

Реализация обработки данных малоуглового рассеяния с использованием веб- и параллельных средств фреймворка ROOT

А.Г. Соловьев¹, Т.М. Соловьева¹, Е.В. Земляная^{1,2}

¹ Объединенный институт ядерных исследований, Дубна, Россия,

² Государственный университет «Дубна», Дубна, Россия

Рассматривается задача получения структурной информации о фосфолипидных мембранах, образующих оболочку однослойных везикул. Эту информацию извлекают путем исследования рассеяния нейтронов или рентгеновских лучей на таких объектах. Получаемые экспериментальные кривые рассеяния приближают затем различными многопараметрическими теоретическими функциями. По результатам этой подгонки оценивают значения параметров везикулы: радиус, толщину оболочки и другие. Для решения подобных задач было разработано веб-приложение FITTER_WEB, основанное на новых инструментах пакета ROOT. FITTER_WEB развернут в облачной инфраструктуре ОИЯИ. Пользователю предоставляется полный контроль над процессом фитирования, а также возможность задействовать встроенные механизмы ROOT для распараллеливания этого процесса.

Ключевые слова: малоугловое рассеяние, облачные технологии, параллельные вычисления, веб-технологии, ROOT.

1. Введение

Метод малоуглового рассеяния рентгеновских частиц и нейтронов широко применяется при исследовании надатомной структуры неупорядоченных систем, таких как макромолекулы в растворе, аморфные сплавы, органические нанокристаллы и другие нанопористые вещества (см. обзор [1] и цитируемую литературу). В малоугловых экспериментах измеряется интенсивность рассеяния при малых величинах модуля вектора рассеяния, существенно меньших межатомных расстояний. Далее экспериментальная интенсивность сравнивается с рассеянием на целом ряде модельных тел. Сначала аппроксимация формы производится в рамках моделей, описывающих простые геометрические формы – сферы, параллелепипеды, эллипсоиды. Если же имеется дополнительная информация о строении частицы (симметрия, наличие внутренних элементов), то можно попытаться найти более тонкие детали внутренней структуры изучаемого объекта. Поэтому чаще всего строится несколько моделей исследуемых частиц, и моделируется рассеяние на них, а затем результат сравнивается с экспериментальной кривой. Количество моделей обычно достаточно велико, поэтому необходимо программное обеспечение, способное быстро рассчитать интенсивность рассеяния целого ряда моделей. Кроме того, в ряде случаев спектр малоуглового рассеяния могут содержать большое количество экспериментальных точек (1000 и более), что приводит к значительному увеличению времени счета. Существенного ускорения работы программ можно добиться, используя распараллеливание алгоритмов и проводя расчеты на многоядерных вычислительных архитектурах.

Для решения задач анализа данных малоуглового рассеяния на нано-структурах различного типа было разработано веб-приложение FITTER_WEB [2], основанное на новых инструментах пакета ROOT [3]. Программный комплекс FITTER_WEB развернут в облачной инфраструктуре Многофункционального информационно-вычислительного Комплекса (МИВК) Лаборатории Информационных Технологий им. М.Г. Мещерякова ОИЯИ [4,5].

Целью данной работы является апробация разработанного инструментария FITTER_WEB применительно к обработке спектров малоуглового рассеяния нейтронов (МУРН) на полидисперсных популяциях везикулярных систем на основе метода разделенных форм-факторов (РФФ), активно используемого для исследования однослойных везикул различного типа (см., например, недавнюю работу [6] и цитируемую литературу).

2. Приложение FITTER_WEB

Приложение FITTER_WEB предназначено для фитирования экспериментальных данных выбранной теоретической многопараметрической функцией. Реализация этого приложения основана на фреймворке ROOT, являющемся стандартом в обработке данных физики высоких энергий. Использование же новых инструментов ROOT7 [7] позволило реализовать его, как веб-приложение. FITTER_WEB был развернут в облачной инфраструктуре ОИЯИ [5], для чего была создана виртуальная машина с восемью ядрами (процессор Intel Xeon E5-2650 v4 @ 2,20 ГГц), операционной системой Ubuntu 22.04.1 и 32 ГБ оперативной памяти [8]. FITTER_WEB доступен по адресу <http://fitter.jinr.ru>.

Окно приложения показано на рис. 1. Пользователю предоставляется возможность выбрать файл с экспериментальными данными и файл с ROOT-скриптом, задающим теоретическую функцию. Пользователь может редактировать данный скрипт или создать новый, отвечающий его конкретной задаче.

Фитирование проводится по методу наименьших квадратов. Минимизация осуществляется с помощью входящих в состав ROOT инструментов для минимизации функционалов, включая пакет MINUIT [9, 10]. Для управления опциями фитирования используется входящий в ROOT7 компонент FitPanel. С его помощью пользователь может выбрать метод и алгоритм минимизации, установить начальные значения параметров, диапазон их изменения, область фитирования и т. д. Результат фитирования может быть сохранен в текстовом или графическом виде.

Каждая процедура фитирования запускается в отдельном процессе. Таким образом несколько пользователей могут работать с FITTER_WEB одновременно, причем каждый пользователь может запустить несколько сеансов фитирования, открыв его в разных вкладках браузера.

Процедура фитирования хорошо поддается распараллеливанию с помощью встроенных механизмов ROOT:

- параллелизма на уровне задач;
- параллелизма на уровне данных.

Возможность задействовать эти механизмы, как по отдельности, так и совместно, предоставляется самому пользователю.

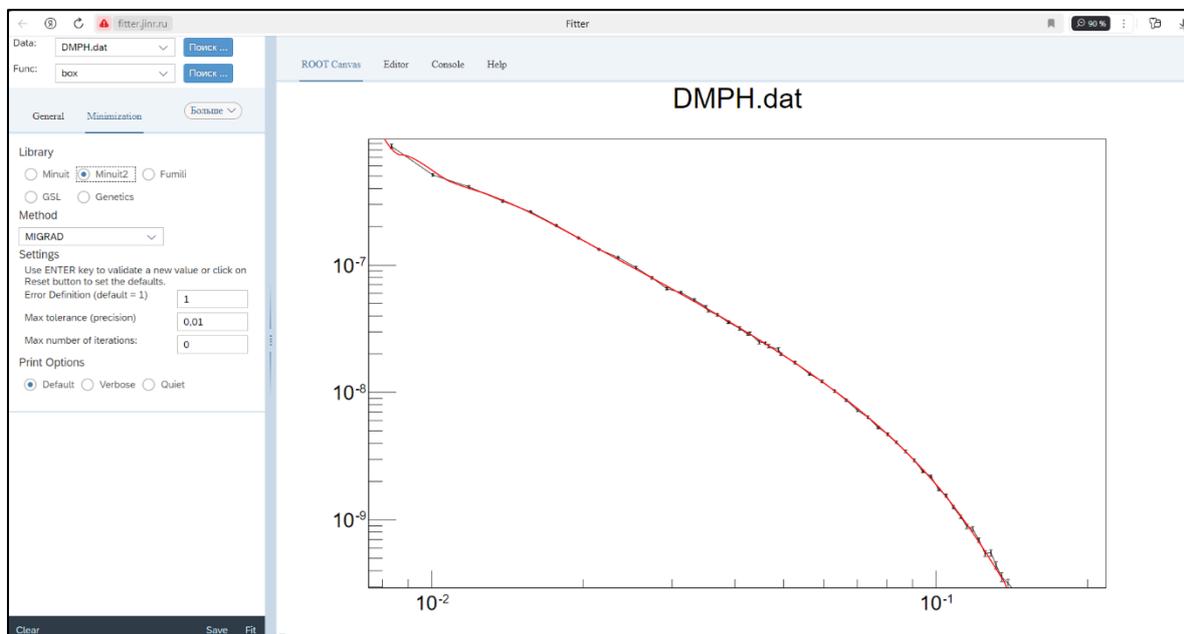


Рис. 1. Окно приложения FITTER_WEB

Параллелизм на уровне задач реализован в ROOT через многопоточность. По умолчанию ROOT работает в последовательном режиме. Но некоторые его компоненты, к которым относятся и фитирование, имеют дополнительные реализации, которые могут распределять работу

по нескольким потокам, чтобы использовать несколько ядер машины. Чтобы активировать использование именно этих многопоточных реализаций, пользователь должен вызвать `ROOT::EnableImplicitMT()`. При использовании `FITTER_WEB` этот вызов можно поместить в скрипт с реализацией теоретической функции. Поскольку дальнейшее происходит «под капотом», параллелизмом управляет сам `ROOT`, в этом случае говорят о *неявной* многопоточности.

Параллелизм на уровне данных использует векторизацию [11]. Векторизация позволяет загрузить несколько значений определенного типа в SIMD массив и одновременно выполнить над ними одну и ту же операцию. Для того чтобы воспользоваться этим механизмом в `FITTER_WEB`, пользователь должен модифицировать скрипт теоретической функции, так чтобы ее аргумент и возвращаемое значение были векторными типами (`ROOT` определяет новые типы данных `ROOT::Double_v` и `ROOT::Float_v`, представляющие собой SIMD-векторы двойной и одинарной точности соответственно).

Процесс распараллеливания, выполняемый в каждом из этих алгоритмов, представлен на рис. 2. Реализация основана на шаблоне MapReduce [12]. На первом этапе происходит разбивка данных на фрагменты. Затем метод `Map` класса `ROOT::TThreadExecutor` распределяет обработку этих фрагментов по имеющимся потокам.

Если пользователь предоставил векторизованную функцию, которая принимает SIMD-типы `ROOT` в качестве параметров, количество вызовов пользовательской функции при каждой такой обработке сокращается пропорционально размерности SIMD-массива.

На заключительном этапе, после того как все фрагменты данных обработаны, метод `Reduce` класса `ROOT::TThreadExecutor` объединяет результаты всех подзадач в общий результат.

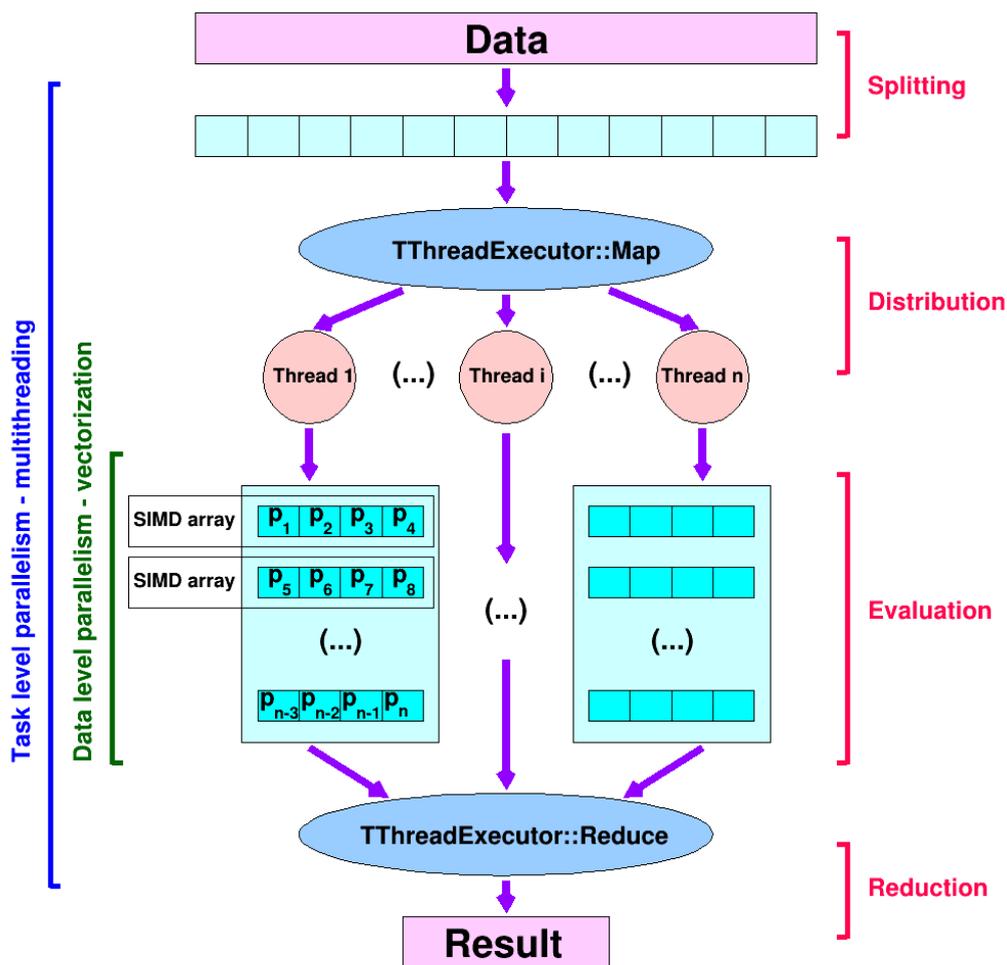


Рис. 2. Схема работы встроенных механизмов ROOT для распараллеливания фитирования

3. Модель полидисперсной везикулярной системы

Фосфолипидные везикулы используются в фармакологии и косметологии для транспорта лекарственных или косметически активных препаратов. Поэтому получение информации об их структуре и свойствах в различных средах имеет важное практическое значение. Однослойные везикулы представляют собой полые внутри нано-размерные объекты, имеющие близкую к сферической форму и образующие в воде или жидких растворах популяции, размеры которых в общем случае могут быть разными. Оболочка везикул представляет собой бислой, состоящий из липидных молекул, имеющих гидрофильную полярную головку и длинные гидрофобные углеводородные цепочки [1]. При этом все неполярные углеводородные «хвосты» находятся в центральной части мембраны, в то время как полярные головки располагаются в наружных областях мембраны, соприкасаясь с молекулами воды (раствора) как снаружи, так и внутри везикулы.

Метод РФФ для анализа структуры везикулярных систем по данным МУРН подробно излагается в ряде работ, включая [1, 6, 13], и основан на факторизации выражения амплитуды МУРН на везикулярной системе в приближении сферической симметрии. В предположении, что все везикулы имеют одинаковый радиус R (монодисперсная популяция), толщину бислоя d и что $R \gg d$, сечение рассеяния на монодисперсной популяции однослойных везикул можно записать как

$$\frac{d\Sigma}{d\Omega_{mon}}(q) = nF_s(q, R)F_b(q, \rho)S(q, R), \quad (1)$$

где q – модуль вектора рассеяния, n – число везикул в единице объема, $S \approx 1$ – структурный фактор, отражающий взаимодействие везикул в популяции [1], F_s – зависящий только от радиуса R форм-фактор сферической везикулы с бесконечно тонкой толщиной бислоя, а F_b – форм-фактор симметричного липидного бислоя, зависящий только от толщины бислоя d (с учетом $R \gg d$ мы пренебрегаем кривизной бислоя). Формфакторы $F_s(q, R)$ и $F_b(q, d)$ определяются следующими выражениями:

$$F_s(q, R) = \left(4\pi \frac{R}{q} \sin(qR) \right)^2, \quad (2)$$

$$F_b(q, \rho) = \left(\int_{-\frac{d}{2}}^{\frac{d}{2}} \Delta\rho(x) \cos(qx) dx \right)^2. \quad (3)$$

В формуле (3) $\Delta\rho(x)$ – разность (контраст) между плотностью длины рассеяния фотона от липидного бислоя везикулы (ρ) и от окружающей жидкости (ρ_0). При использовании многих функций, часто применяемых для моделирования структуры бислоя поперек мембраны, интеграл (3) можно вычислить аналитически, что существенно упрощает расчеты.

Для учета полидисперсности радиуса везикул используется несимметричное распределение Шульца:

$$G(R, < R >) = \frac{R^m}{m!} \left(\frac{m+1}{< R >} \right)^{m+1} \exp \left[-\frac{(m+1)R}{< R >} \right]. \quad (4)$$

Тогда макроскопическое сечение $d\Sigma(q)/d\Omega$ для полидисперсной популяции везикул среднего радиуса $< R >$ имеет вид:

$$\frac{d\Sigma}{d\Omega}(q) = \frac{\int_{R_{min}}^{R_{max}} \frac{d\Sigma}{d\Omega_{mon}}(q, R, < R >) G(R, < R >) dR}{\int_{R_{min}}^{R_{max}} G(R, < R >) dR}. \quad (5)$$

С учетом некогерентного фона I_b и разрешения спектрометра реально измеряемое значение макроскопического сечения $\frac{d\Sigma}{d\Omega}(q)$ записывается как:

$$\frac{d\Sigma}{d\Omega}(q)_{cor} = \frac{d\Sigma}{d\Omega}(q) + \frac{1}{2} \Delta^2 \frac{d^2}{dq^2} \left[\frac{d\Sigma}{d\Omega}(q) \right] + I_b. \quad (6)$$

Здесь Δ – второй момент функции разрешения спектрометра.

Параметры подгоняются под экспериментальные данные МУРН путем минимизации невязки между экспериментальными и теоретическими значениями МУРН. Здесь $\delta(q_i)$ – ошибки экспериментальных данных; N – число экспериментальных точек; k – количество фитируемых параметров:

$$\chi^2 = \frac{1}{N-k} \sum_{i=1}^N \left(\frac{\frac{d\Sigma}{d\Omega}(q_i) - \frac{d\Sigma}{d\Omega}(q_i)_{exp}}{\delta(q_i)} \right)^2. \quad (7)$$

В нашем случае параметрами, определяемыми при решении задачи фитирования, являются величины m , d , R , I_B , n , а также параметры функции ρ , определяющей структуру бислоя. В данной работе мы рассмотрим три варианта функции ρ , моделирующие плотность длины рассеяния поперек мембраны (бислоя) везикулы.

4. Результаты расчетов

В данной работе представлены результаты РФФ-анализа спектра МУРН на полидисперсной популяции однослойных везикул димеристоила фосфатидилхолина в тяжелой воде. Измерения проведены на малоугловом спектрометре ЮМО в Лаборатории нейтронной физики им. Дзержелова ОИЯИ. Для анализа спектров МУРН на основе модели РФФ, определяемой формулами (1–7), производился вызов соответствующей C++ функции в рамках веб-приложения FITTER_WEB.

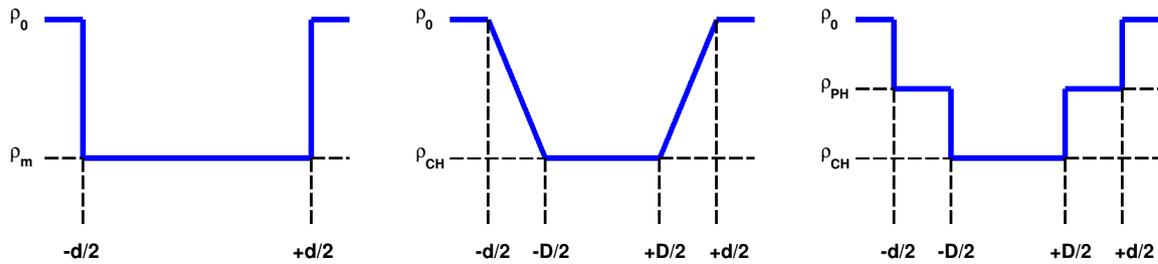


Рис. 3. Функция плотности для модели box (слева), hh (в центре) и step (справа)

В расчетах использовались следующие модели, наиболее часто используемые для анализа данных МУРН на везикулярных системах. В модели box предполагается однородная структура поперек бислоя: $\rho = \rho_m$ поперек бислоя. Такая плотность обычно используется для первоначальных расчетов с целью предварительной оценки значений толщины бислоя, среднего радиуса и полидисперсности везикул в образце. Гидрофобно-гидрофильная модель hh учитывает разницу между рассеянием в области полярных головок на границах бислоя (гидрофильная область) и в центральной области углеводородных цепочек (гидрофобная область). Здесь используется кусочно-линейная аппроксимация плотности в гидрофильной области и введена в качестве параметра толщина области D углеводородных цепочек (CH). Ступенчатая модель step также учитывает различие между рассеянием в области полярных головок и углеводородных цепочек и предполагает, что молекулы раствора проникают только в область полярных головок бислоя на краях бислоя. Значения ρ_0 и ρ_{CH} – известны. Для данного эксперимента $\rho_0 = 6.37 \times 10^{+10}$, $\rho_{CH} = -0.36 \times 10^{+10}$.

Таблица 1. Параметры фитирования для моделей box, hh и step

Параметры	Значения для модели box	Значения для модели hh	Значения для модели step
χ^2	1.22	1.058	1.056
$R, \text{ \AA}$	264.6 ± 3.5	263.16 ± 5	263.2 ± 5
$d, \text{ \AA}$	36.9 ± 0.13	41.5 ± 0.4	41.2 ± 1.3
m	8.2 ± 0.5	8.2 ± 0.5	8.2 ± 0.5
$n, \text{ см}^{-3}$	$(2.7 \pm 0.09) \times 10^{+14}$	$(1.98 \pm 0.16) \times 10^{+14}$	$(2.23 \pm 0.21) \times 10^{+14}$
$I_B, \text{ см}^{-1}$	$(1.42 \pm 0.08) \times 10^{-2}$	$(1.26 \pm 0.12) \times 10^{-2}$	$(1.26 \pm 0.12) \times 10^{-2}$
$\rho, \text{ см}^{-2}$	$\rho_m = (5.27 \pm 0.09) \times 10^{+10}$		$\rho_{PH} = (2.19 \pm 0.54) \times 10^{+10}$
$D, \text{ \AA}$		28 ± 2.4	$18,2 \pm 0,99$

Полученные значения параметров, представленные в таблице 1, согласуются с результатами ранее проведенного анализа аналогичного спектра, опубликованными в работе [13].

На рисунке 1 в окне веб-приложения можно видеть согласие расчетной кривой с экспериментальными данными для модели box. Рисунок 4 показывает аналогичные результаты для моделей hh и step.

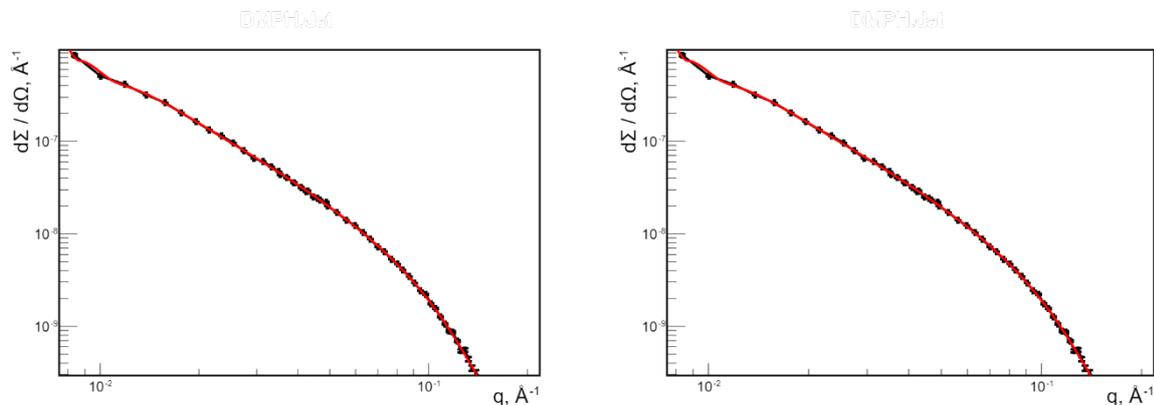


Рис. 4. Сечение рассеяния для модели hh слева и step справа

Для исследования эффекта параллелизма были проведены измерения времени работы веб-приложения при расчетах с разным количеством потоков (threads), задаваемых процедурой `ROOT::EnableImplicitMT(threads)`. Ускорение вычислений (speedup) по сравнению с расчетами в однопоточном режиме растет с увеличением числа потоков и достигает 6 раз. Расчеты производились при условии, что веб-приложение в каждую конкретную единицу времени обслуживает только одного пользователя.

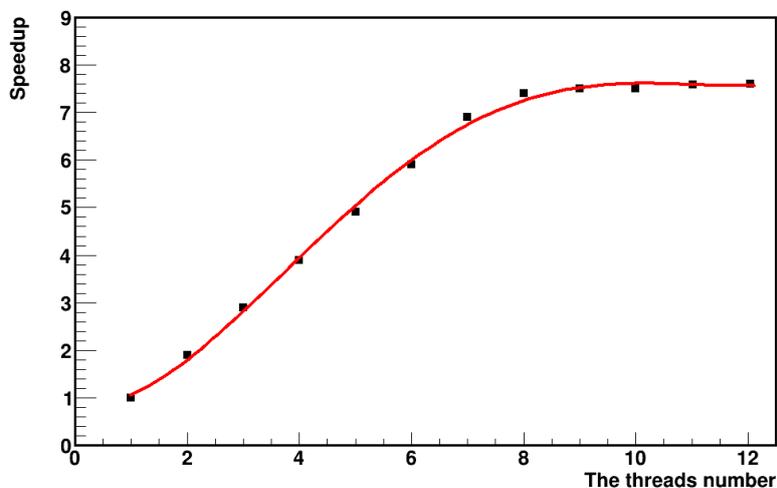


Рис. 5. Зависимость коэффициента ускорения от числа потоков

Показанные на рисунке 5 результаты соответствуют оценкам эффективности распараллеливания процесса фитирования экспериментальных данных, полученным для феррожидкости с наночастицами Fe3O4 в воде и керосине, представленным в работе [2].

5. Заключение

Разработано приложение FITTER_WEB, предоставляющее возможность фитировать экспериментальные данные теоретическими функциями. FITTER_WEB развернут в облачной инфраструктуре для свободного доступа и может эффективно использоваться для решения широкого круга задач. Приложение было успешно апробировано для анализа структуры полидисперсных везикулярных систем на основе обработки данных малоуглового рассеяния. Для ускорения вы-

числений в приложении используется неявная многопоточность и векторизация функций. В рассмотренном здесь случае, когда экспериментальный спектр состоит из 70 точек, ускорение вычислений достигает величины более чем в 7 раз, что является хорошим результатом с учетом того, что расчеты велись на 8-ядерной машине. Можно ожидать, что при анализе спектров данных с большим количеством точек на вычислительных системах с большим количеством ядер эффект от использования параллелизма будет более значительным.

Авторы выражают благодарность сотрудникам МИВК за вычислительную поддержку облачной инфраструктуры ОИЯИ.

Литература

1. Киселев М.А. Методы исследования липидных наноструктур на нейтронных и синхротронных источниках // Физика элементарных частиц и атомного ядра. 2011. Т. 42, № 2. С. 579–630.
2. Соловьев А.Г., Соловьева Т.М., Куклин А.И., Балашою М. Разработка веб-приложения для фитирования данных спектрометра малоуглового рассеяния нейтронов // Параллельные вычислительные технологии (ПаВТ'2023): Короткие статьи и описания плакатов. (Санкт-Петербург, 28–30 марта 2023 г.). Челябинск: Издательский центр ЮУрГУ, 2023. С. 206–214. DOI: 10.14529/pct2023.
3. Brun R., Rademakers F. ROOT – An Object-Oriented Data Analysis Framework // Nucl. Inst. & Meth. in Phys. Res. A. 1997. Vol. 389. P. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.
4. Adam G., Bashashin M, et al. IT-ecosystem of the HybriLIT heterogeneous platform for high-performance computing and training of IT-specialists // CEUR Workshop Proceedings (CEUR-WS.org, ISSN:1613-0073, Published by RWTH Aachen University). 2018. Vol. 2267. P. 638–644.
5. Baranov A.V., Balashov N.A., Kutovskiy N.A. et al. JINR cloud infrastructure evolution // Phys. Part. Nuclei Lett. 2016. Vol. 13, No. 5. P. 672–675. DOI: 10.1134/S1547477116050071.
6. Киселев М.А., Земляная Е.В., Жабицкая Е.И., Башашин М.В., Иваньков А.И. Исследование возможностей анализа везикулярной структуры нанолечарств на основе ФТНС по данным малоуглового рассеяния нейтронов // Поверхность. Рентгеновские, синхротронные и нейтронные исследования. 2023. Т. 17, № 1. С. 3–8. DOI: 10.31857/S1028096023010119.
7. Betsou I., Linev S., Bellenot B., Couet O. New ROOT graphics language // Journal of Physics: Conference Series. 2020. Vol. 1525. P. 012061. DOI: 10.1088/1742-6596/1525/1/012061.
8. Соловьев А.Г., Соловьева Т.М., Лукьянов К.В. Развертывание веб-приложения для фитирования экспериментальных данных в облачной инфраструктуре ОИЯИ // Физика элементарных частиц и атомного ядра. 2024. Т. 55, № 3. С. 1–5.
9. James F., Roos M. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations // Comput. Phys. Commun. 1975. Vol. 10. P. 343–367.
10. James F.W., Winkler M. MINUIT User's Guide. URL: <https://api.semanticscholar.org/CorpusID:121865778> (дата обращения 15.01.2024).
11. Valls X., Moneta L., Amadio G., Tsang A. New developments in the ROOT fitting classes // EPJ Web of Conferences. 2019. Vol. 214. P. 05043 DOI: 10.1051/epjconf/201921405043.
12. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters // OSDI'04: Sixth Symposium on Operating System Design and Implementation. San Francisco. CA. 2004. P. 137–150.
13. Kiselev M.A., Zemlyanaya E.V., Aswal V.K., Neubert R. What can we learn about the lipid vesicle structure from the small-angle neutron scattering experiment? Investigation DMPC vesicle structure by small angle neutron scattering // European Biophysics Journal. 2006. Vol. 35, № 6. P. 477–493.