

ПРОЦЕССОРЫ ЭЛЬБРУС ДЛЯ НАУЧНЫХ ВЫЧИСЛЕНИЙ**ELBRUS PROCESSORS FOR SCIENTIFIC COMPUTING**

Зубанков А.А., разработчик, ООО «Научно-исследовательский центр цифровых технологий», Москва, Россия

Zubankov A.A., Developer, LTD «Scientific research center of digital technologies», Moscow, Russia

Аннотация. Современная наука невозможна без компьютерных вычислений – от обработки результатов измерений до моделирования процессов. Во всех передовых исследовательских центрах для этих задач служат вычислительные кластеры, в основном, построенные на базе процессоров архитектуры x86. Программное обеспечение изначально разрабатывалось для популярных архитектур и для работы на отечественных процессорах Эльбрус требует портирования. В работе будут представлены портированные на архитектуру Эльбрус программы, позволяющие проводить анализ и обработку физических данных и моделирование.

Ключевые слова: программное обеспечение, обработка физических данных, процессоры Эльбрус.

Abstract. Modern science is impossible without computer calculations - from processing measurement results to modeling processes. All advanced research centers use computing clusters for these tasks, mostly based on x86 processors. The software was originally developed for popular architectures and requires porting to work on domestic Elbrus processors. In the paper the programs ported to the Elbrus architecture, which allow to perform physical data analysis, processing and modeling were presented.

Keywords: software, physical data processing, Elbrus processors.

Отечественные процессоры Эльбрус разрабатываются компанией МЦСТ [1]. Их отличительной особенностью является собственная архитектура e2k с очень длинной машинной командой VLIW. Это не позволяет процессорам Эльбрус без портирования нативно запускать большую часть программного обеспечения, в том числе и для анализа физических данных. Современная физика не может обойтись без сложных и многомерных вычислений, что особенно заметно в физике высоких энергий. Ведущие международные исследовательские центры – Европейский центр ядерных исследований CERN, Центр по исследованию ионов и антипротонов FAIR, Объединенный институт ядерных исследований ОИЯИ, Брукхейвенская национальная лаборатория BNL и другие, для анализа физических данных используют вычислительные кластеры, построенные в основном на базе x86 процессоров. Портирование ПО на российские процессоры Эльбрус сделает возможным импортозамещение даже в таких узких и требовательных областях, связанных с фундаментальными исследованиями.

Своя собственная архитектура e2k дает широкие возможности для оптимизации кода, в отличие от x86 [2]. Оптимизирующий компилятор lcc позволяет использовать особенности

процессора для ускорения работы кода на C/C++. Кроме того, процессоры Эльбрус обладают Режимом Безопасных Вычислений, обеспечивающим дополнительный контроль поведения программ на C/C++.

Портирование ПО для научных расчетов выполнялось в рамках операционной системы Collaboration OS (CollabOS) на базе Linux. Ключевой особенностью CollabOS является поддержка пакетной базы на основе CentOS 9 и Epeel, что обеспечивает легкий переход на процессоры Эльбрус с других дистрибутивов Linux, основанных на RedHat: CentOS, Fedora и Scientific Linux. Поддержка языков C, C++, Fortran, Python 3.9, Perl, Java 8 и 11, PHP, Lua, Ruby, R, Rust позволяет осуществлять привычную работу с кодом. На данный момент операционная система поддерживает процессоры 5 и 6 поколений Эльбрус 8CB, 2C3, 16C.

Одним из наиболее широкоиспользуемых фреймворков для обработки данных в физике высоких энергий, ядерной физике, физике элементарных частиц и космологии является CERN ROOT [3]. Он обладает широким инструментарием статистического анализа, линейной алгебры, средств фитирования, средств четырехвекторных вычислений, а также инструментами многовариантного анализа данных, то есть использования нейронных сетей. ROOT позволяет проводить статистически обоснованный научный анализ и визуализацию больших объемов данных.

Объектно-ориентированная среда для анализа данных ROOT имеет собственный встроенный интерактивный интерпретатор C/C++ – cling, построенный на основе инфраструктуры LLVM и Clang. Для портирования cling использовался LLVM-13, модифицированный МЦСТ [4]. Портированный LLVM использует проприетарный оптимизирующий e2k-бэкенд из компилятора lcc.

Дополнительным преимуществом актуальной версии ROOT является интеграция языков Python и R, а также поддержка БД MySQL, PostgreSQL и SQLite. Портированный ROOT позволяет производить сложные вычисления и построения, что было протестировано на основе данных с эксперимента BM@N на NICA [5] на рисунке 1.

Также было выполнено портирование генератора событий физики высоких энергий Pythia8. Он позволяет получать из теории и моделей детальное описание жестких и мягких взаимодействий, распределения партонов, ливни партонов в начальном и конечном состояниях, множественные взаимодействия, фрагментацию и распад. Выходные файлы могут быть использованы в ROOT для сравнения экспериментальных данных с моделированием, что необходимо в современных реалиях физики высоких энергий.

Другим мощным инструментом, доступным на e2k и обладающим широким спектром статистических и графических методов, является язык R. Он позволяет проводить статистические тесты, линейное и нелинейное моделирование, анализ временных рядов, классификацию и кластеризацию. Также, как и ROOT, он обладает режимом интерпретатора, что существенно упрощает работу с данными. Для решения ресурсоемких задач реализована возможность подключать и вызывать код на языках C, C++ и Fortran во время выполнения.

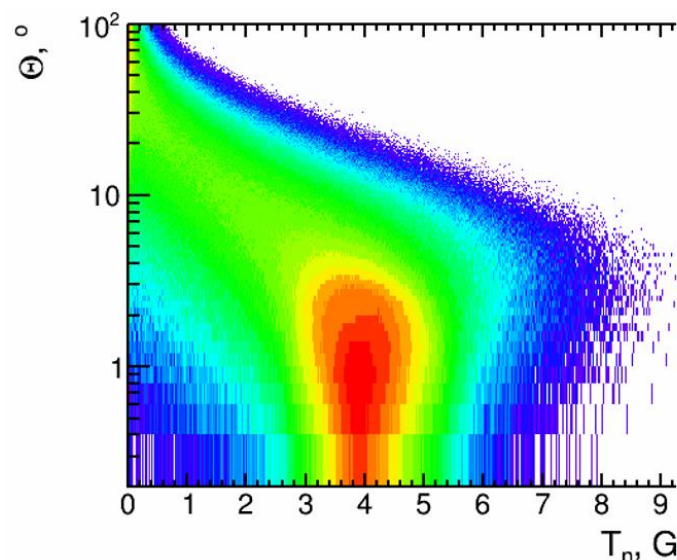


Рисунок 1 – Двумерное распределение выходов нейтронов по кинетическим энергиям и углу вылета в результате столкновения пучка Хе с энергией 3,8 ГэВ/нуклон с мишенью CsI в эксперименте BM@N из модели UrQMD-AMC, построенное в портированном на e2k ROOT

Один из наиболее популярных языков программирования Python также позволяет проводить анализ данных благодаря библиотекам. NumPy дает возможность проводить быстрые числовые и матричные вычисления, а инструментарий SciPy и Matplotlib – выполнять построения графиков, гистограмм и проводить статистический анализ и аппроксимацию. Как отмечалось ранее, Python также может взаимодействовать с ROOT через модуль PyRoot.

Другим популярным мультиплатформенным инструментом для визуализации данных и математических функций является Gnuplot. Его можно использовать как в интерактивном режиме в графическом режиме или в командной строке, так и в неинтерактивных приложениях, таких как веб-скриптинг. Высокоуровневый язык для численных вычислений GNU Octave использует для построения графиков Gnuplot. Octave обладает большим математическим инструментарием, совместим со многими скриптами Matlab и легко расширяется пользовательскими функциями, написанными на Octave, C, C++, Fortran и других языках.

Несмотря на существенное отличие архитектуры e2k, в рамках операционной системы CollabOS был портирован ряд ПО, востребованного в научно-исследовательских организациях, в частности в ядерной физике, физике высоких энергий, физике элементарных частиц и космологии. Портированы: программы для обработки данных, статистики и моделирования – ROOT, R, SciPy, NumPy, Pythia8, LHAPDF, Octave; для построения графиков, гистограмм и диаграмм – ROOT, R, Matplotlib, Gnuplot; для верстки текста статей – TeXLive, AbiWord; языки программирования – C, C++, Fortran, Python 3.9, Perl, Java 8 и 11, PHP, Lua, Ruby, R, Rust, Octave. Благодаря этому достигается широкая совместимость проводимых научных расчетов на отечественных процессорах Эльбрус 8СВ и 2С3 с другими архитектурами.

СПИСОК ЛИТЕРАТУРЫ

1. Архитектура Эльбрус // АО «МЦСТ». – URL: http://www.mcst.ru/elbrus_arch (дата обращения: 26.03.2025).
2. Нейман-заде, М. И. Руководство по эффективному программированию на платформе «Эльбрус» / М. И. Нейман-заде, С. Д. Королёв // АО «МЦСТ». – 2024. – Выпуск 1.2. – URL: http://mcst.ru/doc/elbrus_prog/elbrus-prog-1.2_2024-02-28.pdf (дата обращения: 26.03.2025).
3. Brun, R. ROOT – An Object Oriented Data Analysis Framework / R. Brun, F. Rademakers // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. – 1997. – Vol. 389, Iss. 1-2. – P. 81-86. – DOI: 10.1016/S0168-9002(97)00048-X.
4. Проект по портированию LLVM на e2k-платформу // GitFlic. – URL: <https://gitflic.ru/project/e2khome/llvm-project-e-2-k> (дата обращения: 26.03.2025).
5. Zubankov, A. Performance study of the Highly Granular Neutron Detector prototype in the BM@N experiment / A. Zubankov [et al.] // arXiv:2503.12624 [physics.ins-det]. – URL: <https://arxiv.org/abs/2503.12624> (дата обращения: 26.03.2025).

REFERENCES

1. Architecture Elbrus // JSC «MCST». – URL: http://www.mcst.ru/elbrus_arch (date of reference: 26.03.2025).
2. Neyman-zade, M. I. Guide to efficient programming on the Elbrus platform / M. I. Neyman-zade, S. D. Korolev // JSC «MCST». – 2024. – Iss. 1.2. – URL: http://mcst.ru/doc/elbrus_prog/elbrus-prog-1.2_2024-02-28.pdf (date of reference: 26.03.2025).
3. Brun, R. ROOT – An Object Oriented Data Analysis Framework / R. Brun, F. Rademakers // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. – 1997. – Vol. 389, Iss. 1-2. – P. 81-86. – DOI: 10.1016/S0168-9002(97)00048-X.
4. Project on porting LLVM to e2k platform // GitFlic. – URL: <https://gitflic.ru/project/e2khome/llvm-project-e-2-k> (date of reference: 26.03.2025).
5. Zubankov, A. Performance study of the Highly Granular Neutron Detector prototype in the BM@N experiment / A. Zubankov [et al.] // arXiv:2503.12624 [physics.ins-det]. – URL: <https://arxiv.org/abs/2503.12624> (date of reference: 26.03.2025).