# Вычислительная химия на российских грид-полигонах: текущее состояние, проблемы и перспективы

Волохов В.М., **Варламов Д.А.**, Волохов А.В., Пивушков А.В., Покатович Г.А., Прохоров А.И.

***Институт проблем химической физики РАН***

**ПаВТ'2013, Челябинск, 2013**

# Вычислительная химия – ненасытный потребитель вычислительных ресурсов

1. **Первые две** научные задачи, использовавшие петафлопсную производительность – это задачи вычислительной химии: расчеты электронной структуры высокотемпературных сверхпроводников и исследования эффекта магнито-сопротивления в наночастицах (метод Monte-Carlo) – “Jaguar”, Oak Ridge, США – 1,64 Pf;
2. Примеры ресурсных требований: исследования свойств воды в низкоразмерных системах - до $8 \cdot 10^6$ CPU(core)-часов (Argonne National Laboratory); исследования химических катализаторов – до $30 \cdot 10^6$ CPU-часов в год – “Jaguar”, 2010;
3. Типичный FULL docking белковых лигандов: 200 атомов x 300 000 конфигураций x 1000 core-часов – более **300 Pflops**!!!
4. Большинство суперкомпьютерных центров США (San Diego, Ohio, Urbana-Illinois etc.) предоставляют до **40%** вычислительных мощностей для нужд биохимии, молекулярного моделирования, квантовой химии, нанотехнологических расчетов

# Суперкомпьютерные центры, специализирующиеся на квантово-химических и молекулярно-динамических расчетах

1. The EPSRC (Engineering and Physical Sciences Research Council), UK служба National Service for Computational Chemistry Software, Великобритания, http://www.nsccs.ac.uk – >250 Tflops
2. The National Resource for Biomedical and Chemistry Supercomputing (NRBSC, http://www.nrbsc.org, США) – расчет молекулярных систем от 20000 до 120000 атомов, обеспечивает доступ внешних пользователей к вычислительным ресурсам до 100 000 CPU-часов на структуру;
3. Chemical Computing Group (http://www.chemcomp.com), Квебек, Канада – >120 Tflops
4. Lawrence Berkeley National Laboratory (США, http://www.lbl.gov/csd), подразделение вычислительной химии – > 450 Tflops;
5. Lehrstuhl für Theoretische Chemie der Technische Universität München, Мюнхен, Германия (http://www.lrz.de/services/software/chemie) – собственный Linux кластер (>96 Tflops) плюс постоянно выделенные ресурсы Supercomputer Center;
6. Swiss National Supercomputing Centre (http://www.cscs.ch) – для химических расчетов выделены ресурсы от десятков Tflops до суб Pflops

# Важно!

- Для построения многомерных потенциальных поверхностей, адекватно описывающих химические реакции, нужно провести $10^2$ – $10^{\text{много}}$ независимых *ab initio* расчетов
- Для исследования динамики химической реакции с использованием классических траекторных расчетов нужно рассчитать до $10^9$ независимых траекторий
- Для численного исследования функции $F(x_1,x_2\ldots x_n)$ в области изменения параметров $x_1,x_2,\ldots,x_n$ при разбиении диапазона изменения каждого параметра на 10 ячеек нужно провести $10^n$ независимых расчетов *F*

# Вывод №1

- Независимые расчеты помимо суперкомпьютеров необходимо проводить на географически распределенных ресурсах, т.е. в GRID средах
- Необходимо разработать технологии запуска множества параллельных приложений на удаленных ресурсах

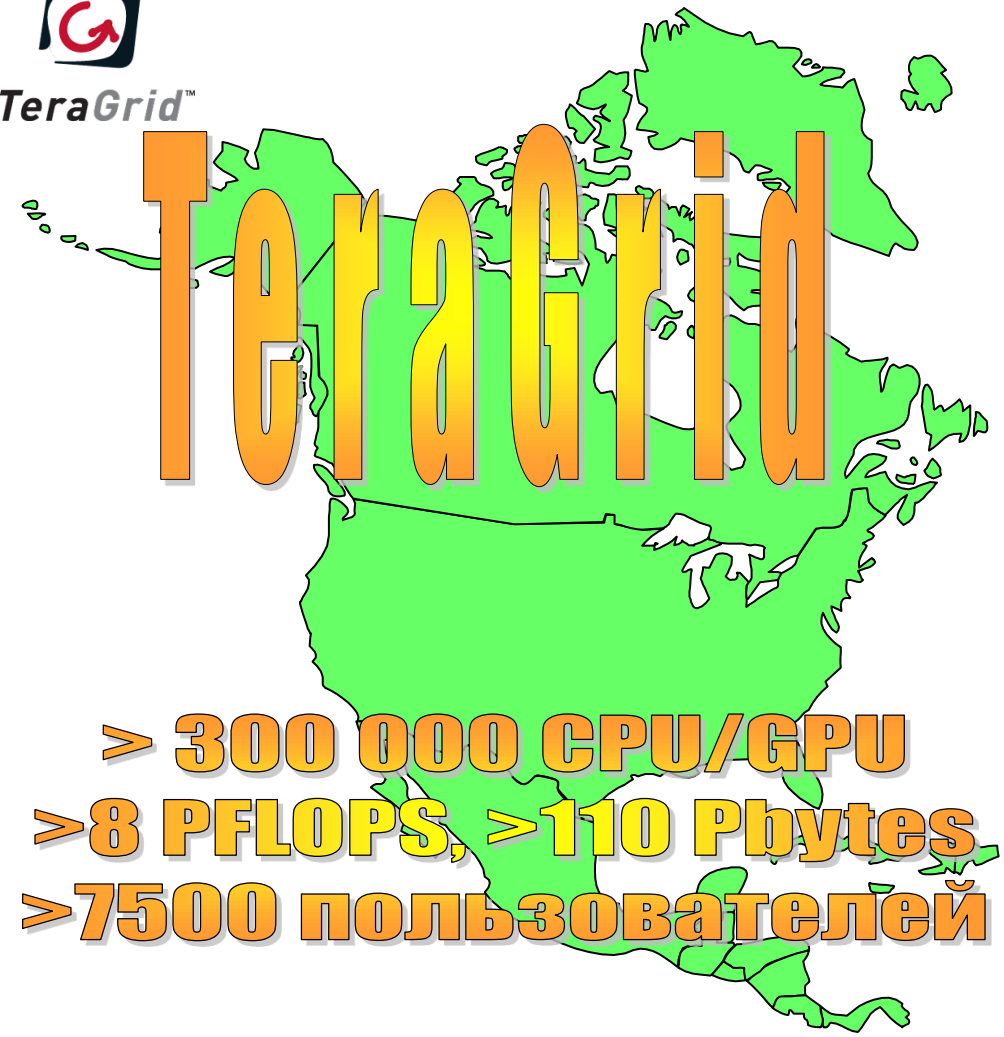
# Крупнейшие мировые проекты

ЯДЕРНАЯ ФИЗИКА
БИОИНФОРМАТИКА
ФАРМАЦЕВТИКА
КВАНТОВАЯ ХИМИЯ
АСТРОНОМИЯ/АСТРОФИЗИКА
МЕТЕОРОЛОГИЯ
МАТЕРИАЛОВЕДЕНИЕ
ГЕОЛОГИЯ/ГЕОФИЗИКА
ФИНАНСЫ
АВТОМОБИЛЕСТРОЕНИЕ
и еще более 20 отраслей

> 250 000 CPU/GPU
7 PFLOPS, 100 Pbytes
300 сайтов в 50 странах
>40 000 пользователей

**Финансирование ЕС:**
2004-2006 – 32 M€
2007-2008 – 46 M€
Сейчас в год централизованно выделяется около 18 M€ + проекты

# Наиболее значительные ГРИД проекты



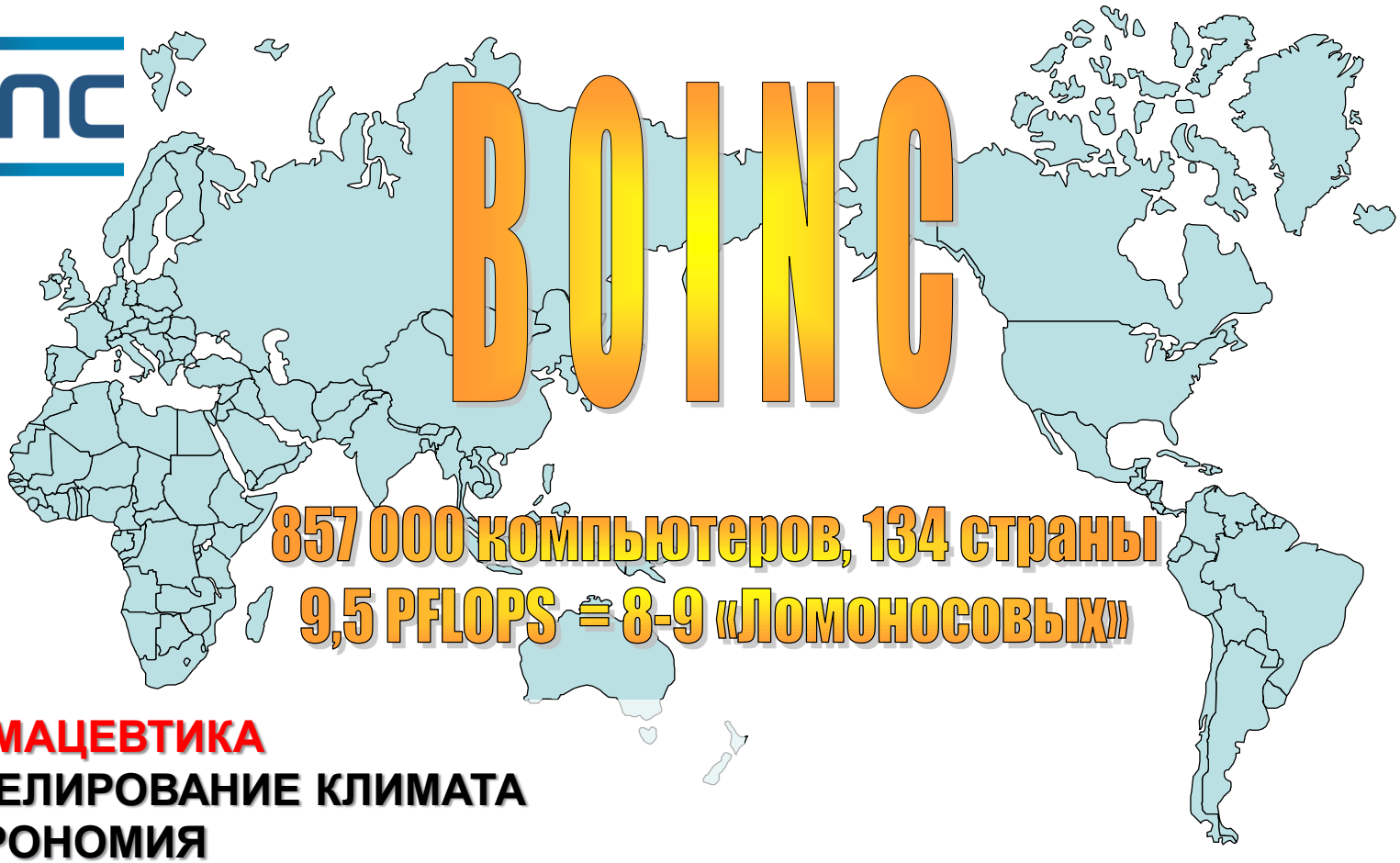

| | |
|---|---|
| МОЛЕКУЛЯРНАЯ БИОЛОГИЯ, ФАРМАЦЕВТИКА | 23% |
| ФИЗИКА | 23% |
| ХИМИЯ | 21% |
| АСТРОНОМИЯ | 10% |
| МЕТЕОРОЛОГИЯ | 7% |
| ТЕХНОЛОГИИ | 6% |
| МАТЕРИАЛОВЕДЕНИЕ | 4% |
| МАТЕМАТИКА И МОДЕЛИРОВАНИЕ | 2% |
| ГЕОЛОГИЯ И ГЕОФИЗИКА | 1% |

**Финансирование NSF**

| | |
|---|---|
| 2001 | 53 M$ |
| 2003 | 35 M$ |
| 2005 | 150 M$ - 5 лет |
| 2011 | 121 M$ - 5 лет |

С 2011 – **Extreme Science and Engineering Discovery Environment (XSEDE)**

# Наиболее значительные ГРИД проекты




- ФАРМАЦЕВТИКА
- МОДЕЛИРОВАНИЕ КЛИМАТА
- АСТРОНОМИЯ
- ГЕНЕТИКА
- МОЛЕКУЛЯРНАЯ БИОЛОГИЯ
- МАТЕМАТИКА И КРИПТОГРАФИЯ
- КВАНТОВАЯ ХИМИЯ

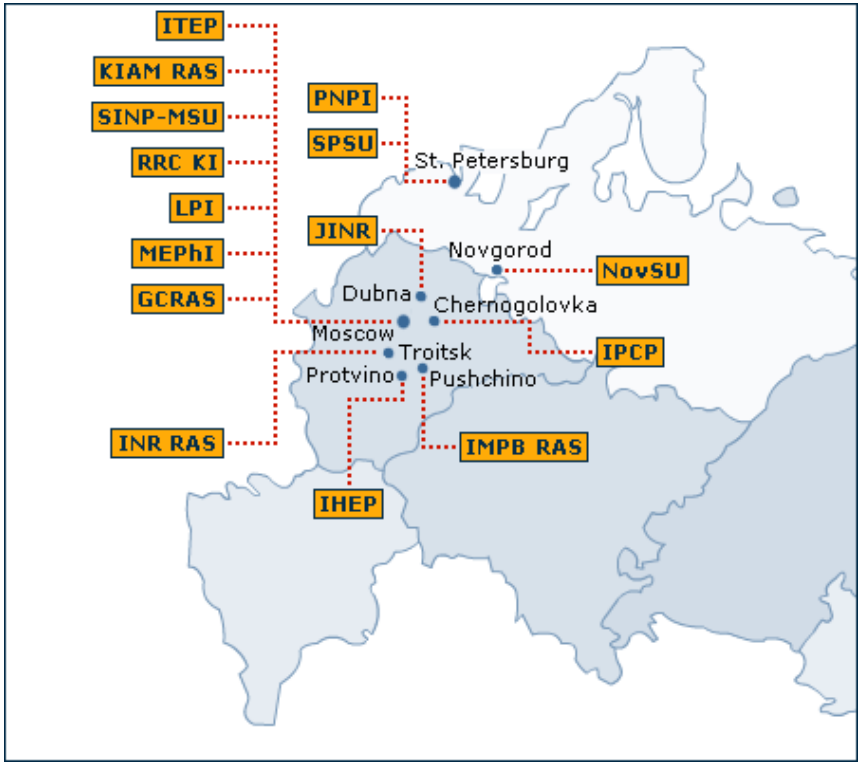
**Всего более 120 проектов**

Россия – 11 место, более 40000 пользователей

* Boinc = “Berkeley Open Infrastructure for Network Computing”

# Российские грид-проекты и ИПХФ в них

## Российский сегмент EGI – National Grid Initiative
## 2003-2013 гг. (ранее EGEE-RDIG, http://rus.egee-rdig.ru)




ИПХФ РАН в течение 2005-2010 участвовал в качестве полнофункционального ресурсного центра в работе российского сегмента одного из крупнейших грид-проектов в мире: EGEE. Работы проводились в рамках виртуальной организации **Rgstest.**

Большинство разработанных в ИПХФ грид-технологий были впервые созданы и апробированы именно **на полигоне EGEE-RDIG (сейчас EGI-[RU-NGI] – European Grid Infrastrucure, российский национальный сегмент) .**

**Rgstest** в течении ряда лет был прекрасной площадкой для проведения технологических разработок и их тестирования, предоставляя до 400-500 CPU и 8-10 Тб на основе middleware **LCG2/gLite**.

Были получены значительные наработки по проведению квантово-химических расчетов в грид-средах как с использованием предустановленных и описанных в рамках ВО прикладных пакетов (GAMESS-US, VASP), так и с использованием авторских программ и оригинальных технологий.

# Российские проекты и ИПХФ в них

**СКИФ-Полигон**

**Научно-техническая программа Союзного государства РФ-РБ 2007-2010 годы, "Разработка и использование программно-аппаратных средств Грид-технологий перспективных высокопроизводительных (суперкомпьютерных) вычислительных систем семейства “СКИФ”**

В ИПХФ был создан Ресурсный сайт СКИФ-Полигона (категория «А») на **базе Unicore 6.2 Middleware**

Обеспечен доступ (на время проекта) к 10-12 ресурсам (ИПС РАН, Cyberia Томск, СевКазГУ, Нижегородский ГУ – всего до 120 CPU, около 3-4 Тфлопс)

Адаптированы квантово-химические пакеты, апробированы все разработанные технологии («пучки», «контейнеры», виртуализация ресурсов и приложений)

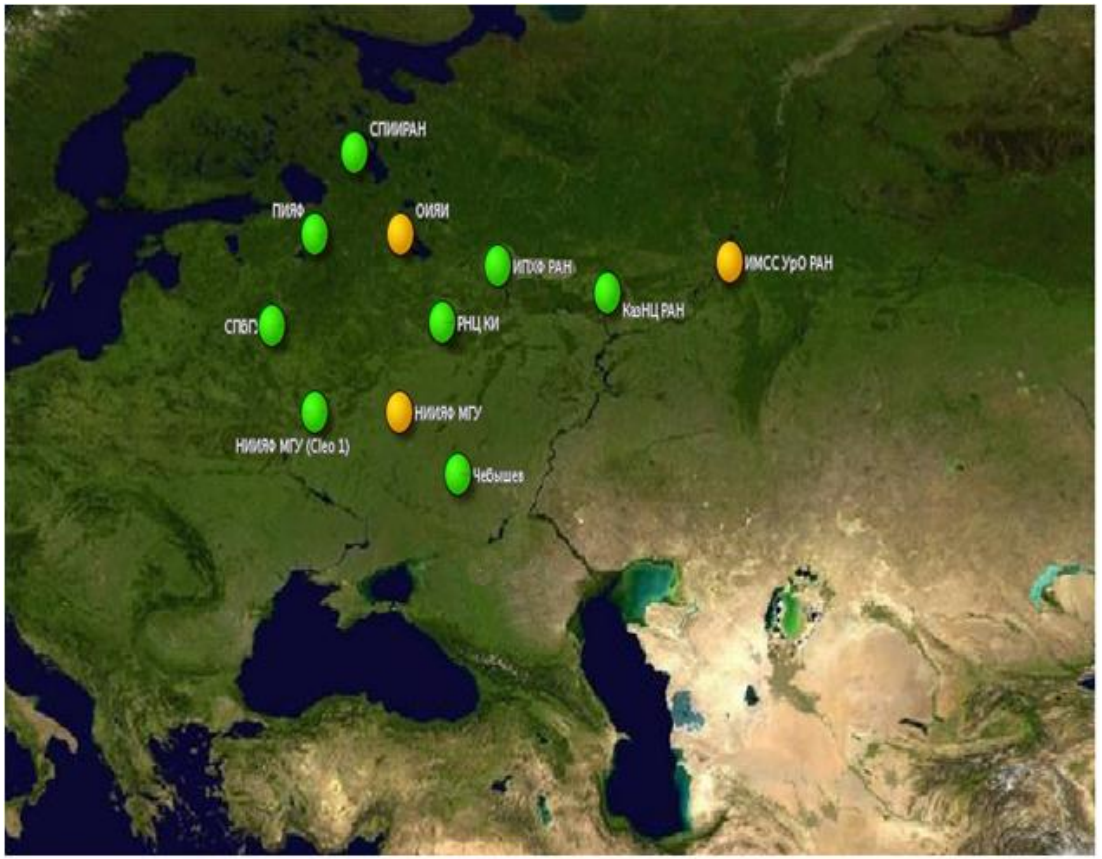
# Российские проекты и ИПХФ в них

## Национальная Нанотехнологическая Сеть (http://www.ngrid.ru) 2010-2012 гг.

**Участники ГридННС**

- НИВЦ МГУ («Чебышев»)
- НИИЯФ МГУ
- ОИЯИ (Дубна)
- РНЦ “Курчатовский Институт”
- Казанский НЦ РАН
- ПИЯФ
- ***ИПХФ РАН***
- ИМСС УрО РАН (Пермь)
- СПбГУ
- ВЦ ДВО РАН (Хабаровск)
- СПИИРАН

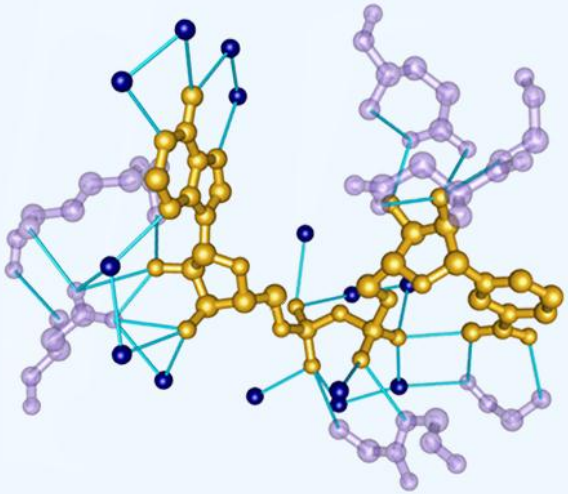
и другие ВУЗы, НИИ РАН, инновационные и коммерческие предприятия



# Участие ИПХФ в грид-проектах

## Национальная Нанотехнологическая Сеть и Виртуальные Организации (VO) в области вычислительной химии

ИПХФ созданы и поддерживаются 3 Виртуальные организации "NanoChem", "Gamess", "Gaussian" для нужд вычислительной химии (http://nanogrid.icp.ac.ru/virtual.php, http://ngrid.ru/ngrid/gridnnn/volist)



На основе модифицированного Globus Toolkit 4 был обеспечен доступ более 200 пользователей к вычислительным ресурсам сети ГридННС (до 40-50 Тф и до 10000 CPU) и большому объему прикладного ПО, в том числе квантово-химического, поддерживаемому инфраструктурой ГридННС

Основной целью проекта ***"Расчет физико-химических свойств наноструктур и высокомолекулярных соединений"*** является исследование ***"Расчет электронного состояния, строения и физико-химических свойств наноструктур и высокомолекулярных соединений с использованием квантово-химического ПО в Грид-среде на распределенных параллельных ресурсах"***.

На базе GRID центра ИПХФ постоянно работает сертифицированный ресурсный сайт ГридННС (на основе модифицированного Globus Toolkit – среда ГридННС) с поддержкой ряда установленных прикладных квантово-химических пакетов (до 10 ППП) и 3-х Виртуальных Организаций, связанных с вычислительной химией

## Квантово-химические пакеты, реализованные в рамках ГридННС

- Gaussian 03
- GAMESS US, Firefly
- CPMD
- Dalton 2012
- NAMD
- NWChem
- Abinit
- VASP
- PWScf

## Молекулярная динамика

- LAMMPS
- OpenMX
- GROMACS

# Пилотная зона российской грид-системы для высокопроизводительных вычислений (на базу ФГУП «Восход») 2011-2012 гг.

**Участники пилотной грид-системы**

- НИИЯФ МГУ
- ОИЯИ (Дубна)
- РНЦ “Курчатовский Институт”
- Казанский НЦ РАН
- ПИЯФ
- ***ИПХФ РАН***
- ИМСС УрО РАН (Пермь)
- СПбГУ
- ВЦ ДВО РАН (Хабаровск)

и другие ВУЗы, НИИ РАН, инновационные и коммерческие предприятия

На основе модифицированного Globus Toolkit 5 был обеспечен доступ более 100 пользователей к вычислительным ресурсам пилотной грид-сети (до 20-30 Тф и до 3000 ядер), к большому объему прикладного ПО, в том числе квантово-химического, поддерживаемому инфраструктурой пилотной грид-сети

# Оригинальные научно-технические разработки ИПХФ в области грид-технологий

1. Разработка высокоуровневых WWW-интерфейсов к сложным прикладным квантово-химическим пакетам для работы с ними в условиях грид-сред (в том числе с применением базовых шаблонов)
2. Создание и управление «пулами» («пучками») грид-заданий для работы с большими задачами на равномерных «сетках» данных или параметров (с использованием многопараметрических задач или прикладных пакетов типа GAMESS). Они могут быть представлены в виде объединения большого количества параллельно выполняемых независимых друг от друга задач;
3. «Виртуализация» грид-приложений и использование динамически формируемых «вычислителей» (контейнеров);
4. Разработка способов виртуализации вычислительных ресурсов и грид-сервисов.

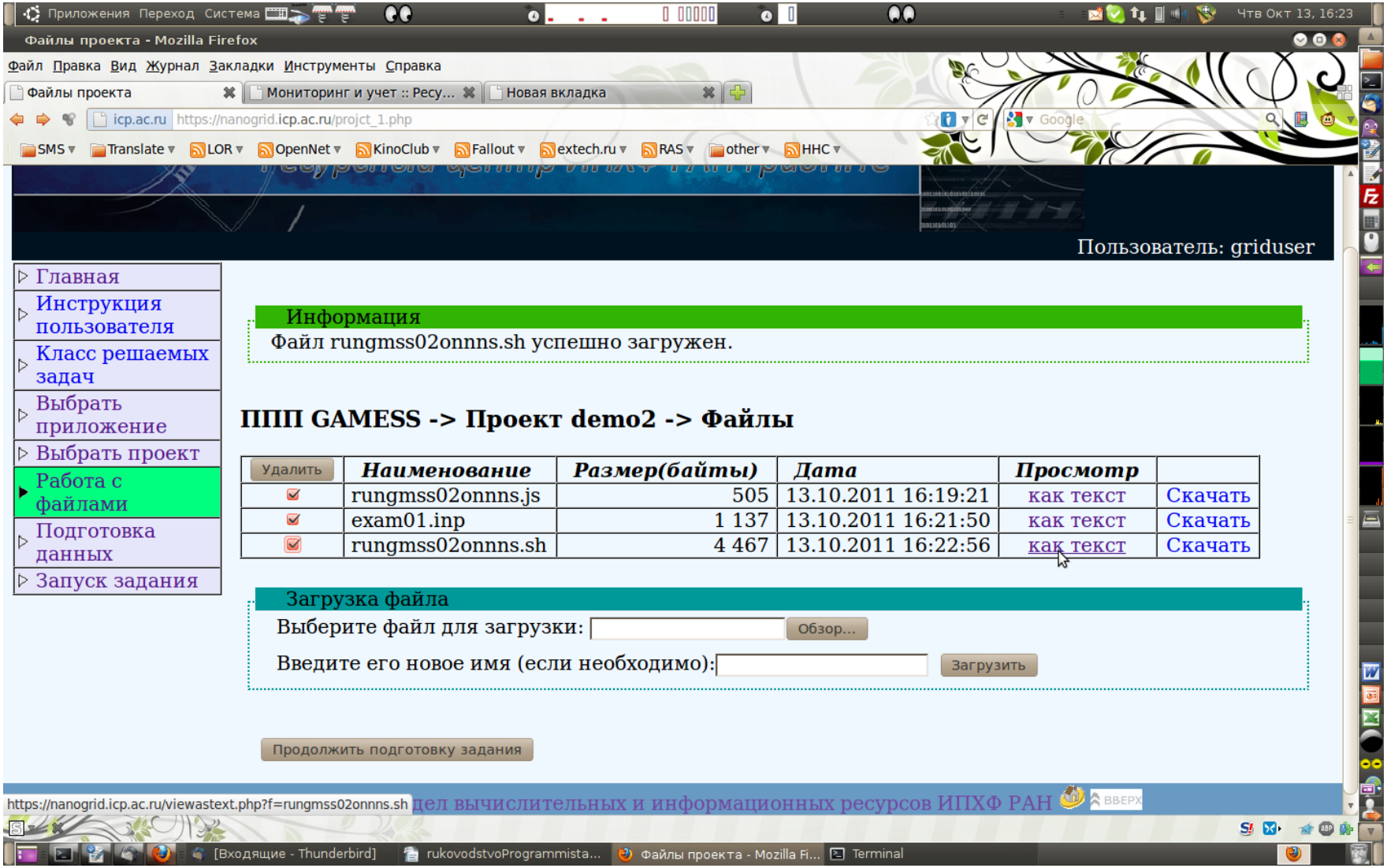
**В рамках ГридННС и пилотной грид-сети созданы Web-интерфейсы** (http://nanogrid.icp.ac.ru , http://webgrid.icp.ac.ru – грид-порталы ИПХФ) для работы с квантово-химическими прикладными пакетами ***GAMESS-US, Gaussian, NAMD, NWChem, PWscf и др.*** в виртуальных организациях **ГридННС и пилотной зоны российской грид-системы для высокопроизводительных вычислений**

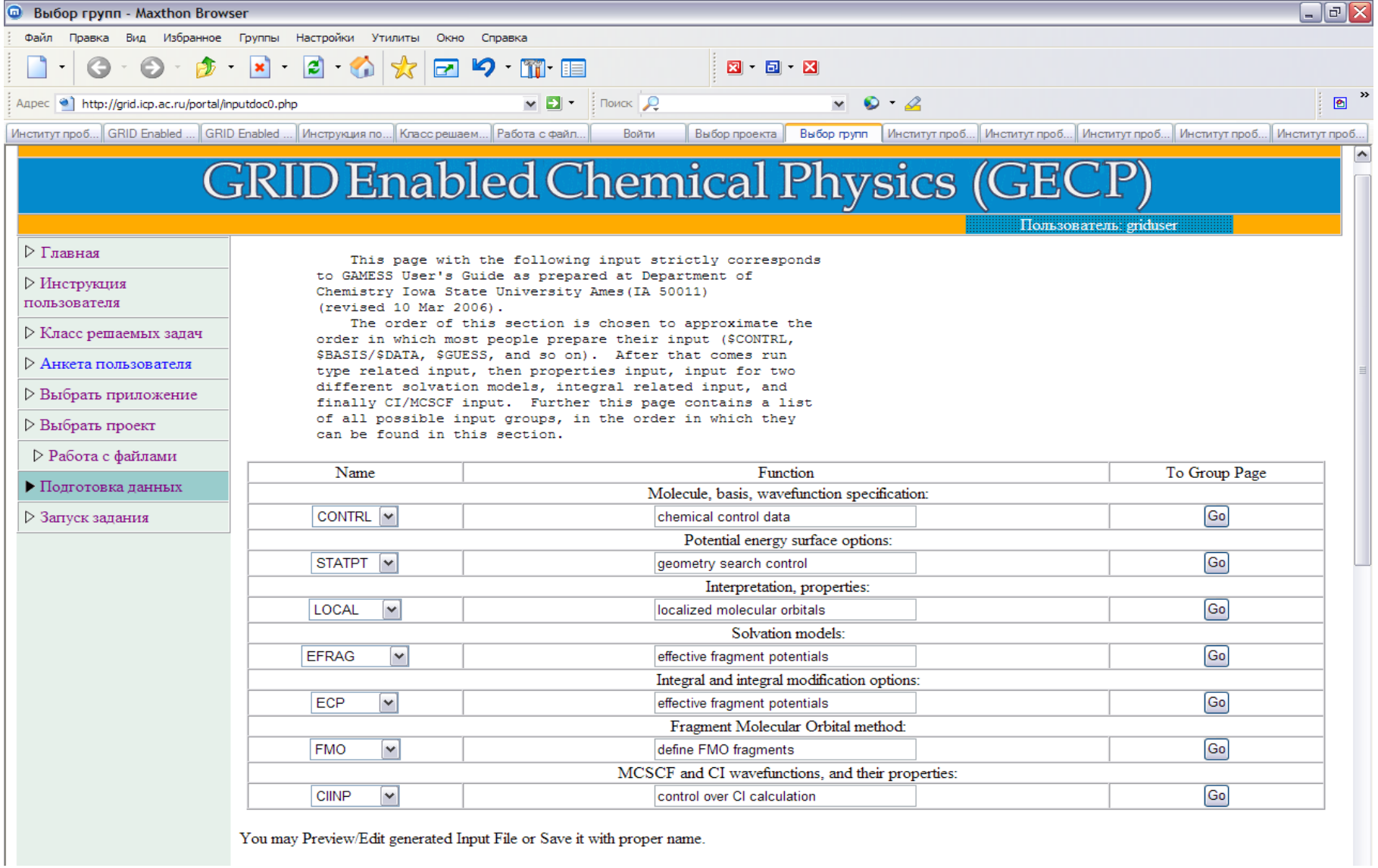
Дружелюбные конечному пользователю (исследователю) Web-интерфейсы к прикладным программным пакетам значительно снижают трудоемкость работы пользователя с подобными пакетами в условиях распределенных вычислительных сред.

Созданные интерфейсы позволяют пользователю работать через Интернет-браузер и осуществлять следующие **действия**:

- авторизовать пользователя для запуска комплекса и проводить его сертификацию в виртуальной организации;
- подготовить задание (включая создание и редакцию начальных данных и конфигурационных файлов) в соответствие с требованиями пакета;
- запустить прикладной пакет в инфраструктуре грид-полигона (при необходимости – на произвольном или избранном GRID ресурсе);
- вести мониторинг выполнения задания (включая останов и перезапуск);
- по завершении – получить результаты счета.

# Пример модуля Web-интерфейса квантово-химического пакета GAMESS (работа с файлами)

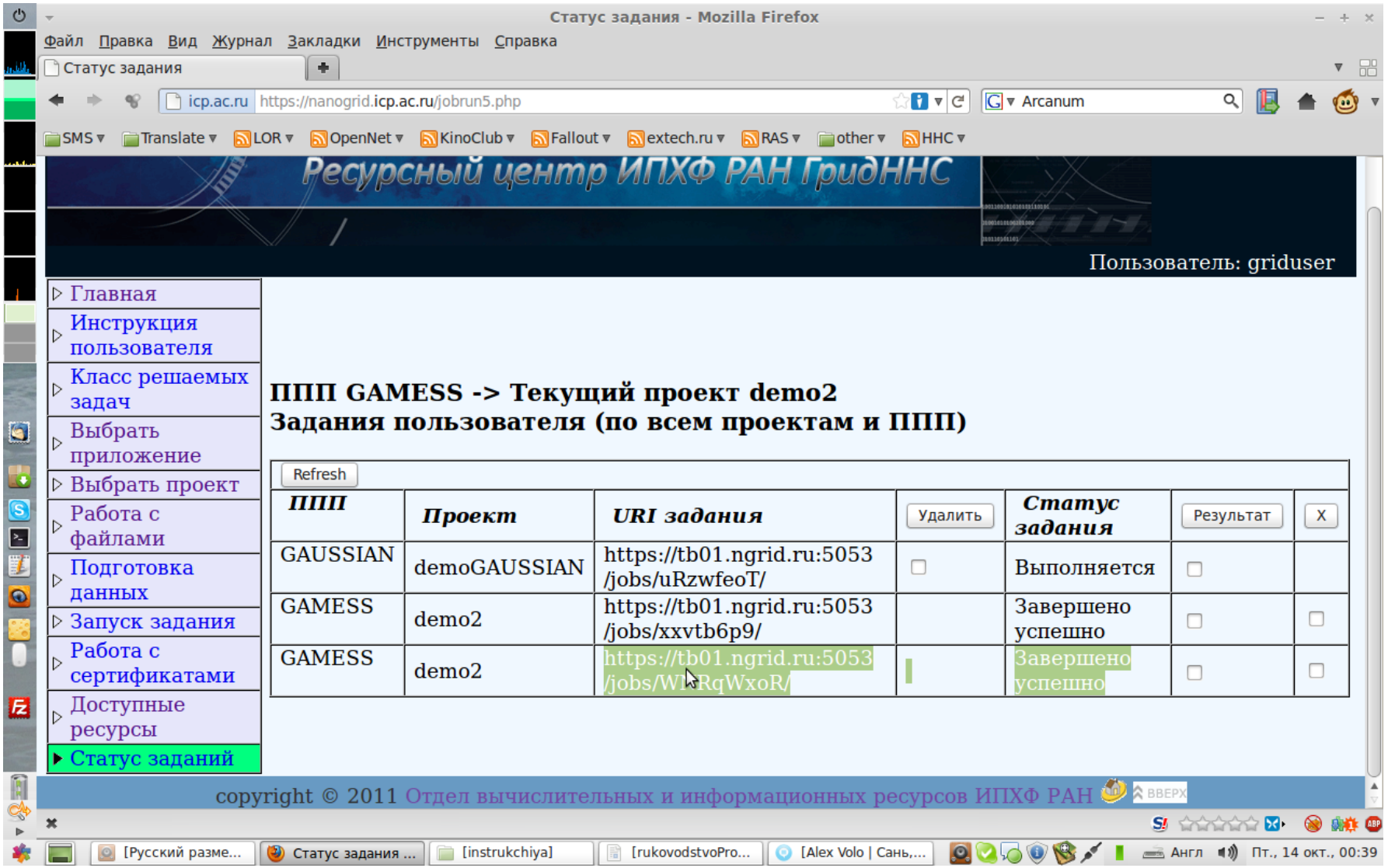





# Веб-интерфейс для полномасштабной подготовки данных, запуска и мониторинга квантово-химического пакета GAMESS-US и многопараметрических задач в распределенных средах

**ПаВТ'2013, Челябинск, 2013**

# Пример модуля Web-интерфейса квантово-химического пакета GAMESS (мониторинг выполнения)



# Глобальные трудности создания Российского GRID

- Неготовность большинства исследователей к постановке и решению масштабных задач, требующих GRID вычислений;
- Отсутствие кредитоспособных заказчиков в лице государства и бизнеса – не путать с хорошо финансируемыми *краткосрочными* проектами.

Деньги нужны для постоянной поддержки GRID полигонов – см.выше для зарубежных полигонов

# Локальные проблемы создания, развития и поддержки GRID сервисов

- Высокие экономические затраты на создание и эксплуатацию вычислительных ресурсов и инфраструктуры, отсутствие стабильной поддержки Российских грид-полигонов;
- Отсутствие заинтересованности со стороны производства и бизнеса:
  - общая низкая культура проведения технологических разработок в большинстве НПО и промышленных организаций;
  - неумение (и нежелание) использовать результаты математического моделирования и высокопроизводительных расчетов.
- Нехватка квалифицированных специалистов как со стороны вычислителей (системщики, программисты), так и со стороны пользователей (химики, технологи);
- Практическое отсутствие отечественных прикладных пакетов --> высокая стоимость ПО, лицензионные ограничения, сокращенная область использования, трудности освоения;
- Неурегулированность правовых и экономических отношений между собственниками вычислительных сервисов (в том числе GRID) и потребителями

# Предлагаемые способы решения проблем

- Стимулирование пользователя (промышленного в первую очередь) для использования вычислительных GRID сервисов
  - отсутствие или минимальная оплата пользования GRID ресурсами на первых стадиях инновационных разработок;
  - обучение пользователей – как очное, так и в режиме интерактивного on-line -> создание информационных web-ресурсов по обучению работе с интерфейсами прикладных пакетов и с GRID ресурсами;
  - введение в планы обучения прикладных специалистов (инженеров, химиков и т.п.) дополнительных курсов обучения работы с прикладными пакетами ПО и проведение студентами обязательных расчетов реальных научно-технических и инженерных задач с использованием суперкомпьютеров и GRID структур;
  - создание дружелюбных, с большим количеством базовых шаблонов, интерфейсов (WWW-ориентированных) к GRID ресурсам;
  - введение государством в качестве обязательного условия сертификации применения средств математического моделирования и тестирования для разработки ряда товарных продуктов и технологических процессов.

# Предлагаемые способы решения проблем (продолжение)

- **Государственные политики и стимулирование разработчиков прикладных GRID сервисов и ресурсов**
  - создание государством (или долевое участие) вычислительной инфраструктуры (суперкомпьютеры, кластеры средних масштабов, высокоскоростные каналы связи и т.п.);
  - компенсация государством части эксплуатационных расходов суперкомпьютерных центров;
  - налоговые льготы на разработку отечественного прикладного ПО, создание ориентированных на это инновационных парков;
  - совершенствование нормативно-правового регулирования в области стратегических информационных технологий;
  - в области государственного образования: (а) расширение количества бюджетных мест в вузах для специальностей «системное администрирование/программирование», «прикладное программирование» и т.п.; (б) использование в процессе обучения широкого круга студентов и аспирантов вузов для разработки необходимого системного и прикладного ПО; (в) проведение повышения квалификации вузовских преподавателей по данным специальностям

# Выводы

В рамках Ресурсного Грид-Центра ИПХФ, интегрированного в Национальную Нанотехнологическую Сеть и пилотную зону российской грид-системы для высокопроизводительных вычислений, был реализован доступ к ряду квантово-химических и молекулярно-динамических прикладных пакетов для проведения широкомасштабных расчетов на российских грид-полигонах с использованием как упрощенных, так и проблемно-ориентированных высокоуровневых web-интерфейсов, разработан ряд оригигальных грид-технологий

**СПАСИБО ЗА ВНИМАНИЕ!!!!**

# УЧАСТИЕ ИПХФ В ГРИД-ПРОЕКТАХ

## Федеральные целевые программы

1. Государственный контракт № 07.514.11.4019 от 23 сентября 2011 г. с ИПХФ РАН в рамках ФЦП «**Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы**» по теме «**Разработка технологии проведения высокопроизводительных расчетов в области вычислительной химии в различных распределенных средах с применением методов виртуализации приложений и ресурсов**» – основной исполнитель
2. ФЦП "**Развитие инфраструктуры наноиндустрии в Российской Федерации на 2008-2011 годы**", проект «**Создание Национальной нанотехнологической сети (ГридННС)»,** государственный контракт № 16.647.12.2031 от 13.05.2011 г. – соисполнитель
3. Государственный контракт в рамках ФЦП «**Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы**» на тему **«Развитие пилотной зоны российской грид-системы для высокопроизводительных вычислений, в том числе в интересах федеральных ядерных центров»**, тематика: **«Адаптация пакетов прикладных программ для работы в пилотной инфраструктуре грид-системы. Разработка методов проведения грид-вычислений с использованием динамически формируемых сред исполнения грид-заданий»** – соисполнитель

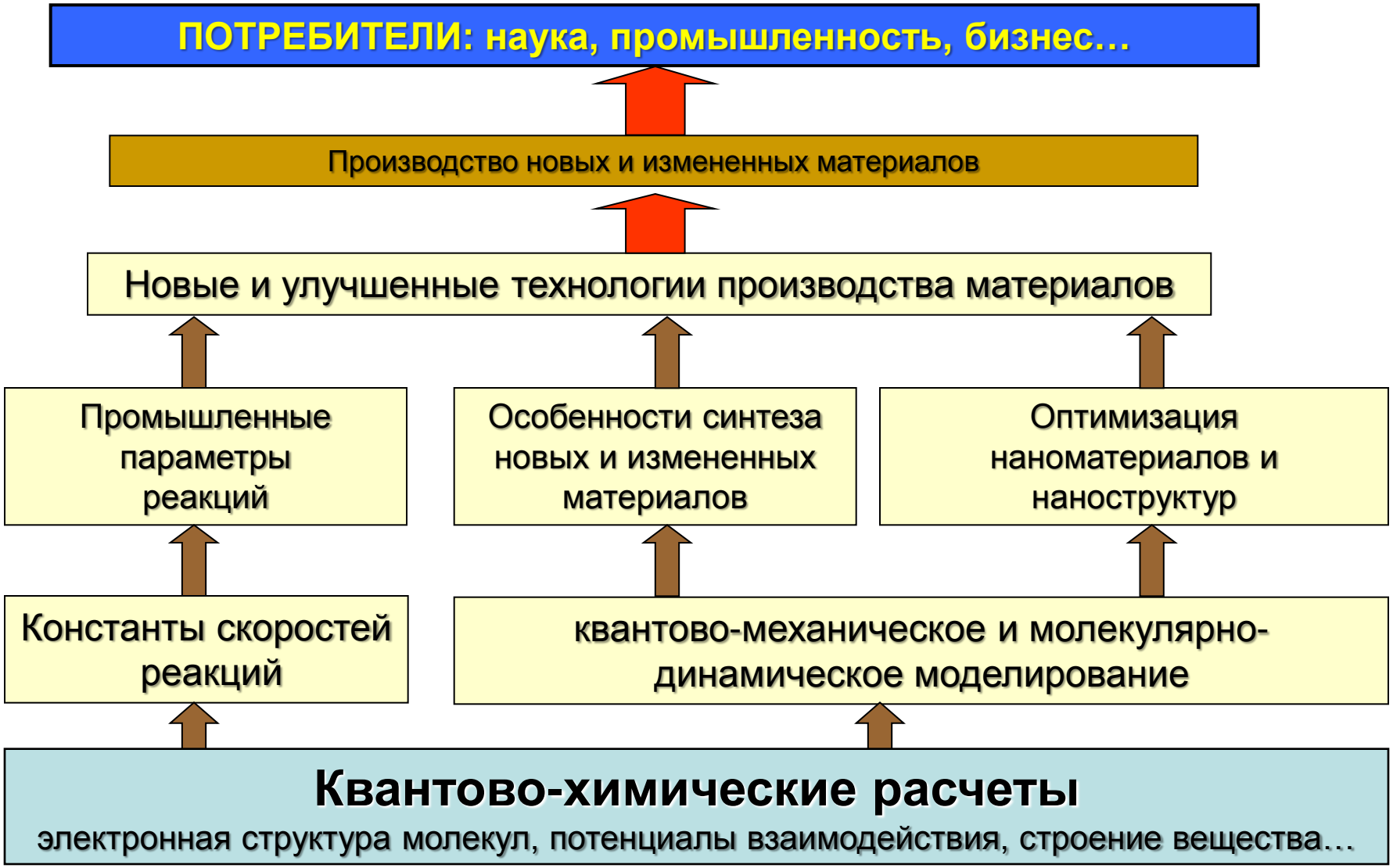
# Программы фундаментальных исследований Президиума РАН

1. Программа № 13 фундаментальных исследований Президиума РАН на 2009-2013 годы «**Проблемы создания национальной научной распределенной информационно-вычислительной среды на основе развития GRID технологий и современных телекоммуникационных сетей**», проект «**Исследование методов виртуализации вычислительных сред и приложений в области вычислительной химии. Динамическое формирование параллельных программных сред на распределенных ресурсах**»;
2. Программа № 27 фундаментальных исследований Президиума РАН на 2009-2013 годы «**Основы фундаментальных исследований наноматериалов**», проект «**Самоорганизация наноразмерных материалов и процессы их взаимодействия с адсорбируемыми соединениями: компьютерное моделирование в параллельных и распределенных GRID средах терафлопного уровня**»

# Гранты Российского Фонда фундаментальных исследований

1. Грант РФФИ № 11-07-00686-а «**Разработка методов динамического формирования параллельных вычислительных сред на различных типах грид ресурсов (на примере приложений квантовой химии)**»

# Химия – наука о материалах, т.е. о мире, в котором мы живем



# Иерархический многомасштабный подход к моделированию объектов от квантового до макроуровня

**Уровень сплошных сред и уровень конструкций**

**определения параметров метода конечных элементов, реализующего модели сплошных сред и модели уровня конструкций, на основе теории механизмов и машин и теории сложных систем**

**Мезоскопический уровень**

**вязкость, теплопроводность, коэффициенты трения, волновые процессы в элементах, мезо- и макроскопические процессы в материалах и т.д.**

**Кинетический уровень**

**системы из 1000 и до 1 млн.атомов, т.е. объекты размером 100-1000 нм$^3$**

a) эволюция квантовых неравновесных систем, состоящих из сотен кластеров
b) методы молекулярной динамики (уравнения классической механики) для атомов и малых молекул, рассматриваемых как классические частицы

**Квантово-статистический уровень**

**расчет моделей с учетом ближайшего окружения кластеров атомов и малых молекул**

**Квантово-механический уровень**

**моделирование методами *ab initio* молекул и малых кластеров из 10-100 атомов (возможные в материале фазы)**

# Упрощенная схема грид-вычислений с использованием высокоуровневых интерфейсов к ППП (среды gLite, Unicore, гридННС)

# Выводы

В ИПХФ РАН разрабатывается ряд вычислительных грид-технологий, направленных на увеличение эффективности проведения расчетов в области вычислительной химии в грид-средах. Введение подобных технологий:

- Существенно ускоряет проведение квантово-химических расчетов и их масштабируемость
- повышает эффективность использования грид-ресурсов
- упрощает их администрирование
- значительно повышает надежность управляющих узлов
- понижает расходы на поддержку грид-инфраструктуры.

Разрабатываемые технологии защищены в качестве основных компонентов объекта интеллектуальной собственности «**Специализированный проблемно-ориентированный пакет вычислительных сервисов различного уровня для решения задач вычислительной и квантовой химии в распределённых сетях ГРИД**», авторское свидетельство №