Вычислительный эксперимент на суперкомпьютерах

Якобовский М.В.

Институт прикладной математики
им. М.В.Келдыша РАН

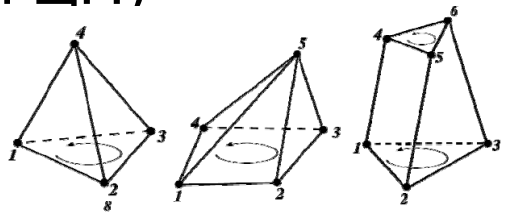
ВМК МГУ им. М.В.Ломоносова

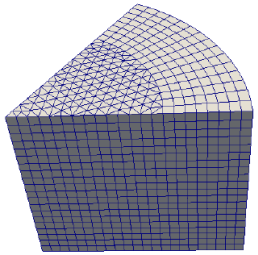
lira@imamod.ru

www://lira.imamod.ru

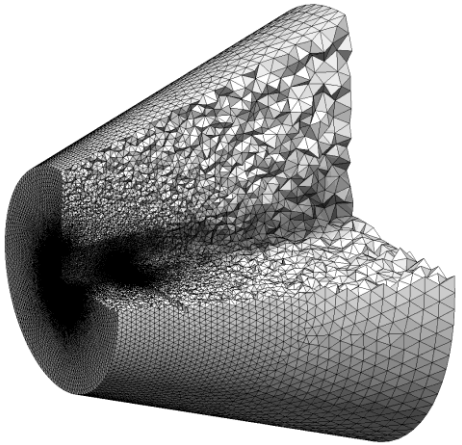
Методы, пакеты, инструменты

- GIMM (ПРАН)
- GIMM NANO (ФЦП)

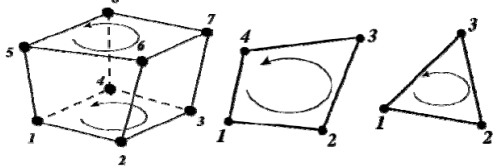




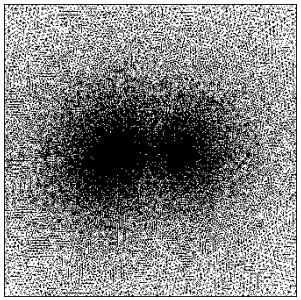
- NOISEtte

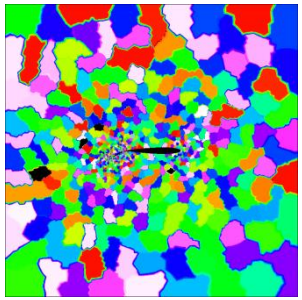


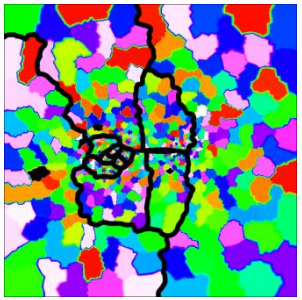
- MARPLE3D



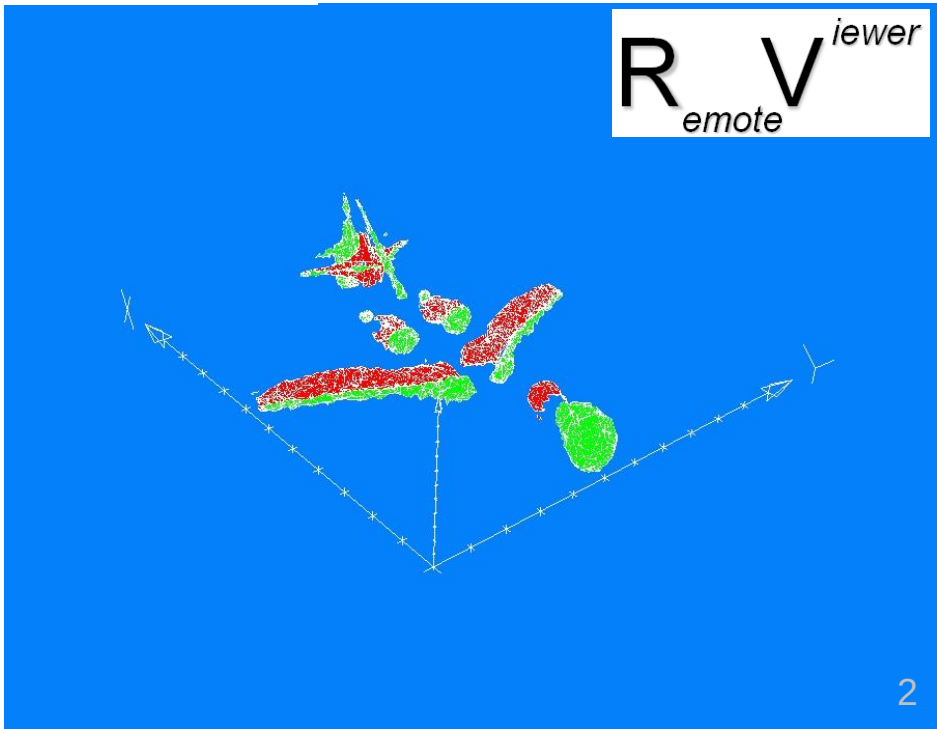
- Декомпозиция сеток







- Автоматизация разработки параллельных программ DVM, DVMH

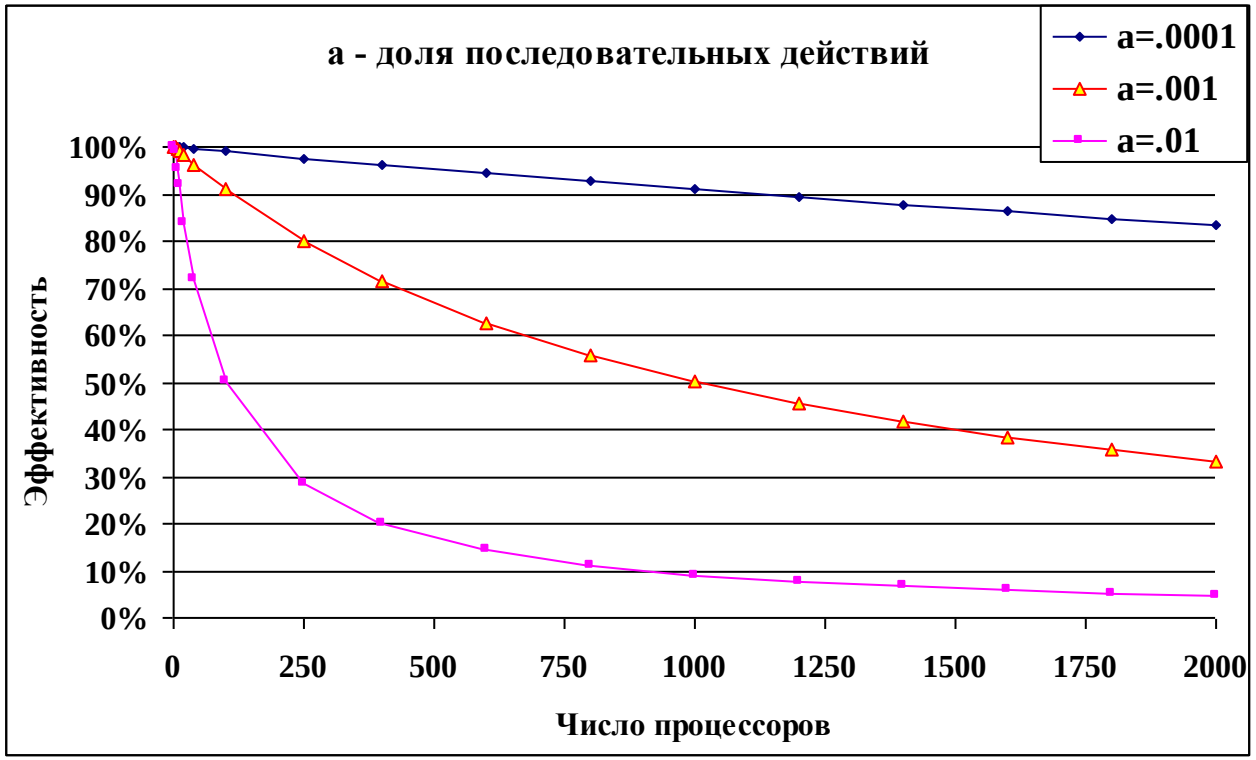


Ограничения

- Закон Амдаля

$$S(p) = \frac{1}{a + \frac{1-a}{p}}$$

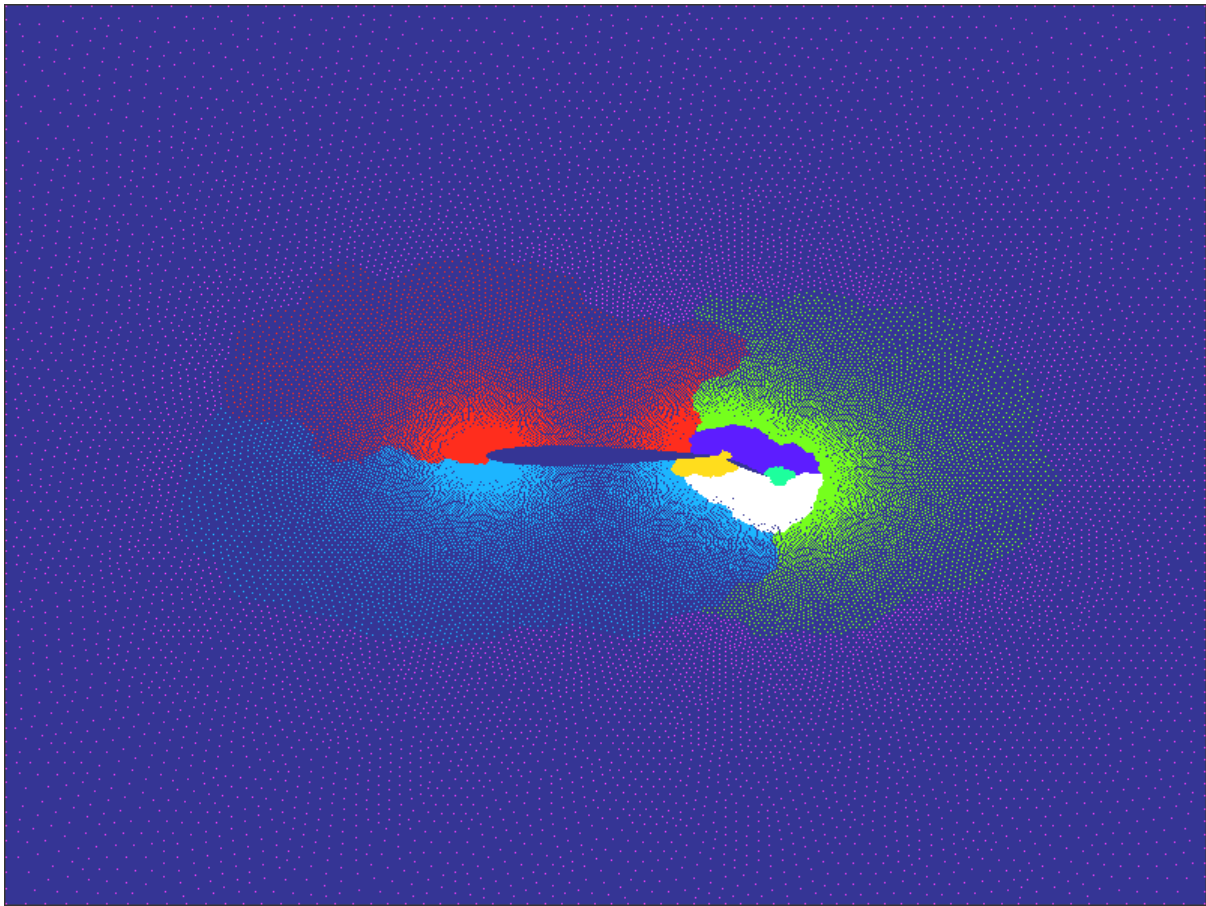
$$E(p) = \frac{1}{1 + a(p-1)}$$

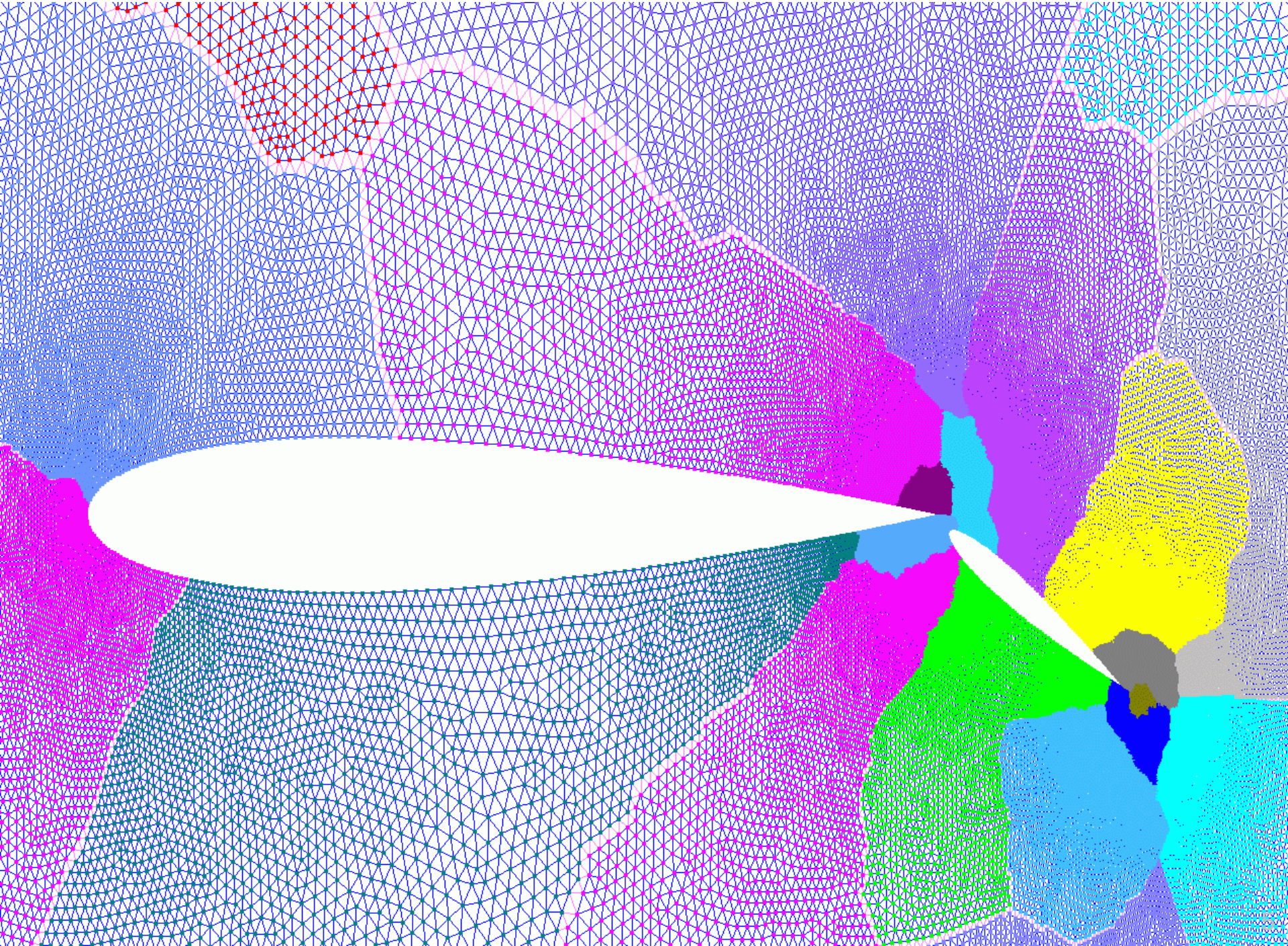


- Пакетный режим исполнения и отладки приложений
- Процедуры авторизованного доступа к удаленным системам
- Высокая динамика изменения конфигурации суперкомпьютеров
- Несоизмеримость ресурсов рабочей станции пользователя и суперкомпьютера

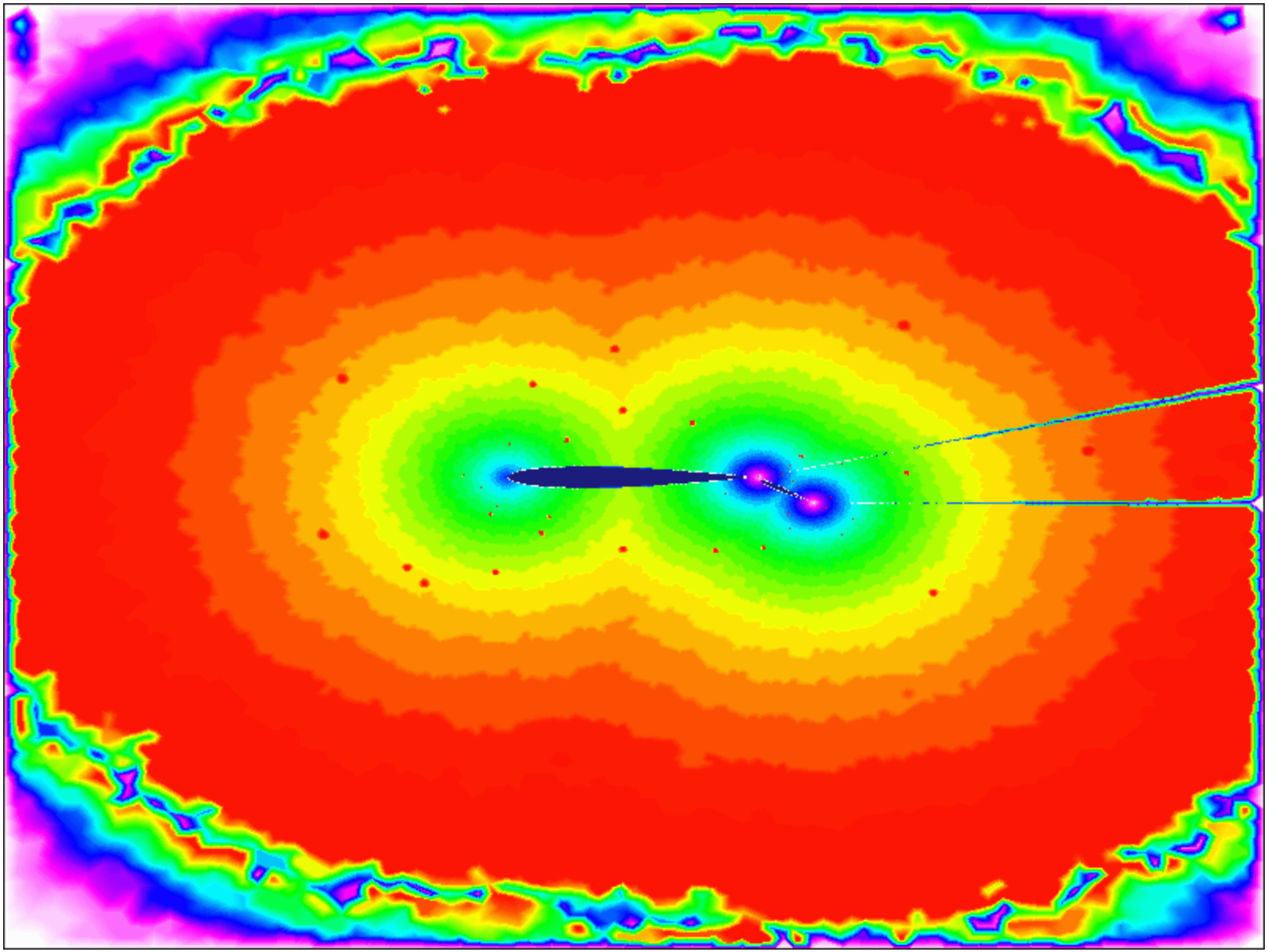
Статическая балансировка загрузки

- Критерии декомпозиции
- Инкрементный алгоритм декомпозиции
- Иерархическая обработка больших сеток

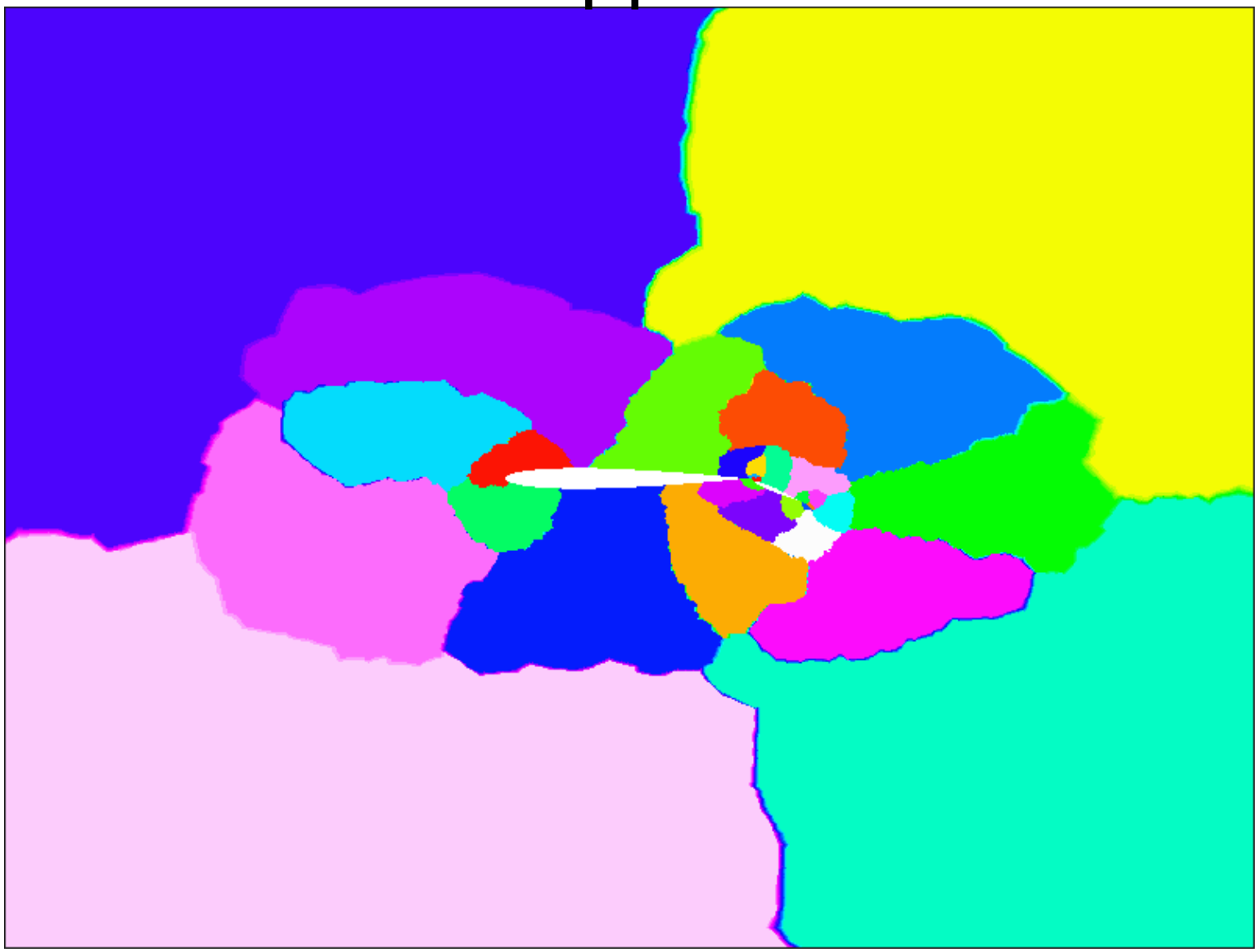




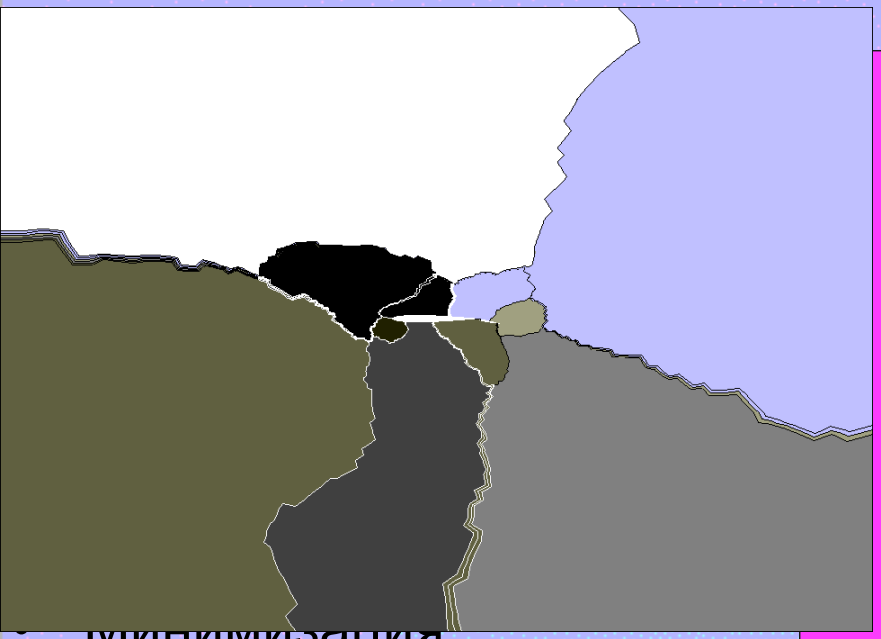
Простое разбиение на 32 домена

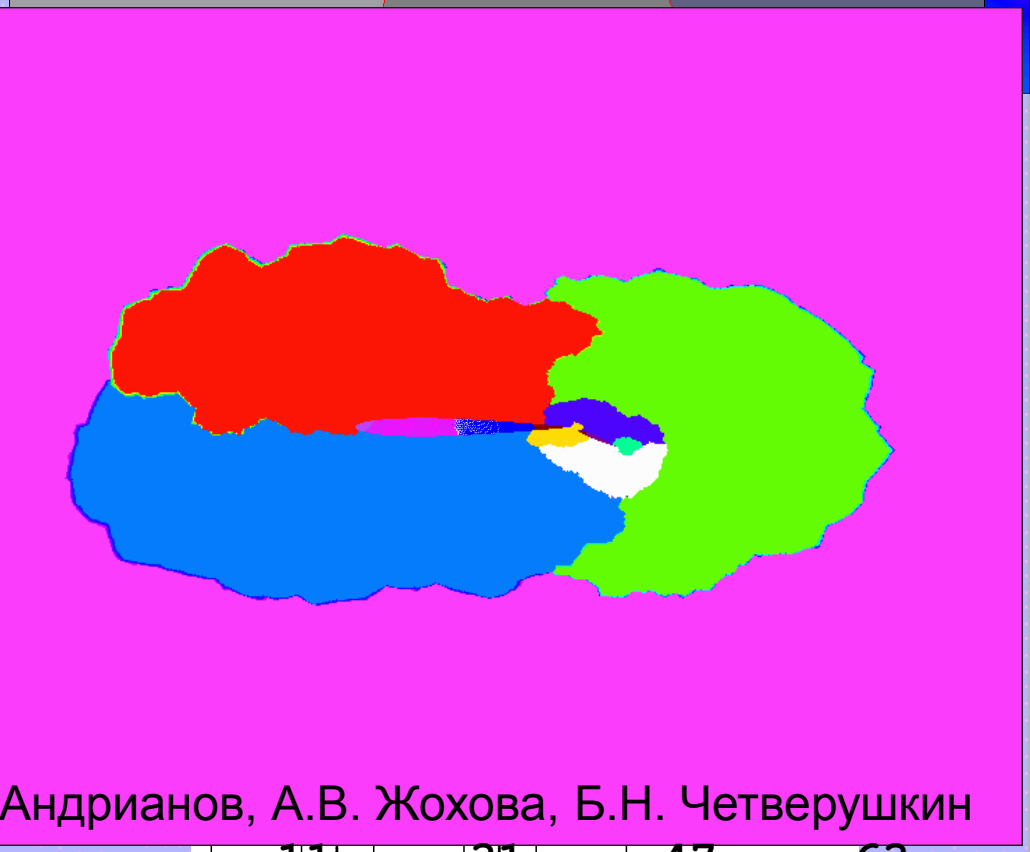


Рациональное разбиение на 32 домена



Критерии декомпозиции графов



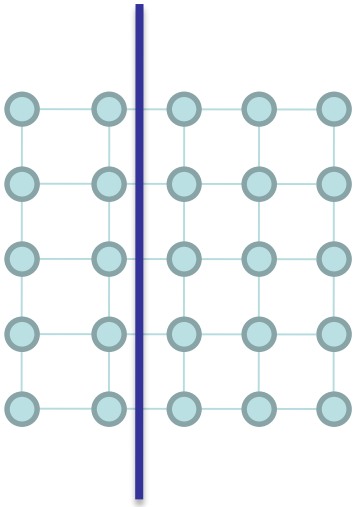


- Минимизация максимальной степени доменов
- Обеспечение связности доменов
- Обеспечение связности множества внутренних узлов доменов

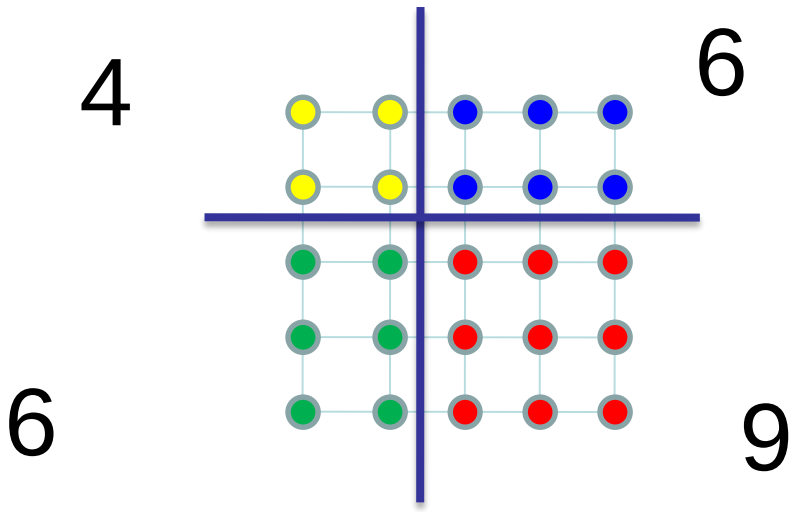
А.Н. Андрианов, А.В. Жохова, Б.Н. Четверушкин

Процессоров	11	31	47	63
New_sort	13.59	5.59	4.38	4.16
METIS	13.61	11.00	11.10	10.56

Декомпозиция сетки из 25 узлов на 2 части



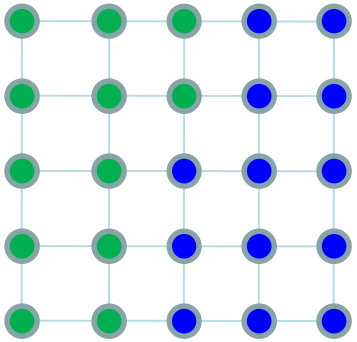
Декомпозиция решетки 5 x 5 на 4 части



Дисбаланс 9/4=2.25

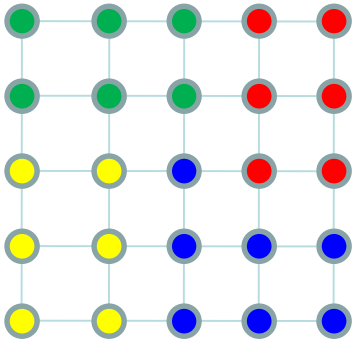
25/4 = 4 ? 6 ? 9

- Декомпозиция решетки 5 x 5 на 2 домена
- Дисбаланс 13/12 : 8%



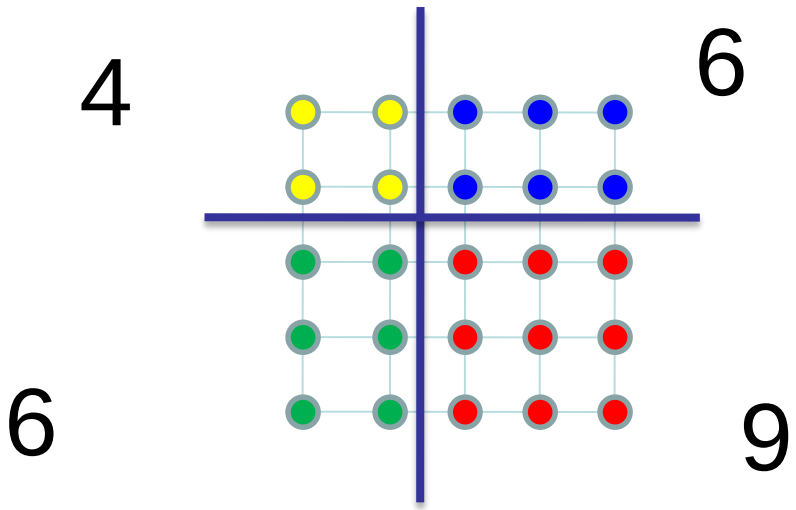
25/4 = 4 ? 6 ? 9

- Декомпозиция решетки 5 x 5 на 4 домена
- Дисбаланс 7/6 : 17%



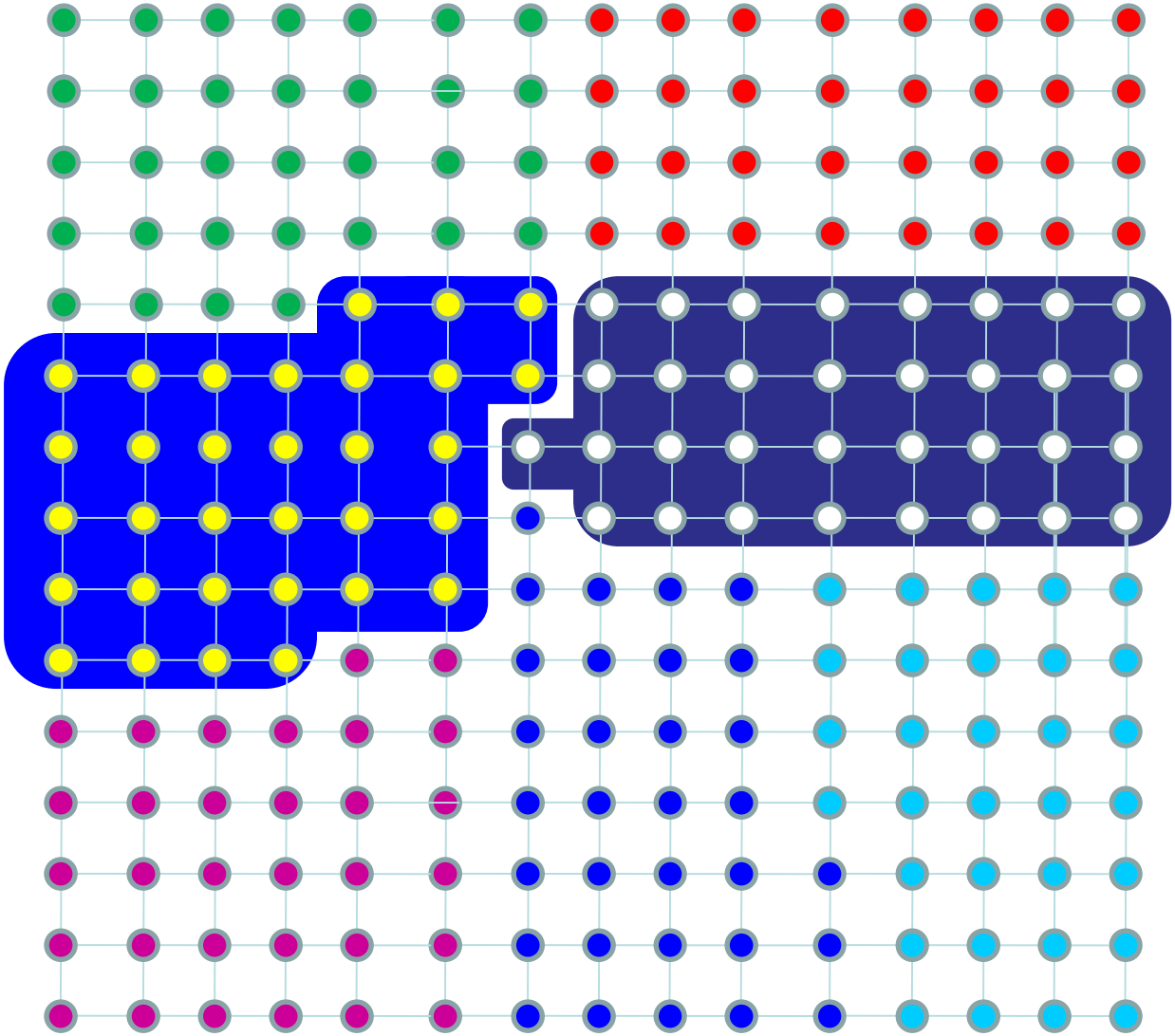
25/4 = 4 ? 6 ? 9

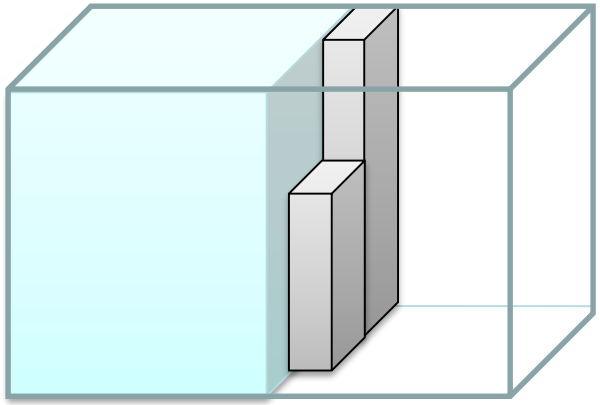
- Декомпозиция решетки 5 x 5 на 4 домена

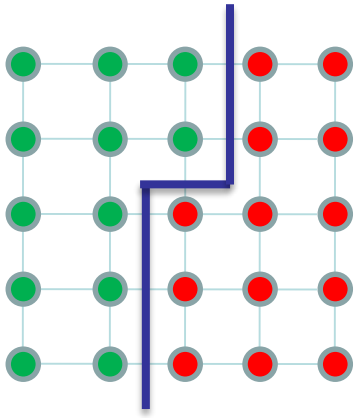


- Дисбаланс 9/4=2.25

Декомпозиция сетки 25x25 на 7 частей







Разрезано ребер

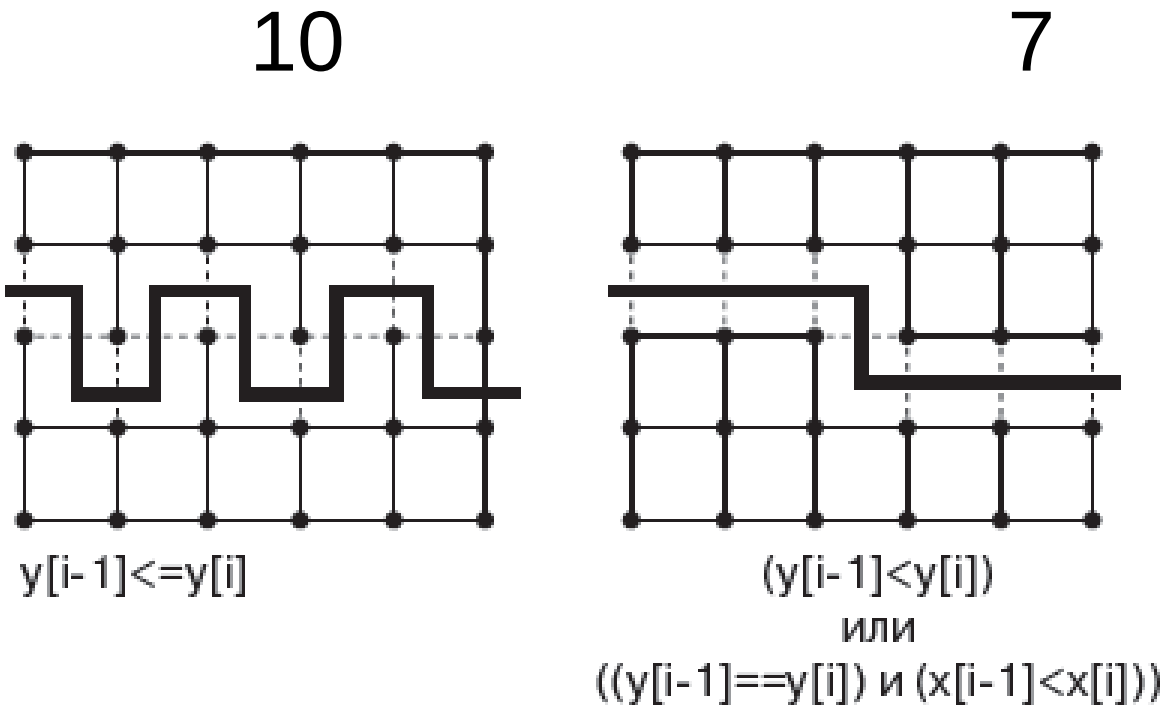
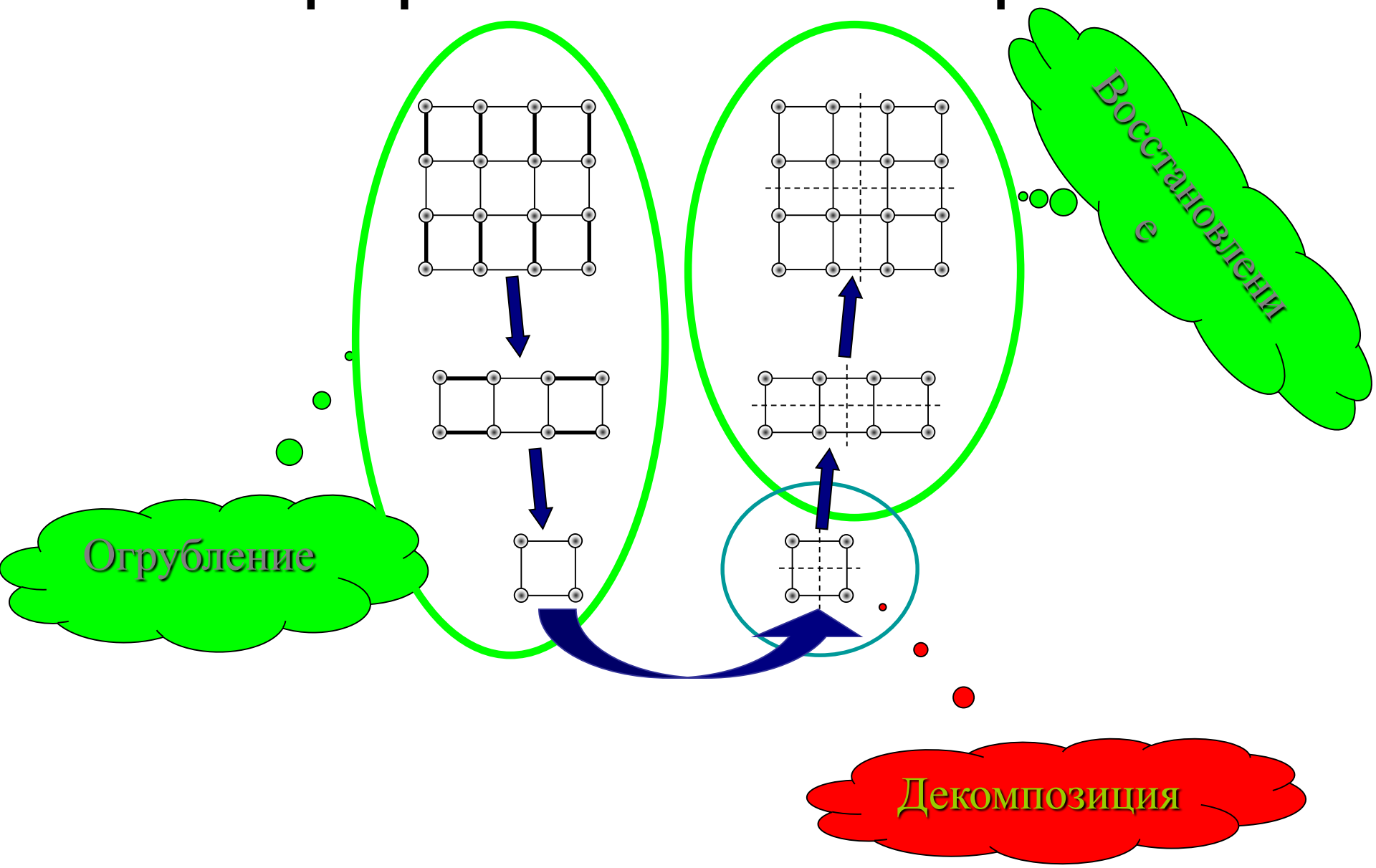


Рис. 76. Координатная бисекция

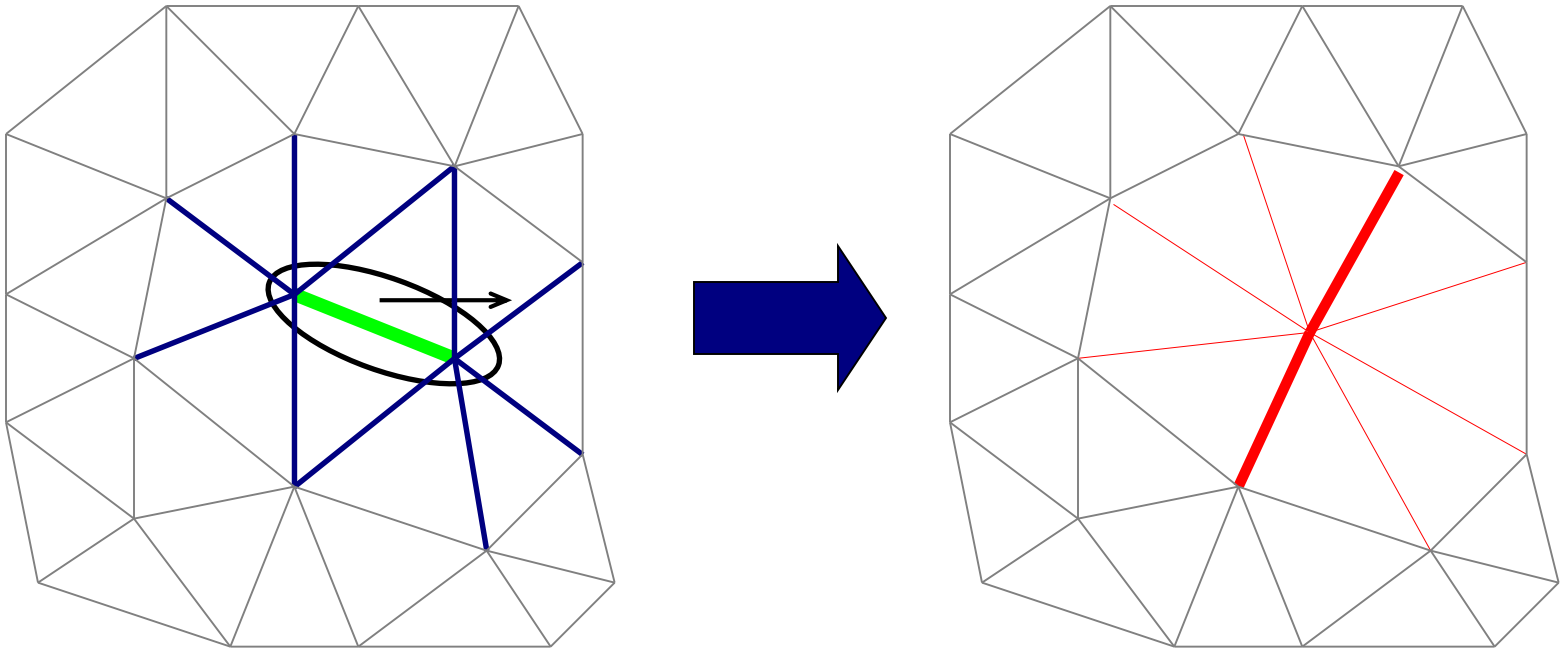
Пакеты декомпозиции графов

Chaco	Bruce Hendrickson Robert Leland
ParMETIS	George Karypis Vipin Kumar
PARTY	Robert Prais, et al.
JOSTLE	Chris Walshaw, et al.
SCOTCH	Francois Pellegrini

Иерархический алгоритм



Огрубление графа



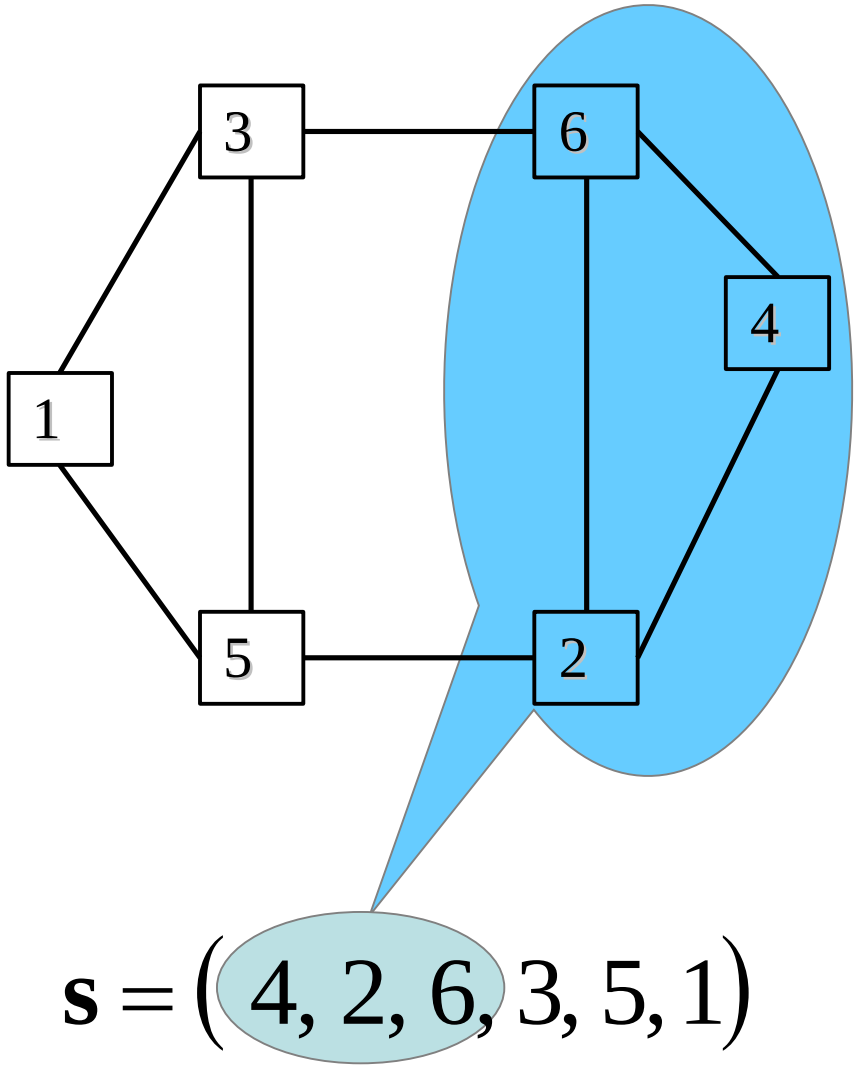
Спектральный метод

$$L = \begin{pmatrix} 2 & 0 & -1 & 0 & -1 & 0 \\ 0 & 3 & 0 & -1 & -1 & -1 \\ -1 & 0 & 3 & 0 & -1 & -1 \\ 0 & -1 & 0 & 2 & 0 & -1 \\ -1 & -1 & -1 & 0 & 3 & 0 \\ 0 & -1 & -1 & -1 & 0 & 3 \end{pmatrix}$$

$\lambda = \{0,1,3,3,4,5\}$ $\quad Lq = \lambda q$

$\lambda_2 = 1$ - Минимальное ненулевое

$q_2 = (2, -1, 1, -2, 1, -1)$



$\mathbf{s} = (4, 2, 6, 3, 5, 1)$

Спектральная бисекция

Пометим вершины метками 1 и -1

q[i]={-1,1}

при разбиении вершин на два равномощных множества

$$\sum_{i=1}^{|V|} q_i = 0$$

число разрезанных ребер

$$|E_c| = \frac{1}{4} \sum_{i,j:e_{i,j} \in E} \left(q_i - q_j\right)^2$$

Спектральный метод

Сумма квадратов меток $\sum_{i=1}^{|V|} q_i^2 = q^T q = |V|$

$$l_{ij} = \begin{cases} -1, & e_{ij} \in E,\ i \neq j, \\ \sum_{k \neq i} l_{ik}, & i = j, \\ 0, & \text{иначе} \end{cases}$$

$Lq = \lambda q$ для собственных значений и векторов число разрезанных ребер будет равно

$$|E_c| = \frac{1}{4} q^T L q = \frac{1}{4} q^T \lambda\ q = \frac{1}{4} \lambda |V|$$

Разбиение вершин на два множества

Для минимизации $|E_c| = \frac{1}{4}\lambda|V|$ следует найти минимальное собственное число и соответствующий ему собственный вектор – вектор Фидлера

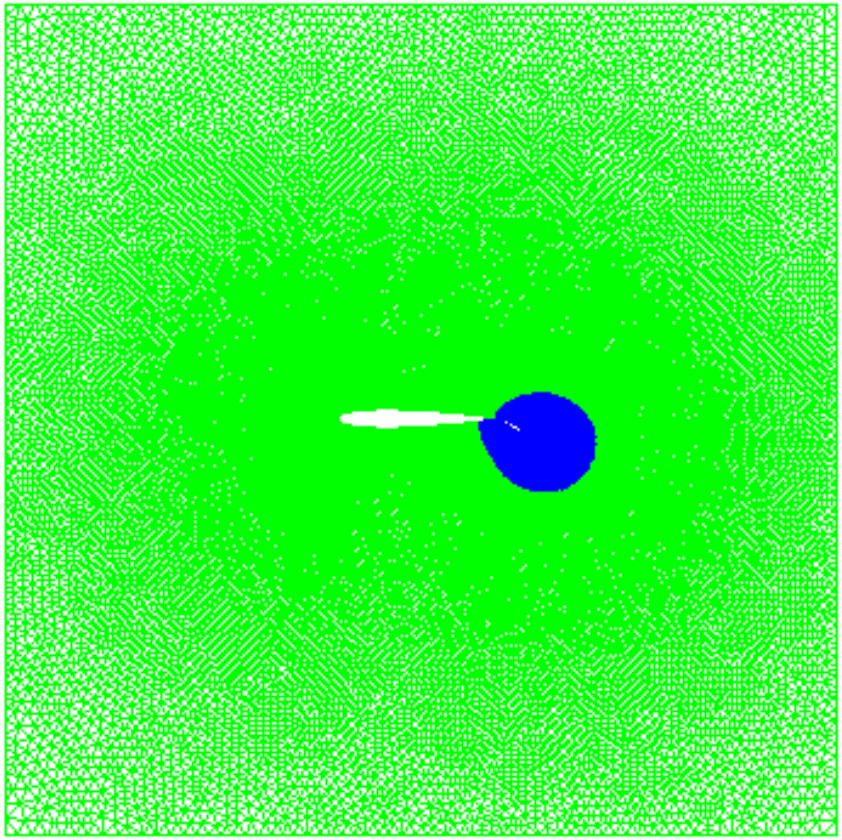
Он ортогонален вектору соответствующему нулевому $\lambda_0 = 0$ - единичному вектору

$$L\psi = \lambda_1\psi \qquad e\psi = 0$$

Следовательно $\sum_{i=1}^{|V|} q_i = 0$ - множества {-1} и {1} содержат одинаковое число вершин

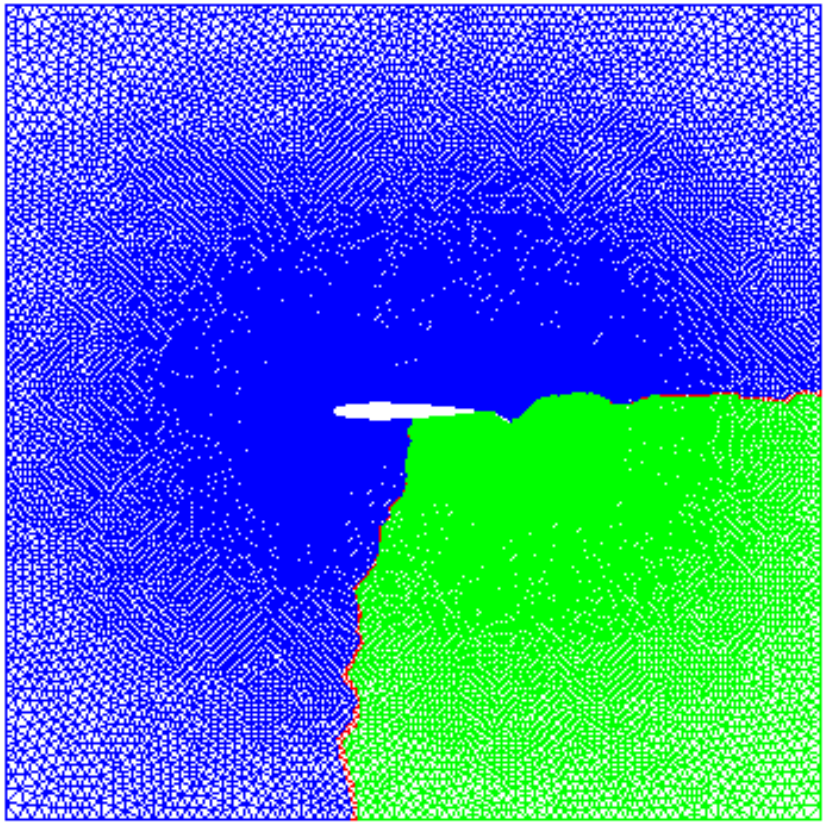
Метод спектральной бисекции

Spectral Partition



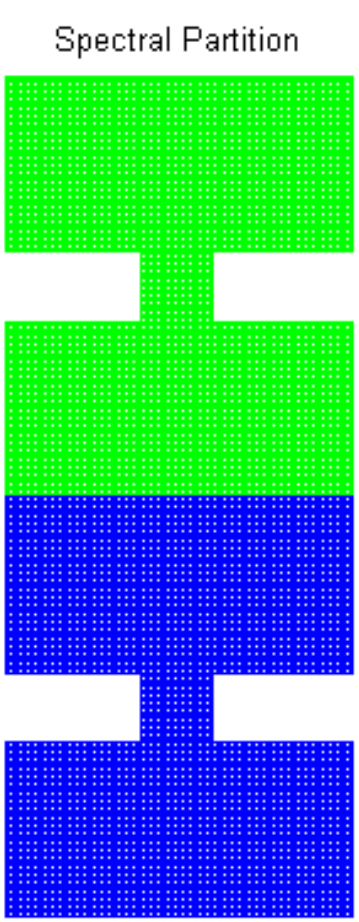
655 cut edges

Metis Partition

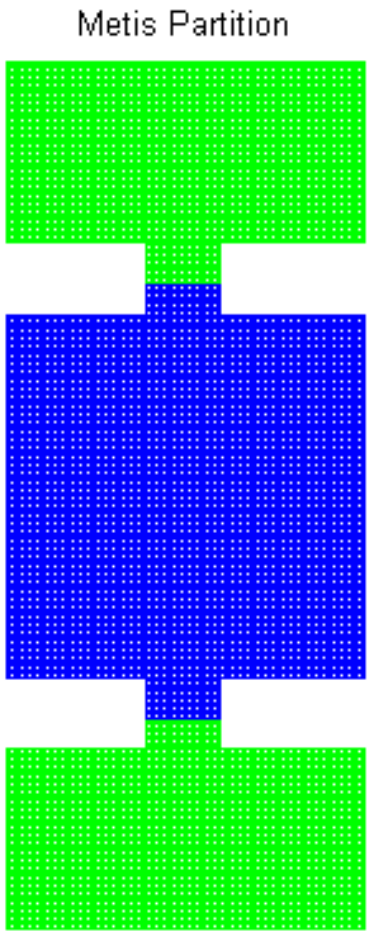


524 cut edges

Метод спектральной бисекции

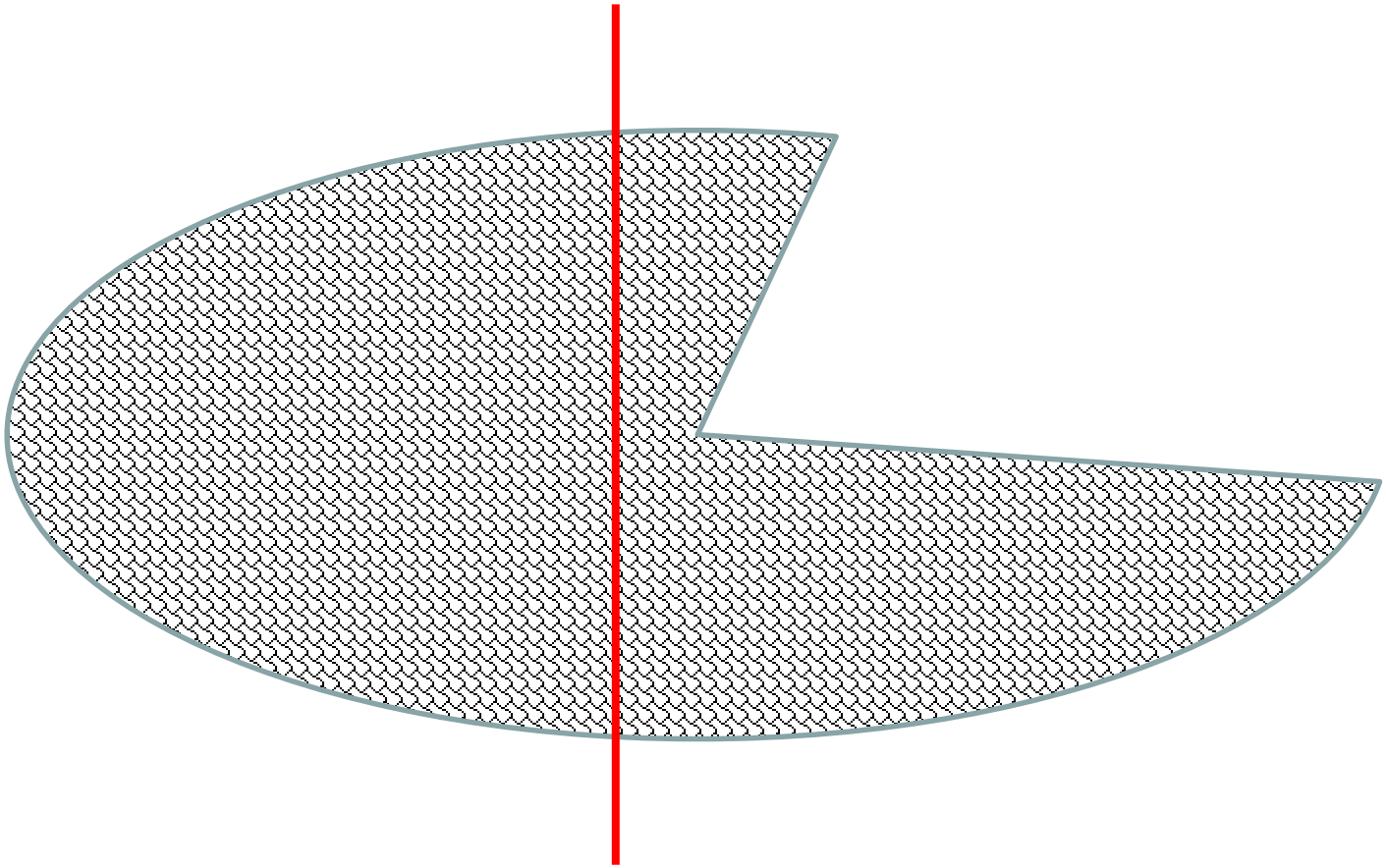


100 cut edges

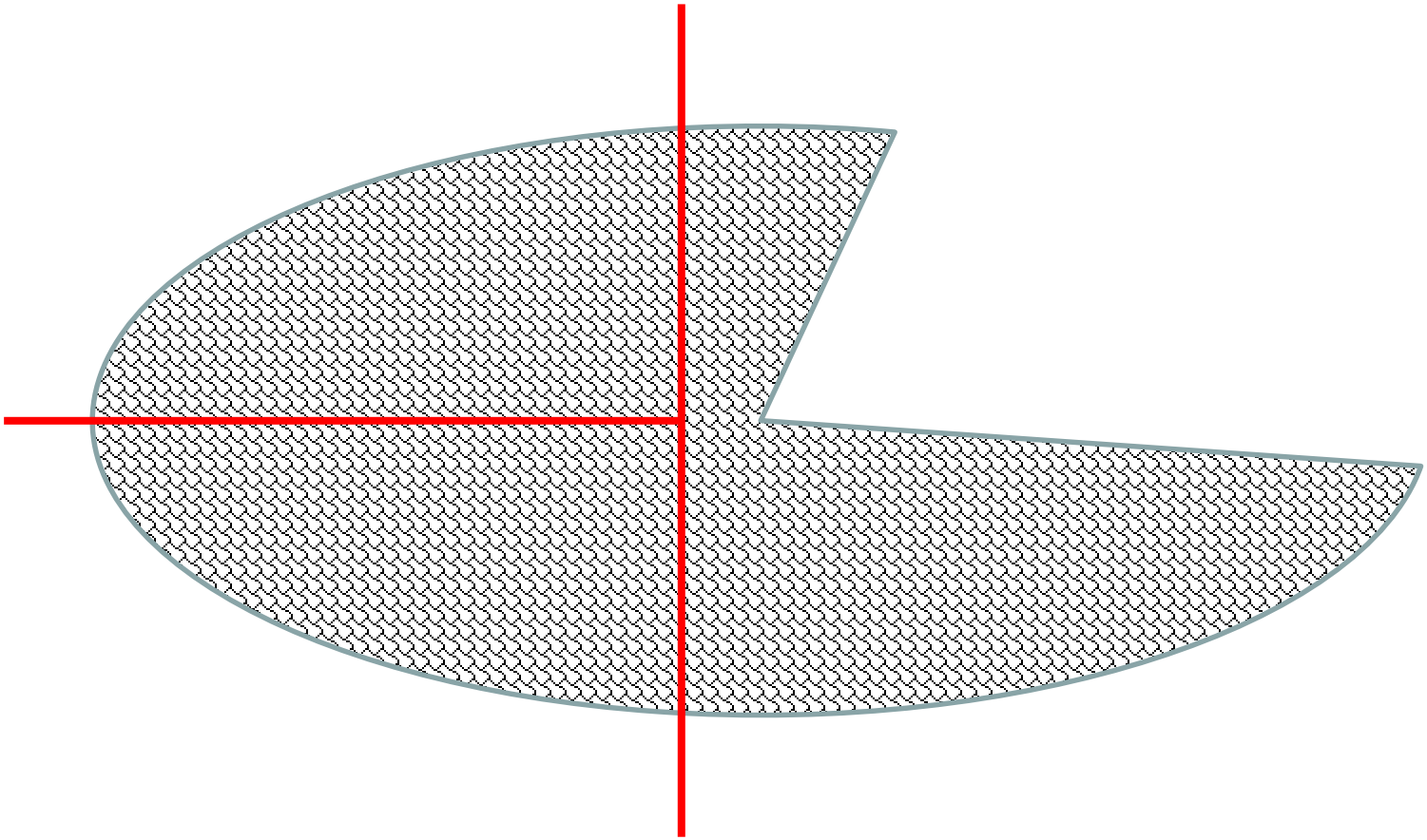


42 cut edges

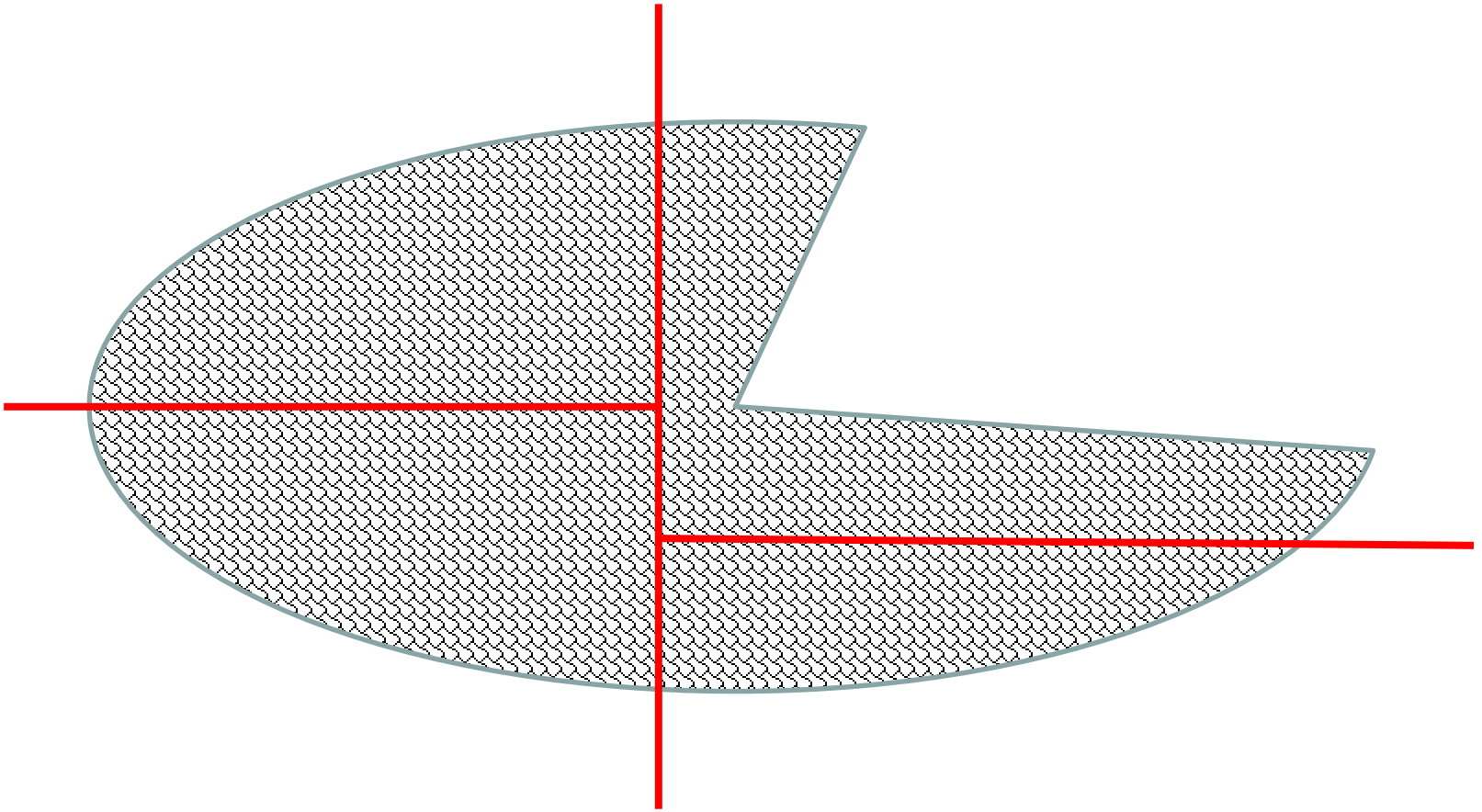
Как разрезать граф на 4 части?



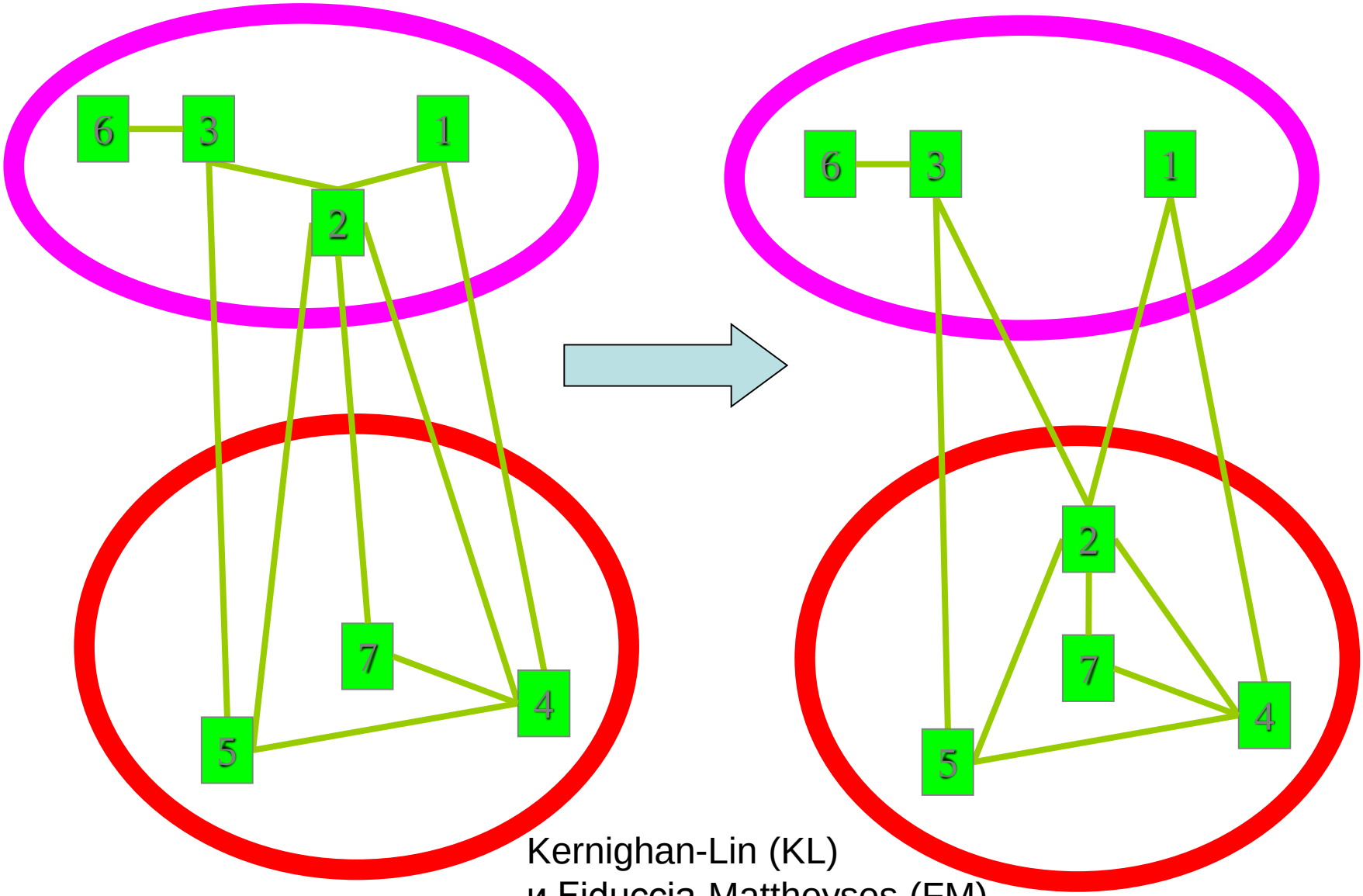
Как разрезать граф на 4 части?



Как разрезать граф на 4 части?



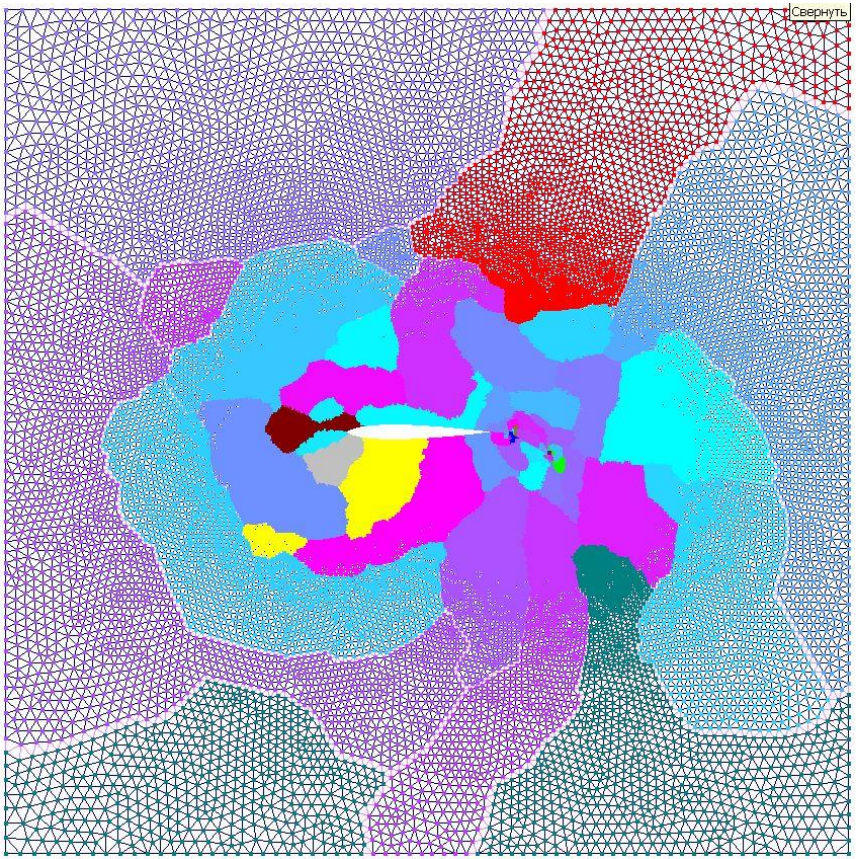
Локальное уточнение



Kernighan-Lin (KL)
и Fiduccia-Mattheyses (FM)

Связность важна:

- **алгоритмы решения систем линейных уравнений**
- **компрессия сеточных данных**
- **алгоритм композиции подобластей[1]**
- **распараллеливание[2] методики ТИМ-2D**

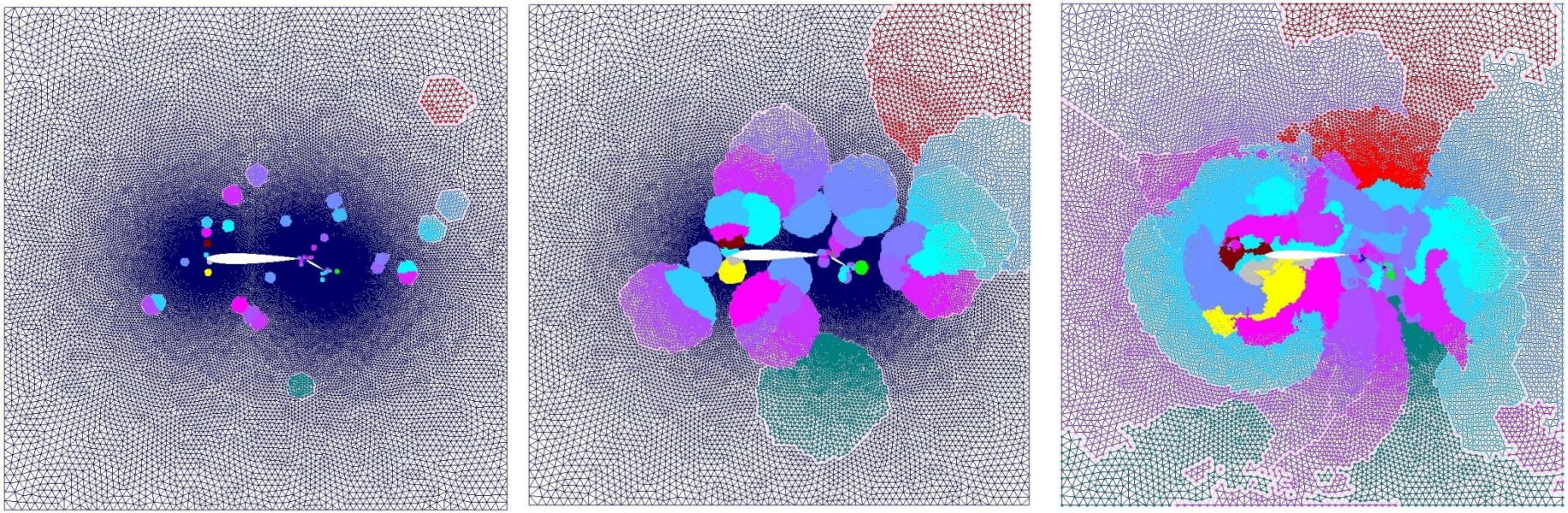


[1] А. И. Илюшин, А. А. Колмаков, И. С. Меньшов. Построение параллельной вычислительной модели путем композиции вычислительных объектов // Математическое моделирование. 2011. Т. 23. № 7. 97-113.

[2] А. А. Воропинов. Декомпозиция данных для распараллеливания методики ТИМ-2D и критерии оценки ее качества // Вестник ЮУрГУ. Серия «Математическое моделирование и программирование:», вып. 4. 2009. №37(170). 40-50.

Инкрементный алгоритм декомпозиции графов

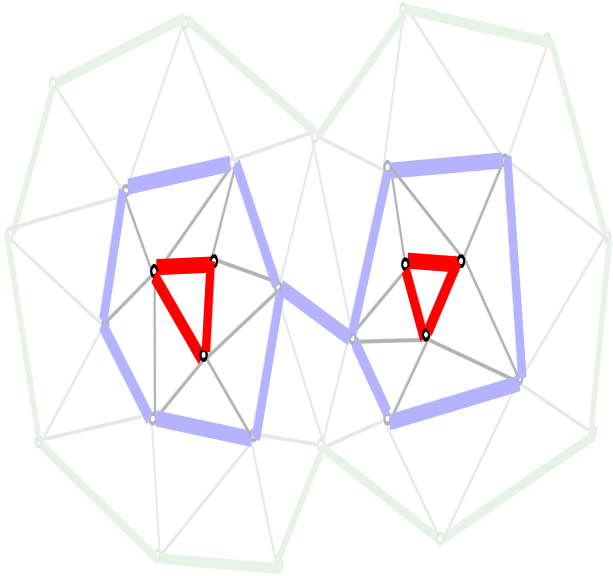
- **инкрементный рост доменов**
- **диффузное перераспределение вершин между доменами**



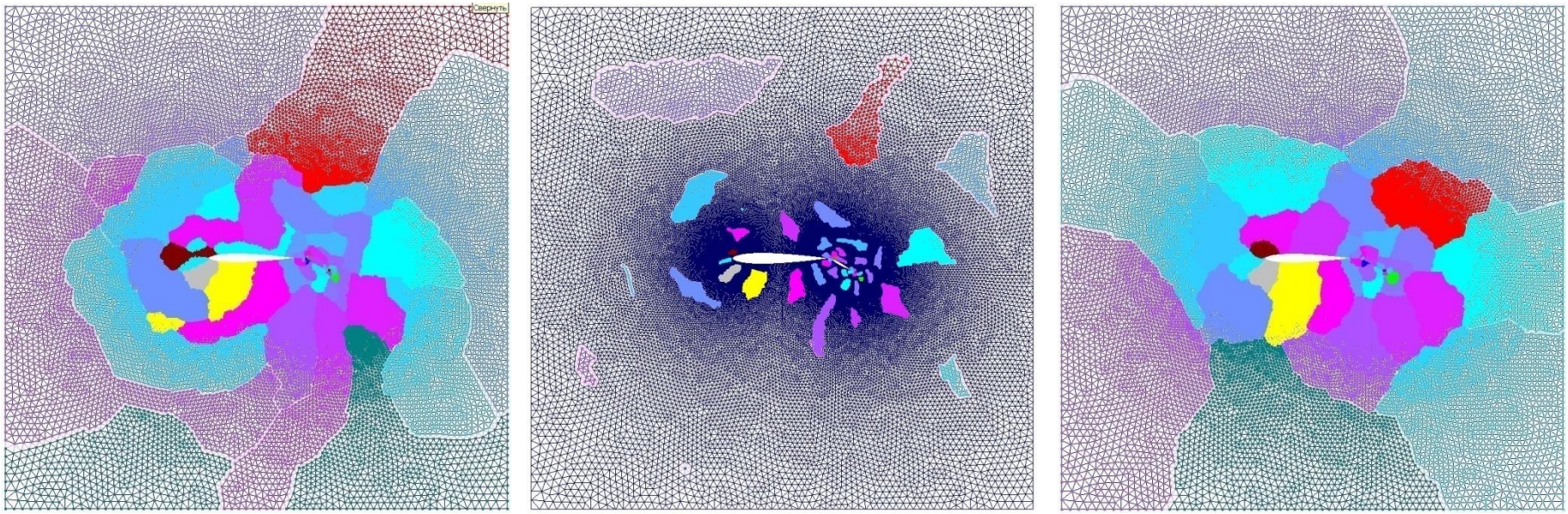
сетка вокруг крыла самолета с закрылком

Инкрементный алгоритм

- **локальное уточнение доменов**
- **проверка качества доменов**
- **освобождение части вершин плохих доменов**

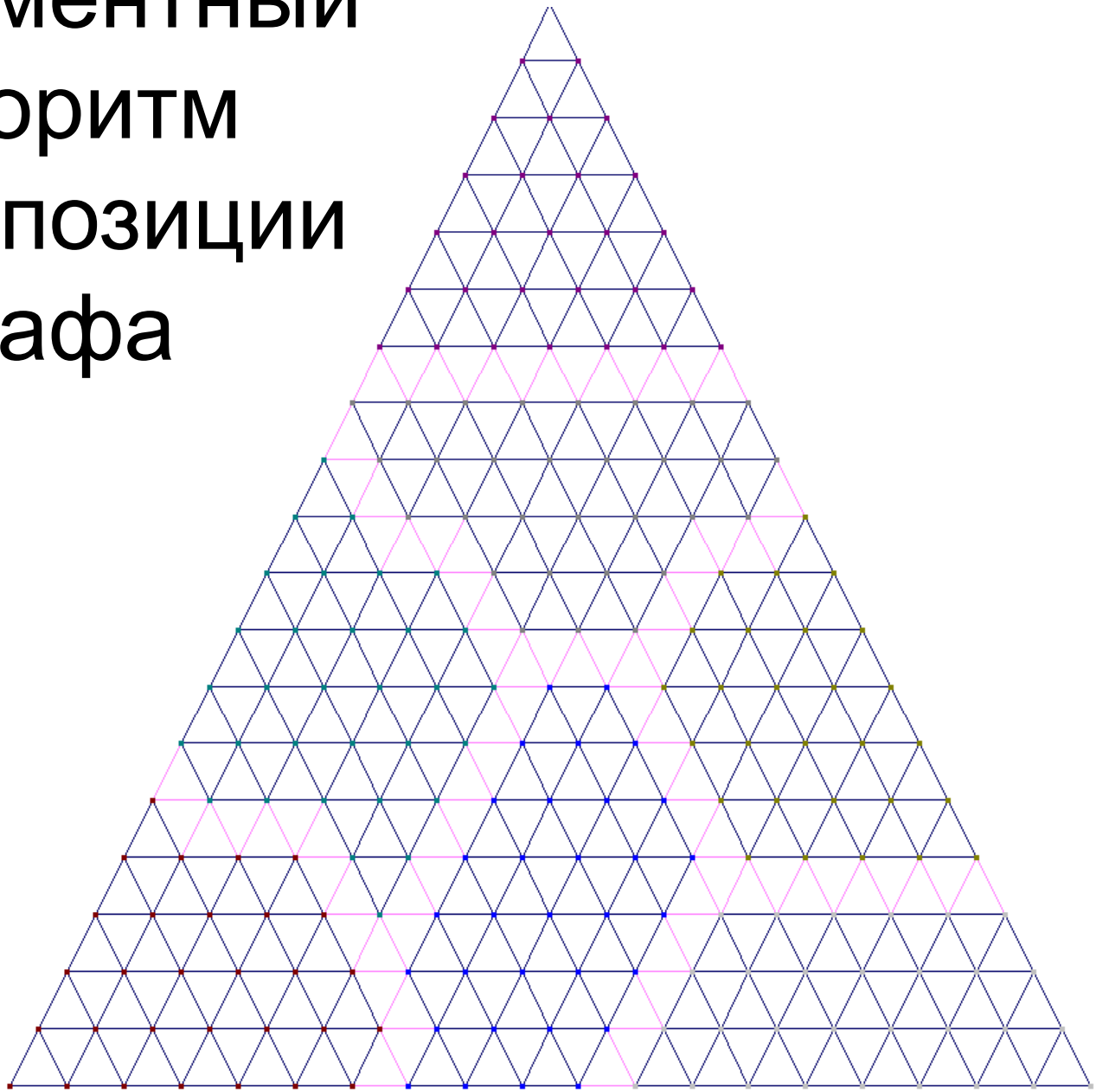


$$T_{k+1} = \mathbf{A}T_k \setminus T_k \setminus T_{k-1}, \quad T_0 = \phi$$

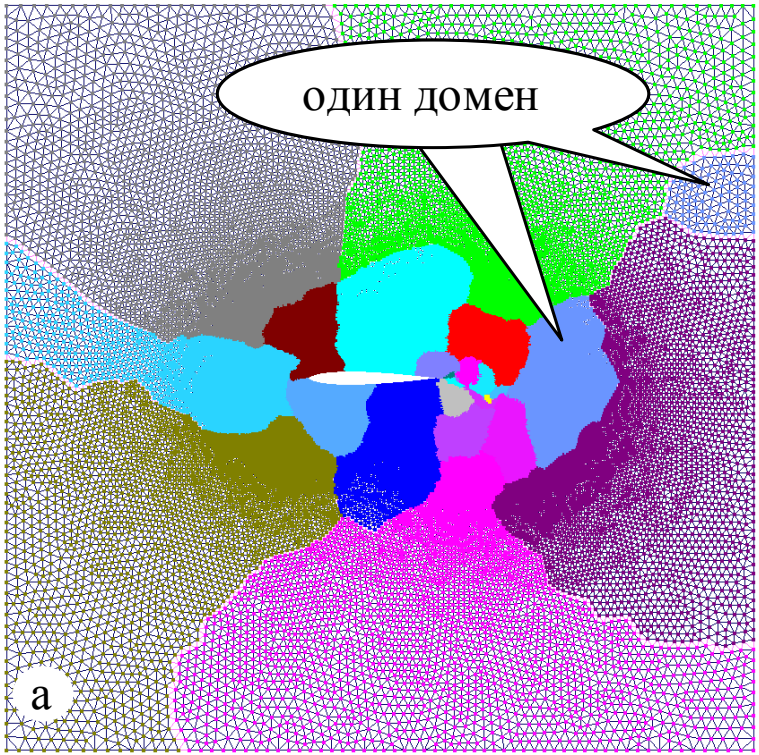


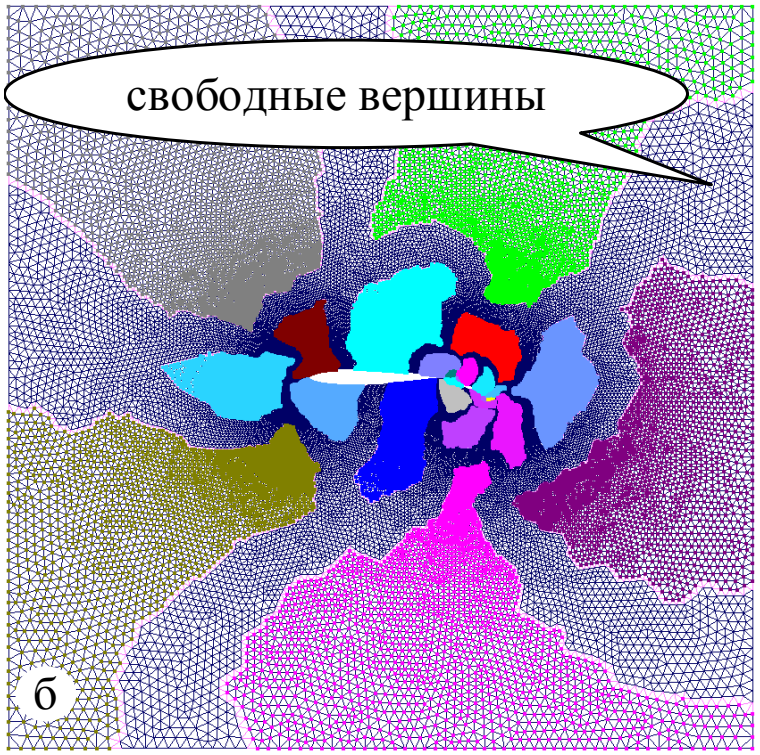
сетка вокруг крыла самолета с закрылком

Инкрементный алгоритм декомпозиции графа

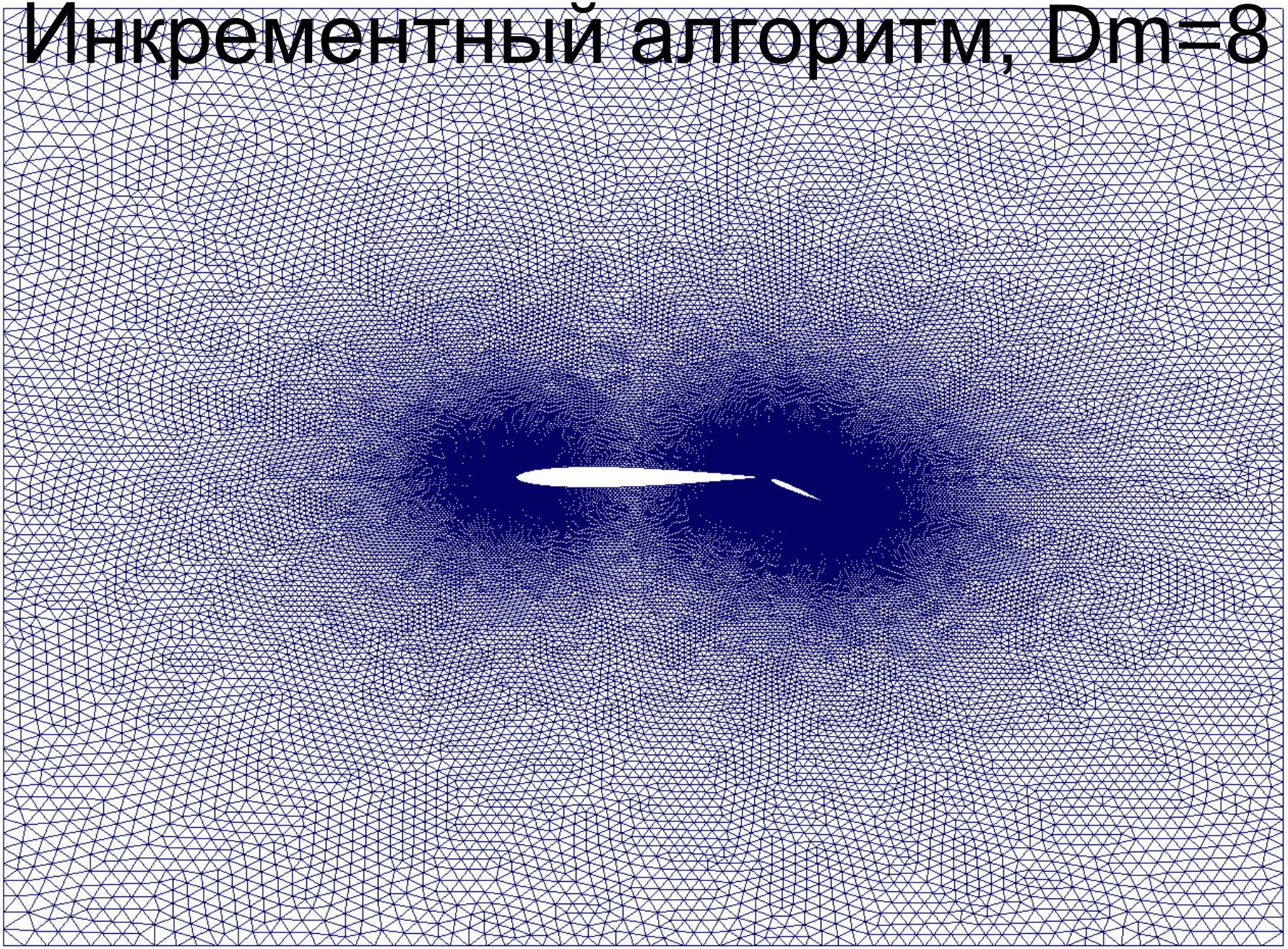


Редуцирование доменов

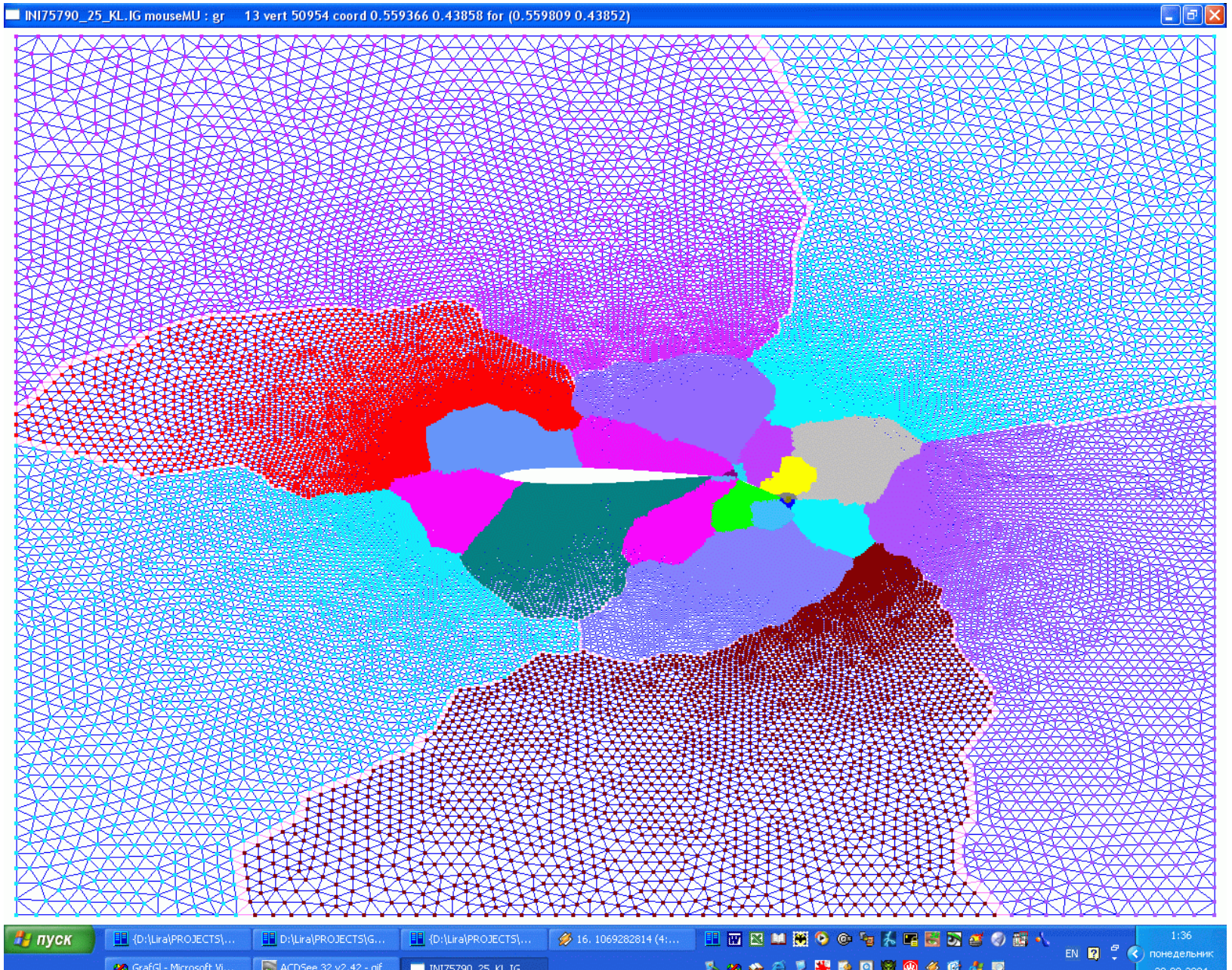




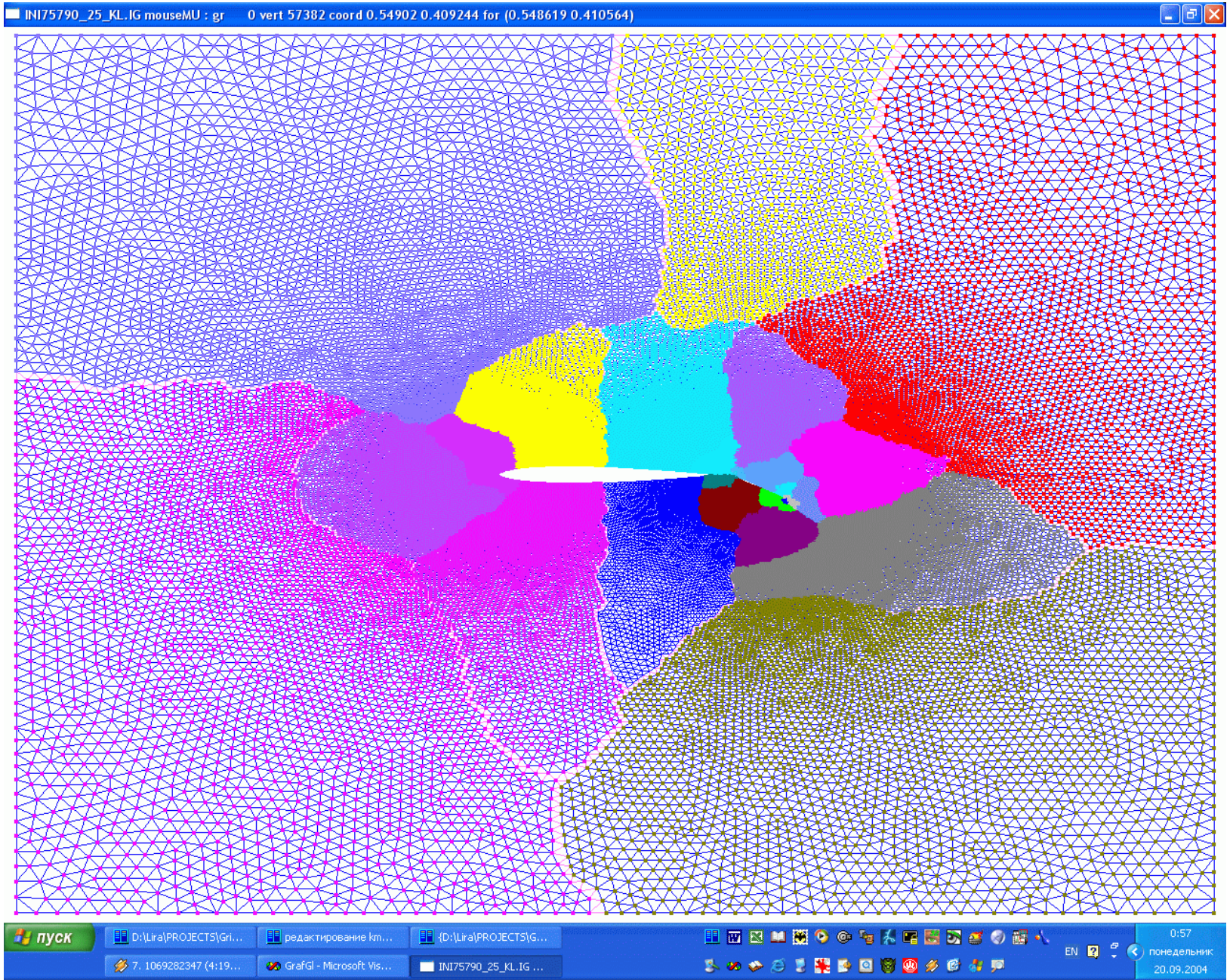
Инкрементный алгоритм, Dm=8



Инкрементный алгоритм, Dm=25

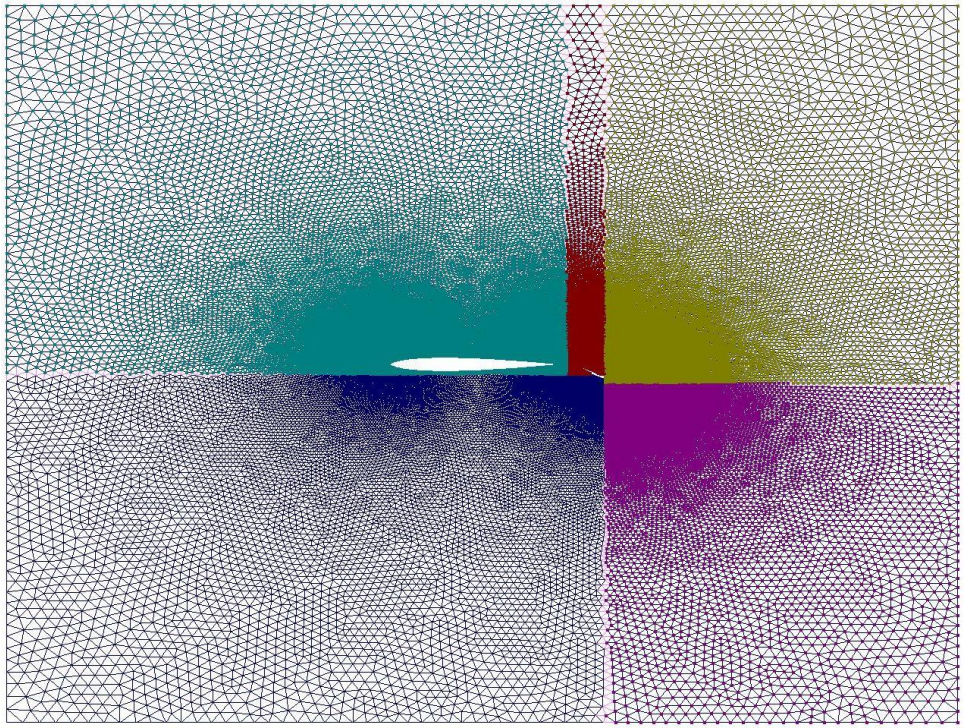


Kmetis, Dm=25

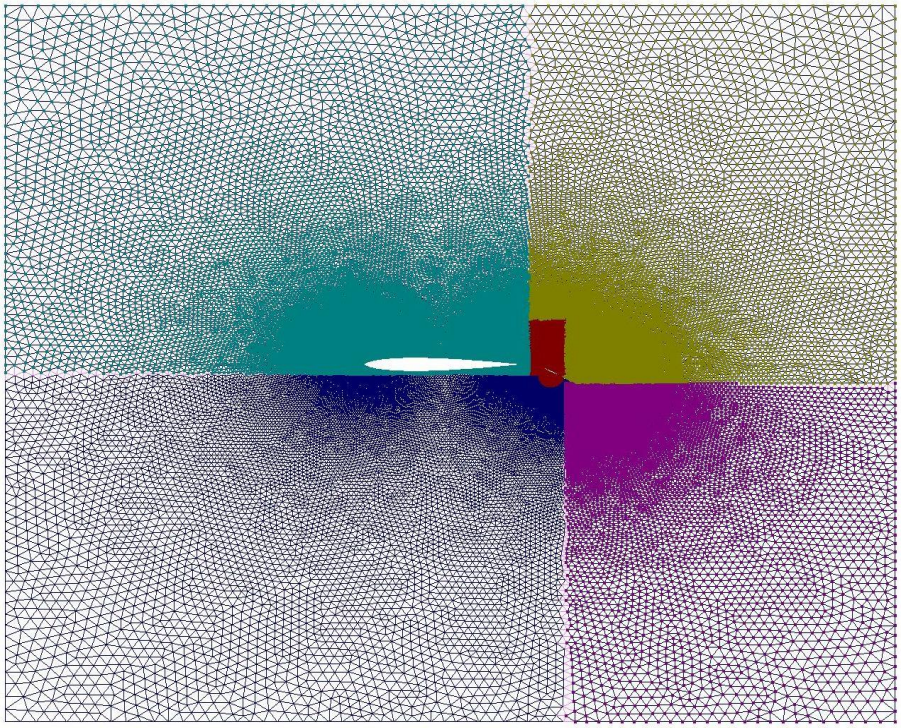

INI75790_25_KL.IG mouseMU : gr 0 vert 57382 coord 0.54902 0.409244 for (0.548619 0.410564)
пуск
D:\Lira\PROJECTS\Gri...
редактирование km...
{D:\Lira\PROJECTS\G...
7. 1069282347 (4:19...
GrafGl - Microsoft Vis...
INI75790_25_KL.IG ...
EN
0:57
понедельник
20.09.2004

Треугольная сетка из 75790 вершин (пространство вокруг крыла)

результат геометрической декомпозиции на 5 групп
(в дальнейшем каждый процессор считывает свою группу вершин)

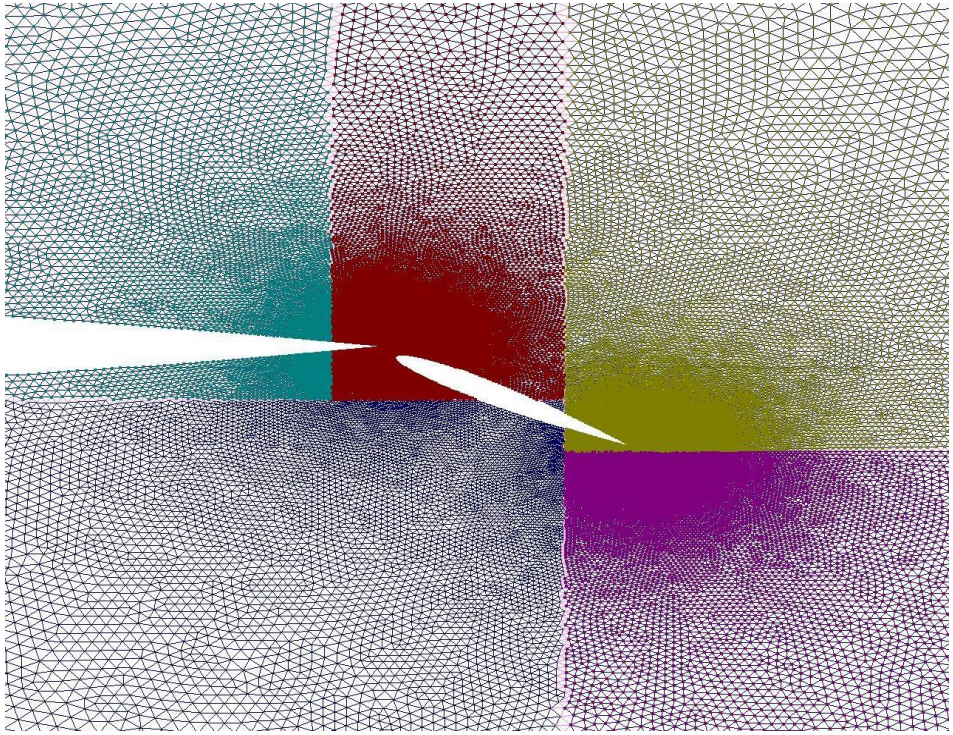


результат перераспределения малых блоков вершин

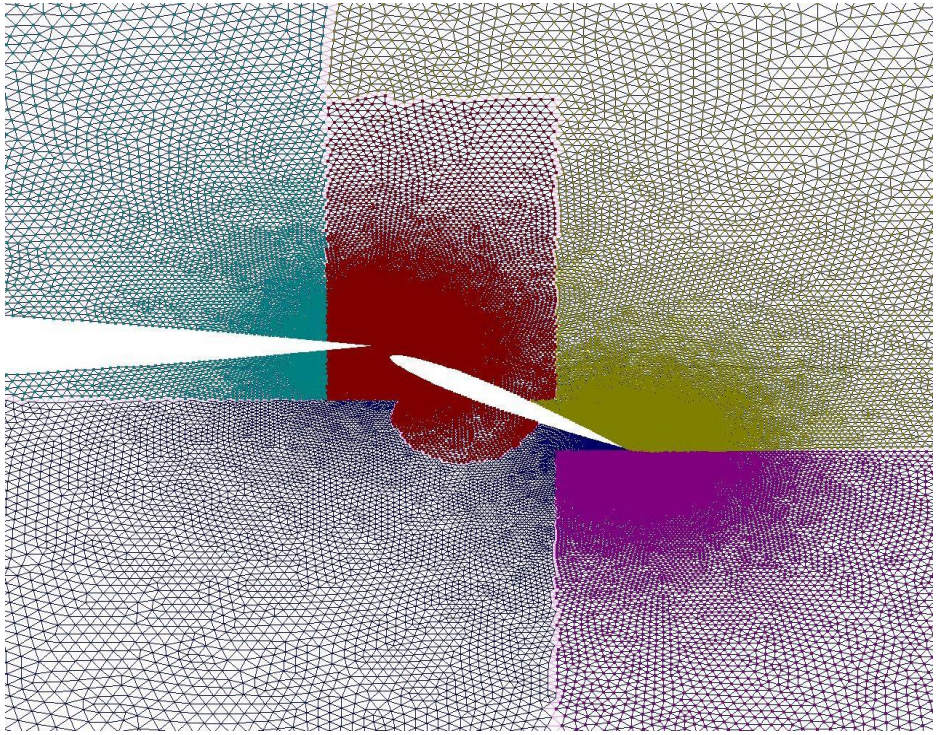


Фрагмент треугольной сетки из 75790 вершин

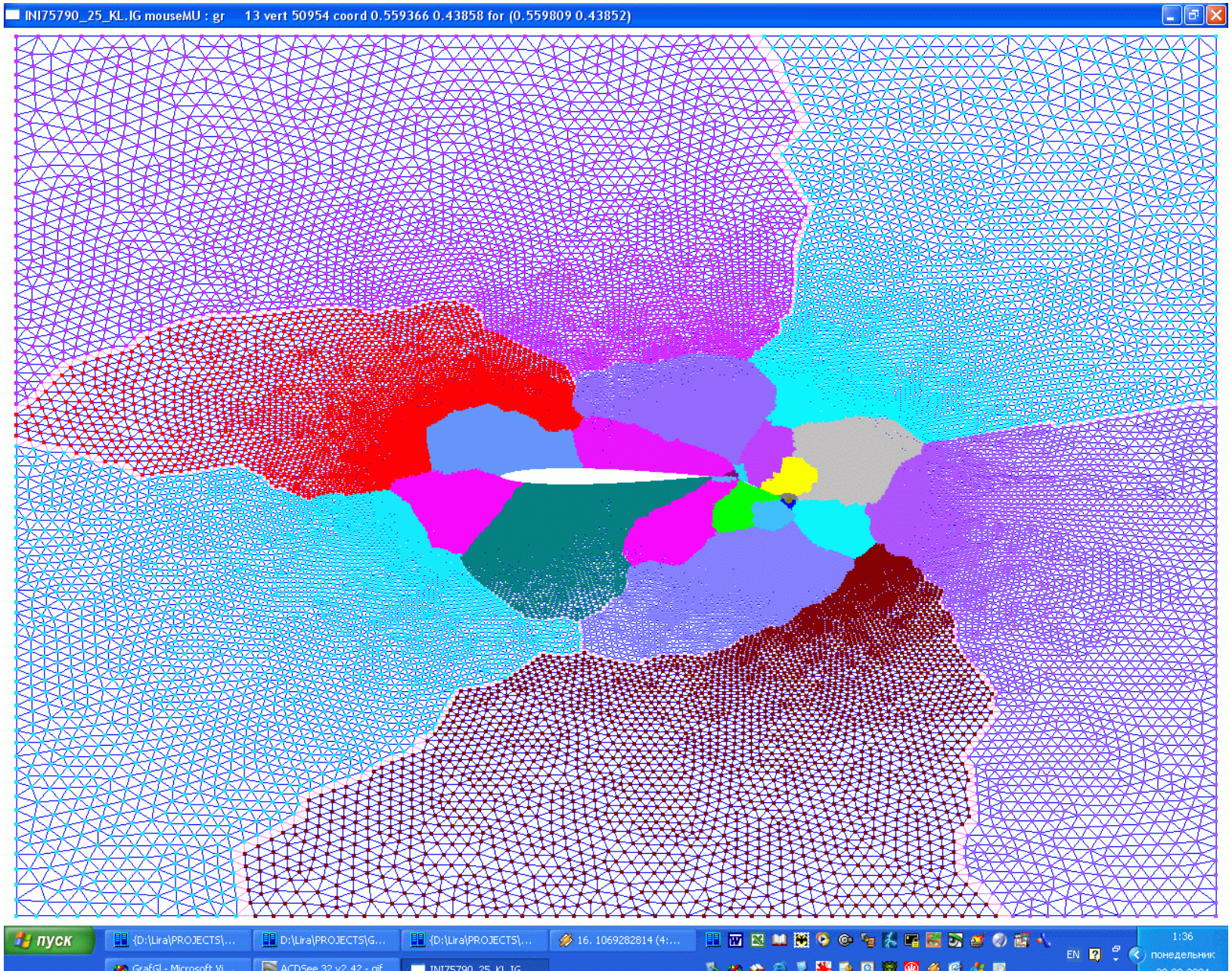
результат геометрической декомпозиции



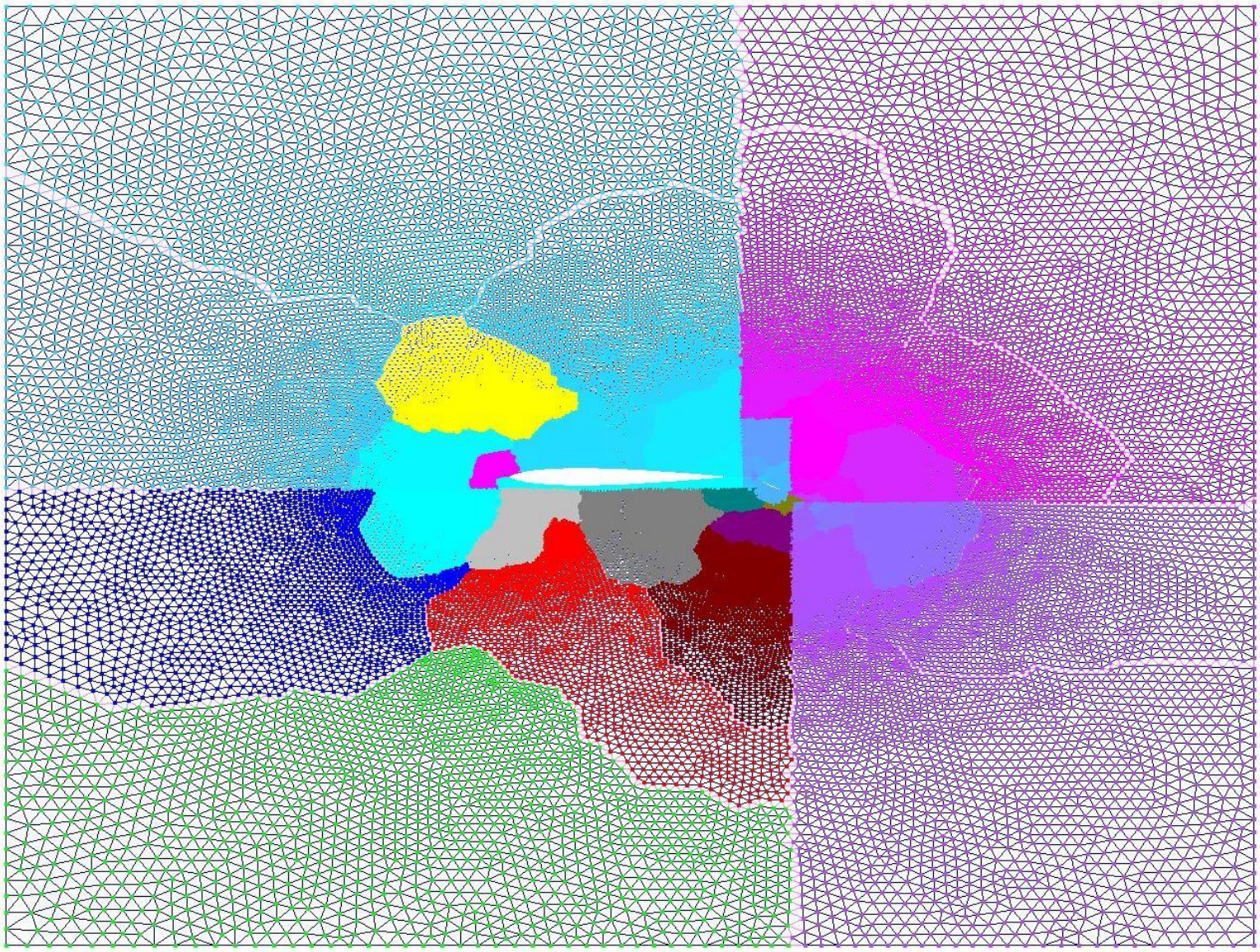
результат перераспределения малых блоков вершин



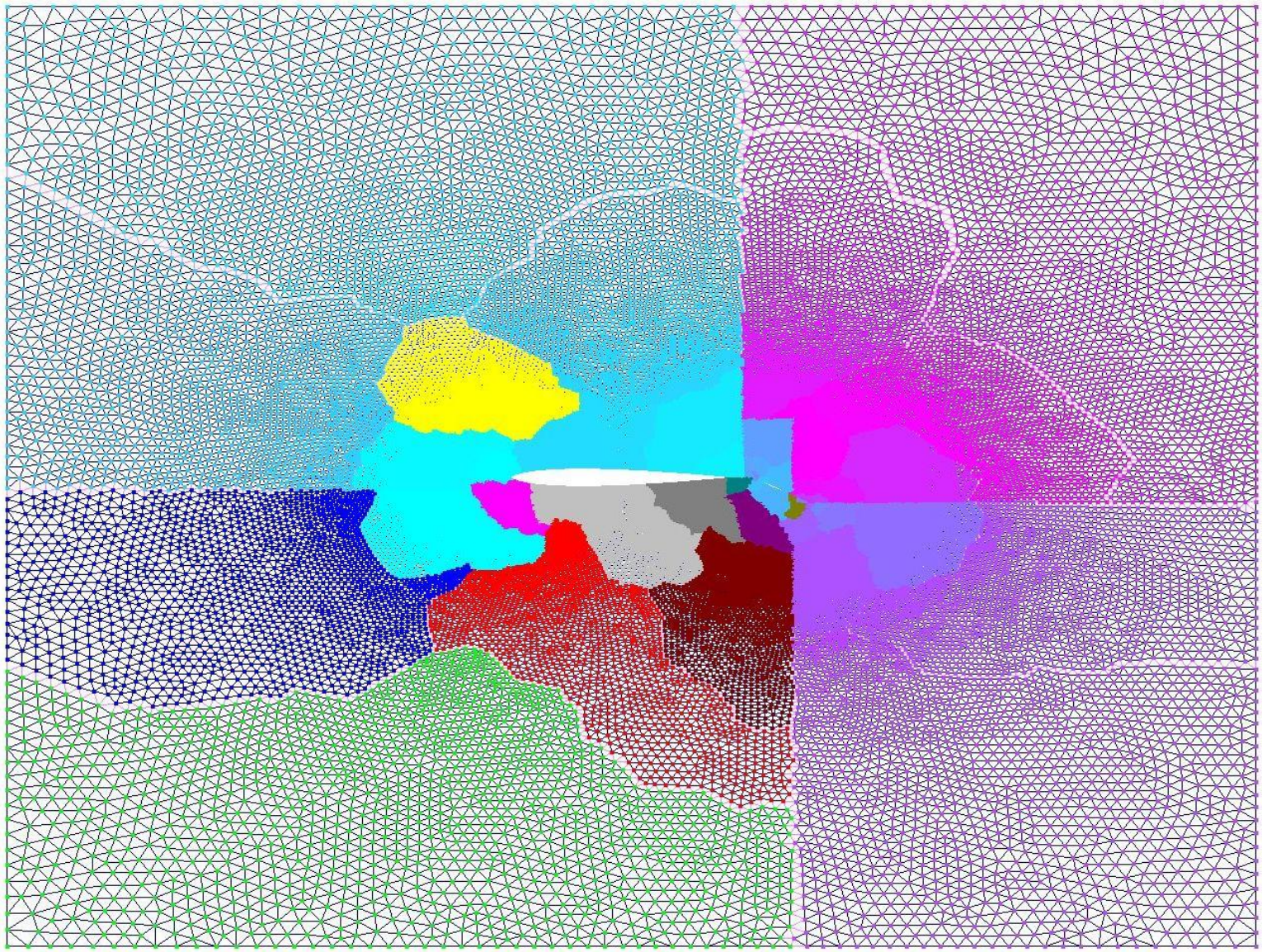
Инкрементный алгоритм, Dm=25



Результат локального разбиения сетки из 75790 вершин на 50 доменов на 5 процессорах



Результат сбора плохих групп доменов и их повторного разбиения

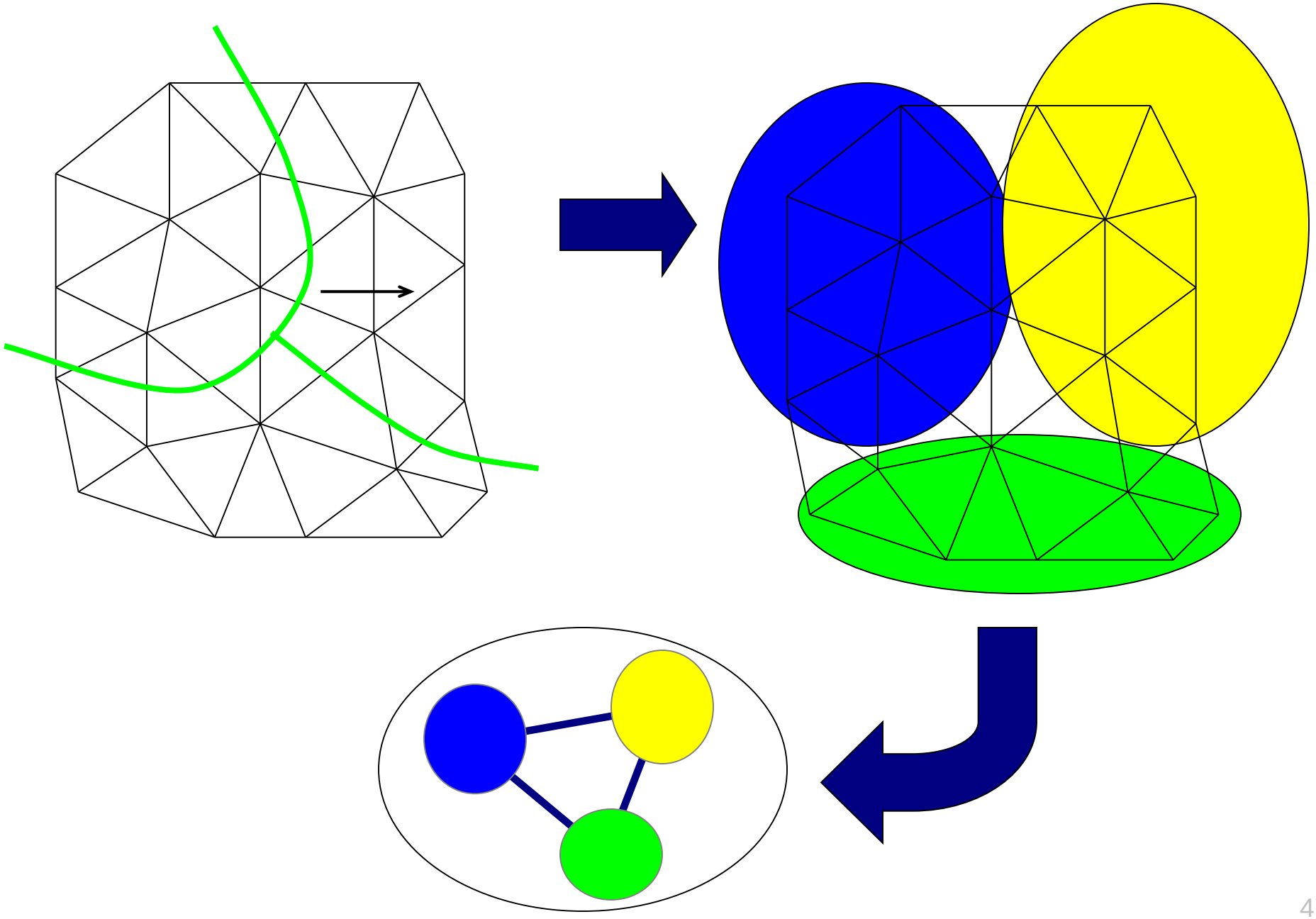


Разбиение тетраэдральной сетки, содержащей $2 \cdot 10^8$ узлов, на 125 процессорах

- вычисления производились на кластере СКИФ МГу (1250 4-хядерных процессоров, 60 TFlop/s)

		геометрическая декомпозиция		ParMETIS	
число доменов		80 000		20 000	
время		21 сек.		10 сек.	
число вершин в домене		2612	2613	2 328	10 932
мин.	макс.				
среднее число связей с соседними доменами		14		14	
число некомпактных доменов		229		364	

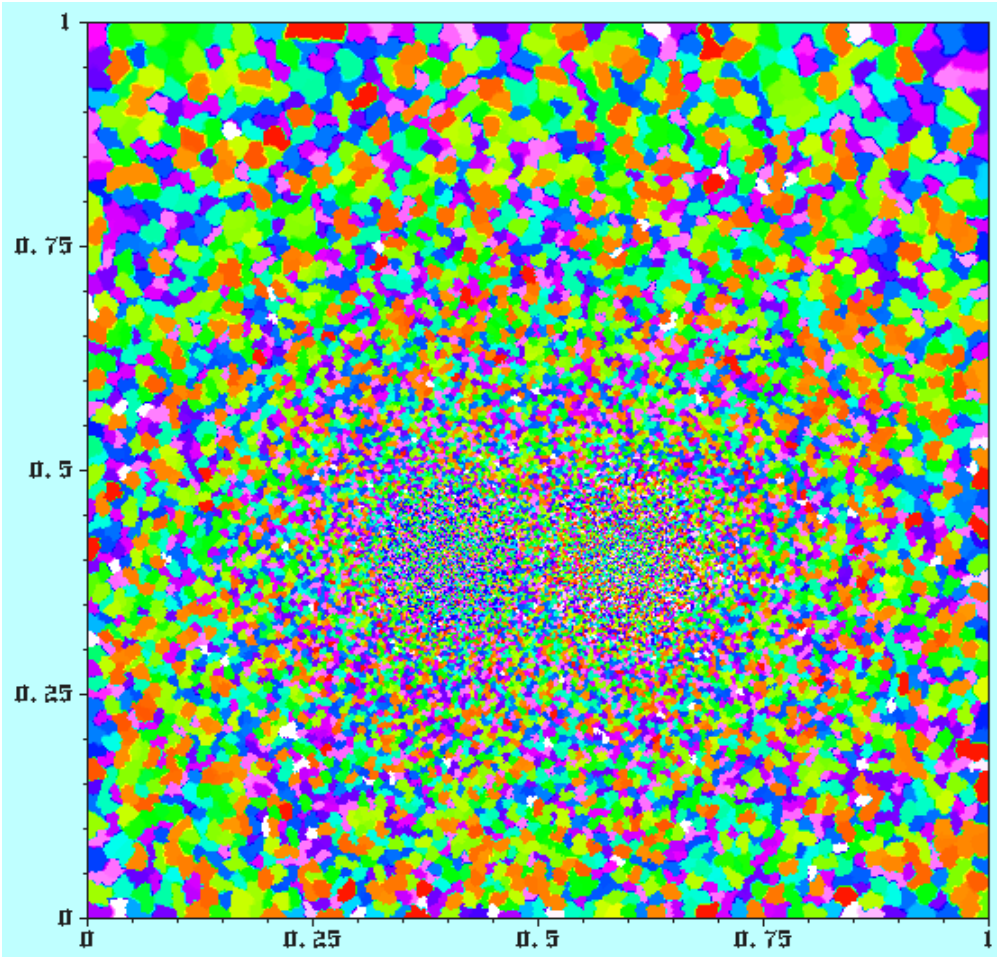
Формирование макрографа



Сетка микродоменов

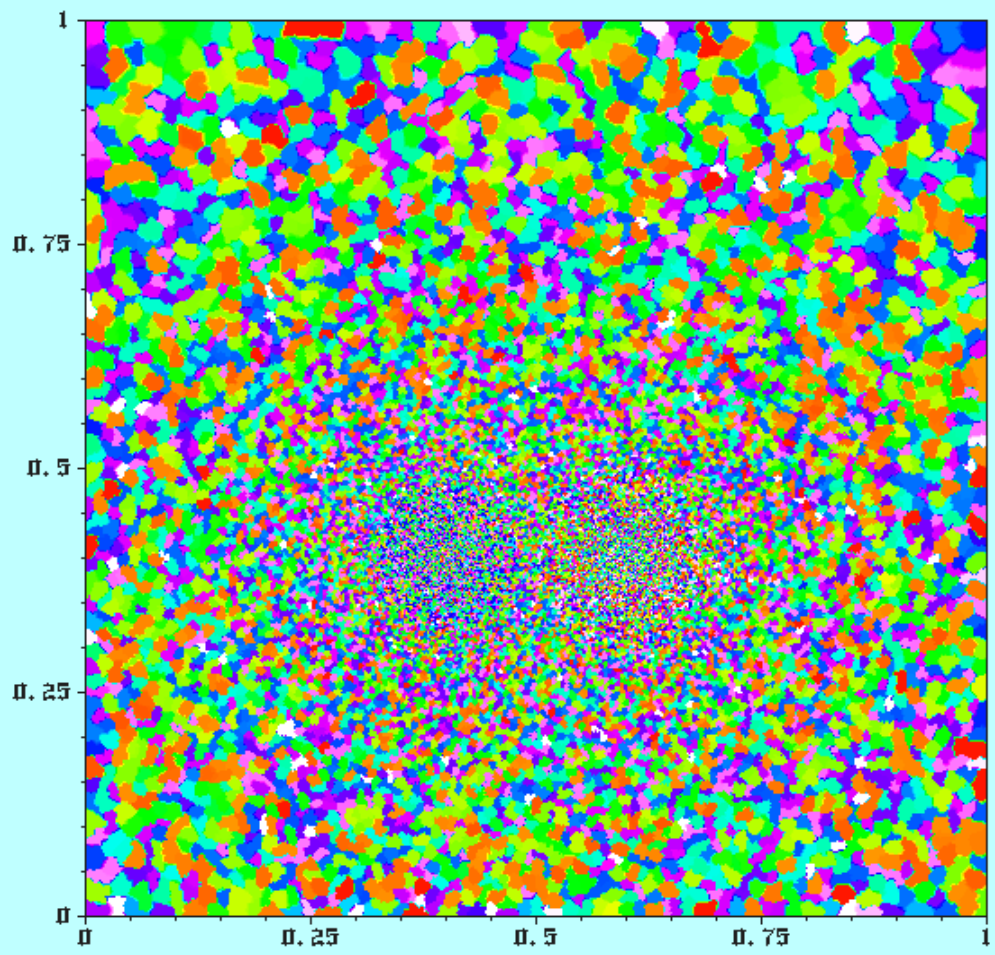
51 538 микродомен

в каждом около 20 узлов



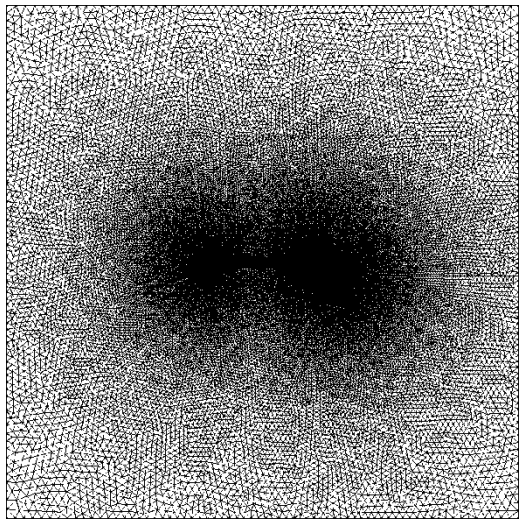
Сетка микродоменов

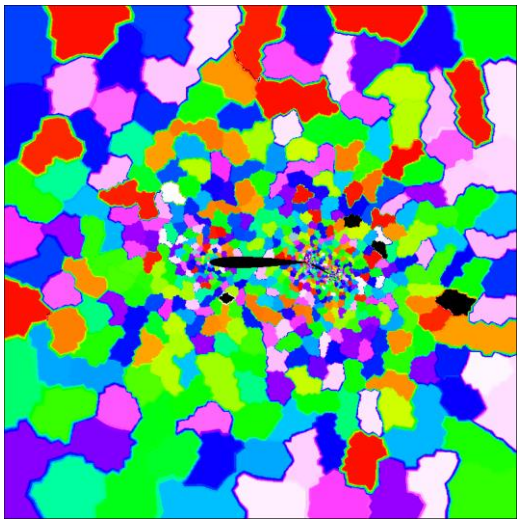
вес	число	% отн. число
12	3	0.01%
13	3	0.01%
14	15	0.03%
15	33	0.06%
16	228	0.44%
17	1 373	2.66%
18	5 433	10.54%
19	11 713	22.73%
20	14 218	27.59%
21	11 069	21.48%
22	5 737	11.13%
23	1 505	2.92%
24	192	0.37%
25	13	0.03%
26	2	0.00%
27	1	0.00%

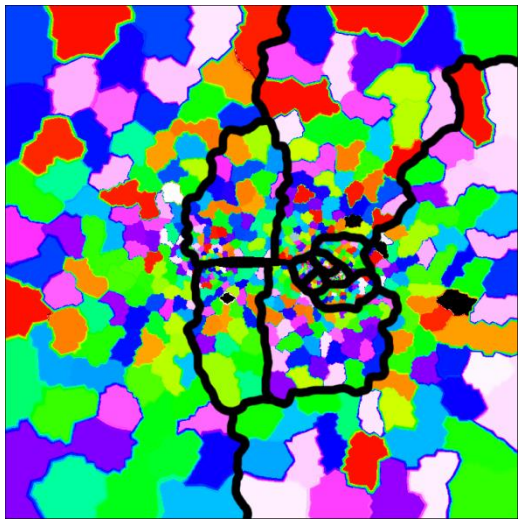


51 538 микродомен

Двухуровневое разбиение







I

II

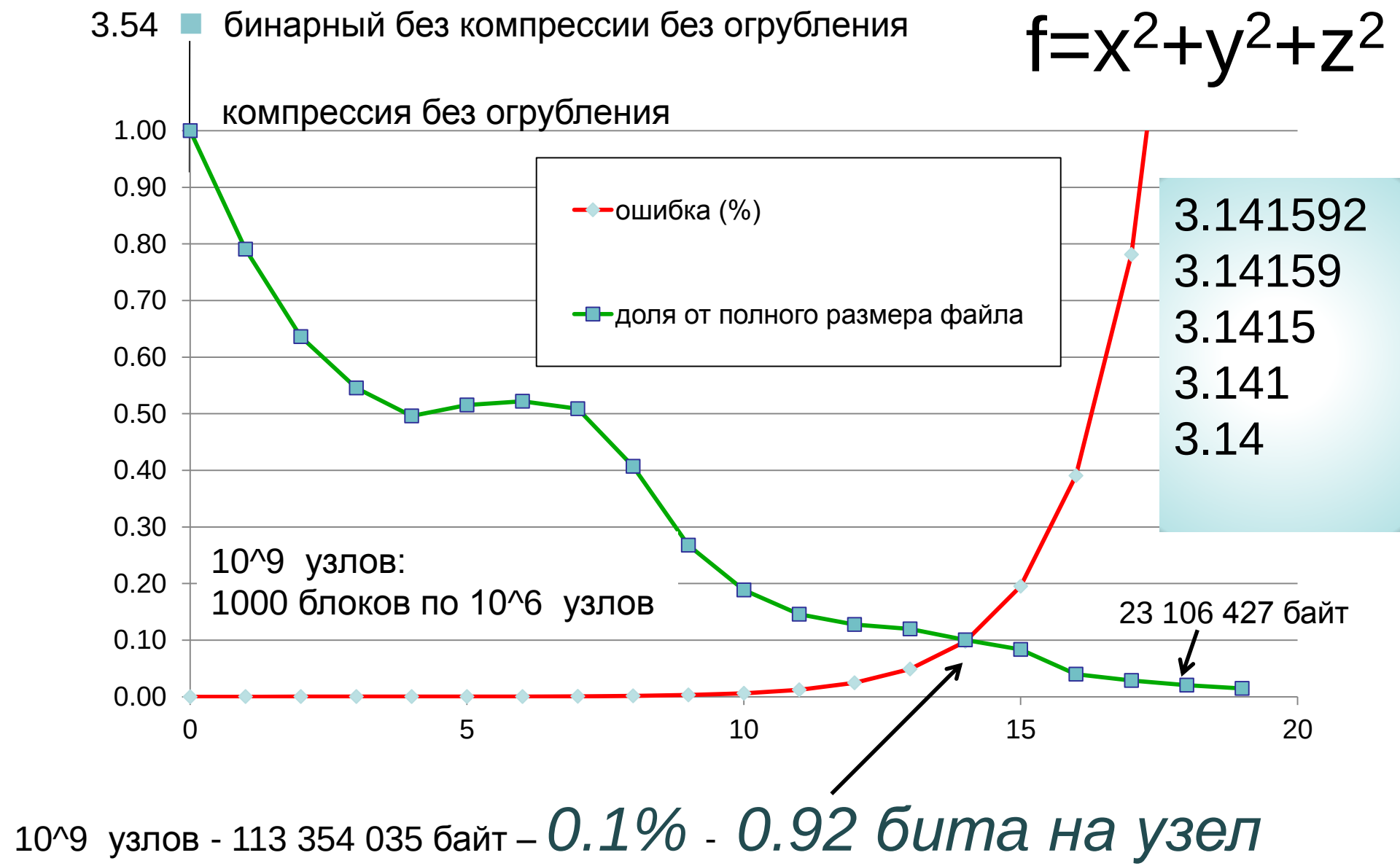
Сетка предварительно разбивается на большое число *микродоменов,* образующих *макрограф*

Вершины *макрографа* распределяются по процессорам

Метод эффективен для сверхбольших сеток

Отсечение младших бит мантиссы

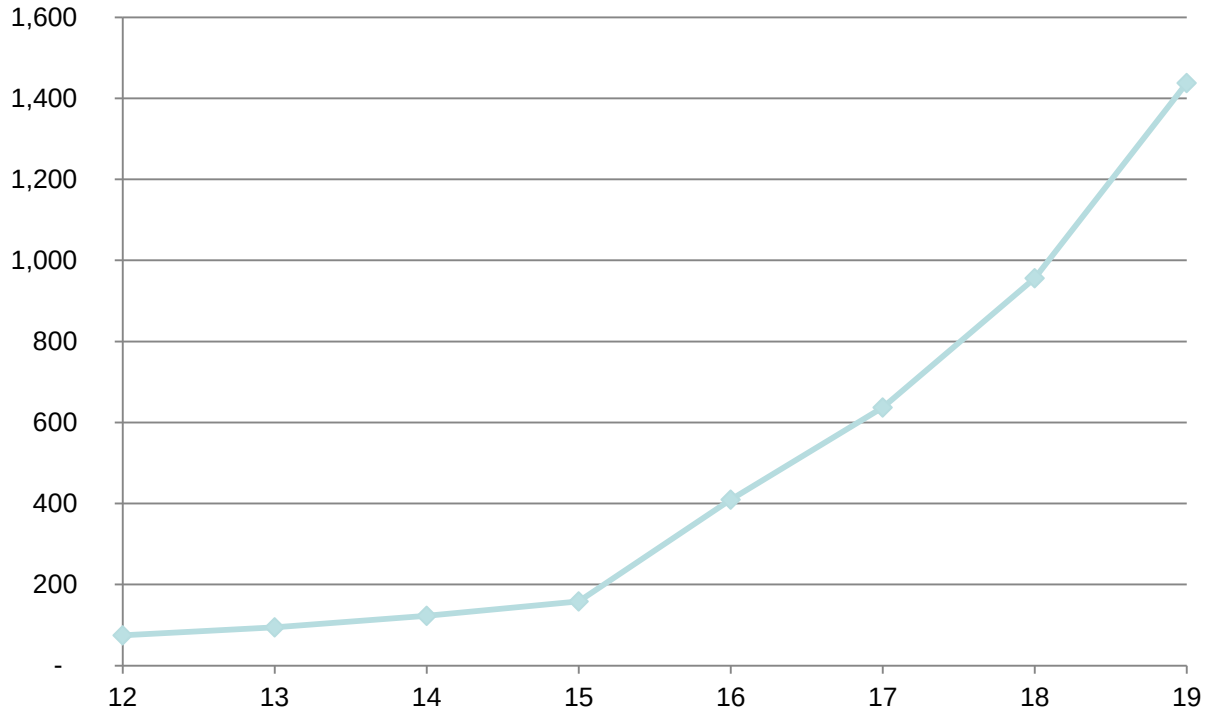
$f=x^2+y^2+z^2$



10^9 узлов - 113 354 035 байт – *0.1% - 0.92 бита на узел*

Зависимость коэффициента сжатия от числа усеченных бит

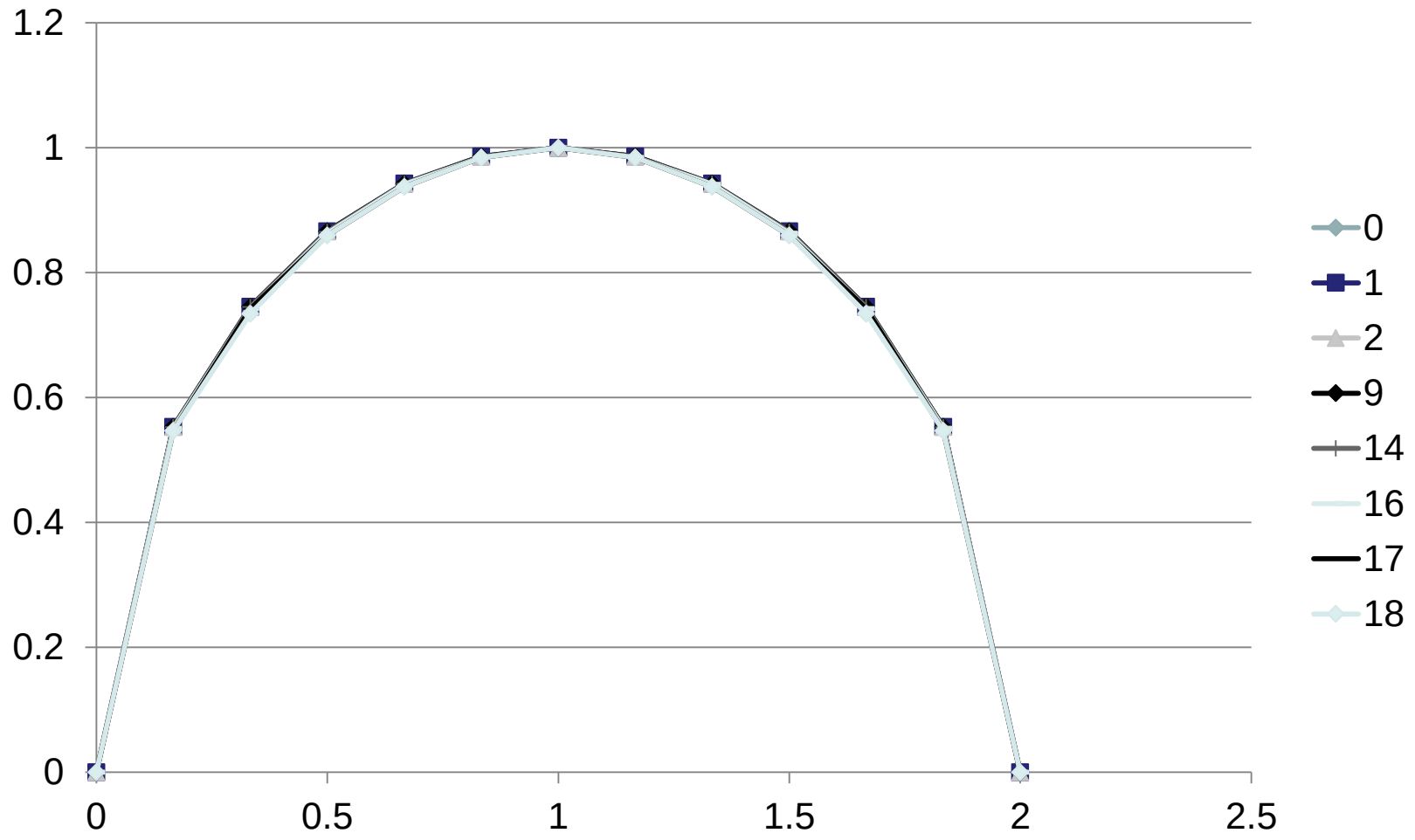
Сетка: 1000 x 3500 x 150 = 525 млн узлов



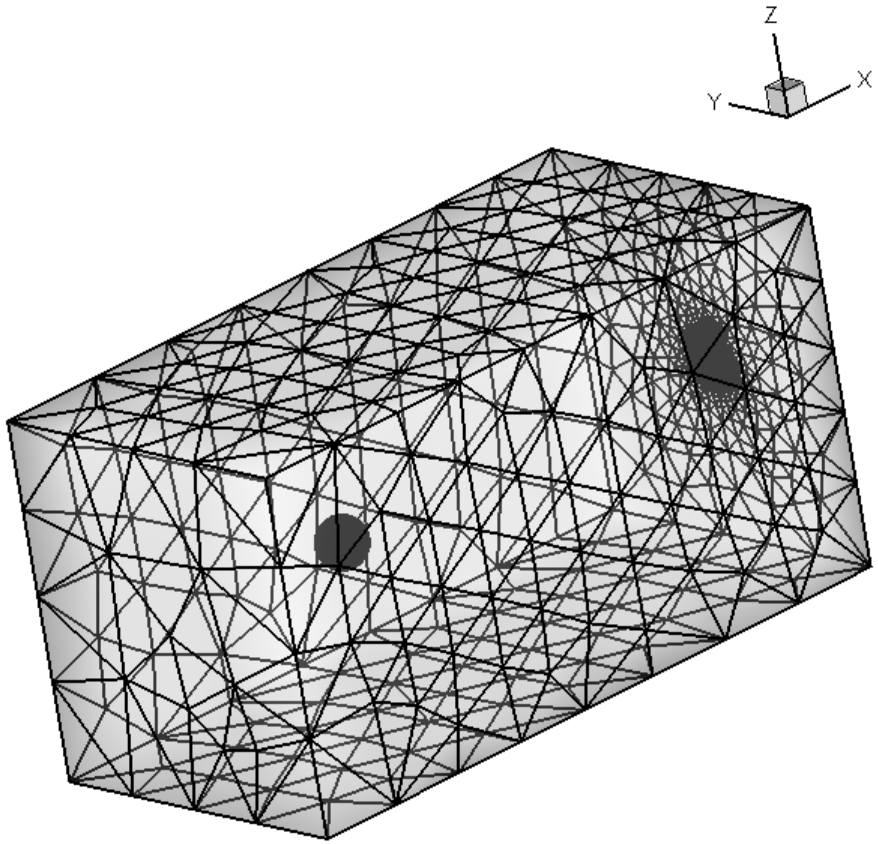

```
28 244 379   w101_reduced 12.bjn
22 340 718   w101_reduced 13.bjn
17 228 023   w101_reduced 14.bjn
13 339 249   w101_reduced 15.bjn
 5 171 208   w101_reduced 16.bjn
 3 321 150   w101_reduced 17.bjn
 2 213 949   w101_reduced 18.bjn
 1 471 818   w101_reduced 19.bjn

   793 457   w101grid.bjn
```

Огрубление данных

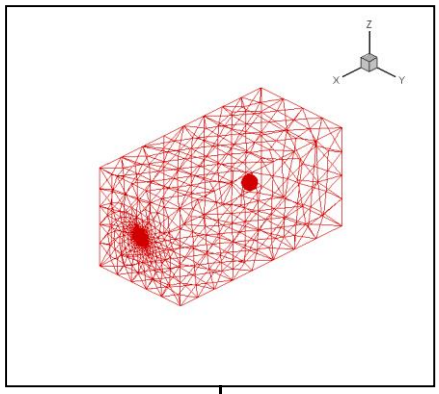


Тетраэдральная сетка

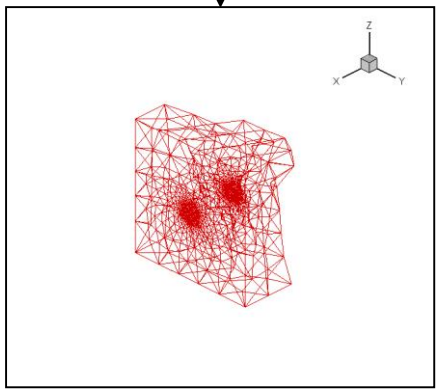


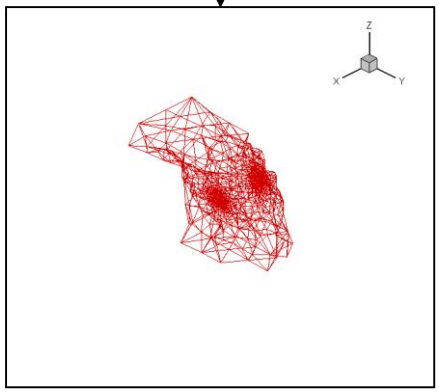
Сетка подготовлена Ю.В.Василевским

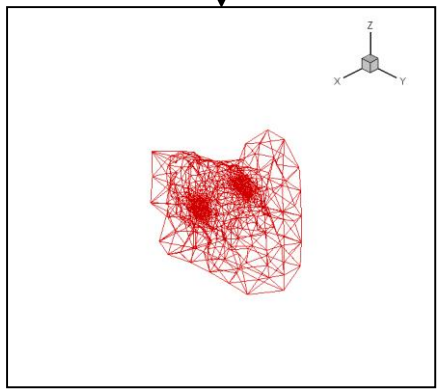
Декомпозиция тетраэдральной сетки

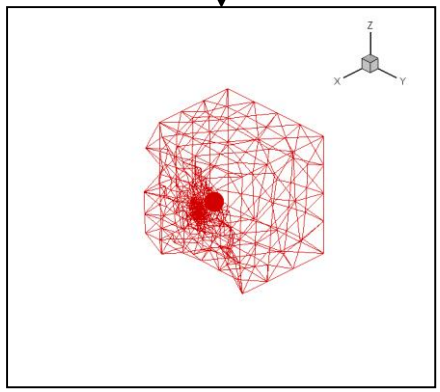


Расчетная сетка:
70 300 узлов, 401 418 тетраэдров



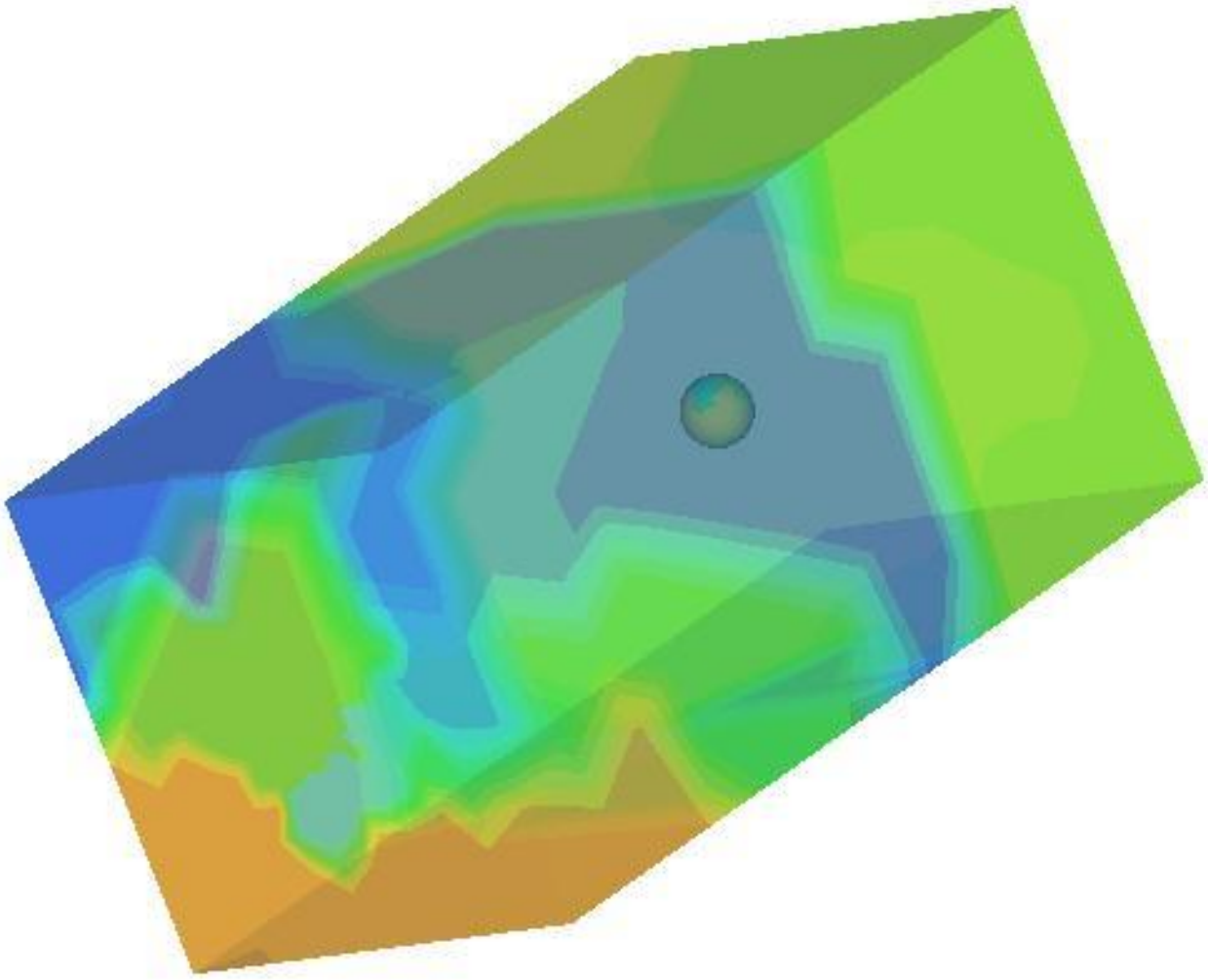




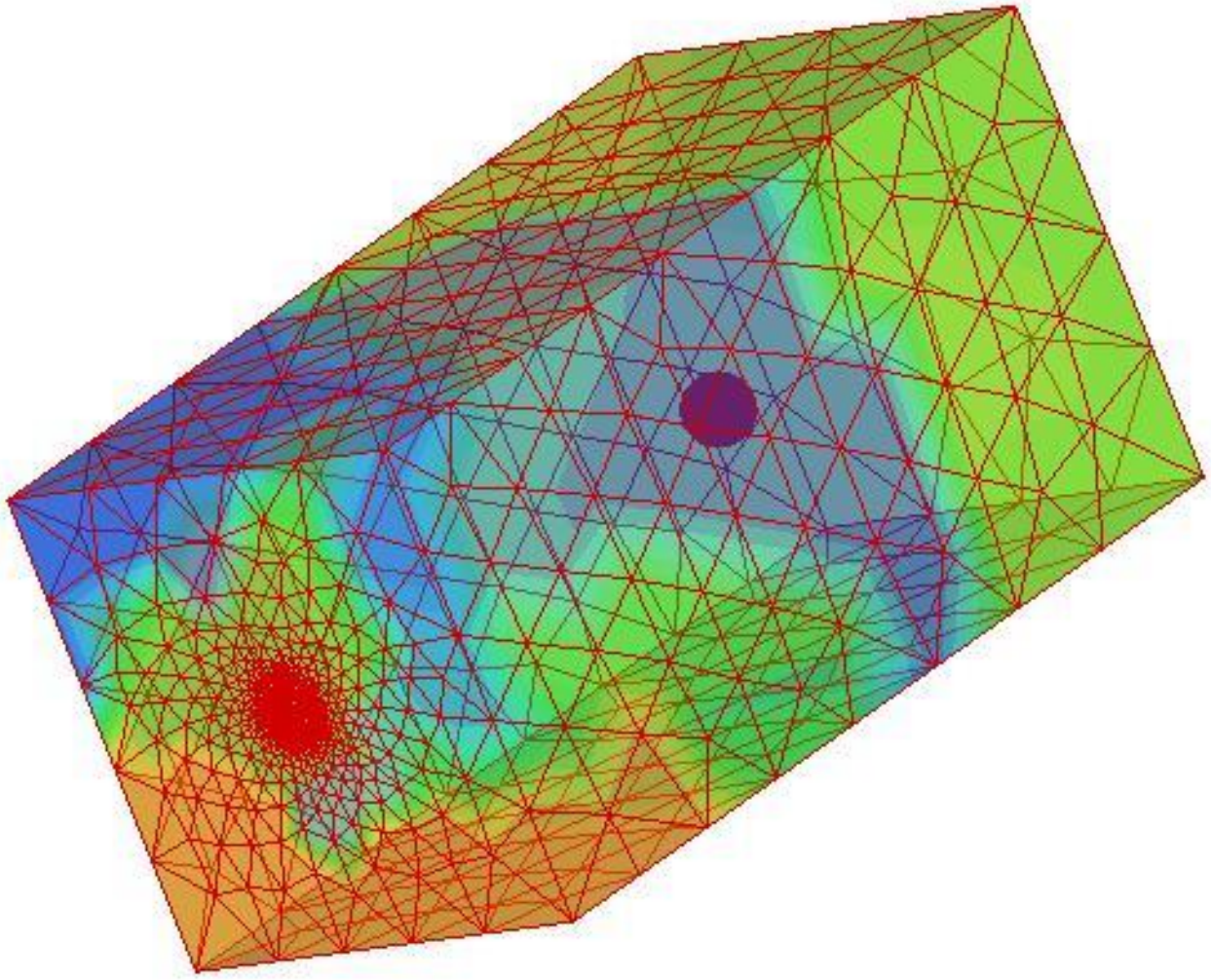


Число доменов	Число разрезанных ребер (%)	Эффективность метода конечного объема	Эффективность метода конечных элементов
4	1.54	0.96	0.94
10	4.50	0.92	0.88
40	10.18	0.87	0.78

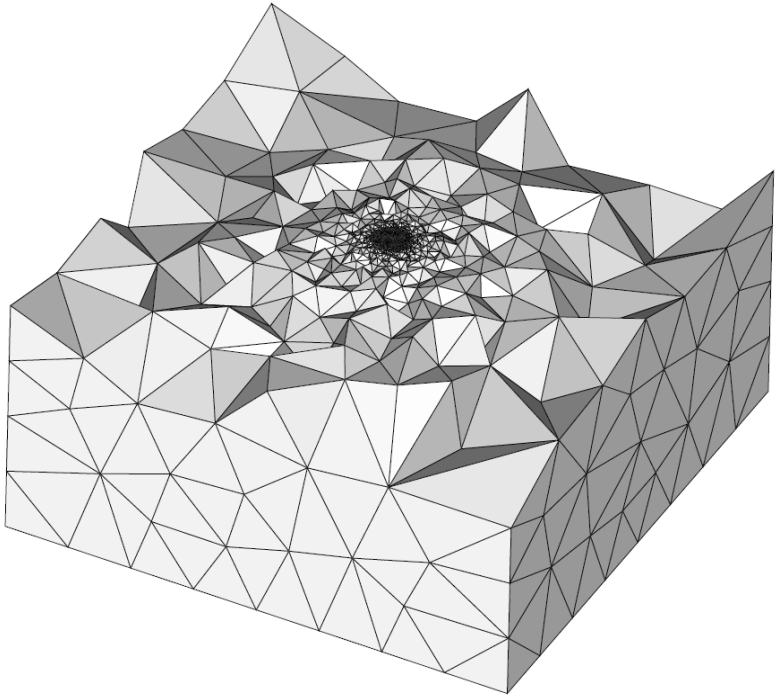
Разбиение на 12 доменов

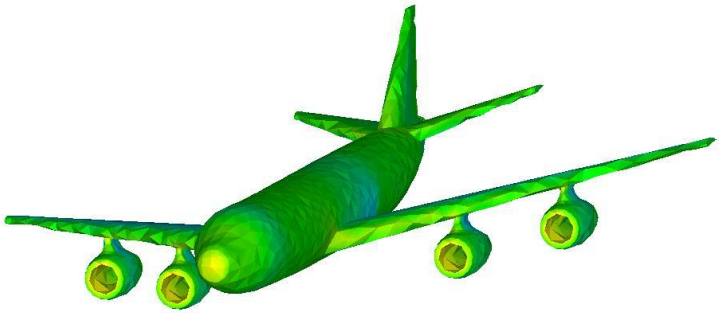


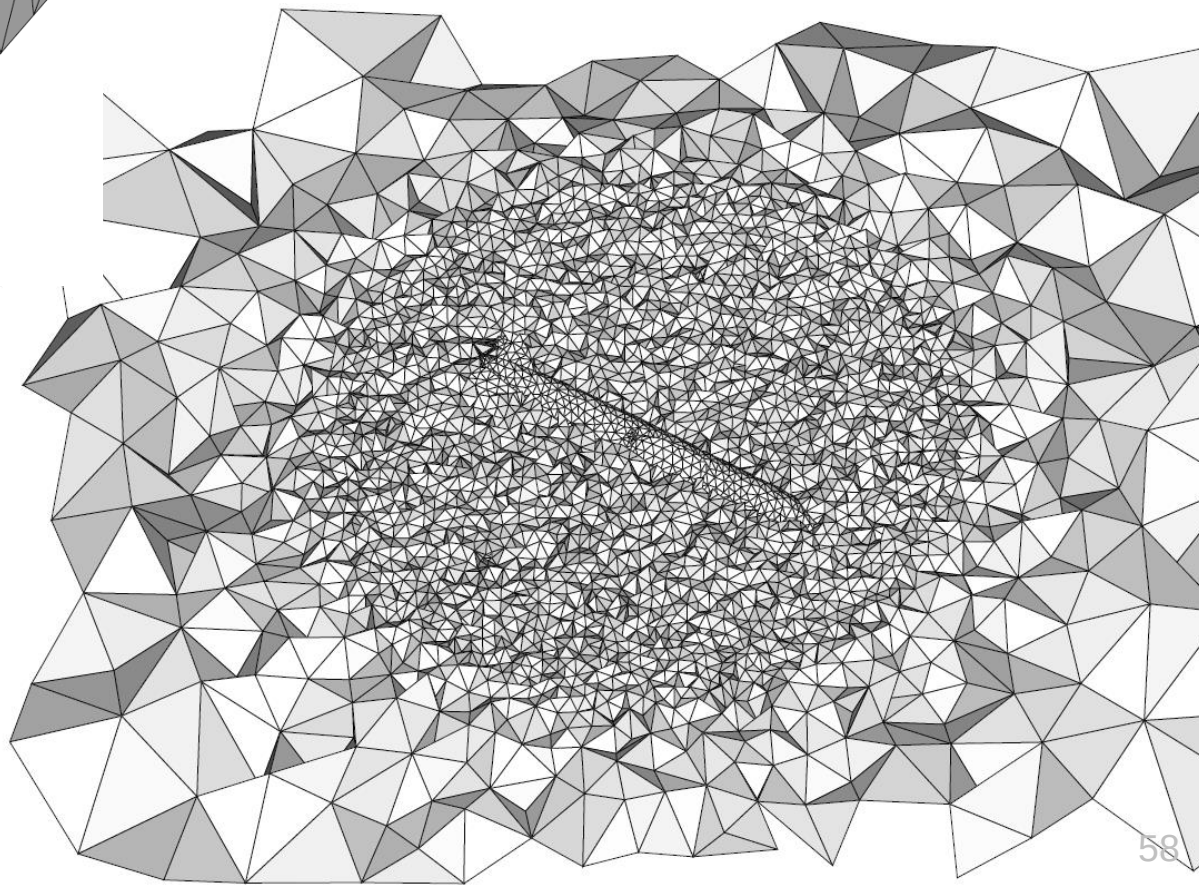
Поверхностная сетка и разбиение



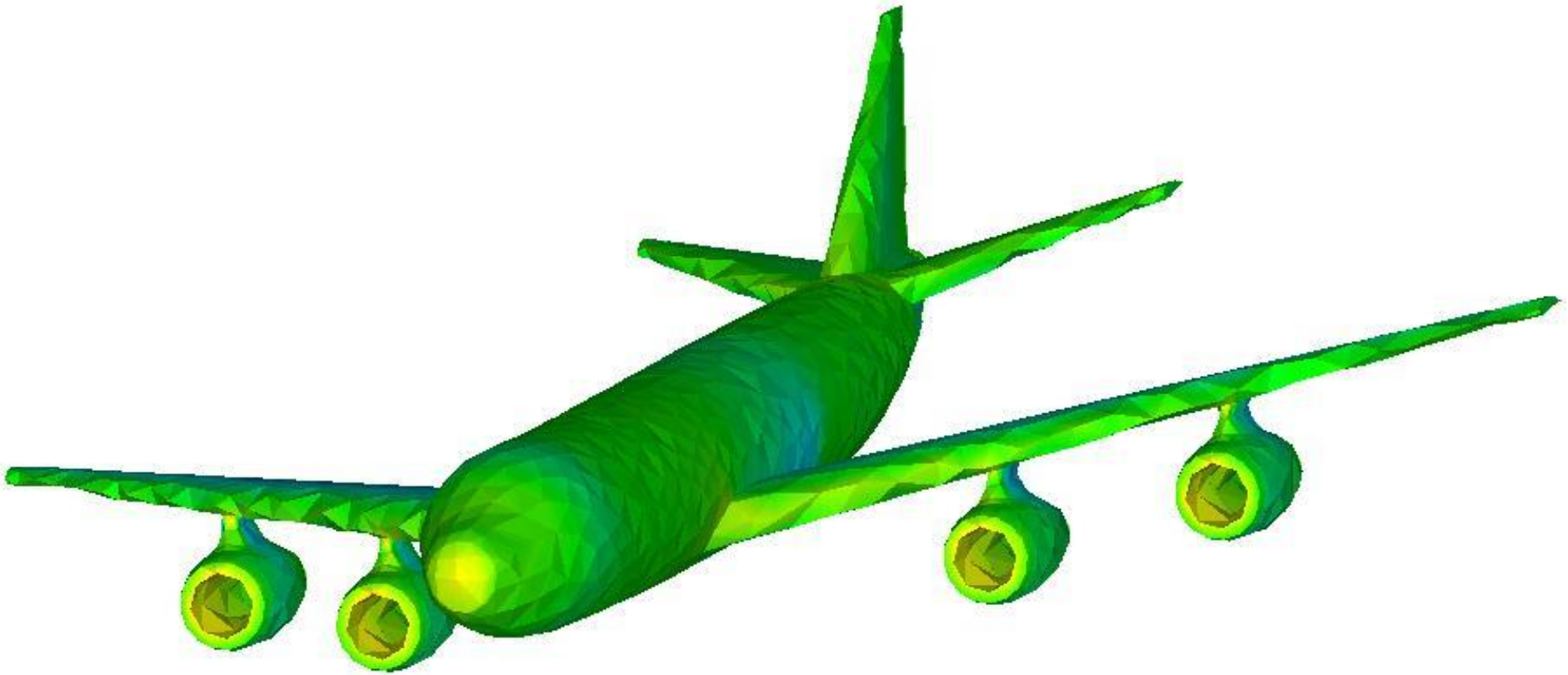
Сетки Неструктурированные



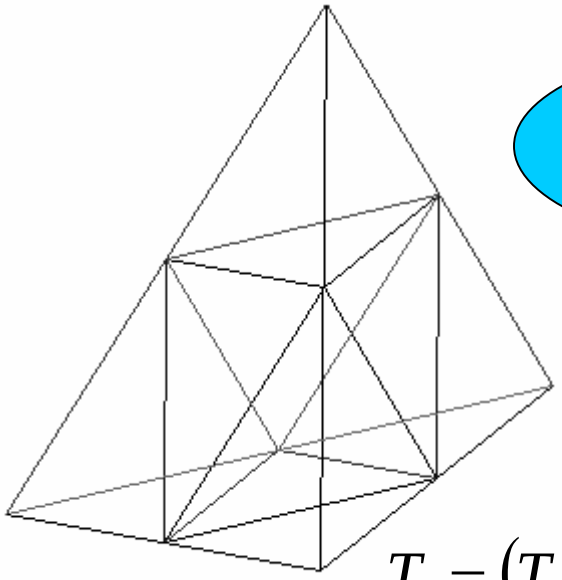


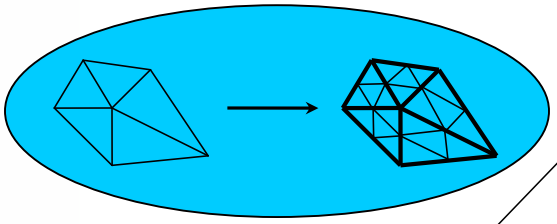


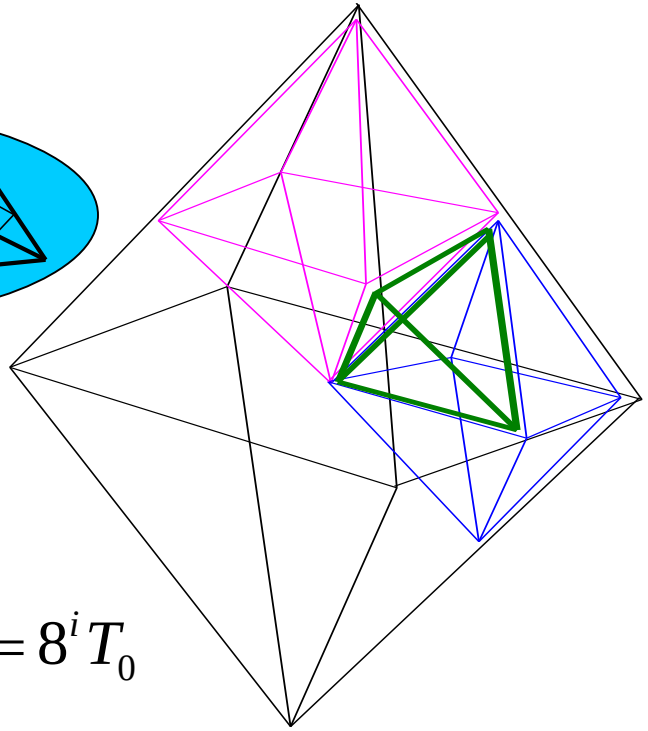
Тетраэдральные сетки 10^8 узлов



Измельчение тетраэдральной сетки



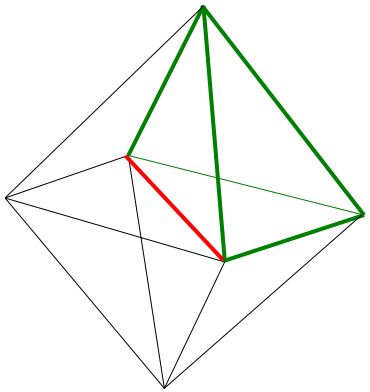




$$T_i = (T_0, 0)\begin{pmatrix} 4 & 1 \\ 8 & 6 \end{pmatrix}^i \begin{pmatrix} 1 \\ 4 \end{pmatrix} = 8^i T_0$$

Разбиение тетраэдра:

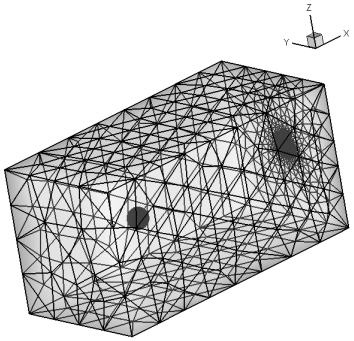
- 4 пирамиды
- 1 октаэдр

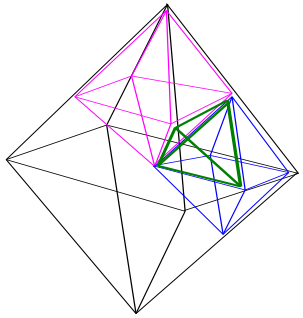


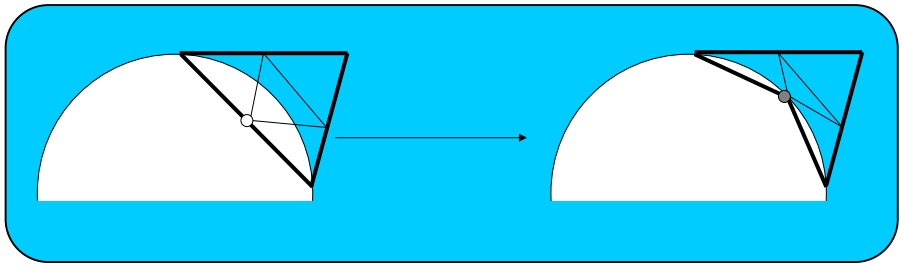
Разбиение октаэдра:

- 8 пирамид
- 6 октаэдров

Тетраэдральная сетка







70 300 -> 34 422 954 узлов

401 819 * 8^3 -> 205 731 328 тетраэдров

Бинарный формат без сжатия - 4.1 Гбайт

500 микродоменов, 44 файла, со сжатием gzip

(словарного сжатия Зива-Лемпела) - 1.8 Гбайт

Время расчета шага на 44 процессорах Xeon 3,06 Ghz

- 23 секунды

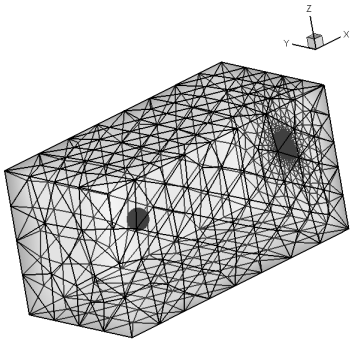
Время чтения данных – 20 секунд

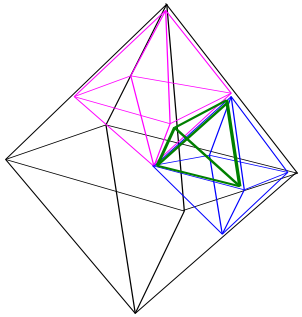
Зависимость объема хранимых данных от числа микродоменов

Число микродоменов	1	50	1000	1500	2000	2500	3000
Размер описания (МБ)	124	127	145	152	158	163	168

38 350 -> 2 356 196 узлов

219 034 * 8^2 -> 14 018 176 тетраэдров

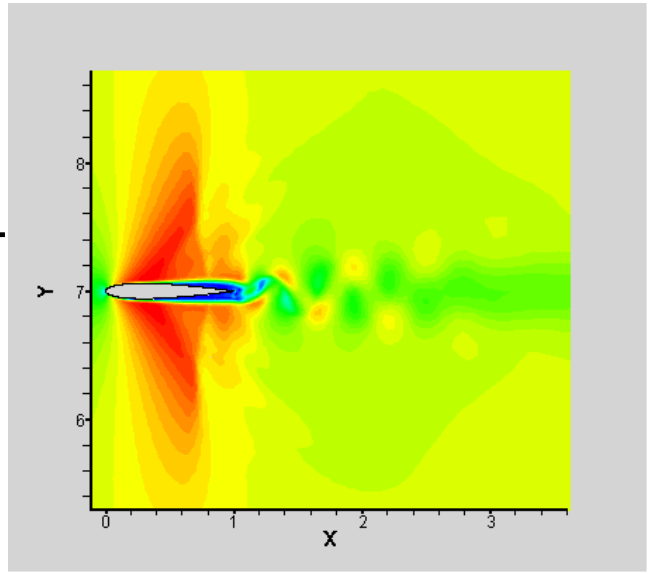


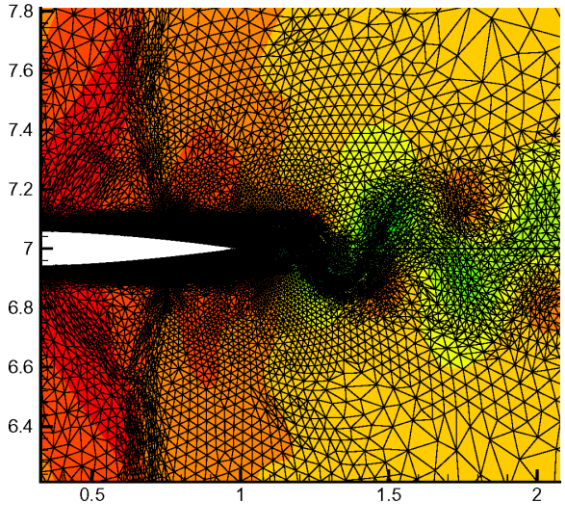


На 35% больше чем 124

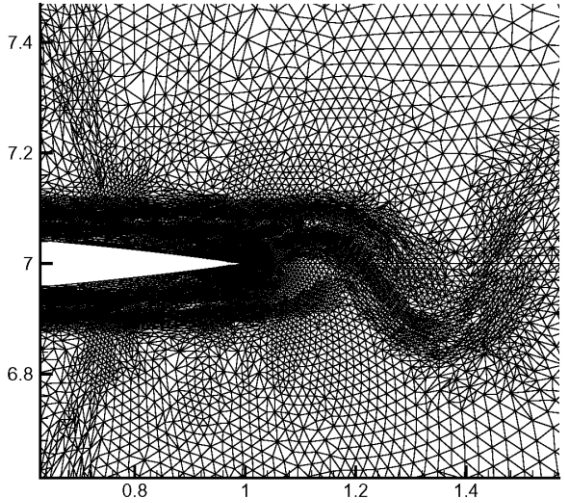
Использование адаптивной сетки

Обтекание профиля NACA001
(M=0.85, Re=10^4)
под нулевым углом атаки:





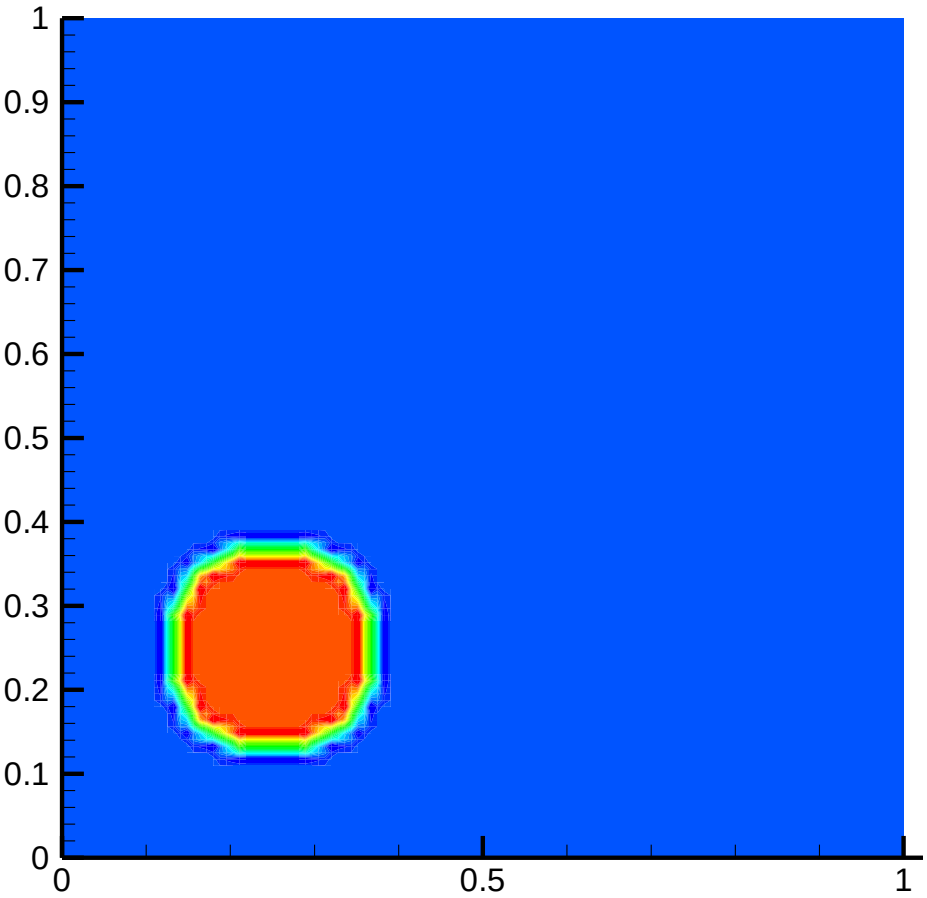
Поле продольной скорости

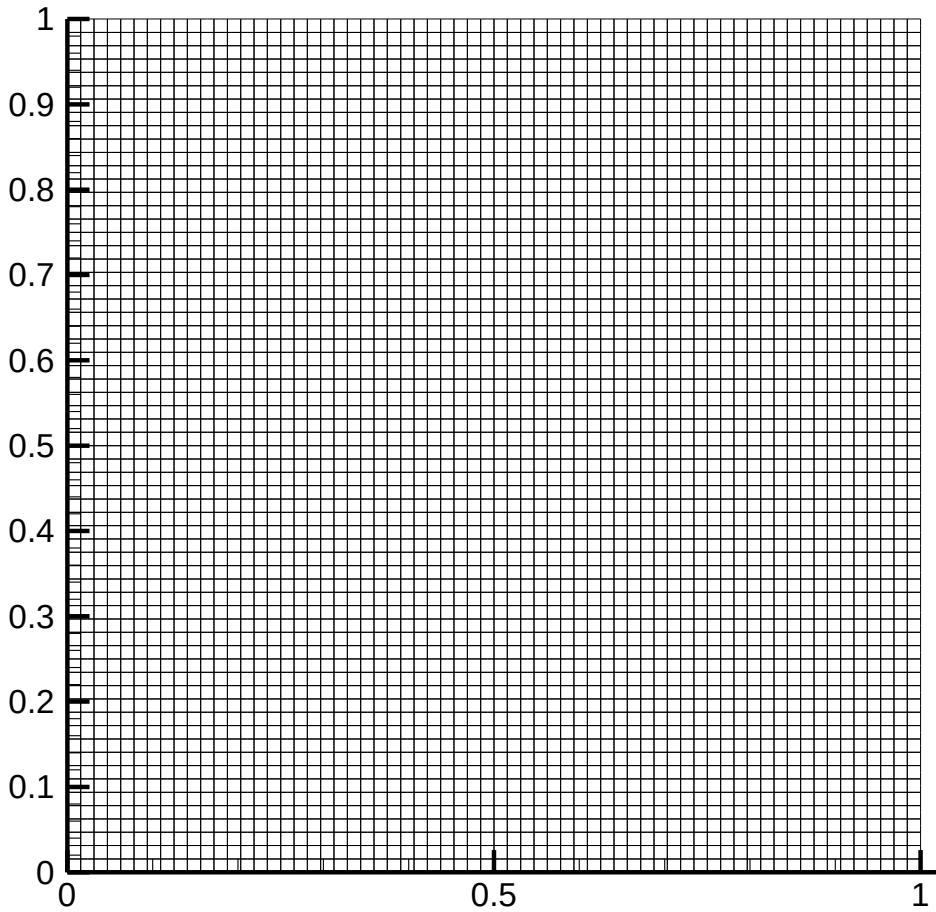


Фрагмент сетки

Равномерная сетка

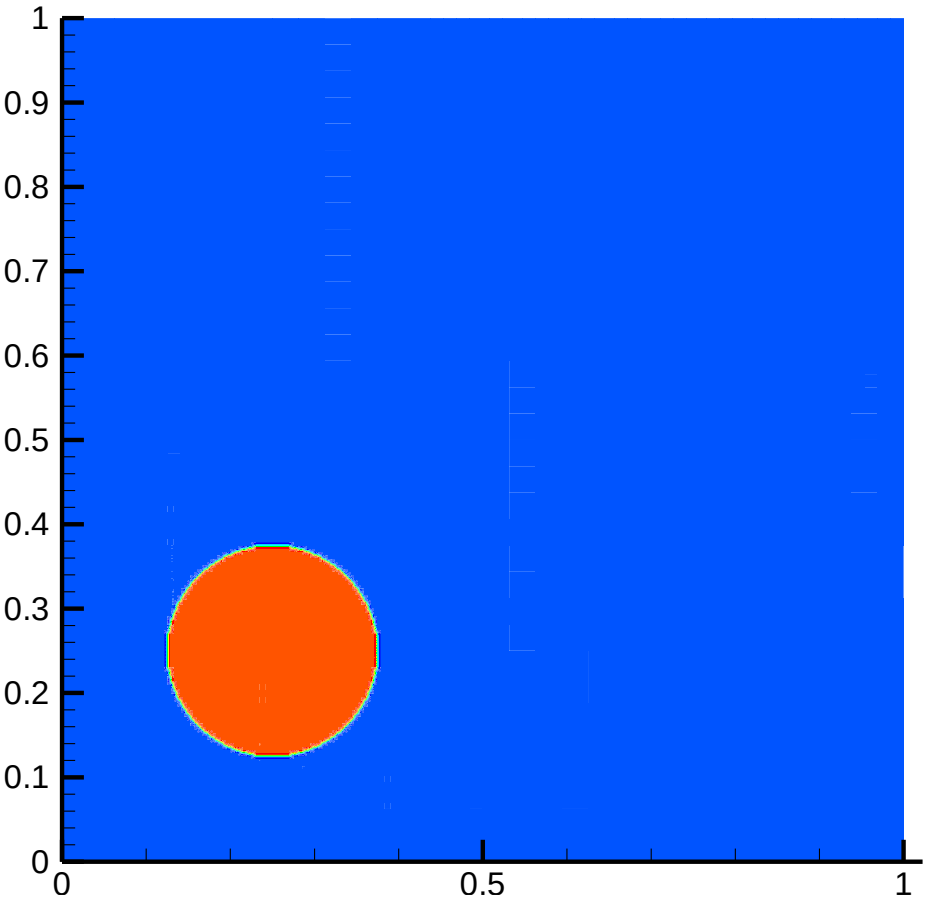
Слева – ??*круглое*?? пятно примеси

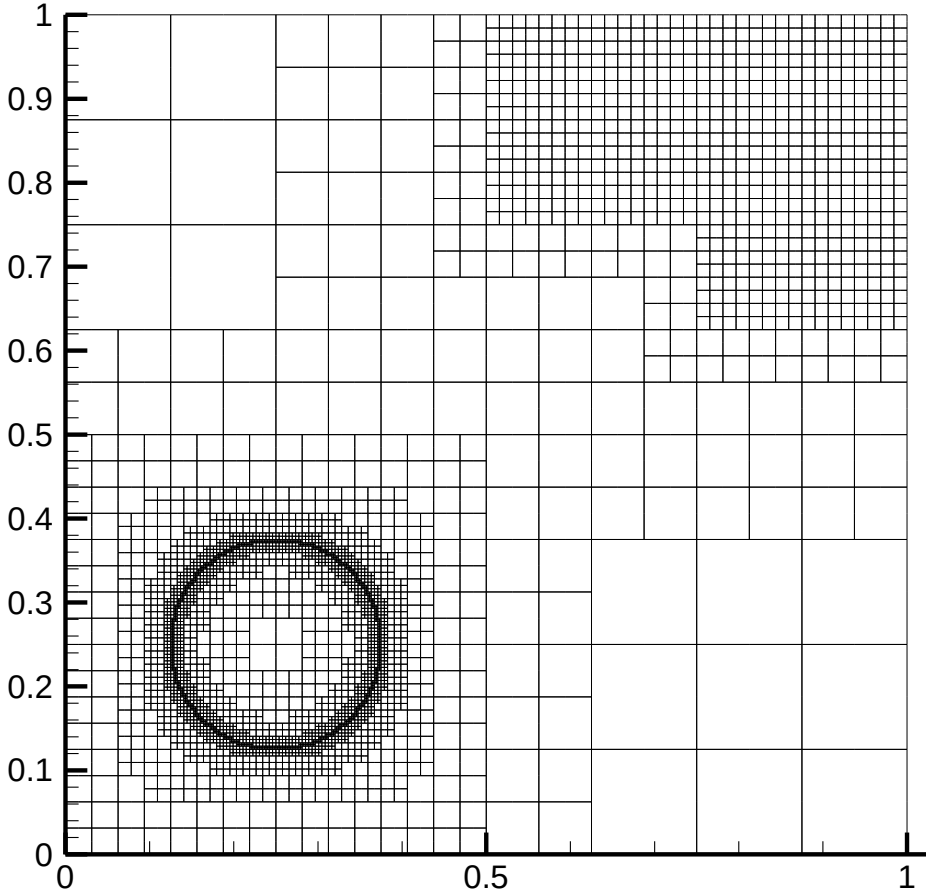




Адаптивная сетка

Слева – ***круглое*** пятно примеси

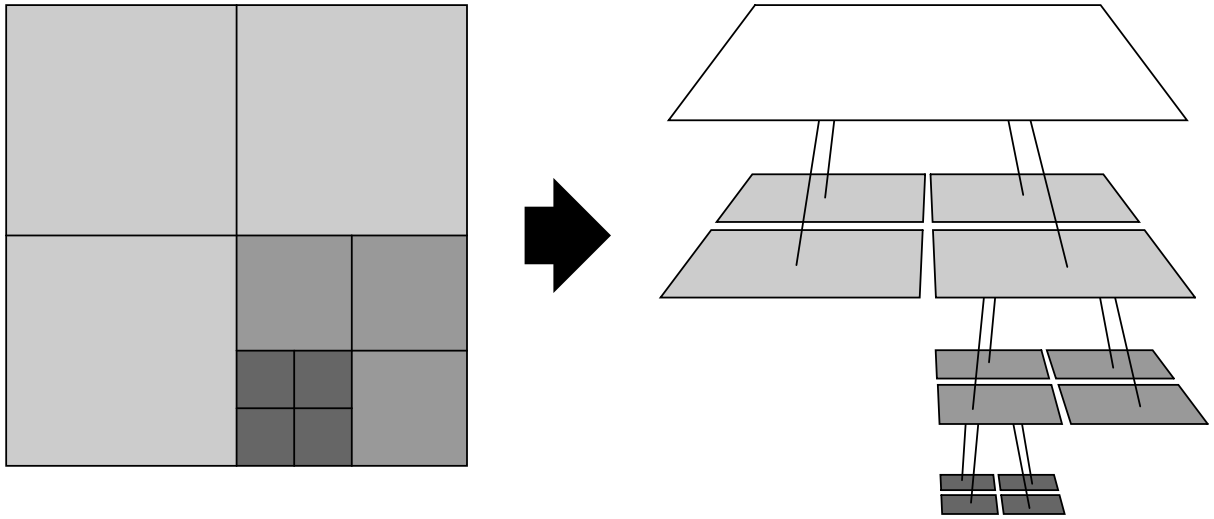




Адаптивные декартовы сетки

- Вначале сетка состоит из одной прямоугольной ячейки
- Каждая ячейка может быть разделена на четыре ячейки одинакового размера
- Если ячейки когда-то составляли одну ячейку, то они могут быть объединены обратно
- Каждая ячейка хранит величину, описывающую среднее значение неизвестной функции в пределах ячейки (метод конечных объёмов)

При данных предположениях сетку удобно хранить в виде четверичного дерева:

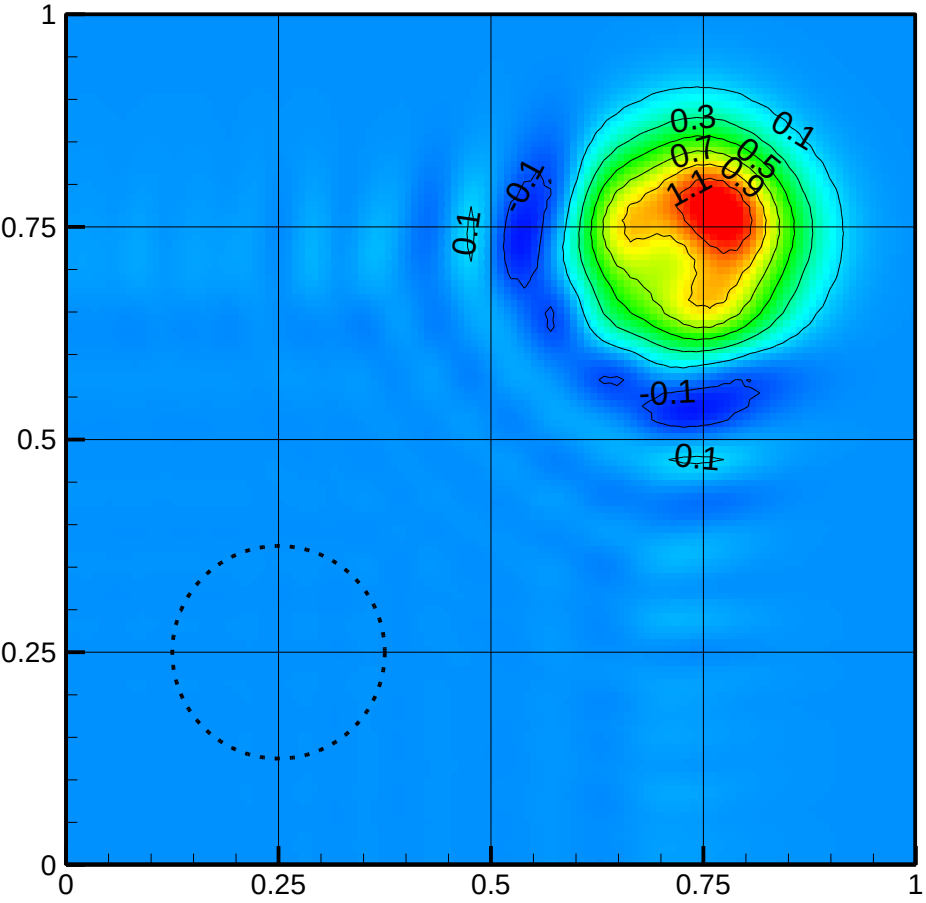


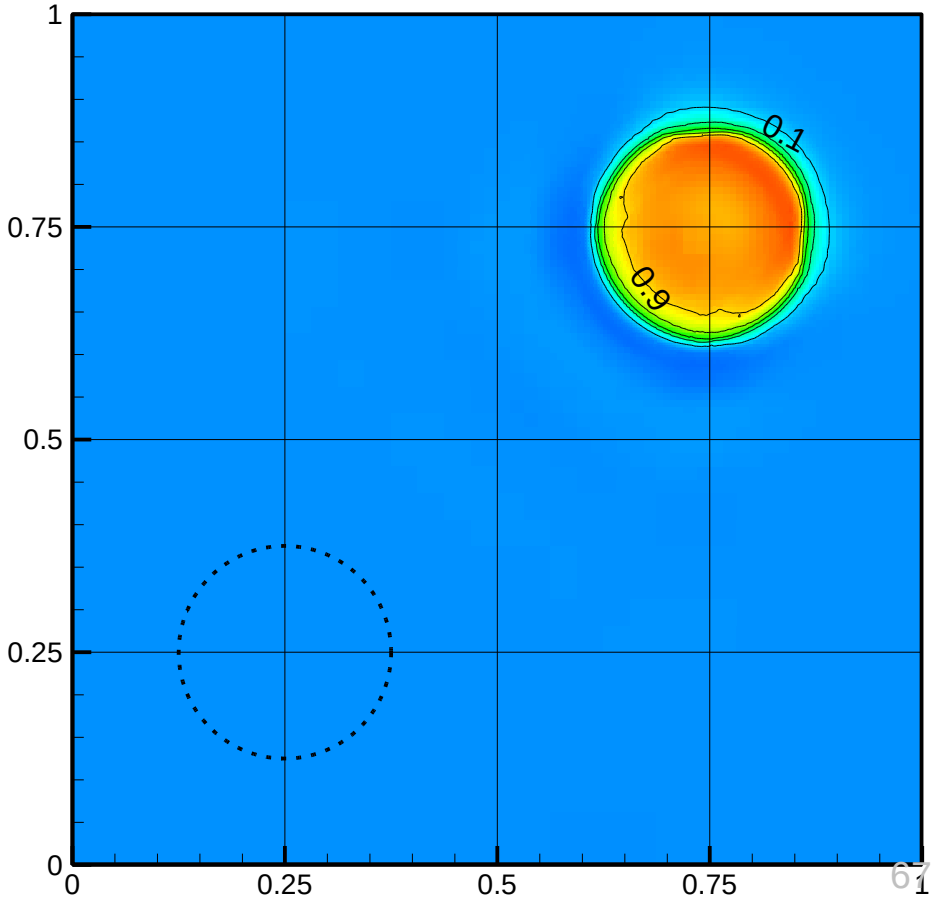
Дополнительные ограничения на размеры ячеек:

- Задан максимально допустимый размер ячеек
- Задан минимально допустимый размер ячеек
- Размеры соседних ячеек должны различаться не более, чем в 2 раза

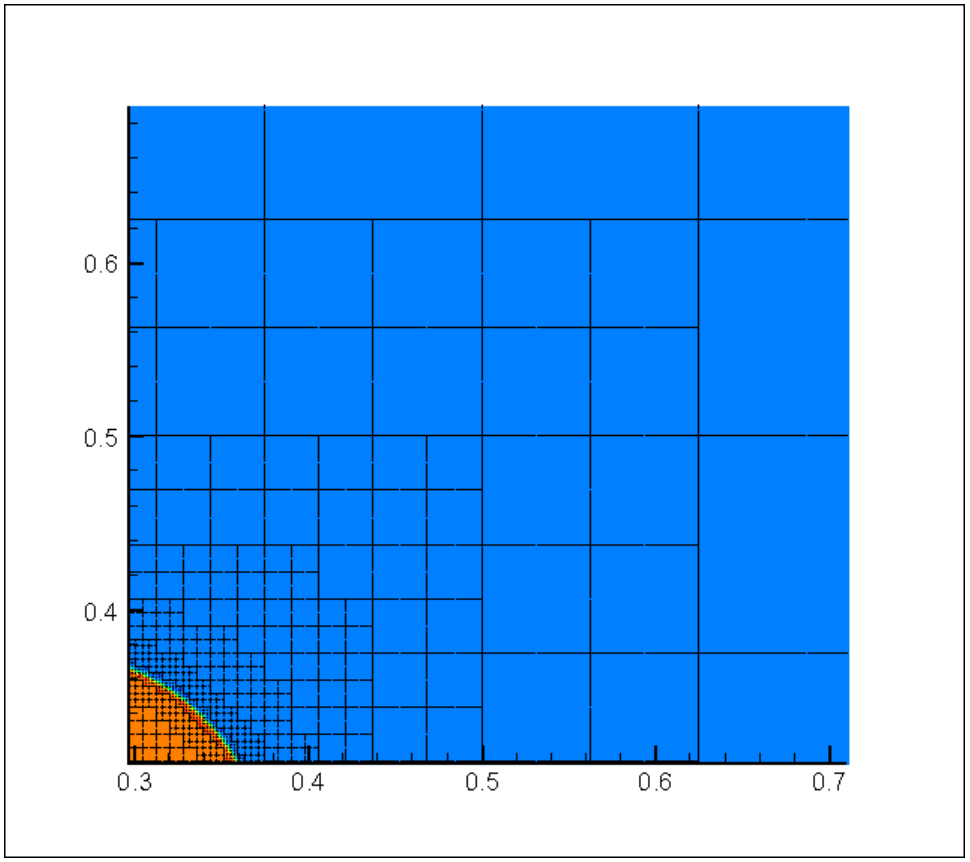
Сравнение с равномерной сеткой

На рисунках показаны результаты решения простейшей задачи переноса на равномерной (слева) и адаптивной (справа) сетках с одинаковым числом ячеек (4096 штук). Скорость переноса направлена под углом 45° к линиям сетки; начальное условие показано пунктиром

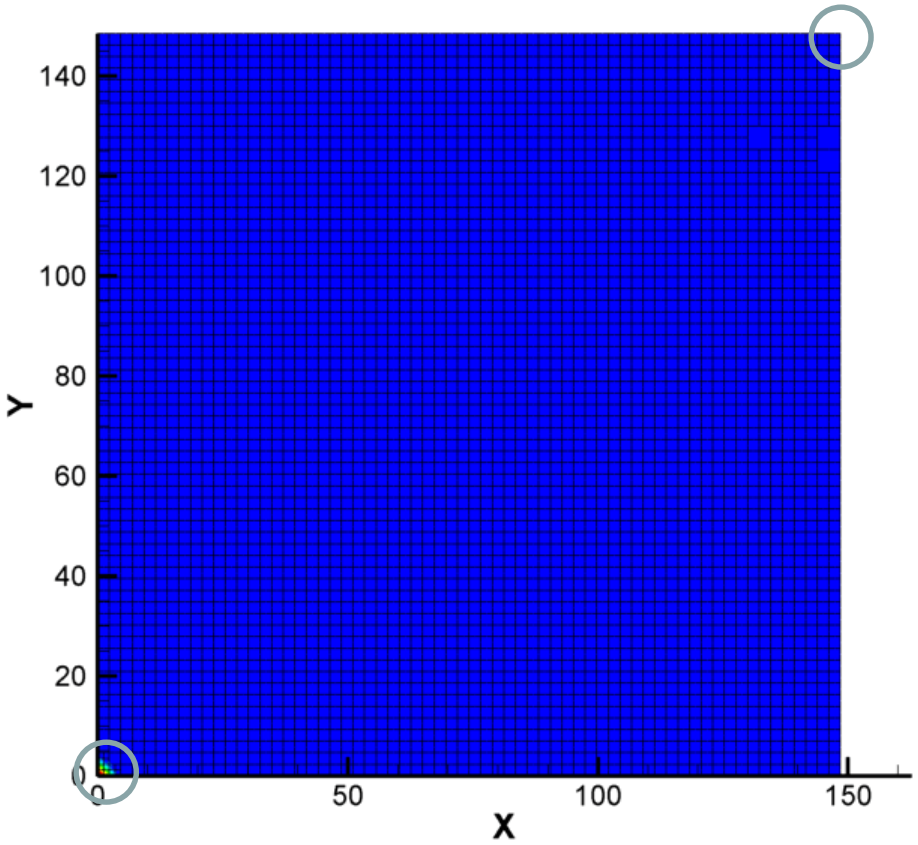




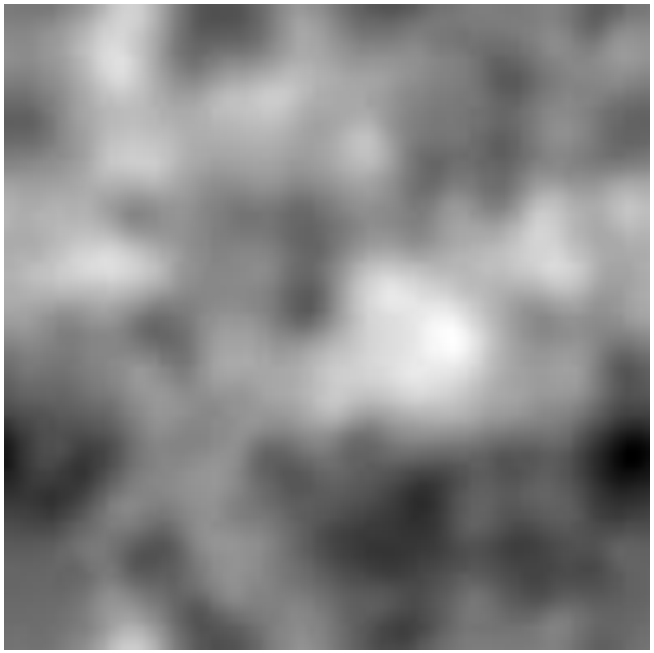
Адаптивная сетка



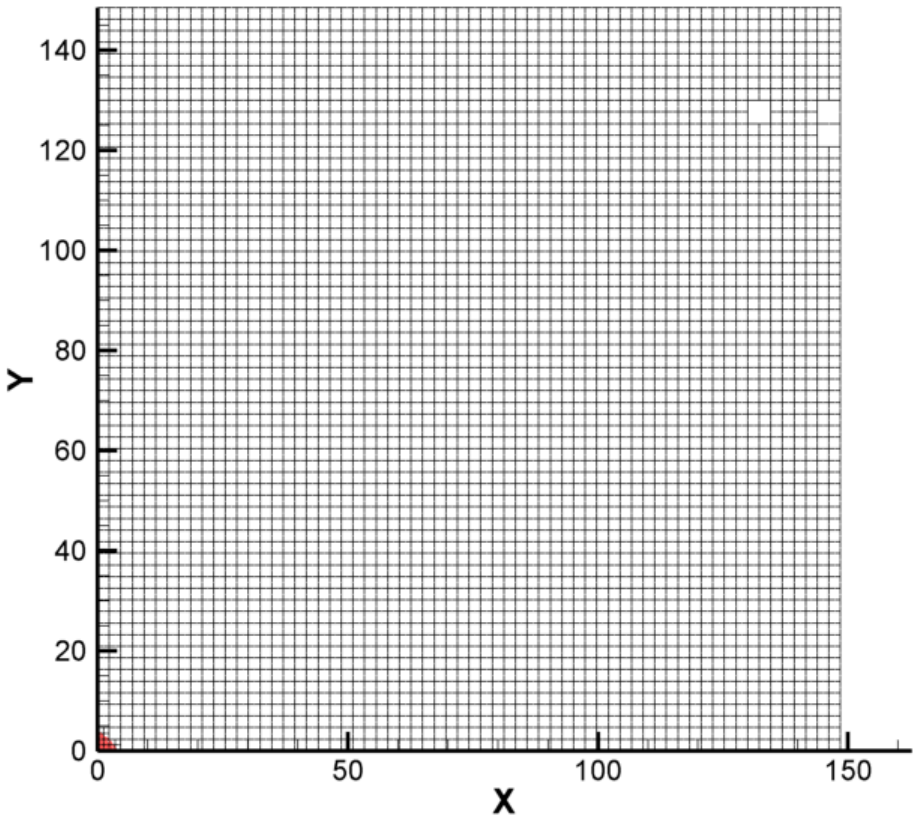
Решение двумерной задачи фильтрации нефтеводяной смеси в области с неоднородной проницаемостью



В юго-западном углу находится скважина, нагнетающая воду, в северо-восточном углу — добывающая скважина.
5-ти точечная схема
Поле проницаемости с разбросом значений на 4 порядка).

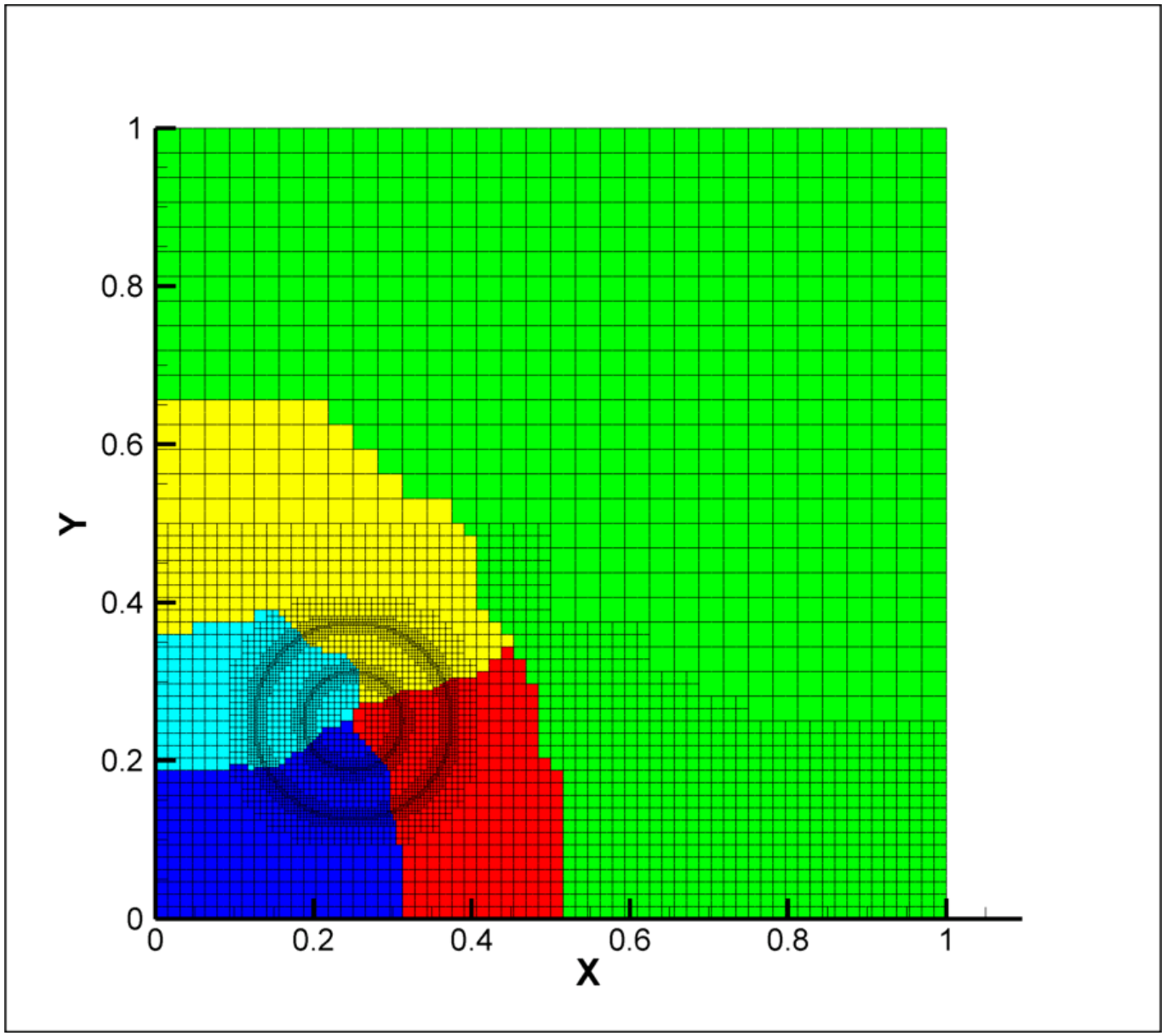


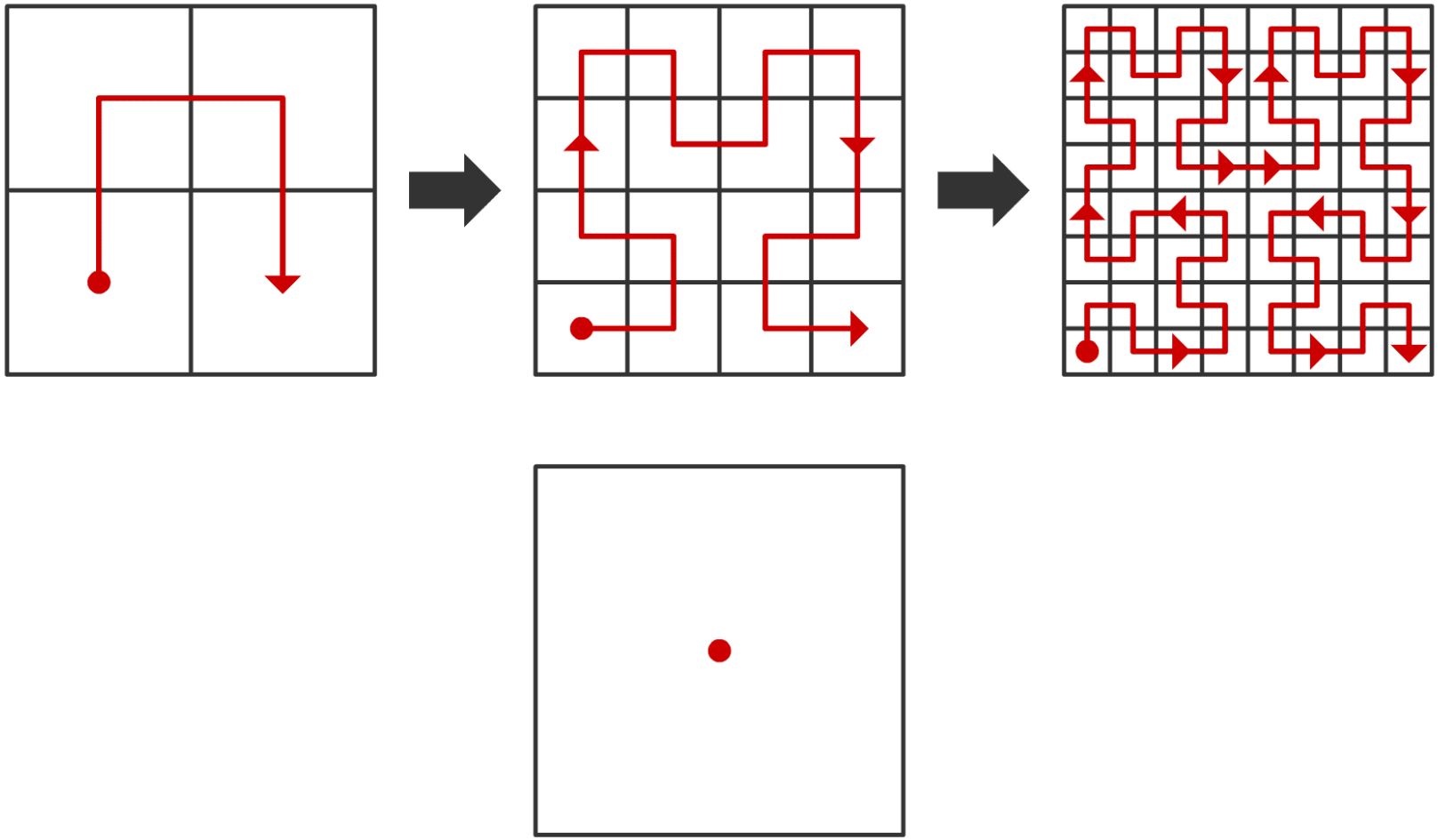
Решение двумерной задачи фильтрации нефтеводяной смеси в области с неоднородной проницаемостью



В юго-западном углу находится скважина, нагнетающая воду, в северо-восточном углу — добывающая скважина.
5-ти точечная схема
Поле проницаемости с разбросом значений на 4 порядка).

Декомпозиция пакетом Metis





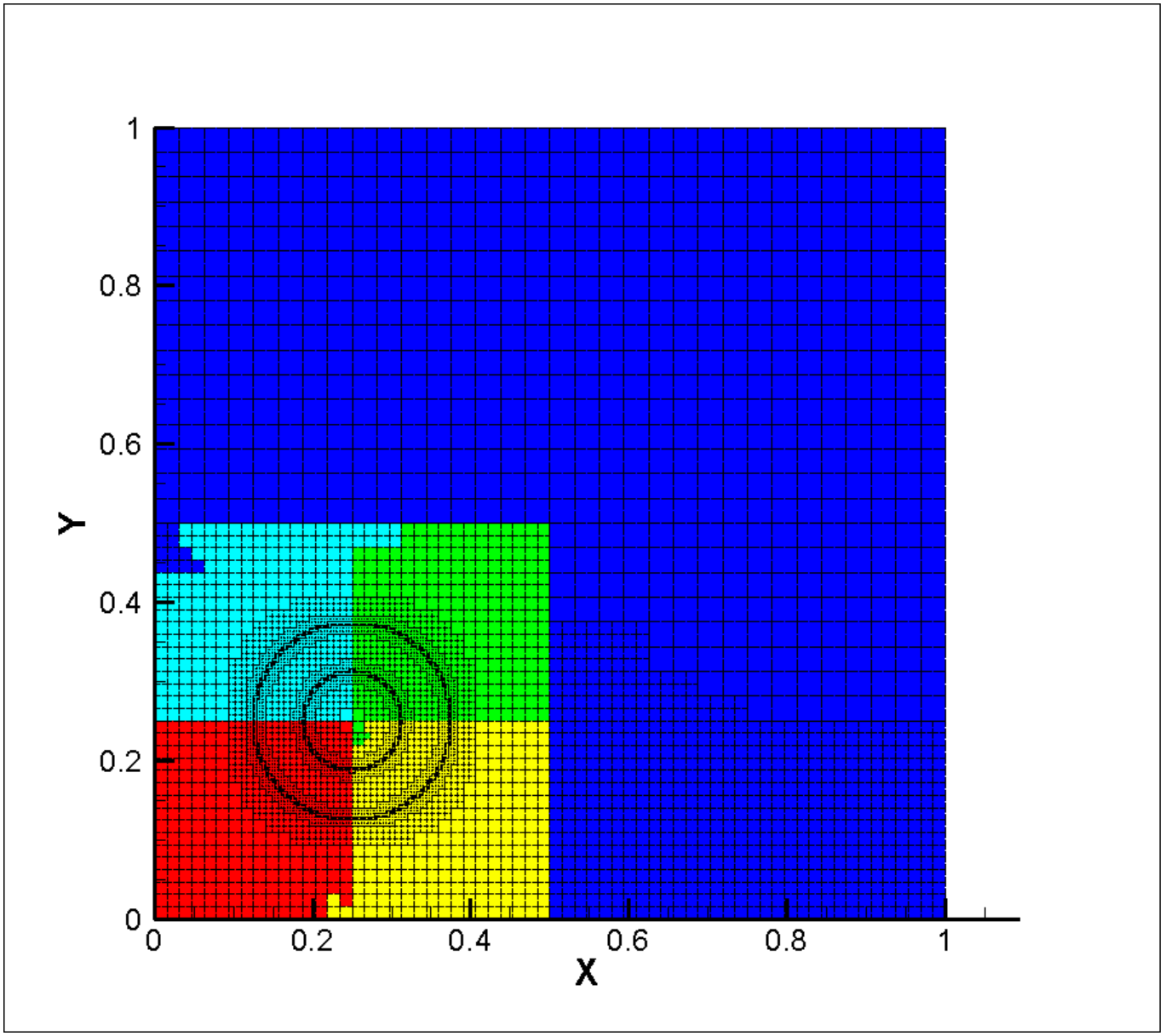
Hilbert-curve ordering

This ordering can be built by simple recursive procedure.

When mesh changes locally, Hilbert curve changes locally too.

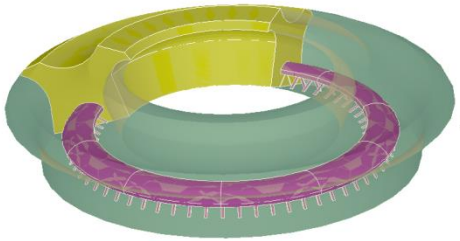
It cannot be used for parallel computations due to chain dependence of elements.

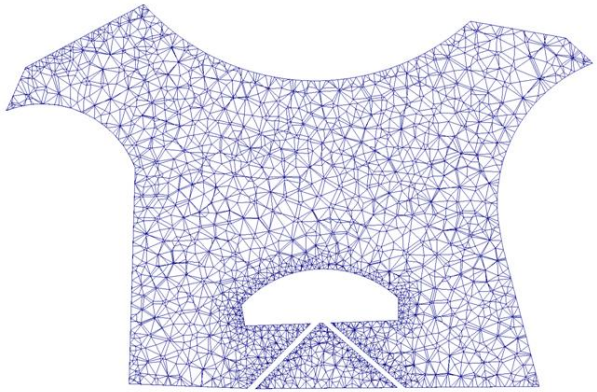
Гильберта

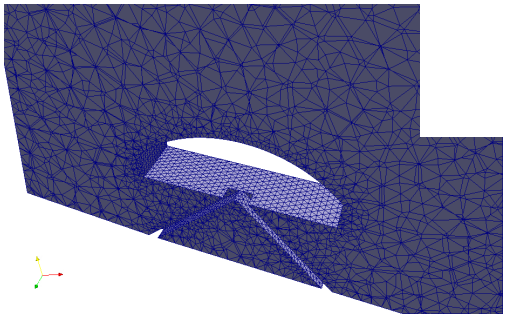


Расчетные сетки

Дивертор токамака (divertor)

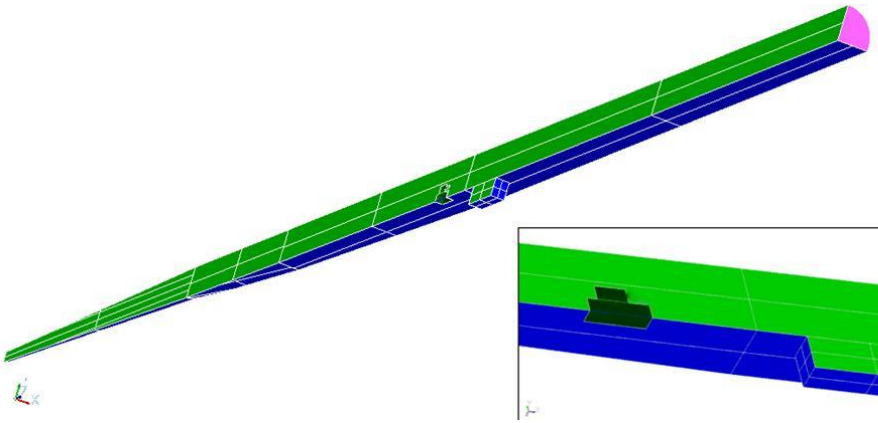






- Тетраэдральная сетка (3 миллиона ячеек)
- сгущение сетки вблизи мелких объектов
- 256 доменов

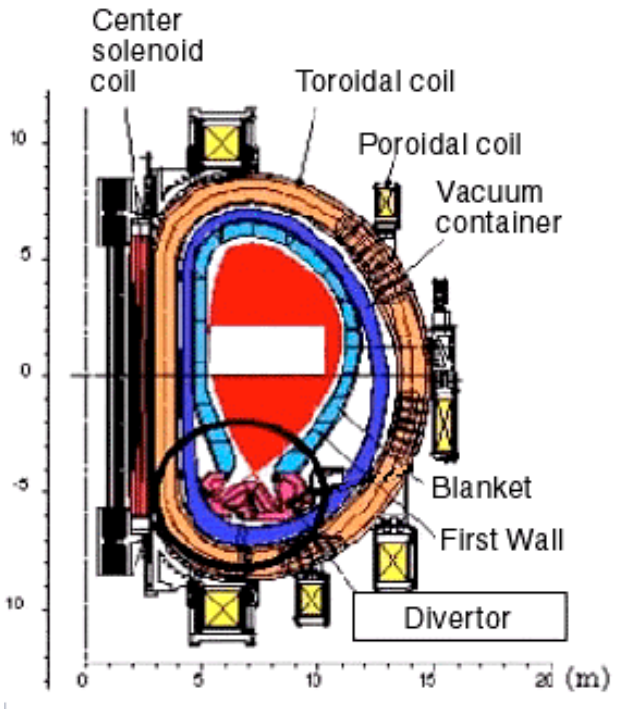
Ударная труба (tube)



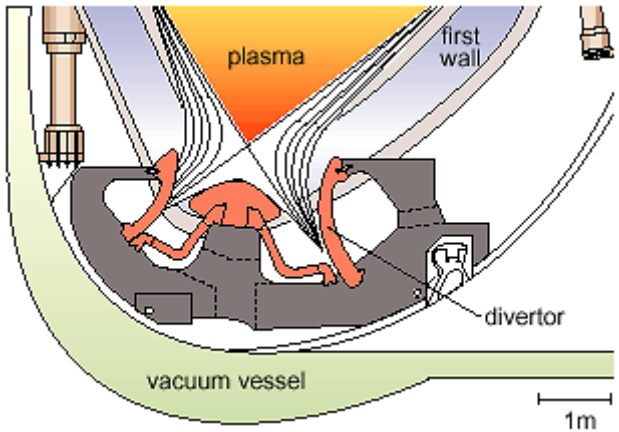
- Тетраэдральная сетка (более 25 миллионов ячеек)
- сгущение сетки вблизи мелких объектов
- 4096 доменов

The immediate task is the development of the model for turbulent heat and mass transfer in ITER divertor (MHD + turbulence + radiative transfer).

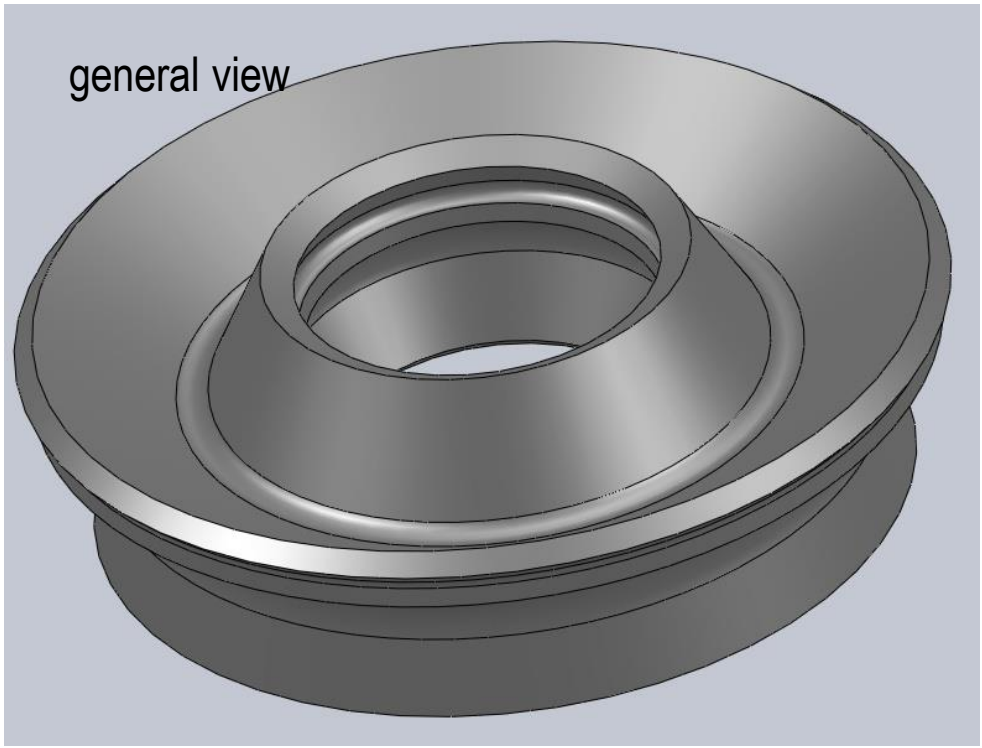
Cross-section of ITER

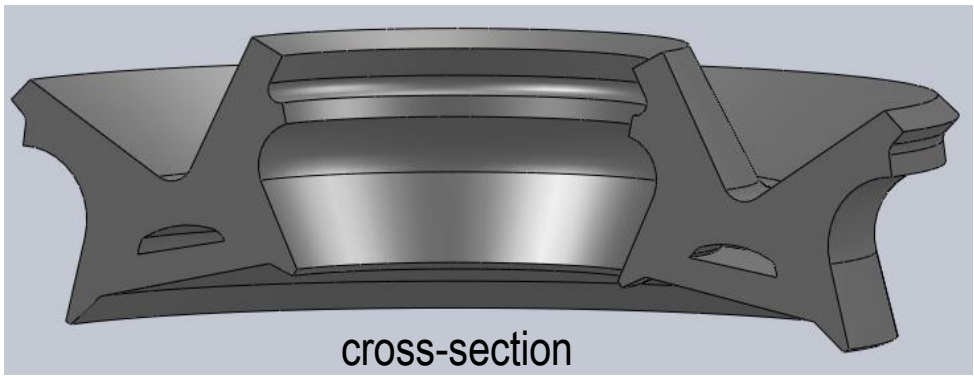


Magnified view of divertor area



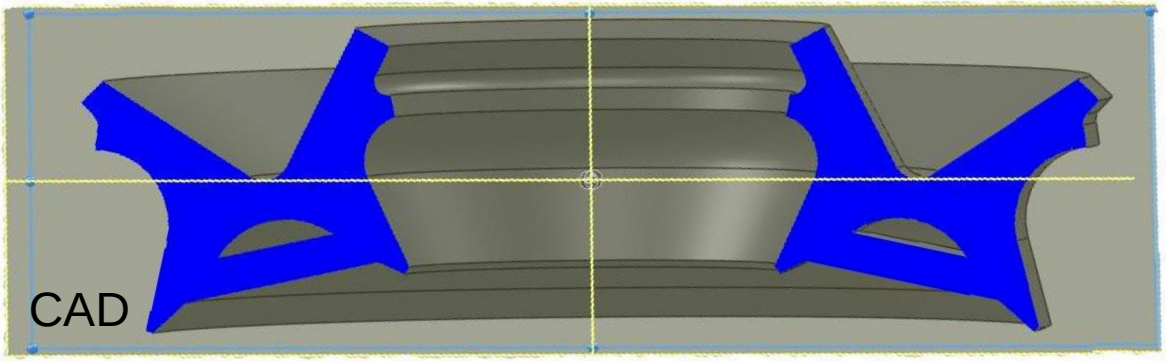
CAD images

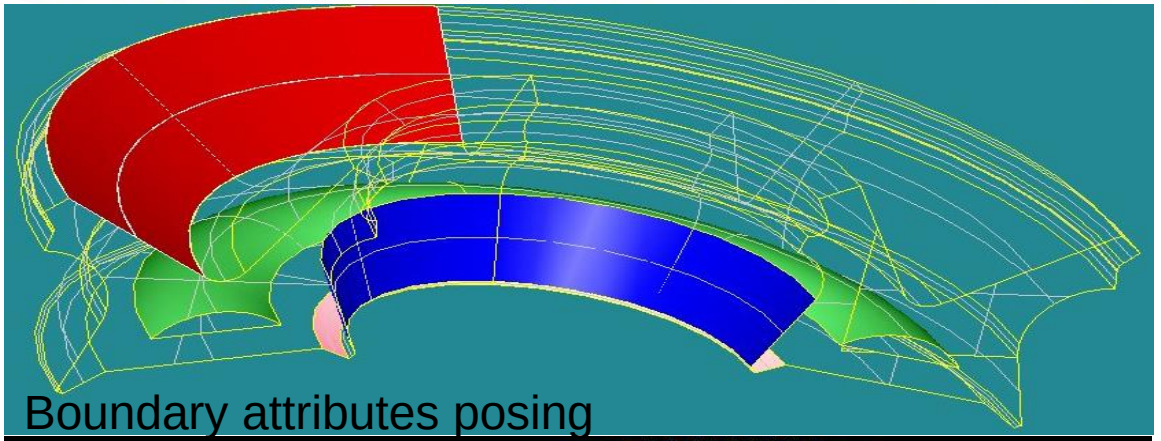


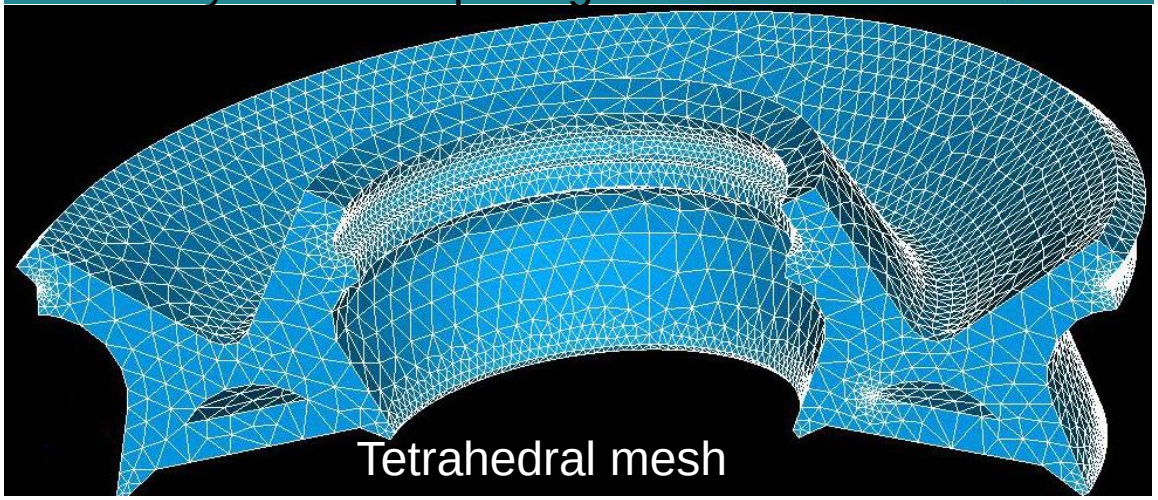


Turbulent
heat and mass transfer
in ITER divertor:

From CAD model
to computational mesh

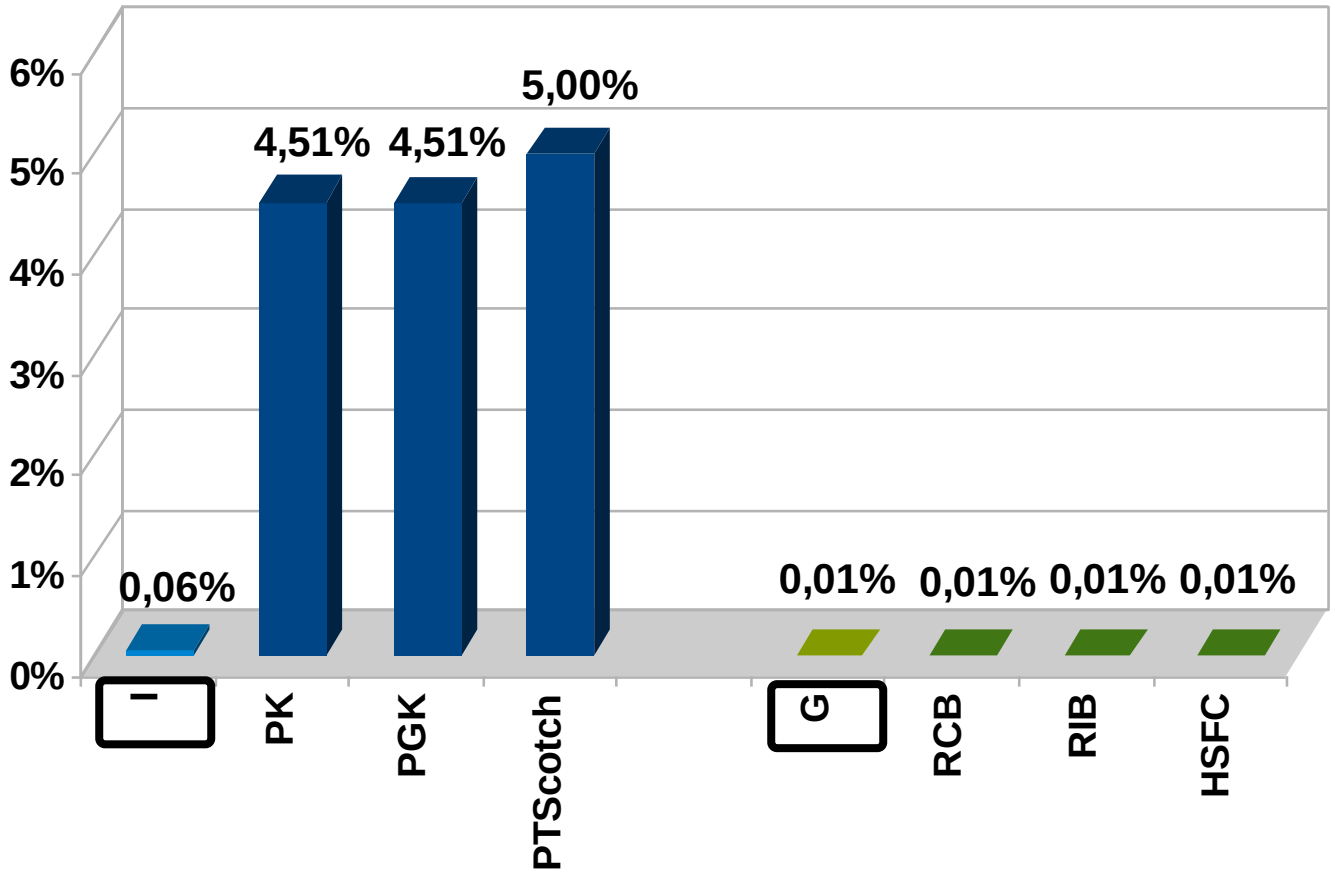




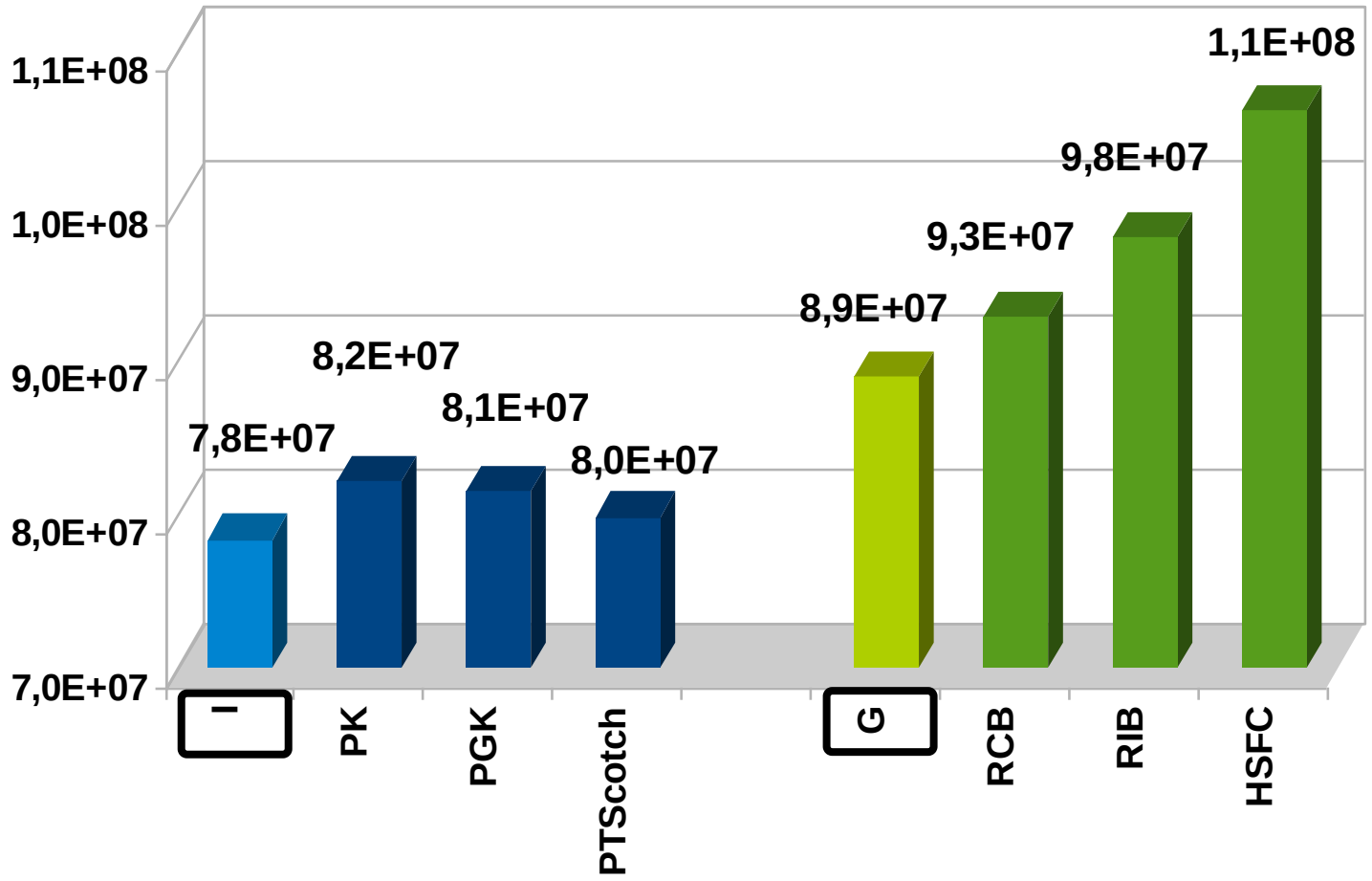


Initial tetrahedral mesh before refinement is shown.
The resulting mesh includes 10^8 cells and more.

Дисбаланс числа вершин в доменах: избыток вершин (boom)

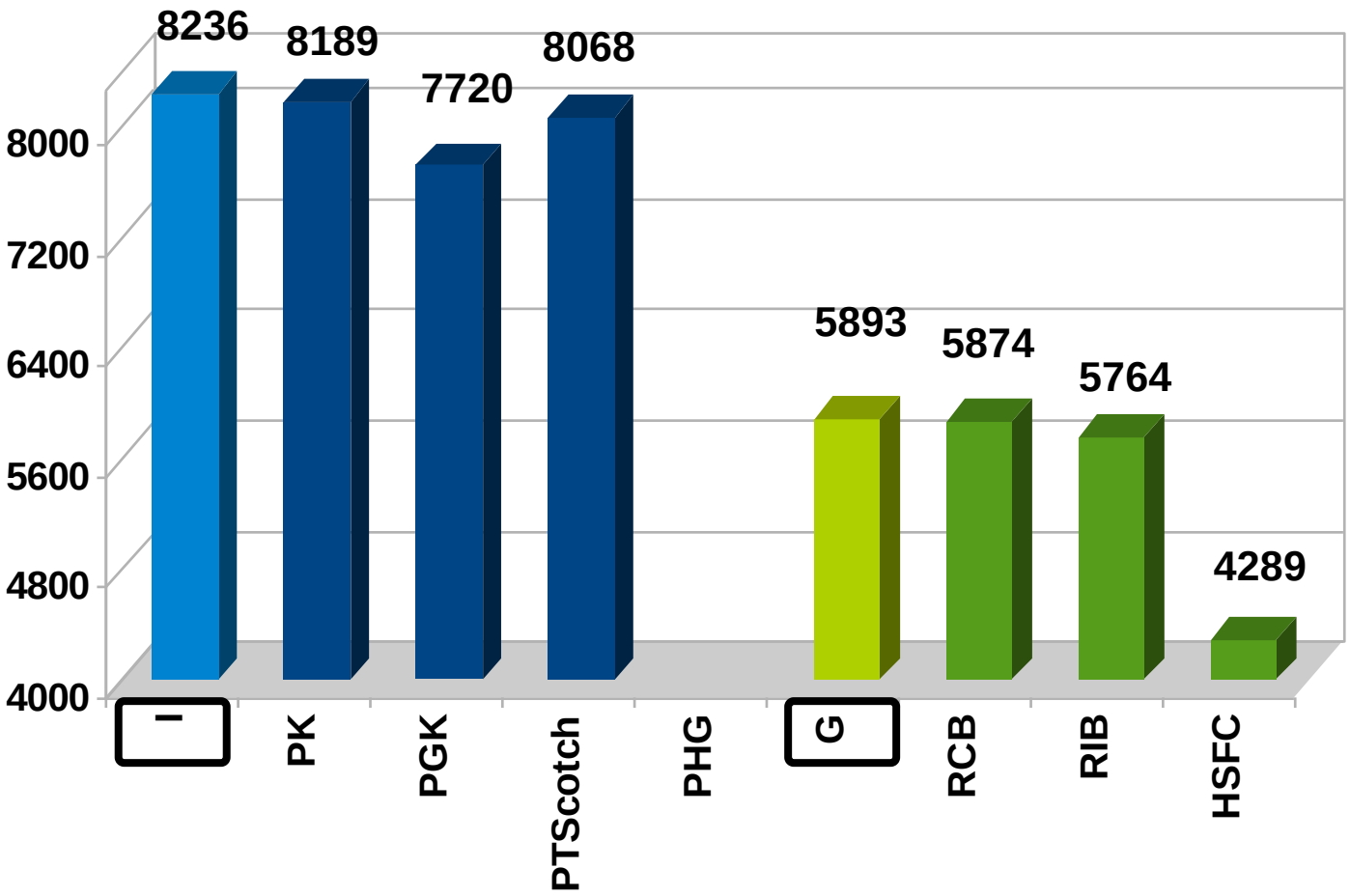


Число разрезанных ребер (boomL)



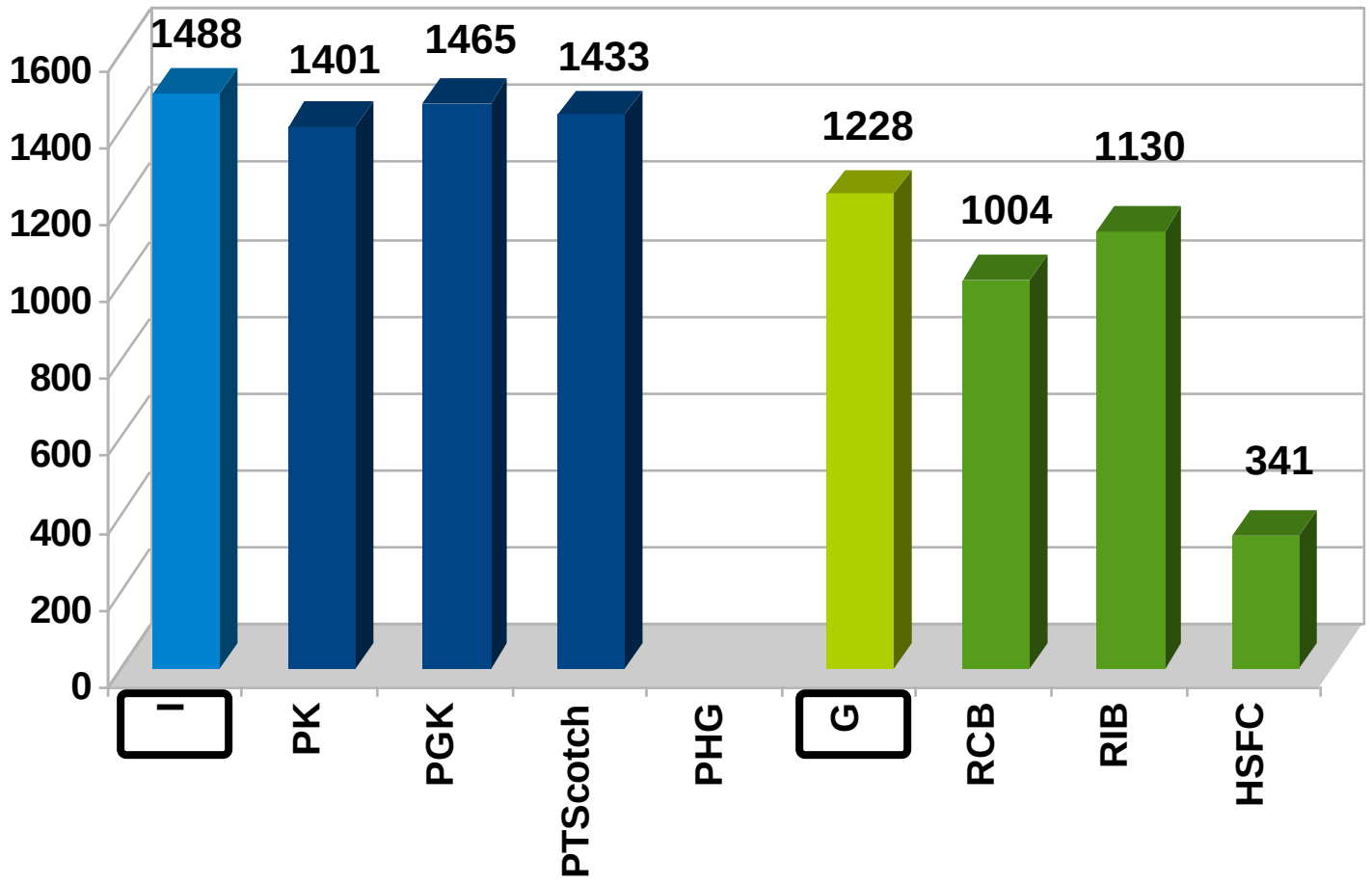
78

Число шагов по времени (divertor)



79

Число шагов по времени (tube)



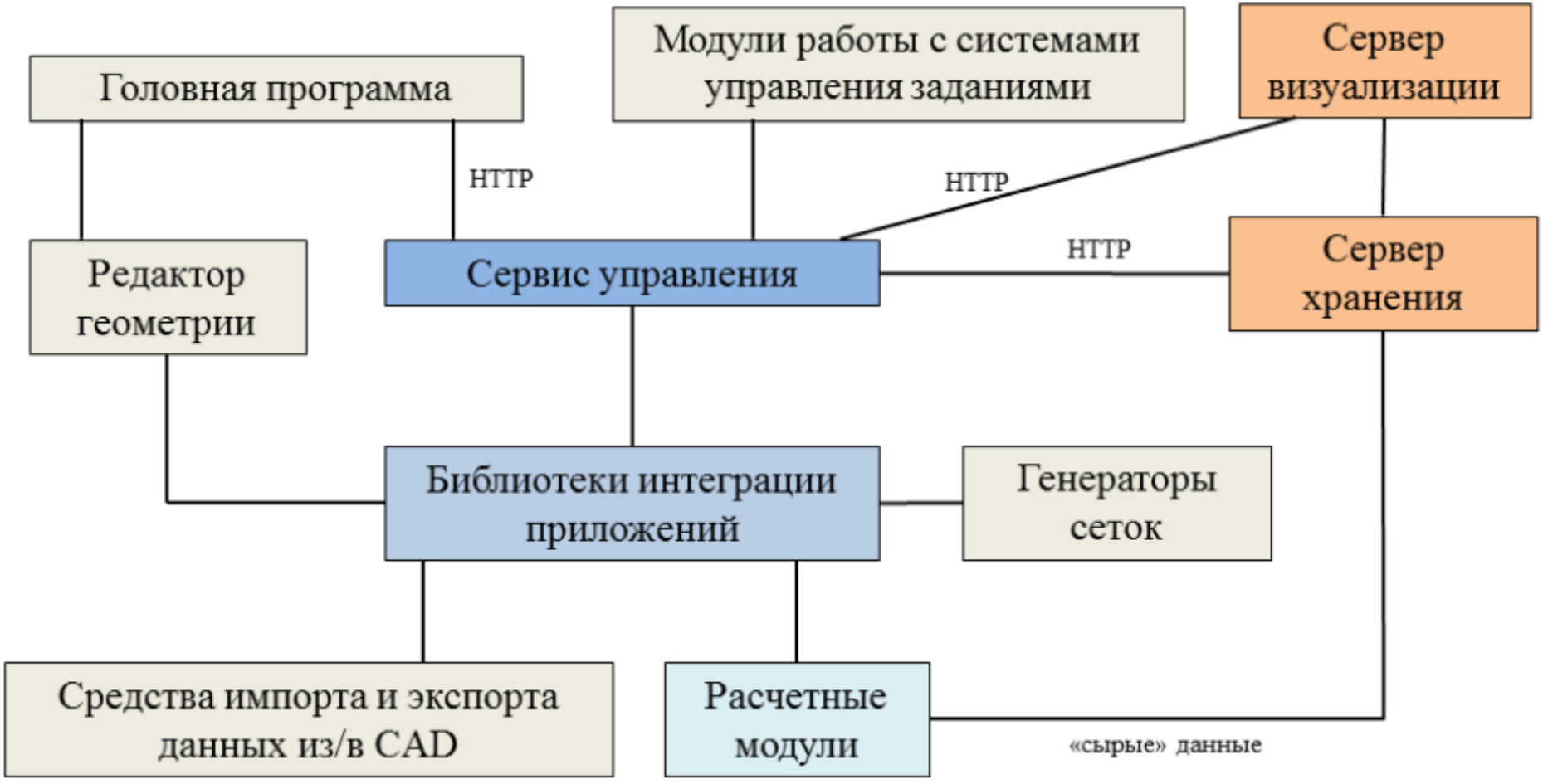
80

Декомпозиция графов – основной инструмент начального распределения данных и вычислительной нагрузки по процессорам

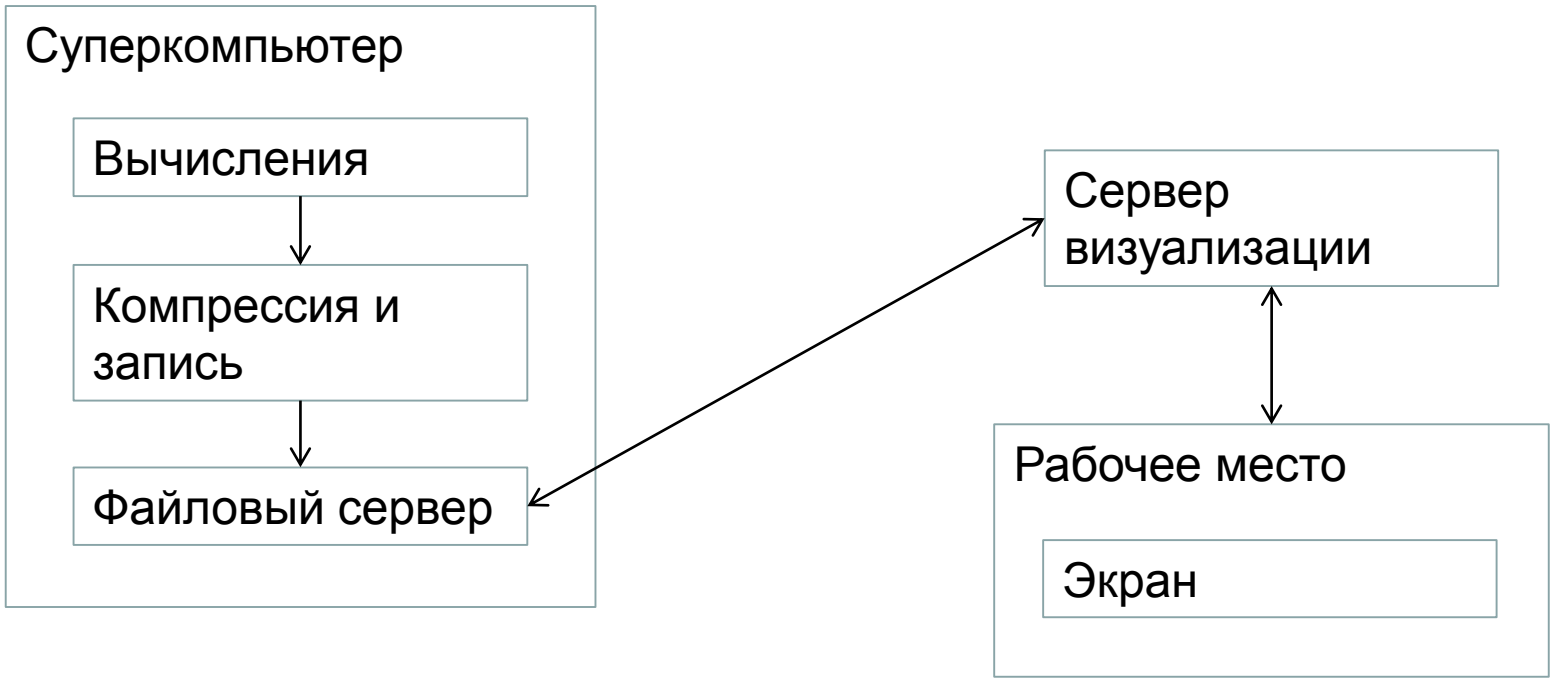
Точное решение соответствующей задачи невозможно, но эмпирические методы позволяют получать приемлемые результаты

Gimm nano

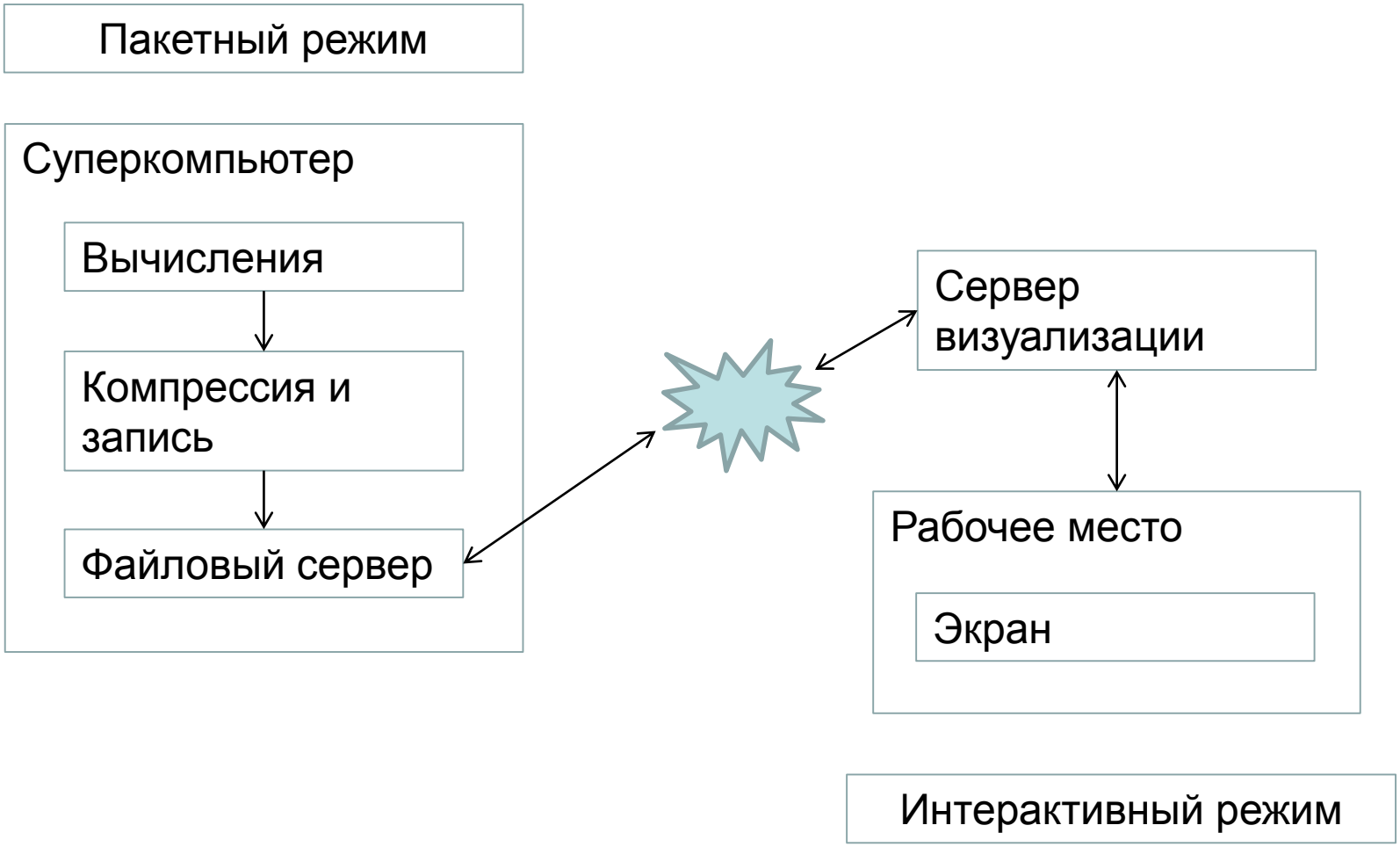
Укрупнённая компонентная модель комплекса



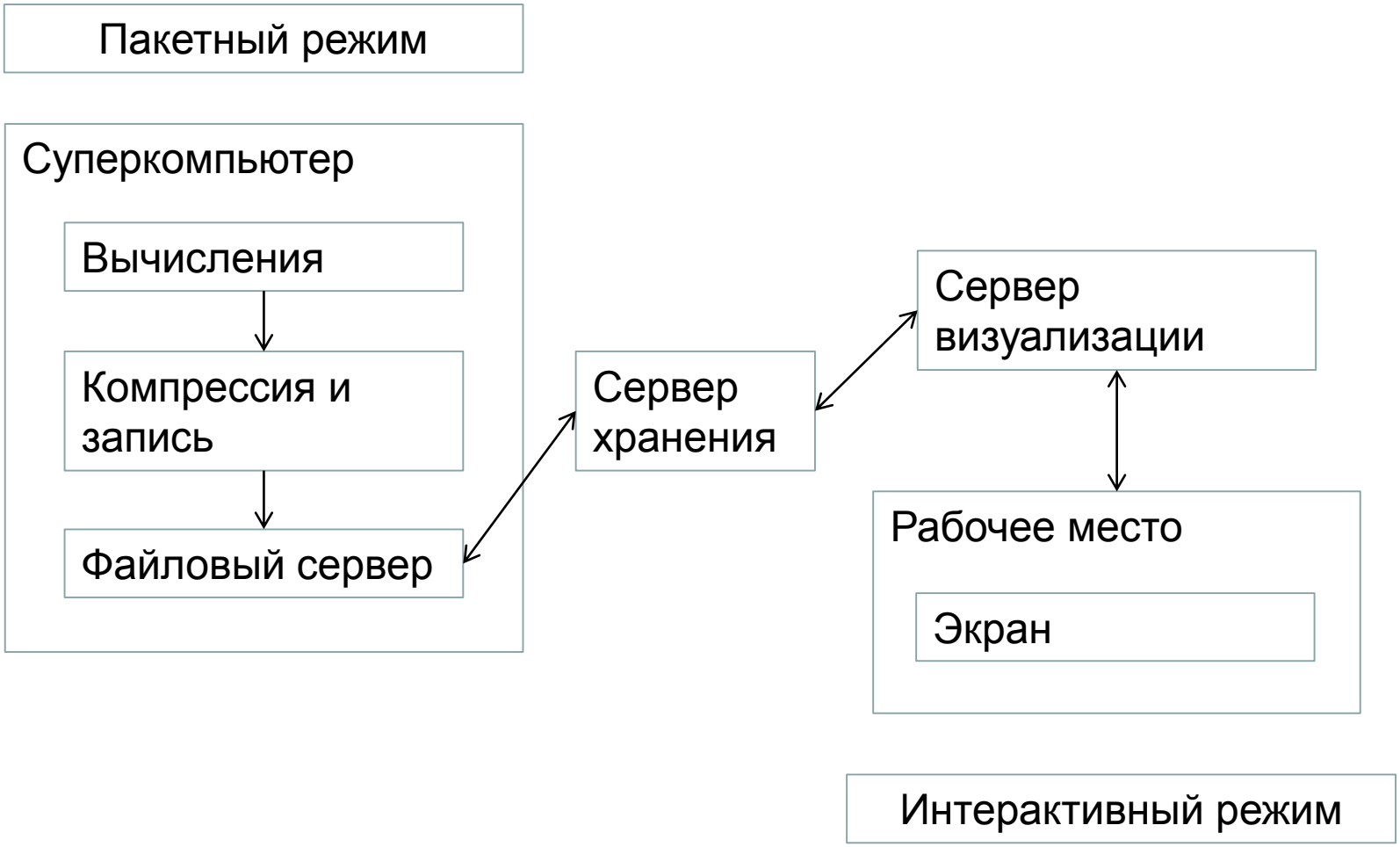
Обработка результатов



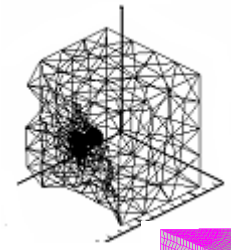
Обработка результатов

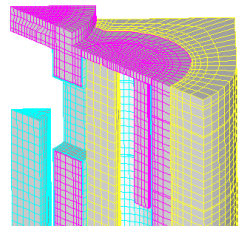


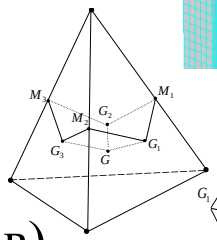
Обработка результатов

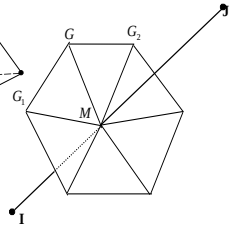


- **Сетки**: нерегулярная (неструктурированные) сетки смешанных типов элементов, блочные и блочно-структурированные сетки
- **Параллельная реализация** для систем с распределенной и общей памятью, гибридных систем
- **Автоматизация** создания параллельных программ для гибридных систем
  - DVM – общая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





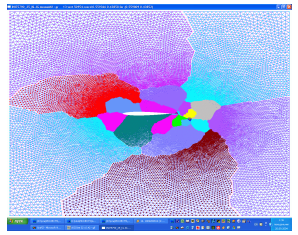




$$\iint_{\partial\Omega_i}\left(\mathbf{F}_1(\mathbf{Q})n_x+\mathbf{F}_2(\mathbf{Q})n_y+\mathbf{F}_3(\mathbf{Q})n_z\right)d\sigma=\iint_{\partial\Omega_i}\mathbf{F}\,d\sigma=\sum_{j\in\Omega_i}\Phi_{ij}\left(\mathbf{F}_i,\mathbf{F}_j,\mathbf{n}_{ij}\right)$$

Основные характеристики:

- Реалистичная 3D геометрия
- Нерегулярные сетки большого объема (10^8 - 10^9 - … 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



Контакты

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

web: http://lira.imamod.ru