Применение HPC-систем в промышленности и научных исследованиях

Публикация продолжает тему (SN №№ 1/30, 2007; 2/31, 2007) – использование HPC-систем – одну из определяющих для развития всей IT-отрасли и представляет обзор перспектив ее развития в ведущих отраслях промышленности. О применении HPC-систем рассказывают руководители и ведущие специалисты ФГУП "ММПП "Салют", НИИ эпидемиологии и микробиологии им. Н.Ф. Гамалеи РАМН и др.

ФГУП "ММПП "Салют": без развития методов численного моделирования поддержание конкурентоспособности газотурбинной отрасли России уже невозможно

Елисеев Д.Н., Дружинин Е.А.

ФГУП "ММПП "Салют" является одним из крупнейших государственных предприятий авиационной промышленности в части разработки и производства авиационных двигателей для военной и гражданской авиации, высокоэкономичных энергетических установок. Предприятие обладает уникальной научно-производственной базой. В настоящее время на предприятии идут работы по изготовлению и сервисному обслуживанию авиадвигателей для новейших военных самолетов семейства Су, модификации и ремонту двигателей самолетов Су и МиГ, изготовлению узлов и деталей авиационных двигателей, устанавливаемых в самолетах конструкторских бюро Туполева, Антонова, Яковлева. Большое внимание уделяется производству промышленных энергетических установок. ФГУП "ММПП "Салют" занимается также изготовлением различных товаров общепромышленного и народного потребления.

Основным зарубежным заказчиком по линии военно-технического сотрудничества является Китайская Народная Республика.

Дополнительно ММПП "Салют" изготавливает и поставляет комплектующие и узлы для гражданских авиадвигателей для таких фирм, как "Пратт-Уитни" (Канада), МТУ (Германия), "Снекма" (Франция).

Основным заказчиком энергетических установок выступают ОАО "Газпром", РАО ЕЭС России и другие энергетические компании России и стран СНГ.

11 августа 2007 г. Президент РФ подписал Указ о создании интегрированной структуры на базе ФГУП "ММПП "Салют". Целью создания новой структуры является концентрация интеллектуальных и производственных ресурсов, для реализации перспективных программ в области газотурбостроения, в том числе создания авиадвигателей нового поколения. Основным компонентом стратегии развития ФГУП "ММПП "Салют" является программа создания единого информационного пространства, внедрение информационных технологий (ИТ) на всех этапах конструирования, производства и эксплуатации авиационных и промышленных двигателей. В настоящий момент ФГУП "ММПП "Салют" ведет последовательное создание консолидированного вычислительного центра масштаба, соответствующего образованной структуре.

В конце 2002 г. на предприятии начат проект по внедрению кластерных технологий в инженерном анализе и подготовке существующей инфраструктуры для ее успешного применения в конструкторских подразделениях. В феврале 2003 г. проведен запуск в промышленную эксплуатацию системы распределенных вычислений, таким образом, предприятие получило первый в отрасли кластер на платформе Intel – XEON. В настоящий момент на предприятии функционируют несколько специализированных комплексов на базе процессоров Intel XEON и Intel Itanium 2 совместно с системой хранения конструкторской документации общей емкостью 35 Тбайт. В августе 2007 г. приобретен кластер на базе новейших четырехядерных процессоров Intel® Xeon® серии 5300 (Clovertown) с частотой 3 ГГц. Такое решение продемонстрировало наилучшие результаты, как по абсолютной производительности, так и по соотношению цена/производительность комплекса в целом. Также достигнута необходимая предприятию производительность вычислительного комплекса, равная 4,5...5 Терафлоп. Производительность приобретенного кластера соответствует 4-й позиции в текущем списке 50 наиболее мощных вычислительных систем СНГ.

В настоящий момент предприятие работает над созданием авиационного двигателя нового поколения. Потребность России в газотурбинных двигателях постоянно растет. Поскольку сферы применения такой продукции не ограничиваются военной и гражданской авиацией, энергетика, нефтегазодобыча, железнодорожный транспорт и судостроение ожидают разработки газотурбинных силовых установок нового поколения. Качественное улучшение эксплуатацион-

ных характеристик новых двигателей можно получить, только создав полную и точную математическую модель объекта разработки. Только проведение математического моделирования процессов, проходящих в двигателе в целом, позволит достичь поставленной цели. Ранее для решения таких задач существующих вычислительных мощностей завода было недостаточно. На мощных рабочих станциях инженеры вынуждены ожидать результатов расчетов в течение нескольких суток, либо проводить детерминированные расчеты объекта разработки.

Если рассматривать процесс создания авиационного двигателя, проходивший в 70-80 годы, то можно обратить внимание на очень длительный проведения научно-исследовательских и опытно конструкторских работ, доходивших до 10 лет. Количество изготовленных экземпляров опытных образцов двигателей в некоторых случаях было более 40 шт. Только после изготовления такого числа опытных машин заказчик получал серийный образец продукции. Финансирование разработок вело государство и основной задачей было получение рабочего двигателя. В настоящий момент положение на рынке газотурбинной техники таково, что если время реализации проекта затягивается, то конкурентоспособность изделия будет потеряна. В изменившейся ситуации методы разработки были пересмотрены. Создание дорогостоящих опытных или стендовых экземпляров необходимо свести к минимуму. В идеальном случае первый изготовленный в металле экземпляр продукции должен быть передан заказчику. Сейчас предприятие обладает уникальным опытом проведения разработки новой продукции и испытания ее численными методами. Численный метод испытаний заключается в расчете или моделировании основных характеристик объекта разработки с помощью специализированного программного обеспечения инженерного анализа. Таким образом, проведение многочисленных испытаний новой продукции на начальном этапе возможно БЕЗ изготовления материальной части.

Применение кластерных вычислительных технологий в процессе создания авиационных и газотурбинных двигателей позволяет повысить конкурентоспособность изделия путем существенного сокращения времени разработки и экономии материальных ресурсов, а также за счет количества опытных образцов изделий и натурных испытаний. Суммарная экономия составляет миллионы долларов и тысячи человеко-часов.

Выбор подходящей архитектуры вычислительного комплекса, его монтаж, наладку можно предоставить крупной IT-компании-интегратору. Но в то же время необходимо учитывать, что архитектура будущего вычислительного комплекса является объектом многокритериальной оптимизации. Специалисты предприятия при выборе архитектуры описали около 20 критериев оптимизации. Среди таких очевидных, как электрическая мощность, тепловыделение, были определены приоритетные классы решаемых задач – газовая динамика, тепломассообмен, динамическая прочность и связанные между собой перечисленные дисциплины. Также на выбор архитектуры влияет и геометрическая конфигурация рассчитываемых моделей. Оптимизация комплекса велась с привлечением специалистов европейского центра высокопроизводительных вычислений компании Intel.

Сегодня основная задача, которую необходимо решать предприятию – создание необходимой инфраструктуры для эффективного функционирования кластера и создания необходимой нормативной базы для решения конструкторских задач. На текущий момент не существует проработанных типовых решений таких задач и сотрудникам предприятия приходится решать их самостоятельно. Это естественно, т.к. любое крупное или среднее машиностроительное предприятие имеет уникальную организацию бизнес-процессов и внутреннюю структуру. Для внедрения технологии расчетов с применением кластеров необходимо тщательное описание методик подготовки электронных моделей, проведения самих расчетов. На нашем предприятии работа инженеров-конструкторов, инженеров-расчетчиков ведется с системой управления конструкторским документооборотом (PDM). Использование PDM позволяет сформировать на предприятии непротиворечивую базу электронных моделей и архива конструкторской документации, поддерживать ее в актуальном состоянии.

Следующим этапом развития инженерных вычислений на предприятии является проработка методов многоцелевой оптимизации конструкции изделия.

НИИ эпидемиологии и микробиологии им. Н.Ф. Гамалеи РАМН: использование суперкомпьютеров мощностью 100 Тфлоп/с и более при разработке новых активных наноструктур позволяет экономить сотни млн долл.

Карягина А.С., Ройтберг М.А.

Математизация и компьютеризация биомедицинских исследований развивается быстрыми темпами, начиная с 70-х годов XX века – как в нашей стране, так и во всем мире. По мере роста компьютерных мощностей возрастала и сложность решаемых задач: в каждый момент времени потребности задач, стоявших на повестке дня, несколько опережали возможности компьютеров. Так же обстоит дело и сейчас, когда биологам стали доступны суперпроизводительные компьютеры с быстродействием в сотни терафлопс. Параллельно с развитием компьютеров развивались и используемые в биомедицинских исследованиях программные средства; это относится как к развитию алгоритмов обработки данных, так и к организации пользовательского интерфейса.

Первые применения математико-компьютерных методов связаны с развитием математического моделирования. Различные явления – такие, как иммунитет, взаимодействие организмов в различных биоценозах, свертывание крови и многие другие – могут быть описаны с помощью систем обыкновенных дифференциальных уравнений, которые, в свою очередь, могут быть исследованы с помощью компьютеров. Первые модели, по необходимости, описывали изучаемую систему с помощью небольшого (2–4) количества величин. Развитие компьютеров и появление суперкомпьютеров сделало возможным исследование моделей, содержащих сотни и даже тысячи переменных. Примером таких моделей являются используемые в молекулярной биологии и, в частности, при разработке лекарств, расчеты методами молекулярной динамики. В этих расчетах индивидуально определяется движение каждой из исследуемого ансамбля молекул.

Модели, оперирующие небольшим количеством переменных, сохраняют свое значение при исследовании явлений с качественной точки зрения. Как правило, такие модели исследуются на персональных компьютерах, причем отдельные расчеты могут требовать сутки и более машинного времени. Применение суперкомпьютеров в этом случае позволит

Карягина А.С. – доктор биол. наук, вед. научн. сотр. лаборатории биологически активных наноструктур НИИ эпидемиологии и микробиологии им. Н.Ф. Гамалеи РАМН.

свести время таких расчетов к секундам. В этом случае можно будет вести работу в диалоговом режиме, оперативно исследовать поведение изучаемой системы при различных значениях параметров и, как следствие, эффективность научной работы существенно увеличится.

В последние десятилетия наиболее быстроразвивающейся областью биологии является молекулярная биология, имеющая многочисленные биотехнологические и медицинские приложения (генная инженерия, генная терапия, молекулярная диагностика, разработка новых лекарств и пр.). Компьютеры весьма широко применяются в молекулярно-биологических исследованиях – как фундаментальных, так и прикладных.

Прежде всего, это относится к созданию и использованию молекулярно-биологических баз данных. Наибольшее значение (по объему и частоте использования) имеют базы, содержащие первичные структуры биополимеров (ДНК и белков) – символьные последовательности, где символы изображают соответствующие мономеры. Объем нуклеотидной базы данных – десятки миллиардов

символов, в том числе – полностью расшифрованные последовательности геномов человека (около 3 млрд нуклеотидов), мыши, арабидопсиса, риса и ряда других организмов (сотни бактерий и десятки высших организмов). База данных последовательностей белков включает последовательности сотен тысяч белков общей длиной около 100 млн символов (аминокислот).

Основной тип запросов к этим базам – поиск последовательностей, сходных с данной. Работа с этими базами требует больших компьютерных мощностей. При этом доступность суперкомпьютеров позволит использовать относительно более времяемкие, но и более чувствительные алгоритмы поиска (в настоящее время для поиска используются эвристические алгоритмы, основанные на предварительном индексировании баз данных). Задача поиска в базах данных, допускающая высокую степень распараллеливания, – естественная область применения современных суперкомпьютеров.

Единая база данных, содержащая пространственные структуры белков, ДНК, их комплексов между собой и с другими молекулами – лигандами (PDB) – еще один важнейший источник данных, используемых в молекулярно-биологических исследованиях. Эти данные играют решающую роль в понимании функционирования белков и механизмов патогенеза, обусловленных нарушением структуры биологических макромолекул. Данные о пространственной структуре белков лежат в основе современной технологии разработки новых лекарственных средств, включающей в качестве одного из этапов так называемый компьютерный (computer aided) drug design. В настоящее время пространственная структура экспериментально определена для небольшой части белков (для белков человека, по разным оценкам, – 5–10%), однако в последние годы темпы работ по определению новых пространственных структур существенно выросли. Отметим, что само получение координат атомов по данным рентгеноструктурного анализа или экспериментов на основе ядерного магнитного резонанса – сложная (и до конца не решенная) алгоритмическая задача, требующая значительных компьютерных мощностей.

В последнее время широко используются и другие базы данных – по метаболическим путям, внутривидовым мутациям и их связям с различными болезнями, по научной литературе и др.

Практически все исследования, проводимые в молекулярной биологии, медицинской и общей генетике, биохимии и т.п., используют те или иные из перечисленных выше баз данных и, следовательно, невозможны без применения высокопроизводительной вычислительной техники. При этом для ряда задач анализ данных о первичных и пространственных структурах биополимеров играет основную роль. К таким задачам, в частности, относятся:

1. Распознавание участков геномов, кодирующих белки.

2. Сравнительный анализ последовательностей белков и нуклеиновых кислот.

3. Расшифровка пространственной структуры белков и их комплексов методами рентгеноструктурного анализа (РСА) и ядерного магнитного резонанса (ЯМР).

4. Сравнительный анализ пространственных структур белков и их комплексов.

5. Предсказание пространственной структуры белков, для которых известна лишь последовательность образующих их аминокислотных остатков (фолдинг).

6. Моделирование структуры биологических макромолекул.

7. Компьютерный drug design.

Первые шесть задач – одни из ключевых задач фундаментальной биологической науки, имеющие при этом непосредственный выход в практическую область, связанную с диагностикой и терапией многих заболеваний наследственного и инфекционного характера. Существо этих задач и компьютерные ресурсы, необходимые для их решения, будут кратко рассмотрены ниже. Седьмую задачу, успешная реализация которой имеет огромное значение для разработки новых лекарственных средств, мы рассмотрим более детально.

Распознавание генов – участков геномов, кодирующих белок – наибольший интерес представляет для эукариотических (имеющих клеточное ядро) организмов (для прокариотических – не имеющих ядра – организмов ее можно считать практически решенной). Участки геномов эукариот, соответствующие одному белку, имеют сложную структуру, состоящую из чередующихся фрагментов, кодирующих фрагменты белка и ничего не кодирующих. Дополнительная сложность состоит в том, что один и тот же ген часто кодирует одновременно несколько альтернативных вариантов белка. Вместе с тем, решение этой задачи является ключевым моментом при аннотировании (описании генной структуры) геномов. Следующий этап аннотирования связан с предсказанием (с последующей экспериментальной проверкой) функций выявленных генов. При этом основным является метод, указанный в пункте 2 обсуждаемого списка – *сравнительный анализ последовательностей белков и нуклеиновых кислот*. Каждый вновь полученный фрагмент генома, каждую аминокислотную последовательность предсказанного белка нужно сравнить со всем массивом данных, полученных ранее и хранящихся в специализированных базах данных, с целью выявления похожих последовательностей. Помимо задач аннотации, при этом получают важнейшую информацию для медицинской практики. Примером результата такого рода исследований, имеющего огромную значимость для медицины, является обнаружение с помощью сравнительного анализа генов мутаций (изменений или нарушений в последовательности ДНК и белков), ассоциированных с наследственными заболеваниями.

Объем информации, подвергающейся анализу при вышеописанных исследованиях, очень большой (для сравнения: длина генома человека и млекопитающих – миллиарды нуклеотидов) и постоянно возрастает. Имеются данные о сотнях тысяч генов (последовательности некоторых генов независимо определены несколько раз разными группами исследователей и для разных организмов). Комбинаторные задачи, решаемые при сравнении белковых последовательностей и последовательностей ДНК, отличаются по сложности вычислений ввиду того, что белки имеют 20-буквенный код, вместо 4-буквенного кода последовательностей ДНК. Сложность вычислений при множественном сравнении последовательностей и сравнении длинных последовательностей возрастает на порядки. Распараллеливание вычислений и использование суперкомпьютеров позволит в сотни и тысячи раз повысить эффективность решения этих важнейших задач современной биологии. Для решения таких задач требуются суперкомпьютеры производительностью в сотни терафлопс.

Расшифровка пространственной структуры белков и их комплексов – важнейшая задача современной биологии, лежащая в основе понимания структуры и механизма действия белков. В настоящее время она решается в основном методами рентгеноструктурного анализа

Ройтберг М.А. – канд.физ.-мат.наук., заведующий лабораторией прикладной математики Института математических проблем биологии РАН.

и ядерного магнитного резонанса. Помимо сложных экспериментальных процедур, она включает использование специального математического аппарата и высокопроизводительных компьютеров. Объем и сложность вычислений значительно возрастают при анализе комплексов, включающих несколько макромолекул, представляющих особый интерес для биологии и медицины. Отдельной вычислительной задачей, требующей больших мощностей, является *сравнительный анализ полученных пространственных структур со структурами, расшифрованными ранее*. В настоящее время суперкомпьютеры мало используются при решении задач определения пространственных структур и их сравнения. Однако по мере увеличения сложности исследуемых молекулярных комплексов и разработки более изощренных (но и требующих больших ресурсов) математических методов, роль суперкомпьютеров будет возрастать.

Экспериментальные методы расшифровки пространственной структуры белков крайне сложны и дорогостоящи. Кроме того, для целого ряда белков они непри-

менимы. В связи с этим теоретический подход к получению информации о структуре белков – пространственное сворачивание белков, для которых известна лишь последовательность образующих их аминокислотных остатков (фолдинг) – представляется очень перспективным. Однако решение такой задачи "в лоб" связано с поиском глобального минимума функции десятков тысяч переменных и не может быть найдено даже с использованием самых современных суперкомпьютеров. Обходные пути решения этой проблемы связаны с построением структуры белка на основе данных о структурах похожих белков, полученных экспериментально, с последующим уточнением структуры. Таким образом, на современном этапе развития структурной биологии решение задачи фолдинга фактически сводится к моделированию структуры биологических макромолекул на основе известных гомологов. Даже такой "упрощенный" способ решения проблемы требует применения высокопроизводительной компьютерной техники. Вычислительные эксперименты, связанные с моделированием структуры биологических макромолекул требуют огромных вычислительных мощностей. Расчет одной структуры может занимать сутки и месяцы. Существенное ускорение может быть достигнуто при переходе к параллельным вычислениям.

Рассмотренные выше методы биологии – анализ геномов, сравнительный анализ белков и ДНК, изучение пространственной структуры и моделирование белков – лежат в основе рационального подхода по разработке новых лекарственных средств в качестве одного из основных инструмента, использующего так называемый *компьютерный drug design*.

Процесс создания нового лекарственного соединения включает в себя следующие этапы: (1) выбор и валидация мишени (например, белка), на которую будет направлено действие нового лекарства; (2) поиск низкомолекулярных соединений, обладающих способностью нужным образом взаимодействовать с мишенью; (3) сравнительное исследование этих соединений в экспериментах и выбор лучших кандидатов; (4) проведение клинических испытаний. Два первых этапа могут включать в себя компьютерные эксперименты.

Наиболее благоприятная ситуация наблюдается в том случае, если для выбранного белка-мишени известна пространственная структура, изучен механизм действия белка и его роль в патогенезе. В том случае, если пространственная структура белка-мишени неизвестна, ситуация бывает двоякой: когда имеется белок с известной пространственной структурой, похожий на белок-мишень по аминокислотной последовательности, и когда такого белка не существует. В первом случае строят модель пространственной структуры белка-мишени, укладывая белковую цепочку в пространстве тем же способом, как в похожем белке, с последующей оптимизацией структуры. Во втором случае можно пытаться определить пространственную структуру белка-мишени экспериментально, с помощью методов РСА или ЯМР, но, как было сказано выше, это дорогостоящие и длительные эксперименты и успех не гарантирован.

Существенным моментом является, помимо выявления мишени, проведение валидации выбранной мишени, которая, как правило, осуществляется с помощью различных экспериментальных подходов. При этом мишень тем или иным способом инактивируют (на уровне гена или на уровне белка) и изучают появляющиеся симптомы. Один из современных методов валидации мишени предусматривает характеристику ее взаимодействия с тем или иным соединением методом плазмонного резонанса. Обычно валидацию проводят с использованием специальных генетических линий лабораторных животных, однако, окончательным этапом (который можно достичь далеко не всегда) является, несомненно, демонстрация того, что модификация или ингибирование мишени приводит к устойчиво воспроизводимым симптомам у человека.

На этапе валидации одной из важных задач является задача определения места связывания низкомолекулярного соединения белком-мишенью, которая в ряде случаев может быть решена с применением высокопроизводительной компьютерной техники. В некоторых случаях она решается только экспериментально или не решается, и тогда следующий этап работы, связанный с перебором низкомолекулярных соединений, которые могут взаимодействовать с белком-мишенью, сильно усложняется, поскольку при переборе соединений приходится исходить из предположения, что каждое из них может взаимодействовать с любым местом на поверхности молекулы-мишени. Понятно, что в данном случае требуется использование суперкомпьютеров максимально доступной вычислительной мощности.

После того, как тем или иным способом получена пространственная структура белка-мишени и выяснено (или не выяснено) место взаимодействия низкомолекулярного вещества, переходят к следующему этапу – так называемому докингу. При этом из баз данных химических соединений последовательно берут соединения-кандидаты и делают попытку присоединить их к молекуле-мишени, как бы "обкатывая" интересующие участки поверхности. Низкомолекулярное вещество при этом необходимо поворачивать разными сторонами с целью подобрать положение максимального соответствия поверхностей мишени и низкомолекулярного соединения. При этом, разумеется, учитывают не только геометрические характеристики, но и химические свойства атомных групп на обоих взаимодействующих поверхностях, такие, как заряд, способность образовывать водородные связи непосредственно или через молекулы растворителя и т.д. Полученные комплексы подвергают молекулярно-динамическим и квантово-химическим обсчетам. Имеются различные варианты методов, некоторые из которых позволяют несколько сократить объем требуемых вычислений за счет уменьшения точности вычислений (например, вместо "гибких" взаимодействующих молекул, рассматриваются "жесткие", не проводятся квантово-химические расчеты). Однако, учитывая хотя бы то обстоятельство, что примерный объем базы химических соединений, которые анализируются таким способом, составляет 10^{14} молекул, даже при использовании самых современных, например, 100-терафлопных суперкомпьютеров время счета составляет для каждой мишени от нескольких часов или дней до нескольких недель.

Таким образом, компьютерная разработка лекарств – это естественная область применения суперкомпьютеров. Это определяется (1) сложностью и трудоемкостью анализа взаимодействия данного потенциального лиганда с данным белком-мишенью и (2) необходимостью анализа большого количества потенциальных лигандов и возможностью независимо и параллельно обрабатывать разные лиганды. Примером успешного применения компьютерного drug design, проведенного по описанной выше схеме, включающей моделирование структуры молекулы-мишени с последующим докингом, является получение ингибиторов протеазы вируса иммунодефицита человека для лечения СПИДа и ингибиторов ренина для лечения эссенциальной гипертензии и др. *(Srinivasan N. et.al., In: Protein Structure Prediction, Oxford Univ. Press., 1996, 111).*

В настоящее время компьютерный drug design, по-видимому, переживает период, характерный для многих задач, связанных с применением компьютеров для решения "интеллектуальных" задач – таких, как машинный перевод, алгоритмы игры в шахматы и т.п. Первый этап в исследовании таких задач характеризуется выработкой различных (иногда противоречащих друг другу) идей и подходов, попыткой полностью решить глубоко нетривиальную проблему "кавалерийским наскоком". На втором этапе, происходит отбраковка ложных идей, синтез всего полезного и, главное, формулировка, возможно, менее амбициозных, но весьма практически полезных (и даже необходимых), достижимых целей. Так, никто сейчас не пытается научить компьютер глубоко оценивать шахматную позицию, как это делал М.М. Ботвинник. Однако не столь глубокие, но нетривиальные способы оценки качества позиции в сочетании с эффективными алгоритмами перебора вариантов, базами данных по стандартным позициям и возросшей мощностью компьютеров привели к тому, что ведущие шахматные программы регулярно обыгрывают ведущих шахматистов мира.

Подобный переход сейчас переживает и компьютерный drug design. Его целью становится не расчет единственно верного лекарства, а помощь в подборе возможных кандидатов для последующего химического анализа (скрининга). При этом компьютерные и химические методы могут применяться итеративно. Так, выбрав путем скрининга лучший из отобранных с помощью компьютера лигандов-кандидатов, можно снова провести компьютерный поиск среди молекул, которые можно получить небольшим "шевелением" отобранного лиганда (или нескольких лигандов).

При разработке новых лекарств основное время занимает скрининг химически синтезированных лекарственных соединений и их последующее клиническое тестирование: именно на них тратится и большая часть средств. Так, процесс разработки нового лекарственного средства занимает от 5 до 16 лет. Затраты на клинические испытания одного соединения-кандидата составляют более $100 млн. Суммарная стоимость разработки одного лекарственного препарата, с учетом препаратов, не достигших рынка, часто превышает $1 млрд. Использование компьютерных методов в ряде случаев позволяет снизить стоимость разработки в несколько раз.

Таким образом, эффективное использование высокопроизводительной вычислительной техники (суперкомпьютеров) в ближайшем будущем, вероятно, во многом будет определять прогресс в современной молекулярной биологии и базирующихся на ней биомедицинской химии и фармакологии. Прежде всего, это касается задач, требующих решения большого числа относительно трудоемких и независимых однородных подзадач. Примерами таких задач являются рассмотренные нами молекулярная динамика, поиск в базах данных и компьютерная разработка лекарств.

Обеспечение биомедицинских исследований необходимыми вычислительными ресурсами, в том числе суперкомпьютерами, критически важно для успешного развития этих исследований.

ОАО "Туполев": использование HPC-систем позволило бы сократить минимум вдвое время проектирования самолета

Слободчиков А.С. *– начальник центра компьютерных технологий,* ***Толстов К.В.*** *– зам. начальника центра "Техпроекты"*

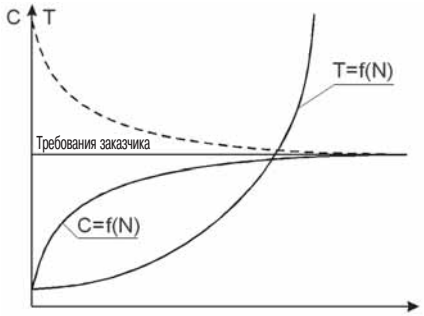
Основная цель конструкторского бюро – создание конкурентоспособной продукции в условиях жесточайшей конкуренции. Как следствие этого, основными задачами КБ являются:

- повышение привлекательности продукции за счет повышения уровня комфорта, снижение затрат на техническое обслуживание и ремонт, повышение топливной экономичности и др.;
- сокращение сроков подготовки производства (изготовление технологической оснастки, ЧПУ программ, и т.д.);
- выпуск продукции, соответствующей не только Российским авиационным правилам, но и требованиям международных норм (экология, шум, безопасность полета, эргономика).

В этой связи "Туполев" стремится создать как для всего семейства самолетов в целом, так и для каждого самолета в отдельности полное электронное определение изделия, в которое входили бы не только геометрические, но и все расчетные модели, расчеты, документация, наземное оборудование и т.д. Полное электронное определение необходимо для оценки влияния изменений, вносимых в конструкцию или системы конкретного самолета, а, соответственно, и совместимость данных изменений с авиационным комплексом в целом. При созданном полном электронном определении изделия хотелось бы получать такие оценки хотя бы в течение недели.

Использование CAD/CAM/CAE систем снимает часть вопросов, возникающих в связи с этим, а также заставляет заботиться о наращивании производительности вычислительного комплекса. Геометрия самолета, сделанная достаточно подробно, уже является неудобоваримой для работы с ней целиком. Предъявление электронного макета, комиссии, авиационного комплекса целиком (с системами наземного обслуживания, движением по полосе и рулежным дорожкам, анализ аварийных и отказных ситуаций, оценка различных ситуаций в режиме REAL TIME) пока не реализуемо.

В процессе проектирования также ощущается необходимость в мощном вычислительном комплексе. Процесс создания практически любого изделия является итеративным. На графике изображены зависимости C= (N) и T= (N) где:



C – степень соответствия изделия требованиям заказчика, относительная величина;

T – время, затраченное на проработку данного варианта;

N – порядковый номер итерации.

В данной ситуации HPC-системы являются необходимыми. При проектировании самолета производится большое количество расчетов: аэродинамические, прочностные, ресурсные и т.д. , а также моделируются процессы работы различных систем и поведение самолета при их работе. Решение многих задач при проектировании самолетов требует от вычислительного комплекса производительности, равную нескольким десяткам TFlops, так как решение этих задач хотелось бы получить хотя бы в течение одной ночи.

Использование HPC-систем позволило бы сократить минимум вдвое время проектирования самолета и подготовки его производства, увеличив при этом уровень детализации расчетов, степень достоверности (сходимость с реальностью), точность. По оценкам специалистов КБ, на данный момент существует необходимость в наращивании мощности вычислительной системы до 5 TFlops с перспективой дальнейшего увеличения. При получении данной производительности появляется возможность перехода на качественно новый уровень проектирования.

На данный момент ощущается острая нехватка финансирования, подготовленных кадров в этой области и пакетов, ориентированных на работу в данных технологиях, применимых в авиастроении.

Санкт-Петербургский Государственный Морской Технический Университет: HPC-технологии позволят перевести решение ряда практически важных задач из группы – "исследования" в группу "использование при конструировании"

Журава В.М., Фрумен А.И.

Перспективный путь развития Российской экономики и судостроения в частности – это овладение "высокими технологиями", включая HPC-технологии.

Проведение параллельных вычислений на многопроцессорных кластерах позволит перевести решение ряда практически важных задач из группы – "исследования" в группу "использование при конструировании", что повысит безопасность системы "Корабль" на всех этапах ее жизненного цикла.

Следует отметить, что модель безопасности системы "Корабль" является многокомпонентной, с изменяемым во времени составом; компоненты ее имеют различную физическую природу; между компонентами имеется как физическая, так и информационная связь. Система имеет многоуровневую структуру, включающую в себя непрерывные, дискретные или гибридные компоненты. Учитывая комплексность воздействия разрушающих факторов на иерархически организованную систему "Корабль", наиболее качественную оценку безопасности дает имитационное моделирование, требующее колоссальных затрат времени и других ресурсов ЭВМ.

Применение HPC-технологий требует написания или закупки специальных пакетов прикладных программ (ППП), реализующих алгоритмы параллельных вычислений. Наибольший эффект достигается при использовании численных методов, базирующихся на матричных операциях, итерационных процедурах и ряде других аналогичных методов.

Имеется широкий спектор отечественных и зарубежных ППП, реализующих современные численные методы и позволяющие решать сложные задачи прочности и надежности судовых конструкций, задачи гидроаэроупругости и акустики, но требующие значительных ресурсов ЭВМ. Прежде всего, это методы конечных элементов (МКЭ); суперэлементов (МСЭ); граничных элементов (МГЭ); частиц в ячейках; крупных частиц в ячейках; панелей; дискретных вихрей; графическое программирование итерационных процедур решения систем дифференциальных уравнений (имитационное моделирование) и пр.

Внедрение в практику проектирования HPC-технологий позволит определить реальные предельные нагрузки и соответствующие коэффициенты запаса прочности, внешние нагрузки при неста-

ционарных (ударных) режимах эксплуатации морской техники и многое другое.

С точки зрения обучения владением сложными пакетами прикладных программ можно ввести три уровня: базовый (начальный), продвинутый (расширенный) и профессиональный. При этом следует иметь в виду, что для ряда сложных пакетов программ необходима профессиональная пользовательская среда (минимум 5–6 человек) для сохранения уровня владения соответствующим пакетом. В бакалавриате, на наш взгляд, можно обеспечить только базовый уровень владения ППП, а в магистратуре – продвинутый для студентов, проявивших необходимые способности. Профессиональный же уровень владения ППП достигается специалистом самостоятельно в ходе решения необходимых промышленности (предприятию) задач при наличии соответствующей пользовательской среды. Профессиональному росту также способствует учеба в аспирантуре и соответствующие курсы повышения квалификации специалиста.

Отметим, что при решении проектных задач нередко возникают трудности с анализом получаемых результатов, его адекватности физическому смыслу решаемой задачи. Это связано, в том числе, и с демографическим разрывом работающих технических специалистов: специалисты в возрасте 22–26 лет хорошо разбираются в компьютерных технологиях, а в возрасте 50–55 лет и более – в смысле задачи.

Важно отметить, что современные ППП достаточно дороги (сотни тысяч долларов и более) и купить их может себе позволить только крупное предприятие. Это касается в перую очередь пакетов, реализующих алгоритмы параллельных вычислений. Существуют более дешевые и ограниченные по возможностям учебные версии ППП для некоторых пакетов.

Например, программа ANSYS (г. Хьюстон, штат Пансильвания, США), реализующая МКЭ для решения задач сплошных сред, была разрешана для продажи в РФ после 1991 г.; имеет вариант использования на многопроцессорных вычислительных системах (до 64 процессоров за пределами США) и стоит в 5 раз дороже, чем однопроцессорный вариант.

Частично эту проблему может помочь решить создание учебно-научно-производственных центров при университетах, которые смогут обслуживать несколько промышленных предприятий и одновременно решать задачи подготовки кадров в области промышленных информационных технологий для этих и других предприятий.

Вывод: внедрение высокопроизводительных HPC-технологий, обеспечение качества многоуровневой подготовки и переподготовки специалистов возможно только при сотрудничестве университета с промышленностью как в области решения сложных научно-технических задач создания современной морской техники, так и в области подготовки специалистов для промышленности при необходимой финансовой поддержке государства.

"НПО "Сатурн" удесятеряет IT-мощности

Июль 2007 г. – НПО "Сатурн", компании КРОК и IBM объявили о старте совместного проекта по созданию суперкомпьютера производительностью 8 Тфлопс. Соответствующее соглашение о сотрудничестве подписано в ходе Международного авиационно-космического салона 2007.

В 2005 г. НПО "Сатурн" запустил в эксплуатацию свой первый вычислительный кластер на базе серверов IBM xSeries, который сегодня загружен расчетными задачами на 100% и возникла реальная необходимость на порядок увеличить существующие вычислительные мощности. Запуск нового суперкомпьютера намечен на конец 2007 – начало 2008 г. Специалисты НПО "Сатурн" рассчитывают, что затраты времени на внедрение нового суперкомпьютера сократятся в 2 раза по сравнению с предыдущим проектом.

Юрий Зеленков, зам. директора по информационным технологиями НПО "Сатурн" отметил: "Ввод в эксплуатацию предыдущего кластера позволил увеличить наши общие вычислительные мощности в 50 раз. Реализация проекта, о котором мы говорим сегодня, увеличит наши мощности еще в 8–10 раз. Таким образом, мы увеличиваем наши ресурсы, как минимум, в 400 раз за два года. И для НПО "Сатурн" это не просто увеличение вычислительных мощностей, а серьезный вклад в повышение конкурентоспособности как на российском, так и международном рынке".

ФГУП "ММПП "Салют" запускает суперкомпьютер

Июль 2007 г. – ФГУП "ММПП "Салют" заключило контракт на создание высокопроизводительного кластера из 50 двухпроцессорных серверов RX200 S3 производства Fujitsu Siemens Computers на базе новейших четырехъядерных процессоров Intel® Xeon® 5365 с частотой 3,0 ГГц. Поставку, интеграцию и настройку всего комплекса будет осуществлять компания-интегратор "АйТеко". Производительность вычислительного комплекса должна составить 4,5–5 Тфлопс. Система будет использоваться, в частности, для инженерного анализа и необходимых расчетов в рамках проекта по созданию российского авиадвигателя нового поколения.

В настоящий момент на предприятии используются несколько специализированных комплексов на базе платформ Intel Xeon и Intel® Itanium® 2, объединенных с системой хранения конструкторских данных общей емкостью 35 Тбайт.

Качественное улучшение эксплуатационных характеристик нового двигателя можно получить, только создав полную и точную математическую модель объекта разработки. По анализам конструкторов, для решения поставленной задачи объем расчетной сетки необходимо довести до 250 млн элементов.

IBM Information Server Blade для интеграции информации

Август 2007 г. – Корпорация IBM представила новую версию Information Server Blade для интеграции информации в распределенной корпоративной среде на базе кластера из blade-серверов. Решение будет доступно с октября с.г. (не-blade версия поставляется с ноября 2006 г.).

IBM Information Server Blade – это полностью интегрированное решение, включающее blade-системы IBM, программную платформу для интеграции данных IBM Information Server и услуги по внедрению. Данное решение предназначено для реализации проектов любой сложности в области интеграции данных, включая консолидацию, слияния и приобретения, бизнес-анализа или организации хранилищ данных.

Система Information Server Blade работает под управлением ОС Red Hat Linux и создана на базе серверов IBM BladeCenter HS21 с процессорами Dual-Core Intel Xeon. Применение в этой энергоэффективной системе стандартных для отрасли процессоров с низким напряжением питания обеспечивает снижение объемов потребляемой электроэнергии и требований к охлаждению, в сравнении с более масштабными системами.

Чтобы упростить управление и эффективное использование grid-вычислений и средств виртуализации, решение Information Server Blade включает ПО IBM Systems Director, предоставляющее пользователям информационную панель для централизованного анализа и администрирования всех рабочих нагрузок и управления всеми физическими и виртуальными машинами. Кроме того, это решение обеспечивает гладкое, интегрированное управление grid-ресурсами с использованием ПО Tivoli Workload Scheduler LoadLeveler, позволяющего легко управлять рабочими нагрузками для всех blade-серверов. ПО Tivoli Workload Scheduler LoadLeveler обеспечивает высокопроизводительную обработку рабочих нагрузок и эффективное использование ресурсов в grid-кластерах. В случае увеличения потребностей в вычислительных ресурсах можно просто подключить дополнительные blade-серверы, а ПО Tivoli Workload Scheduler LoadLeveler может использоваться для управления распределением рабочих нагрузок для множества grid-систем.