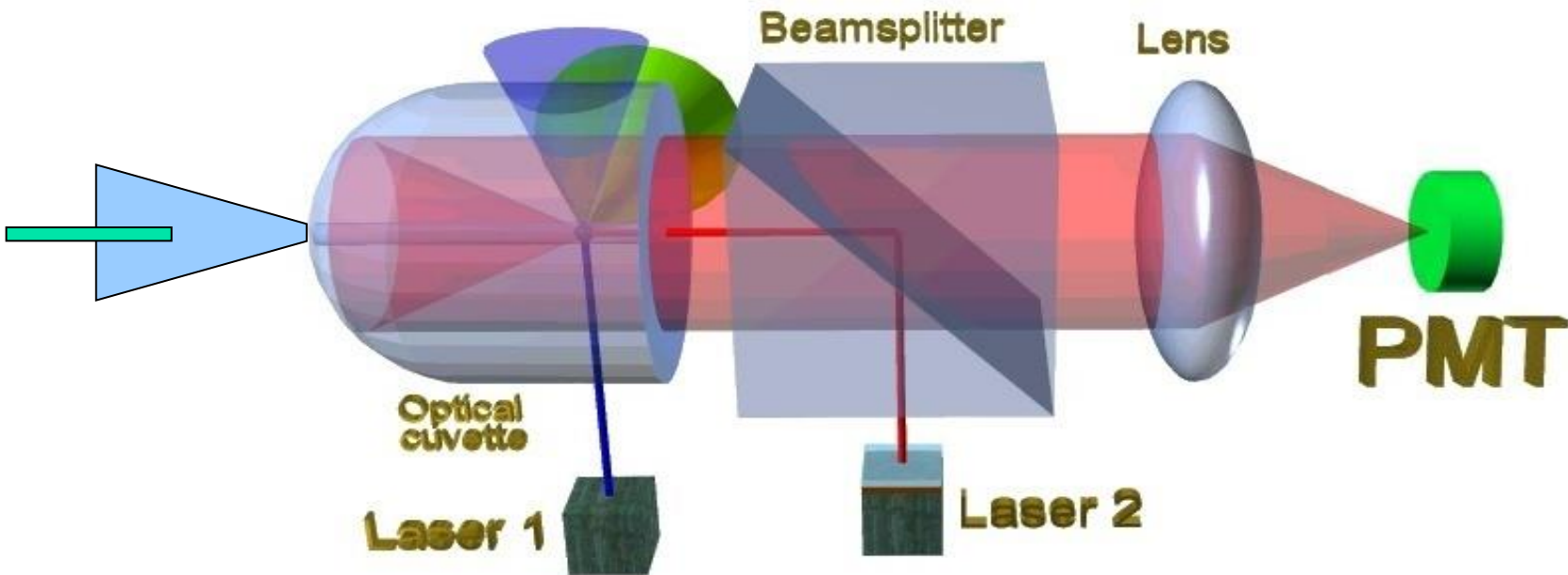


План

- Экспериментальная техника
- Практические объекты
- Глобальная оптимизация
- Особенности наших обратных задач
- Множество стартовых точек
- Интерполяция по базе данных
- Адаптивное построение базы данных

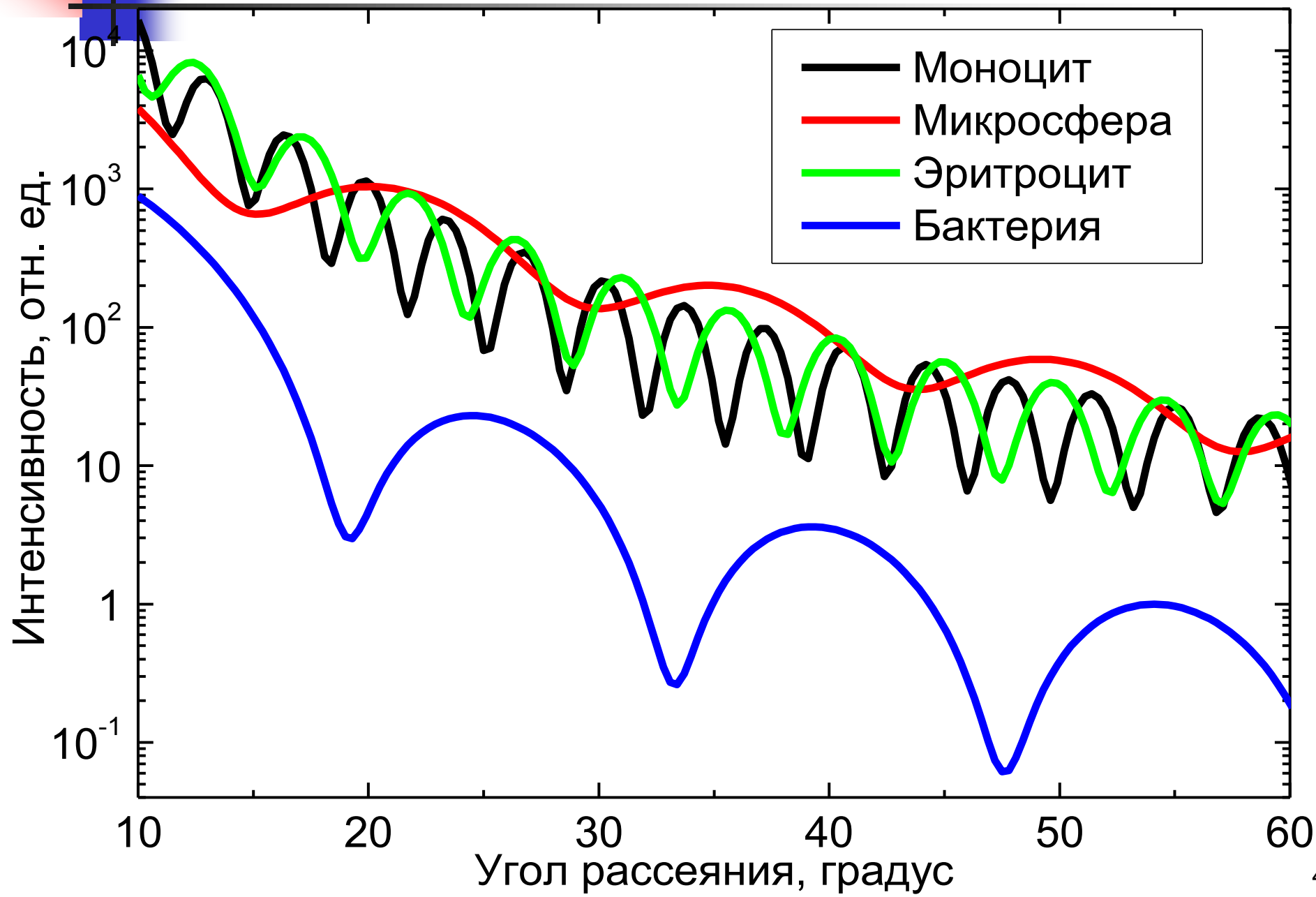
Сканирующий проточный цитометр

- Измеряет индикатрису светорассеяния и сигнал флуоресценции одиночных частиц.



$$I(\theta) = \frac{w(\theta)}{2\pi} \int_0^{2\pi} d\varphi (S_{11}(\theta, \varphi) + S_{14}(\theta, \varphi)) \quad w(\theta) = \frac{1^\circ}{\theta} \exp\left[-2 \ln^2(\theta / 54^\circ)\right]$$

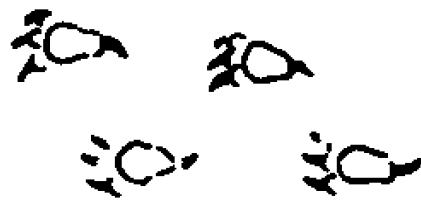
Индикатриса светорассеяния



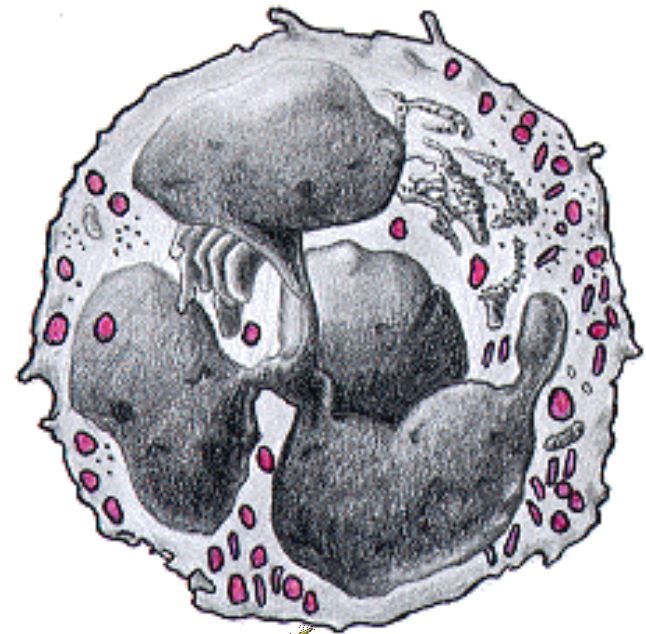
Обратная задача



Дракон



Следы



Прям



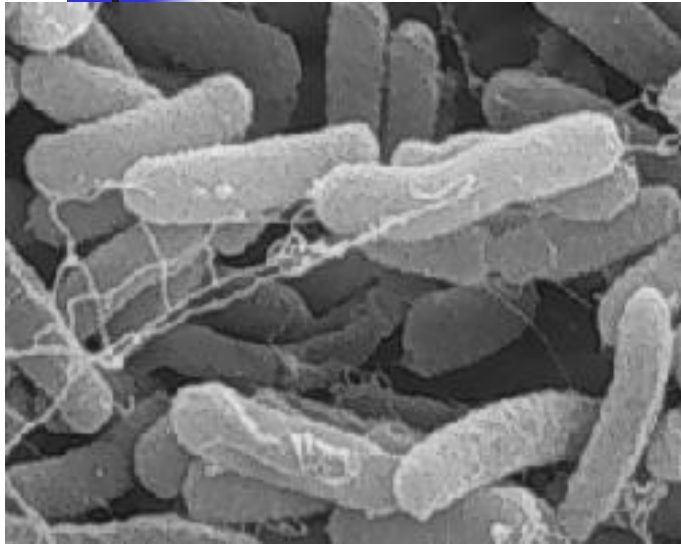
Обратная Дракон



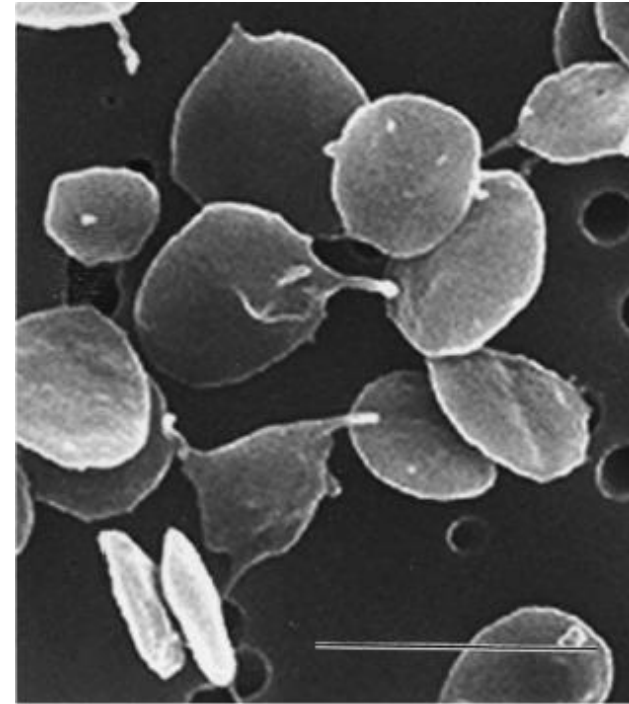
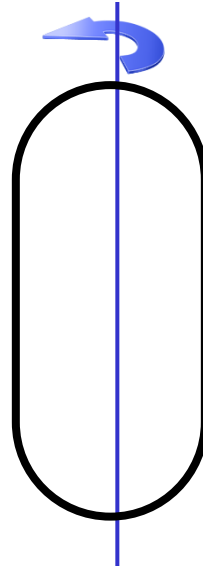
Компромисс

- Используем приближённые морфологические модели, описываемые всего несколькими параметрами.
- Задача – измерить эти параметры и оценить погрешности полученных значений для каждой частицы.

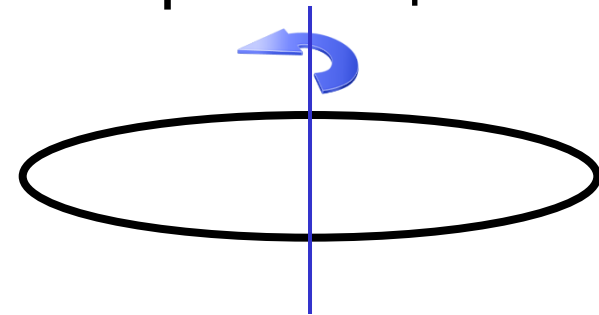
Типичные несферические объекты



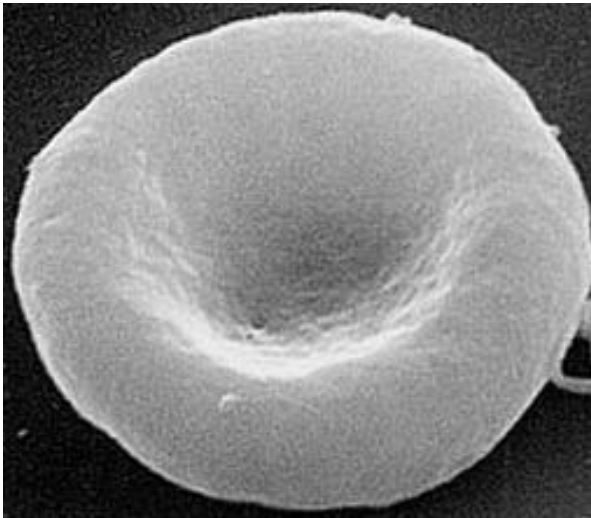
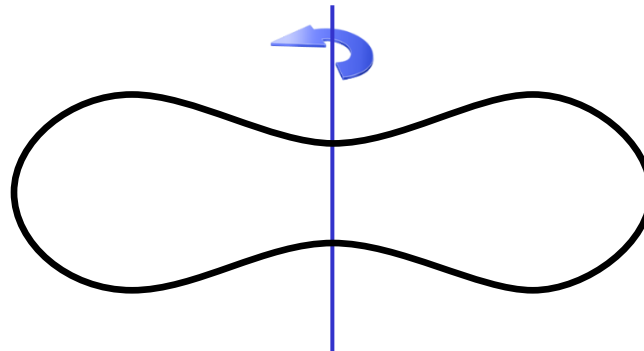
E. Coli



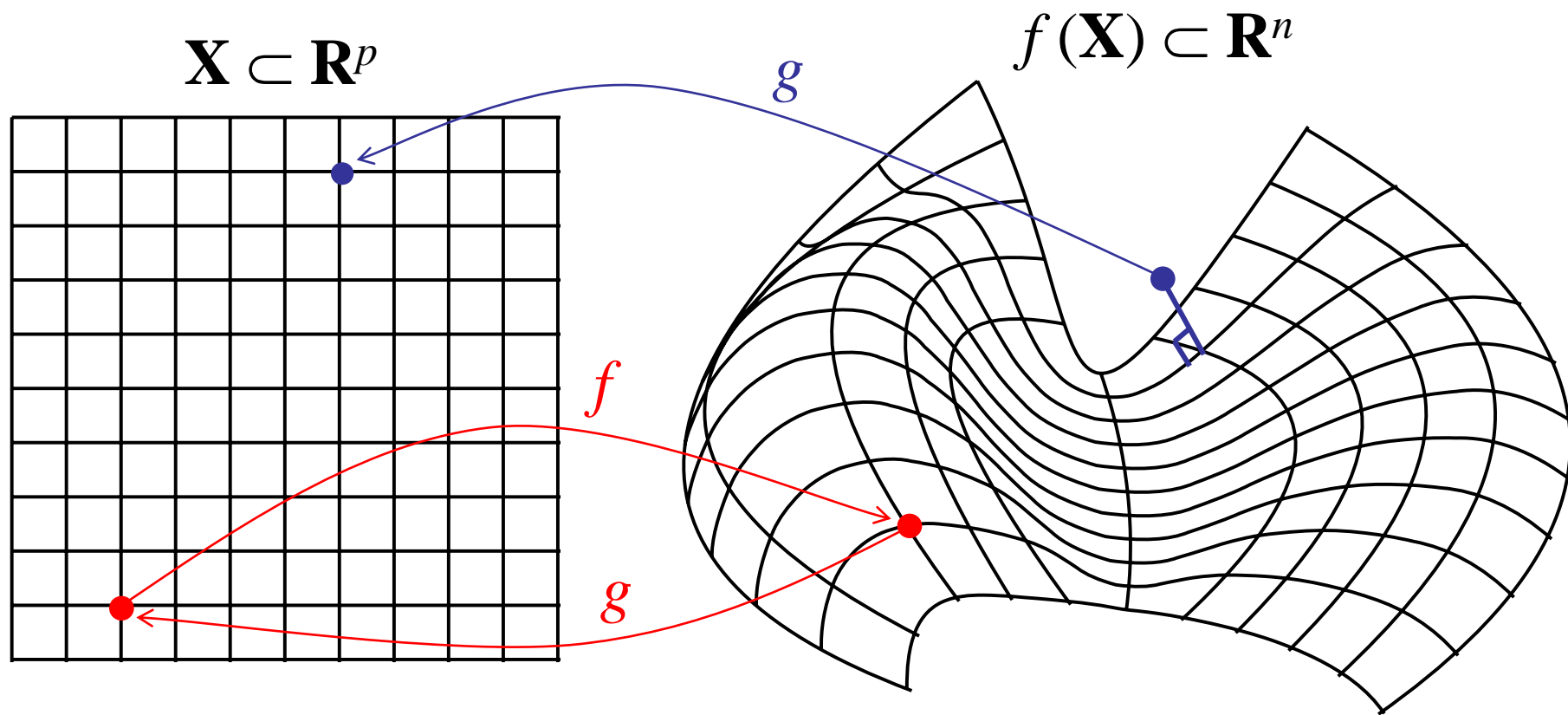
тромбоцит



Эритроцит



Обратная задача и регрессия



- Обратное отображение $g(y) = f^{-1}(y)$, $y \in f(X)$
- Можно доопределить для всех $y \in \mathbf{R}^n$ (метод наименьших квадратов) $g(y) = \operatorname{argmin}_{x \in X} \|y - f(x)\|$

Математическая постановка

- Задано прямое отображение $f: \mathbf{X} \rightarrow \mathbf{R}^d$, $\mathbf{X} \subset \mathbf{R}^p$, $d \geq p$
- Найти обратное отображение $f^{-1}: f(\mathbf{X}) \rightarrow \mathbf{X}$
- Неитерационные методы («аналитические») существуют только для шаров и не могут быть обобщены на другие формы

Ludlow & Everitt, *J. Optical Soc. Am. A* **17**, 2229–2235 (2000).

Semyanov et al., *Appl. Opt.* **43**, 5110–5115 (2004).

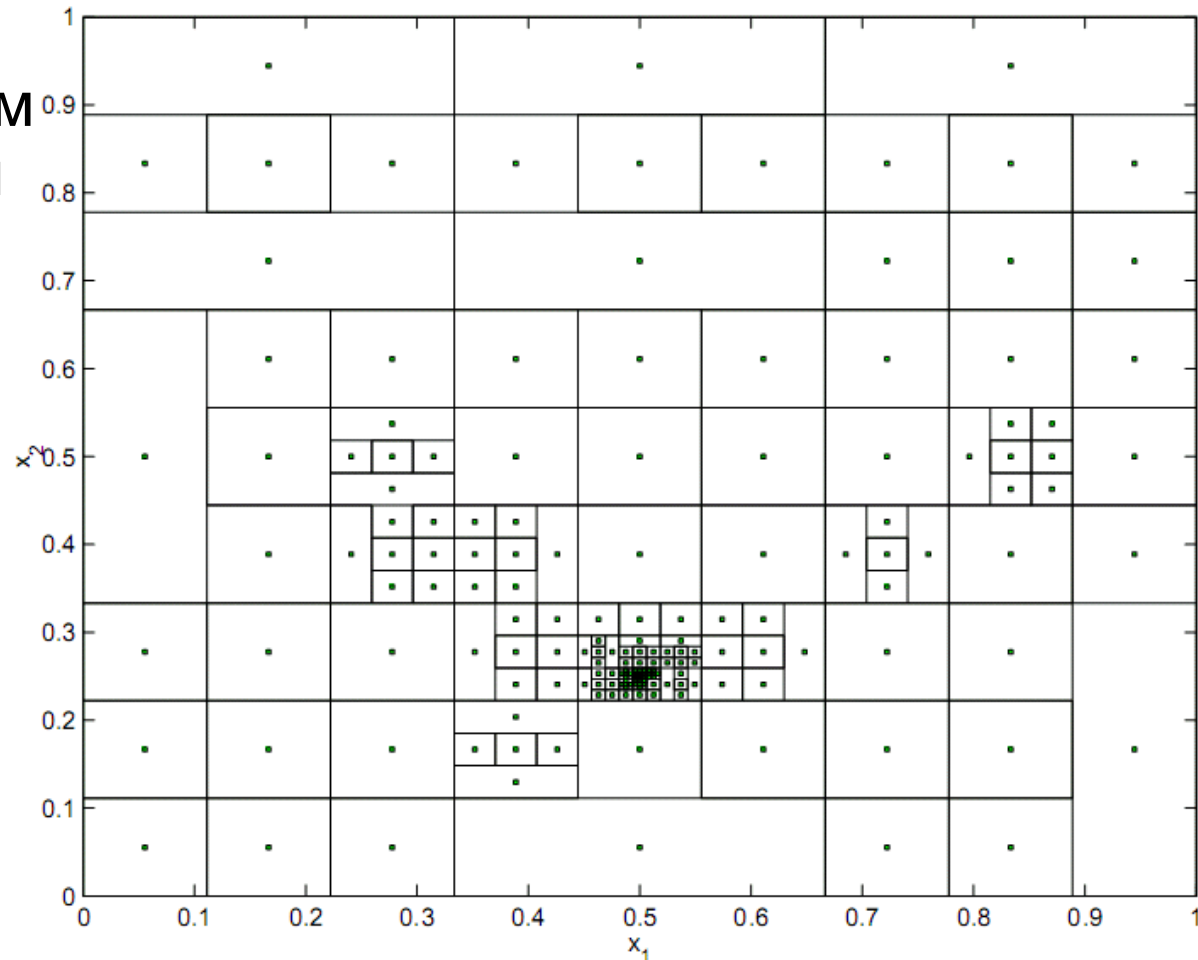
- Общий подход – оптимизация:

$$g: \mathbf{R}^d \rightarrow \mathbf{X}, \quad g(y) = \arg \min_{x \in \mathbf{X}} \underbrace{\|y - f(x)\|}_{S_y(x)}$$

- Минимизируемая функция $S_y(x)$ имеет много минимумов
 \Rightarrow требуется глобальная оптимизация

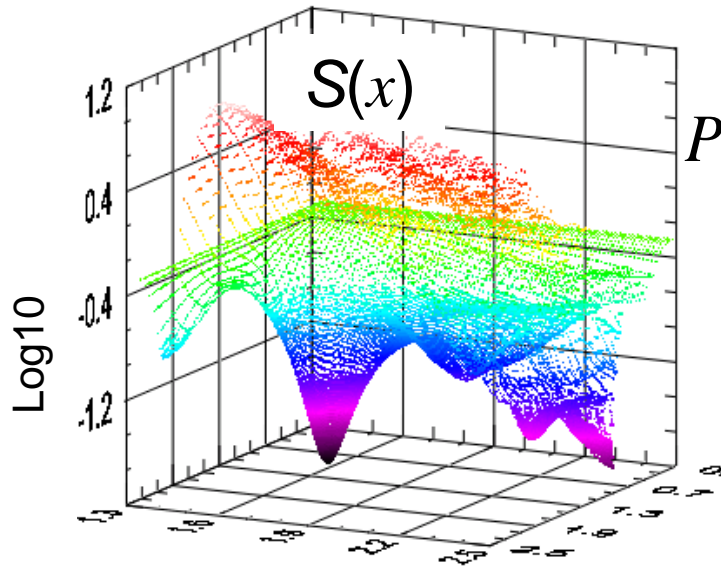
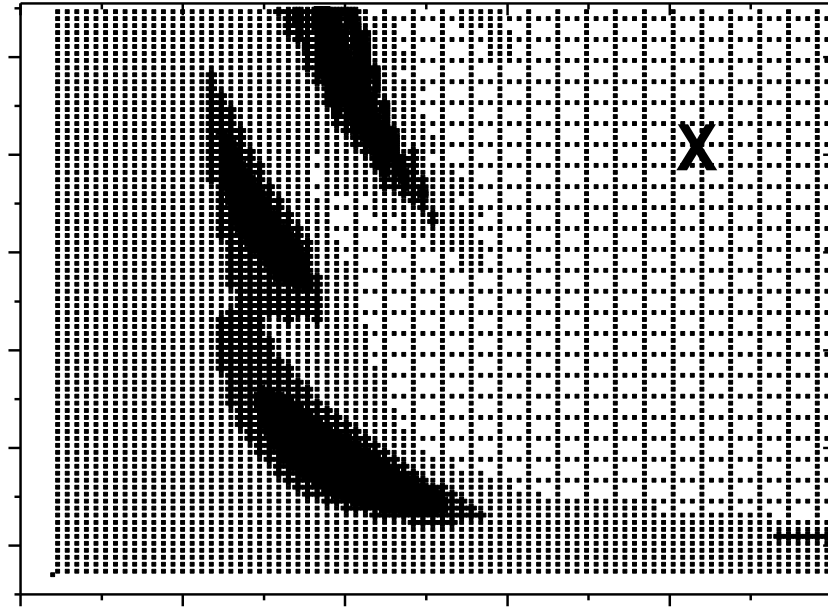
Метод глобальной оптимизации DiRect

- Рекурсивно делит пространство параметров, начиная с некоторого гиперкуба.
- Новое разделение выбирается с учётом значения функции и размера текущего разбиения.
- Очень много вычислений, но прописывает всю поверхность целевой функции (глобальная оптимизация).

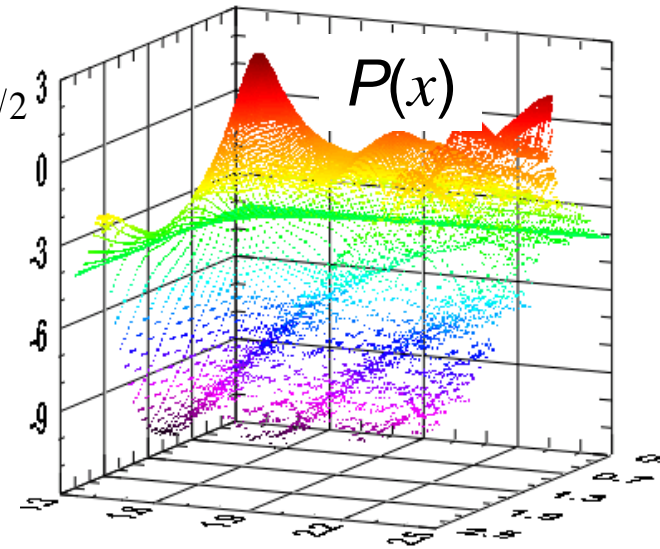
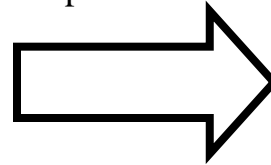


Погрешности определения параметров

- Минимизация $S(x)$
- Байесовский подход

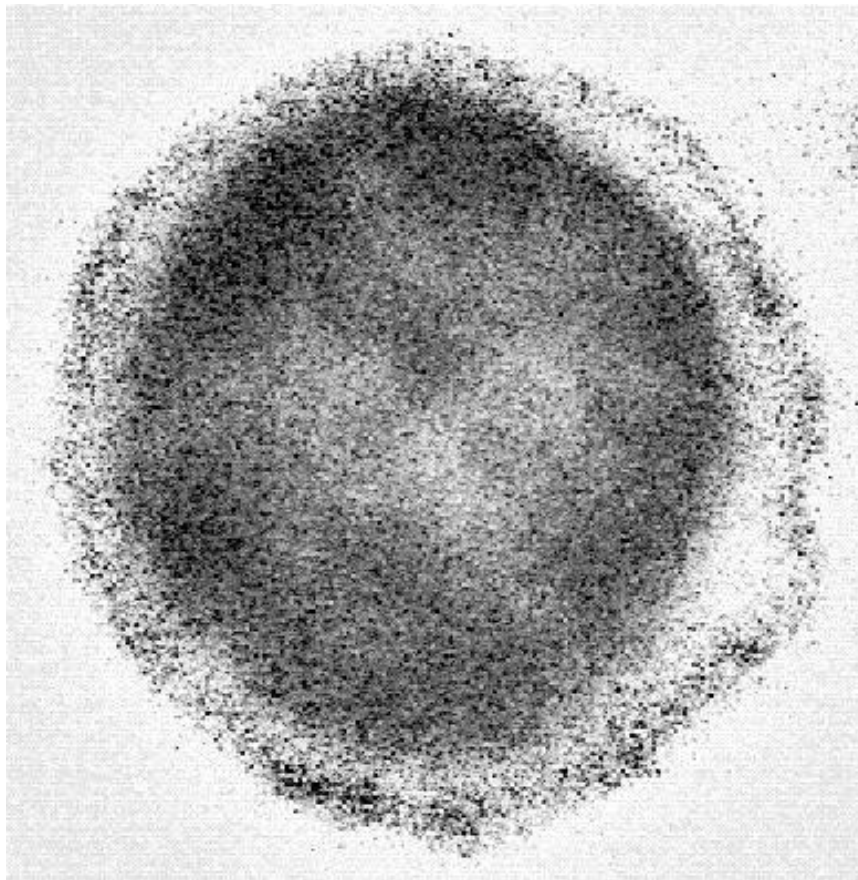


$$P(x|y_{\text{exp}}) = \kappa [S(x)]^{-n/2}$$

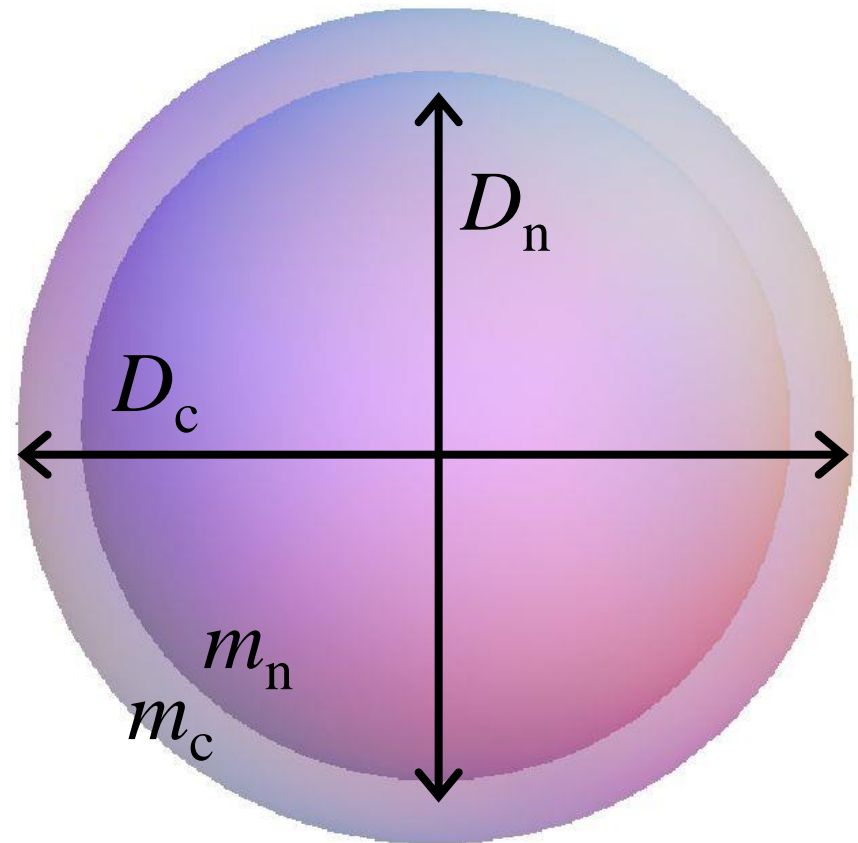


Пример – лимфоциты

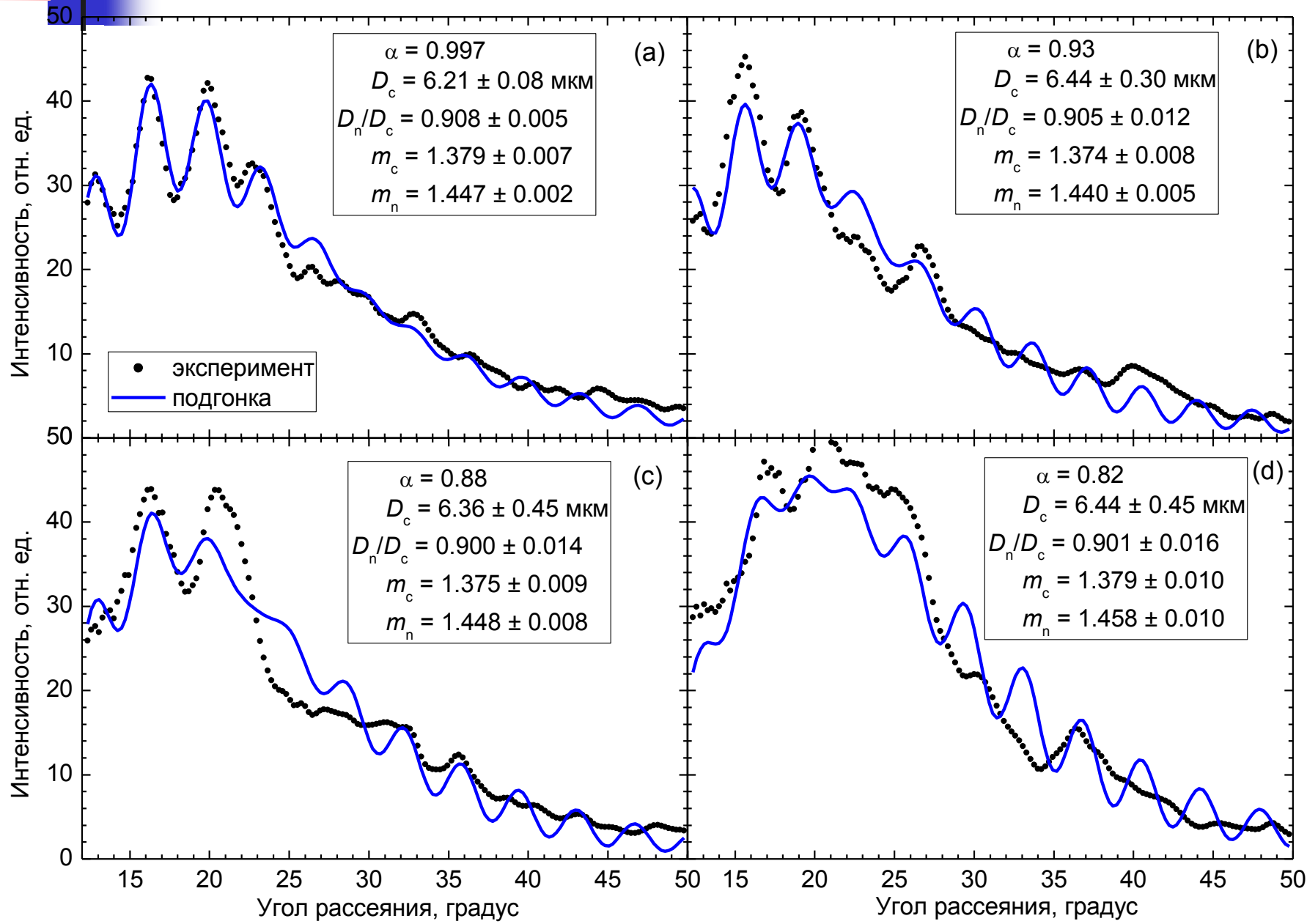
Изображение лимфоцита с
конфокального микроскопа



Модель:
двухслойный шар



Типичные результаты Т-лимфоцитов



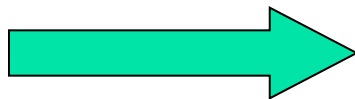


Особенности

- Одна модель объекта
- Тысячи клеток в пробе
- Типичные времена:
 - Измерение одной клетки ~ 1 мс
 - Решение прямой задачи ~ 1 мс – 1 мин
 - Прямая подгонка (регрессия) ~ 1 с – 1 сутки
- Обратная задача – обращение конечномерного отображения.
- Исследование обратного отображения – один раз много вычислений.

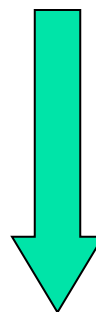
Подход к «медленным» задачам

Модель объекта
+ диапазоны для
параметров



База данных
теорет. сигналов

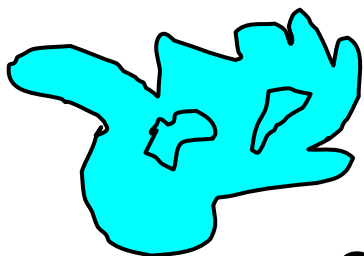
↑
Решение
прямой
задачи



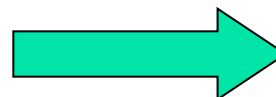
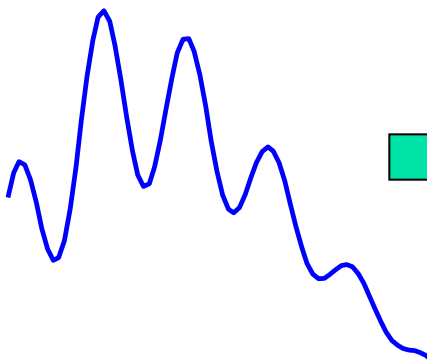
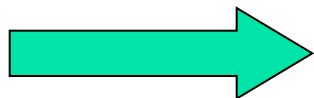
Метод
обращения



Параметры
объекта



↑
Эксперимент.
установка



Анализ отображения $f(x)$

- **Непрерывный.** Набор стартовых точек $\{z_i\}$
 - Для любого y определяем $z_0 = \operatorname{argmin}_{z \in \{z_i\}} \|y - z\|$
 - Локальная (градиентная) оптимизация, начиная с z_0 , сходится к глобальному минимуму – $g(y)$
 - Обычно, требуется только несколько вычислений $f(x)$

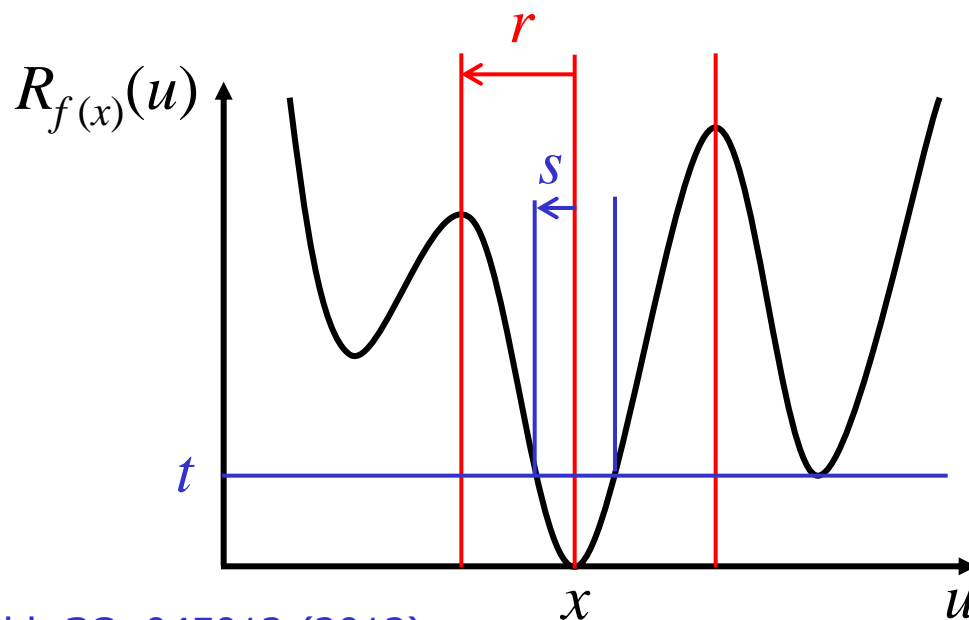
Непрерывный метод

- Предполагаем, что реально провести “достаточно плотное” исследование прямого отображения $f(x)$

$$r(x) = \sup \left\{ r > 0 \mid B(x, r) \subset X, \forall u \in B(x, r) \langle \nabla_u S_{f(x)}(u), u - x \rangle > 0 \right\}$$

$$t(x) = \inf_{u \in X \setminus B(x, r(x))} S_{f(x)}(u)$$

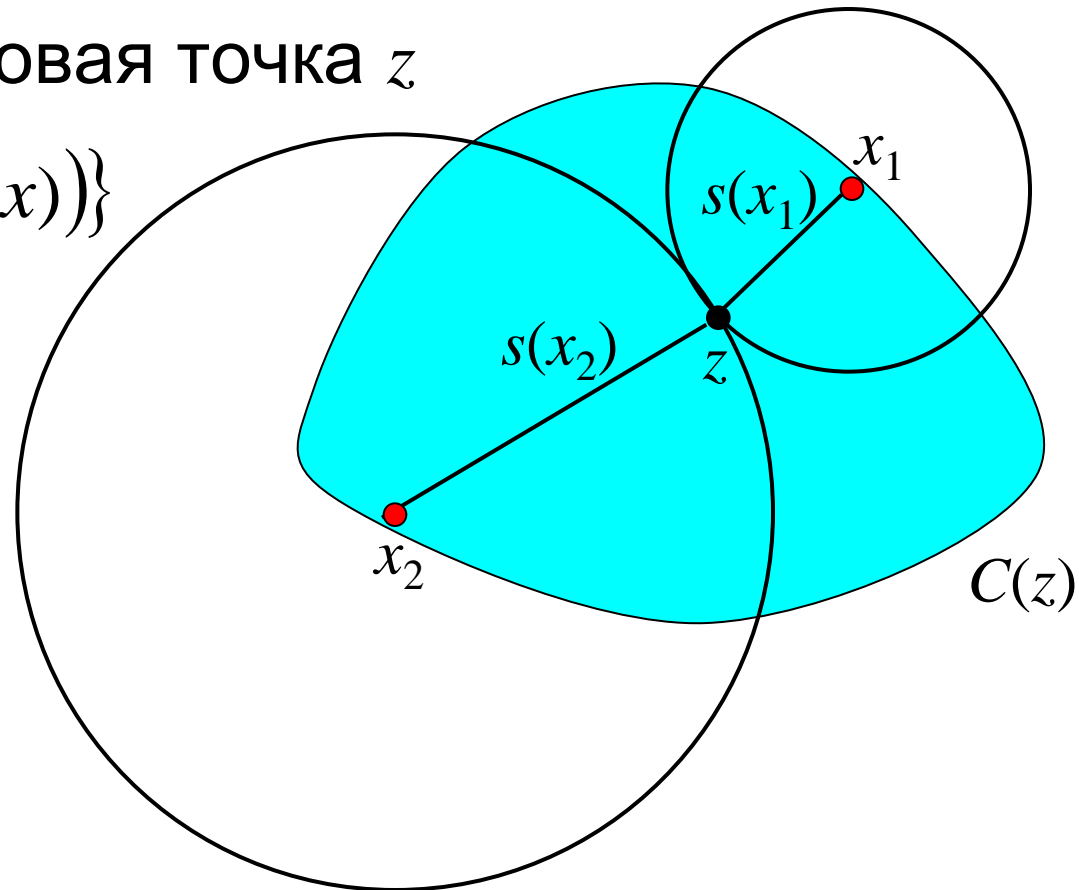
$$s(x) = \sup \left\{ s > 0 \mid \forall u \in B(x, s) S_{f(x)}(u) < t(x) \right\}$$



Зона покрытия стартовой точки

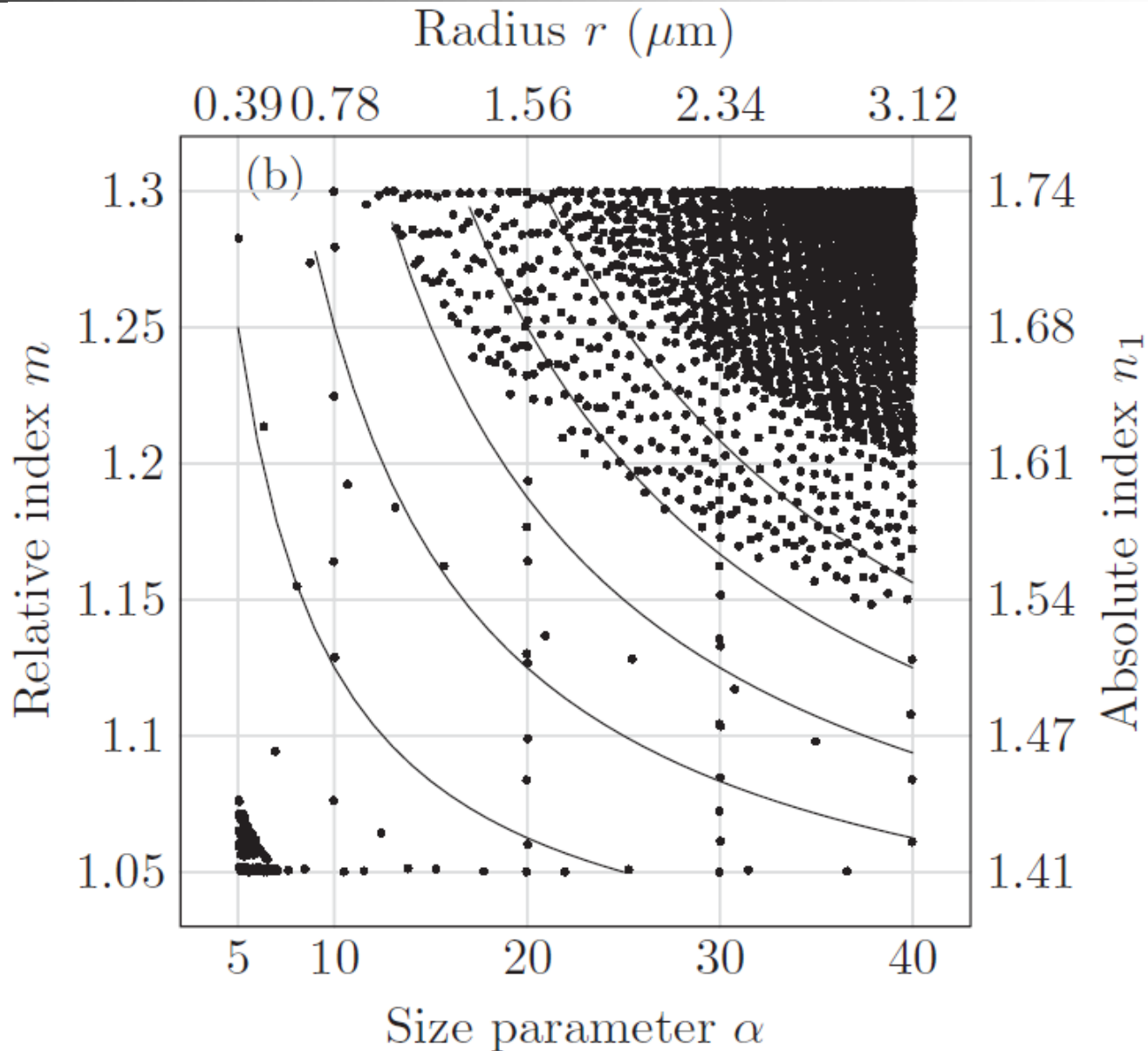
- Решение $x \rightarrow$ стартовая точка z

$$C(z) = \{x \in X \mid z \in B(x, s(x))\}$$



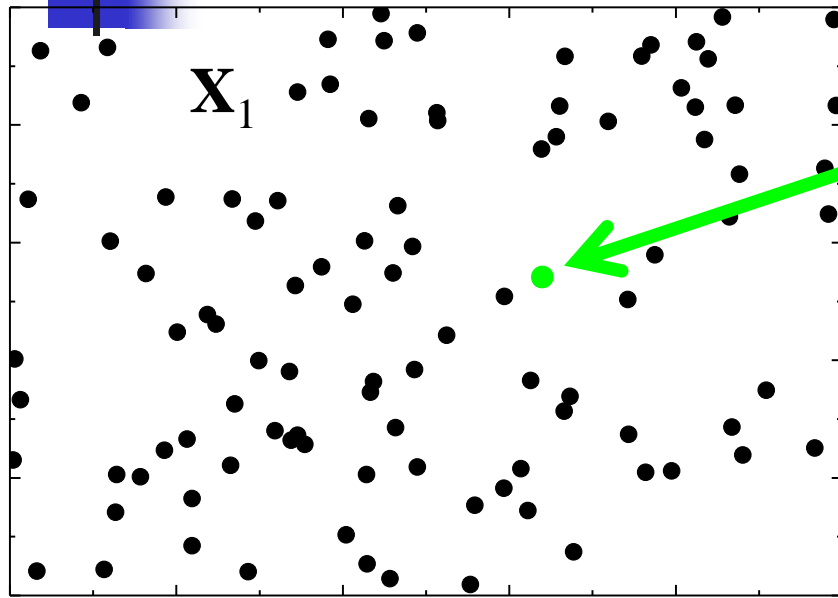
- Строим множество $\{z_j\}$ итерационно, пока не удовлетворим $X \subset \cup_j C(z_j)$

Начальные точки (2458) для шаров

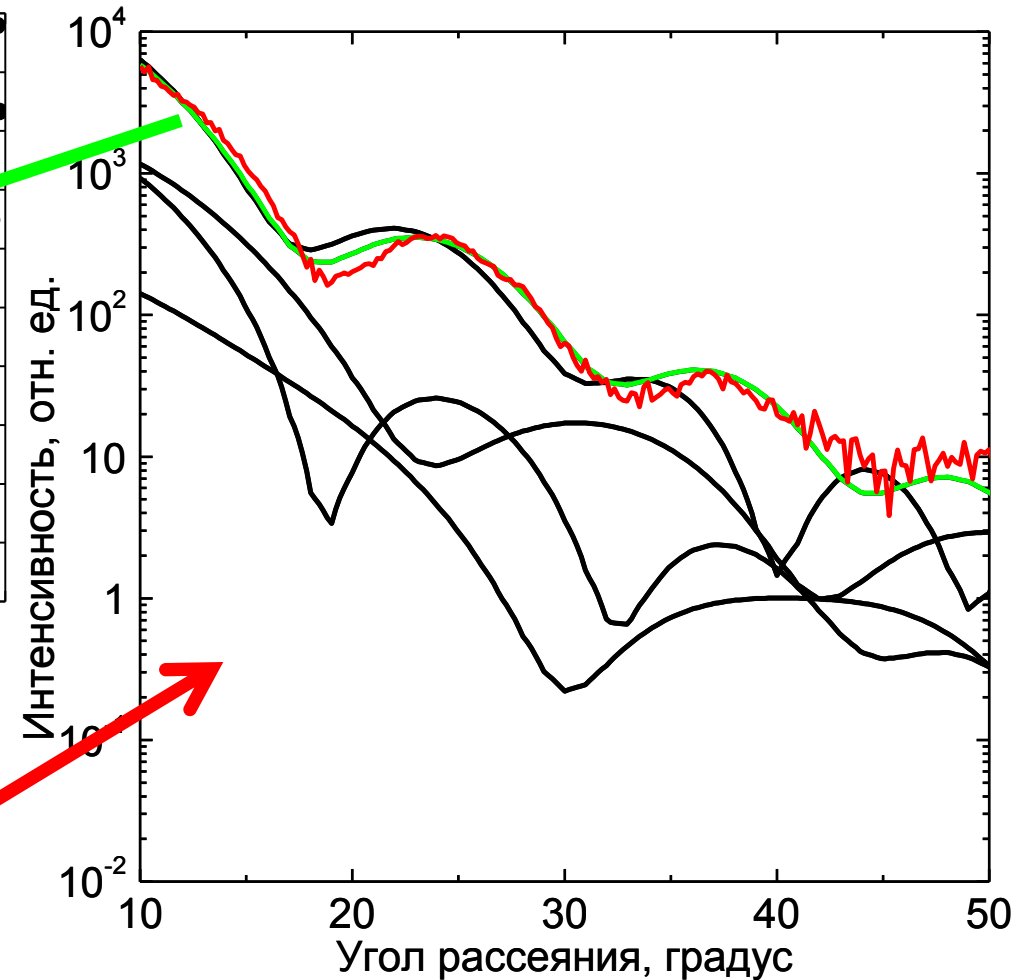
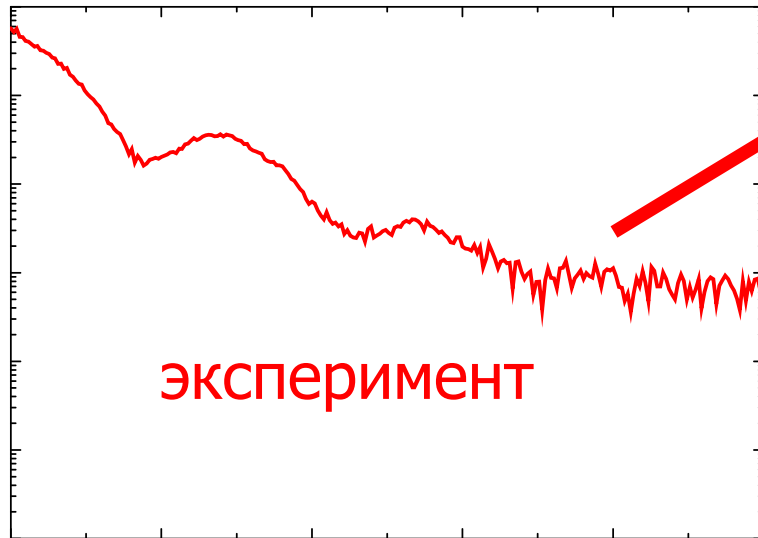


Интерполяция по базе данных

Характеристика 2



Характеристика 1

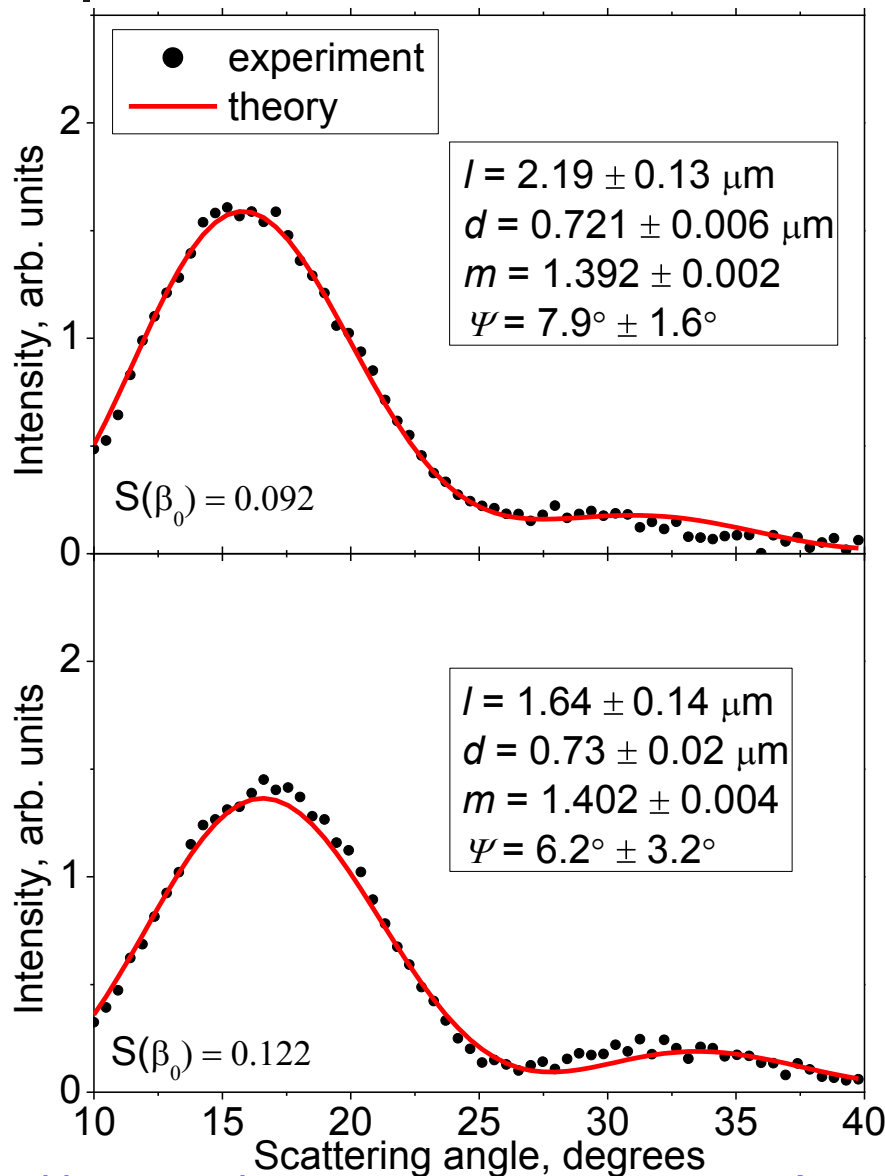


$$g_1 : \mathbf{R}^d \rightarrow X_1 \in X,$$

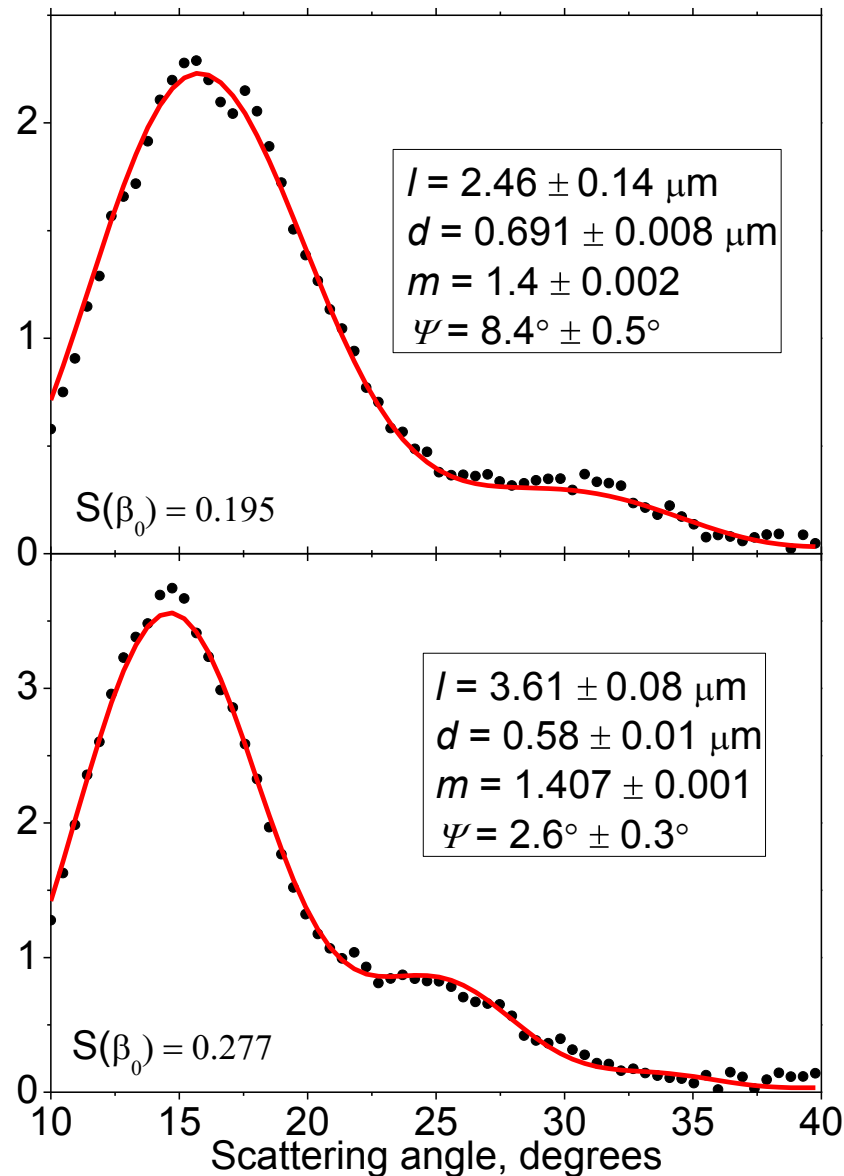
$$g_1(y) = \arg \min_{x \in X_1} \|y - f(x)\|$$

Пример – характеристика бактерий

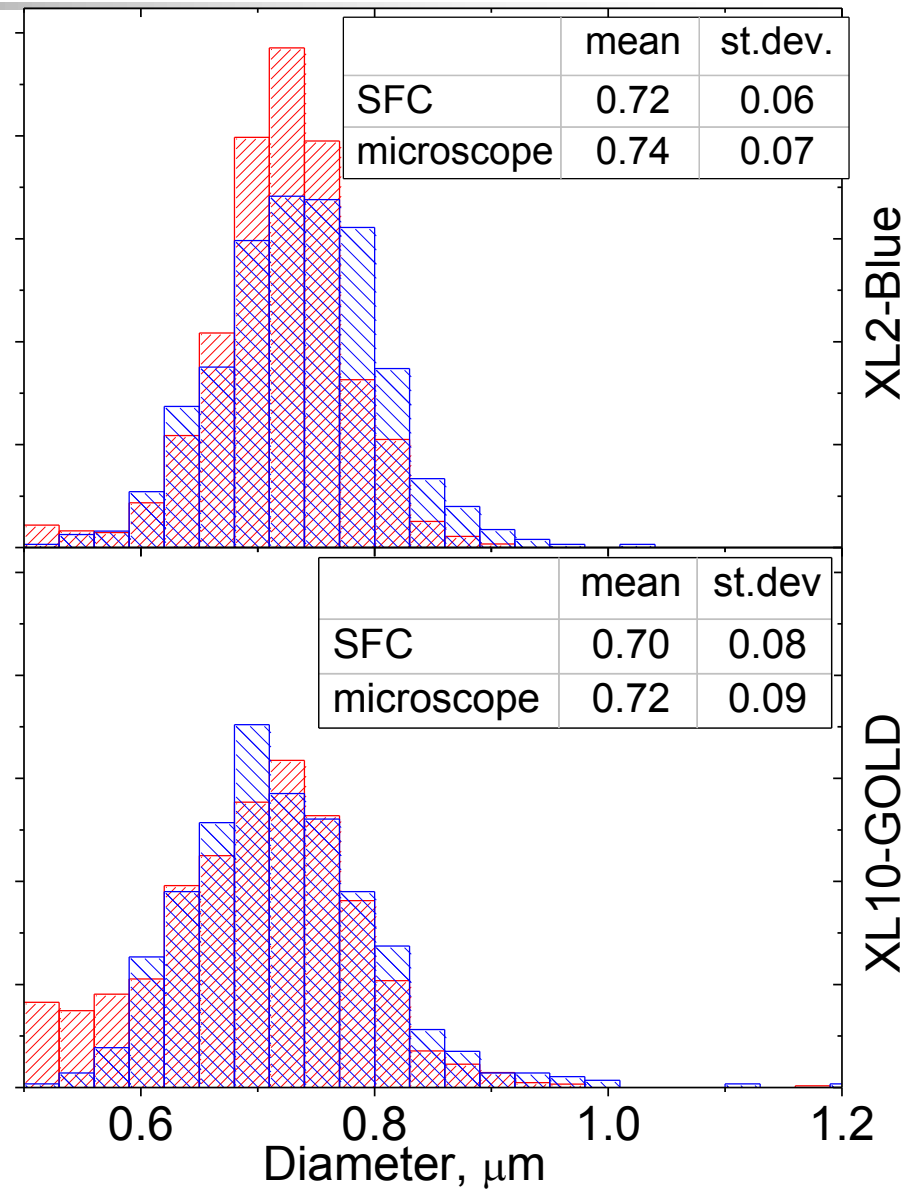
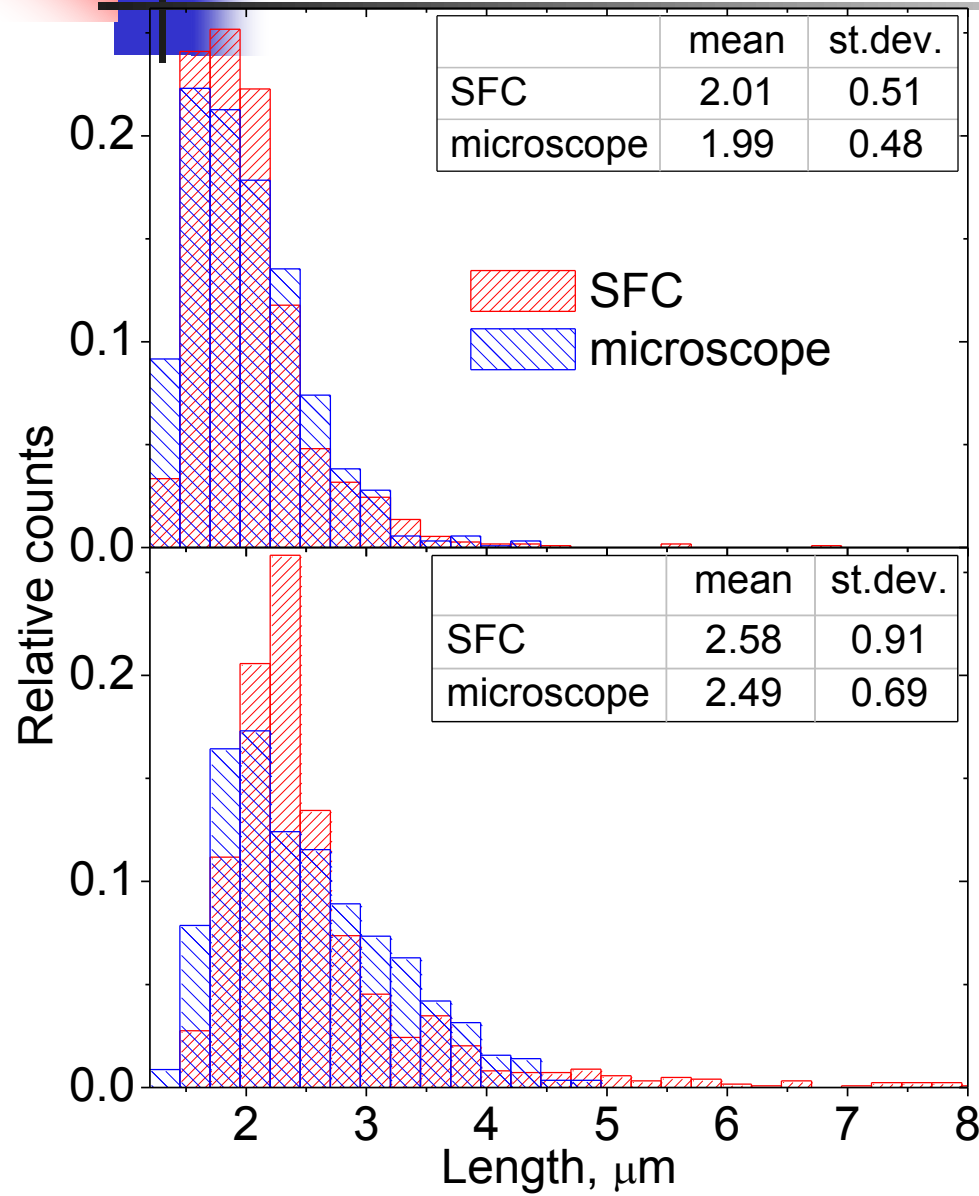
XL2-Blue



XL10-GOLD



Сравнение с микроскопом

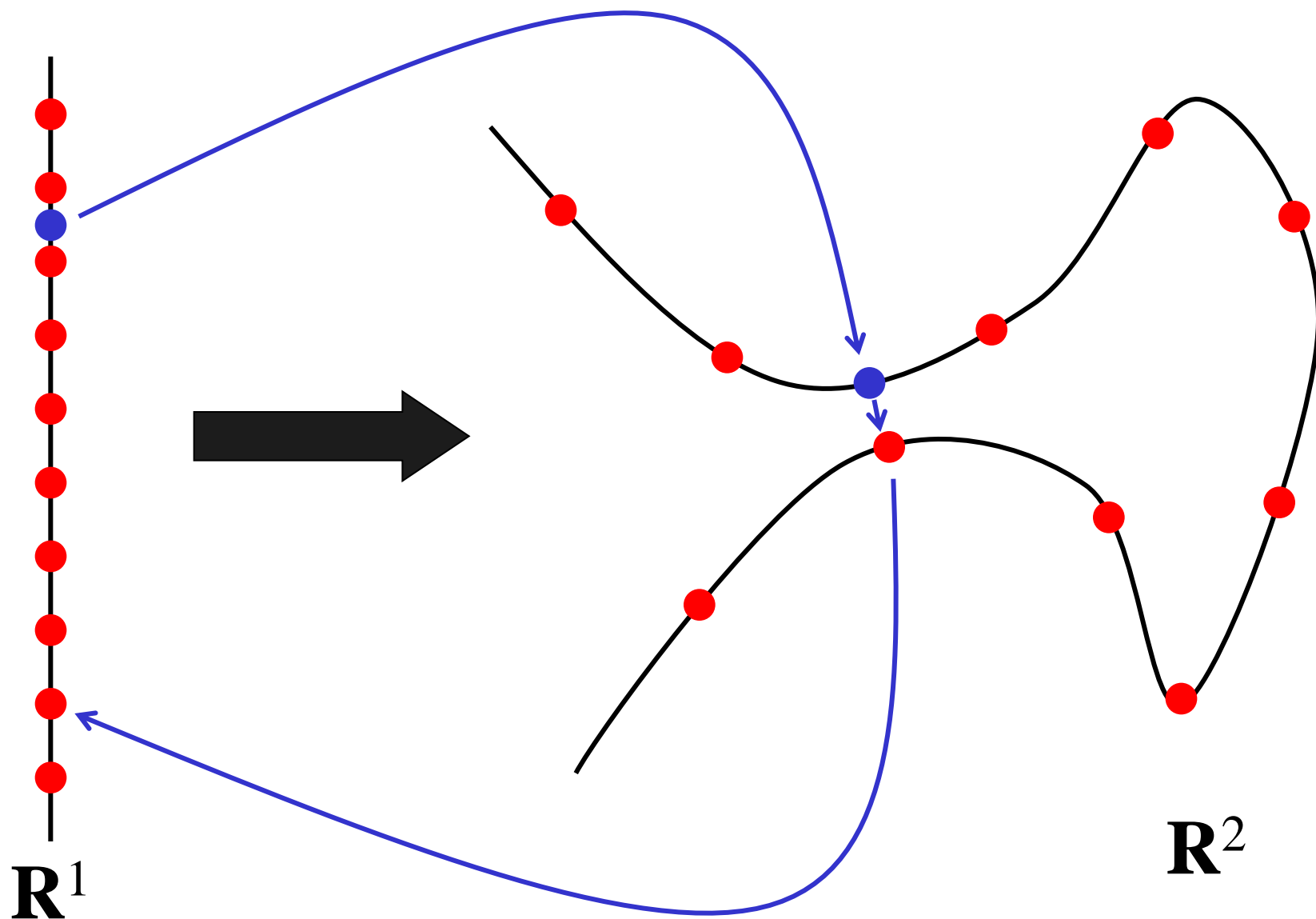




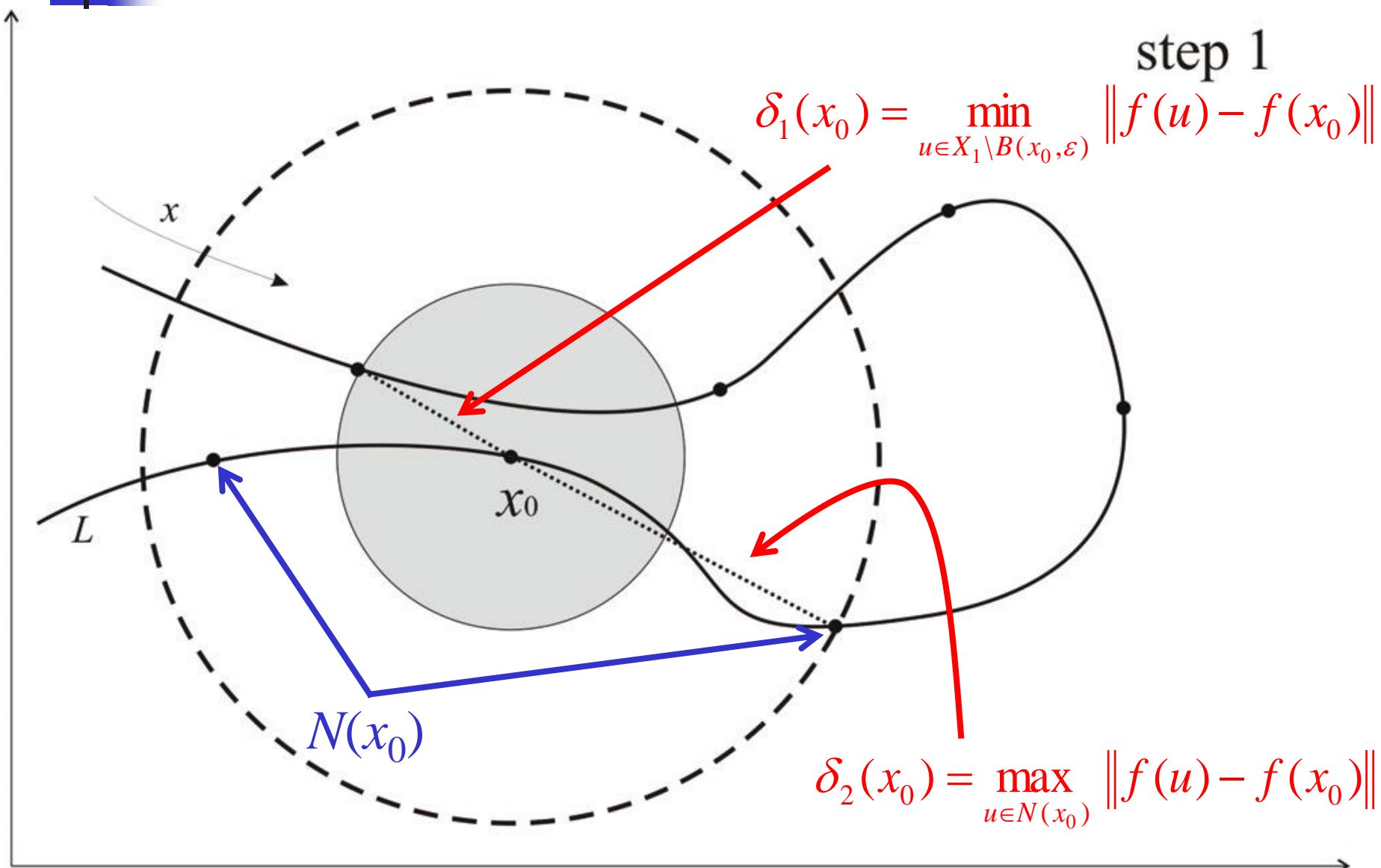
Адаптивное построение базы данных

- Информация об отображении доступна лишь в виде «небольшого» количества значений $f(x)$
- В общем случае нет никакой гарантии
- Практическая цель – удовлетворительная точность, т.е. избежать огромных ошибок
- База данных строится итерационно (добавляются точки) пока не удовлетворен критерий точности

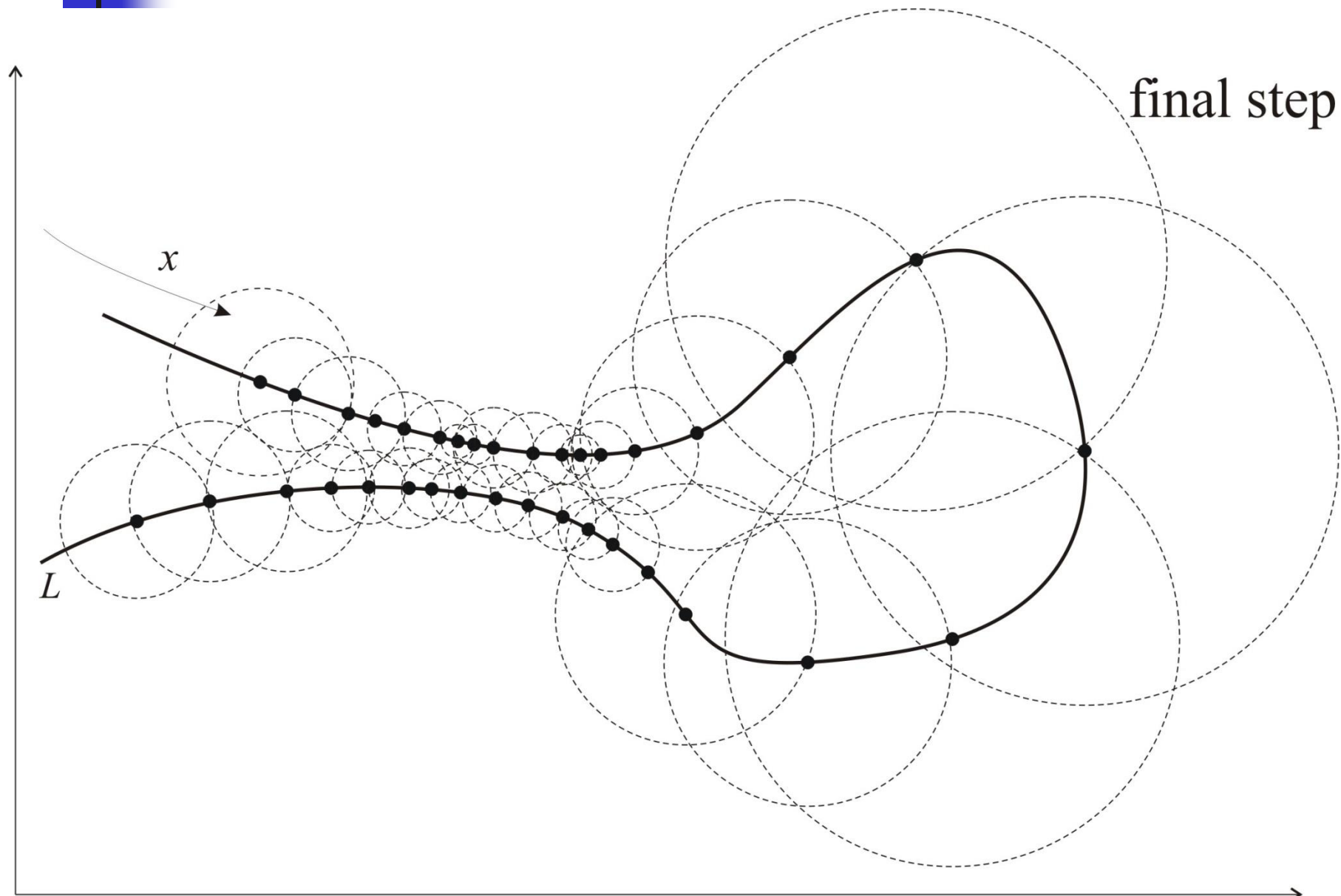
Пример огромной ошибки $f: \mathbf{R}^1 \rightarrow \mathbf{R}^2$



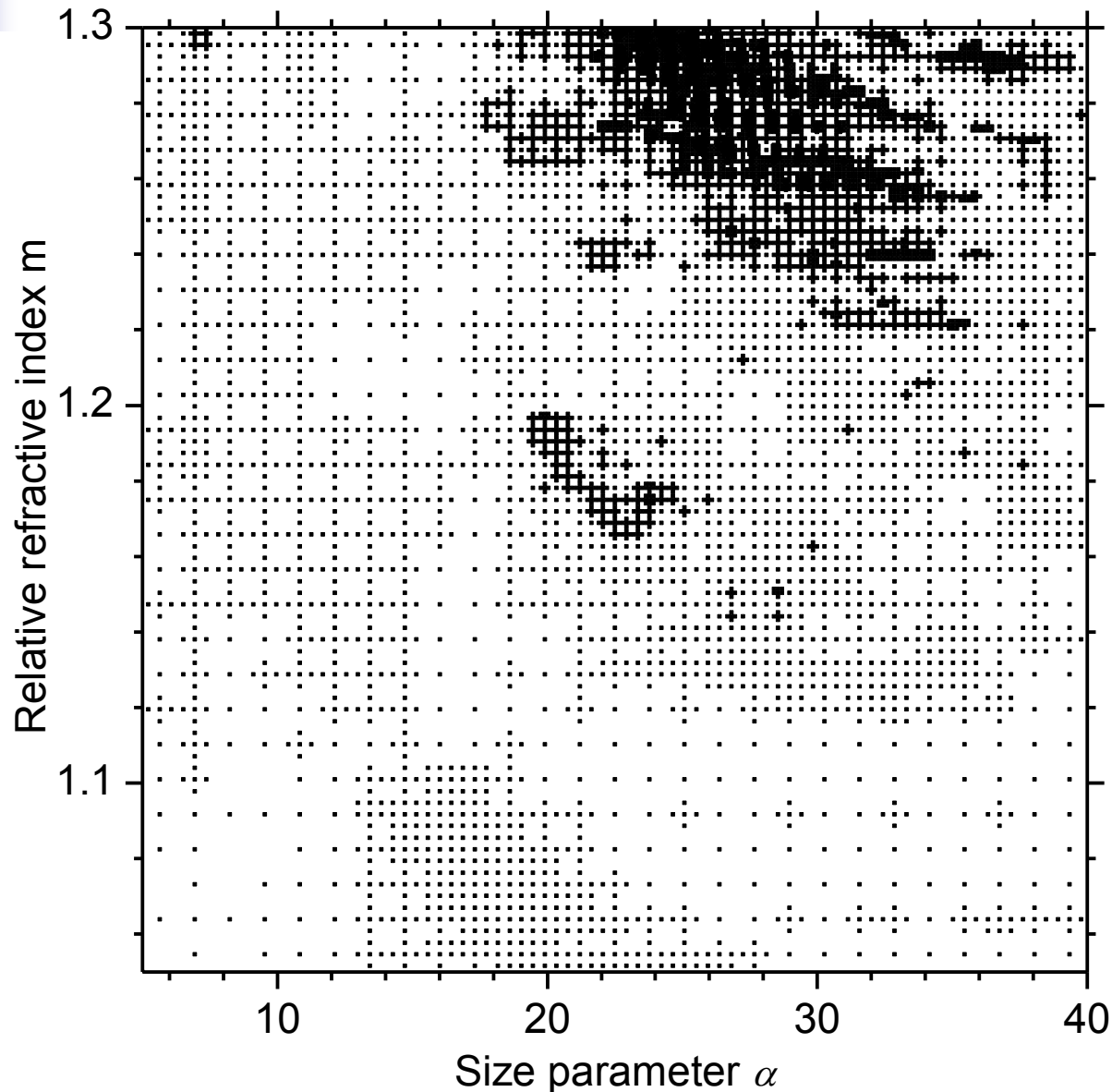
Итоговая цель: $\delta_2(x_0) < \delta_1(x_0)$



Итоговая база данных



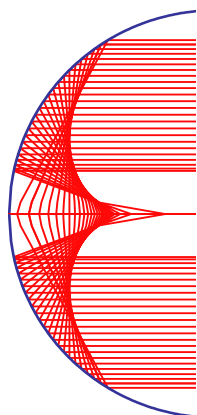
База данных для шаров (9366)





Спасибо!

Максим Юркин



Институт Химической Кинетики и Горения СО РАН,
Лаборатория Цитометрии и Биокинетики

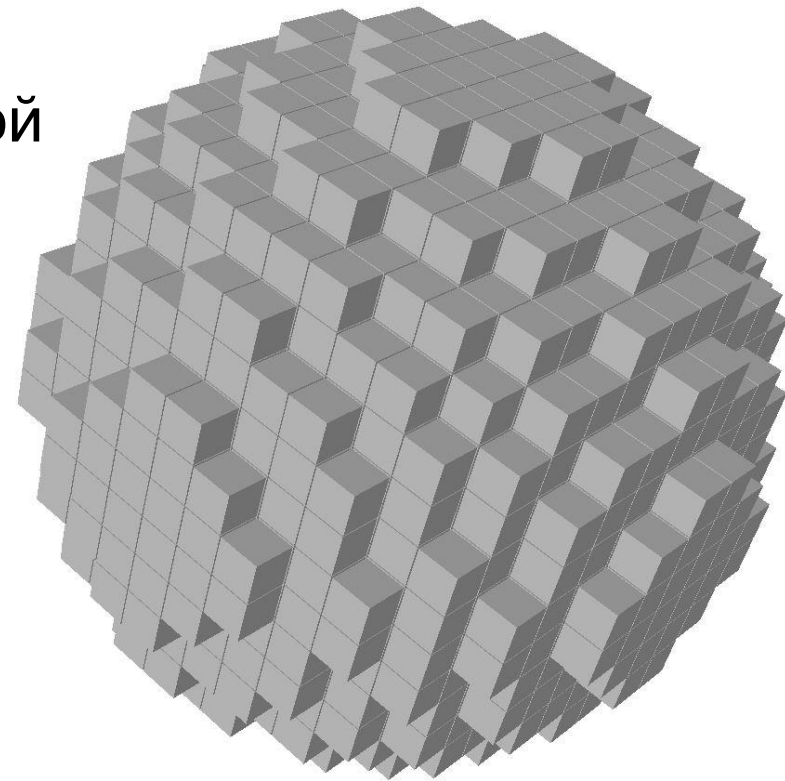
<http://cyto.kinetics.nsc.ru/>

тел: +7(383)3333240

yurkin@gmail.com

Метод дискретных диполей (МДД)

- Решается интегральное уравнение Максвелла в частотной области с помощью объёмной дискретизации
- Объект произвольной формы и структуры делится на маленькие диполи
- Диполи взаимодействуют между собой и с падающим излучением
⇒ система линейных уравнений
⇒ определяется поляризация диполей
- Далее – любые измеряемые характеристики светорассеяния



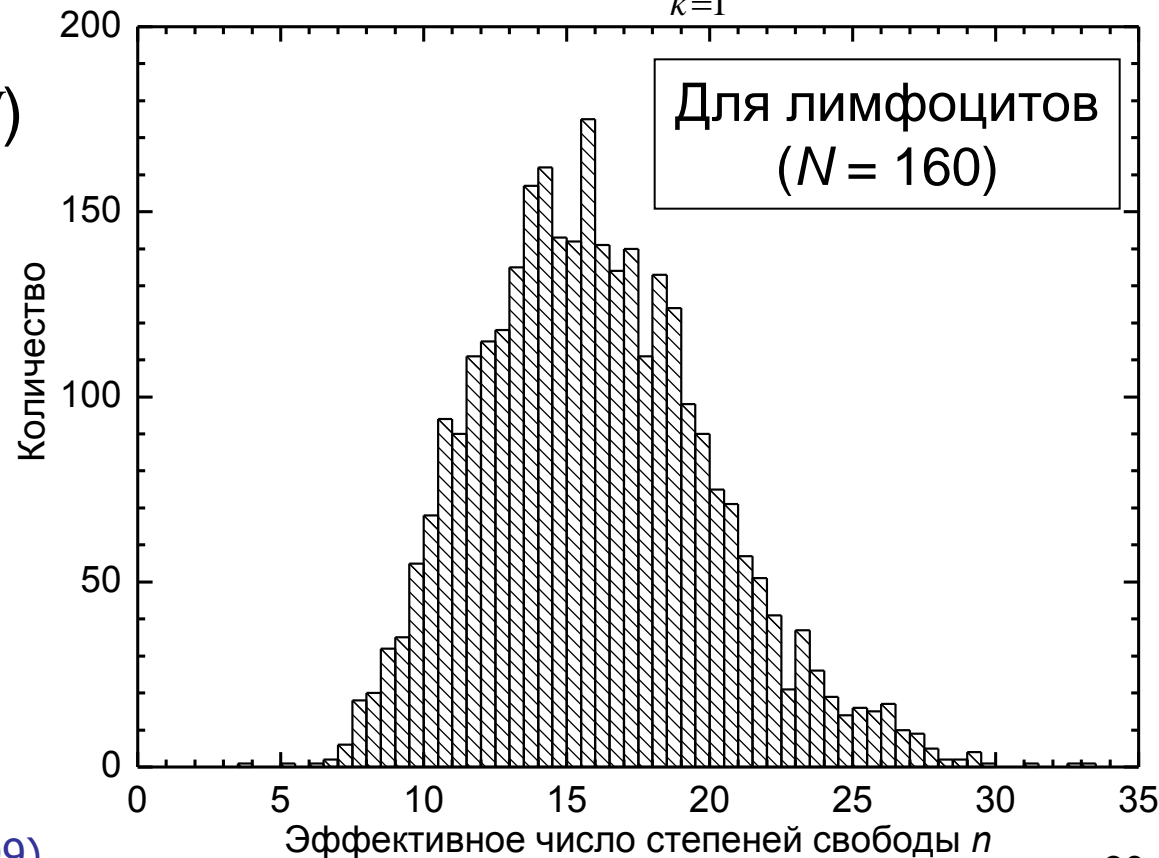
Эффективное число степеней свободы

- N зависимых измерений $\Leftrightarrow n$ независимых измерений (приблизительно)

$$n = \frac{N^2}{N + 2 \sum_{k=1}^{N-1} (N-k) \rho_k^2}$$

- Точно при полной зависимости ($n = 1$) и независимости ($n = N$)

- Описывает увеличение погрешностей при увеличении зависимости



Погрешности характеристики

- Тот же подход, что и при глобальной оптимизации
- Главное отличие, что $S(\beta)$ известно на наборе значений β , равномерно распределённых в \mathbf{B} (а не концентрированных вокруг β_0)
- Вычисление интегралов от $S(\beta)$ методом Монте-Карло приводит к определённой погрешности

$$\langle f(\beta) \rangle = \int_{\mathbf{B}} f(\beta) P(\beta) d\beta$$

$$\langle f(\beta) \rangle \approx \bar{g} / \bar{h}, \quad \bar{x} \equiv \frac{1}{N} \sum_{i=1}^N x_i, \quad g_i = f(\beta_i) h_i, \quad h_i = N p_i S(\beta_i)^{-k_{\text{eff}}/2}$$

$$\text{Var}[\bar{x}] = \frac{\text{Var}[x]}{N} \approx \frac{\text{SV}[x]}{N}, \quad \text{SV}[x] = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Характеризация тромбоцитов

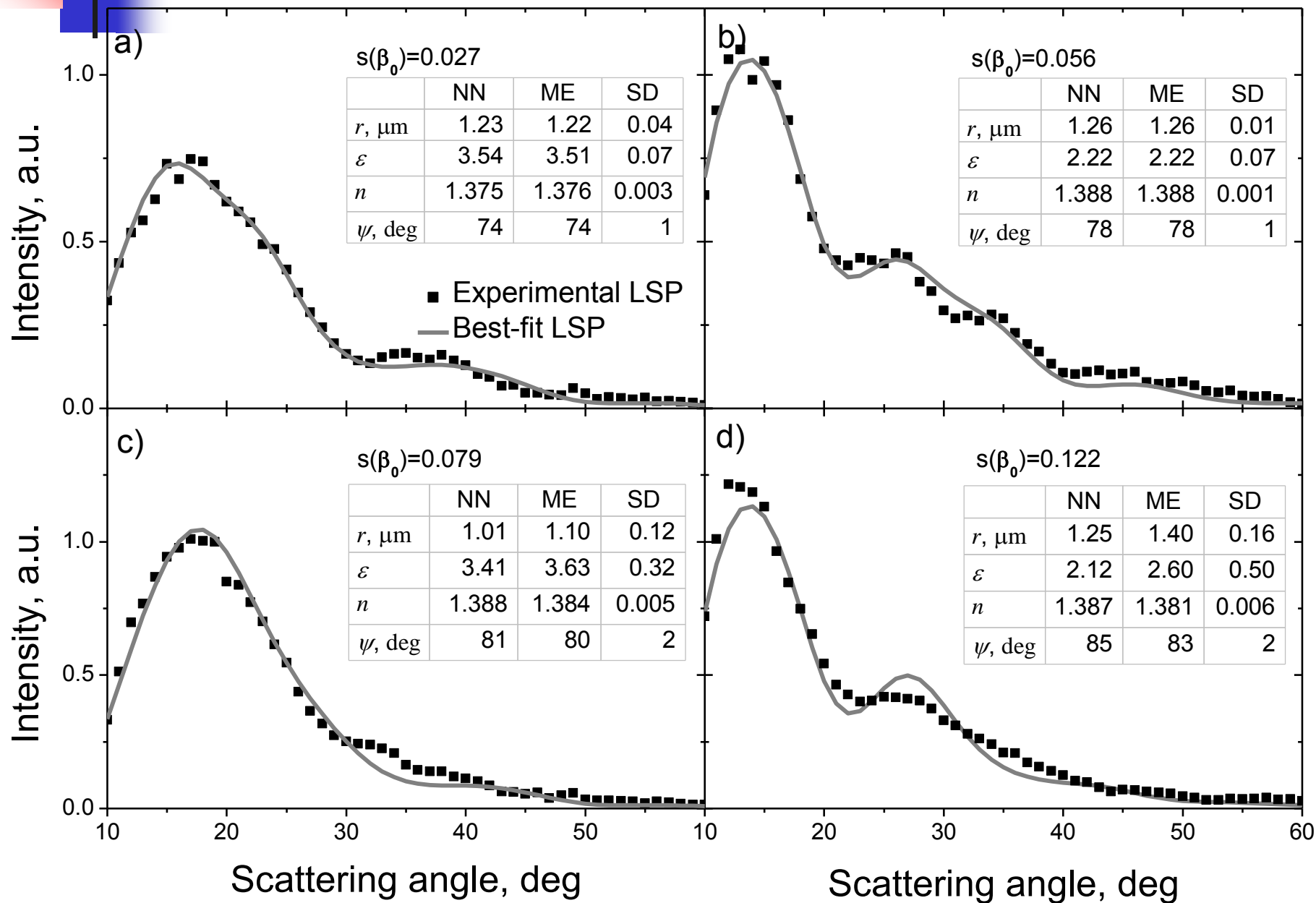
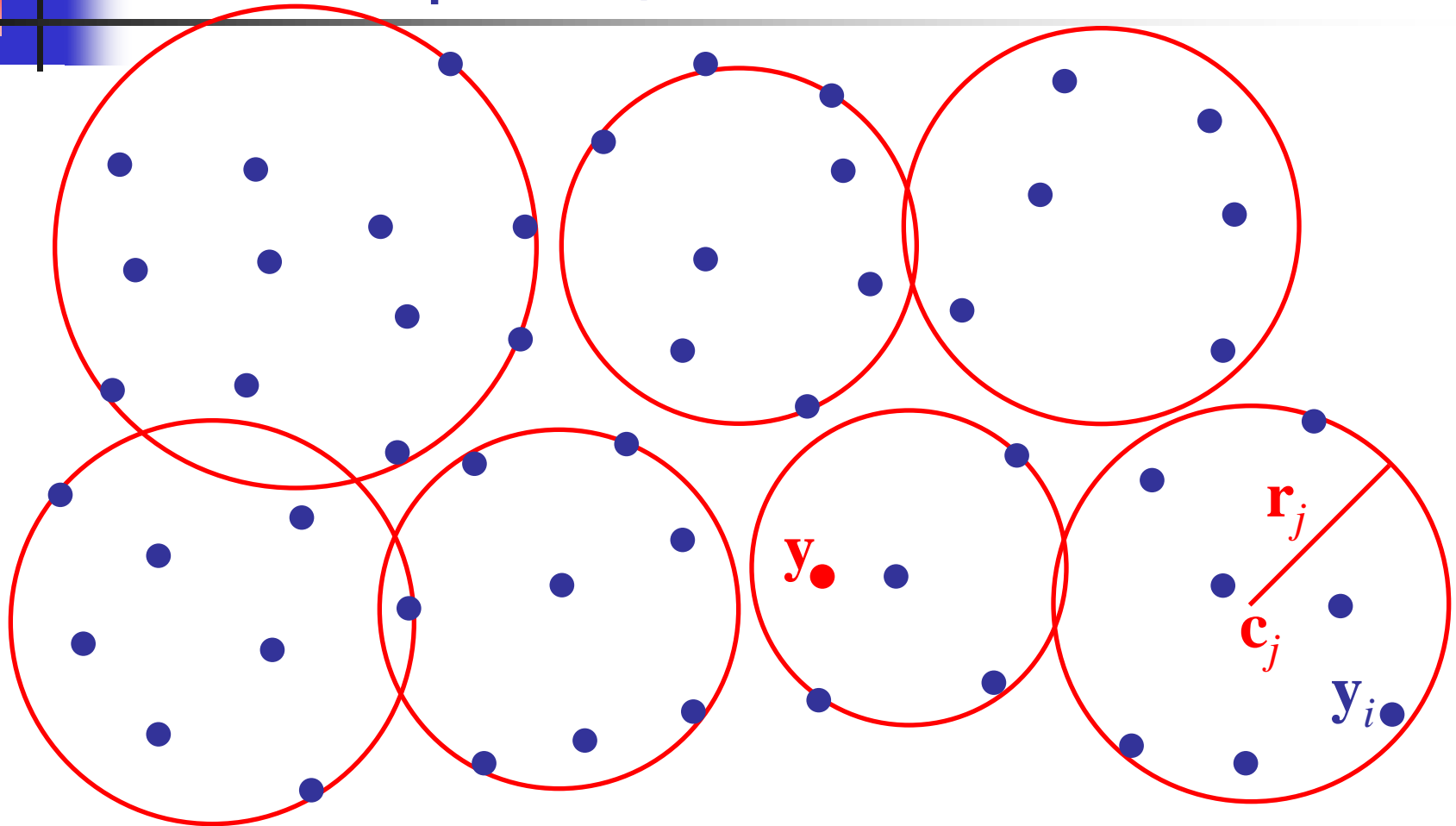




Таблица значений (база данных)

- Построение таблицы – много вычислений, но один раз.
- Вместо регрессии интерполяция методом ближайших соседей.
- Пример:
 - Таблица из 100 000 индикатрис (по 61 точке).
 - Обработка 1 экспериментальной индикатрисы (нахождение ближайшей) ~ 100 мс
- Чтобы еще быстрее – кластеризация базы данных

Кластеризация



- Рекуррентно уменьшаем минимум M ($\|y - y_i\|$ или $\|y - c_i\| + r_i$)
- Перебираем кластеры. Если $\|y - c_i\| < M + r_i$, то весь кластер можно пропустить, иначе перебираем все элементы кластера.
- Еще быстрее иерархическая (многоуровневая) кластеризация