# ПРОШЛОЕ И НАСТОЯЩЕЕ РОССИЙСКИХ СУПЕРКОМПЬЮТЕРОВ

## ПО МАТЕРИАЛАМ ДОКЛАДОВ ТРЕТЬЕГО НАЦИОНАЛЬНОГО СУПЕРКОМПЬЮТЕРНОГО ФОРУМА

**М.Шейкин** max.shaking@ya.ru

Во второй части обзора пленарных докладов, прозвучавших на третьем Национальном суперкомпьютерном форуме*, расскажем о российских высокопроизводительных вычислительных системах и программном обеспечении для них.

Чем больше ядер участвует в расчетах, тем эффективнее они выполняются. Это утверждение справедливо в отношении небольшого количества ядер, но по мере возрастания сложности и увеличения мощности вычислительного кластера появляются проблемы, которые могут свести на нет преимущества суперкомпьютерных технологий.

Еще в 1967 году выдающийся проектировщик вычислительных систем Джин Амдал сформулировал закон, гласящий: "В случае, когда задача разделяется на несколько частей, суммарное время ее выполнения на параллельной системе не может быть меньше времени выполнения самого длинного фрагмента". Иными словами, линейного роста производительности можно добиться лишь на полностью параллельных задачах. На практике это недостижимо, так как в любом случае некоторые фрагменты задач будут выполняться последовательно. Более того, при наращивании количества ядер производительность многопроцессорной вычислительной системы (МВС) достигает некоего пика, после чего начинает падать (рис.1). Обусловлено это тем, что структура вычислительной системы жесткая, и при большом числе ядер возрастают накладные расходы на управление потоками данных, дают о себе знать неравномерность загрузки ядер, "бутылочные горлышки" при доступе к общей памяти, сети и т.д.

Кроме того, по мере увеличения количества вычислительных модулей проявляется еще одна серьезная проблема. Один терафлопс вычислительной мощности на современной аппаратной базе – это примерно один киловатт мощности и четверть кубометра объема. Кластеры с первых позиций топ-500 нуждаются в десятках мегаватт мощности и занимают сотни кубометров помещений вычислительных центров. К 2020 году прогнозируется появление экзафлопсных МВС, которые потребуют отдельных электростанций для питания и огромных зданий стометровой высоты для их размещения. Столь внушительные параметры означают и непомерно высокие финансовые затраты.

Улучшить характеристики вычислительных систем на один-два порядка и обойти описанные подводные камни позволяет использование проблемно-ориентированной архитектуры. Вычислительные элементы таких систем объединяются в соответствии с поставленной задачей, благодаря чему обеспечивается оптимальный обмен данными между ними.

Понятно, что перестраивать традиционные кластерные системы под каждую задачу – крайне трудоемкое и дорогостоящее занятие. Но сравнительно несложно сделать это на реконфигурируемых вычислителях. Примечательно, что концепция оптимизации систем под конкретные задачи была реализована в той или иной степени еще в середине прошлого века (аналоговые вычислительные машины, цифровые дифференциальные анализаторы). Единственным примером ее серийного воплощения в нашей стране стала многопроцессорная вычислительная система ПС-2000. С 1980 года было выпущено 180 таких машин, которые работали на разных предприятиях и в учреждениях. В частности, вычислительный комплекс на базе восьми ЭВМ ПС-2000 с 1986 по 1997 годы использовался в системе предварительной обработки телеметрической информации Центра управления космическими полетами (ЦУП). Благодаря хорошему распараллеливанию задач эта машина была одной из лучших с точки зрения архитектуры вычислительных систем, созданных в СССР, а также первой в мире серийной высокопроизводительной ЭВМ.

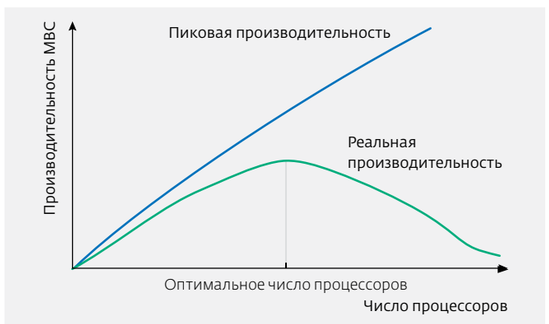


**Рис.1.** Зависимость производительности МВС от числа ядер

Тем не менее отсутствие подходящей элементной базы, позволяющей быстро перестраивать конфигурацию вычислительной системы, сдерживало развитие реконфигурируемых вычислительных систем (РВС). Лишь с появлением микросхем программируемой логики (ПЛИС) концепция гибкой вычислительной сети смогла проявить себя в полной мере. ПЛИС позволяет создавать внутри себя вычислительные элементы необходимой и достаточной для конкретной задачи конфигурации и связать их так, чтобы время передачи информации между узлами было минимизировано, а на обслуживание сети не расходовались лишние ресурсы.

### РВС В РОССИИ

Тема вычислительных комплексов на основе ПЛИС была затронута в трех пленарных и нескольких секционных докладах, прозвучавших на форуме (не это ли свидетельствует об актуальности концепции реконфигурируемых систем?). Об отечественных разработках в этой области рассказывали директор НИИ многопроцессорных вычислительных систем Таганрогского государственного радиотехнического института, доктор технических наук Игорь Анатольевич Каляев и заместитель директора по научной работе НИИ "Квант" (г. Москва) Виктор Станиславович Горбунов. Вычислительные модули на основе ПЛИС также входят в аппаратный комплекс "Грифон", разработанный в ЗАО "НПФ "Доломант", но об этом ниже.

Основа вычислительной системы на ПЛИС – модуль, содержащий одну или несколько плат с микросхемами программируемой логики. Платы

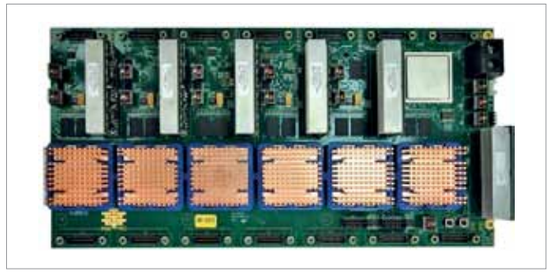

**Рис.2.** Базовая плата "Плеяда"

в рамках одного модуля и модули в составе комплекса объединены в "виртуальную ПЛИС", в едином поле которой и выполняются вычисления. Например, базовая плата ММ475, разработанная в 2011 году в ООО "Научно-исследовательский центр СуперЭВМ и нейрокомпьютеров" (ООО "НИЦ СЭ и НК", г. Таганрог), включает восемь ПЛИС Xilinx Virtex-6 (47,5 млн. вентилей каждая) и позволяет создать до 1,5 тыс. элементарных процессоров IEEE-754 с тактовой частотой 350 МГц. При потребляемой мощности 300 Вт суммарная производительность платы достигает 720 GFlops при вычислениях с одинарной точностью и 340 – с двойной.

Плата следующего поколения "Плеяда" (рис.2) включает шесть ПЛИС Virtex-7. Будучи значительно компактнее платы ММ475, "Плеяда" столь же производительна. На основе этих плат созданы вычислительные модули "24 V7–750" (рис.3) производительностью 2,78 TFlops при потребляемой мощности 1400 Вт.

В свою очередь, модули "24 V7–750" стали основой вычислительной машины РВС-7 – одной из последних разработок ООО "НИЦ СЭ и НК".

**Рис.3.** Базовый модуль "24 V7–750", содержащий четыре платы "Плеяда"

* См.: Электроника: НТБ, 2015, № 1, с. 132–144.

**Таблица 1.** Характеристики суперкомпьютера "Ломоносов" и РВС

| Суперкомпьютер | "Ломоносов" | РВС-7 |
|---|---|---|
| Количество стоек | 26 | 3 |
| Занимаемая площадь, м$^2$ | 252 | <5 |
| Энергопотребление, кВт | 2800 | 90 |
| Стоимость, млн. руб. | 2670 | 360 |
| Производительность, TFlops | 137* | 160 |

* Средняя производительность в задачах с интенсивным обменом данными по состоянию на ноябрь 2014 года.

Производительность одной стойки, включающей 36 модулей, достигает полутора петафлопс при максимальной потребляемой мощности всего 50 КВт.

Для доказательства преимуществ РВС перед традиционными суперЭВМ И.А.Каляев привел пример обескураживающего падения производительности лучшего отечественного суперкомпьютера при решении задач с интенсивным обменом данных (табл.1). В таких случаях три стойки РВС-7 справляются с вычислительными задачами не хуже "Ломоносова"! Столь красноречивые цифры говорят сами за себя.

Еще одна интересная разработка ООО "НИЦ СЭ и НК" – персональный реконфигурируемый компьютер "Калеано" (рис.4), предназначенный для обработки данных, поступающих по сети Gigabit Ethernet без поддержки IP-протоколов. Основа "Калеано" – вычислительное поле из шести ПЛИС


**Рис.4.** Реконфигурируемый компьютер "Калеано"

и управляющая ЭВМ Contron COM-Express, в задачи которой входят ввод/вывод данных, подготовка и отладка программы для вычислительного поля. ПЛИС соединены между собой каналами LVDS, к каждой подключен модуль динамической памяти емкостью 256 МБ. ЭВМ "Калеано" выпускается в двух модификациях: "Калеано-К" на базе ПЛИС Kintex-7 XC7K160T и "Калеано-V" на ПЛИС Virtex-7 (табл.2).

В докладе В.С.Горбунова рассказывалось о реконфигурируемой моделирующей вычислительной системе (МГВС), созданной в НИИ "Квант". Ее отличие от описанных выше ЭВМ на ПЛИС – изначальная ориентация на определенную задачу: МГВС предназначена для моделирования суперкомпьютеров экзафлопсного класса (с числом ядер до нескольких миллионов) и средств их программирования. Архитектура МГВС не отличается от аналогичных вычислительных систем на ПЛИС,

**Таблица 2.** Характеристики настольных реконфигурируемых компьютеров "Калеано"

| ЭВМ | "Калеано-К" | "Калеано-V" |
|---|---|---|
| Число и тип ПЛИС | 6 Kintex-7 | 6 Virtex-7 |
| Общее количество эквивалентных вентилей в ПЛИС вычислительного поля, млн. | 96 | 288 |
| Объем оперативной памяти, Гб | 3 | |
| Производительность вычислительного модуля, одинарная/двойная точность, Гфлопс | 150/75 | 440/220 |
| Тактовая частота, МГц | 330 | 400 |
| Скорость обмена данными по каналу Ethernet, Гбит/с | 1 | |
| Частота передачи данных по LVDS между ПЛИС вычислительного поля, МГц | 900 | 1200 |
| Потребляемая мощность, Вт | 200 | 320 |
| Габаритные размеры, мм | 480×270×70 | |
| Стоимость, млн. руб. | 1,3 | 2 |





**Таблица 3.** Производительность реконфигурируемых суперЭВМ ООО "НИЦ СЭ и НК"

| Изделие | Тип ПЛИС | Производительность вычислительной платы (одинарная/двойная точность), GFlops | Производительность вычислительного модуля (одинарная/двойная точность), GFlops | Производительность стойки, двойная точность, TFlops |
|---|---|---|---|---|
| "Орион-5" 2009 г. | Virtex-5 | 250/85 | 1000/340 | 19,2–28,8 |
| "Ригель" 2012 г. | Virtex-6 | 400/125 | 1600/500 | 34,5–51,8 |
| "Тайгета" 2013 г. | Virtex-7 | 900/300 | 3600/1200 | 68–100 |
| "Скат" ~2016 г. | UltraScale | 7250/2500 | 82500/30000 | 1000–1250 |

однако докладчик особо подчеркнул весьма привлекательные перспективы применения в качестве основы РВС новых ПЛИС Xilinx серии UltraScale (US). Их преимущества – в два с лишним раза большее количество ячеек, сниженное энергопотребление при значительно меньшей стоимости по сравнению с аналогами предыдущего поколения.

На основе ПЛИС Kintex-US (XCKU040/060/075) в НИИ "Квант" был создан вычислительный модуль для МГВС "Топаз-3". Восемь ПЛИС модуля через коммутатор объединены каналами PCI Express 1.0 или 2.0. Функции конфигурирования и управления вычислительным полем выполняет отдельная ПЛИС Virtex-6. Взаимодействие плат и объединение систем обеспечиваются через коммутатор, оснащенный двумя внешними портами PCI-Express.

ПЛИС UltraScale рассматривались и в качестве следующей ступени развития компактных реконфигурируемых ЭВМ "Калеано". Предполагается, что новая компонентная база повысит производительность системы почти вдвое, при этом энергопотребление возрастет всего в 1,3 раза.

В ООО "НИЦ СЭ и НК" разрабатывается также концепция перспективного РВС с погружным охлаждением. Создана плата вычислительного модуля "Скат-8" на основе ПЛИС Virtex UltraScale (100 млн. эквивалентных вентилей каждая), потребляющая 800 Вт энергии. В вычислительном модуле высотой 3U размещаются 16 таких плат, погруженных в электрически нейтральный жидкостный хладагент. Приток и охлаждение хладагента обеспечивают насосная группа и теплообменник, установленные в каждом модуле. Производительность стойки с 12 модулями "Скат-8" достигает 333 TFlops при потреблении всего 144 КВт. При столь же красноречивых цифрах многословные комментарии излишни. Стоит лишь обратить внимание на резкое повышение производительности перспективных РВС на современной компонентной базе (табл.3).

## ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ РВС НА ОСНОВЕ ПЛИС

Что же препятствует широкому внедрению реконфигурируемых вычислительных систем? В отличие от традиционных компьютеров и кластеров, при работе с РВС на основе ПЛИС требуется разработка не только программного алгоритма вычислений, но и конфигурации системы в вычислительном поле. Программирование РВС выполняется в два этапа: схемотехник разрабатывает структуру под конкретную задачу, а затем прикладной программист создает параллельную программу, определяющую потоки данных в структуре. Для программирования ПЛИС обычно применяются среды разработки типа Xilinx ISE, Altium Designer и др., не рассчитанные на создание проекта в нескольких ПЛИС. Поэтому традиционные способы создания конфигурации РВС на ПЛИС очень трудоемки – работа может растянуться на месяцы. Чтобы не отпугнуть пользователей высокопроизводительных ЭВМ такими сложностями, "делом чести" для создателей РВС становится разработка программного обеспечения, позволяющего программистам и инженерам преодолеть этот барьер.

Для упрощения работы с РВС ООО "НИЦ СЭ и НК" предлагает комплекс программного обеспечения на основе языка программирования COLAMO, созданного в 1987 году И.И.Левиным (НИИ МВС ЮФУ) – одним из соавторов доклада И.А.Каляева. Язык COLAMO предназначен для описания вычислительной структуры РВС в виде фрагментов информационного графа задачи, каждый из которых является вычислительным конвейером потока операндов. Комплекс включает в себя:

- транслятор программы на COLAMO в информационный граф параллельной прикладной программы;
- синтезатор масштабируемых схемотехнических решений на уровне логических ячеек ПЛИС Fire! Constructor, отображающий полученный информационный граф на архитектуру РВС, размещающий его по кристаллам ПЛИС и автоматически синхронизирующий фрагменты графа в разных кристаллах;
- библиотеку IP-ядер, соответствующих операторам языка COLAMO.

Таким образом, решения отечественных разработчиков аппаратного и программного обеспечения позволяют применять РВС на основе ПЛИС для различных прикладных вычислений. Благодаря высокой вычислительной мощности вкупе с экономичностью и относительно низкой стоимостью РВС можно рассматривать в качестве кандидата на получение статуса "компьютеры нового поколения", которые заменят существующие кластерные системы. Единственная преграда на пути их распространения – более сложный, можно сказать, непривычный процесс программирования алгоритмов для реконфигурируемых систем.

## МОДУЛЬНАЯ ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА "ГРИФОН"

Российская компания "НПФ "Доломант" представила на форуме одну из своих последних разработок – модульную высокопроизводительную вычислительную систему "Грифон", предназначенную для встраиваемых применений. О ней рассказал Петр Владимирович Галаган, заместитель технического директора ЗАО "НПФ "Доломант".

В начале своего выступления докладчик отметил, что высокопроизводительные вычисления – стратегическая составная часть политики любого государства в области информационных технологий. В условиях все большей роботизации и виртуализации как вполне мирных производственных процессов, так и военной техники возможность быстрых вычислений становится столь же необходимой, как и точные инструменты и оружие. А в условиях санкционных запретов на поставки не только ряда компонентов, но и технологических решений перед российскими производителями встает задача разработки отечественных высокопроизводительных вычислительных систем.

Гетерогенная вычислительная платформа "Грифон" предназначена для работы в жестких условиях окружающей среды (стойкость к воздействию внешних факторов в соответствии с ГОСТ РВ 20.39.304–98). Ударопрочное исполнение позволяет применять ее для создания практически любых встраиваемых систем, в том числе бортовых.

При создании системы "Грифон" разработчики руководствовались тремя принципами: компактности, модульности и гетерогенности. В результате на свет появилась очень гибкая, универсальная и надежная вычислительная система.

К габаритам и массе встраиваемых систем предъявляются жесткие требования. Вычислительный блок "Грифон" собран в корпусе стандартной высоты 3U, что позволяет как устанавливать его в стандартные стойки, так и встраивать в ограниченные пространства транспортных средств (рис.5). В зависимости от условий эксплуатации можно выбирать систему с кондуктивной, принудительной воздушной или жидкостной системой охлаждения.

Чтобы обеспечить универсальность системы, то есть возможность гибкой настройки под конкретные задачи, было решено сделать ее модульной. В каждый вычислительный блок устанавливается до девяти модулей, в частности:

- CPC510/512 – модуль центрального процессора Intel Core i7;
- VIM556 – модуль графического процессора nVidia Quadro;
- FPU500 – модуль реконфигурируемого вычислителя на базе ПЛИС Xilinx Virtex-6/7;
- модули коммутации PCI-E, сетевые контроллеры Ethernet и т.д.

В качестве сети межмодульного взаимодействия применяется шина открытого стандарта

**Рис.5.** Вычислительный модуль системы "Грифон"





PCI Express 3.0. Специально для "Грифона" разработаны сетевые свитчи с пропускной способностью до 640 Гб/с. Возможно соединение вычислительных модулей по принципу "каждый с каждым" аналогично связям в суперкомпьютерных кластерах; это позволяет создавать на основе "Грифона" высокопроизводительные вычислительные системы. С помощью сетевых адаптеров 10 Gigabit Ethernet можно объединять в сеть несколько блоков.

Архитектура системы "Грифон" позволяет комбинировать модули различных типов так, как это требуется для решения конкретной задачи. Иными словами, система может быть гетерогенной, сочетающей процессоры x86, графические ускорители и вычислительные поля на ПЛИС. На основе "Грифона" можно создавать универсальные вычислительные системы, РВС, а также их комбинации, подобные описанному выше реконфигурируемому компьютеру "Калеано".

Как и в случае с РВС на основе ПЛИС, без соответствующего программного обеспечения новая система не получит широкого распространения. В докладе П.В.Галагана подчеркивалась важность разработки программных компонентов системы вкупе с аппаратными – пользователи не должны испытывать серьезных трудностей при освоении новых вычислительных средств. Созданный для "Грифона" комплекс программного обеспечения упрощает работу с гетерогенными аппаратными модулями системы. Ее пользователям не обязательно изучать механизмы низкоуровневого взаимодействия блоков для всех их сочетаний. Так, микропроцессорное межмодульное взаимодействие обеспечивается сетевым драйвером с транспортом IP по шине PCI Express, библиотеками сокетов с транспортом по PCI Express и отображением участков памяти одного модуля на другой. Для работы с массивом ПЛИС написаны специальные драйверы, а графические процессоры поддерживают знакомый программистам CUDA SDK.

Докладчик отметил открытость архитектуры системы "Грифон". Любой производитель может самостоятельно создавать модули с интерфейсом открытого стандарта PCI Express. Это выгодно отличает "Грифон" от аналогов типа MicroTCA, VPX и др., полностью или частично основанных на закрытых интерфейсах. Доступность спецификаций интерфейса исключает несовместимость модулей, а также ситуации, когда разработчик системы переходит на новые стандарты и прекращает поддержку изделий предыдущих серий. Более того, "НПФ "Доломант" предлагает всем заинтересованным в развитии "Грифона" любые формы сотрудничества и приветствует создание сторонними разработчиками новых модулей для системы.

Применение в качестве основного интерфейса шины PCI Express имеет и в некотором роде стратегическое значение. PCI Express – широко распространенный стандарт, поддерживаемый многими производителями элементной базы по всему миру. Поэтому проблем с подбором компонентов не возникнет даже при жестких ограничениях на поставки, вполне возможных при ухудшении политического климата. А на волне актуальной сегодня темы импортозамещения начались работы по созданию модуля для системы "Грифон" на основе процессоров отечественного производства "Эльбрус".

В конце выступления П.В.Галаган подчеркнул, что качественные и функциональные характеристики разработанной и произведенной в России системы "Грифон" полностью удовлетворяют требованиям мирового рынка встраиваемых решений, а в некоторых случаях даже превосходят их.

## ПАКЕТ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ FLOW VISION

Говоря о программном обеспечении для высокопроизводительных ЭВМ, нельзя не упомянуть отечественный пакет для моделирования физических процессов Flow Vision, первая версия которого была представлена в 1991 году. Его создатели, сотрудники Института автоматизации проектирования и Института математического моделирования РАН (г. Москва), поставили перед собой задачу разработать инструмент для выполнения сложных расчетов в различных областях машиностроения. Flow Vision разрабатывался под конкретные проекты и на средства, выделяемые заказчиками; одним из них была корпорация "РКК Энергия", которая использовала пакет для расчетов в рамках проекта Sea Launch (рис.6).

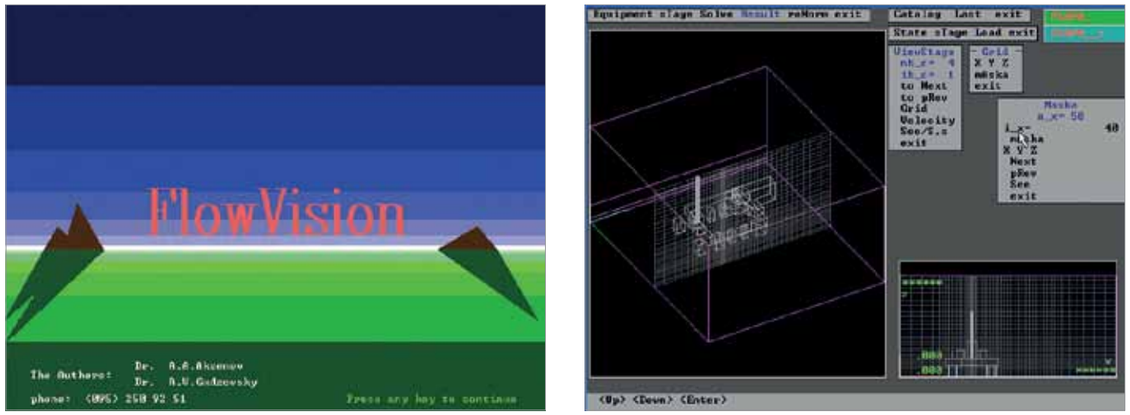


**Рис.6.** Заставка первой версии Flow Vision и окно расчета для Sea Launch

Написанный на Фортране Flow Vision 1.0 работал в MS-DOS – самой популярной в то время вычислительной среде в России. Уже тогда в пакете была реализована параллельность вычислений для транспьютерных систем – всего 128 процессов! Особенностью первой версии программы стало применение ступенчатой сетки на границах, что позволило снизить количество расчетов.

В 1994 году команда создателей Flow Vision переходит в компанию "ТЕСИС", где продолжается работа над проектом. Вторая версия Flow Vision, созданная на языке C++, работала в ОС Windows 95. В 2000 году выходит коммерческая версия пакета, и постепенно Flow Vision становится известен за рубежом. В частности, в 2003 году были заключены многолетние контракты с крупнейшим американским производителем шин Goodyear и шведской машиностроительной компанией Atlas Copco. Тогда же компания Dassault встроила средства Flow Vision в среду математического моделирования Abaqus FEA, что закрепило за отечественной разработкой статус продукта мирового уровня. Конечно, не оставались в стороне и отечественные заказчики – НИКИЭТ, "РКК Энергия" и т.д. О качестве второй версии Flow Vision можно судить по тому, что, несмотря на прекращение ее разработки в 2010 году и поддержки – в 2012-м, она все еще используется в различных организациях.

Одновременно с ростом популярности второй версии Flow Vision велась разработка нового пакета с тем же названием, но иной идеологии. В 2006 году был представлен Flow Vision HPC, рассчитанный на многопроцессорные вычисления. Это главное его свойство означало, что из "обычной" программы для персонального компьютера Flow Vision превращается в глобальный и, возможно, стратегический проект. С выходом третьей версии Flow Vision разработчики (наконец-то!) получили и государственную поддержку в рамках различных федеральных целевых программ.

В основу обновленного Flow Vision были положены принципы параллельных вычислений на всех этапах выполнения алгоритма, кроссплатформенности и масштабируемости. Flow Vision 3 стал универсальной платформой, на базе которой можно было создавать расчетное программное обеспечение любого назначения. Нужно особо отметить междисциплинарность пакета – на стыке многих научных дисциплин появляется возможность моделировать физические процессы так близко к реальности, как это только возможно в рамках математических моделей.

Из прочих вех истории Flow Vision стоит отметить интеграцию в 2010 году решателя Flow Vision в программный комплекс Autodesk CFDesign, предназначенный для выполнения гидродинамических задач, и сотрудничество с ВНИИЭФ в разработке программного комплекса ЛОГОС.

Рассказ о российском программном продукте не случайно был выбран в качестве завершающего. Увы, несмотря на огромный интеллектуальный потенциал России, еще очень долго успехи нашей страны в суперкомпьютерной отрасли будут измеряться тем, насколько удалось сократить отставание от ведущих стран. На равных сотрудничать и конкурировать с зарубежными разработчиками отечественные компании могут лишь в области создания программного обеспечения, и история пакета Flow Vision – отличный тому пример.