

Экзафлопсные вычисления: модели, алгоритмы, программные комплексы

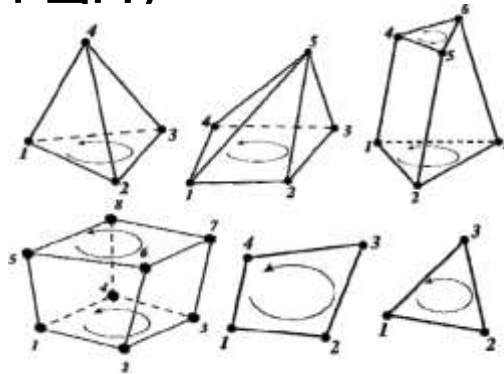
Б.Н.Четверушкин, *М.В. Якобовский*



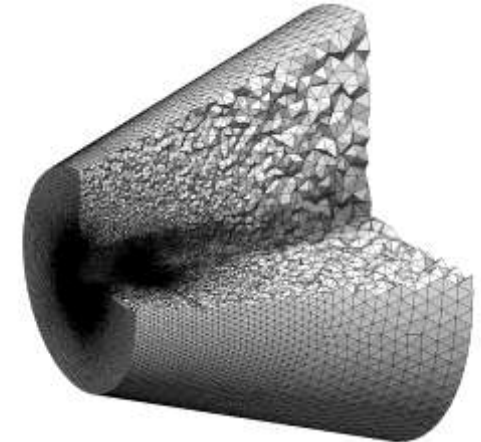
Институт прикладной математики
им. М.В.Келдыша РАН

Методы, пакеты, инструменты

- GIMM (ПРАН)
- GIMM NANO (ФЦП)

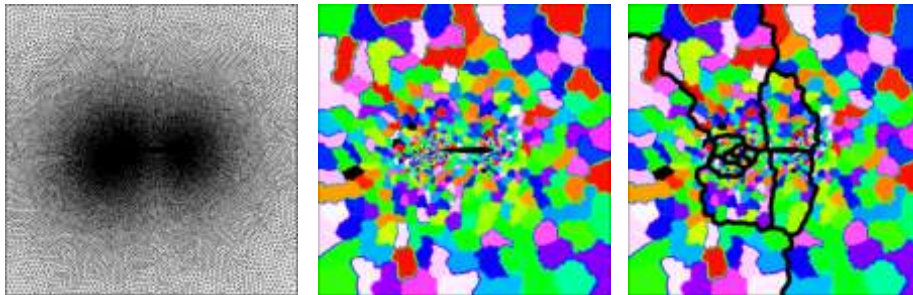


- NOISEtte

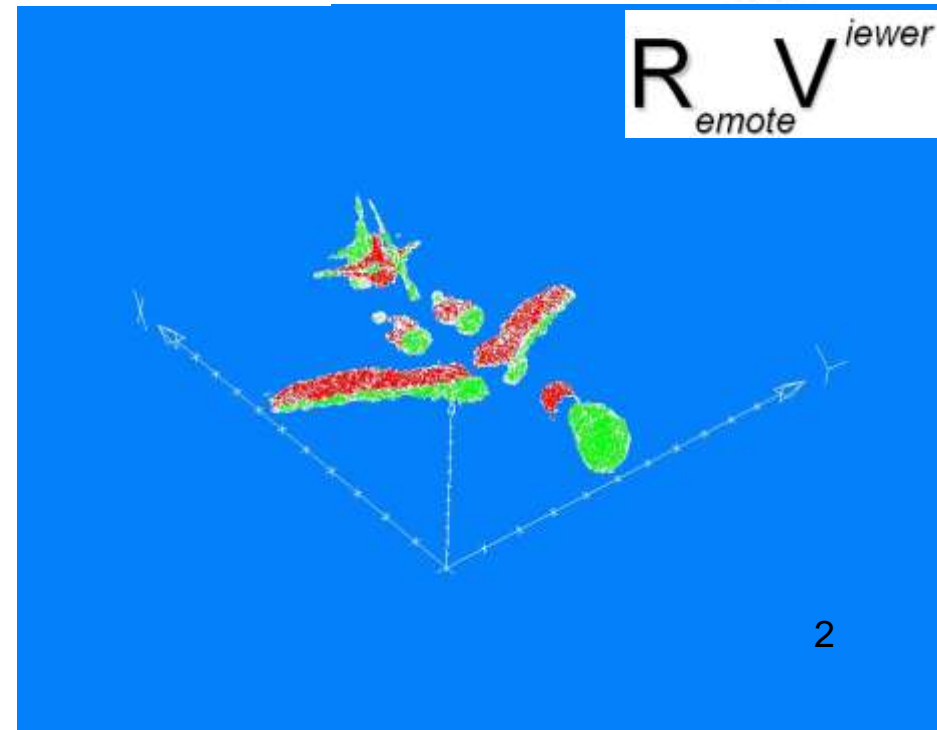


- MARPLE3D

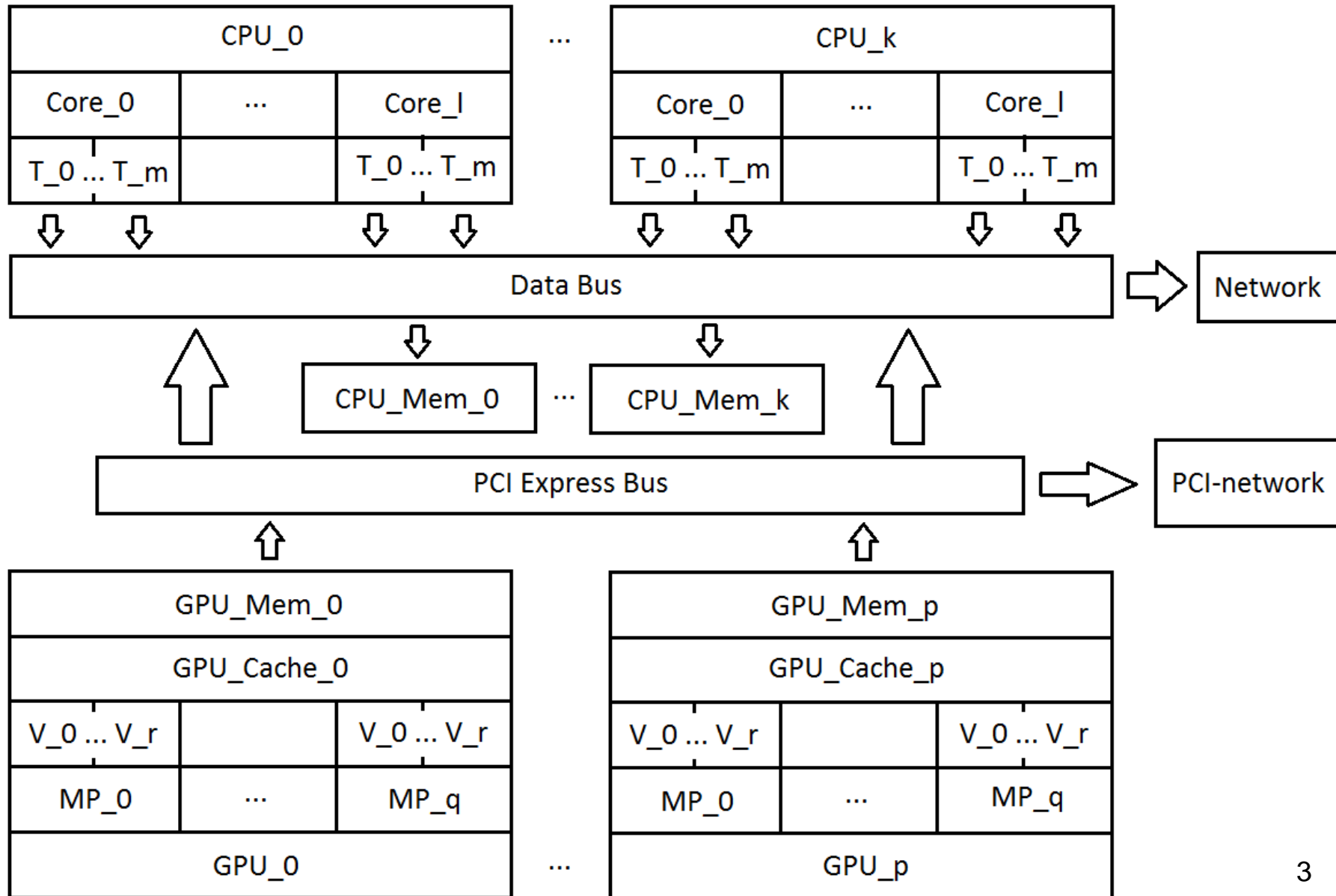
- Декомпозиция сеток



- Автоматизация разработки параллельных программ DVM, DVMH



Структура гибридной ВС



Три класса вычислительных систем (ВС)

- настольные и/или персональные компьютеры (ПК),
- бюджетные кластеры (БК) лабораторий и небольших организаций,
- суперкомпьютерные вычислительные системы (СКВС) крупных организаций и центров коллективного пользования (ЦКП).

В основном ВС отличаются:

- По количеству независимых вычислительных узлов
- По способам их интеграции в единую ВС.

Архитектура современного вычислительного узла этих систем де-факто является *гибридной*.

Программные средства гибридной архитектуры

Операционные системы (32/64 bit):

- **Unix и Linux**
- **Windows**

Основные языки разработки:

- **C/C++**
- **Fortran**

Основной параллельный инструментарий:

- **GRID**
- **MPI**
- **OpenMP**
- **Shmem**

Программное обеспечение GPU:

- **CUDA драйвер**
- **CUDA Toolkit**
- **CUDA SDK**
- **OpenCL**
- **OpenACC**

Основной полигон:
Суперкомпьютер К-100
(ИПМ им. М.В.Келдыша РАН)

К-100 - масштабируемый гибридный
вычислительный комплекс
на базе серийно выпускаемых
графических ускорителей GP GPU
пиковой производительностью ~
107 TFlop/s



64 вычислительных узла
128 процессоров Xeon X5670 (768 ядер)
ОП 6144 ГБ,

192 графических процессора Fermi C2050
(86016 ядер CUDA и 576 ГБ спец. памяти),

8 коммутаторов PCI-Express с пропускной способностью до 700 МБ/с
(ФГУП «Квант»).

Каверны, регулярная сетка

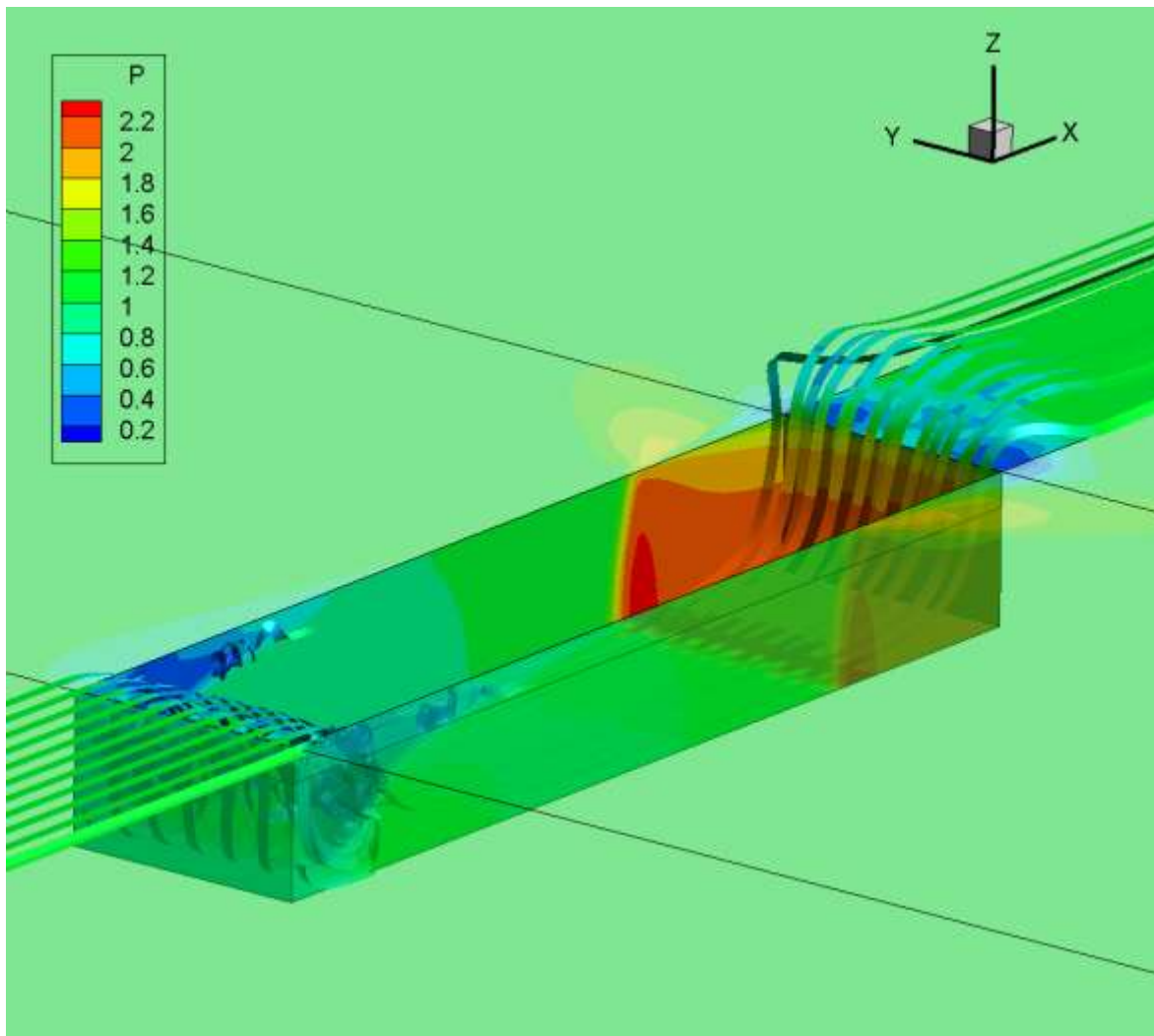


<http://www.keldysh.ru/events/fly/>



<http://www.buran.ru/htm/terlost.htm>

Моделирование течения в каверне

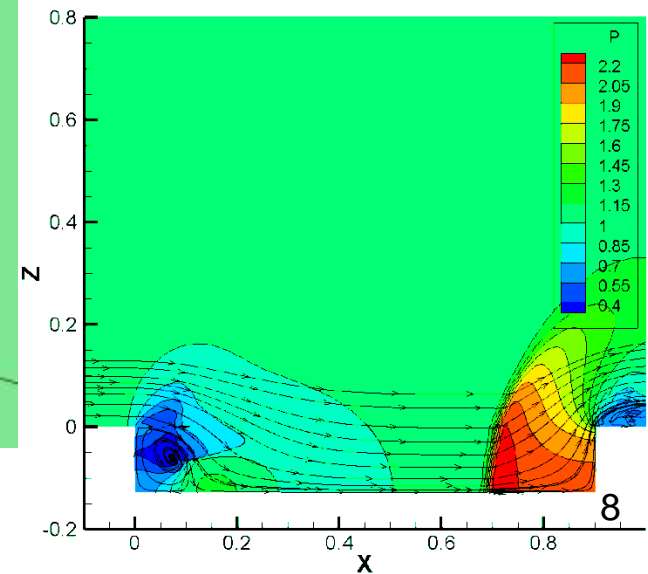


Мгновенные линии тока в центральном сечении каверны

Уровни давления

Число Маха набегающего потока $M=0.8$

число Рейнольдса $Re=10^6$



Эффективность параллельного расчета при различном числе ускорителей (GPU) На СК «Ломоносов»

Сетка: 1024 блока, общее число ячеек сетки – 32 млн

GPUs	4	8	16	32	64	128	256	512
Time, s	414.8	209.8	110	62.9	33.9	16.6	8.7	4.76
SpeedUp	4.0	7.9	15.1	26.4	48.9	100.0	190.7	348.6
Efficiency %	100.0	98.9	94.3	82.4	76.5	78.1	74.5	68.1

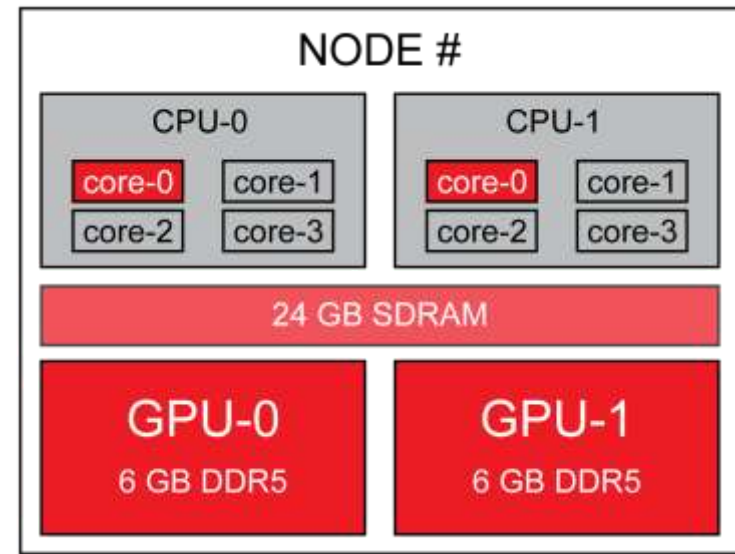
Расчет только на GPU,
CPU-ядра – исключительно
управление графическими
процессорами.

Данные располагаются целиком
в памяти GPU

14.25 - Ускорение 512 GPU (4.76 с)
по сравнению 512 ядрами CPU (67.68 с)

420 000 000 ячеек

67% эффективности на 684 GPU
(по отношению к 16 GPU)



Моделирование течения в каверне

Сравнение производительности GPU и CPU

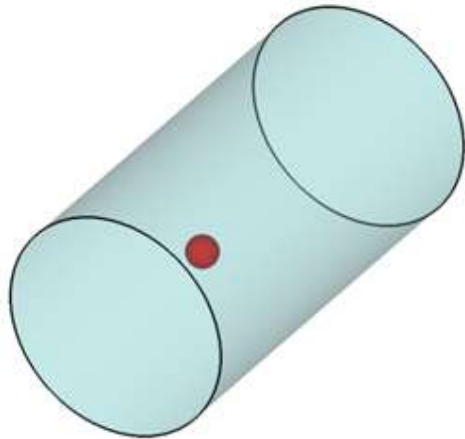
Размер сетки ~400 млн. узлов

	512 GPUs (256 Nodes)	512 CPU-cores (64 Nodes)
Time, s	4.76	67.68
SpeedUp		14.25

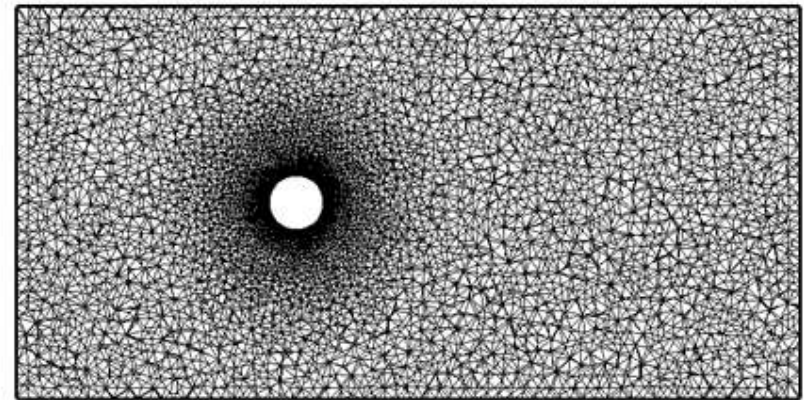
Таким образом, при использовании только CPU-ядер потребовалось бы в 3.5
раза больше процессорных узлов – порядка 900
(больше 7000 ядер)

Обтекание сферы

Расчетная область задачи представляет собой цилиндр круглого сечения (длина цилиндра $L = 14$, диаметр торцевой окружности $D = 7$), внутри которого находится сфера единичного диаметра. Ось содержащего расчетную область цилиндра сонаправлена с осью OX и вектором скорости набегающего потока. Число Маха в набегающем потоке равно $M = 0.1$, что позволяет вести сравнение результатов с точным решением аналогичной задачи для несжимаемой жидкости.



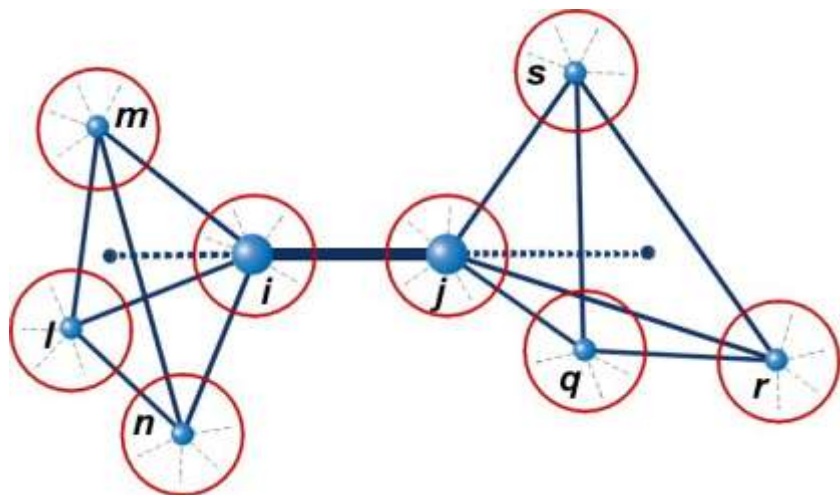
а) расчетная область задачи



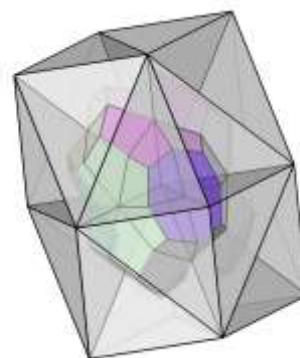
б) сетка в срединном сечении расчетной области

Дискретная модель расчетной области представляет собой сгущающуюся к поверхности обтекаемого тела сетку, содержащую 119286 узлов и 679339 тетраэдров. Использование сетки с однотипными элементами в данном случае обусловлено желанием минимизировать потери производительности программного обеспечения, *связанные с возможным дисбалансом загрузки при вычислении конвективных потоков.*

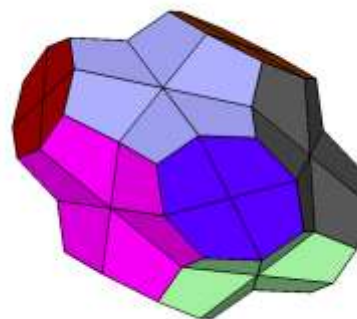
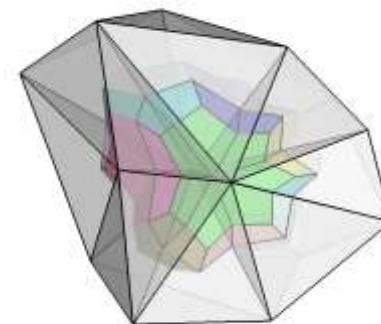
3D контрольные объемы



Декартова сетка



Неструктурированная тетраэдральная сетка



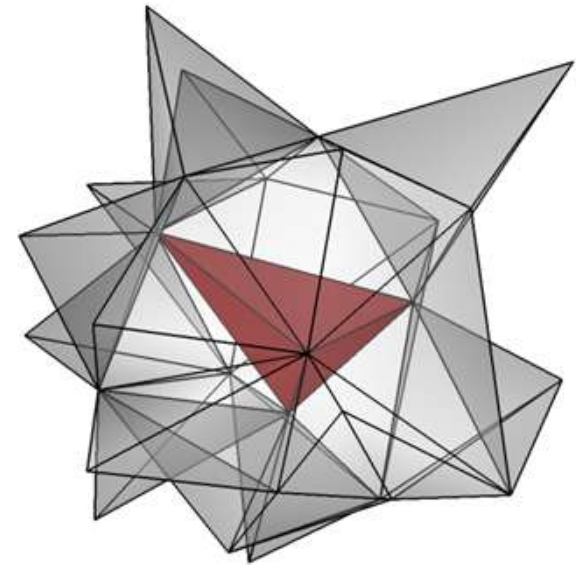
3D шаблон высокого порядка:
Противопоточные тетраэдры + соседи

Обтекание тела сверхзвуковым потоком газа

Метод контрольных объёмов

Повышение порядка точности

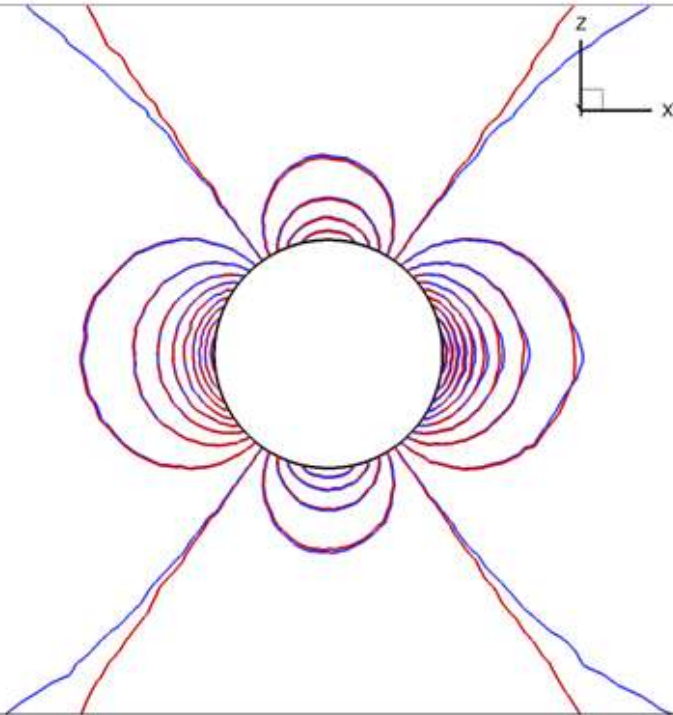
$$\begin{pmatrix} A_{1D_{0,0}} A_{1D_{0,1}} \dots A_{1D_{0,NG_{1D}-1}} \\ A_{1D_{1,0}} A_{1D_{1,1}} \dots A_{1D_{1,NG_{1D}-1}} \\ A_{1D_{2,0}} A_{1D_{2,1}} \dots A_{1D_{2,NG_{1D}-1}} \\ \dots \\ A_{1D_{S,0}} A_{1D_{S,1}} \dots A_{1D_{S,NG_{1D}-1}} \end{pmatrix} \begin{pmatrix} f_{G_{1D_0}} - f_{1D} \\ f_{G_{1D_1}} - f_{1D} \\ f_{G_{1D_2}} - f_{1D} \\ \dots \\ \dots \\ \dots \\ f_{G_{1D_{NG_{1D}-1}} - f_{1D}} \end{pmatrix} = \begin{pmatrix} C_{f_{1D}}^0 \\ C_{f_{1D}}^1 \\ C_{f_{1D}}^2 \\ C_{f_{1D}}^3 \\ C_{f_{1D}}^4 \\ C_{f_{1D}}^5 \\ C_{f_{1D}}^6 \\ C_{f_{1D}}^7 \\ C_{f_{1D}}^8 \end{pmatrix}$$



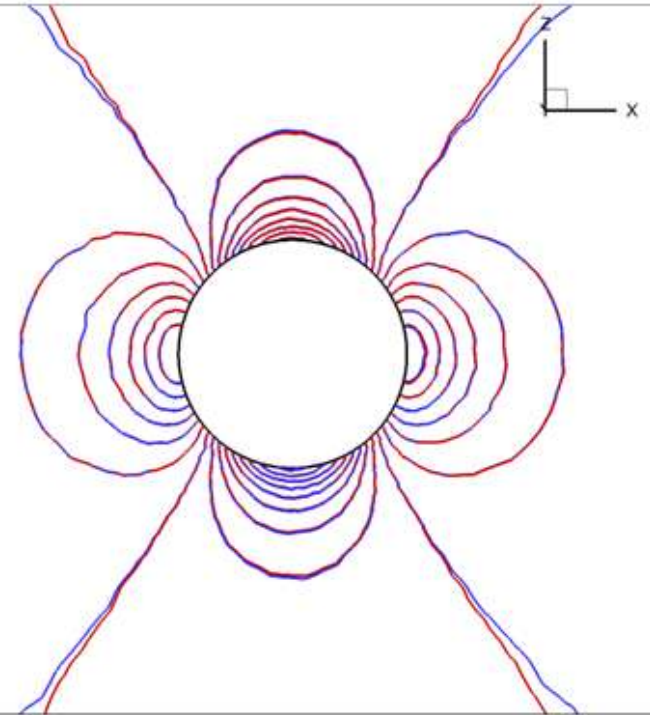
Результаты расчётов

И.В.Абалакин, С.А. Суков

Картины течения вблизи поверхности обтекаемого тела, построенные по результатам расчета задачи методом установления, на фоне точного решения аналогичной задачи для случая несжимаемой жидкости:



а) изолинии продольной скорости



б) изолинии давления

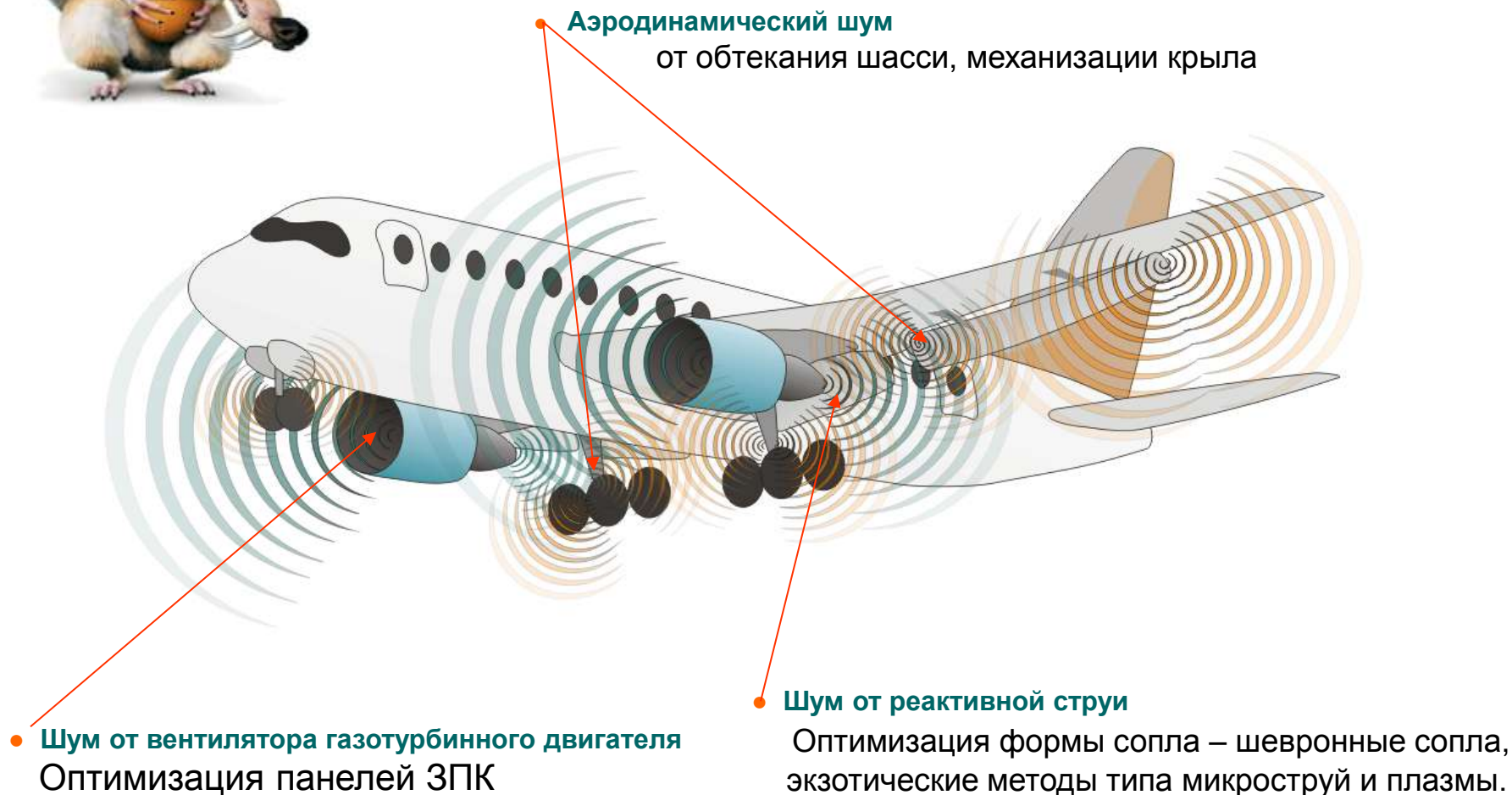
Сравнение результатов численного расчета с точным решением задачи в срединном сечении расчетной области ($Y=0$).
Линии синего цвета – результат расчета, линии красного цвета – точное решение.

Комплекс программ NOISEtte для расчета задач аэродинамики и аэроакустики

Т.К. Козубская, И.В.Абалакин,
А.В.Горобец, С.А.Суков,
А.П.Дубень, П.А.Бахвалов



Разработан в ИММ / ИПМ им. М.В. Келдыша РАН,
при участии INRIA, Sophia-Antipolis, Франция

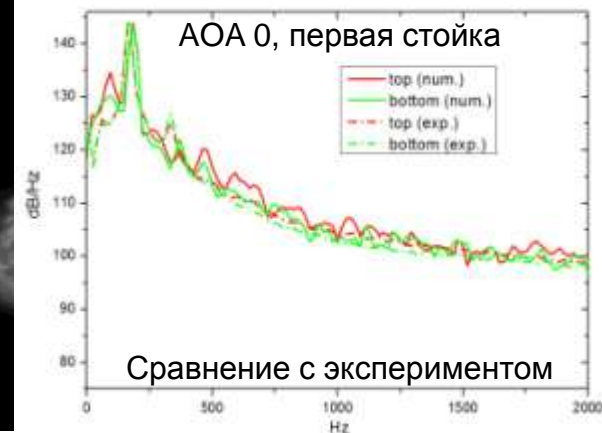
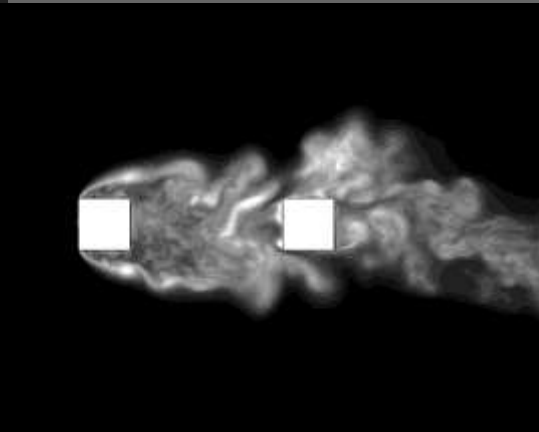
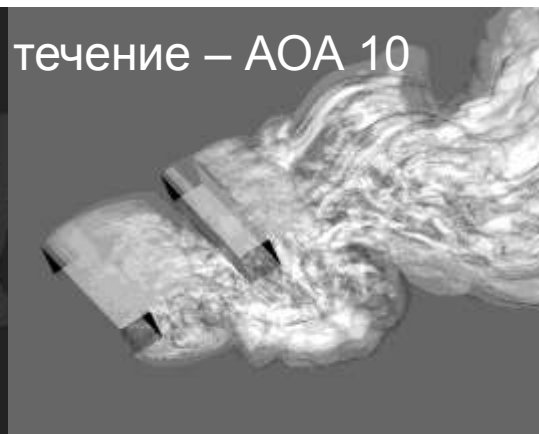
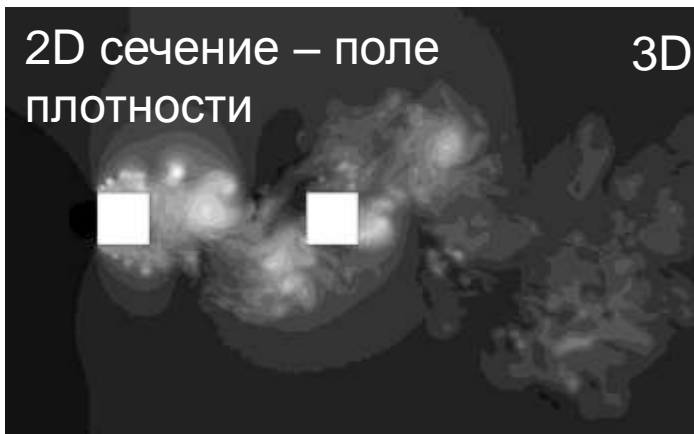


Взаимодействие турбулентности с двумя стойками

Т.К. Козубская, И.В.Абалакин,
А.В.Городец, С.А.Суков,
А.П.Дубень, П.А.Бахвалов

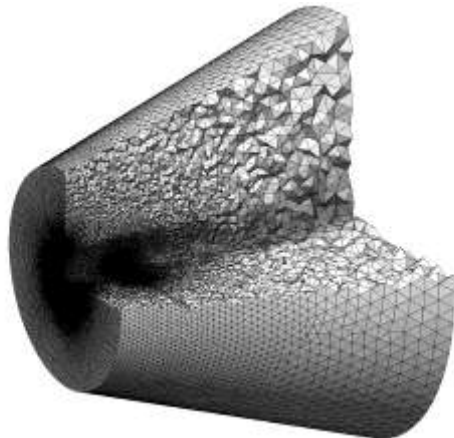
Воспроизводятся механизмы возникновения аэродинамического шума при обтекании стоек шасси самолета

- $M = 0.2$; $Re = 200000$; углы атаки: 0, 10, 15; сетки до 13 млн. узлов, 80 млн. тетраэдров
- **Параметры численного метода:** Схема повышенного порядка с центрами в узлах, неявное интегрирование по времени 2го порядка, линеаризация по Ньютону, модель турбулентности DDES, далее поле FW/H



Расчеты по локализации источников звука в турбулентном следе

Совместные с ЦАГИ исследования обтекания цилиндра дозвуковой струей, имитируется обтекание стойки шасси самолета



Неструктурированная
адаптированная
тетраэдральная сетка



Изоповерхности модуля
скорости

Воспроизводятся механизмы генерации аэродинамического шума при обтекании шасси самолета.

Расчеты и соответствующие эксперименты, выполненные в ЦАГИ, нацелены на определение позиции источников шума в турбулентном следе за цилиндром.

- **Параметры задачи**

Число Маха: $M=0.206$

Число Рейнольдса: $Re = 14000$

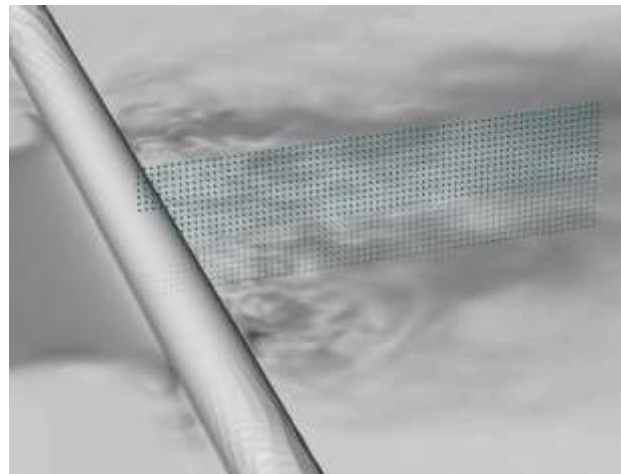
Размер сетки: до 16 млн. узлов, **100 млн.**

тетраэдров

Период интегрирования по времени: 1700

Количество контрольных точек: 1400

Задействовано процессоров: **до 12800**



Расположение массива
контрольных точек
в следе за цилиндром



Поле модуля скорости, вид сверху



Поле модуля скорости, вид сбоку

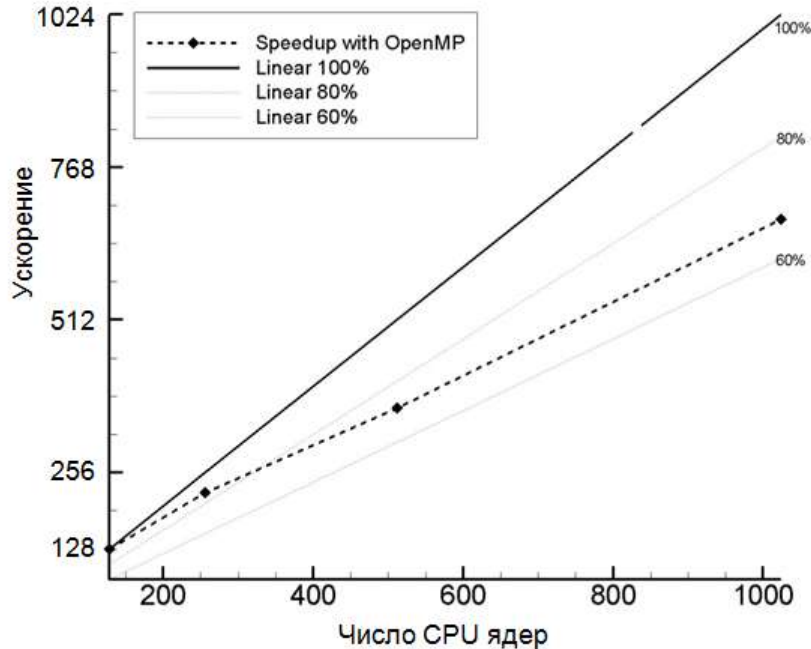


Поле завихренности, вид сбоку

Тесты на суперкомпьютере Ломоносов до 12800 ядер

Двухуровневое распараллеливание MPI+OpenMP

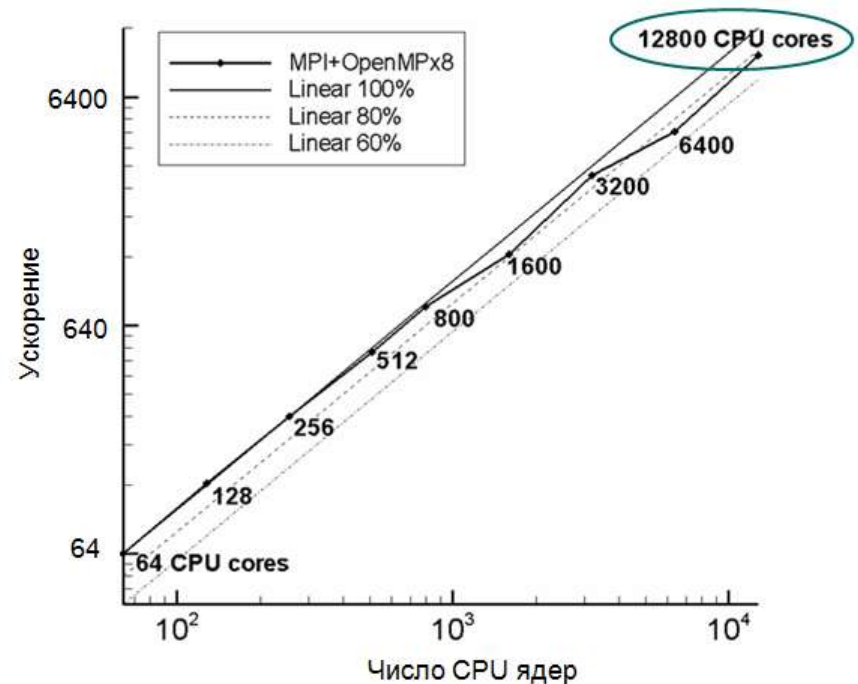
- **MPI работает на первом уровне**, объединяя группу параллельных процессов, запущенных на разных узлах суперкомпьютера.
- **OpenMP работает на втором уровне – внутри узлов**, обеспечивая распараллеливание на многопроцессорных узлах в рамках модели с общей памятью.



Ускорение OpenMP: число MPI процессов фиксировано 128, число OpenMP нитей меняется от 1 до 8.

Задача об обтекании цилиндра струей

Тесты выполнены на примере задачи об обтекании цилиндра дозвуковой струей. Сетка 16M узлов, 100M тетраэдров, 4-шаговый явный метод Runge-Kutta 4-го порядка по времени, схема повышенного порядка по пространству.

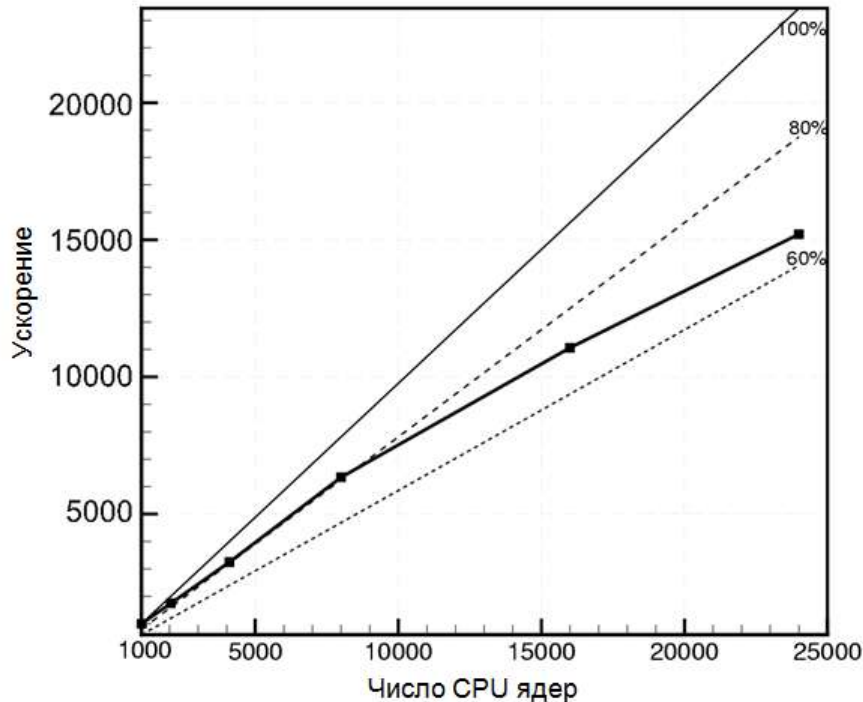


Ускорение MPI, 8 OpenMP нитей. (логарифмическая шкала)

Характерное время вычислений: 26.8 сек. на шаг по времени на 64 ядрах, 0.175 сек. на 12800, ускорение 152.6.

Тесты на суперкомпьютере Ломоносов до 30 000 ядер

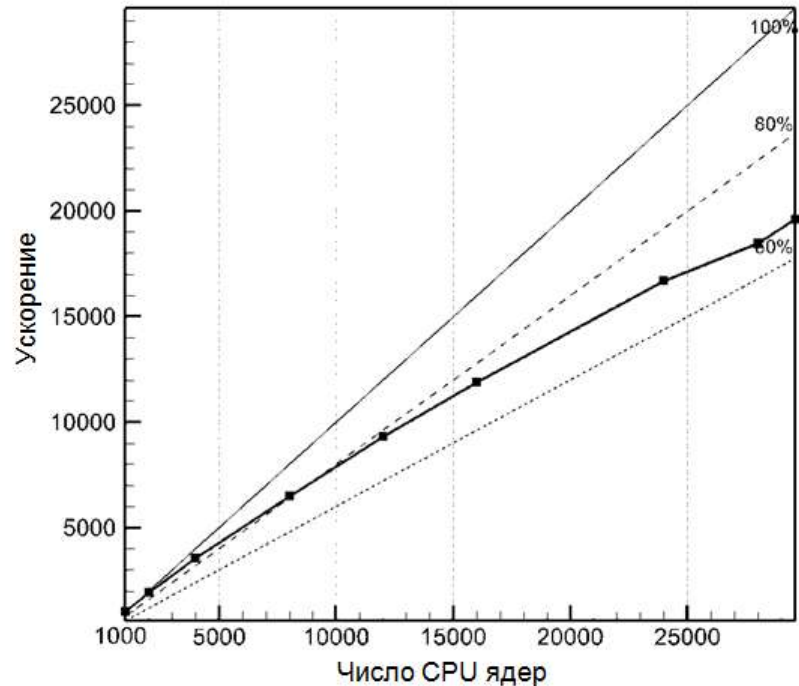
- **Сетка 16 миллионов узлов / 100 млн. тетраэдров**
Схема с центрами в узлах повышенного порядка



Характерное время вычислений:

на 1 шаг Рунге-Кутты 0.65 сек. на 1024 ядрах и 0.04 сек. на 24000 ядрах, ускорение ~16, параллельная эффективность ~65%

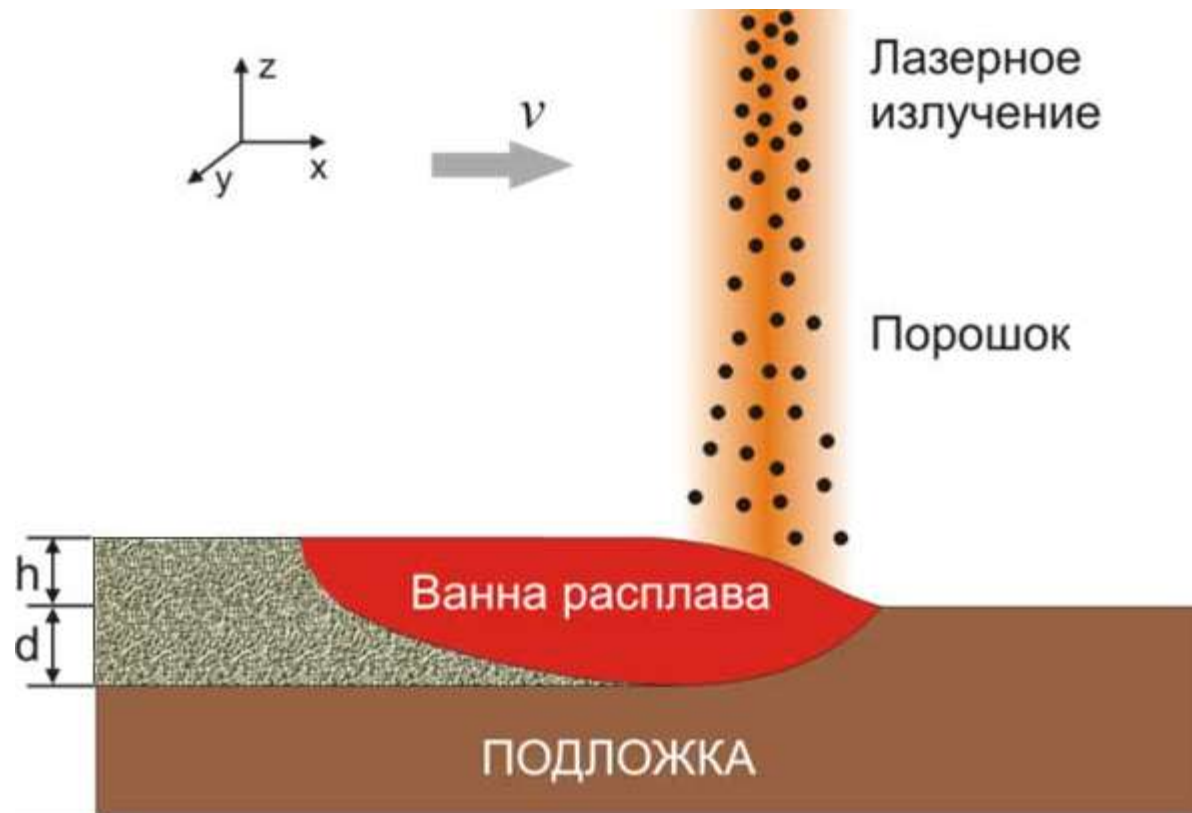
- **Сетка 1.5 миллиарда контрольных объемов**
Схема с центрами в элементах



Характерное время вычислений:

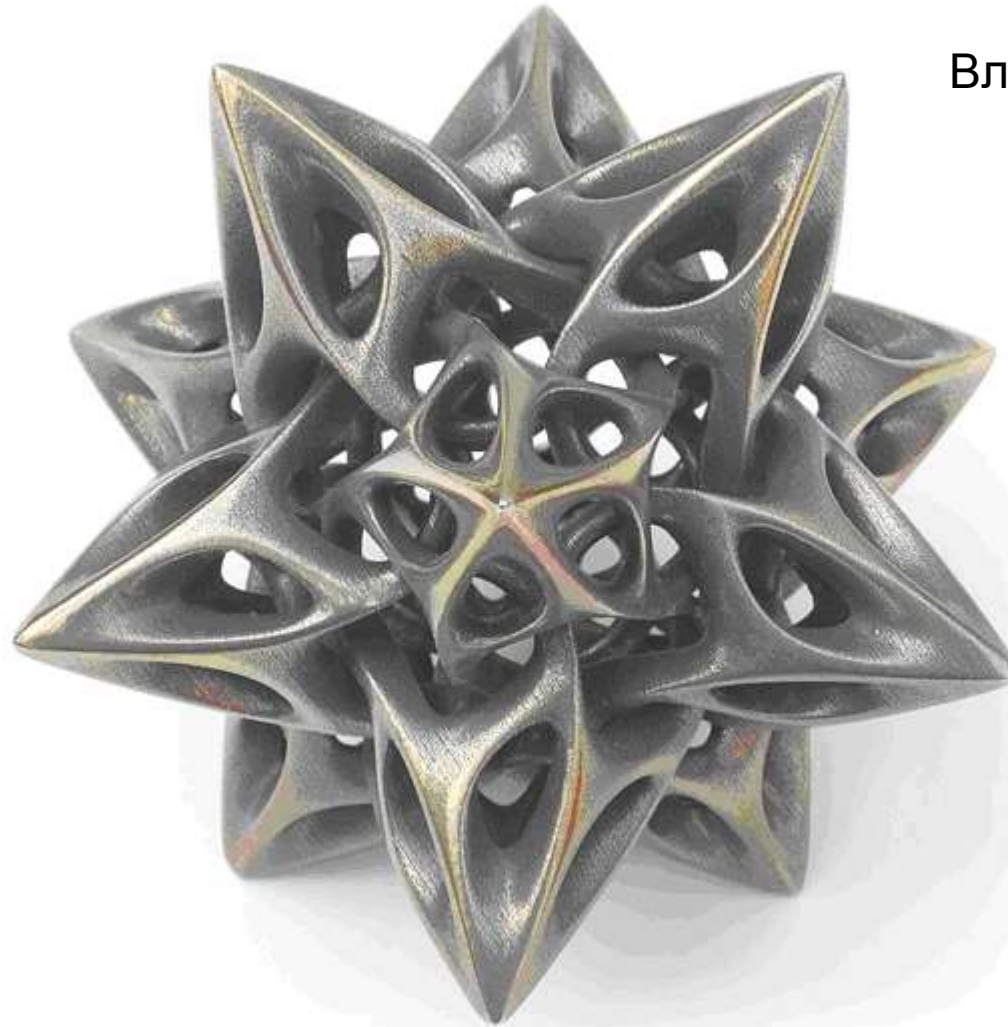
0.4 сек. на шаг по времени на 1000 ядрах, 0.02 сек. на 29600, ускорение 19.6, параллельная эффективность 67%.

Моделирование температурного поля при послойном нанесении порошка на подложку методом СЛС



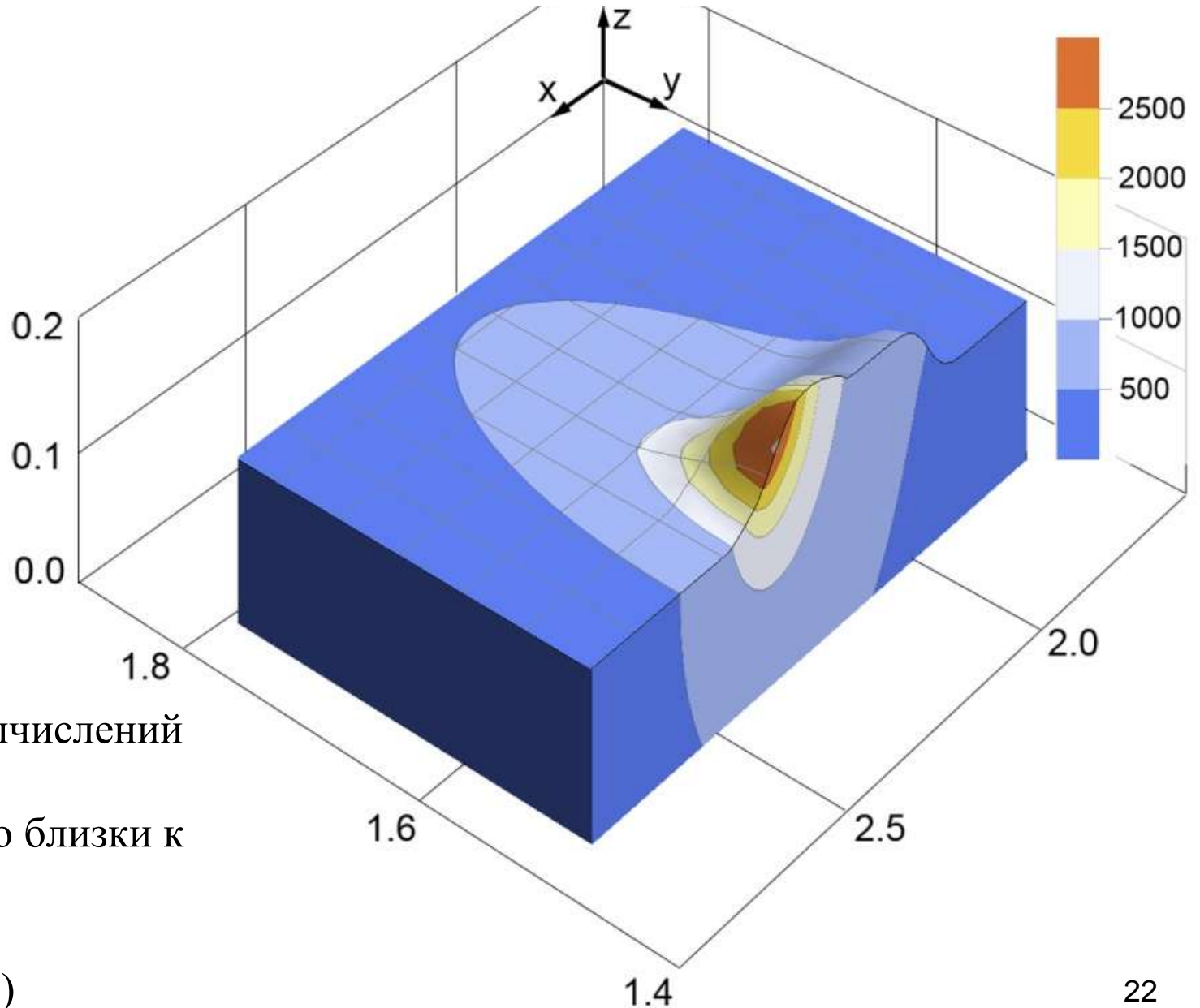
Селективное лазерное спекание *сталь и бронза : додекаэдр*

Владимир Булатов
2007



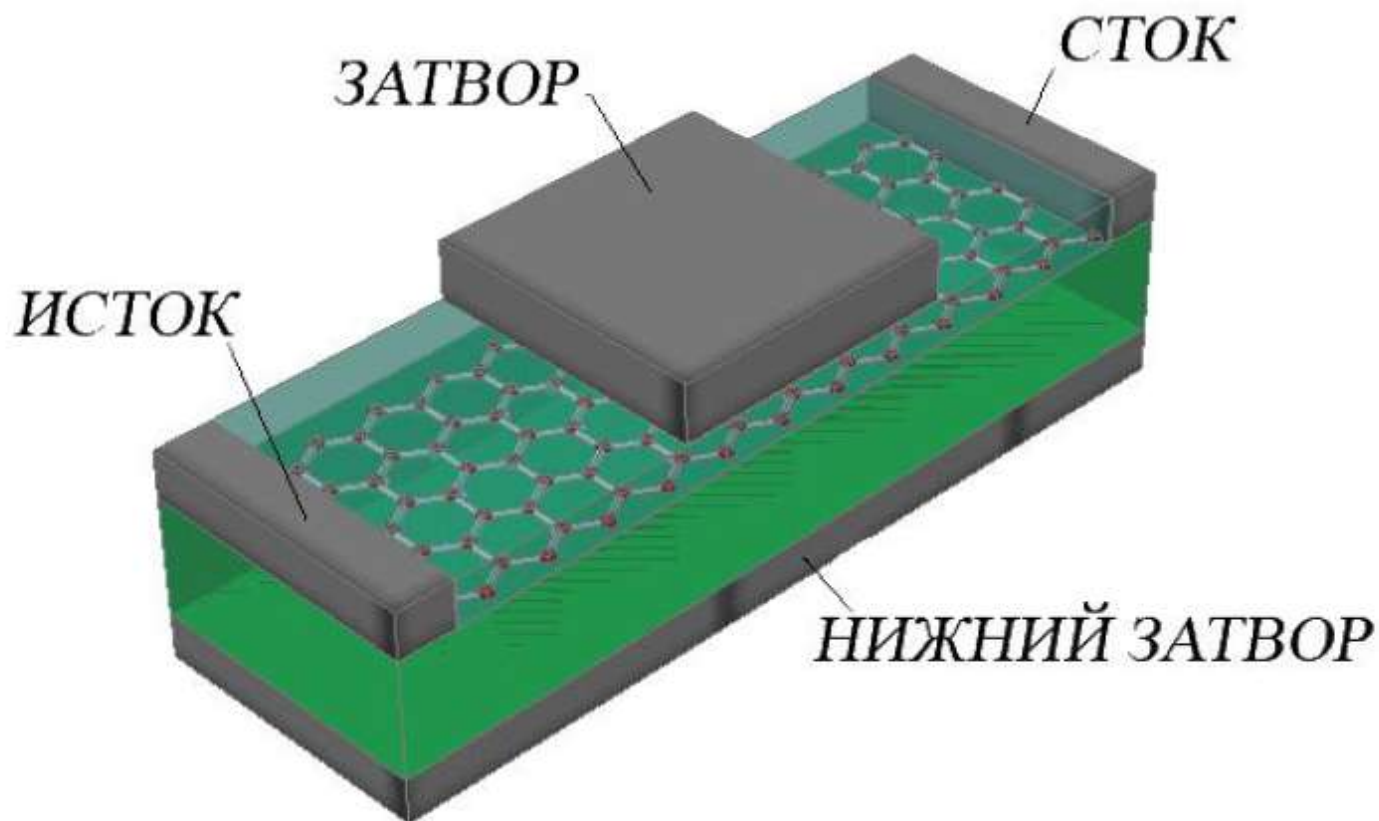
<http://www.bridgesmathart.org/art-exhibits/jmm08/bulatov.html>
<http://honeybuild.com/guides/selective-laser-sintering/>

Сопоставление с экспериментом



Результаты вычислений
качественно и
количественно близки к
измеренным в
эксперименте
(ИПЛИТ РАН)

Моделирование полевого нанотранзистора на базе графена на основе квантовой гидродинамики



Структура полевого транзистора на основе графена.

Погрешность вычислений

Таблица 1. Относительная погрешность вычислений на последовательности сеток.

Программа	Сетка 1	ОЗУ, %	Сетка 2	ОЗУ, %	Δ , %
GIMM_APP_Silicon_Field_Transistor	513x513x513	52.3	2049x2049x4609	52.3	2.394
GIMM_APP_Graphene_FET_1	4000x16000	48.3	48000x192000	48.3	2.093
GIMM_APP_Graphene_FET_2	4000x16000	49.1	48000x192000	49.1	2.127
GIMM_APP_Quantum_Bit_States	121x121x401	52.5	481x481x3601	52.5	2.153
GIMM_APP_Silicon_Field_Emitter_2D	20633104	57.6	3033062542	58.3	2.451
GIMM_APP_Silicon_Field_Emitter_3D	19537920	60.0	2803693118	61.5	2.318
GIMM_APP_Graphene_Field_Emitter	2401x2401	53.1	28801x28801	53.1	2.074
GIMM_APP_Carbon_Field_Emitter_2D	18691072	56.9	2747583128	57.4	2.261
GIMM_APP_Carbon_Field_Emitter_3D	24484032	63.2	3513454428	63.6	2.382
GIMM_APP_Powder_2D	4001x4001	46.3	48001x48001	46.3	1.938
GIMM_APP_Powder_3D	161x161x161	48.2	641x641x1441	48.2	2.155
GIMM_APP_Crystal_2D	1001x1001	46.5	12001x12001	46.5	1.954
GIMM_APP_Crystal_3D	101x101x101	48.6	401x401x901	48.6	2.231

Проведено исследование точности 13 разработанных прикладных программ.

**Относительная точность вычислений контролируется
методом расчёта на последовательности сеток.**

**Ошибка воспроизведения решений, заложенных в математических моделях,
не превышает 3%**

Эффективность вычислений

Таблица 2. Эффективность параллельных вычислений, измеренная на последовательности сеток.

Программа	Сетка 1	ОЗУ, %	Сетка 2	ОЗУ, %	E, %
GIMM_APP_Silicon_Field_Transistor	513x513x513	52.3	2049x2049x4609	52.3	83,2
GIMM_APP_Graphene_FET_1	4000x16000	48.3	48000x192000	48.3	87.4
GIMM_APP_Graphene_FET_2	4000x16000	49.1	48000x192000	49.1	86.8
GIMM_APP_Quantum_Bit_States	121x121x401	52.5	481x481x3601	52.5	81.4
GIMM_APP_Silicon_Field_Emitter_2D	20633104	57.6	3033062542	58.3	87.1
GIMM_APP_Silicon_Field_Emitter_3D	19537920	60.0	2803693118	61.5	84.3
GIMM_APP_Graphene_Field_Emitter	2401x2401	53.1	28801x28801	53.1	85.2
GIMM_APP_Carbon_Field_Emitter_2D	18691072	56.9	2747583128	57.4	86.8
GIMM_APP_Carbon_Field_Emitter_3D	24484032	63.2	3513454428	63.6	84.7
GIMM_APP_Powder_2D	4001x4001	46.3	48001x48001	46.3	82.5
GIMM_APP_Powder_3D	161x161x161	48.2	641x641x1441	48.2	81.0
GIMM_APP_Crystal_2D	1001x1001	46.5	12001x12001	46.5	83.3
GIMM_APP_Crystal_3D	101x101x101	48.6	401x401x901	48.6	82.7

Моделирование задач управляемого ядерного синтеза на ЭВМ эксафлопного уровня производительности

Эксафлопная инициатива: NuFuSE - Nuclear Fusion Simulations at Exascale

Россия: Проект РФФИ 11-07-93939-ДЖИ8-а

Проблема:

Полномасштабное по времени и пространству моделирование плазмы в установках управляемого термоядерного синтеза на основе кинетических и радиационно-магнитогидродинамических моделей неравновесной плазмы высокой плотности энергии.

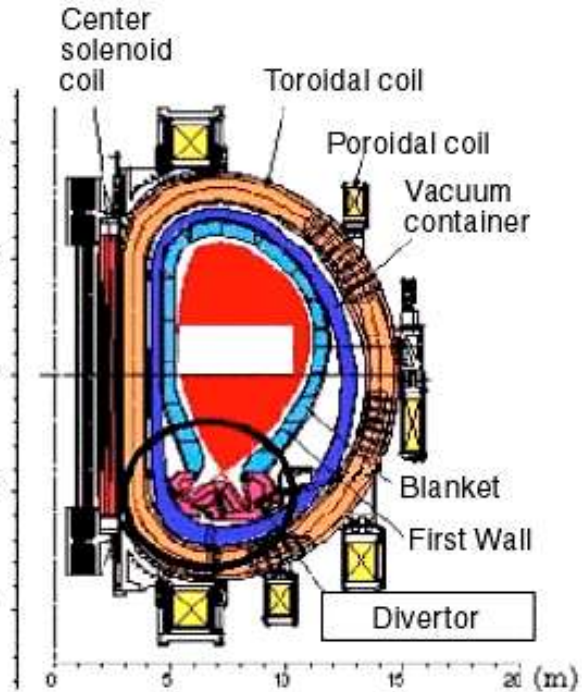
Задачи:

Создание и реализация на ЭВМ сверхвысокой производительности комплекса математических и компьютерных моделей динамики неравновесной плазмы для анализа пристеночных плазменных процессов в условиях экспериментальных термоядерных устройств, включая:

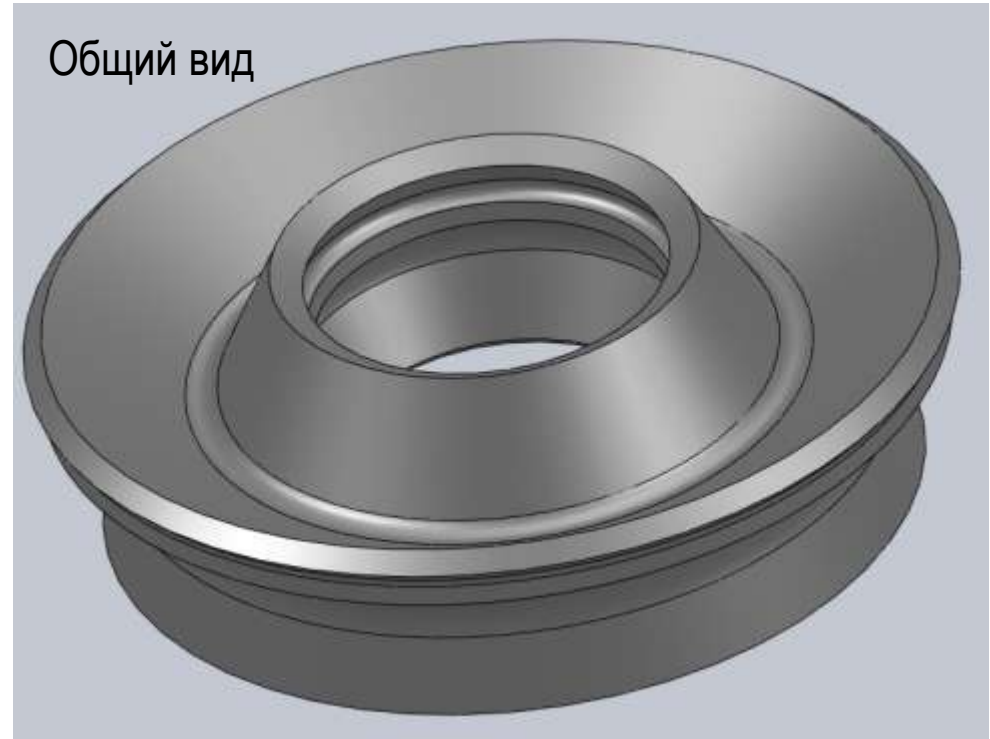
- самосогласованную трехмерную модель радиационной магнитной газовой динамики в двухжидкостном приближении,
- модель кинетики населенностей уровней ионов, включая неравновесное распределение электронов и радиационные процессы,
- модель динамики многокомпонентной плазмы,
- модели нестационарных процессов в плазме на основе кинетического уравнения Больцмана,
- многогрупповые по спектру модели переноса энергии излучением,
- статистические квантово-механические модели вещества.

Ближайшая задача – создание модели турбулентного
тепломассопереноса в диверторе ITER.

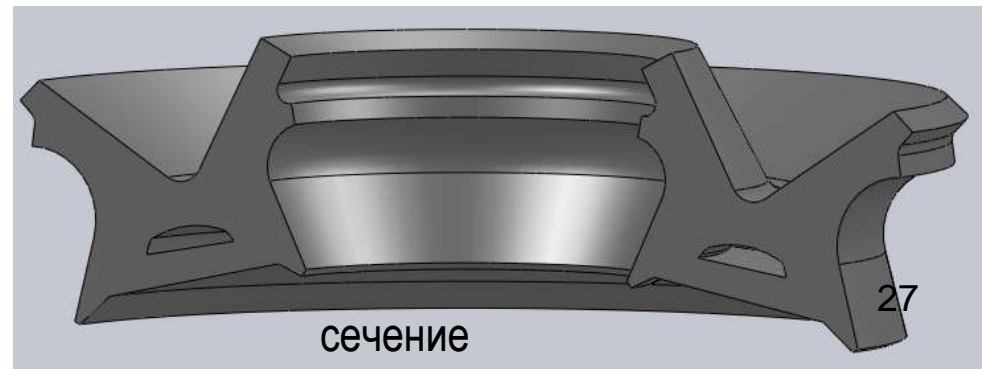
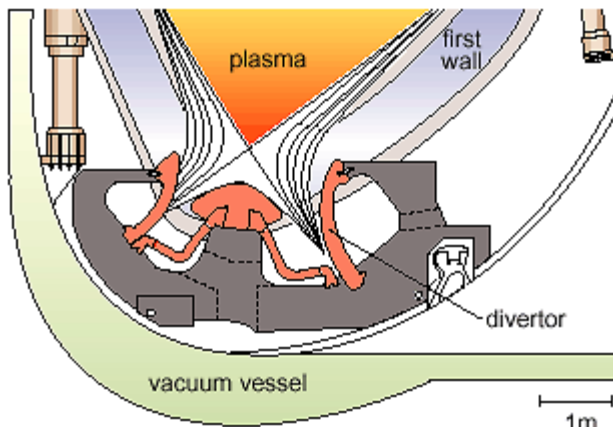
Поперечный разрез ITER



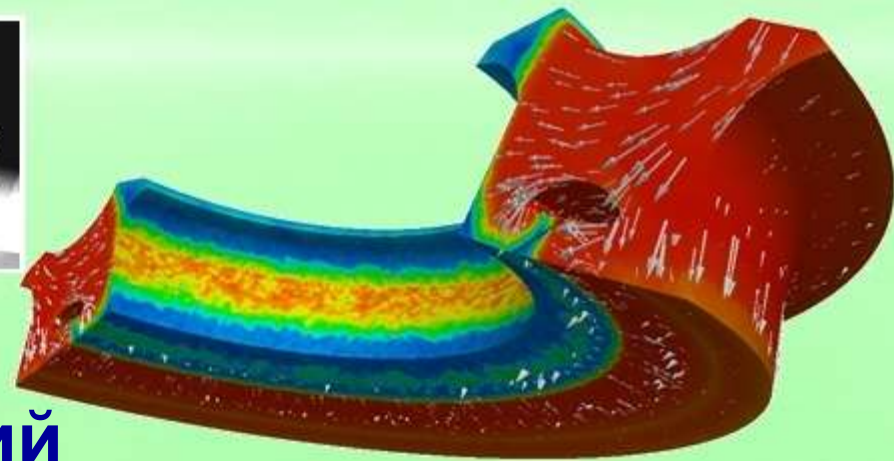
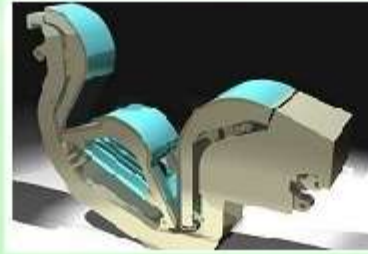
CAD-модель



Область дивертора

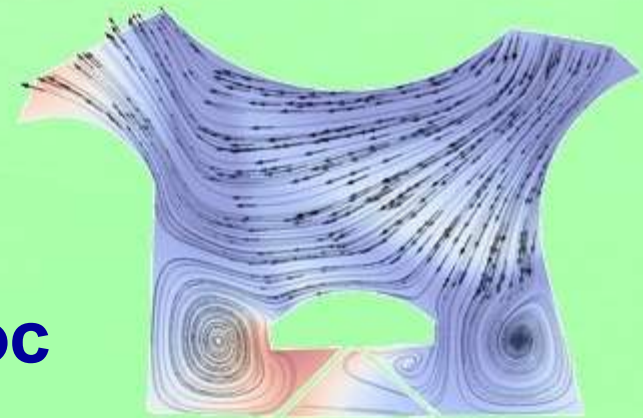


Моделирование газоплазменных потоков в диверторе токамака ITER



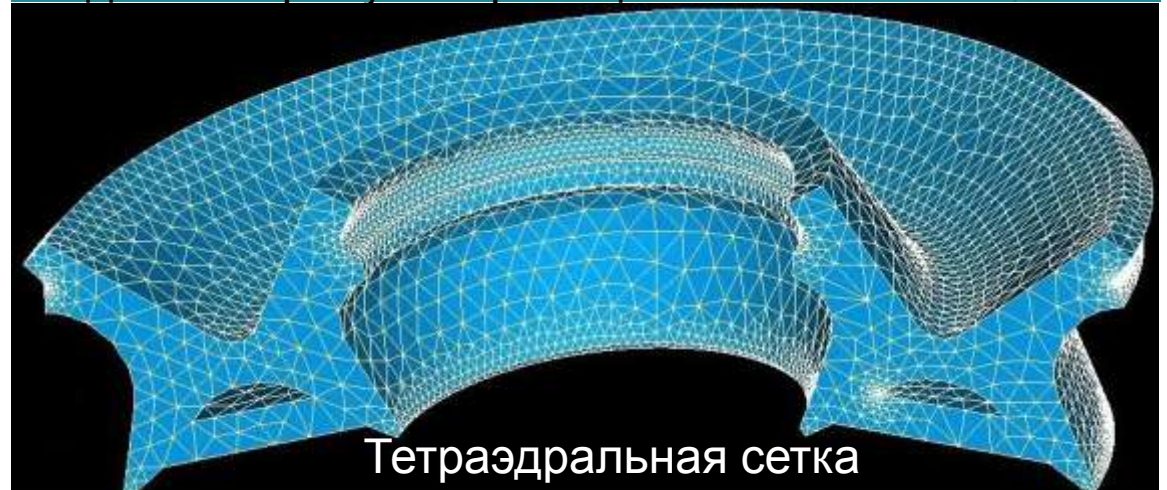
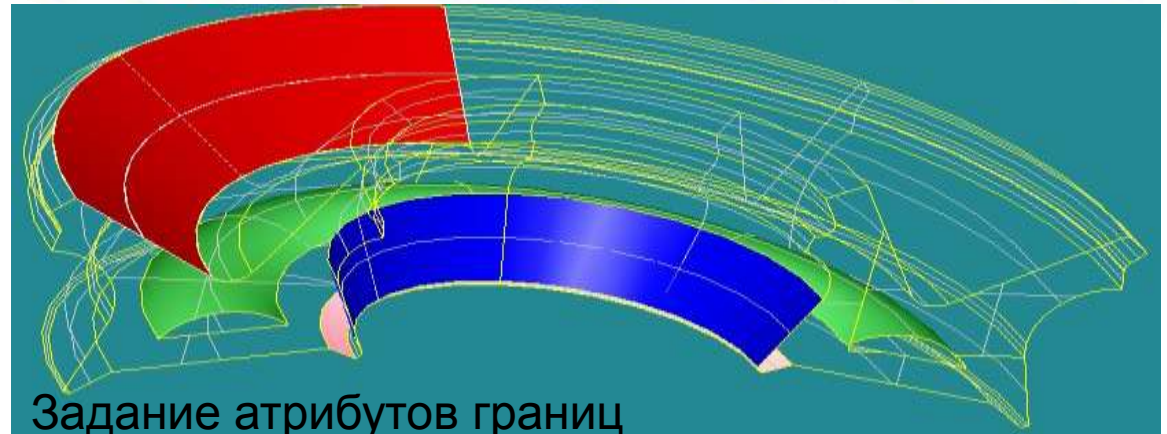
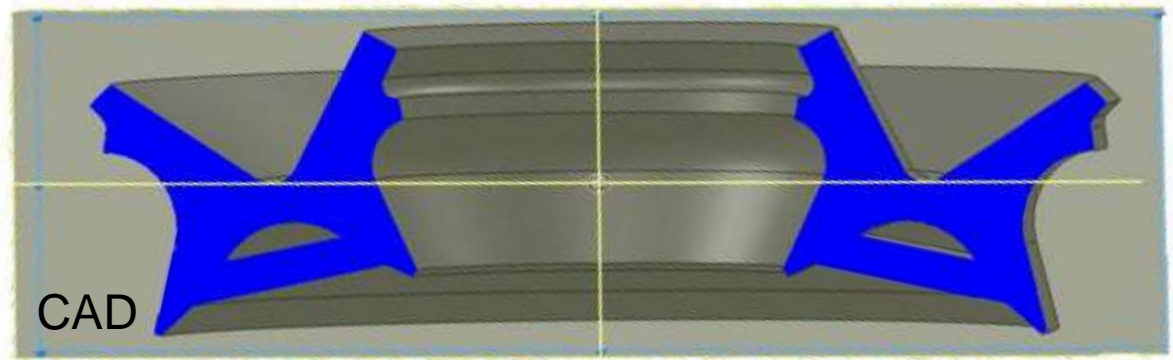
Полная система уравнений газовой динамики:

- Турбулентная вязкость
- Радиационный и кондуктивный теплоперенос
- Табличные УРС



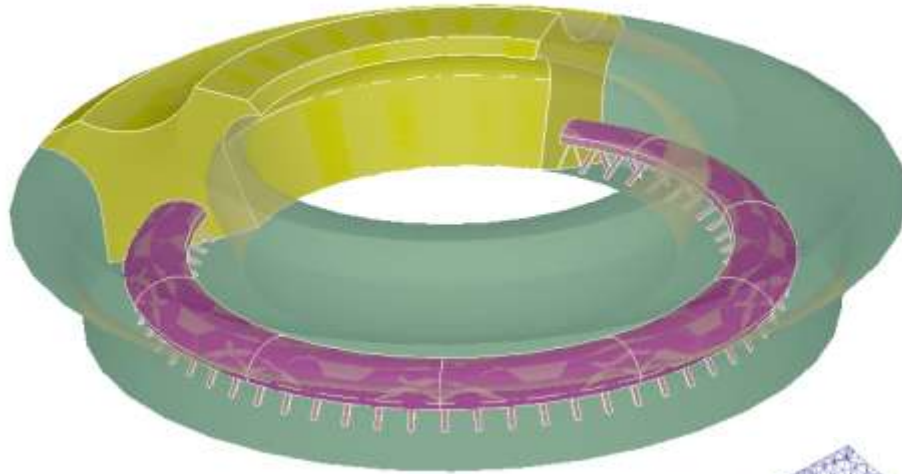
турбулентный
тепломассоперенос
в диверторе ITER :

от CAD - модели к
расчетной сетке



Показана первоначальная тетраэдральная сетка до измельчения (SALOME).
Результирующая сетка содержит 10^8 ячеек и более.

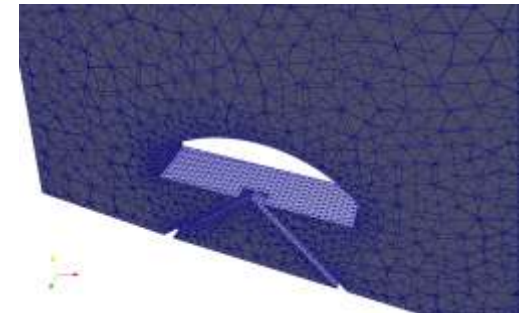
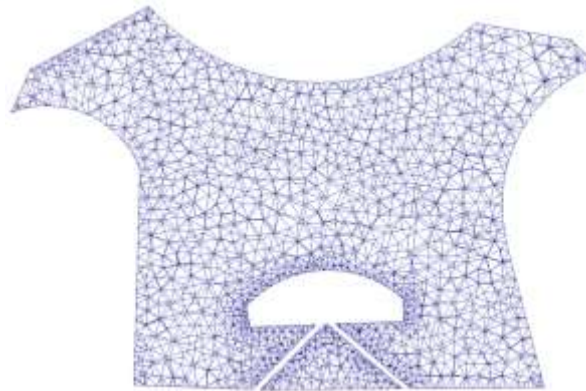
Дивертор токамака (divertor)



Область сложной геометрии

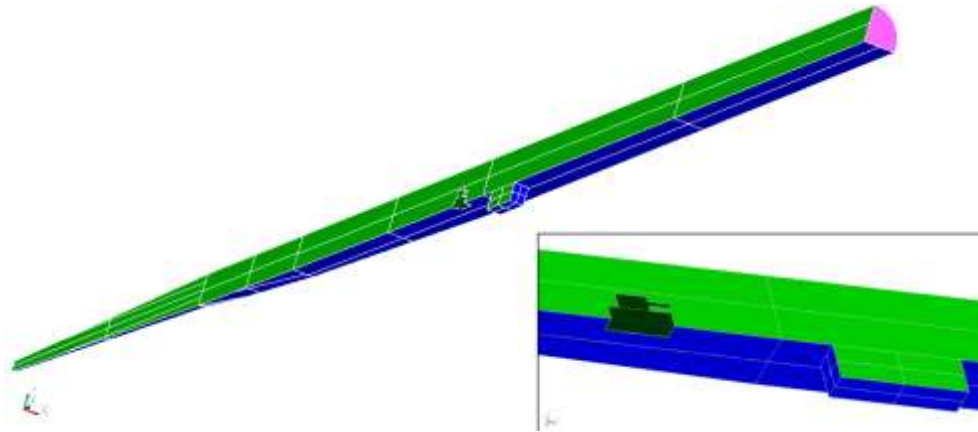
Тетраэдральная сетка
(3 миллиона ячеек)

Сгущение сетки вблизи
мелких объектов



- 256 доменов
- 8 ядер на декомпозицию

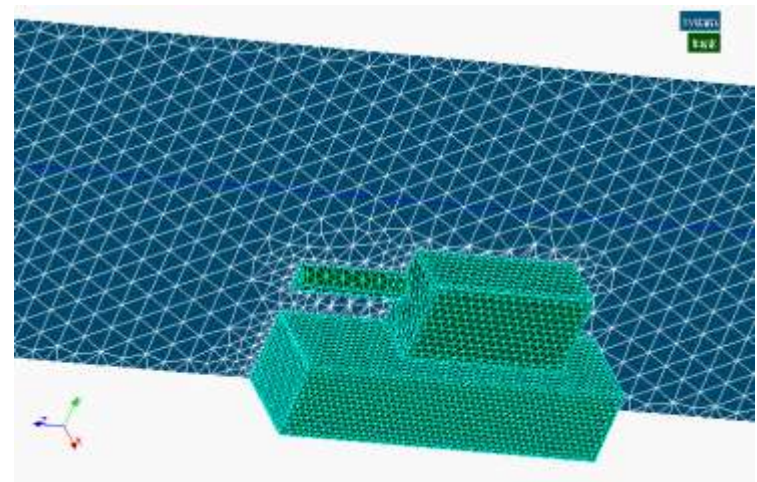
Ударная труба (tube)



Область сложной геометрии
Разномасштабная модель
Тетраэдральная сетка
(более 25 миллионов ячеек)

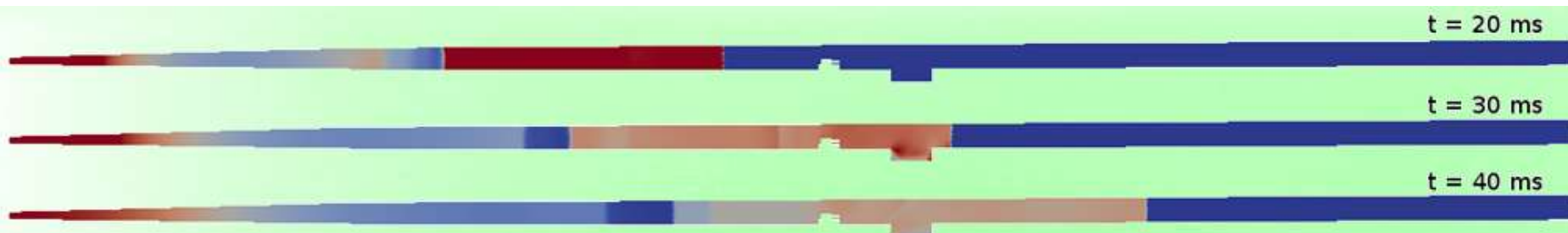
Сгущение сетки вблизи
мелких объектов

- 4096 доменов
- 16 ядер на геометрические алгоритмы
- 128 ядер на алгоритмы декомпозиции графов



Моделирование распространения ударной волны от взрыва химического взрывчатого вещества в протяженном сооружении с нетривиальной геометрией

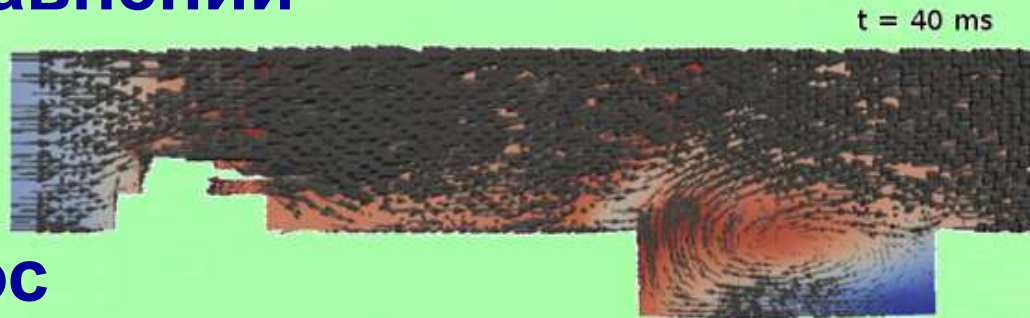
Проектируемый экспериментальный комплекс 500м x 20м (10см)



Полная система уравнений газовой динамики:

- Турбулентный тепло-массоперенос

- Многокомпонентность

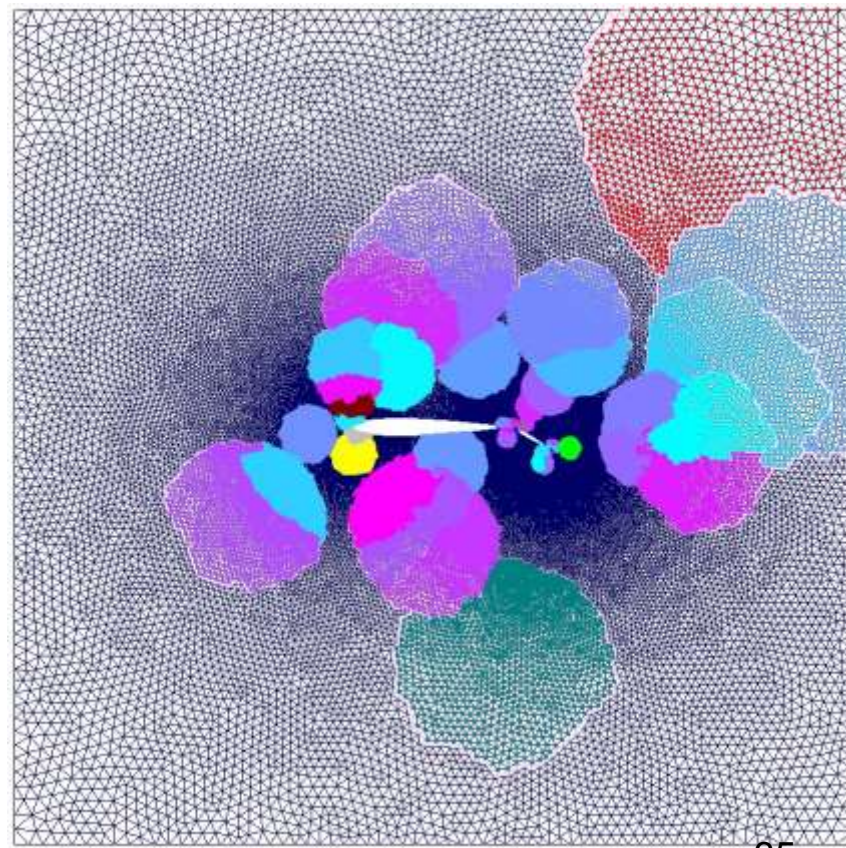
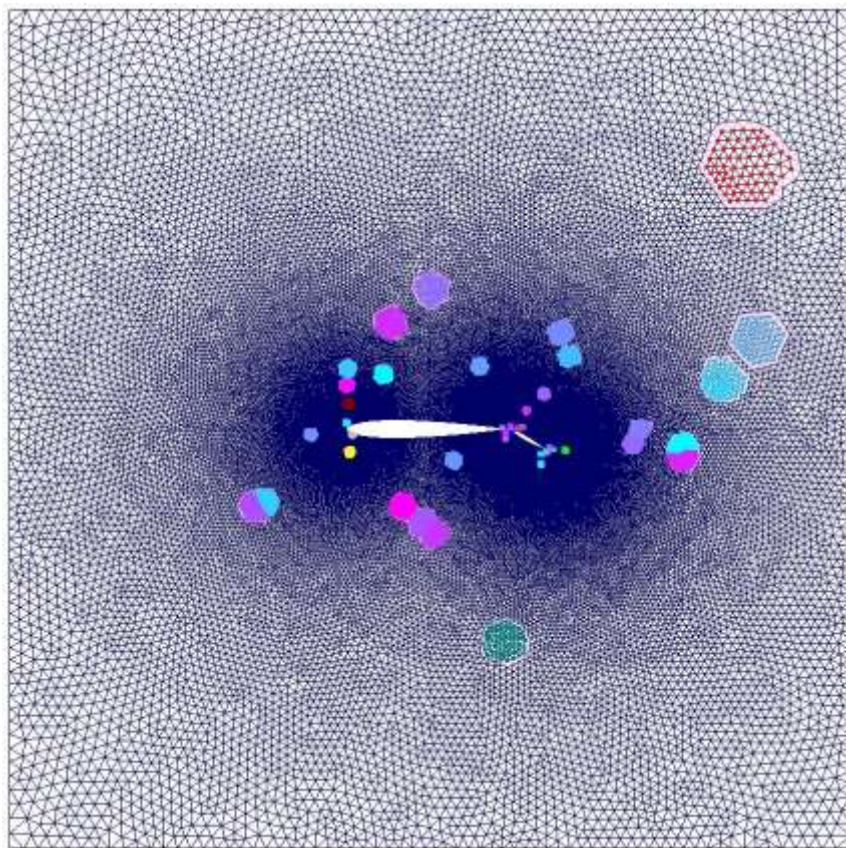


- **Дуальные графы:**
 - **число вершин $2.8 \cdot 10^6 - 1.2 \cdot 10^8$**
 - **число ребер $2.3 \cdot 10^7 - 1.0 \cdot 10^9$**
- **Сравнение разбиений и эффективности расчетов:**
 - **GRIDSPIDERPAR**
 - **PARMETIS**
 - **ZOLTAN**
 - **PT-SCOTCH**
- **Суперкомпьютеры:**
 - **MBC-100K (227,94 TFlop/s)**
 - **"Ломоносов" (1700 TFlops)**
 - **«Helios» (1524.1 TFlop/s)**

- На задачах газовой динамики проведено тестирование разбиений, полученных методами пакетов GRIDSPIDERPAR, PARMETIS, ZOLTAN и PT-SCOTCH
- Сравнивалась эффективность параллельного счета физических задач пакетом MARPLE3D (ИПМ им. М.В.Келдыша РАН) при распределении сеток по ядрам в соответствии с различными разбиениями

Инкрементный алгоритм декомпозиции графов

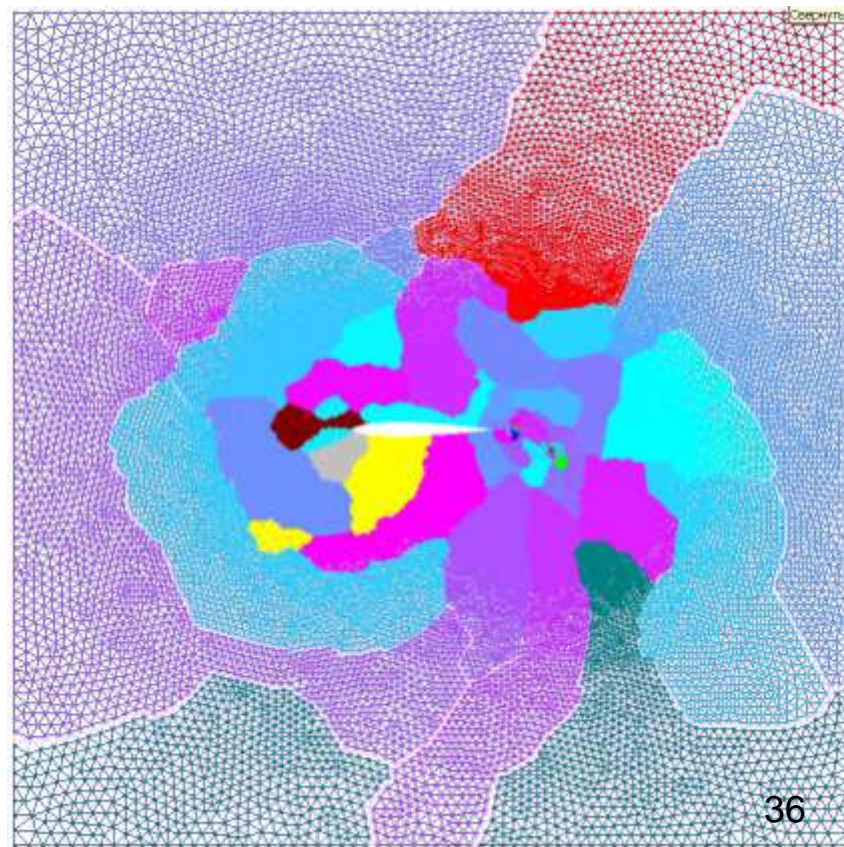
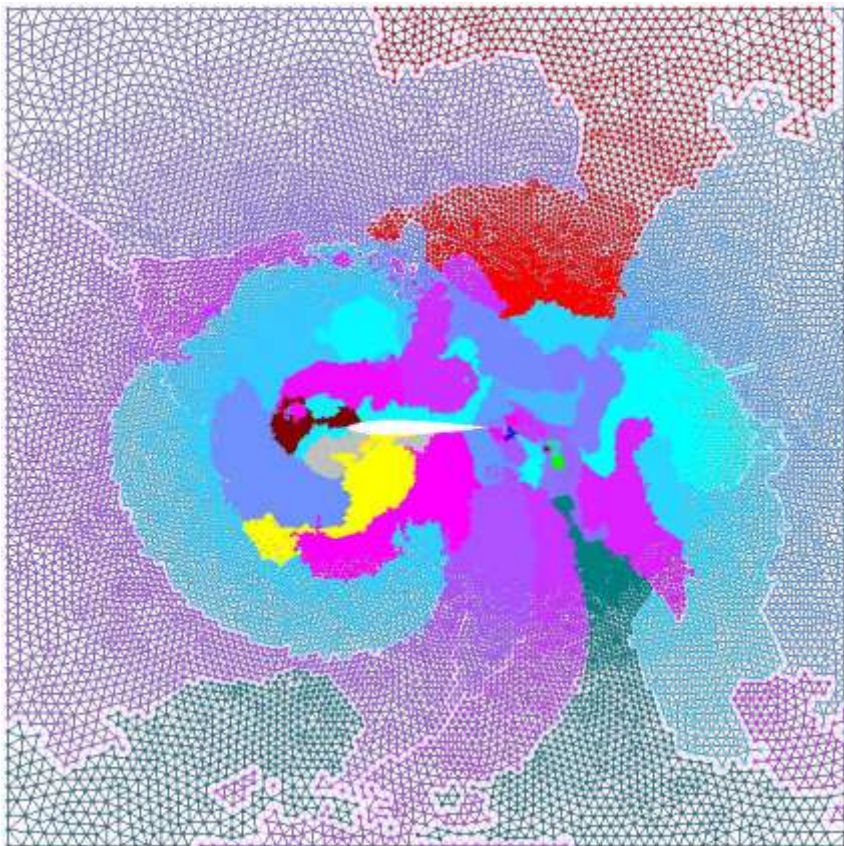
- **инкрементный рост доменов**



сетка вокруг крыла самолета с закрылком

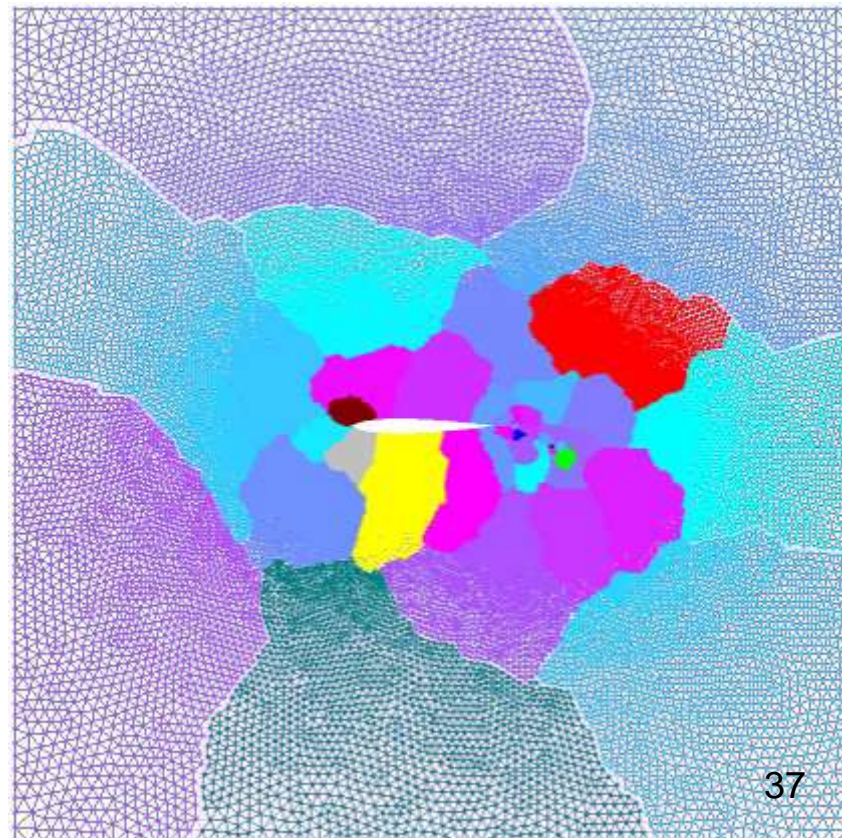
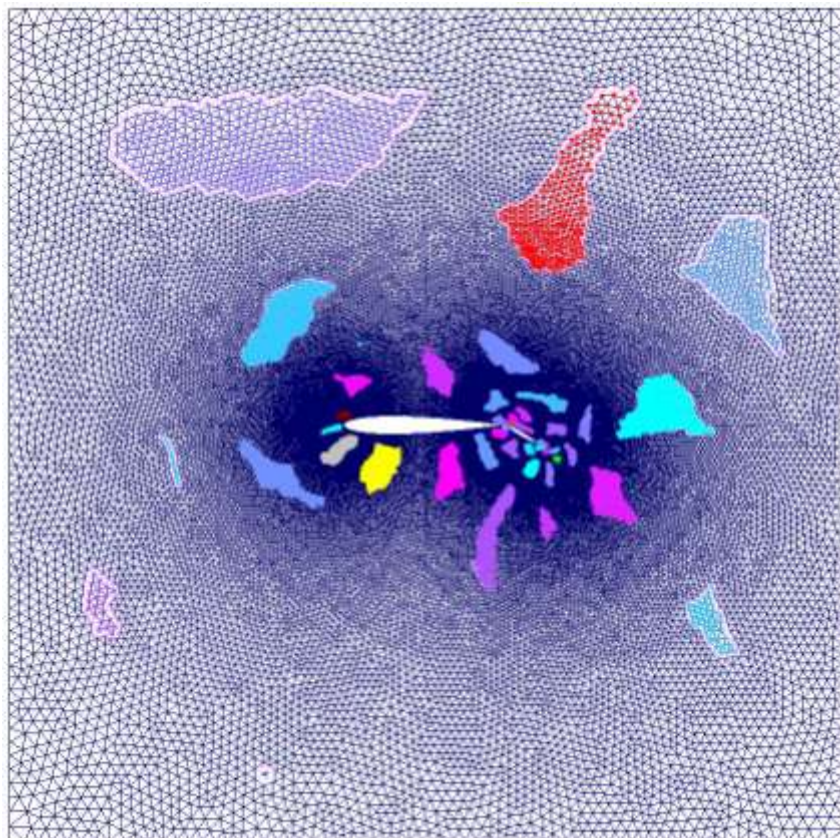
Инкрементный алгоритм декомпозиции

- диффузное перераспределение вершин между доменами
- локальное уточнение доменов



Инкрементный алгоритм декомпозиции

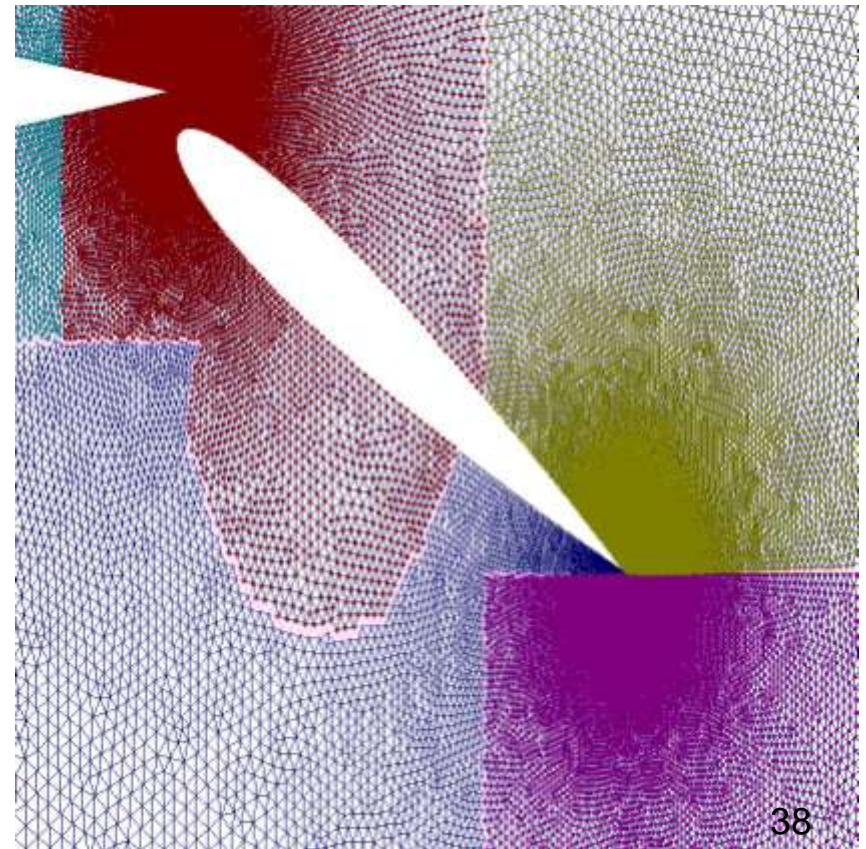
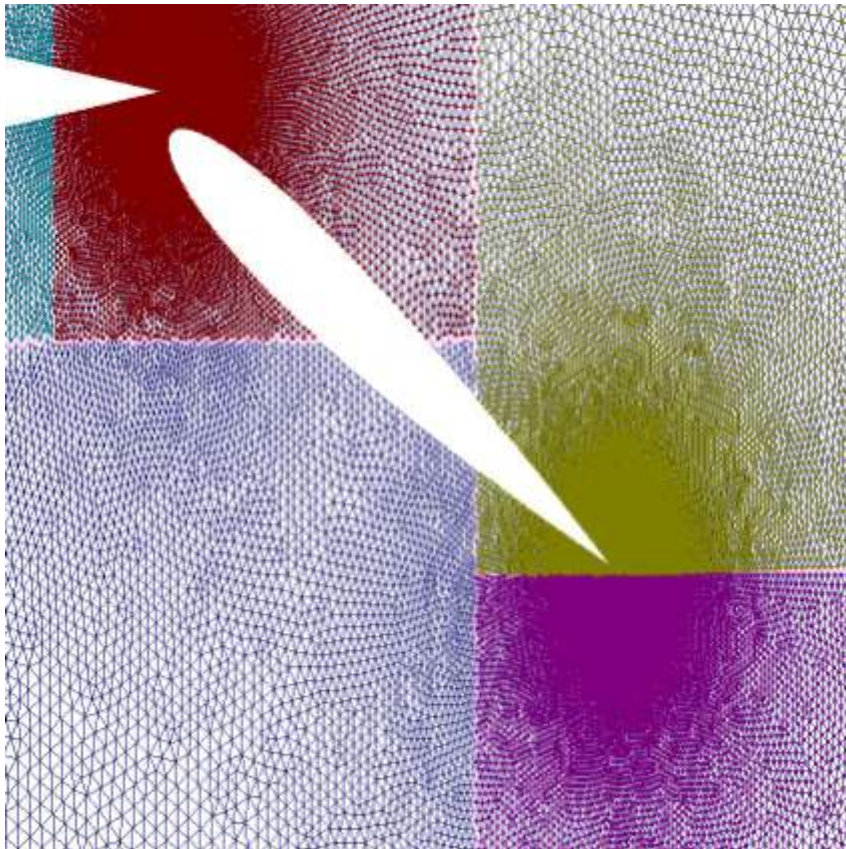
- проверка качества доменов
- освобождение части вершин плохих доменов
- результирующее разбиение



Параллельный инкрементный алгоритм

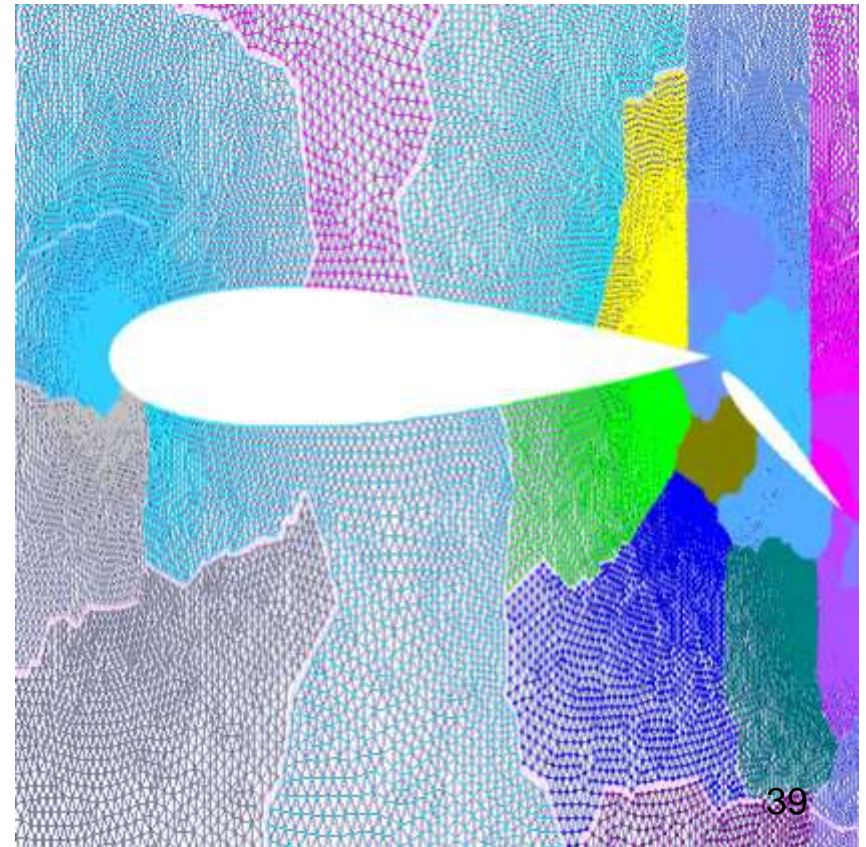
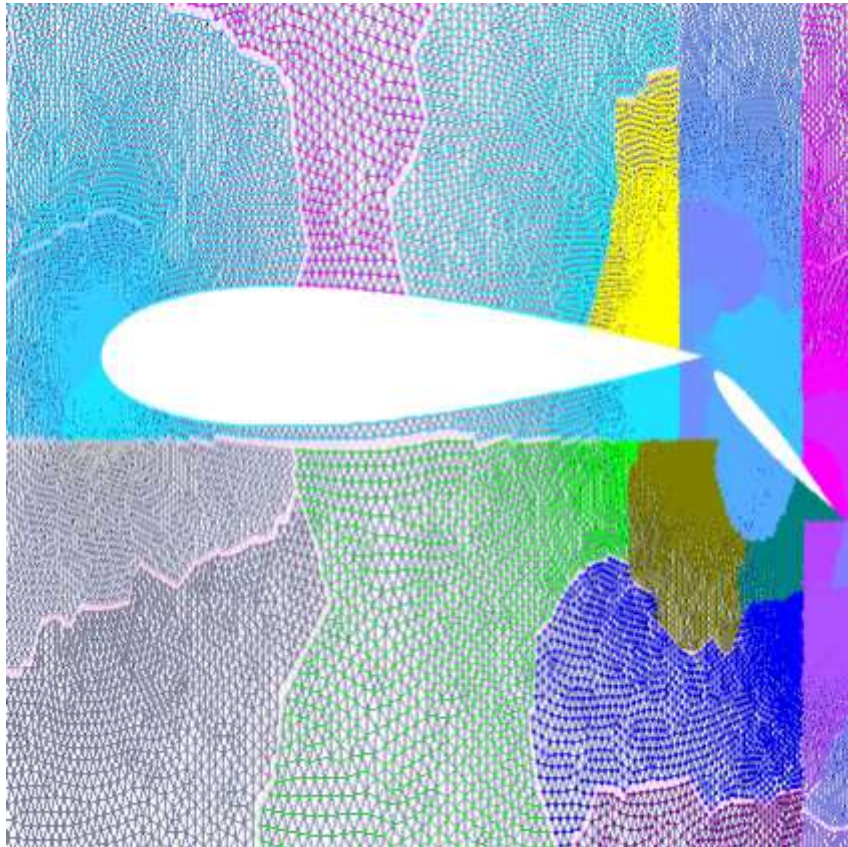
- геометрическое разбиение сетки по процессорам
- перераспределение малых блоков вершин

сетка вокруг крыла самолета с закрылком

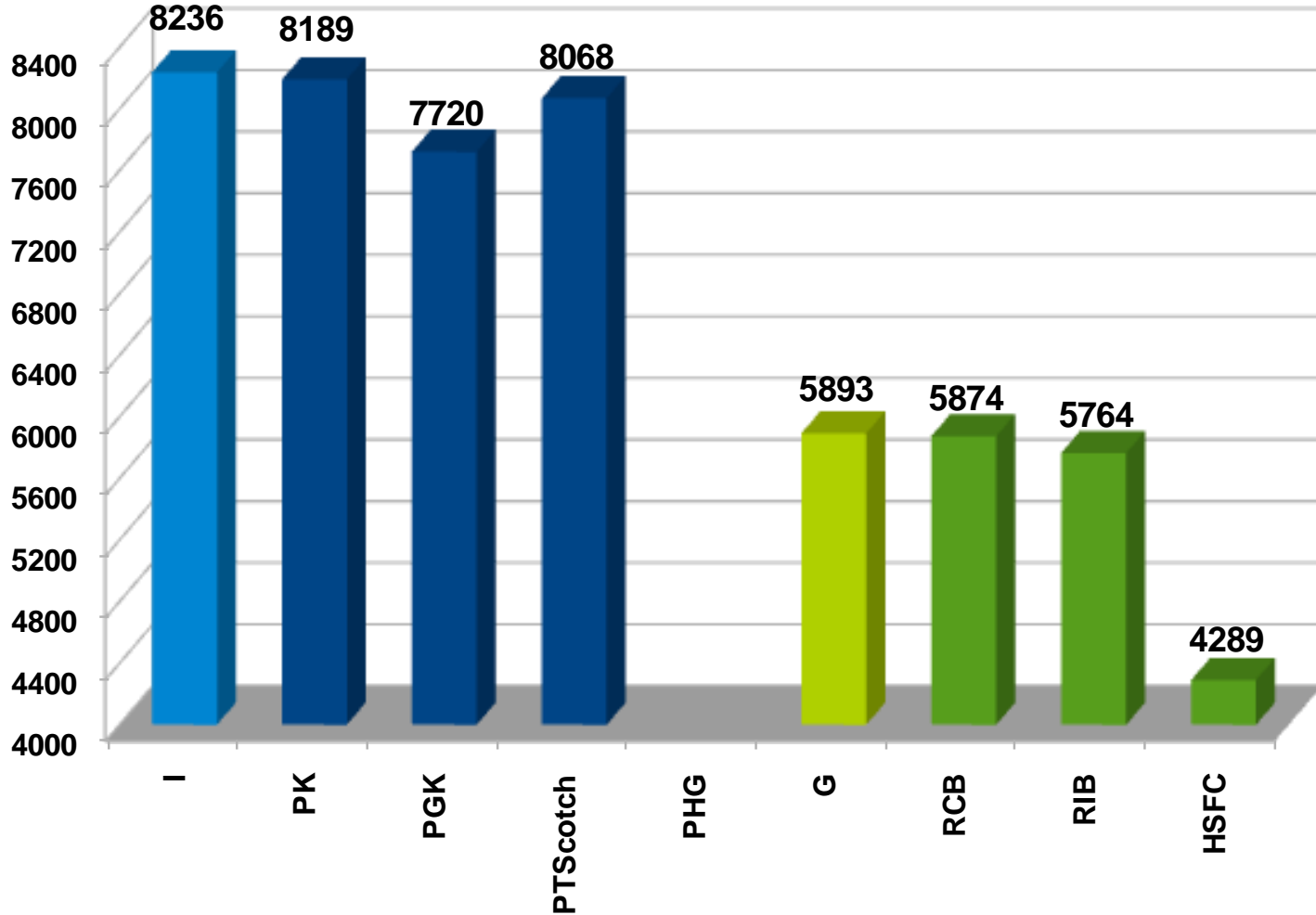


Параллельный инкрементный алгоритм

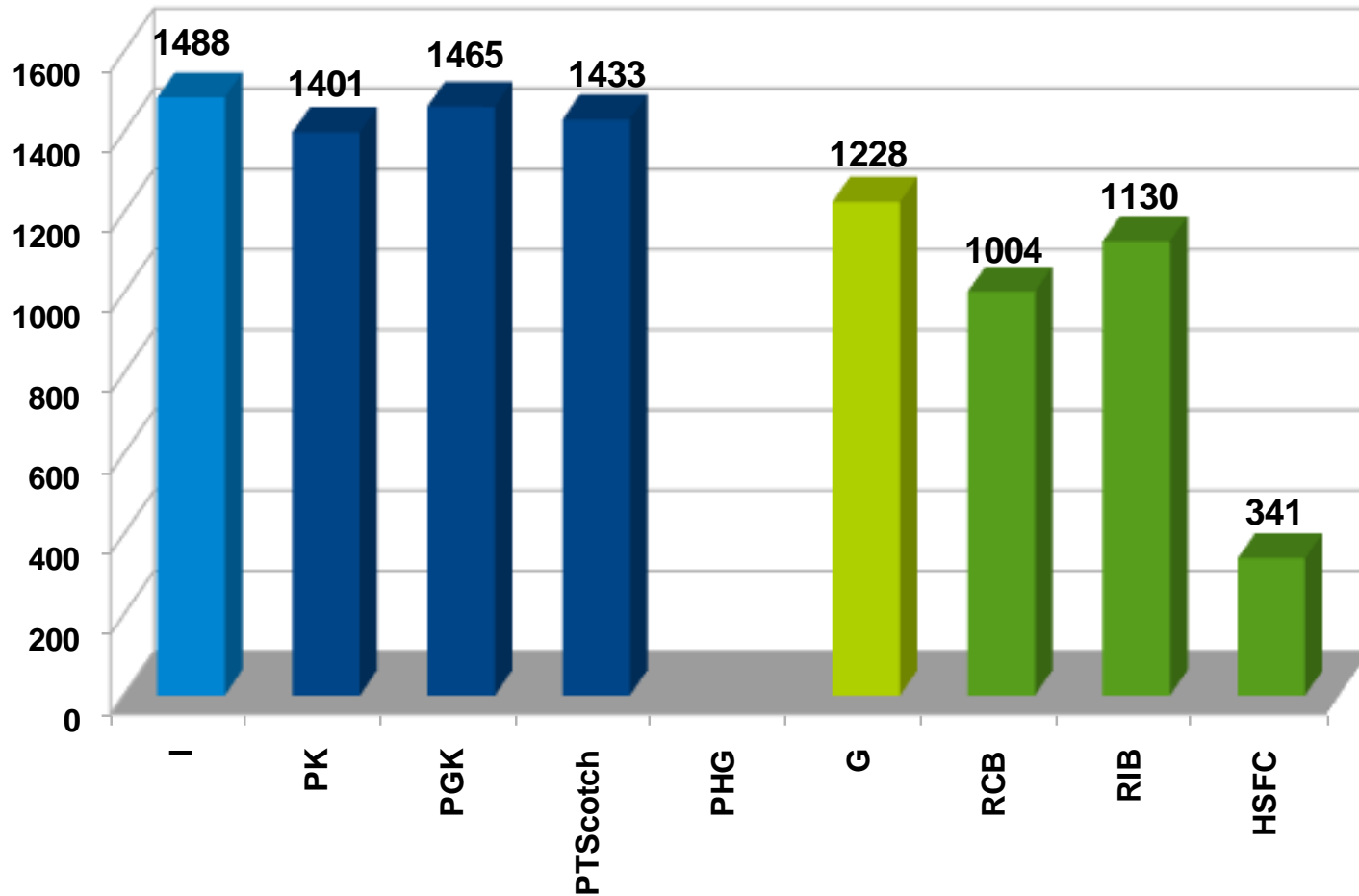
- локальное разбиение
- сбор плохих групп доменов и их повторное разбиение



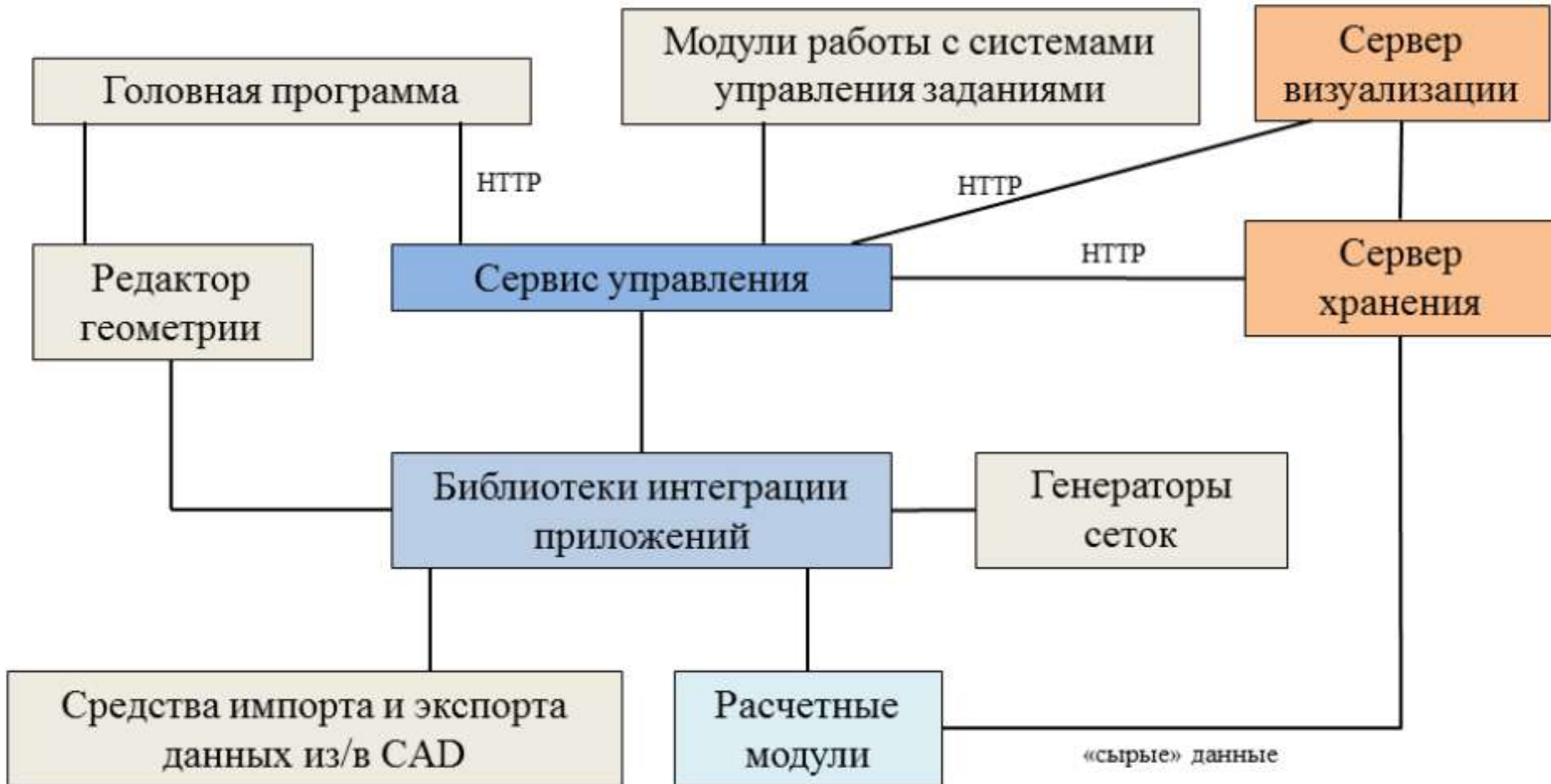
Число шагов по времени (divertor)



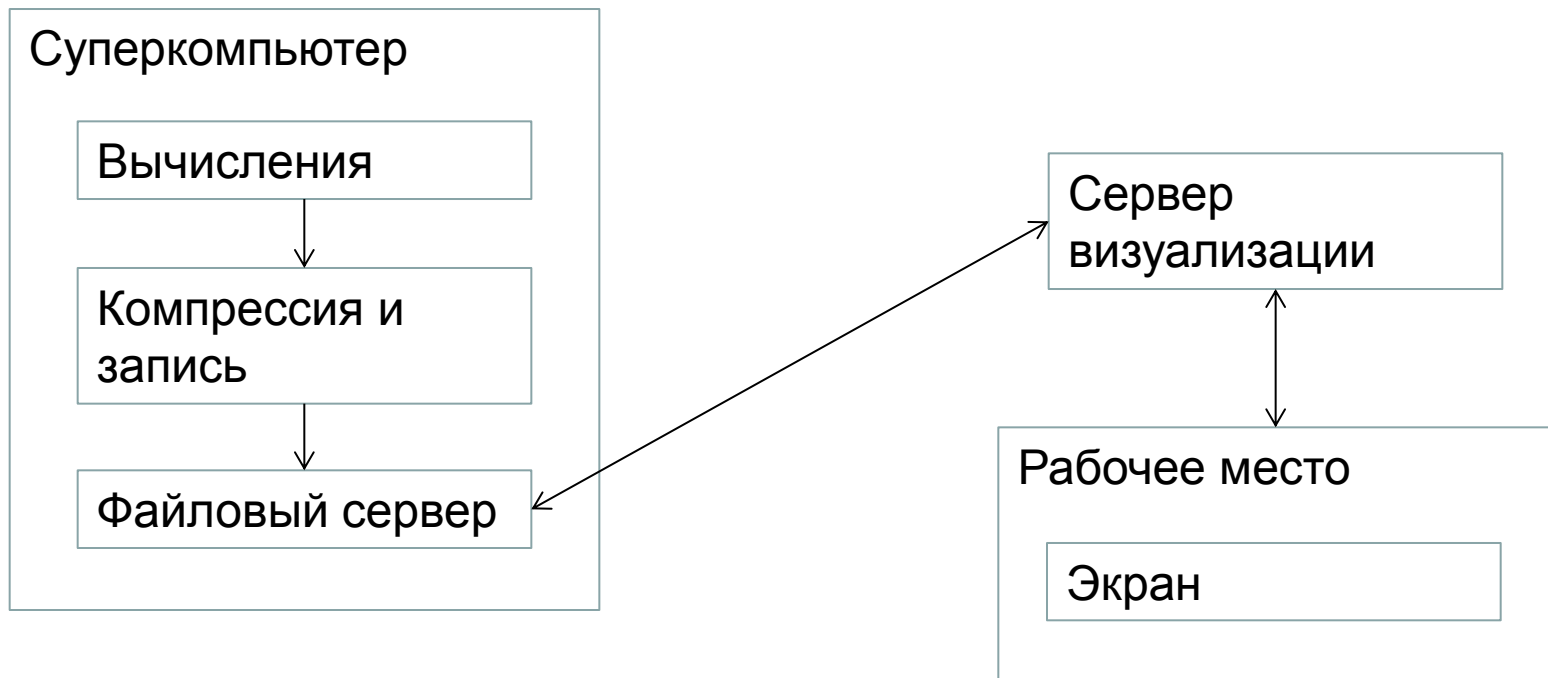
Число шагов по времени (tube)



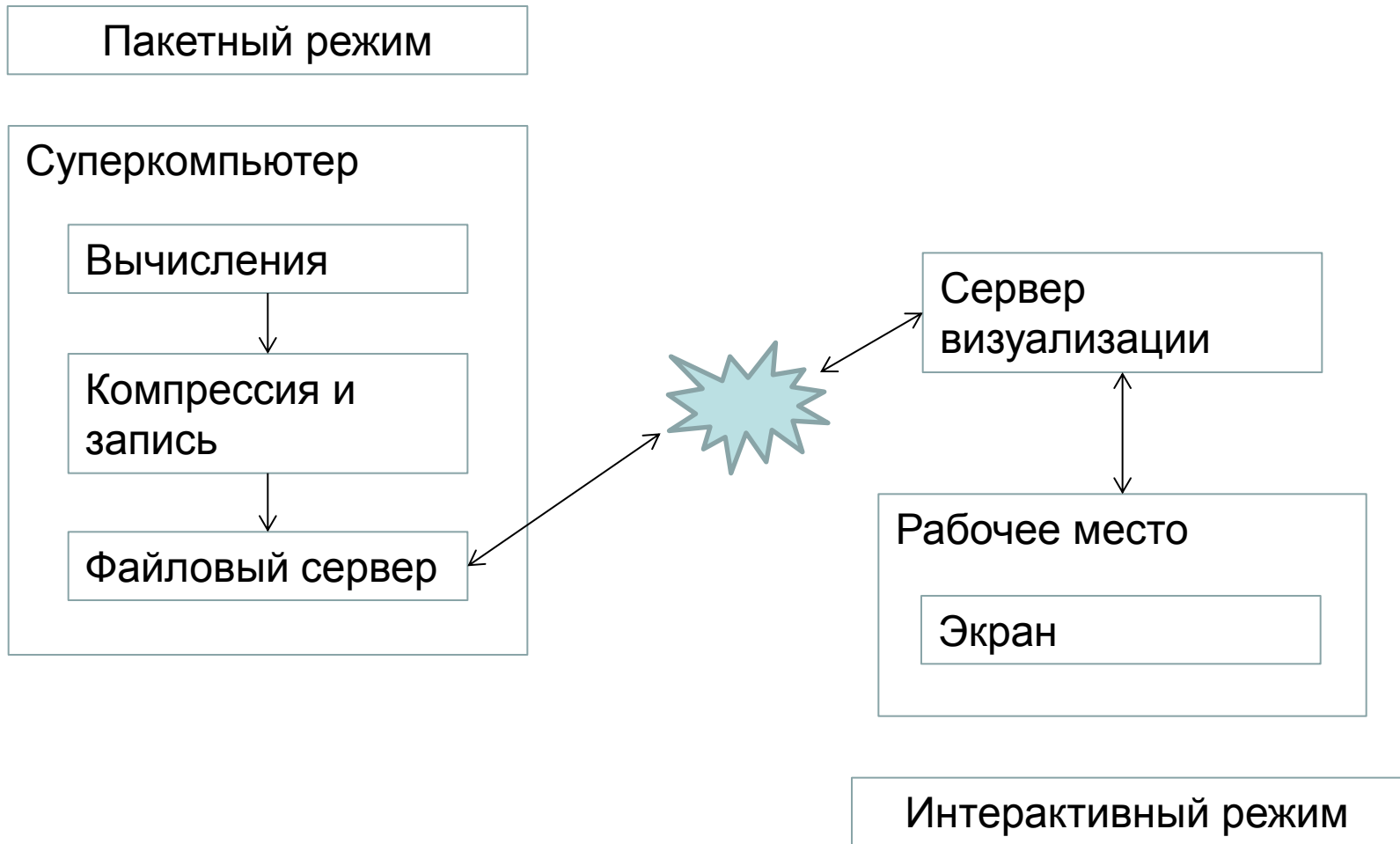
Укрупнённая компонентная модель комплекса



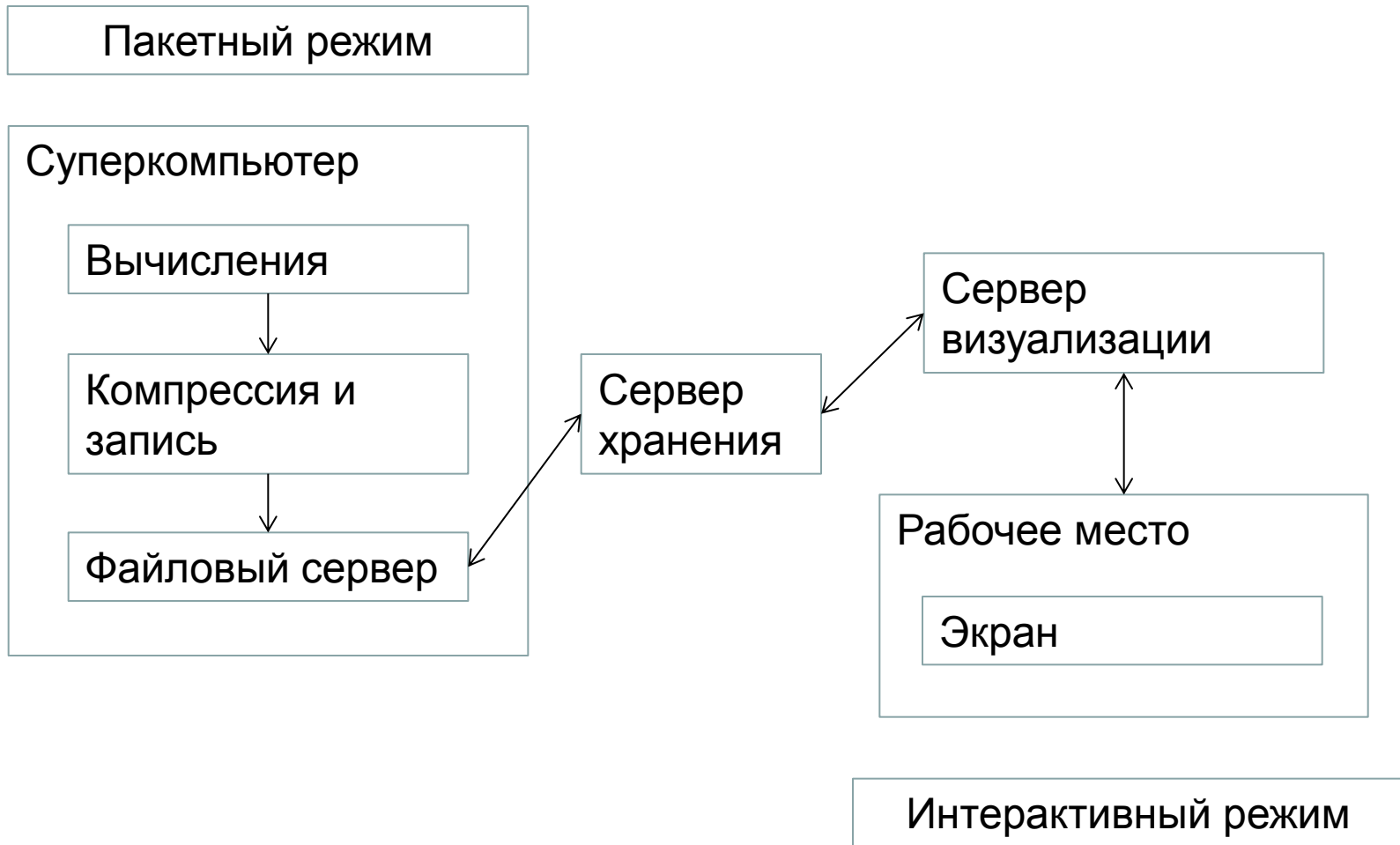
Обработка результатов



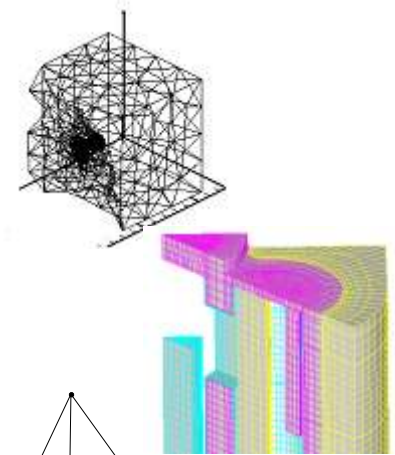
Обработка результатов



Обработка результатов

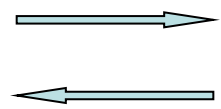
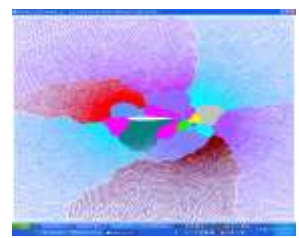
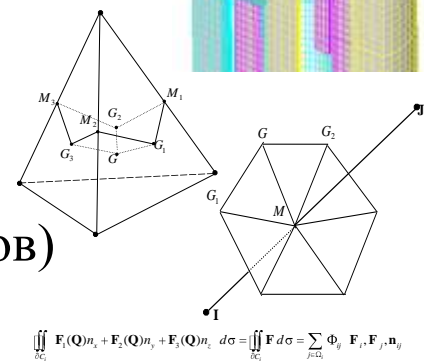


- **Сетки:** нерегулярная (неструктурированные) сетки смешанных типов элементов, блочные и блочно-структурированные сетки
- **Параллельная реализация** для систем с распределенной и общей памятью, гибридных систем
- **Автоматизация** создания параллельных программ для гибридных систем
 - DVM – общая и распределенная память
 - DVMH - DVM + GPU



Основные характеристики:

- Реалистичная 3D геометрия
- Нерегулярные сетки большого объема ($10^8 - 10^9 - \dots$ элементов)
- Параллелизм на всех этапах решения задачи
- Инструменты:
 - компрессия и ввод-вывод больших объемов данных
 - декомпозиция сеток
 - визуализация сеток и сеточных данных
- Интерактивное взаимодействие пользователя с удаленными вычислительными ресурсами



Новое качество, обеспечиваемое комплексом

**Суперкомпьютер – дефицитный
ресурс крайне ограниченного
доступа**

**Гибридный пакет позволяет
выполнить полный
предсказательный расчет**

Финансовая поддержка работ

- Программы РАН
- Госконтракт Минобрауки РФ № 07.524.12.4019 (07.05.2012)
- Проект РФФИ 11-07-93939-ДЖИ8-а
- Проект РФФИ 13-01-12073-офи-м
- Проект РФФИ 11-07-00779-а

Якобовский М.В.

проф., д.ф.-м.н.,

зав. сектором

«Программного обеспечения
многопроцессорных систем и
вычислительных сетей»

Института прикладной математики им.
М.В.Келдыша Российской академии наук

[mail: lira@imamod.ru](mailto:lira@imamod.ru)

[web: http://lira.imamod.ru](http://lira.imamod.ru)