

Ю. С. Затуливетер, Е. А. Фищенко, С. Е. Артамонов, В. А. Козлов

Элементы стратегии опережения и архитектурные предпосылки к созданию однокристалльного ускорителя массовых вычислений общего назначения на базе архитектуры отечественного многопроцессорного компьютера ПС-2000

АННОТАЦИЯ. Анализируются причины, проявления и индустриальные проблемы кризиса структурного насыщения микропроцессорных архитектур. Рассматривается состояние рынка однокристалльных компьютеров-ускорителей общего назначения для задач с массовым параллелизмом. Предлагается концепция развития отечественной линии высокопараллельных компьютеров ПС-2000, которые можно рассматривать как первых представителей общедоступных многопроцессорных систем класса GP (General Purpose). В развитие архитектуры ПС-2000 приводится описание свойств однокристалльной масштабируемой и комплексируемой многопроцессорной системы ПС-2000М, обладающей архитектурным потенциалом опережения. Приводится сопоставление архитектуры ПС-2000М с современными архитектурами однокристалльных компьютеров с массовым параллелизмом.

Ключевые слова и фразы: компьютерный кризис, архитектуры с массовым параллелизмом, элементная база, однокристалльные компьютеры-ускорители, массовые вычисления общего назначения, масштабируемость, комплексируемость, стратегия опережения.

Введение

Одной из главных причин гарантированного отставания в сферах высокопроизводительных компьютеров является отсутствие отечественной элементной базы (ЭБ), отвечающей современным и перспективным требованиям. В рамках суперкомпьютерной про-

© Ю. С. ЗАТУЛИВЕТЕР, Е. А. ФИЩЕНКО, С. Е. АРТАМОНОВ, В. А. КОЗЛОВ, 2013

© ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ ИМЕНИ В.А. ТРАПЕЗНИКОВА РАН, 2013

© ООО «ИДМ» (INTERNATIONALDESIGN&MARKETING, LTD.), 2013

© **ПРОГРАММНЫЕ СИСТЕМЫ: ТЕОРИЯ И ПРИЛОЖЕНИЯ**, 2013

блематики и, особенно, ExaScale Computing особую роль играют однокристалльные существенно многопроцессорные ускорители. В условиях «теплового барьера», который сделал невозможным дальнейшее наращивание прежними темпами рабочей частоты, ЭБ в виде однокристалльных многопроцессорных ускорителей становится основным резервом увеличения производительности за счёт наращивания многопроцессорного параллелизма. Примерами зарубежной ЭБ, которая обеспечивает восхождение суперкомпьютеров по петафлопсной лестнице, являются однокристалльные многопроцессорные ускорители класса GP (General Purpose — общего назначения), такие как GP GPU (nVIDIA, AMD/ATI), Intel MIC и др.

Достижение значимых позиций на мировом компьютерном рынке — одна из приоритетных задач возрождения отечественного компьютеростроения. На первый взгляд она кажется нерешаемой, но на современном этапе развития мирового компьютеростроения вступают в действие фундаментальные рыночные факторы, которые требуют изменения базовых моделей развития индустрии массового производства компьютеров и программ. А это влечёт необходимость кардинальной переоценки сложившихся приоритетов и соответствующих структурных изменений мирового компьютерного рынка. Инвестиционная политика должна учитывать новые рыночные факторы долгосрочного влияния непреодолимой силы.

Вступление в ВТО может устранить дискриминационные барьеры на путях выхода на мировой высокотехнологичный компьютерный рынок с новыми отечественными разработками ЭБ в части архитектур многопроцессорных ускорителей, но это предъявляет высочайшие требования к уровням их конкурентоспособности. Особое значение обретают проекты, которые опираются на превосходящие фундаментальные научные заделы и успешный опыт практических наработок, и направлены на разработку отечественной ЭБ и организацию ее массового производства на основе полупроводниковых технологий глубокого нанометрового диапазона. Такие

разработки способны составить основу для формирования новых рыночных ниш высокотиражной наукоёмкой продукции.

В области высокопроизводительных и массовых компьютеров конкурентоспособность наукоёмких ноу-хау может достигаться только в рамках сложившегося на мировом рынке разделения труда с привлечением современных полупроводниковых технологий массового производства интегральных схем глубокого нанометрового диапазона 40–28–20–14–10нм.

Передовые СБИС-технологии перешли на промышленное освоение глубокого нанометрового диапазона и уже предоставляют 1–3 млрд. и более транзисторов на кристалле. Однако эффективное использование этого ресурса в целях повышения производительности посредством простого увеличения числа классических микропроцессорных ядер на кристалле не имеет перспектив.

Впервые массовое компьютеростроение оказалось в ситуации архитектурного кризиса: микропроцессорные архитектуры исчерпали свой потенциал параллелизма уже на уровне 10–50 млн. транзисторов, а системообразующий потенциал многоядерных архитектур с фоннеймановскими микропроцессорами совершенно недостаточен для высокоэффективного использования потенциальных возможностей современных и перспективных СБИС-технологий.

Новейшие разработки однокристалльных «неймановских» многопроцессорных архитектур с массовым параллелизмом (сотни процессоров на кристалле и более) всё ещё находятся в начальной стадии индустриального становления и развития.

Наиболее известным примером промышленного производства таких архитектур служат новейшие графические процессорные устройства (GPU) известных производителей nVIDIA и AMD. Однако оснований считать, что архитектуры, изначально ориентированные на использование в составе видеоплат ПК, или других узкопрофильных применений, сохраняют высокую эффективность на других классах задач с массовым параллелизмом, пока нет. Проблемы поиска и обоснования конкурентоспособных многопроцессорных архитектурных решений, отвечающих требованиям массо-

вых применений в широких классах задач, в наступившем десятилетии становятся одними из наиболее приоритетных.

По сути, речь идёт о развитии новой ниши компьютерного рынка — высокопроизводительных однокристалльных компьютеров общего назначения с массовым параллелизмом. Этот рынок охватит весь диапазон применений — от массовых устройств мобильной связи и встраиваемых систем управления, до суперкомпьютеров производительностью 1–1000 Пфлопс и более.

Это стратегически важное направление развития компьютеростроения, в котором отставание в части СБИС-технологий на несколько поколений может в рамках международного разделения труда компенсироваться обладанием ноу-хау в части компьютерных архитектур и способов их индустриального программирования.

Следует отметить, что после длительного периода сверхбыстрого прогресса сложилась следующая ситуация: имеются перспективные СБИС-технологии, для которых пока нет ни совершенных многопроцессорных архитектурных решений высокой эффективности, обеспечивающих наращивание реальной производительности пропорционально росту числа транзисторов и количества процессоров (ядер), ни индустриальных средств автоматизации их программирования, в полной мере отвечающих сложившимся требованиям массового производства компьютеров и программ. В отсутствие эффективных архитектурных решений одностороннее лидерство в технологиях СБИС уже не даёт, как прежде, безусловного превосходства.

Уникальность ситуации в том, что в условиях непреодолимых ограничений на рост рабочих частот компьютерная индустрия вынуждена осваивать массовое производство кристаллов с существенно многопроцессорными архитектурами. В отсутствие альтернативных путей наращивания производительности пропорционально росту числа транзисторов на кристалле для выведения таких архитектур в сферы массового производства/потребления требуются «нестандартные» подходы и к архитектурам, и к способам их индустриального программирования.

Можно говорить о том, что в настоящее время на мировом рынке высокопроизводительных вычислений продолжает оставаться явный дефицит высокоэффективных многопроцессорных архитектур для задач с массовым параллелизмом. Это обстоятельство позволяет нам предложить к рассмотрению апробированную многолетней практикой эффективного промышленного применения отечественную многопроцессорную архитектуру ПС-2000 [1], [2], [3], [4], которая может быть положена в основу как фундамент для построения высококонкурентных однокристалльных многопроцессорных компьютеров класса GP.

Цели данной работы:

- проанализировать причины и индустриальные проблемы внутрикомпьютерного кризиса и определить некоторые ключевые тенденции развития компьютеростроения в сфере высокопроизводительных вычислений;
- в сравнении с современными архитектурами однокристалльных компьютеров с массовым параллелизмом рассмотреть пути конкурентоспособного развития архитектурной линии ПС-2000 в ориентации на передовые технологии глубокого нанометрового диапазона.

1. Компьютерный рынок в преддверии кардинальных перемен

Компьютерный рынок — не только одна из наиболее доходных и динамично развивающихся сфер мирового рынка. Массовые компьютеры, связанные глобальными сетями, проникают во все сферы жизнедеятельности. Сверхбыстрое развитие этого рынка влияет на мировую экономику в целом. Благодаря глобальному информационному пространству, носителем которого являются компьютерные сети, он оказывает тотальное системообразующее влияние на все стороны жизни. Меняются базовые ценности и приоритеты, зарождаются новые траектории развития мировой социосистемы, которые потребуют изменений структуры мирового рынка и социальных отношений. Компьютеростроение стало ключевым фактором социально-экономического развития. Качество жизни прочно свя-

зано с уровнем технического развития компьютерных технологий, которые становятся доступными всё большему числу людей.

Растущая зависимость социосферы от компьютерных технологий имеет и обратную сторону. Своевременен нерешаемые фундаментальные проблемы компьютеростроения, связанные с глубинными проявлениями внутрикомпьютерного кризиса, приводят к неравномерному развитию ключевых сфер компьютерного рынка, что ведёт к нарушению системно-технических балансов в развитии собственно компьютерной среды и компьютерного рынка. В условиях неустранённого внутрикомпьютерного кризиса утрачиваются перспективы устойчивого развития компьютерного рынка и, как следствие, мирового. Инвесторы утрачивают долгосрочные ориентиры. Растут инвестиционные риски, а вместе с ними и избыток бездействующих финансовых ресурсов, что усугубляет нестабильность. И чем масштабнее проявления внутрикомпьютерного кризиса, тем больше потери упущенной выгоды.

Анализ ключевых тенденций развития компьютерной среды и причин нарастающих проявлений внутрикомпьютерного кризиса [5] показывает, что прежние принципы её формирования и механизмы развития приблизились к исчерпанию своего системообразующего потенциала.

В первых поколениях классических микропроцессоров, которые, как известно, реализуют универсальную модель последовательных вычислений Дж. фон Неймана, повышение производительности осуществлялось не только за счёт увеличения рабочей частоты, но и на структурно-архитектурном уровне путём увеличения разрядности машинных слов и аппаратного распараллеливания алгоритмов выполнения арифметических операций. Производительность в расчёте на транзистор [5] при этом быстро росла пропорционально числу транзисторов, см. рис.1. Это говорит о том, что классическая модель последовательного счёта на структурно-архитектурном уровне имела скрытые резервы внутреннего параллелизма (разрядность, параллелизм арифметических операций, специализированных устройств типа ММХ, кэширование потоков

данных и команд, конвейеризация операций и команд, предсказание условных переходов и др.).

Компьютерный рынок – не только одна из наиболее доходных и динамично развивающихся сфер мирового рынка. Массовые компьютеры, связанные глобальными сетями, проникают во все сферы жизнедеятельности. Сверхбыстрое развитие этого рынка влияет на мировую экономику в целом. Благодаря глобальному информационному пространству, носителем которого являются компьютерные сети, он оказывает тотальное системообразующее влияние на все стороны жизни. Меняются базовые ценности и приоритеты, зарождаются новые траектории развития мировой социосистемы, которые потребуют изменений структуры мирового рынка и социальных отношений. Компьютеростроение стало ключевым фактором социально-экономического развития. Качество жизни прочно связано с уровнем технического развития компьютерных технологий, которые становятся доступными всё большему числу людей.

Растущая зависимость социосферы от компьютерных технологий имеет и обратную сторону. Своевременно нерешаемые фундаментальные проблемы компьютеростроения, связанные с глубинными проявлениями внутрикомпьютерного кризиса, приводят к неравномерному развитию ключевых сфер компьютерного рынка, что ведёт к нарушению системно-технических балансов в развитии собственно компьютерной среды и компьютерного рынка. В условиях неустранённого внутрикомпьютерного кризиса утрачиваются перспективы устойчивого развития компьютерного рынка и, как следствие, мирового. Инвесторы утрачивают долгосрочные ориентиры. Растут инвестиционные риски, а вместе с ними и избыток бездействующих финансовых ресурсов, что усугубляет нестабильность. И чем масштабнее проявления внутрикомпьютерного кризиса, тем больше потери упущенной выгоды.

Анализ ключевых тенденций развития компьютерной среды и причин нарастающих проявлений внутрикомпьютерного кризиса [5] показывает, что прежние принципы её формирования и механизмы развития приблизились к исчерпанию своего системообразующего потенциала.

В первых поколениях классических микропроцессоров, которые, как известно, реализуют универсальную модель последовательных вычислений Дж. фон Неймана, повышение производительности осуществлялось не только за счёт увеличения рабочей частоты, но и на структурно-архитектурном уровне путём увеличения разрядности машинных слов и аппаратного распараллеливания алгоритмов выполнения арифметических операций. Производительность в расчёте на транзистор [5] при этом быстро росла пропорционально числу транзисторов, см. Рис. 1. Это говорит о том, что классическая модель последовательного счёта на структурно-архитектурном уровне имела скрытые резервы внутреннего параллелизма (разрядность, параллелизм арифметических операций, специализированных устройств типа MMX, кэширование потоков данных и команд, конвейеризация операций и команд, предсказание условных переходов и др.).

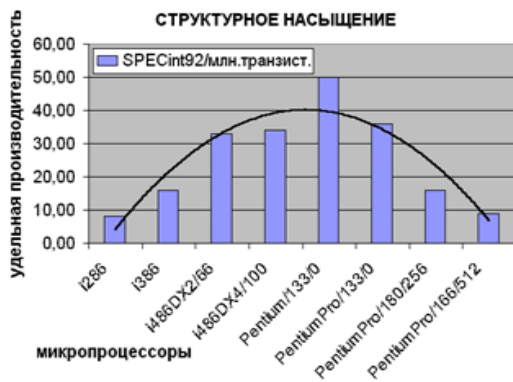


Рис. 1. Структурное насыщение микропроцессорных архитектур

Из Рис. 1 видно, что максимальное значение этого показателя эффективности микропроцессорных архитектур фирмы Intel достигнуто на первом Пентиуме (Pentium I, 3.1 млн. транзисторов). Изначально ограниченные внутренние резервы параллелизма последовательной модели были, в значительной мере, исчерпаны.

Можно утверждать, что в диапазоне 3–25 млн. транзисторов на кристалле было достигнуто структурное насыщение микропроцессорных архитектур.

Компьютерная индустрия в середине 90-х вошла в начальную фазу кризиса классической модели последовательного счета. Последующие поколения одноядерных микропроцессоров в связи с исчерпанием скрытых резервов параллелизма модели последовательных вычислений быстро утрачивали компоненту наращивания производительности за счёт сверхбыстрого роста количества транзисторов на кристалле. Темпы роста производительности обеспечивались, главным образом, за счёт увеличения рабочей частоты, связанного с уменьшением размеров транзисторов. Ценой роста рабочей частоты стало непропорционально высокое потребление энергии.

Налицо растущее обесценивание возможностей прогрессирующих полупроводниковых технологий — главного двигателя компьютерного прогресса. Контраргументация о резком снижении стоимости каждого транзистора с ростом степени интеграции не отменяет архитектурного кризиса микропроцессоров, а лишь объясняет недолгосрочную конъюнктурную рентабельность массового производства микропроцессорных кристаллов, заполненных по большей части «безработными» транзисторами.

Компьютеростроение и компьютерный рынок приблизились к критической фазе своего развития, когда для дальнейшего прогресса необходимо принципиальное обновление компьютерных первооснов. Опережающего прогресса полупроводниковых технологий уже недостаточно.

Компьютерный рынок вступает в период кардинальных структурных перемен, которые неизбежно приведут к смене поколений лидеров компьютерной индустрии, существенному пересмотру приоритетов в инвестиционных процессах. Преимуществ нынешних лидеров компьютерного рынка, основанных на прежних достижениях, уже недостаточно для открытия и обустройства новых сфер массового влияния. Масштабы необходимых структурных перемен в большинстве случаев превысят адаптационные возможности ли-

деров, которые сильно ограничены длинными шлейфами прежних обязательств перед миллиардами клиентов. На этапе смены системообразующих принципов строить новое будущее и одновременно тащить растущий груз прошлого становится нереальным. Будущее вступает в противоречие с прошлым и требует новых правил функционирования компьютерного рынка.

2. Причины и проявления внутрикомпьютерного кризиса

Суть глубинного внутрикомпьютерного кризиса в следующем. В основе современного рынка массовых компьютеров и программ лежат два крупнейших достижения 20-го века:

- классическая модель универсальных последовательных вычислений — модель Дж. фон Неймана, которая дала старт компьютерной эпохе (конец 40-х прошлого столетия);
- микроэлектронные технологии массового производства полупроводниковых интегральных схем, которые компьютерную эпоху сделали достоянием всего человечества.

Модель фон Неймана — это свод простых логико-технических правил автоматического выполнения любых алгоритмов в последовательном режиме «команда-за-командой». Они были положены в основу первых универсальных компьютеров. Они стали основой микропроцессорной революции и до сих пор остаются единым и единственным логическим «стандартом» индустрии массовых компьютеров и программ. Главное достоинство этой модели — простота и эффективность машинной реализации универсальных вычислений.

В последовательной модели вычислений в каждый момент исполняется одна команда (операция), что позволяет называть её скалярной моделью вычислений. Вычислительный процесс выглядит как последовательная траектория точечных (скалярных) событий, каждое из которых представляет исполнение одной операции. Уникальное достоинство классической модели, при этом, состоит в том, что она на инженерном уровне предлагает логически простейший и, в то же время, практически эффективный механизм

управления их реализацией. Поэтому именно она легла в основу микропроцессорной революции.

Технологии полупроводниковых интегральных схем дали долгосрочную материальную основу (в виде кремниевых кристаллов) для массовой реализации компактных и недорогих универсальных компьютеров, основанных на классической модели последовательных вычислений. В основе таких компьютеров лежат различные реализации однокристалльных процессоров, которые стали называться «микропроцессорами». Эти миниатюрные, быстро развивающиеся микроэлектронные устройства лежат в основе массового компьютеростроения и трёх десятилетий компьютерной революции. Темпы компьютерного прогресса стали определяться сверхвысокими скоростями развития полупроводниковых технологий массового производства интегральных схем, которые выражаются известным законом Мура: «Количество транзисторов на кристалле удваивается каждые 1.8–2 года».

Определяющей тенденцией в развитии технологий интегральных схем является уменьшение размера транзисторов и толщины проводников. Их уменьшение даёт двойной эффект. Во-первых, чем меньше транзисторы, тем быстрее они могут срабатывать и чем меньше расстояние между ними, тем скорее обмены сигналами, что позволяет увеличивать рабочие частоты, а значит и вычислительную производительность. Во-вторых, увеличивается плотность размещения транзисторов на поверхности кристалла. Быстрое увеличение количества транзисторов на кристалле (квадратичный рост с уменьшением линейных размеров транзисторов) открывает возможности для аппаратного наращивания параллелизма вычислительных устройств, что также служит ключевым фактором повышения производительности.

На начальном этапе развития микропроцессоров повышение производительности осуществлялось одновременно как за счёт увеличения рабочей частоты, так и на структурно-архитектурном уровне наращивания параллелизма. Во второй половине 90-х (после появления микропроцессора PentiumI, см. РИС. 1) резервы аппаратного реализуемого параллелизма классической модели последо-

вательного счёта были в значительной мере исчерпаны. Сохраняя на уровне программистов главный принцип управления — «команда-за-командой», конвейерный параллелизм (суперскалярность) также ограничен числом операций, задаваемых командой (считывание команды, операндов, вычисление, запись результата). По мере исчерпания скрытого параллелизма модели последовательного счета продолжающийся по закону Мура рост числа транзисторов в рамках классической модели перестал трансформироваться в пропорциональное прибавление производительности. Удельная производительность микропроцессора в расчёте на транзистор в линейке Intel достигла максимума на микропроцессоре PentiumI и в следующих поколениях она только снижалась, причём с нарастающей скоростью увеличения числа транзисторов на кристалле по закону Мура.

Таким образом, внешняя рыночная сторона внутрикомпьютерного кризиса проявляла себя в кардинальном снижении вычислительной отдачи от крайне дорогостоящих и быстро растущих инвестиций в новые поколения полупроводниковых технологий. С середины 90-х снижение удельной производительности каждого транзистора микропроцессоров новых поколений вело, по сути, к обесцениванию фундаментальных инвестиций в полупроводниковые технологии. Но некоторое время на фоне растущих прибылей от снижения себестоимости и миниатюризации, которые давали основу для расширения потребительских сфер компьютерного рынка, обесценивание происходило незаметно для бизнеса (без снижения доходности). Это были ранние проявления начальной фазы внутрикомпьютерного кризиса.

В течение последующих 10 лет, вплоть до середины 00-х кризис классической модели последовательных вычислений и микропроцессорных архитектур развивался в латентной форме. Компьютерная индустрия, развиваясь за счёт расширения сфер применения массовых компьютерных устройств, игнорировала нарастающие проявления структурного кризиса. Исчерпав резервы скрытого от программистов параллелизма классической модели, она стала до-

вольствоваться увеличением производительности новых поколений однопроводных микропроцессоров лишь за счёт наращивания рабочей частоты. А сверхбыстрый рост «избыточных» транзисторов успешно прятали в многоуровневых кэшах.

В соответствии с ростом рабочих частот (1–4 ГГц), достигавшемся посредством уменьшения размера транзисторов, росло быстроедействие, которое позволяло ускорять выполнение всех последовательных программ (в одинаковой мере новых и старых) и обеспечивать коммерческий успех дальнейшей смены поколений микропроцессоров на рынке.

Активная реклама обходила вниманием катастрофическое снижение удельной производительности в расчёте на транзистор, которое осталось совершенно неизвестным как для бизнеса и потребителя, так и инвесторов. Разработчики, не имея единой универсальной модели параллельных вычислений, которая могла бы составить рыночную альтернативу классической последовательной, поддерживали инерционное движение в рамках модели последовательного счёта, шлифуя накатаные, но уже бесперспективные в долгосрочном горизонте, микропроцессорные решения.

Однако беззаботный период «гладкого» развития в рамках классической модели и «лёгких» прибылей к середине 00-х завершился. С освоением технологий 90–65нм количество транзисторов на кристалле приблизилось к миллиарду. Рабочая частота достигла 4 Гц и более, но при этом энергопотребление транзисторов увеличилось настолько, что воздушное охлаждение перестаёт справляться с отводом тепла. Дальнейшее наращивание производительности массовых микропроцессоров за счёт увеличения частоты стало экономически неоправданным. Кроме того, огромный и сверхбыстро растущий сектор мобильных устройств особо остро нуждается в энергоэффективных методах повышения производительности.

Тепловой барьер лишил возможности повышения производительности за счёт увеличения частоты. Стратегическая ловушка структурного насыщения микропроцессорных архитектур захлопнулась окончательно.

Чтобы хоть как-то наращивать производительность промышленности вынужденно ответила многоядерными кристаллами. Появились 2–4–8-ми ядерные микропроцессоры. Однако они решили проблемы наращивания производительности лишь в незначительной степени. Если 2-х ядерные давали рост производительности почти в 2 раза, то каждое новое ядро добавляло производительности все меньше и меньше. В большинстве применений максимальное количество ядер не превышает 4. Этот параллелизм реализуется на системном уровне не всегда эффективно. Кроме того, проблема «узкого горла» общей памяти для растущего числа ядер в принципе не позволяет существенно увеличивать их количество. Современные многоядерные кристаллы не стали адекватным ответом внутрикомпьютерному кризису структурного насыщения микропроцессорных архитектур.

Приведём наглядный пример, который показывает масштабы кризиса структурного насыщения. Технология вчерашнего дня 45 нм предоставляет на кристалле около 1 млрд. транзисторов. На таком кристалле можно было бы разместить более 300 ядер с эффективной архитектурой Pentium I (3.1 млн. транзисторов). Однако из-за «узкого горла» памяти все 300 ядер будут работать медленнее, чем каждое ядро в отдельности. Отсюда видна «цена» архитектурного кризиса классической модели вычислений — снижение коэффициента полезного использования транзисторов в сотни раз.

Латентный период развития кризиса структурного насыщения «внезапно» для подавляющего большинства закончился. Универсальные многоядерные микропроцессоры, которые выпускаются массовыми тиражами с середины 00-х, не устранили проблемы структурного насыщения микропроцессорных архитектур.

Отсутствие долгосрочных перспектив развития такой многоядерности выражается в следующем:

- ограниченный параллелизм в обменах между ядрами и памятью не позволяет наращивать производительность пропорцио-

нально числу ядер, рост производительности прекращается уже на нескольких ядрах;

- прямым следствием структурного насыщения микропроцессорных архитектур стал кризис индустриальных технологий программирования, в основе которых десятилетиями оставалась модель последовательных вычислений.

Массовое производство столь бесперспективной «всего-лишь-несколько-ядерной» архитектуры стало молчаливо-безличным признанием компьютерной индустрией того свершившегося факта, что системообразующий потенциал классической модели последовательного счёта и реализующих её одноядерных микропроцессорных архитектур исчерпан, а её полноценной постнеймановской замены всё ещё нет.

Отсутствие публичных признаний в этом нерядовом явлении понятно. В течение 30 лет гладкого асфальта инвесторы привыкли, форсируя педаль газа, каждые 3–4 года удваивать свои вложения в полупроводниковые технологии. Однако прежняя магистраль развития достигла предела, а нынешние и, тем более, грядущие удвоения многомиллиардных инвестиций — дело всё более непростое. Нужны ясные перспективы и понятные дорожные карты. В их отсутствие инвестор начинает пересматривать риски и вспоминать о тормозах. А это никому не нужно — ни компьютерной индустрии, ни потребителям, ни политикам, ни самим инвесторам. Всем необходимо иное — как можно быстрее и с наименьшими потерями вывести рынок на новый компьютерный мэйнстрим.

В отсутствие единой постнеймановской модели массовое компьютеростроение до сих пор не имеет ключевой опоры для дальнейшего прогресса. Компьютерной индустрии необходима новая модель развития, которая должна прийти на замену классической и стать логическим «стандартом» индустрии массового производства уже не последовательных, а высокопараллельных компьютеров и программ. Внутрикомпьютерный кризис достиг своего пика, но пути выхода из него только предстоит отыскать.

3. Индустриальные проблемы внутрикомпьютерного кризиса

Начальный, относительно гладкий, потому сравнительно лёгкий, этап тридцатилетнего развития массового компьютеростроения в рамках единой и простейшей модели последовательного счёта закончился. Бизнесом собрано почти всё, что скрывалось в тонком поверхностном слое компьютерного прогресса, на который распространяется действие классической модели последовательного счёта. Более глубокие пласты тотальной компьютеризации на многие порядки более прибыльны, но требуют новых моделей и инструментов. Но их уже невозможно собрать в рамках классической компьютерной аксиоматики.

С опозданием на десятилетие компьютерный рынок вынужденно приступает к поиску индустриальных моделей параллельных вычислений, ориентированных на реализацию посредством многопроцессорных архитектур, которые оказались бы способными стать основой формирования нового, уже постнэймановского и постмикрпроцессорного, мэйнстрима в массовом производстве компьютеров и программ.

Для индустриализации нового мэйнстрима на смену классической — простейшей, ввиду своей скалярности, модели универсального счёта, должны прийти другие модели и архитектуры, которые регламентируют универсальные вычисления в пространстве параллельных вычислительных процессов. Фундаментальное отличие от классических последовательных траекторий скалярных вычислительных событий в том, что в пространстве параллельных процессов доминирует новое измерение, характеризующее множественность вычислительных событий, обозначаемая термином «параллелизм». Параллелизм предполагает, что в каждый момент одновременно исполняется множество операций (команд). Чем больше таких операций способна предоставить вычислительная задача, тем более высокими уровнями параллелизма должны располагать ком-

пьютерные архитектуры. Только так достигается рост производительности параллельных вычислений.

Модели параллельных вычислений и соответствующие им архитектуры начали активно разрабатываться ещё в 60-е годы. С тех пор их наработано огромное количество. Но почти все они создавались вне индустриальных требований практической реализуемости компьютеров и программ в массовых тиражах и, поэтому, не могут составить основу для нового индустриального мэйнстрима.

Массовый компьютерный рынок с одной стороны крайне динамичен в поисках и охвате новых сфер сбыта, а с другой — консервативен, поскольку отягощён колоссальной инерцией сопровождения наработанных продуктов. Целостное развитие уходящего мэйнстрима компьютеростроения обеспечивалось системообразующим потенциалом классической модели последовательных вычислений. Теперь, когда этот потенциал почти вычерпан до дна из-за отсутствия общей базовой модели параллельных вычислений, адекватной требованиям массового производства компьютеров и программ, резко возрастают риски утраты целостности компьютерного рынка. Упущенное десятилетие, в течение которого внутрикомпьютерный кризис из латентной фазы перерос в запущенную и открытую, требует экстренных мер по предотвращению стихийного распада устаревающих рыночных структур, уже не отвечающих новым вызовам.

Десятилетней задержке с активными поисками новой базовой модели трудно найти оправдание, поскольку характер моделей параллельных вычислений и соответствующих компьютерных архитектур по отношению к классике кардинально меняется. При этом научные проблемы достижения практической эффективности таких моделей и архитектур, отвечающие индустриальным требованиям массового производства компьютеров и программ, крайне сложны и для их решения требуется определённое время. В них доминируют уже не столько задачи совершенствования собственно технологий производства аппаратных и программных средств, сколько фундаментальные проблемы комбинаторной сложности, связанные с поиском эффективных многопроцессорных структур в

условиях математической многовариантности структурно-динамического многообразия параллельных вычислений.

В долгосрочной перспективе продвижения на рынок моделей параллельных вычислений и архитектур, отвечающих требованиям массового производства компьютеров и программ, необходимы новые высокоэффективные многопроцессорные архитектуры и технологии их индустриального программирования. Синергетическим эффектом стартовой точки лавинного роста может стать любое принципиальное продвижение в части однокристалльных компьютеров с высокопараллельными многопроцессорными архитектурами общего назначения.

4. Архитектурный потенциал опережения

Высокопараллельные многопроцессорные архитектуры — это та область знаний, в которых уровень конкурентоспособности изделий не может основываться только на превосходстве полупроводниковых интегральных технологий. Потребительские свойства таких изделий напрямую зависят от глубины научной проработки сложнейших многоаспектных проблем обоснования эффективности многопроцессорных архитектур, особенно тех, которые успешно апробированы обширной практикой промышленного применения.

Проблема в том, что среди огромного количества вариантов допустимого множества параллельных процессов только малое их число обладает достаточной эффективностью. Необходимость поиска эффективных решений становится главной проблемой индустриальных параллельных вычислений, архитектур и технологий программирования. Лишь ничтожная часть этих структурных решений имеет практическую значимость. И пока никто не предложил регулярных методов нахождения таких решений.

Высокопараллельные многопроцессорные архитектуры, выходя за пределы классической модели последовательного счета, требуют от создателей нестандартного научного и здравого инженерного

мышления. Здесь несогласованные мысли могут взлетать очень высоко, чтобы затем разбиться о жесткие реалии.

В этой рекордной области архитектурных тайн и сегодня остается значительно больше, чем найдено решений. И тем ценнее достижения, прошедшие проверку практикой большого промышленного производства и применения.

Оригинальные, выстроенные с большим, опережающим своё время, идеологическим запасом, архитектуры, не уходят в небытие вслед за устаревшей элементной базой.

Сказанное в значительной мере относится к отечественному многопроцессорному компьютеру ПС-2000 [1]–[4]. Это был один из первых в мире суперкомпьютеров, который выпускался большой промышленной серией и имел широкое народнохозяйственное применение в 1980–1997 г.г. Его оригинальная, масштабируемая и комплексируемая архитектура доказала свою высокую вычислительную эффективность и экономическую рентабельность во многих сферах промышленной обработки данных, а также в научно-инженерных расчётах, в системах обработки потоков данных реального времени и больших объёмов данных, в том числе обработки космической телеметрии, гидроакустических сигналов дальнего обнаружения и др. На ряде объектов промышленной обработки данных он более чем успешно конкурировал с полулегально приобретёнными (в обход эмбарго) зарубежными суперкомпьютерами лучших мировых производителей. Уступая в десятки раз по рабочей частоте, ПС-2000 не уступал в реальной производительности и на порядки выигрывал в себестоимости производства и эксплуатации.

Важнейшая особенность высокопараллельной архитектуры ПС-2000 состоит в том, что её эффективность не определяется жёсткой привязкой к особенностям технологической базы. Это нашло подтверждение в том, что, несмотря на более чем скромные возможности задействованной элементной базы, по показателям абсолютной производительности компьютеры ПС-2000 благодаря преимуществам своей архитектуры в 80-е годы длительное время находились на достойных позициях мирового списка суперкомпьютеров

Тор500. При том, что рабочая частота в десятки раз уступала соседям по списку. По потенциалу совершенствования экономического показателя производительность/стоимость эта архитектура остаётся одной из лучших в своём классе и в настоящее время. По признанию зарубежных специалистов [6], [7] ПС-2000 был одним из самых продвинутых и самым быстродействующим советским компьютером.

По сути, ПС-2000 с опережением на 2 десятилетия стал первым в мире многопроцессорным компьютером общего назначения для задач с массовым параллелизмом, отвечающий индустриальным требованиям многотиражного производства, применения и программирования. В настоящее время компьютеры такого рода и назначения принято относить к классу GP (General Purpose).

Современные многопроцессорные аналоги по областям применения, широко представленные на мировом рынке, — это однокристалльные многоядерные графические ускорители GP GPU (General Purpose Graphics Processing Units) с которыми предстоит конкурировать новым поколениям ПС-2000, изначально создавалась и балансировалась под узкие классы задач. Так, производители nVIDIA и AMD (ATI) начинали осваивать массовые рынки с изготовления графических ускорителей GPU (Graphics Processing Units) для ПК. В ходе их развития к середине 00-х сформировался новый рыночный класс однокристалльных многопроцессорных компьютеров класса GP. Примерно в это же время IBM вывела на рынок однокристалльный ускоритель Cell (IBM, Sony и Toshiba) [8], с гибридной многопроцессорной архитектурой, заточенный под игровые приставки.

Именно так, спустя 20 лет после индустриальной премьеры ПС-2000, состоялся новый выход высокопараллельных многопроцессорных архитектур в широкие сферы применений.

Конечно, масштабы промышленного тиражирования продуктов совершенно иные. ПС-2000 — выпускался большой промышленной серией в несколько сотен вычислительных комплексов. Продукция GPU и Cell — миллионные тиражи. Но это не удивительно. Совре-

менный компьютерный рынок — совершенно иная элементная база, иные масштабы сфер потребления. К классу GP также относятся новые многоядерные чипы MISC (Many Integrated Core) фирмы Intel, которые являются развитием трёх предыдущих проектов Larrabee, Teraflops Research Chip и Single-chip Cloud Computer, а также матричные многоядерные процессоры фирмы Tiler и др.

Отсутствие явно лидирующей в сферах массовых применений архитектуры с массовым параллелизмом свидетельствует об отсутствии качественно опережающего архитектурного решения, способного составить конкуренцию в различных сферах по большинству потребительских параметров. В отличие от узкопрофильных многопроцессорных архитектур в основу высокопараллельной архитектуры ПС-2000 изначально положены общие принципы высокоэффективной обработки данных на широких классах задачах с массовым параллелизмом. Она обладает достаточным потенциалом своего развития для конкурентоспособного покрытия большинства классов задач с массовым параллелизмом, отдельные подклассы которых поделены между сегодняшними игроками.

Надо отметить особый вклад узкопрофильных многопроцессорных ускорителей GPU и Cell, с которыми была пройдена наиболее рискованная часть рыночной инновации многопроцессорных архитектур. Уверенный бизнес ускорителей этих классов доказал, что высокопараллельные многопроцессорные архитектуры, несмотря на гораздо более высокую, в сравнении с классическими однопроцессорными компьютерами, сложность программирования, нашли своё место на компьютерном рынке и неуклонно расширяют своё присутствие.

Далее, по мере наполнения массового рынка узкопрофильных многопроцессорных ускорителей, стал формироваться широкий спрос на недорогие ускорители для более широких классов задач. К ним относятся, прежде всего, программируемые ускорители для научно-инженерных задач высокой вычислительной сложности. Такие системы отличаются от узкопрофильных прежде всего тем, что имеют открытые для пользователей системы программирования, которые позволяют им решать свои задачи.

Так на базе рынка узкопрофильных ускорителей сформировался следующий потребительский класс однокристалльных многопроцессорных ускорителей общего назначения — GP.

Значительная часть ускорителей класса GP применяется в настольных суперкомпьютерах и вычислительных серверах, центрах обработки данных, которые всё шире используются в исследовательских и проектных организациях, университетах, медицинских центрах. Такие ускорители применяются и в топовых суперкомпьютерах, обеспечивая освоение диапазона производительности более 1–10 Пфлопс и более.

В настоящее время большая часть ускорителей класса GP представлена расширением архитектур, изначально специализированных на обработку графики. Это класс GP GPU. Одним из лидеров этого класса являются ускорители фирмы nVIDIA — Fermi (40 нм, 512 ядер) [9] и Kepler (28 нм, 1536 ядер) [10], которые используются для обработки данных в различных приложениях с различной степенью эффективности. Это обосновано особенностями архитектуры, которая изначально была ориентирована на графическую обработку. Соответственно, те прикладные задачи, которые близки к структуре алгоритмов графической обработки, выполняются на GPU эффективно (Таблица 1). На других задачах эффективность архитектуры существенно падает.

ТАБЛИЦА 1. Показатели эффективности GP GPU в различных приложениях

Примеры задач, решаемых на GP GPU nVIDIA	Кратное ускорение $g_{GP GPU/CPU}$	КПД, % ¹
Гидрогазодинамика	10	5
Обработка изображений, кодирование видео, компьютерное зрение	3–160	1.5–80

¹ Представлено ускорение относительно времени решения задачи на универсальном многоядерном микропроцессоре. Соотношение пиковых производительностей GP GPU и большинства многоядерных микропроцессоров ~ 200. Отсюда КПД= Ускорение *100%/200.

Примеры задач, решаемых на GP GPU nVIDIA	Кратное ускорение $g_{=GP}$ GPU/CPU	КПД, % ¹
Сейсмическое моделирование (поиск нефти/газа)	30	15
Квантовая химия, моделирование белка	20–80	10–40
Моделирование взаимодействия объектов на молекулярном уровне	17	8.5
Базы данных, поиск, сортировка	2-6	1–3
Предсказание погоды, моделирование климата	20–40	10–20
Анализ и распознавание объектов, слежение за объектами	7–12	3.5–7
Криптография и криптоанализ	1.7–20	1–10

Следует признать, что системы класса GP, получаемые трансформацией из узкопрофильной многопроцессорной архитектуры GPU, изначально ориентированной на ограниченный набор алгоритмов графической видеобработки, не могут рассматриваться как окончательные решения, обеспечивающие высокую вычислительную эффективность на разнообразных классах задач. В частности одним из слабых звеньев такой архитектуры остаётся «узкое» горло к общей оперативной памяти, что является серьёзным ограничением в классах задач с произвольным доступом к данным. Всё это снижает стратегический потенциал конкурентоспособности таких подходов в развитии систем класса GP.

Дальнейшее развитие однокристалльных ускорителей класса GP требует архитектурных решений, которые сохраняют высокую вычислительную эффективность на как можно более широких классах задач с массовым параллелизмом.

Архитектура ПС-2000 изначально ориентирована на эффективное решение широких классов задач с массовым параллелизмом. Архитектура прошла практическую апробацию в течение более 10-летней эксплуатации в различных приложениях на задачах промышленной обработки, где ее эффективность приближалась к 100%. На показанных выше задачах, решаемых на GPU, архитектура ПС-2000 существенно более эффективна (КПД более 80%).

В ТАБЛИЦА 2 сопоставлены характеристики архитектур основных «игроков» рынка высокопроизводительных вычислений и предлагаемого однокристалльного компьютера ПС-2000М.

ТАБЛИЦА 2. Сравнение характеристик архитектур высокопроизводительных вычислителей для массового рынка

Характеристика	IBM Cell	nVIDIA Fermi, Kepler	ПС-2000М (концепция)
Тип архитектуры	Гибридная, ЦПУ и параллельная подсистема	Высокопараллельная	Гибридная, высокопроизводительное ЦПУ и высокопараллельная подсистема
Степень поддержки универсальных вычислений	высокая	низкая	высокая
Технология	65 нм	40 нм–10 нм	40 нм–10 нм
Эффективность связи между ЦПУ и параллельной подсистемой	высокая	очень низкая	высокая
Число ядер ЦПУ	1	-	>1
Возможности структурного масштабирования параллельной подсистемы	низкие	средние (пропускная способность памяти отстает от арифметики)	пропорционально высокие
Число ПЭ в параллельной системе	8 (16–32)	масштабируемое, 512, 1536	масштабируемое, 1024, 2046, ..., 16384
Пиковая производительность параллельной системы, Tflops	0,256	масштабируемая, 1, 3	масштабируемая, 1–2–4–8–16–(32 — на 10 нм)
Гибкость параллельной архитектуры	средняя	низкая	высокая, поддержка набора вычислительных моделей
Организация памяти	разделяемая и распределенная	разделяемая	разделяемая и распределенная
Объем встроенной памяти, МБ	1	0.6	16–512, пропорционально количеству ПЭ
Наличие прямого доступа в память	есть	нет	есть

Характеристика	IBM Cell	nVIDIA Fermi, Kepler	ПС-2000М (концепция)
Наличие встроенного межпроцессорного коммутатора	есть	нет	есть
Возможности межкристального масштабирования	нет	нет	есть

Здесь ЦПУ — центральное процессорное устройство, ПЭ — процессорные элементы.

Процессоры Cell имеют гибридную архитектуру с определенной степенью универсальности, но при отсутствии дальнейшего развития к настоящему моменту обладают относительно низким уровнем пиковой производительности. К недостаткам архитектуры следует отнести невысокую гибкость параллельной архитектуры и отсутствие возможностей ее масштабирования на линейке нанотехнологий.

К недостаткам архитектуры nVIDIA и GPU в целом можно отнести:

- существенную жёсткость архитектуры;
- низкую степень универсальности архитектуры;
- небольшой объём встроенной памяти;
- отсутствие встроенной межпроцессорной коммуникации;
- невозможность прямого ввода/вывода в память;
- отсутствие возможности «бесшовного» комплексирования нескольких GPU в систему.

В силу отсутствия встроенного ЦПУ, GPU не обеспечивают достаточного уровня универсализма вычислений. Поэтому невозможно построить высокопроизводительную компьютерную систему на основе только GPU, необходимо также использовать и универсальные микропроцессоры. Таким образом, одной из проблем эффективного использования GP GPU в высокопроизводительных вычислениях является узкое горло связи между универсальным микропроцессором и GP GPU, что существенно снижает возможности эффективного использования параллельных ресурсов GP GPU в широком диапазоне классов задач и алгоритмов.

Когда размерность решаемой задачи выходит за пределы возможностей аппаратуры GPU, отсутствие встроенной межпроцессорной коммутации и небольшой объем встроенной памяти приводит к необходимости обмена данными через внешнюю память. Это влечет за собой необходимость обеспечения высокой пропускной способности внешней памяти и, как следствие, повышенное энергопотребление всей системы.

Выделим основные свойства архитектуры ПС-2000М, позволяющие прогнозировать её высокую конкурентоспособность относительно современных систем с массовым параллелизмом:

Реализация архитектуры:

- высокое соотношение показателя «пиковая производительность/(транзистор*стоимость* энергопотребление)»;
- «околопиковая» производительность на широких классах задач (70-80% и выше), обеспеченная высокой гибкостью параллельной архитектуры.

Простота аппаратных решений и гибкость управления:

- SIMD из простых исполнительных устройств (АЛУ + регистры);
- VLIW для управления (простота дешифрации, уплотнение команд организует оптимизирующий компилятор);
- эффективная структура межпроцессорных обменов;
- пропорциональная структурная масштабируемость:
 - по числу процессорных элементов;
 - по объему памяти;
 - по ширине внешнего интерфейса ввода/вывода;
 - по частоте (энергосбережение).

Беспшовная комплексируемость:

- полностью программно управляемая;
- использование смесей параллельных и конвейерных структур;
- единый механизм на внутри- и межкристалльных (системных) уровнях;
- встроенные технологии беспшовной сетевой интеграции (позволяют формировать легко масштабируемые облачные кластеры, предназначенные для оказания широкого круга массовых услуг

по доставке пользователям суперкомпьютерных алгоритмических сервисов).

К существенной особенности предлагаемого к реализации однокристалльного компьютера ПС-2000М следует также отнести его гибридную архитектуру, объединяющую в одном микрочипе универсальный микропроцессор и процессор с массовым параллелизмом, что обеспечивает высокий уровень универсализма поддерживаемых вычислений относительно классов решаемых задач.

Пройдя высокорисковую стадию начального формирования на мировом рынке, потребительский класс ускорителей GP уже перешёл в устойчивую стадию развития. Его характерные особенности:

- наличие устойчиво растущего спроса в разнообразных сферах применения и долгосрочных тенденций к их расширению;
- число ведущих производителей невелико (в пределах десятки: среди них nVIDIA, AMD, IBM&Sony&Toshiba, Intel);
- отсутствие однозначно лидирующей архитектуры.

Несмотря на различие представленных на рынке архитектурных подходов, ни один из них не имеет существенного превосходства по всей совокупности значимых потребительских характеристик. Так, вычислительная эффективность, измеряемая коэффициентом полезного использования ядер, в разных классах задач даёт значительный разброс реальной производительности. Отсюда разделение сфер влияния по классам задач: графические видеоплаты (nVIDIA, AMD), игровые приставки (IBM&Sony&Toshiba), научно-инженерные расчёты высокой вычислительной сложности (nVIDIA, AMD, Intel, IBM).

В рамках сложившихся сфер влияния конкуренция идёт не столько на уровне архитектурных принципов, сколько на уровне инженерно-конструкторских способов воплощения чипов в рамках своих устоявшихся архитектурных шаблонов. В нишах, где конкурируют разные игроки, вперёд на короткое время выходит чип ускорителя того производителя, который сумел опередить в темпах проектирования при переходе на очередное поколение полупроводниковой технологии.

Отсутствие явно лидирующей архитектуры свидетельствует об отсутствии качественно опережающего архитектурного решения, способного составить конкуренцию в различных сферах по большинству потребительских параметров.

В отличие от узкопрофильных многопроцессорных архитектур, заложенных в GP GPU, в основу высокопараллельной архитектуры ПС-2000 изначально положены *общие принципы* высокоэффективной обработки данных на широких классах задач с массовым параллелизмом. Она обладает достаточным потенциалом своего развития для конкурентоспособного покрытия большинства классов задач с массовым параллелизмом, отдельные подклассы которых поделены между сегодняшними игроками.

Архитектура ПС-2000М обладает уникальным на фоне современных ускорителей свойством структурной масштабируемости, которое позволяет одновременно с увеличением количества вычислительных ядер *пропорционально* наращивать не только эффективную вычислительную производительность, но и пропускную способность встроенной равномерно распределённой по процессорным элементам памяти с высокопараллельным доступом и произвольной адресацией, снимающей проблемы «узкого горла» как собственно памяти, так и каналов ввода/вывода. Структурная масштабируемость включает в себя и свойства комплексированности как на внутри-, так и на межкристалльном- уровнях.

Структурная масштабируемость открывает дополнительные возможности и в эффективном раскрытии быстро растущего потенциала полупроводниковых технологий глубокого нанометрового диапазона, и в расширении классов эффективно решаемых задач. Одна, вместо многих узкопрофильных, масштабируемая, комплексированная, программно совместимая многопроцессорная архитектурная линейка, которая не уступает им в эффективности по совокупным параметрам, даёт превосходство в охвате разнообразных классов задач, а это важное преимущество в наращивании тиражей и формировании однородного по аппаратно-программным средств

вам массового рынка высокопроизводительных компьютеров и программ.

Доступ к современным полупроводниковым технологиям обеспечивает возможности пропорционального числу транзисторов повышения вычислительной производительности и пропускной способности памяти, снижение удельного энергопотребления, а также себестоимости производства и эксплуатации. Неудовлетворённый спрос массового рынка на общедоступные компьютерные устройства сверхвысокой производительности в классе GP открывает возможности выхода на многомиллионные тиражи, сопоставимые с тиражами традиционных многоядерных микропроцессоров.

Апробированная широкой практикой модель многопроцессорной архитектуры ПС-2000 не утратила своей актуальности до сих пор. Быстро прогрессирующие полупроводниковые технологии предоставляют возможности для максимального раскрытия вычислительного и экономического потенциала этой модели, что даёт долгосрочные перспективы успешной конкуренции в развитии систем класса GP на мировом уровне.

5. Особенности реализации ПС-2000М

Опыт разработки и развития архитектурной линии ПС-2000 составляет основу для возвращения отечественного компьютеростроения на передовые в мире уровни с применением современных полупроводниковых технологий глубокого нанометрового диапазона и соответствующих средств дизайна СБИС. Расчёты показывают, что на становящихся доступными для отечественной полупроводниковой промышленности технологиях 90–65 нм возможна реализация первых кристаллов семейства многопроцессорных компьютеров ПС-2000М с дальнейшим масштабированием на технологии 40–10 нм (см. рис. 2). Такое семейство, благодаря архитектурным преимуществам будет обладать большим запасом конкурентоспособности в диапазоне технологий 90–10 нм. Архитектура ПС-2000М [11] масштабируется по числу процессоров (ядер) на кристалле СБИС от 512 (технология 65нм) до 16К (10 нм) и по производительности однокристалльного компьютера от 1 до 30 Тфлопс и бо-

лее. На Рис. 2 приведена восходящая лестница развития линии однокристалльных компьютеров ПС-2000М на основе свойства структурной масштабируемости её архитектуры при реализации на современных и будущих интегральных технологиях. Важно отметить, что свойство структурной масштабируемости с увеличением количества транзисторов на кристалле обеспечивает пропорциональное улучшение эксплуатационных характеристик — числа процессоров и реальной суммарной производительности, объёмов встроенной памяти и пропускной способности этой памяти как во взаимодействии с процессорами, так и с каналом ввода/вывода.



Рис. 2. Пропорциональное наращивание вычислительных возможностей однокристалльных реализаций высокопараллельной архитектуры ПС-2000М по мере развития промышленных СБИС-технологий

На Рис. 3 приведена динамически перестраиваемая структура ПС-2000М, которая позволяет программной реконфигурацией комплексовать базовые многопроцессорные вычислительные модули с SIMD архитектурой (БВМ) внутри кристалла.

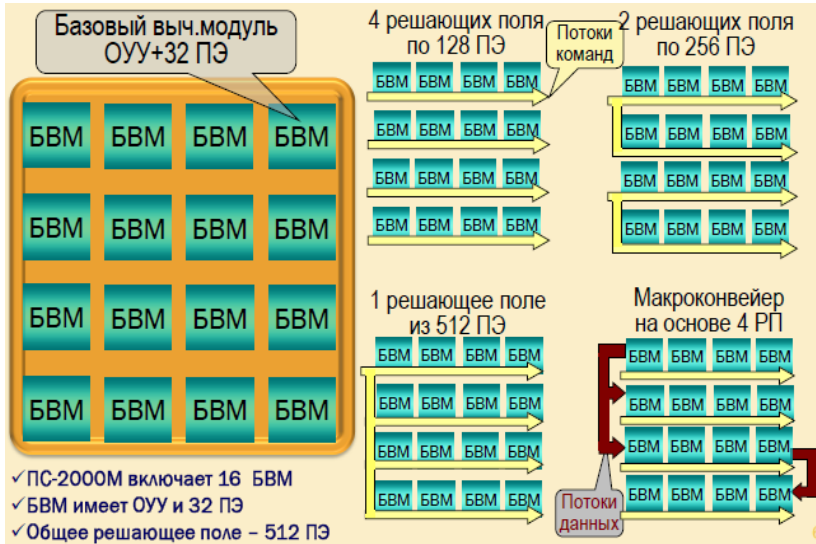


Рис. 3. Multi-SIMD комплексование БВМ внутри кристалла

Семейство однокристалльных компьютеров ПС-2000М может применяться во всем диапазоне средств вычислительной техники двойного назначения — от массовых мобильных компьютерных устройств и встраиваемых бортовых систем сетецентрического управления, до суперкомпьютеров производительностью 1–1000 Пфлосп и более.

Масштабируемость, комплексовуемость и высокая эффективность архитектуры ПС-2000М на широких классах задач с массовым параллелизмом позволит преодолеть отставание отечественного технологического производства СБИС, за счёт:

- возможности производить на имеющихся отечественных производствах (90, 65 нм) микросхем младшие модели семейства (64,

128, 256, 512 ядер), обеспечивающие повышение характеристик мобильных и стационарных систем двойного назначения;

- обладание уникальным проектом масштабируемой в диапазоне 65–10 нм заказной СБИС однокристалльного компьютера производительностью 1–30 Тфлопс и более станет стимулом для обретения технологий глубокого нанометрового диапазона с последующим выпуском старших моделей кристаллов ПС-2000М, которые откроют возможности достижения паритета и превосходства в области высокотехнологичных изделий.

6. Рыночные перспективы ПС-2000М

Рыночные перспективы однокристалльного компьютера ПС-2000М определяются отличительными свойствами его архитектуры, которые дают серьёзные преимущества и по обеспечению высокого КПД вычислений в широком диапазоне приложений, в части структурной масштабируемости в глубоком нанометровом диапазоне технологий изготовления. Эти преимущества в полной мере отвечают современным тенденциям и требованиям к развитию высокопроизводительных вычислений.

На первой стадии реализации проекта предполагается позиционировать его на рынке элементной базы для высокопроизводительных вычислительных систем, включающем сегменты высокопроизводительных компьютеров и рабочих станций, топовых серверов и суперкомпьютеров, микросерверов и дата-центров, используемых для реализации «облачных» и распределённых вычислений.

Кроме этого, однокристалльный компьютер ПС-2000М может использоваться для создания широкой номенклатуры встраиваемых систем, систем управления реального времени, игровых приставок, высокопроизводительных мобильных компьютеров и др.

Первоначально предполагается обеспечить позиционирование разрабатываемого однокристалльного компьютера ПС-2000М на рыночном сегменте высокопроизводительных серверов и суперкомпьютеров, а также сегменте «облачных» и распределённых вычислений. В дальнейшем при выполнении ряда условий возможно про-

никновение на рыночные сегменты персональных компьютеров и рабочих станций и на другие сегменты рынка средств вычислительной техники, включая рынки с практически неограниченным потенциалом потребления, такие как мобильные и встраиваемые компьютерные устройства.

Для эффективного проникновения на все рыночные сегменты, перечисленные выше, необходимо установление партнёрских отношений с ведущими российскими коммерческими и государственными компаниями, а также с зарубежными производителями средств вычислительной техники, разработчиками программного обеспечения в целевых сегментах рынка и прикладных областях.

Заключение

Создание отечественной элементной базы является ключевой народнохозяйственной задачей, которая открывает пути к возрождению отечественного компьютеростроения.

Сбалансированная для реализации на одном кристалле архитектура ПС-2000М обобщает уникальный опыт разработки компьютера ПС-2000. Эффективность его оригинальной, не утратившей актуальность, архитектуры, апробирована многолетней практикой высокопроизводительной обработки промышленных данных. ПС-2000 имеет многое, чтобы стать основой для однокристалльных многопроцессорных компьютеров, которым принадлежит будущее.

С появлением многоядерных кристаллов активизируется спрос на новейшие многопроцессорные архитектуры, сбалансированные под требования однокристалльной реализации. Весьма актуальной поэтому становится задача воплощения масштабируемой архитектуры ПС-2000М на одном кристалле.

Прекращение в середине 90-х годов финансирования научных разработок не остановило развитие архитектурной линии ПС-2000. За упущенные 15 лет в области архитектур с массовым параллелизмом не произошло отставания, поскольку в это время в области высокопроизводительных вычислений за рубежом доминировали кластерные многопроцессорные архитектуры из микропроцессоров, производительность которых росла в основном за счёт увеличения

рабочих частот. Микропроцессоры практически не раскрывают массовый параллелизм вычислительных операций. По сути это было время глубокого застоя в развитии архитектур с массовым параллелизмом.

Ситуация кардинально изменилась с достижением в середине 2000-х теплового барьера кремниевых технологий. Рост частоты стал невозможным из-за опережающего роста тепловыделения. Дальнейшее наращивание производительности прежними темпами на основе микропроцессоров стало невозможным. «Много»-ядерные микропроцессоры не дают прироста производительности уже после 4 ядер. Представленные на компьютерном рынке однокристалльные многопроцессорные ускорители класса GP GPU по своей эффективности на различных классах задач с массовым параллелизмом ещё очень далеки от совершенства

Наш «развал» и их «застой» на этот раз совпали. Сейчас, в условиях кризиса классической модели последовательного счета и острого архитектурного голода на высокоэффективные на широких классах задач многопроцессорные однокристалльные архитектуры система ПС-2000 даёт свой полновесный ответ на вопрос: «Что важнее, — элементная база или архитектура?» Многопроцессорный компьютер с лучшей архитектурой сумеет показать свои достоинства и на слабой элементной базе, а на лучшей ему не будет равных. Архитектура ПС-2000, доказавшая в ходе десятилетнего промышленного применения на разных классах задач с массовым параллелизмом эффективность своих принципов управления массовым параллелизмом, положена в основу нового проекта ПС-2000М как фундамент для построения высококонкурентных однокристалльных многопроцессорных компьютеров общего назначения класса GP.

Главный козырь структурно масштабируемой архитектуры ПС-2000М — сохранение высокой вычислительной эффективности на более широких классах задач с массовым параллелизмом, что достигается, прежде всего, за счёт особенностей архитектуры, обеспечивающей превосходство по значительно большему числу

параметров, и сохраняется при переходе к новым поколениям полупроводниковых технологий.

Фундаментальный характер новых решений в части однокристальных многопроцессорных архитектур и технологий их программирования требует в первую очередь интеллектуального ресурса и поэтапного вхождения в международное разделение труда. Для этого не требуются чрезмерные капиталовложения, что даёт шансы отечественным инвесторам и разработчикам.

На переходе к новому этапу развития компьютеростроения глубинный внутрикомпьютерный кризис выравнивает стартовые позиции. В этом *уникальный* исторический шанс для новых игроков — не только разработчиков, но и инвесторов. При этом следует ясно осознавать, что пространственно-временная щель в будущее открывается лишь для тех, кто её увидит раньше других. И открыта она будет недолго.

Список литературы

- [1] А.с. 751238 СССР. *Многопроцессорная вычислительная система* / Бирюков А.Я., Виленкин С.Я., Жуков В.А., Затуливетер Ю.С., Медведев И.Л., Прангишвили И.В., Голован Н.И., Итенберг И.И., Костелянский В.М., Набатов А.С., Пивоваров Г.Ю., Резанов В.В., Фищенко Е.А. // Бюллетень изобретений. —1983. — № 22.
- [2] Прангишвили И. В., Виленкин С. Я., Медведев И. Л. *Многопроцессорные вычислительные системы с общим управлением*. М.: Энергоатомиздат, 1983. 312 с.
- [3] Затуливетер Ю.С., Фищенко Е.А. *Многопроцессорный компьютер PC-2000* // Открытые системы. —2007. № 9. С.74–79. URL: <http://www.osp.ru/os/2007/09/4570286/>
- [4] Затуливетер Ю.С., Фищенко Е.А. *Многопроцессорный компьютер PC-2000 (Опыт создания и пути развития)*. Научное издание (Препринт). М.: Институт проблем управления РАН, 2012. 86 с. URL: http://www.ipu.ru/sites/default/files/publications/16551/3477-препринт%20пс-2000__2.pdf
- [5] Затуливетер Ю.С. *Компьютерные архитектуры: неожиданные повороты* // Hard 'n' Soft. 1996. № 2. С. 86–94. URL: http://zvt.hotbox.ru/p2_z1.htm.

- [6] Wolcott P., Goodman S. E. *High-Speed computers of the Soviet Union* // IEEE Computer. — 1988. Vol. 21, No 9, pp. 32–41.
- [7] Wolcott P., Goodman S. E. *International perspectives: under the stress of reform high-performance computing in the former Soviet Union* // Communications of the ACM. — 1993. Vol. 36, No 10, pp. 21–24.
- [8] *The Cell architecture*. URL: <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/cellengine/>
- [9] URL: http://www.nvidia.com/content/PDF/fermi_white_papers/NVIDIA_Fermi_Compute_Architecture_Whitepaper.pdf
- [10] URL: <http://www.nvidia.com/content/PDF/kepler/NVIDIA-Kepler-GK110-Architecture-Whitepaper.pdf>
- [11] Артамонов С.Е., Затуливетер Ю.С., Фищенко Е.А. *Предпосылки к созданию однокристалльного многопроцессорного компьютера ПС-2000М производительностью 1–10 Тфлорс* // Параллельные вычислительные технологии (ПаВТ'2011): труды международной научной конференции (Москва, 28 марта –1 апреля 2011 г.). С. 402–410. URL: http://omega.sp.susu.ac.ru/books/conference/PaVT2011/talks/012_zatuliveter_talk.pdf, Челябинск: Издательский центр ЮУрГУ, 2011. 730 с.
URL: <http://omega.sp.susu.ac.ru/books/conference/PaVT2011/short/012.pdf>.

Рекомендовал к публикации

Программный комитет

Второго национального суперкомпьютерного форума НСКФ-2013

Об авторах:



Юрий Семенович Затуливетер

Институт проблем управления имени В.А. Трапезникова РАН.

В.н.с., к.т.н., доцент.

Более 170 печатных работ.

Параллельные и распределённые вычисления, многопроцессорные компьютерные архитектуры, проблемы формирования единого алгоритмического пространства в сетевых вычислительных средах, сетецентрическое управление, социальные аспекты компьютерно-сетевой глобализации.

e-mail: [\(zvt@ipu.rssi.ru\)](mailto:zvt@ipu.rssi.ru)



Елена Алексеевна Фищенко

Институт проблем управления имени В.А. Трапезникова РАН.

В.н.с., к.т.н.

70 печатных работ.

Многопроцессорные архитектуры, распределённые вычисления, сетецентрические системы управления.

e-mail: [\(fish@ipu.ru\)](mailto:fish@ipu.ru)



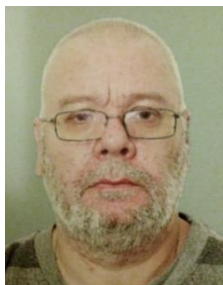
Сергей Евгеньевич Артамонов

ООО «ИДМ» (InternationalDesign&Marketing, Ltd.), г. Зеленоград.

Технический директор.

Современный дизайн в области микроэлектроники и информационных технологий, в том числе в области разработки специализированных и параллельных вычислительных архитектур и микропроцессорных микросхем с ориентацией на изготовление на современных технологиях нанометрового диапазона.

e-mail: [\(Sergey.Artamonov@idm.ru\)](mailto:Sergey.Artamonov@idm.ru)

**Владимир Алексеевич Козлов**

ООО «ИДМ» (InternationalDesign&Marketing, Ltd.), г. Зеленоград.

Директор.

Разработки в области микроэлектроники и информационных технологий по реализации цифровых, аналоговых, смешанных и радиочастотных микросхем, как для отечественных, так и для зарубежных компаний.

e-mail: Vladimir.Kozlov@idm.ru

Образец ссылки на публикацию:

Ю. С. Затуливетер, Е. А. Фищенко, С. Е. Артамонов, В. А. Козлов. Элементы стратегии опережения и архитектурные предпосылки к созданию однокристалльного ускорителя массовых вычислений общего назначения на базе архитектуры отечественного многопроцессорного компьютера ПС-2000 / Программные системы: теория и приложения: электрон. научн. журн. 2014. Т. 4, № 5(19), с. 3–42.

URL: http://psta.psirras.ru/read/psta2013_5_3-42.pdf

Yu. S. Zatuliveter, E. A. Fishchenko, S. E. Artamonov, V. A. Kozlov. Elements of the strategy of outstripping development and architectural prerequisites for the creation of single-chip accelerator for general-purpose massively-parallel computations based on the architecture of domestic multiprocessor computer PS-2000.

ABSTRACT. The causes, manifestations and industrial problems of structural crisis of microprocessor architectures are analyzed. The state of the market of single-chip computer-accelerators for general-purpose massively parallel tasks is considered. The concept of further development of the domestic highly parallel computers PS-2000, which can be regarded as the first representatives of public multiprocessor systems of GP (General Purpose) class, is proposed. The various features of PS-2000M multiprocessor scalable architecture, as the successor of PS-2000, which possesses of great potential for outstripping development, are considered. A comparison of the PS-2000M architecture with modern single-chip architectures for massively parallel computers is given. (*In Russian*)

Key Words and Phrases: computer crisis, a massively parallel architecture, single-chip computer-accelerators, general purpose massively parallel computing, scalability, strategy of outstripping development.