

© 2012 г. Ю.С. ЗАТУЛИВЕТЕР, канд. техн. наук
Е.А. ФИЩЕНКО, канд. техн. наук
(Институт проблем управления им. В.А. Трапезникова РАН, Москва),
С.Е. АРТАМОНОВ, В.А. КОЗЛОВ
(ООО "ИДМ", Москва, Зеленоград)

К РАЗРАБОТКЕ ОТЕЧЕСТВЕННОЙ ЭЛЕМЕНТНОЙ БАЗЫ ДЛЯ ВЫЧИСЛЕНИЙ С МАССОВЫМ ПАРАЛЛЕЛИЗМОМ. ЭЛЕМЕНТЫ СТРАТЕГИИ ОПЕРЕЖЕНИЯ

Анализируются причины, проявления и индустриальные проблемы внутрикомпьютерного кризиса. Рассматривается развитие однокристальных ускорителей класса GP. Предлагается проект масштабируемой и комплексизируемой архитектуры ускорителя класса GP, который обладает архитектурным потенциалом опережения.

TO THE DEVELOPMENT OF DOMESTIC CHIP COMPONENTS (VLSI) FOR COMPUTING WITH MASSIVE PARALLELISM. ELEMENTS OF THE FORWARD TECHNOLOGY STRATEGY / Yu.S. Zatuliveter, E.A. Fishchenko (V.A. Trapeznikov Institute of Control Sciences, Profsoyuznaya 65, Moscow 117342, Russia, E-mail: zvt@ipu.rssi.ru, fish@ipu.ru), S.E. Artamonov, V.A. Kozlov (International Design & Marketing, Ltd., 124498, Zelenograd, Moscow, Russia, 4806, 5, b. 23, E-mail: Sergey.Artamonov@idm.ru, Vladimir.Kozlov@idm.ru). The causes, symptoms and industrial problems of internal computer crisis are analyzed. The development of single-chip accelerator of the class GP is considered. The project of the scalable and integratable architecture of accelerator of the class GP, which has architecture potential of the forward technology strategy, is proposed.

1. Введение

Достижение значимых позиций на мировом компьютерном рынке – одна из приоритетных задач возрождения отечественного компьютеростроения. На первый взгляд она кажется нерешаемой, но, как далее показывается, 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 разработками, но это предъявляет высочайшие требования к уровням их конкурентоспособности. Особое

значение обретают проекты, направленные на формирование новых рыночных ниш высокотиражной наукоёмкой продукции, которые опираются на превосходящие фундаментальные научные заделы и опыт практических наработок.

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.

2. 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 [1]. Без активного приобщения к поиску адекватных ответов на информационные вызовы 21 века невозможно суверенное участие в неизбежном переделе сфер мирового влияния.

Компьютеростроение стало ключевым фактором социально-экономического развития. Качество жизни прочно связано с уровнем технического развития компьютерных технологий, которые становятся доступными всё большему числу людей. В настоящее время эта связь обретает тотальное влияние на мировую социосферу.

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 сложившейся и быстро растущей зависимости социосферы от компьютерной среды дисбалансы передаются и социосистеме в целом. Стагнация ранее ведущих рыночных ниш в должной мере не компенсируется открытием новых. А это упущенная выгода от недоиспользования быстро растущего рыночного потенциала передовых полупроводниковых технологий.

И чем масштабнее проявления внутрикомпьютерного кризиса, тем больше потери. 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.

Анализ ключевых тенденций развития компьютерной среды и причин нарастающих проявлений внутрикомпьютерного кризиса [2] 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, 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 арифметиче-

ских операций. Производительность в расчёте на транзистор [2] 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 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 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 по закону Мура (удвоение за каждые 1.8-2 года). 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 себестоимости каждого транзистора с ростом степени интеграции не отменяет архитектурного кризиса микропроцессоров, а лишь объясняет 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 лидеров, 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.

На переходе к новому этапу развития компьютеростроения глубинный внутрикомпьютерный кризис выравнивает стартовые позиции. В этом *уникальный* исторический шанс для новых игроков. 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.

3. 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 "*микروпроцессорами*". 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 Pentium I) резервы аппаратного реализуемого параллелизма классической модели последовательного счёта были 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 счёта 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 Pentium I 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, получившего название структурного насыщения микропроцессорной архитектуры [2]¹.

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 довольствоваться увеличением производительности новых поколений одноядерных 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 быстродействие, 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 неведомым 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 накатанные, 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 4ГГц 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 промышленность 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 нерегулярных данных в принципе не позволяет увеличивать количество ядер.

¹ К сожалению, тема идентификации глубинного кризиса, впервые открытая в этой публикации, не получила должного развития в работах других авторов. В дальнейшем кризис прогрессировал инкогнито. Только с появления многоядерных микропроцессоров неотвратимость его действия становится очевидной.

Приведём наглядный пример, который показывает масштабы кризиса структурного насыщения микропроцессорных архитектур. Технология вчерашнего дня 45нм предоставляет на кристалле около 1млрд. 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:

1) ограниченный параллелизм в обменах между ядрами и памятью не позволяет наращивать производительность пропорционально числу ядер, рост производительности прекращается уже на нескольких ядрах.

2) 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 мейнстрим.

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 постнеймановского и постмикропроцессорного, 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 компьютерные 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 вырастают 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 однокристальных компьютеров с высокопараллельными многопроцессорными архитектурами общего назначения.

Главным инструментом наращивания производительности становятся такие однокристальные многопроцессорные архитектуры, которые обладают свойствами масштабируемости и комплексности, обеспечивают высокую вычислительную эффективность на широких классах задач с массовым параллелизмом. Такие архитектуры смогут обеспечивать увеличение производительности пропорционально быстро растущему в глубоком нанометровом диапазоне количеству транзисторов на кристаллах за счёт пропорционального увеличения количества ядер.

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 как эти решения находить.

Отсюда особая роль наиболее удачных достижений в части эффективных параллельных архитектур. Их не так много. Одним из весьма удачных решений, является многопроцессорная архитектура отечественного компьютера ПС-2000 [3], ориентированного на решение широких классов задач с массовым параллелизмом. Это был первый в мире широкодоступный суперкомпьютер, который выпускался большой промышленной серией. Его оригинальная, масштабируемая и комплекслируемая 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 в реального времени, а также в системах обработки больших объёмов данных. По признанию зарубежных специалистов ПС-2000 был одним из наиболее успешных отечественных суперкомпьютеров и одним из первых в мире многопроцессорных компьютеров, который выпускался серийно и использовался в многочисленных и разнообразных сферах [6,7].

Современные многопроцессорные аналоги по сферам применимости – однокристальные 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 (видеокарт) для

ПК. В ходе их развития к середине 00-х сформировался новый рыночный класс однокристалльных многопроцессорных компьютеров – GPU (Graphics Processing Units). Примерно в это же время IBM вывела на рынок однокристалльный ускоритель Cell, с гибридной многопроцессорной архитектурой, заточенный под игровые приставки.

Именно так, спустя 20 лет после индустриальной премьеры ПС-2000, состоялся новый выход высокопараллельный 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. Современный компьютерный рынок – совершенно иная элементная база, иные масштабы сфер потребления.

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 – General Purpose (GP)

Значительная часть ускорителей 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 (40нм, 512 ядер) и Kepler (28нм, 1536 ядер). Однако, следует признать, что системы класса GP, получаемые трансформацией из узкопрофильной многопроцессорной архитектуры GPU, изначально ориентированной на ограниченный набор алгоритмов графической видео обработки, 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 нерегулярным 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.

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 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.

Архитектура ПС-2000 обладает уникальным на фоне современных ускорителей свойством структурной масштабируемости, которое позволяет одновременно с увеличением количества вычислительных ядер пропорционально наращивать пропускную способность встроенной хорошо организованной распределённой памяти, а также свойством комплексируемости как на внури-, так и на межкристальном- уровне. А это – принципиально более высокий потенциал и в использовании растущих возможностей полупроводниковых технологий, и в расширении классов эффективно решаемых задач.

Одна, вместо многих узкопрофильных, масштабируемая, комплексируемая, 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 средствам массового рынка высокопроизводительных компьютеров и программ.

Доступ к современным полупроводниковым технологиям обеспечивает возможности пропорционального повышения вычислительной производительности и пропускной способности памяти, снижение удельного энергопотребления, а также себестоимости производства и эксплуатации. Неудовлетворённый спрос массового рынка на общедо-

ступные компьютерные устройства сверхвысокой производительности в классе GP открывают возможности выхода на многомиллионные тиражи, сопоставимые с тиражами традиционных многоядерных микропроцессоров.

Архитектура ПС-2000, доказавшая в ходе десятилетнего промышленного применения, эффективность своих принципов управления массовым параллелизмом [4], положена в основу нового проекта ПС-2000М [4] как фундамент для построения высококонкурентных однокристалльных многопроцессорных компьютеров общего назначения класса GP (General Purpose).

Главный козырь структурно масштабируемой архитектуры ПС-2000 – 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 сохраняются при переходе к новым поколениям полупроводниковых технологий.

Апробированная широкой практикой модель многопроцессорной архитектуры ПС-2000 не утратила своей актуальности до сих пор. Даже наоборот. 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. Основные цели проекта ПС-2000М

- Разработка на базе архитектуры ПС2000 и создание с применением ПЛИС-технологий прототипа высокоэффективной на задачах с массовым параллелизмом, масштабируемой многопроцессорной архитектуры для последующей однокристалльной реализации на технологиях глубокого нанометрового диапазона.
- Экспериментальное обоснование возможности создания на базе архитектуры ПС2000 конкурентоспособного на мировом рынке однокристалльного компьютера для использования в качестве элементной базы высокопроизводительных вычислительных систем, «облачных» и распределённых вычислительных систем и других компьютерных систем.
- Разработка микросхемы однокристалльного компьютера и организация его массового производства на современной нанотехнологии одной из зарубежных фабрик
- Обоснование целевых рыночных сегментов и создание необходимых условий для вывода конкурентоспособного отечественного однокристалльного компьютера на и мировой рынок компьютерной техники.

СПИСОК ЛИТЕРАТУРЫ

1. *Затуливетер Ю.С.* Информационная природа социальных перемен М.: СИНТЕГ, 2001, 132с. URL: http://www.ipu.ru/sites/default/files/publications/Zatuliveter_InfPrirodaSocPeremen_.pdf
2. . *Затуливетер Ю.С.* Компьютерные архитектуры: неожиданные повороты // Hard 'n' Soft. 1996. № 2. С. 86-94. http://zvt.hotbox.ru/p2_z1.htm.
3. *Затуливетер Ю.С., Фищенко Е.А.* Многопроцессорный компьютер ПС-2000 (Опыт создания и пути развития). Научное издание (Препринт). М.: Институт проблем

управления РАН, 2012. 86с. URL:
http://www.ipu.ru/sites/default/files/publications/препринт_пс-2000_.pdf.

4. *Артамонов С.Е., Затуливетер Ю.С., Фищенко Е.А.* Предпосылки к созданию однокристалльного многопроцессорного компьютера ПС-2000М производительностью 1-10 Tflops // Параллельные вычислительные технологии (ПаВТ'2011): труды международной научной конференции (Москва, 28 марта-1 апреля 2011г.). С.402–410. URL:
http://omega.sp.susu.ac.ru/books/conference/PaVT2011/talks/012_zatuliveter_talk.pdf , Челябинск: Издательский центр ЮУрГУ, 2011. 730с. URL:
<http://omega.sp.susu.ac.ru/books/conference/PaVT2011>
5. *Затуливетер Ю.С.* Введение в проблему параметризованного синтеза программ для параллельных компьютеров М.: ИПУ РАН, препринт, 1993, 89с.
6. *Wolcott P., Goodman S.E.* High-Speed computers of the Soviet Union // IEEE Computer. - 1988. Vol. 21, -No 9. P. 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. P. 21-24.