

АЛГОРИТМИЧЕСКИЕ И АРХИТЕКТУРНЫЕ МЕТОДЫ ОПТИМИЗАЦИИ ЭНЕРГОПОТРЕБЛЕНИЯ В ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ И ОБЛАЧНЫХ ВЫЧИСЛЕНИЯХ

А.М. Тюнина, В.В. Кондусова, Д.С. Кукуева, Н.Ю. Заленская

ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

В статье проводится комплексный анализ методов оптимизации энергопотребления в средах высокопроизводительных вычислений (НРС) и облачных платформ. Исследование охватывает многоуровневый подход к энергоэффективности, включая разработку энергооптимизированных алгоритмов, применение политик динамического масштабирования в Kubernetes, использование низкоуровневых механизмов управления питанием процессора и внедрение специализированных метрик для оценки энергетической эффективности. Представлен детальный анализ энергопотребления различных типов процессорных операций и методов доступа к памяти. На основе формул управления мощностью CMOS-схем и сравнительных таблиц энергопотребления операций демонстрируются ключевые факторы влияния на энергоэффективность вычислительных систем. Результаты исследования показывают, что комплексное применение рассмотренных методов позволяет достичь значительного снижения энергопотребления при сохранении требуемого уровня производительности.

Ключевые слова: энергоэффективность, НРС, облачные вычисления, зеленые вычисления, оптимизация энергопотребления, энергоэффективные алгоритмы, динамическое масштабирование, Kubernetes, управление питанием процессора, метрики энергоэффективности.

ALGORITHMIC AND ARCHITECTURAL METHODS FOR OPTIMIZING ENERGY CONSUMPTION IN HIGH-PERFORMANCE AND CLOUD COMPUTING

A.M. Tyunina, V.V. Kondusova, D.S. Kukueva, N.Y. Zalenskaya

The article provides a comprehensive analysis of energy consumption optimization methods in high-performance computing (HPC) and cloud platforms. The research covers a multi-level approach to energy efficiency, including the development of energy-optimized algorithms, the application of dynamic scaling policies in Kubernetes, the use of low-level processor power management mechanisms, and the implementation of specialized metrics to assess energy efficiency. A detailed analysis of the energy consumption of various types of processor operations and memory access methods is presented. Based on the power management formulas of CMOS circuits and comparative tables of energy consumption of operations, the key factors influencing the energy efficiency of computing systems are demonstrated. The results of the study show that the integrated application of the considered methods makes it possible to achieve a significant reduction in energy consumption while maintaining the required level of productivity.

Keywords: energy efficiency, HPC, cloud computing, green computing, energy consumption optimization, energy efficient algorithms, dynamic scaling, Kubernetes, processor power management, energy efficiency metrics.

Современные центры обработки данных и высокопроизводительные вычислительные системы потребляют гигаватты электроэнергии, что создает серьезные экономические и экологические challenges. Рост вычислительных мощностей в областях HPC и облачных технологий сопровождается экспоненциальным увеличением энергопотребления, превращая проблему энергоэффективности из второстепенной в критически важную для устойчивого развития ИТ-инфраструктуры. Традиционные подходы к проектированию вычислительных систем, ориентированные исключительно на максимальную производительность, демонстрируют ограниченную эффективность в условиях растущих требований к энергосбережению и экологической устойчивости. Проблема усугубляется разнородностью рабочих нагрузок в современных дата-центрах и отсутствием унифицированных методов оценки и оптимизации энергопотребления на различных уровнях вычислительного стека. В связи с этим возникает необходимость в разработке комплексных подходов, объединяющих алгоритмические оптимизации, архитектурные решения и системные механизмы управления энергопотреблением. Особую актуальность приобретает исследование взаимосвязей между вычислительной сложностью алгоритмов, архитектурными особенностями процессоров и политиками оркестрации контейнеров в контексте минимизации энергетических затрат.

Главный принцип на уровне приложения — выполнять больше полезной работы на каждый джоуль потребленной энергии. Это достигается не только за счет оптимизации кода, но и за счет выбора или создания специализированных алгоритмов.

Алгоритм с меньшей асимптотической сложностью (например, $O(n \log n)$ вместо $O(n^2)$) в большинстве случаев потребляет меньше энергии для больших объемов данных, так как выполняет меньше вычислительных инструкций.

Различные процессорные инструкции потребляют разное количество энергии. Операции с плавающей точкой, особенно двойной точности, как правило, более энергозатратны, чем целочисленные операции. Векторизация (использование SIMD-инструкций) позволяет выполнить одну энергоемкую инструкцию для нескольких данных, повышая общую эффективность.

Важна также локализация данных. Алгоритмы, оптимизированные для кэширования, минимизируют доступ к оперативной памяти, которая потребляет значительно больше энергии, чем кэш-память процессора.

Таблица 1 – Сравнения операций и их энергопотребления

Тип операции	Относительное энергопотребление
Целочисленная операция	1
Операция с плавающей точкой	3-5
Доступ к кэшу L1	~1
Доступ к кэшу L2	~5-10
Доступ к оперативной памяти	~100-200
Доступ к диску	~1500-10000

Уменьшение частоты снижает производительность, но приводит к нелинейному снижению энергопотребления. Мощность, потребляемая CMOS-схемой, приближенно описывается формулой:

$$P = C * V^2 * f + P_{static}(1)$$

где P — общая потребляемая мощность, C — емкость нагрузки, V — напряжение, f — частота, $P_{statistic}$ — статическая мощность (утечки).

Как видно из формулы, снижение напряжения дает наибольший выигрыш.

Также нужно определить состояния простоя (C-states). Когда ядро процессора бездействует, оно может переходить в состояния с пониженным энергопотреблением (C1, C2, C6 и т.д.). Чем "глубже" состояние, тем меньше энергии потребляется, но тем больше времени требуется на выход из него. Баланс между экономией энергии и производительностью критически важен для энергозатратных приложений.

Чтобы управлять энергоэффективностью, ее необходимо измерять. Наряду с классическими метриками производительности (время выполнения, FLOPS), необходимы специализированные энергетические метрики.

FLOPS — это метрика производительности вычислительных систем, показывающая количество операций с плавающей запятой выполняемых за одну секунду. Данная метрика измеряет операции с числами в формате с плавающей запятой. Располагает основными арифметическими операциями: сложением, вычитанием, умножением, делением. Активно применяется в научных расчетах, инженерных симуляциях, машинном обучении.

Таблица 2 – Энергетические метрики

Метрика	Уровень	Формула	Цель
Performance per Watt	Система, приложение	$FLOPS \text{ Вт}$	Максимизация
Energy per Task	Приложение	$\int P(t)dt$	Минимизация
PUE	Дата-центр	$\frac{\text{Общая энергия}}{\text{Энергия процесса}}$	Минимизация к 1.0
Energy-Delay Product	Система	Энергия * Время	Минимизация

Оптимизация энергопотребления в НРС и облачных средах — это многоуровневая задача, требующая синергии между алгоритмами, системным

программным обеспечением и аппаратным обеспечением. Разработка энергоэффективных алгоритмов, интеллектуальное автоматическое масштабирование в облаках, грамотное использование низкоуровневых механизмов процессора и постоянный мониторинг с помощью адекватных метрик — все это вместе позволяет значительно снизить энергетические затраты, делая высоконагруженные вычисления более устойчивыми и экономически выгодными. Будущее за адаптивными системами, которые в реальном времени балансируют между производительностью, временем выполнения и энергопотреблением, выбирая оптимальный режим работы для каждой конкретной задачи.

Перспективы развития методов оптимизации энергопотребления в НРС и облачных средах связаны с созданием интеллектуальных адаптивных систем, способных в реальном времени балансировать между производительностью, временем выполнения и энергозатратами. Особый потенциал демонстрируют подходы, основанные на машинном обучении для прогнозирования оптимальных режимов работы вычислительных узлов в зависимости от характеристик рабочих нагрузок. Дальнейшие исследования должны быть направлены на разработку энергопрозрачных алгоритмов, позволяющих оценивать и оптимизировать энергопотребление на этапе проектирования вычислительных методов. Важным направлением является также создание стандартизированных интерфейсов для мониторинга и управления энергопотреблением across stack - от уровня приложений до аппаратного обеспечения. Учитывая растущую популярность гетерогенных вычислительных систем, особую актуальность приобретают исследования в области оптимизации распределения задач между различными типами ускорителей (CPU, GPU, FPGA) с учетом их энергетической эффективности. Развитие квантовых и нейроморфных вычислений открывает новые возможности для принципиально новых подходов к энергоэффективным вычислениям, требующих пересмотра традиционных парадигм проектирования вычислительных систем.

Список литературы

1. Антипко, А. В. Облачные вычисления. Модели развертывания систем облачных вычислений / А. В. Антипко // Молодой ученый. – 2023. – № 6(453). – С. 9-10. – EDN PHYRFQ.

2. Алгоритм вычисления преобразования Лапласа-Стилтьеса для времени отклика системы облачных вычислений с гистерезисным управлением / К. Е. Самуйлов, Ю. В. Гайдамака, Э. С. Сопин, А. В. Горбунова // Современные информационные технологии и ИТ-образование. – 2015. – Т. 11, № 2. – С. 172-177. – EDN WAQEZR.

3. Черепенин, В. А. Интеграция и оптимизация систем облачных, туманных и граничных вычислений: моделирование, задержки и алгоритмы / В. А. Черепенин, С. П. Воробьев // Известия высших учебных заведений. Северо-Кавказский регион. Технические науки. – 2024. – № 3(223). – С. 19-25. – DOI 10.17213/1560-3644-2024-3-19-25. – EDN LKVVKM.

References

1. Antipko, A.V. Cloud computing. Models of deployment of cloud computing systems / A.V. Antipko // Young Scientist. – 2023. – № 6(453). – Pp. 9-10. – EDN PHYRFQ.

2. The algorithm for calculating the Laplace-Stieltjes transform for the response time of a cloud computing system with hysteresis control / K. E. Samuilov, Yu. V. Gaidamaka, E. S. Sopin, A.V. Gorbunova // Modern information technologies and IT education. – 2015. – Vol. 11, No. 2. – pp. 172-177. – EDN WAQEZR.

3. Cherepenin, V. A. Integration and optimization of cloud, fog and boundary computing systems: modeling, delays and algorithms / V. A. Cherepenin, S. P. Vorobyov // Izvestiya vysshikh uchebnykh zavedeniy. The North Caucasus region. Technical sciences. – 2024. – № 3(223). – Pp. 19-25. – DOI 10.17213/1560-3644-2024-3-19-25. – EDN LKVVKM.