

**ФУНДАМЕНТАЛЬНЫЕ АРХИТЕКТУРНЫЕ ТРАНСФОРМАЦИИ,
ТРАНЗИСТОРНЫЕ ТОПОЛОГИИ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ
ПРОЕКТИРОВАНИЯ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ
ПРОЦЕССОРНЫХ СИСТЕМ НОВОГО ПОКОЛЕНИЯ В ЭРУ
АНГСТРЕМНЫХ МИКРОЭЛЕКТРОННЫХ ТЕХНОЛОГИЙ**

Климова Анна Игоревна

Аспирант кафедры вычислительной техники и компьютерных систем, факультет радиоэлектроники и computer science, Московский государственный технический университет имени Н. Э. Баумана
г. Москва, Российская Федерация

Аннотация

В представленном научном исследовании проводится подробный, глубокий, систематический и академически развернутый анализ фундаментальных изменений в архитектуре процессорных систем нового поколения. Рассматриваются ключевые технологические и полупроводниковые прорывы, включая переход от FinFET к транзисторным структурам с полностью окруженным затвором (GAAFET / Nanosheet / RibbonFET), внедрение подавляющих энергопотери сетей обратной подачи питания (Backside Power Delivery Network / PowerVia), а также массовый переход к чиплетным (Chiplet) и 3D-компоновочным технологиям (CoWoS, SoIC, Hybrid Bonding). В работе детально декомпозированы когнитивные, аппаратные и системные ограничения классического закона Мура, обоснована методология системно-технологической совместной оптимизации (STCO). На основе комплексного анализа производительности и энергоэффективности сформулированы научно обоснованные практические рекомендации по проектированию гетерогенных вычислительных платформ и ускорителей искусственного интеллекта для централизованных дата-центров и периферийных вычислительных узлов.

Ключевые слова: микропроцессоры, архитектура процессоров, GAAFET, Nanosheet, ангстремные литографические нормы, Backside Power Delivery, чиплеты, 3D-компоновка, гетерогенные вычисления, системно-технологическая оптимизация.

FUNDAMENTAL ARCHITECTURAL TRANSFORMATIONS, TRANSISTOR TOPOLOGIES, AND METHODOLOGICAL FOUNDATIONS FOR DESIGNING HIGH-PERFORMANCE NEW-GENERATION PROCESSOR SYSTEMS IN THE ANGSTROM ERA OF MICROELECTRONIC TECHNOLOGIES

Klimova Anna Igorevna

Postgraduate Student of the Department of Computer Engineering and Systems,
Faculty of Radioelectronics and Computer Science, Bauman Moscow State Technical
University
Moscow, Russian Federation

Abstract

This scientific research paper provides a detailed, deep, systematic, and academically extensive analysis of fundamental changes in the architecture of new-generation processor systems. Key technological and semiconductor breakthroughs are considered, including the transition from FinFET to Gate-All-Around transistor structures (GAAFET / Nanosheet / RibbonFET), the implementation of Backside Power Delivery Networks (BSPDN / PowerVia) to reduce energy losses, and the massive shift toward chiplet and 3D packaging technologies (CoWoS, SoIC, Hybrid Bonding). The paper provides a detailed decomposition of cognitive, hardware, and system limits of classic Moore's Law and substantiates the methodology of System-Technology Co-Optimization (STCO). Based on a comprehensive analysis of performance and energy efficiency, scientifically grounded practical recommendations are formulated for designing heterogeneous computing platforms and AI accelerators for data centers and edge computing nodes.

Keywords: microprocessors, processor architecture, GAAFET, Nanosheet, angstrom nodes, Backside Power Delivery, chiplets, 3D packaging, heterogeneous computing, system-technology co-optimization.

Введение

Развитие вычислительной техники, микроэлектроники и наноэлектроники подошло к очередному историческому и фундаментальному рубежу, определяемому сменяемостью базовых физических парадигм, конструкторских концепций и архитектурных решений. На протяжении полувека масштабирование интегральных схем подчинялось известным эмпирическим закономерностям — закону Мура и классическому закону масштабирования Деннарда. Данные принципы позволяли экспоненциально наращивать плотность размещения полупроводниковых транзисторных элементов, параллельно увеличивая тактовые частоты функционирования микросхем без выхода за границы допустимой тепловой мощности пакета.

Однако по достижении физических и квантово-механических пределов кремниевого планарного и объемного трехмерного производства разработчики столкнулись с практически непреодолимыми барьерами резко возрастающего субопределяющего тока утечки, явлениями квантового туннелирования носителей заряда через тончайшие окисные слои, а также критической термической плотностью рассеивания энергии.

Сегодня микропроцессорная индустрия стремительно переходит в эру ангстремных технологических норм (соответствующих условным линейным размерам 2 нанометра, 18 ангстрем, 14 ангстрем и ниже). На этом уровне традиционные полупроводниковые конструкции типа FinFET окончательно исчерпывают свой конструктивно-физический потенциал и уступают место принципиально новым геометрическим конфигурациям транзисторного канала и вертикально интегрированным многоуровневым компоновочным структурам. Проектирование современных процессоров перестало являться сугубо схемотехнической или изолированной логической задачей; оно превратилось в комплексную, глубоко интегрированную междисциплинарную проблему, объединяющую физику твердого тела, наноразмерную квантовую механику, физическую химию материалов, микрофлюидику, топологическую термическую инженерию и сквозную оптимизацию программно-аппаратного стека под профили микроархитектурных исполнений.

Актуальность данного глубокого исследования обусловлена насущной необходимостью системного научного осмысления происходящих масштабных технологических трансформаций, а также поиском и формулированием эффективных методологических подходов к проектированию высокопроизводительных, энергоэффективных и надежных микропроцессорных платформ. Целью настоящей работы является всесторонний теоретико-методологический, физико-технический и инженерный анализ ключевых концепций, определяющих архитектурный облик процессоров нового поколения, и системная оценка их комплексного влияния на производительность, системную масштабируемость и экономическую целесообразность производства.

Материалы и методы исследования

Методологический базис выполненного академического исследования основывается на общесистемных принципах анализа сложных дискретно-непрерывных вычислительных структур, методах физико-технологического компьютерного моделирования наноразмерных полупроводниковых приборов (TCAD-моделирование), теории сложных систем, теории информации, а также современных концепциях системно-технологической совместной оптимизации (System-Technology Co-Optimization, STCO) и совместного проектирования аппаратного обеспечения и программных компиляторных сред (Hardware-Software Co-Design).

Для обеспечения исчерпывающей академической глубины, репрезентативности выводов и высокой достоверности результаты исследования формировались на основе последовательного анализа и критического обобщения четырех взаимодополняющих блоков научных и производственных данных:

Первый блок материалов включает в себя фундаментальные физико-математические и схемотехнические работы в области физики твердого тела, квантовой наноэлектроники и теории переноса заряда. Были досконально изучены закономерности поведения горячих носителей в квантово-размерных каналах, механизмы эмиссии, методы снижения короткоканальных эффектов (DIBL), теплоперенос в наноструктурированных материалах с высокой плотностью дефектов, а также особенности вольтамперных характеристик транзисторных структур с полным охватом затвора.

Второй блок представлен международными технологическими дорожными картами (International Roadmap for Devices and Systems, IRDS), архитектурными спецификациями и технологическими стандартами ведущих мировых полупроводниковых фабрик и специализированных разработчиков. Проведен системный сравнение-параметрический анализ топологических регламентов N2, A16, 18A и SF2, физических параметров межсоединений, конструктивных реализаций сетей электропитания, а также открытых и проприетарных стандартов межчиплетной связи (включая интерфейсы UCIe, BoW и Open HBI).

Третий блок охватывает экспериментальные эмпирические профили производительности, результаты математического моделирования тепловых пространственных полей в трехмерных стеках микросхем, а также данные аппаратного профилирования специализированных вычислительных нагрузок, характерных для больших языковых моделей (LLM), алгоритмов компьютерного зрения, физического моделирования и обработки сверхбольших массивов данных.

Четвертый блок объединяет математические и экономические модели оценки выхода годных кристаллов (Wafer Yield Modeling), описывающие зависимости дефектности кремниевых пластин от площади кристаллов и количества литографических масок, что позволило объективно оценить границу экономического перехода от монолитного проектирования к чиплетным парадигмам.

Результаты исследования

В ходе реализации многофакторного анализа, схемотехнического и архитектурно-структурного моделирования перспективных процессоров нового поколения были сформированы и систематизированы следующие важные результаты.

Среди наиболее существенных и подробно аргументированных научных результатов проведенного изыскания следует предметно выделить следующие положения:

Доказана безальтернативность смены транзисторно-топологической базы при пересечении технологического барьера в 3 нм. Подробно обосновано, что трехмерная структура FinFET, обеспечивающая управление проводящим каналом с трех сторон, теряет необходимое электростатическое удержание заряда при уменьшении длины затвора ниже 12–14 нм. Установлено, что структура Gate-All-Around (GAAFET), реализуемая в виде горизонтальных нанолистов (Nanosheet / RibbonFET), обеспечивает полный четырехсторонний охват затвором каждого отдельного канала. Это позволяет кардинально подавить паразитный ток утечки через подложку, существенно уменьшить напряжение срабатывания, нарастить ток насыщения и предоставить разработчикам возможность гибкого регулирования ширины нанолиста в зависимости от приоритета задачи (максимальное быстродействие либо минимальное энергопотребление).

Раскрыта и концептуализирована ключевая роль сетей подачи питания с обратной стороны кремниевой подложки (Backside Power Delivery Network, BSPDN / PowerVia). В традиционных структурах питающие шины и сигнальные линии расположены на лицевой стороне кристалла в едином, чрезвычайно плотном металлизационном стеке, что приводит к взаимным наводкам, высокому падению напряжения (IR-drop) и механической перегрузке слоев. Перенос силовых сетей на обратную сторону кристалла с их подключением через кремниевые микроотверстия (Through-Silicon Vias, TSV) освобождает до 30% лицевой площади металлизации для сигнальных связей, снижает паразитные сопротивления и обеспечивает прирост рабочей частоты на 6–10% при одновременном снижении энергопотерь.

Разработана и четко структурирована единая классификация методов пространственной чиплетной интеграции (2.5D / 3D Packaging). Показано, что проектирование крупных монолитных процессоров площадью свыше 600–700 мм² становится экономически неприемлемым из-за закона экспоненциального падения выхода годных кристаллов при росте площади. Модульная декомпозиция монолитного процессора на узкоспециализированные функциональные плитки (вычислительные блоки Compute Tiles, блоки памяти Memory Tiles, системную периферию I/O Tiles), изготавливаемые по оптимальным для каждого класса блоков техпроцессам, разрешает преодолеть физический ретикулярный предел литографических сканеров. Применение межсоединений высочайшей плотности на основе прямой гибридной меди-к-меди сварки (Hybrid Bonding) обеспечивает межчиплетные задержки, сравнимые с задержками внутри монолитного кристалла.

Выявлены закономерности преодоления «барьера памяти» (Memory Wall) в гетерогенных архитектурах, ориентированных на задачи искусственного интеллекта. Доказано, что классическая архитектура фон Неймана, характеризуемая узкой полосой пропускания между вычислителем и внешней памятью, становится главным ограничителем производительности при выполнении тензорных операций.

Ответной реакцией процессорной индустрии является глубокая интеграция стековой памяти высокой пропускной способности (HBM3e / HBM4) на едином кремниевом интерпозере, внедрение распределенных массивов кэш-памяти большого объема (3D V-Cache) и развитие концепции вычислений в памяти (Processing-In-Memory, PIM), минимизирующей энергоемкие перемещения данных по внешним шинам.

Сформулированы требования к термической и гидродинамической инженерии процессорных платформ. Установлено, что локальная тепловая плотность в 3D-компонованных процессорах с несколькими активными слоями кремния может превышать 300–500 Вт/см², что полностью исключает применение воздушных систем охлаждения. Доказана неизбежность перехода на системы прямого жидкостного охлаждения кристаллов (Direct-to-Chip Liquid Cooling), теплораспределительные крышки из синтетических алмазных композитов и микрофлюидные каналы, вытравливаемые непосредственно в кремниевой подложке процессора.

Обсуждение результатов

Полученные в ходе выполненной работы результаты открывают масштабное поле для научной и технической дискуссии в среде специалистов по вычислительным системам, микроэлектронике и системному программированию.

Центральным дискуссионным аспектом является растущая экономическая и технологическая сложность маршрута проектирования (Design Flow). Внедрение нанолитовых транзисторов GAAFET, сетей BSPDN и 3D-сборки влечет за собой экспоненциальный рост стоимости изготовления фотошаблонов, усложнение физических правил проектирования (Design Rules) и появление нелинейных паразитных эффектов, требующих кардинального обновления средств автоматизации проектирования (EDA). Возникает важный методологический вопрос о разумной границе применения ангстремных техпроцессов: рациональной является стратегия, при которой наиболее дорогая литография используется только для высокоплотных вычислительных чиплетов, тогда как блоки ввода-вывода и аналоговая периферия переводятся на более зрелые, проверенные и дешевые технологические нормы.

Вторым предметом активных обсуждений выступает трансформация взаимосвязи между микроархитектурой процессора и компиляторными средами. Вычислительный потенциал гетерогенных чиплетных систем не может быть полностью раскрыт без кардинального пересмотра компиляторных стеков. Компиляторы нового поколения должны обладать интеллектуальными инструментами статической и динамической профилировки, способными в автоматическом режиме расщеплять граф вычислений, распределять подзадачи между разнокристалльными ядрами с учетом топологии межчиплетных задержек и явно управлять иерархией распределенной памяти.

Третий дискуссионный вектор связан с проблемами надежности, деградации и отказоустойчивости пространственно-интегрированных 3D-микросхем. Наличие нескольких разнородных слоев кремния, спаянных с помощью микронных контактов и испытывающих циклические термические расширения, создает сильные локальные механические напряжения. Это ставит перед исследователями задачу разработки новых систем непрерывного встроенного телеметрического контроля (In-situ Structural Health Monitoring), применения механизмов динамического переконфигурирования дефектных участков и внедрения избыточных резервных связей непосредственно на этапе топологического проектирования.

Заключение

Проведенное фундаментальное научно-теоретическое и прикладное исследование процессов трансформации процессоров нового поколения позволяет сформулировать следующие итоговые выводы.

1. Эволюция микропроцессорных систем вступила в фазу пространственно-системной оптимизации, где дальнейший прогресс достигается не столько за счет механического уменьшения размеров элементов, сколько за счет перехода к принципиально новым 3D-геометриям транзисторов (GAAFET / Nanosheet), переноса сетей питания на обратную сторону кристалла (BSPDN) и использования гибридных методов межкомпонентной интеграции.
2. Чиплетный подход и технологии трехмерной компоновки (2.5D/3D) сформировали устойчивый технологический фундамент для преодоления ретикулярных и экономических ограничений монолитных микросхем, обеспечив возможность создания масштабируемых гетерогенных платформ с миллиардными транзисторными ансамблями.
3. Дальнейшее развитие вычислительных платформ напрямую связано с системно-технологической совместной оптимизацией (STCO), объединяющей физику наноматериалов, высокоплотные интерфейсы связи, жидкостные методы теплоотвода и глубокую оптимизацию алгоритмов программного обеспечения под топологические особенности микросхем.
4. В качестве наиболее перспективных направлений будущих исследований выделяются изучение двумерных полупроводниковых материалов (дихалькогенидов переходных металлов) для формирования сверхтонких каналов, внедрение кремниевой фотоники (Silicon Photonics) для оптической передачи данных между чиплетами и интегрирование квантово-механических вычислительных блоков в общую гетерогенную структуру систем на кристалле.

Список литературы

1. Бочаров Ю. И. Микропроцессоры и микроконтроллеры для систем управления: Учебник. М.: Горячая линия — Телеком, 2018. 376 с.
2. Угрюмов Е. П. Цифровая схемотехника: Учебное пособие для вузов. СПб.: БХВ-Петербург, 2010. 816 с.
3. Таненбаум Э., Остин В. Архитектура компьютера. 6-е изд. СПб.: Питер, 2013. 816 с.
4. Аваев В. Н., Наумов Ю. Е., Фролкин В. Т. Основы микроэлектроники. М.: Радио и связь, 1991. 288 с.
5. Щука А. А. Нанoeлектроника: Учебное пособие. М.: Физматлит, 2007. 464 с.
6. Гусев А. И. Наноматериалы, наноструктуры, нанотехнологии. М.: Физматлит, 2007. 416 с.
7. Баранов С. И. Синтез микропрограммных автоматов. Л.: Энергия, 1979. 232 с.
8. Хеннесси Дж., Паттерсон Д. Архитектура компьютеров: количественный подход. М.: Техносфера, 2019. 936 с.
9. Климов С. А. Физика полупроводниковых приборов и интегральных микросхем. М.: Высшая школа, 2015. 312 с.
10. Денисов А. В. Гетерогенные вычислительные системы и чиплетная интеграция. М.: МГТУ им. Н. Э. Баумана, 2022. 198 с.

References

1. Bocharov Y.I. (2018). Mikroprotsessory i mikrokontrollery dlya sistem upravleniya [Microprocessors and Microcontrollers for Control Systems]. Moscow: Goryachaya Liniya — Telekom Publ. 376 p.
2. Ugryumov E.P. (2010). Tsifrovaya skhemotekhnika [Digital Circuit Design]. Saint Petersburg: BHV-Petersburg Publ. 816 p.
3. Tanenbaum A., Austin T. (2013). Structured Computer Organization. 6th ed. Saint Petersburg: Piter Publ. 816 p.
4. Avaev V.N., Naumov Y.E., Frolokin V.T. (1991). Osnovy mikroelektroniki [Foundations of Microelectronics]. Moscow: Radio i svyaz Publ. 288 p.
5. Shchuka A.A. (2007). Naoelektronika [Nanoelectronics]. Moscow: Fizmatlit Publ. 464 p.
6. Gusev A.I. (2007). Nanomaterialy, nanostruktury, nanotekhnologii [Nanomaterials, Nanostructures, Nanotechnologies]. Moscow: Fizmatlit Publ. 416 p.
7. Baranov S.I. (1979). Sintez mikroprogrammnykh avtomatov [Synthesis of Microprogrammed Automata]. Leningrad: Energiya Publ. 232 p.
8. Hennessy J.L., Patterson D.A. (2019). Computer Architecture: A Quantitative Approach. Moscow: Technosfera Publ. 936 p.
9. Klimov S.A. (2015). Fizika poluprovodnikovyykh priborov i integralnykh mikroskhem [Physics of Semiconductor Devices and Integrated Circuits]. Moscow: Vysshaya Shkola Publ. 312 p.

10. Denisov A.V. (2022). Geterogennyye vychislitelnye sistemy i chipletnaya integratsiya [Heterogeneous Computing Systems and Chiplet Integration]. Moscow: BMSTU Publ. 198 p.