

*Параллельные вычисления  
(часть 5)*

*Рациональная декомпозиция  
сеточных графов*

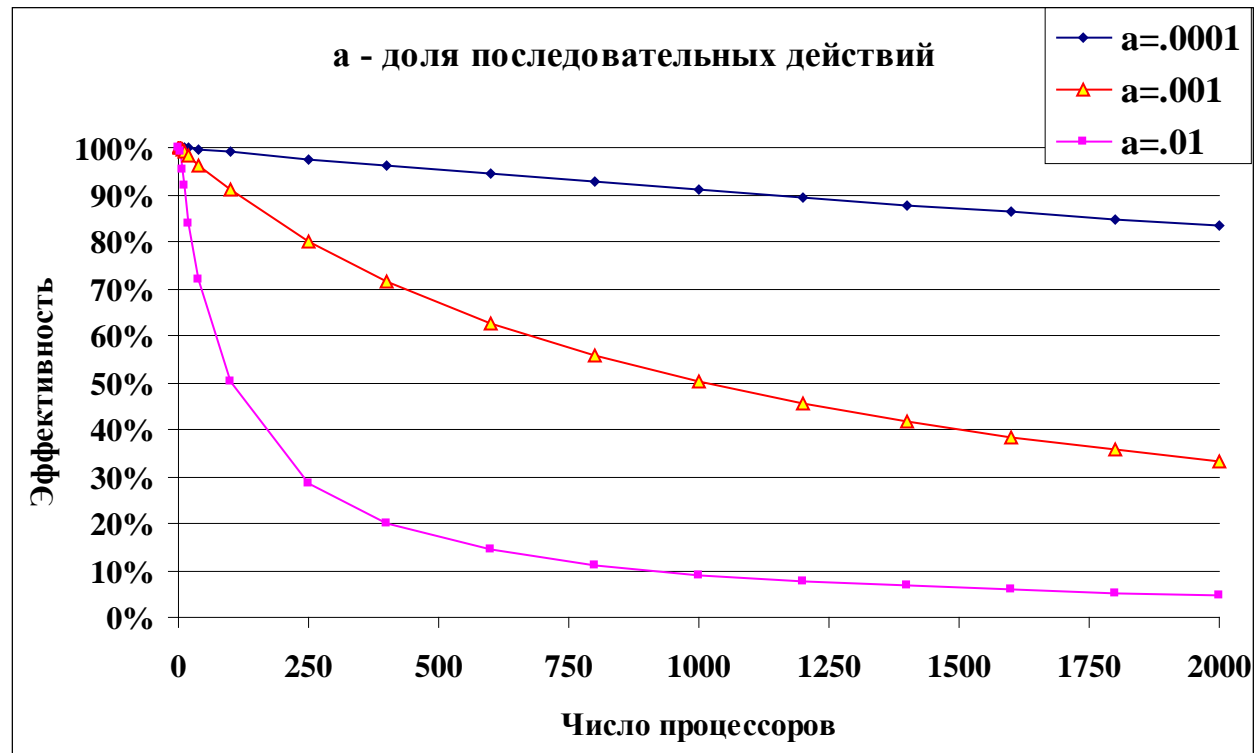
Якобовский Михаил Владимирович

# Ограничения

- Закон Амдаля

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

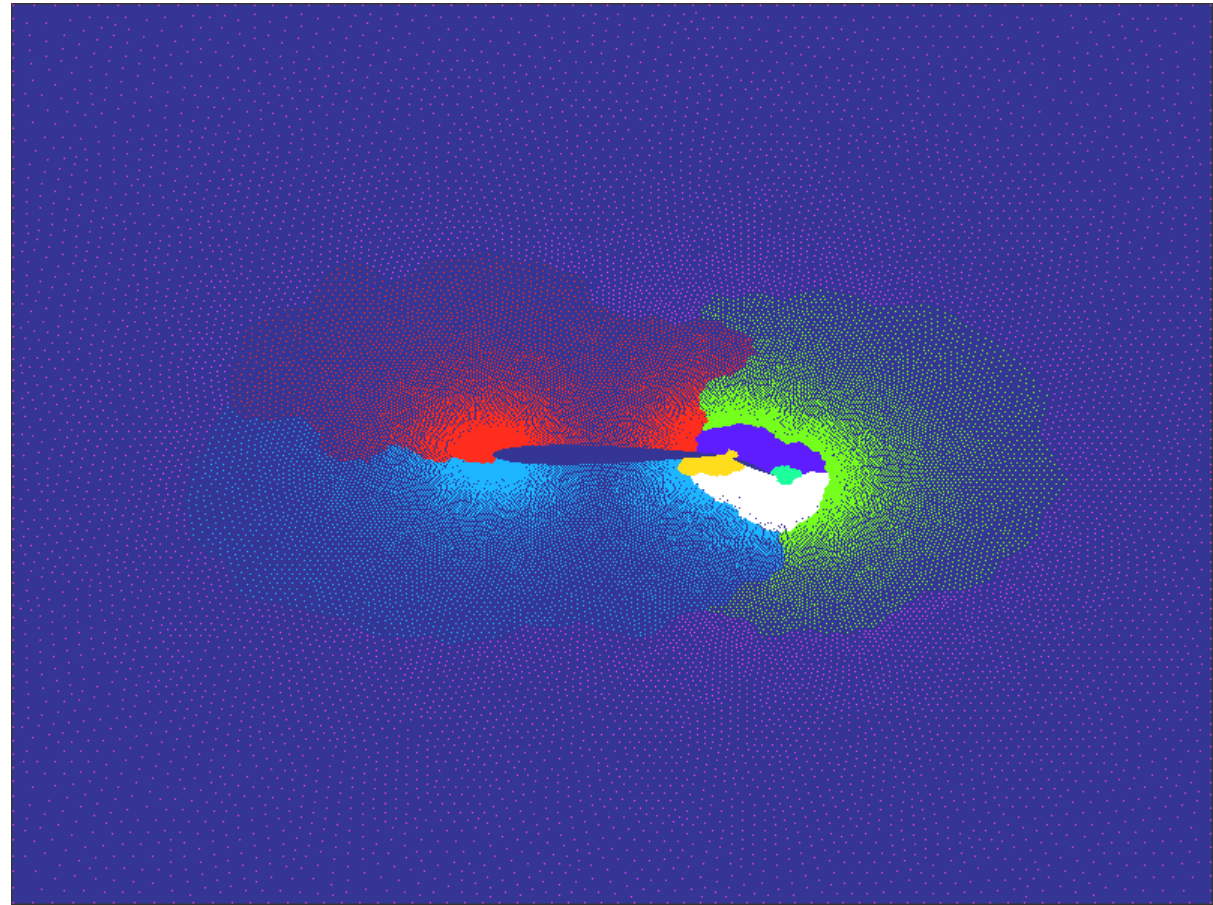
$$E(p) = \frac{1}{1 + a(p-1)}$$

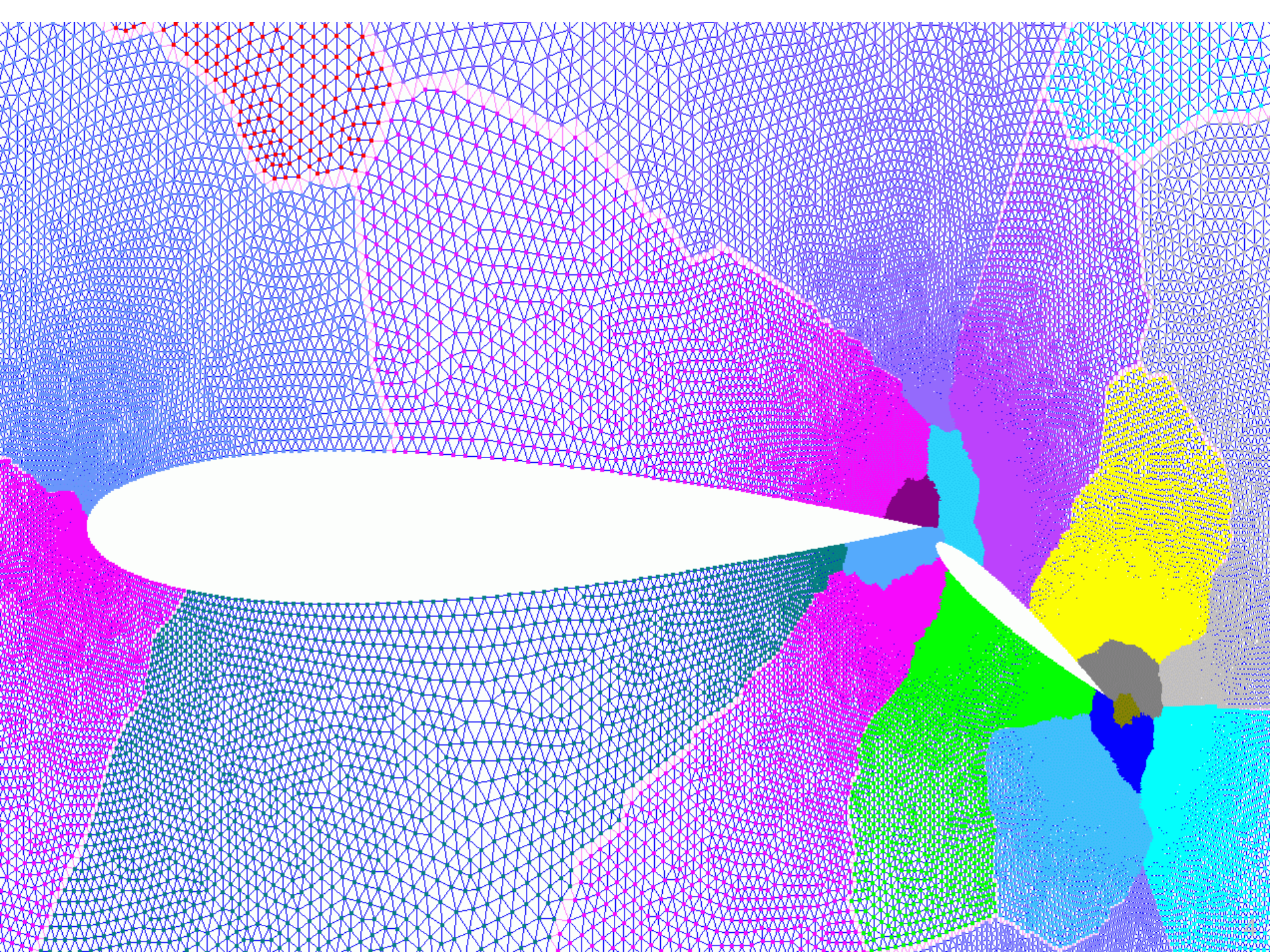


- Пакетный режим исполнения и отладки приложений
- Процедуры авторизованного доступа к удаленным системам
- Высокая динамика изменения конфигурации суперкомпьютеров
- Несоизмеримость ресурсов рабочей станции пользователя и суперкомпьютера

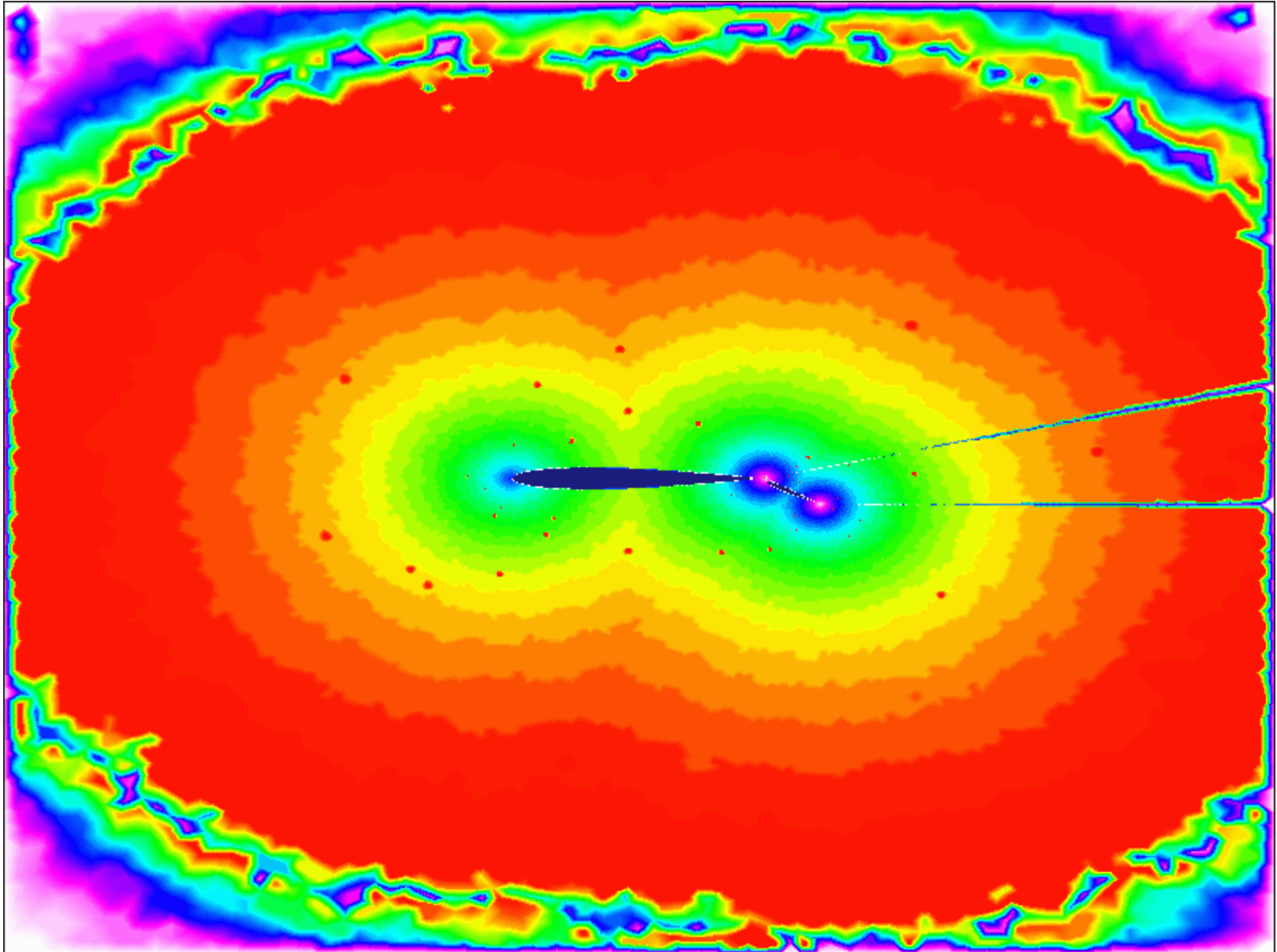
# Статическая балансировка загрузки

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

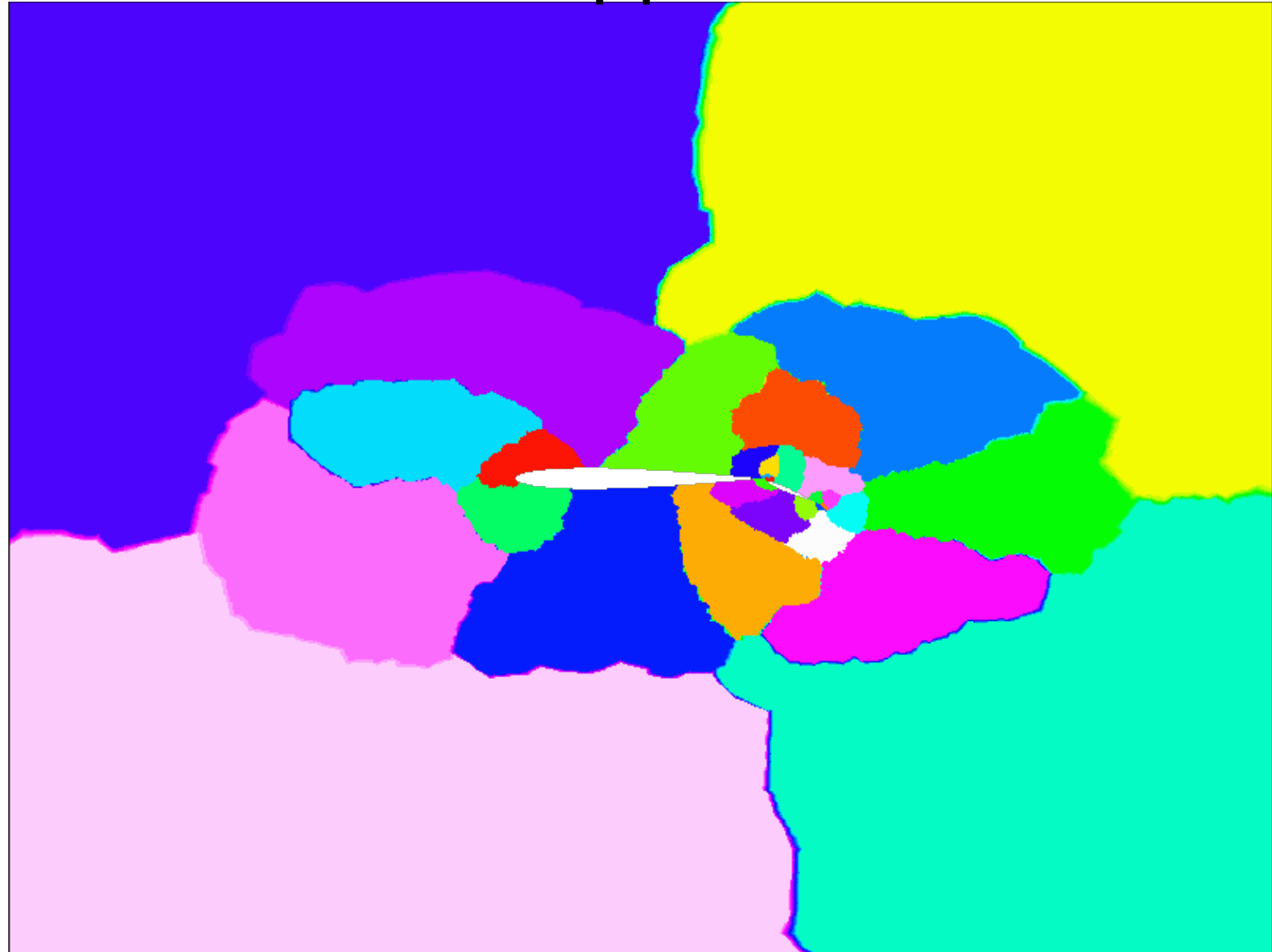




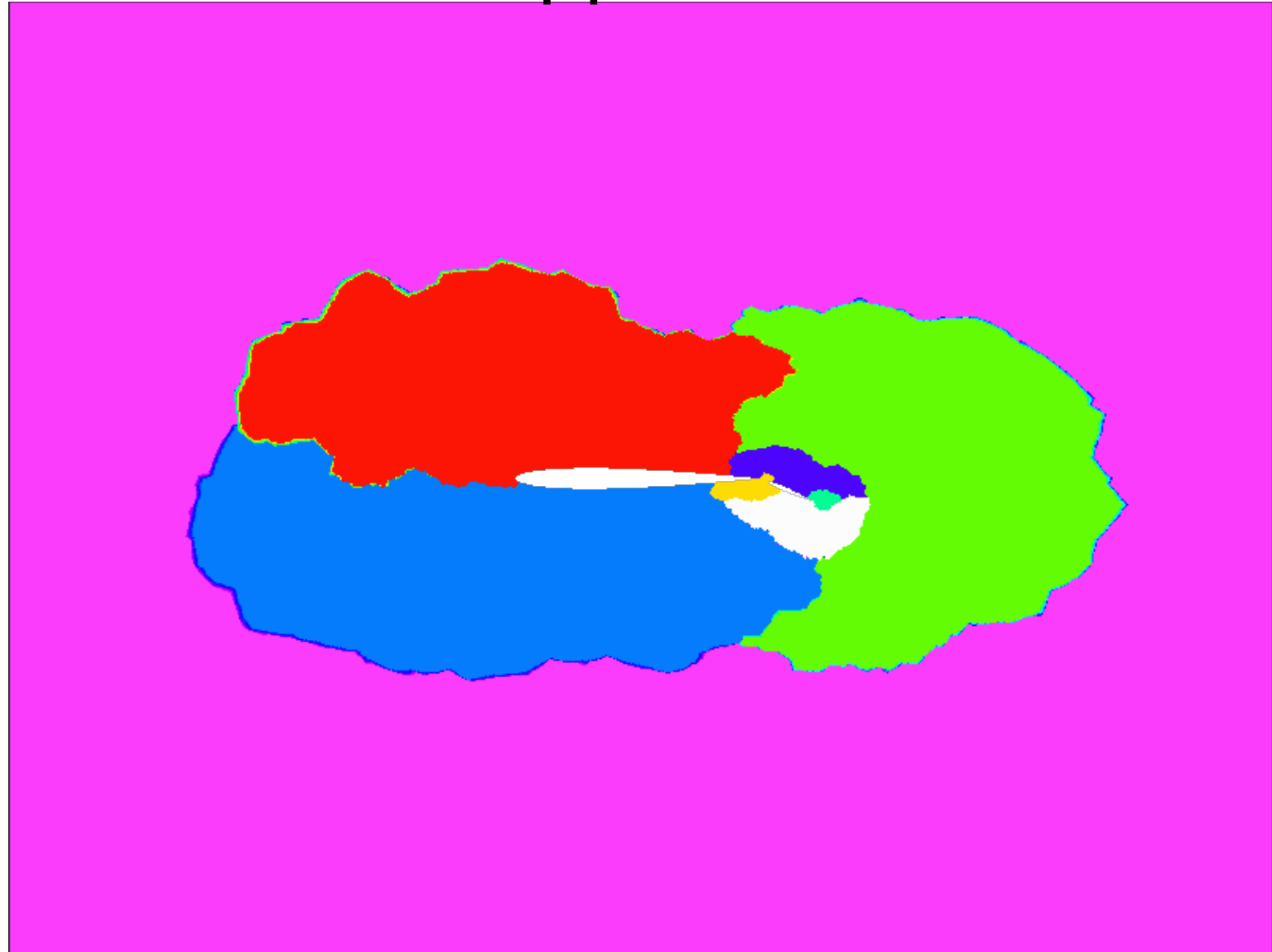
# Простое разбиение на 32 домена



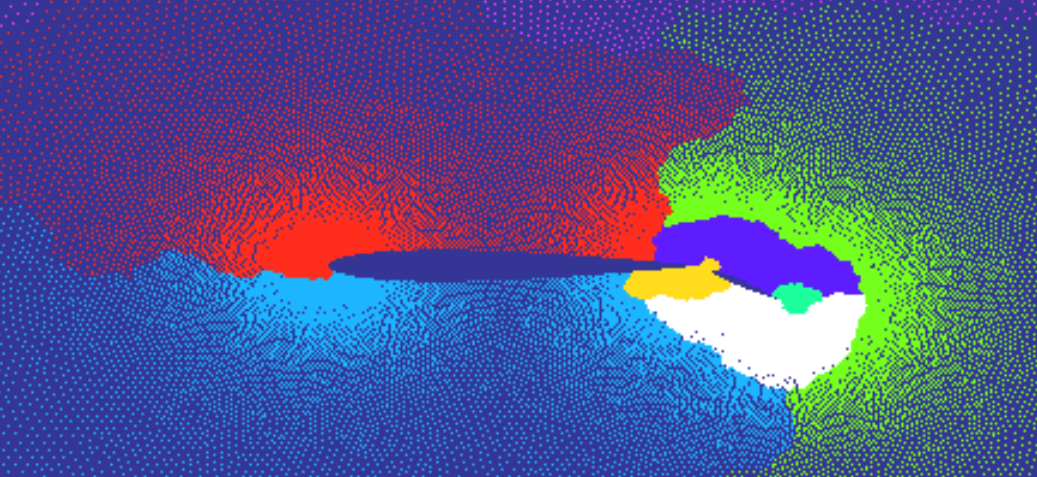
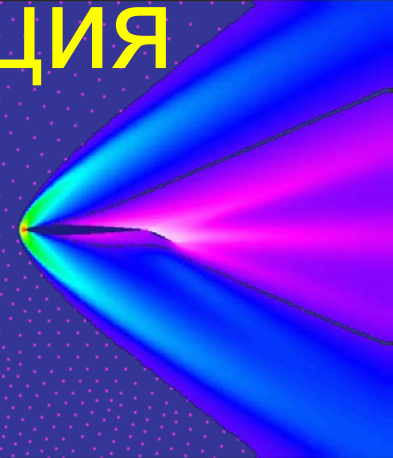
# Рациональное разбиение на 32 домена



# Рациональное разбиение на 8 доменов



# Рациональная декомпозиция сеток

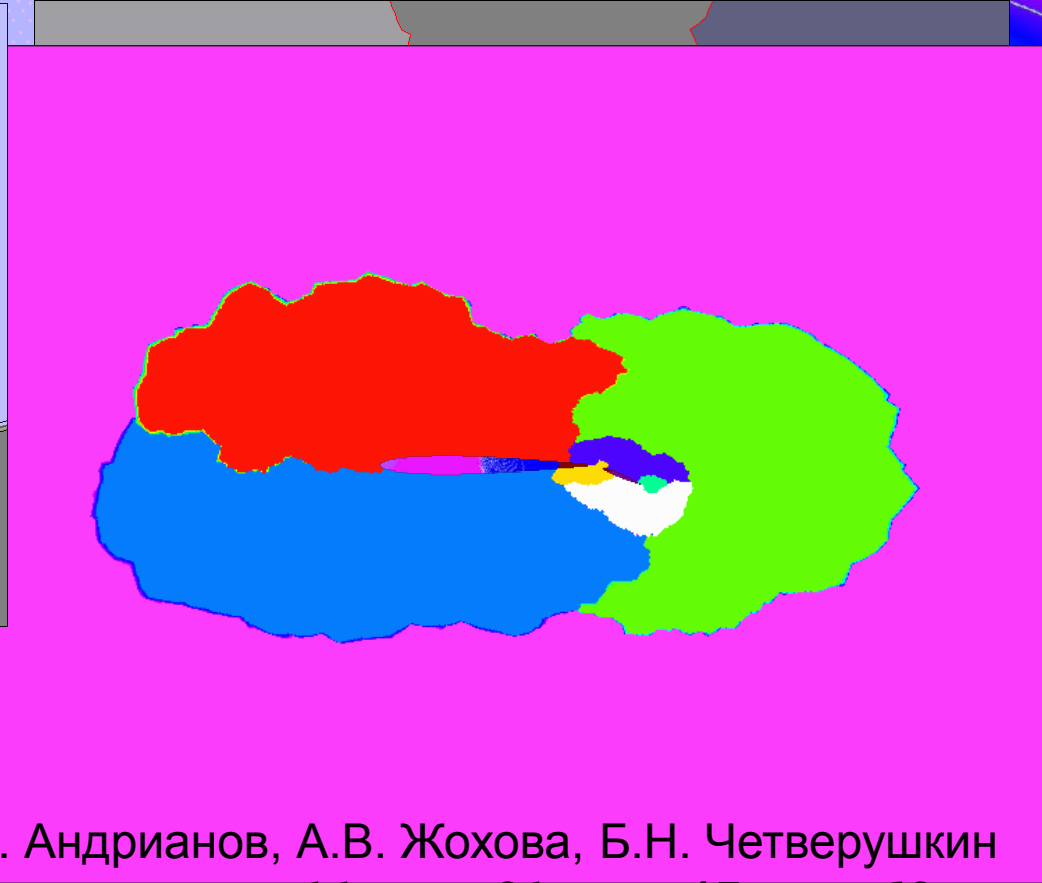
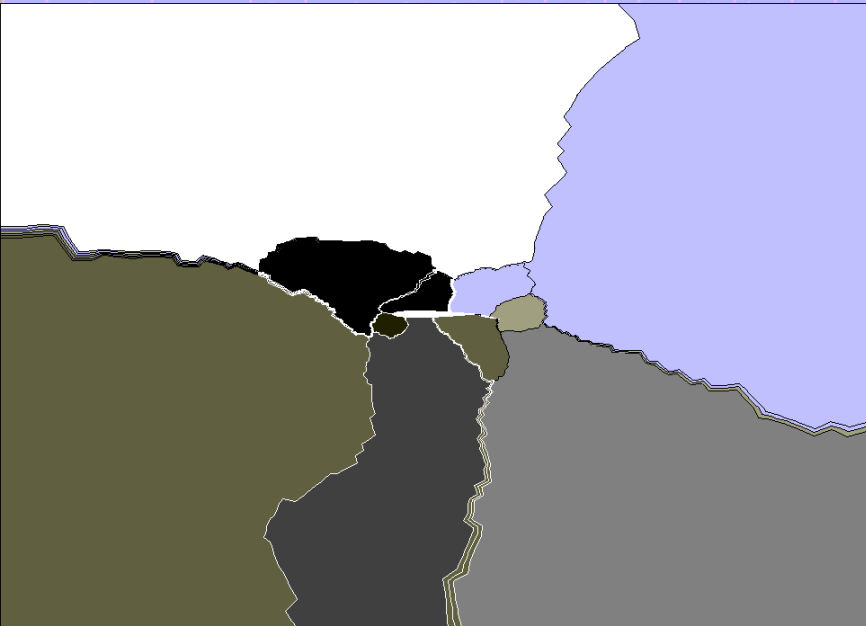


$$G = (V, E), \quad V = \{v_i\}, \quad |V| = n$$

$$V = \bigcup_{k=1}^p V_k, \quad V_i \cap V_j = \emptyset, \text{ при } i \neq j$$

$$\min_{R(V)} \left\{ J = \max_{k=1, \dots, p} \sum_{v_i \in V_k} \left( w(v_i) + \alpha \sum_{v_j \notin V_k} w(v_i, v_j) \right) \right\}$$

# 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

Чему равно  $25/4$  ?

6.25

$$25/4=$$

~~6.25~~

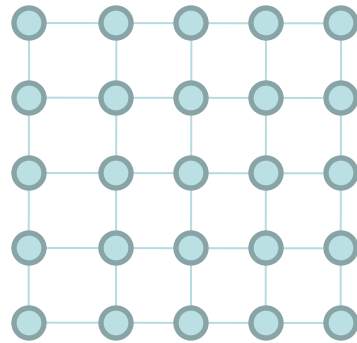
$$25/4=$$

6 ~~6.25~~ 9

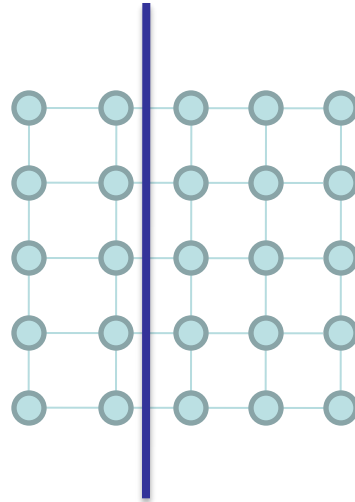
4

$$25/4 = 4 ? 6 ? 9$$

- Разрезать решетку 5 x 5 на 4 части

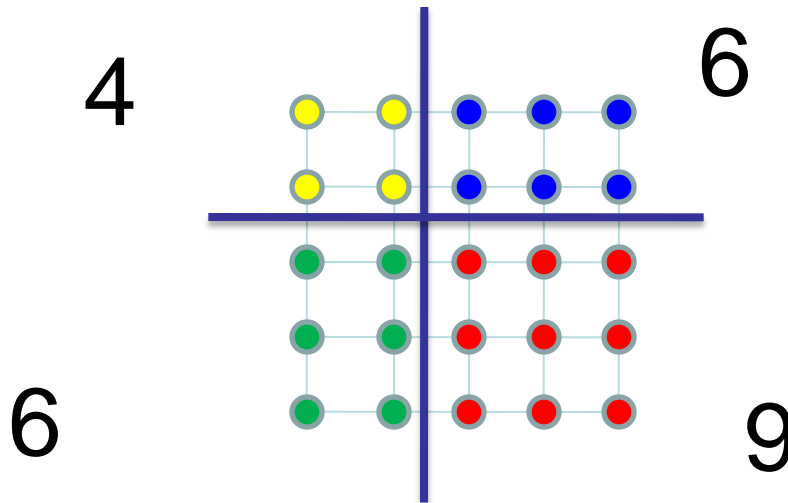


# Декомпозиция сетки из 25 узлов на 4 части



$$25/4 = 4 ? 6 ? 9$$

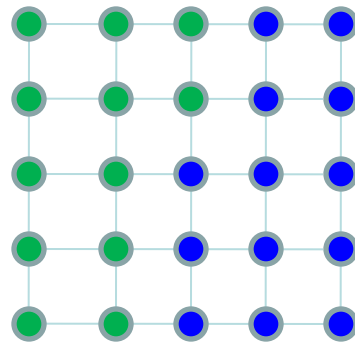
- Декомпозиция решетки 5 x 5 на 4 домена



- Дисбаланс  $9/4=2.25$

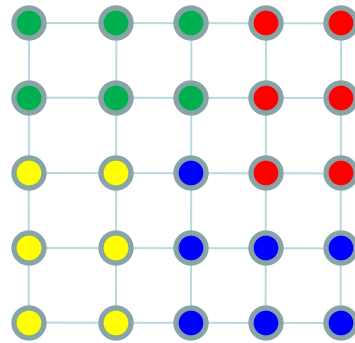
$$25/4 = 4 ? 6 ? 9$$

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



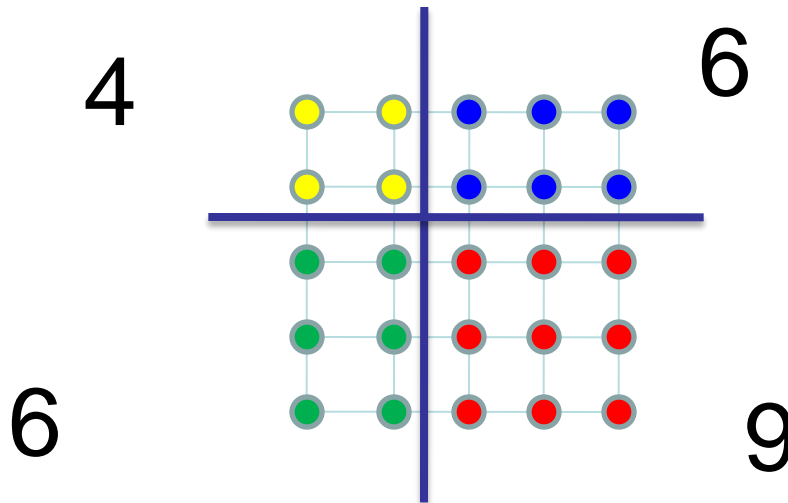
$$25/4 = 4 ? 6 ? 9$$

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



$$25/4 = 4 ? 6 ? 9$$

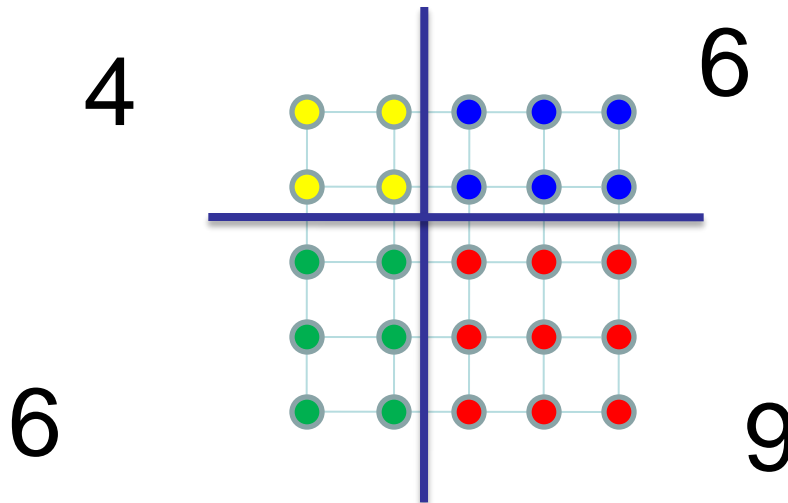
- Декомпозиция решетки 5 x 5 на 4 домена



- Дисбаланс  $9/4=2.25$

$$25/4 = 4 ? 6 ? 9$$

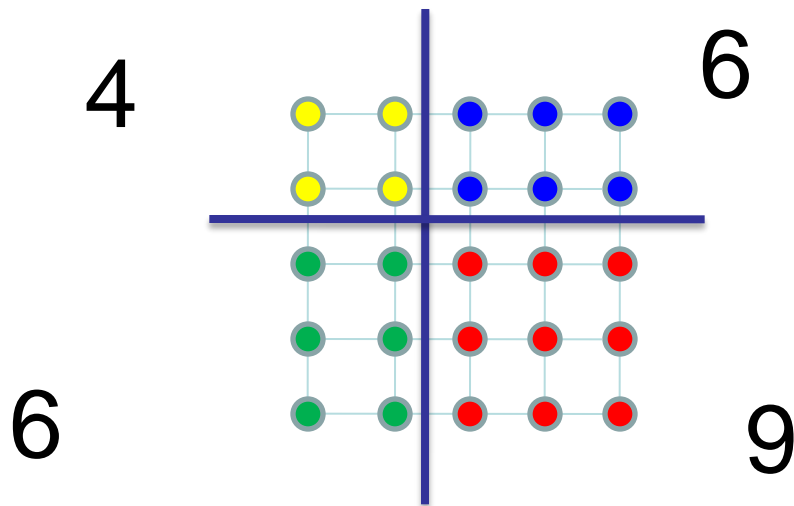
- Декомпозиция решетки 5 x 5 на 4 домена



- ~~• Дисбаланс  $9/4=2.25$~~

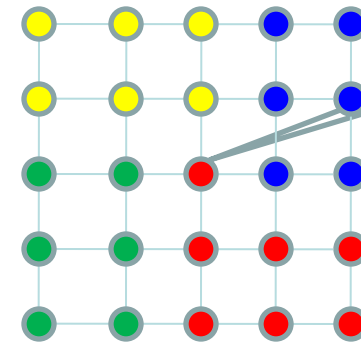
$$25/4 = 4 ? 6 ? 9$$

- Декомпозиция решетки 5 x 5 на 4 домена



Потери

$$9/6.25=1.44$$

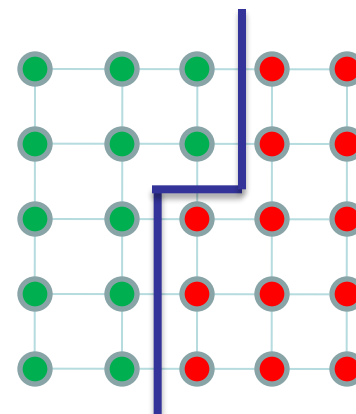
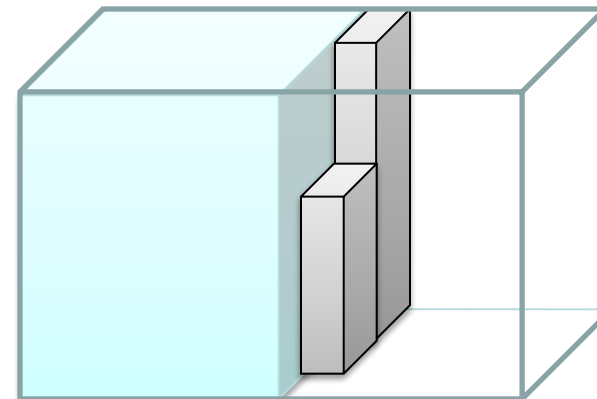
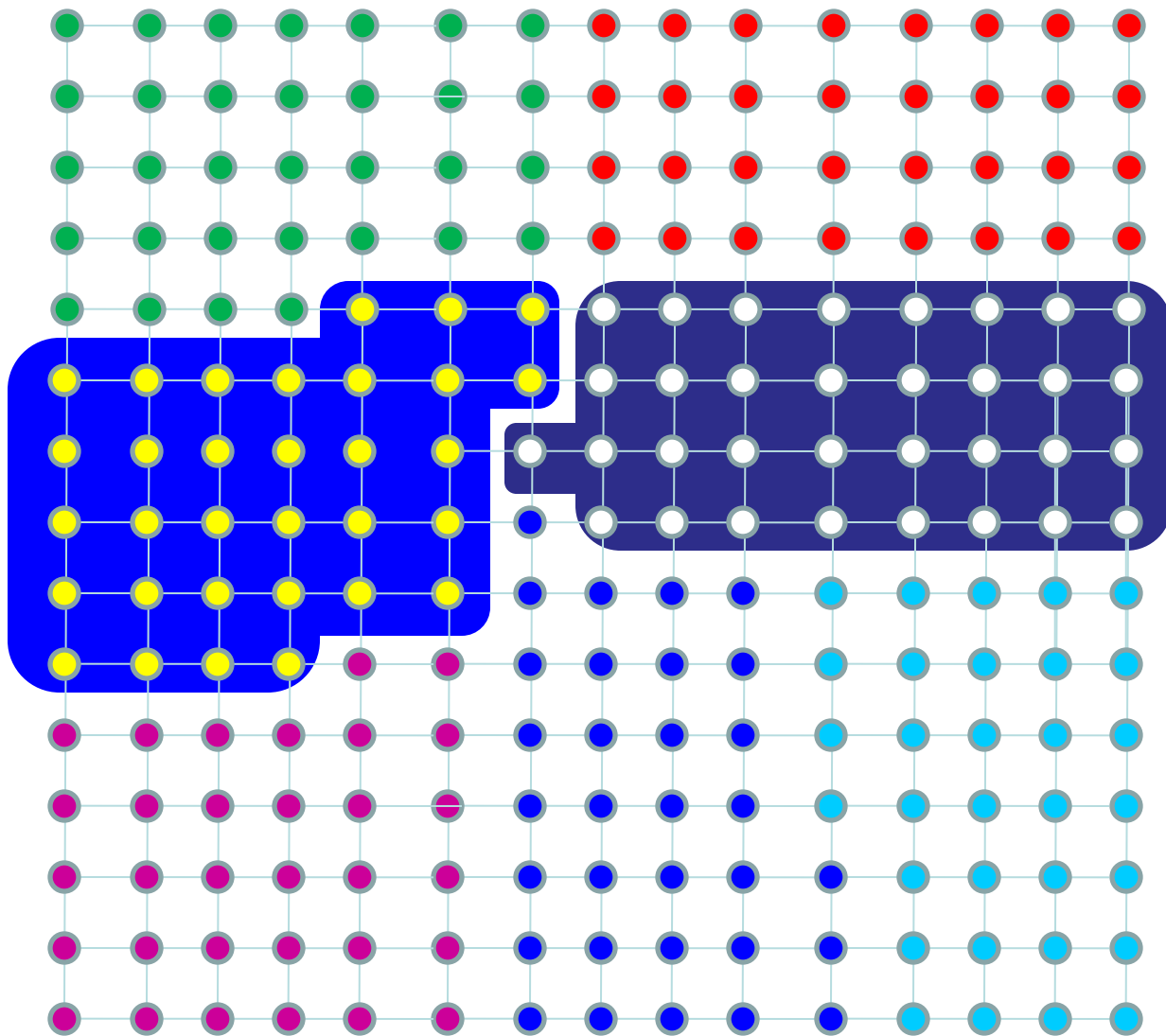


Потери

$$9/7=1.29$$

- ~~• Дисбаланс  $9/4=2.25$~~

# Декомпозиция сетки 25x25 на 7 частей





# Последовательные пакеты декомпозиции сеток

**METIS, Jostle, Scotch, Chaco, Party**

# Параллельные пакеты декомпозиции сеток

**ParMETIS, Jostle, PT-Scotch, Zoltan**

- **нерегулярные сетки, содержащие  $10^9$  и более вершин**

# Пакеты декомпозиции графов

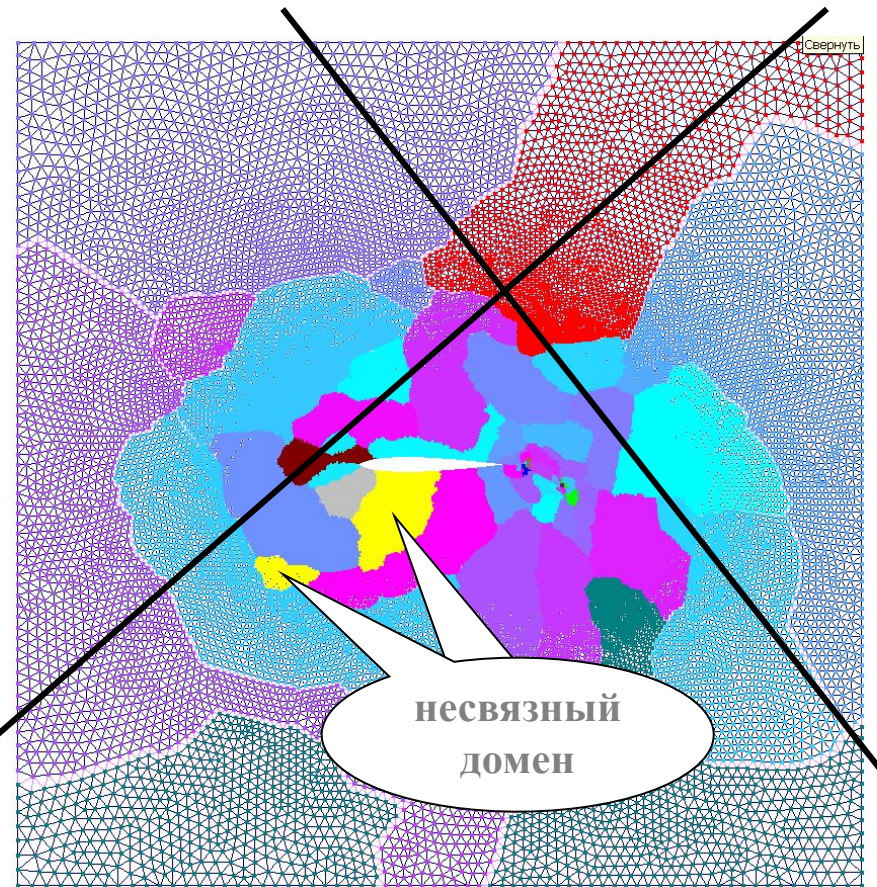
Chaco	Bruce Hendrickson Robert Leland
ParMETIS	George Karypis Vipin Kumar
PARTY	Robert Prais, et al.
JOSTLE	Chris Walshaw, et al.
SCOTCH	Francois Pellegrini

# Недостатки алгоритмов декомпозиции графов, реализованных в существующих пакетах

- **образование несвязных доменов**
- **формирование сильно несбалансированных разбиений (ParMETIS: количества вершин в доменах могут отличаться в 2 раза)**
- **не всегда удастся получить разбиения на большое число микродоменов**

# Связность важна:

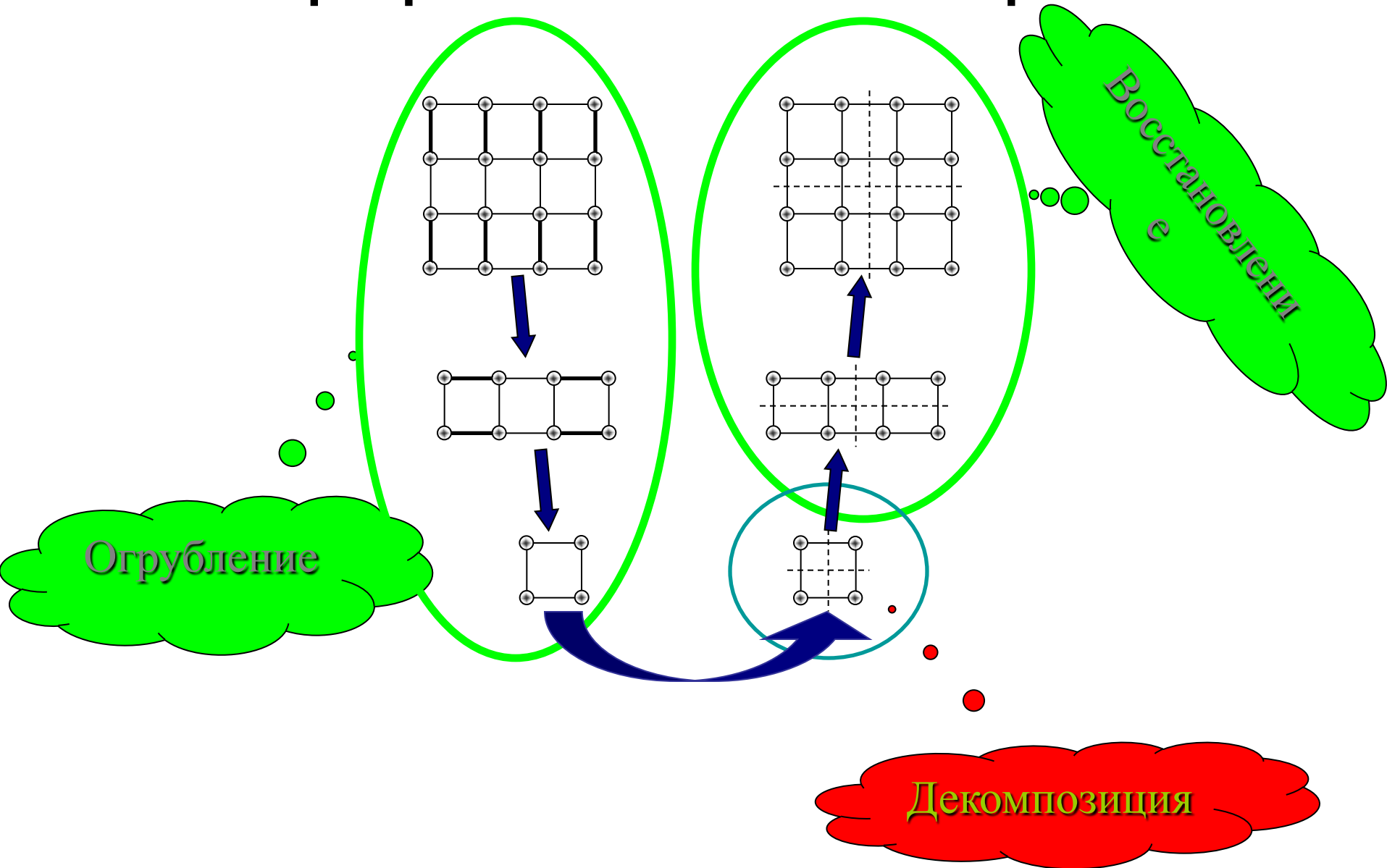
- алгоритмы решения систем линейных уравнений
- компрессия сеточных данных
- алгоритм композиции подобластей<sup>1</sup>
- распараллеливание методики ТИМ-2D<sup>2</sup>



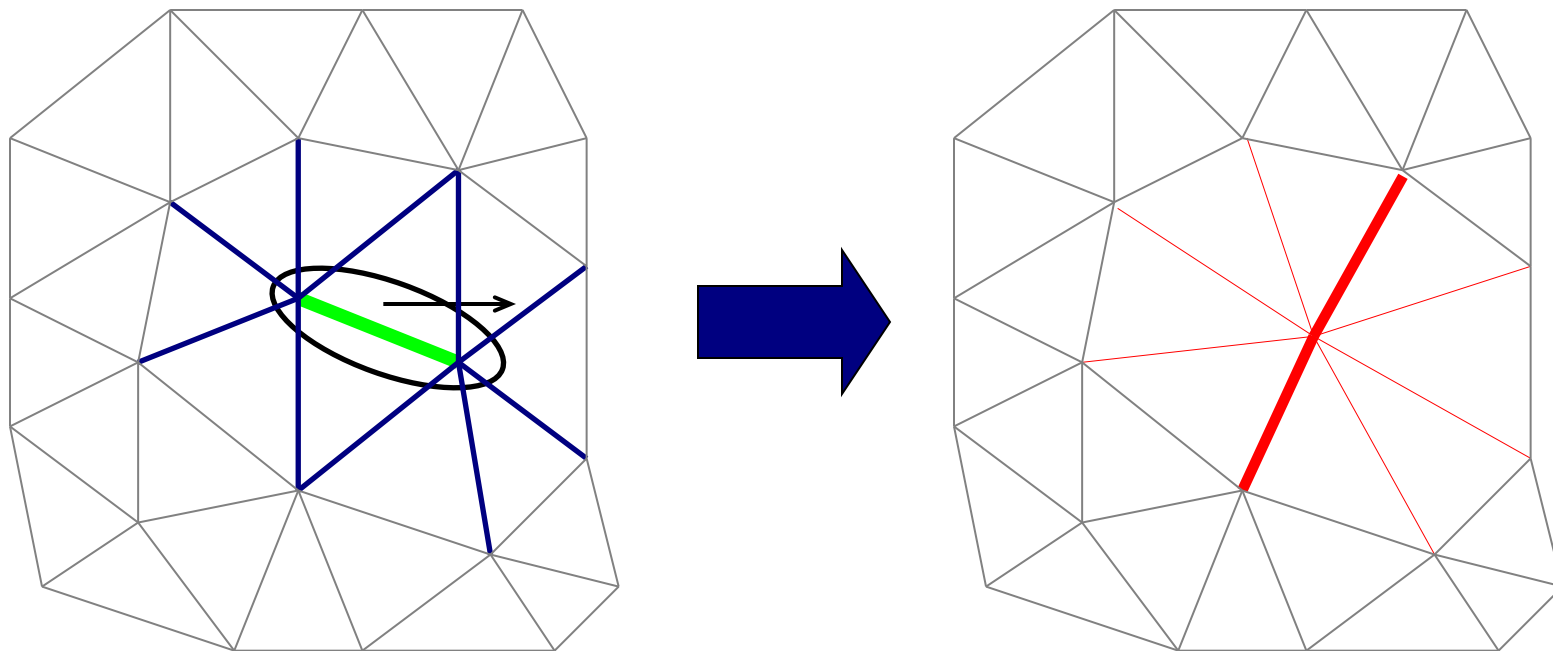
<sup>1</sup> А. И. Илюшин, А. А. Колмаков, И. С. Меньшов. Построение параллельной вычислительной модели путем композиции вычислительных объектов // Математическое моделирование. 2011. Т. 23. № 7. 97-113.

<sup>2</sup> 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.

# Иерархический алгоритм



# Огрубление графа



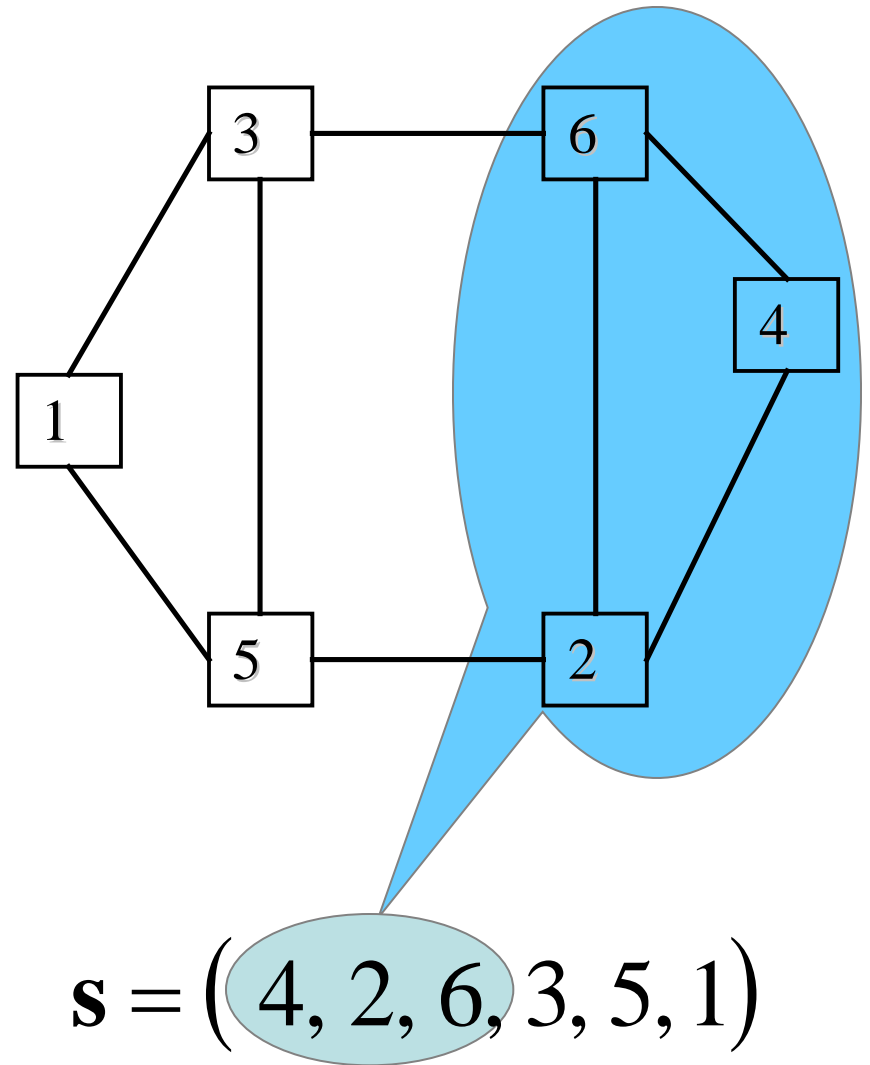
# Спектральный метод

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

$$\lambda_1 = 1$$

$$\lambda_{0-5} = \{0, 1, 3, 3, 4, 5\}$$

$$x_1 = (2, -1, 1, -2, 1, -1)$$



# Спектральная бисекция

Пусть  $q[i] \in \{-1, 1\}$  определяет принадлежность вершины  $i$  одному из двух доменов

Тогда, для равенства числа вершин в доменах, необходимо

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

# Спектральная бисекция

Общее число вершин  $\sum_{i=1}^{|V|} q_i^2 = q^T q = |V|$

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

# Число разрезанных ребер

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

$$Lq = \lambda q$$

$$q^T Lq = \sum_{e_{i,j} \in E} (q_i - q_j)^2 = q^T \lambda q = \lambda |V|$$

Следовательно  $|E_c| = \frac{1}{4} \lambda |V|$

# Разбиение вершин на два множества

Для минимизации  $|E_c| = \frac{1}{4} \lambda |V|$   $\lambda_0 = 0$

следует найти минимальное собственное число и соответствующий ему собственный вектор – вектор Фидлера  $\psi$

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

$L\psi = \lambda_1 \psi$ , следовательно  $e\psi = 0$ ,  $\sum_{i=1}^{|V|} \psi_i = 0$ ,

значит множества  $\{-1\}$  и  $\{1\}$  содержат

# Рекурсивная бисекция

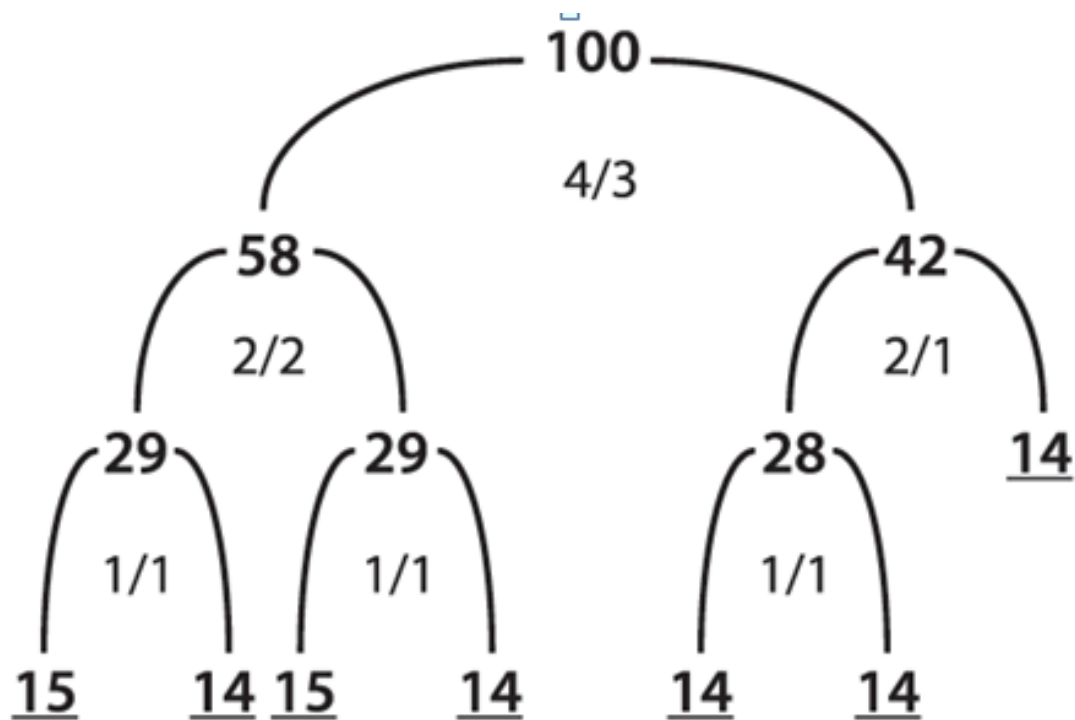
Рекурсивная бисекция

```
RecursiveBisect (graf, n, k)
{
k1 = (k+1) / 2
k2 = k - k1
n1 = n * k1 / k
n2 = n - n1

// разбиение вершин графа на две части
// размером  $n_1$  и  $n_2$ 
Bisect (graf, n, subgraf1, n1, subgraf2, n2)

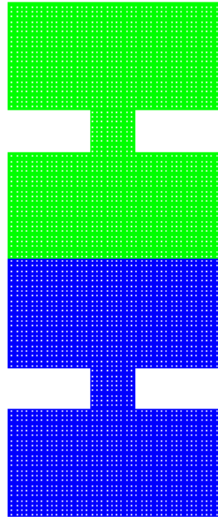
if (k1 > 1) RecursiveBisect (subgraf1, n1, k1)
if (k2 > 1) RecursiveBisect (subgraf2, n2, k2)
}
```

# Бинарное дерево разрезов

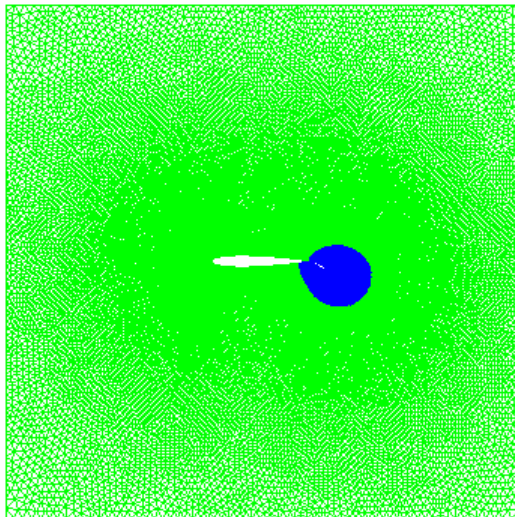


# Метод спектральной бисекции

Spectral Partition

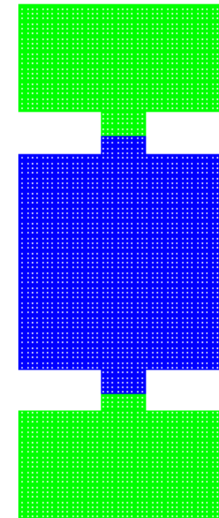


100 cut edges

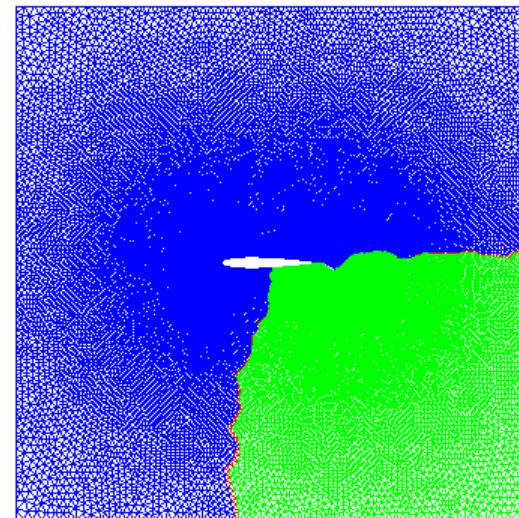


655 cut edges

Metis Partition

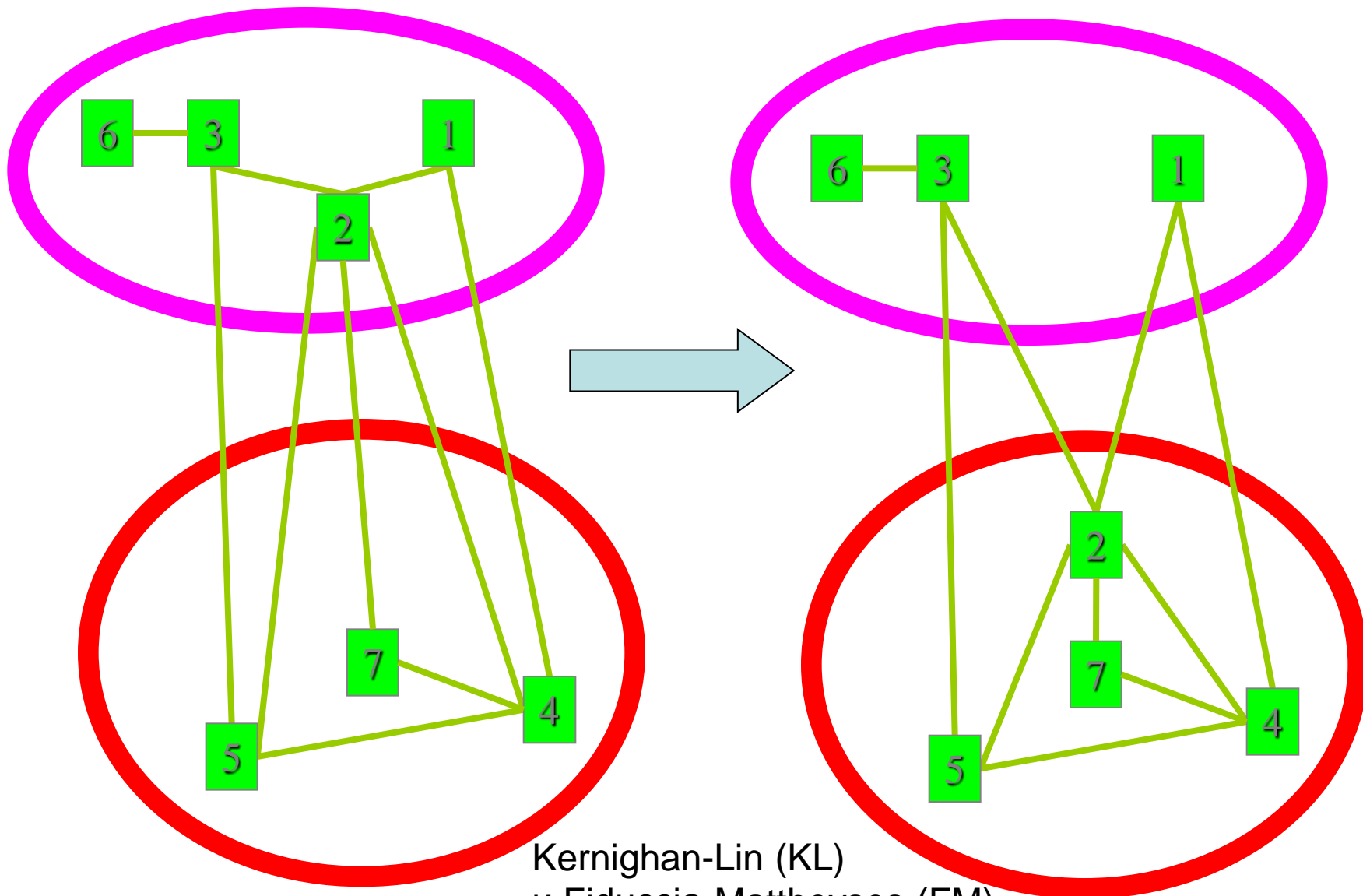


42 cut edges



524 cut edges

# Локальное уточнение

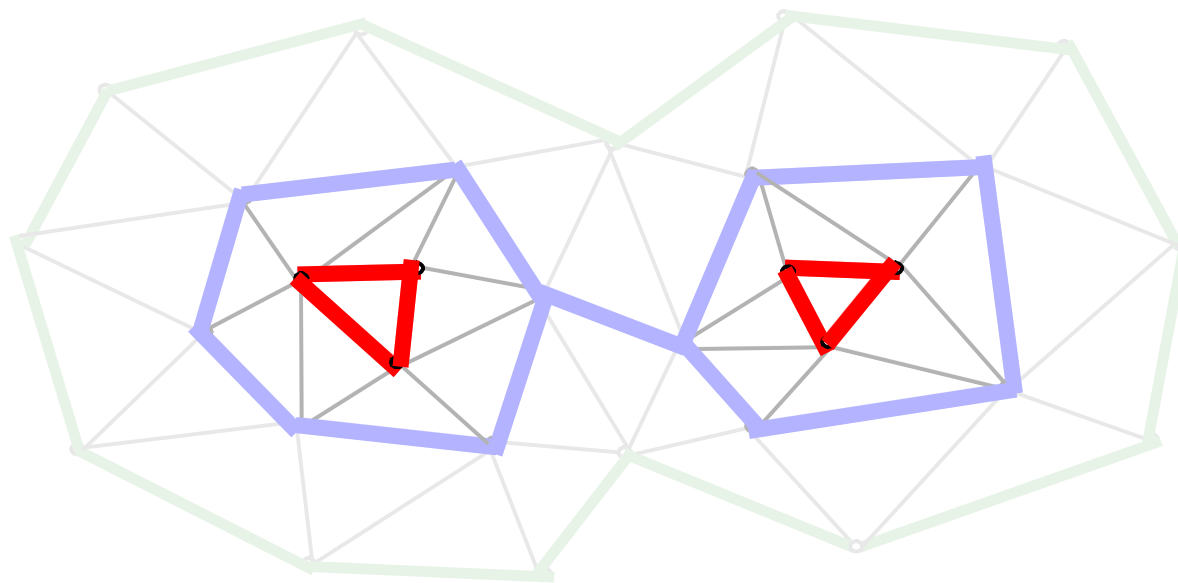


Kernighan-Lin (KL)  
и Fiduccia-Mattheyses (FM)

# Инкрементный алгоритм декомпозиции графа

1. инициализация доменов
2. распределение вершин по доменам методом инкрементного роста
3. локальное уточнение границ сформированных доменов
4. анализ связности ядер сформированных доменов и окончание работы, если заданный уровень связности достигнут
5. перенос части закрепленных за доменами вершин в группу свободных вершин и переход к этапу 2.

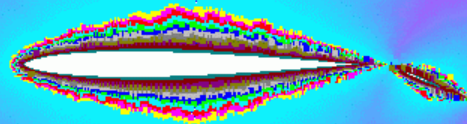
# Связность ядер доменов



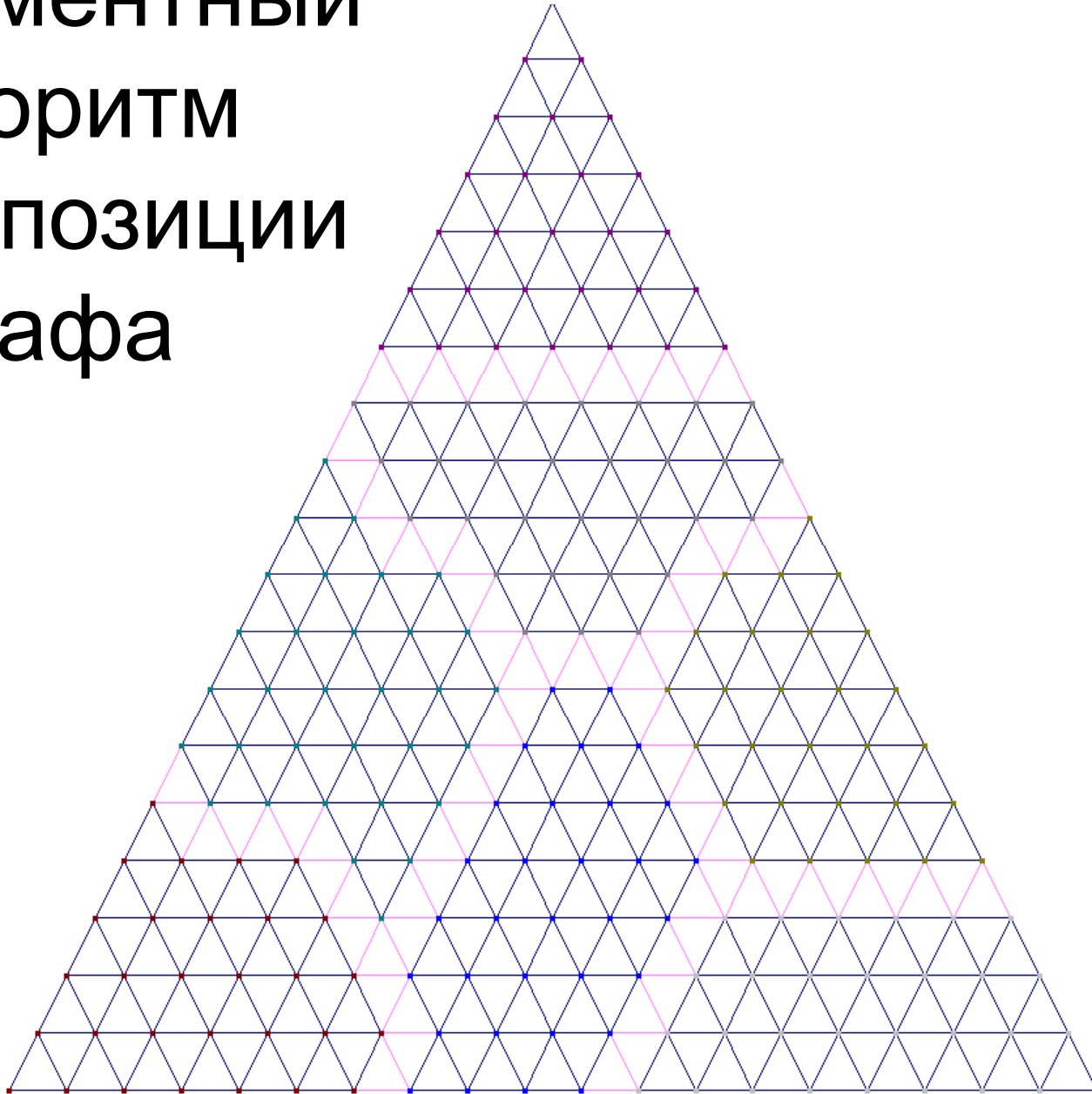
$$d(i) = \min_k : \mathbf{A}^k v_i \cap B \neq \emptyset$$

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

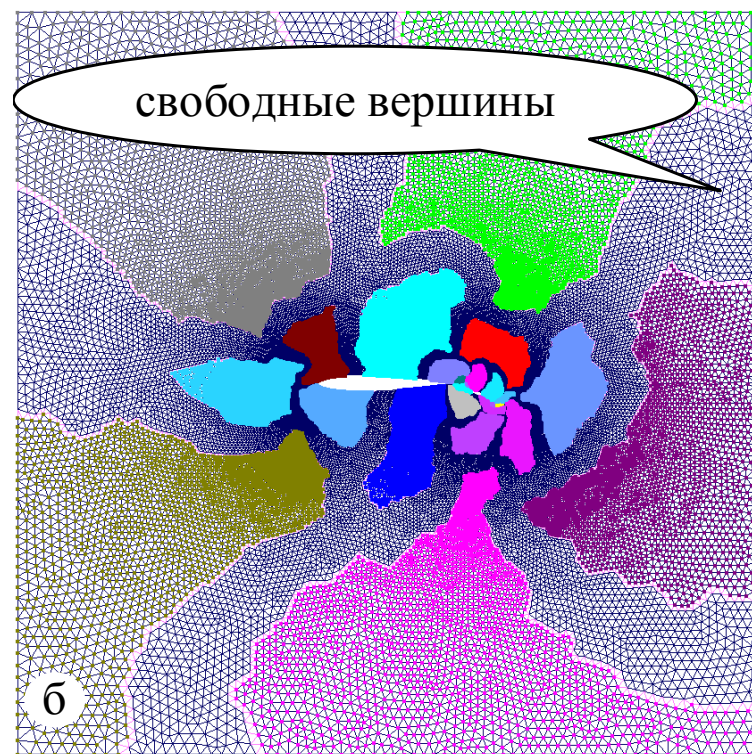
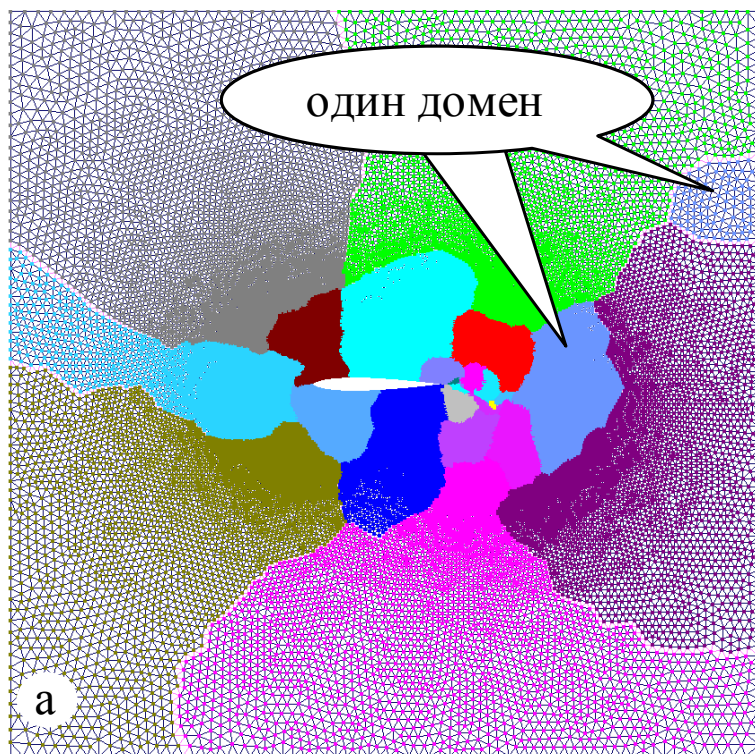
# Оболочки



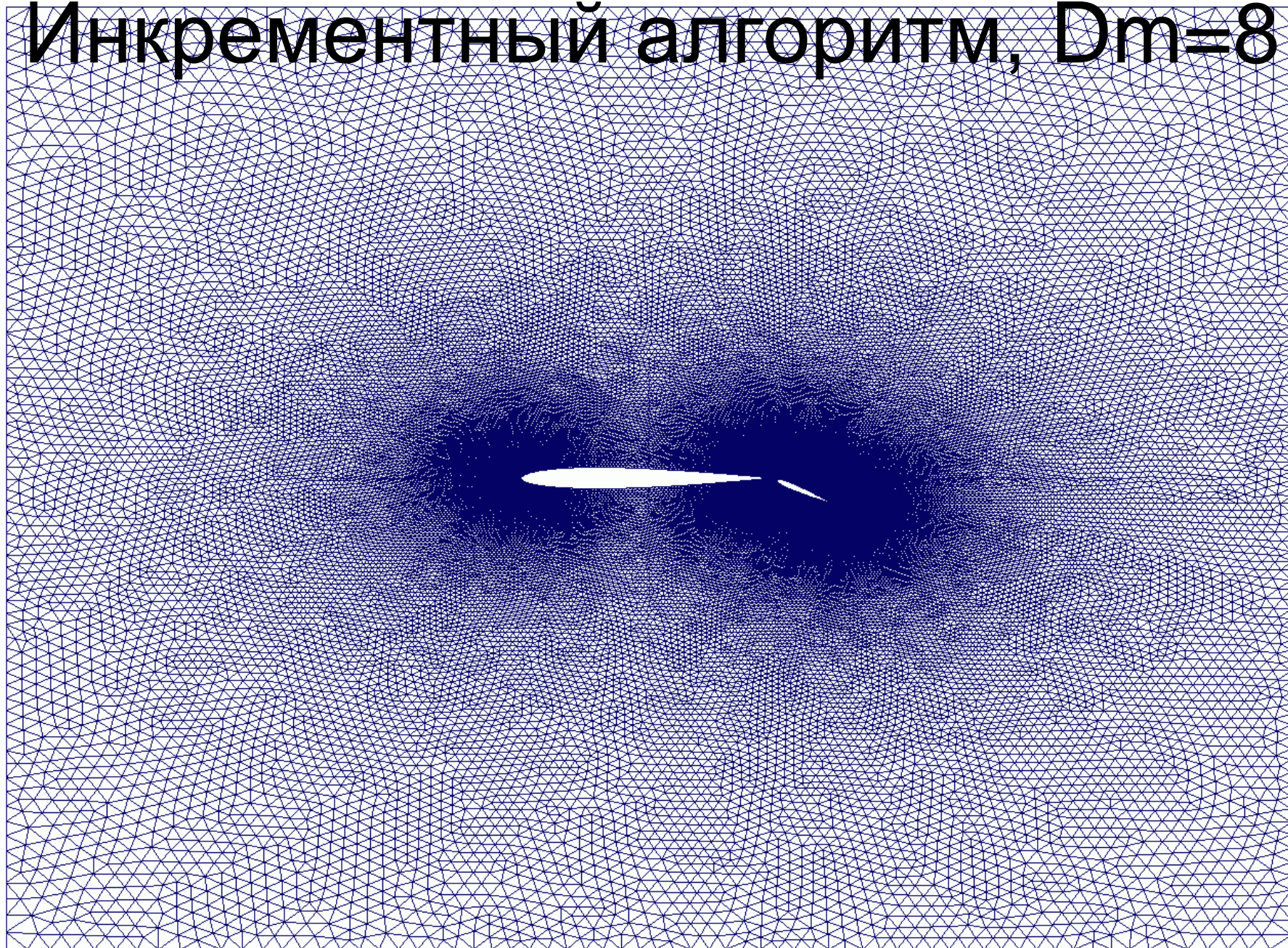
# Инкрементный алгоритм декомпозиции графа



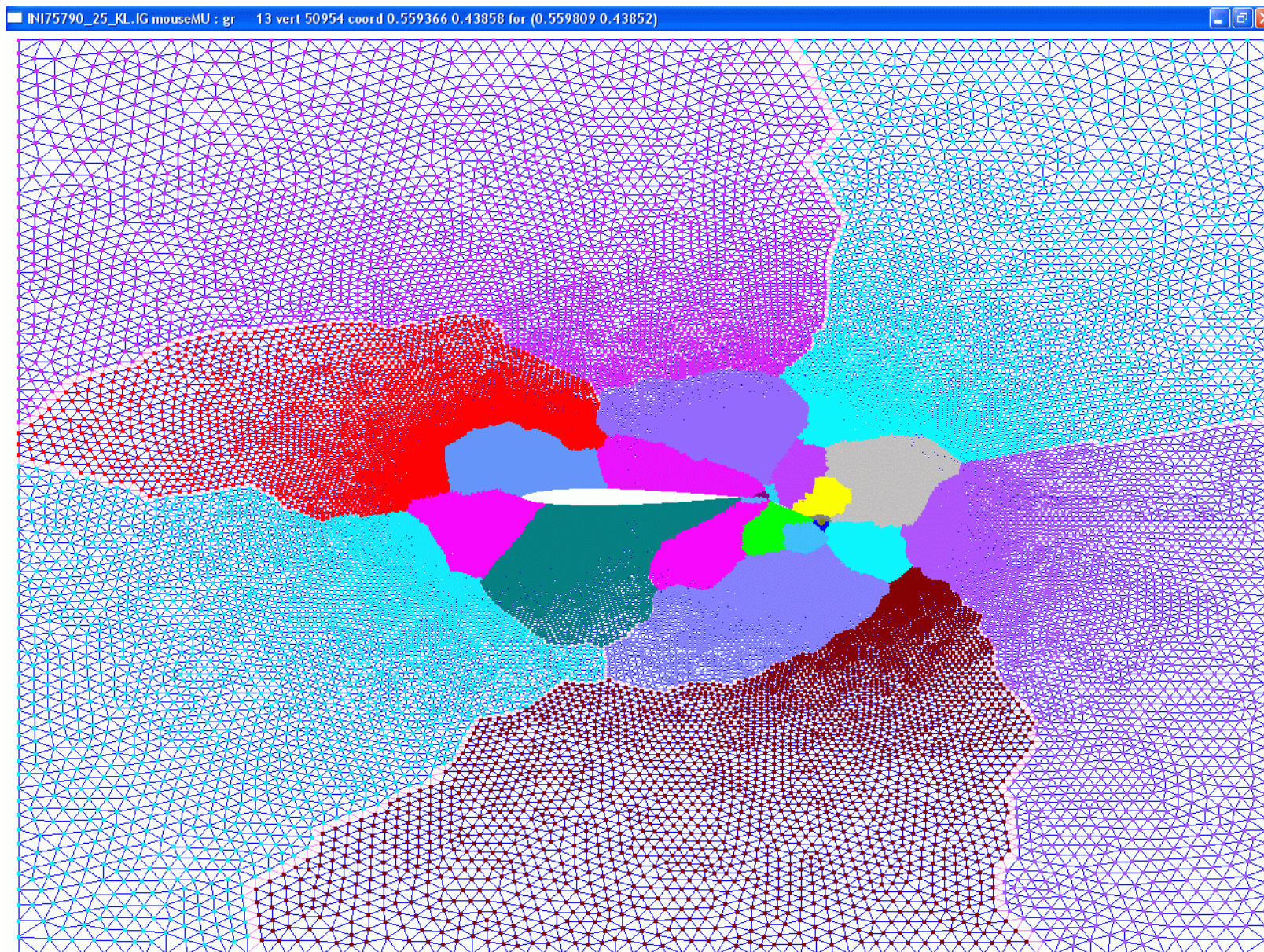
# Редуцирование доменов



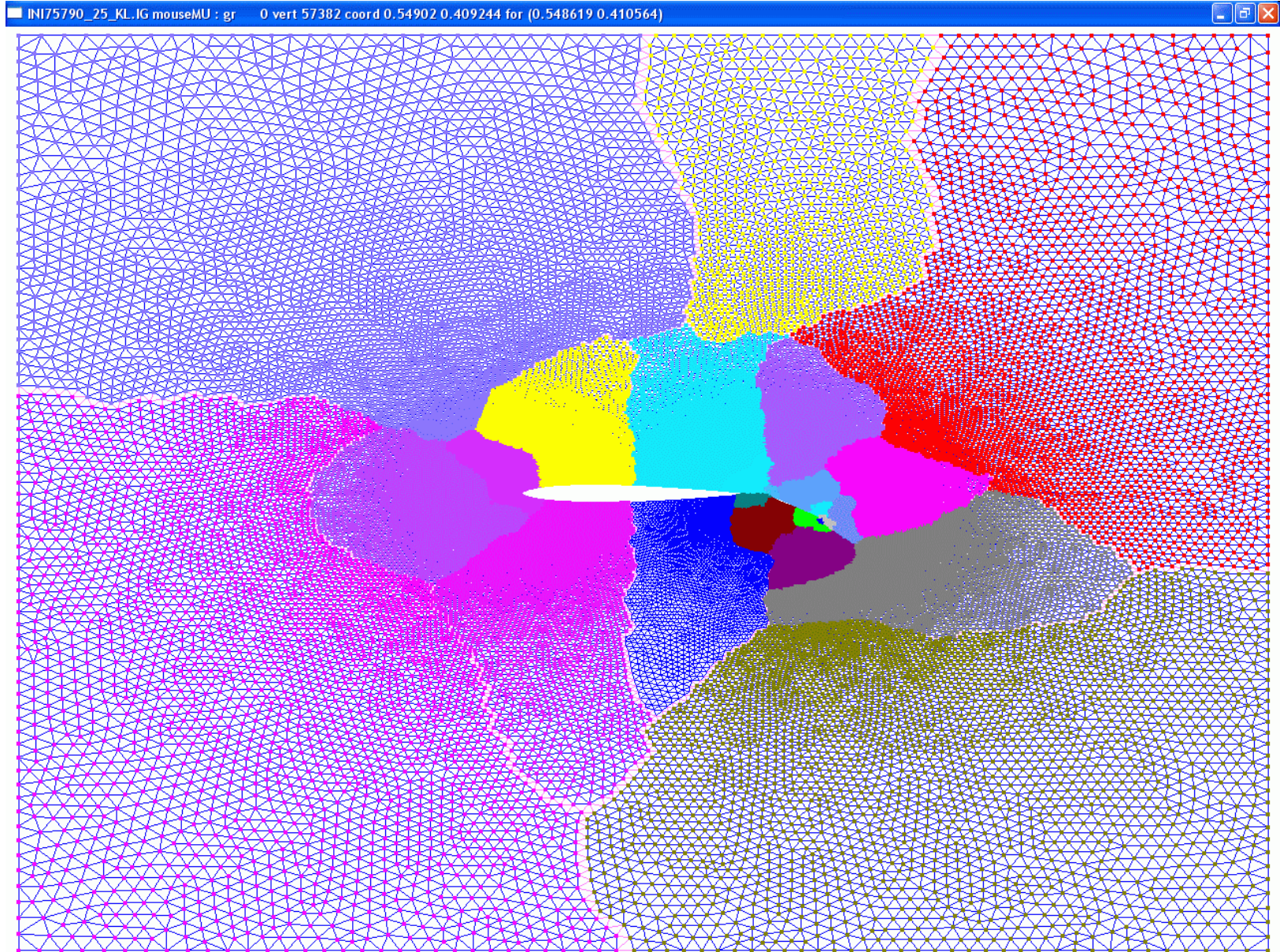
# Инкрементный алгоритм, $Dm=8$



# Инкрементный алгоритм, $Dm=25$



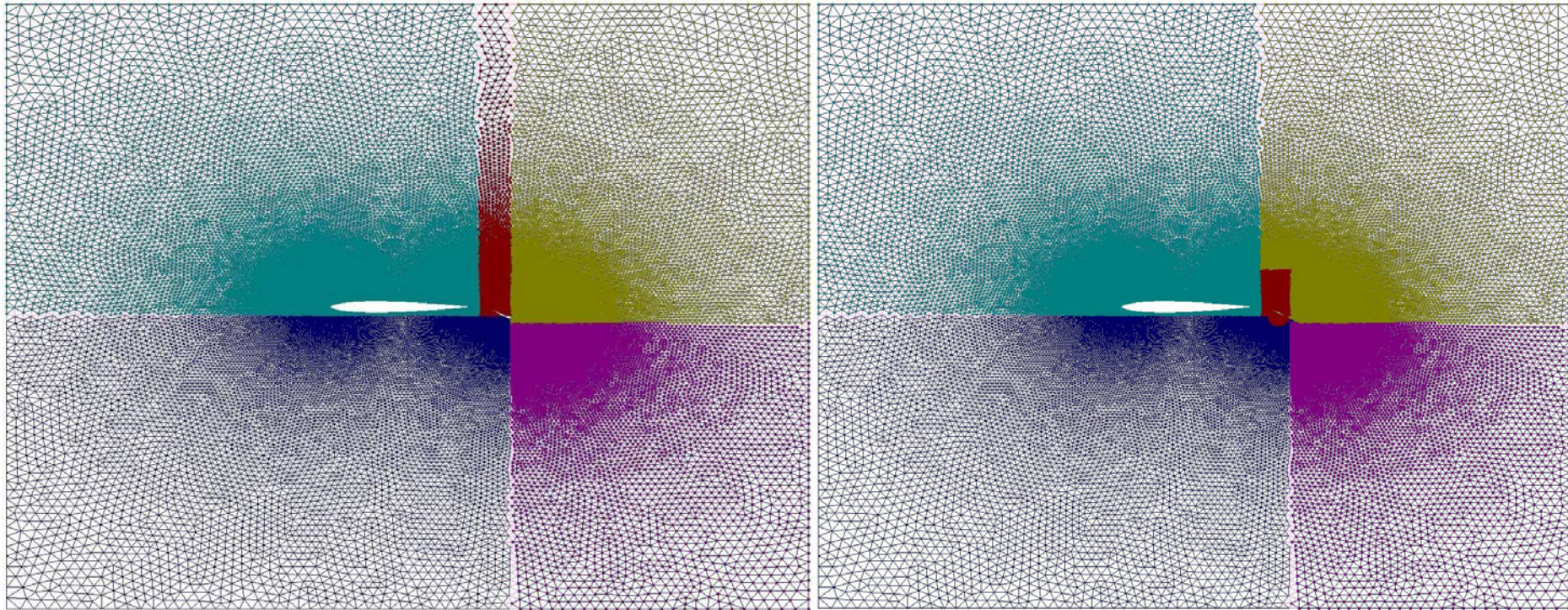
# Kmetis, Dm=25



# Треугольная сетка из 75790 вершин (пространство вокруг крыла)

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

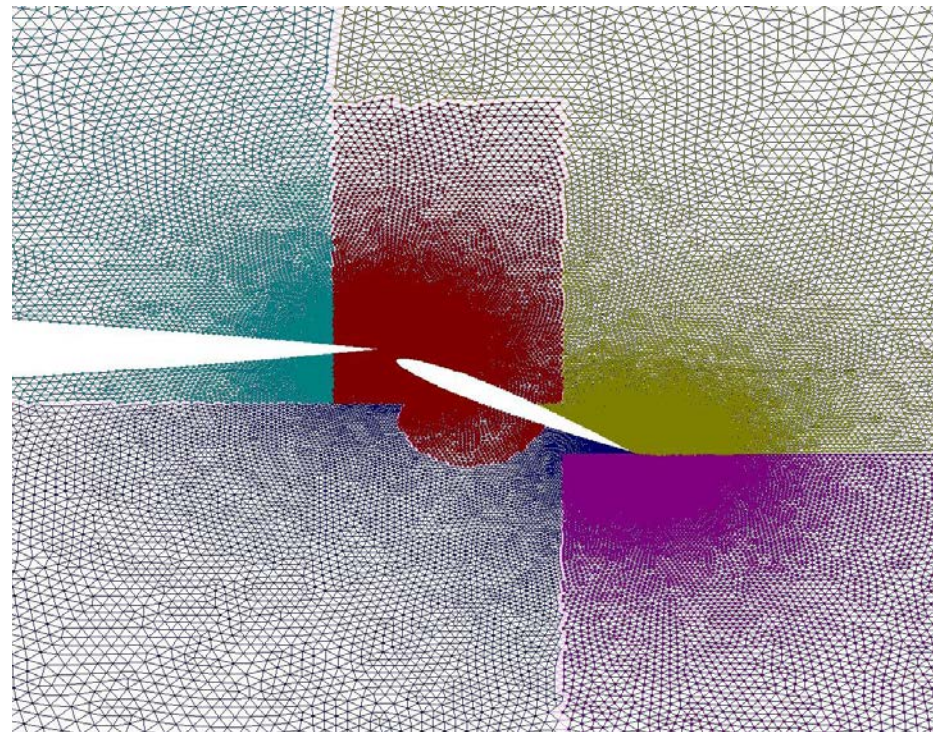
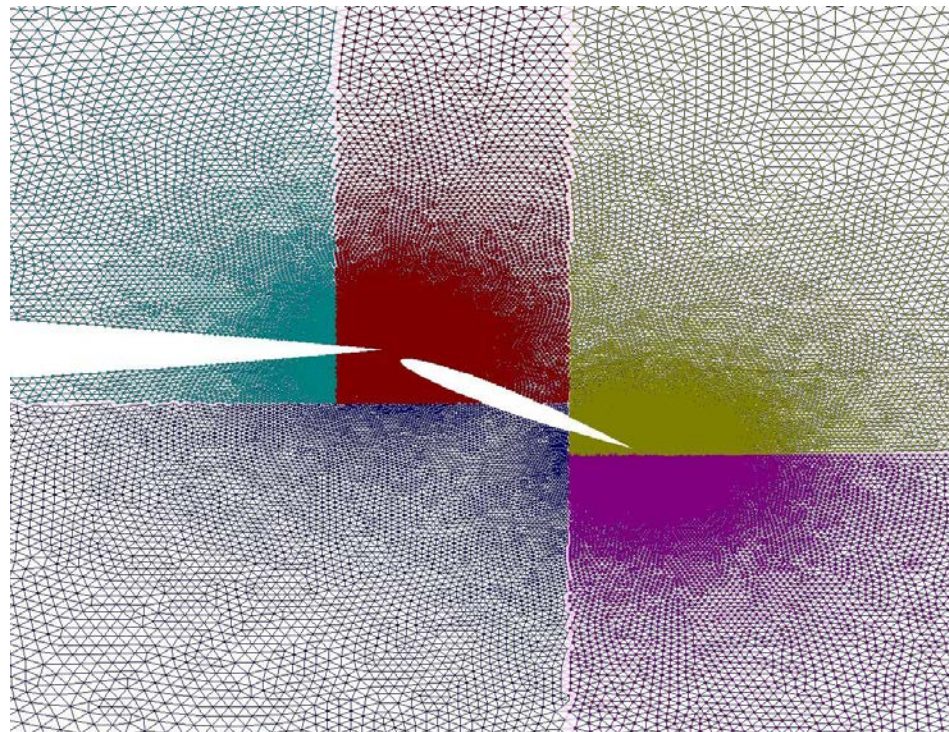
результат перераспределения  
малых блоков вершин



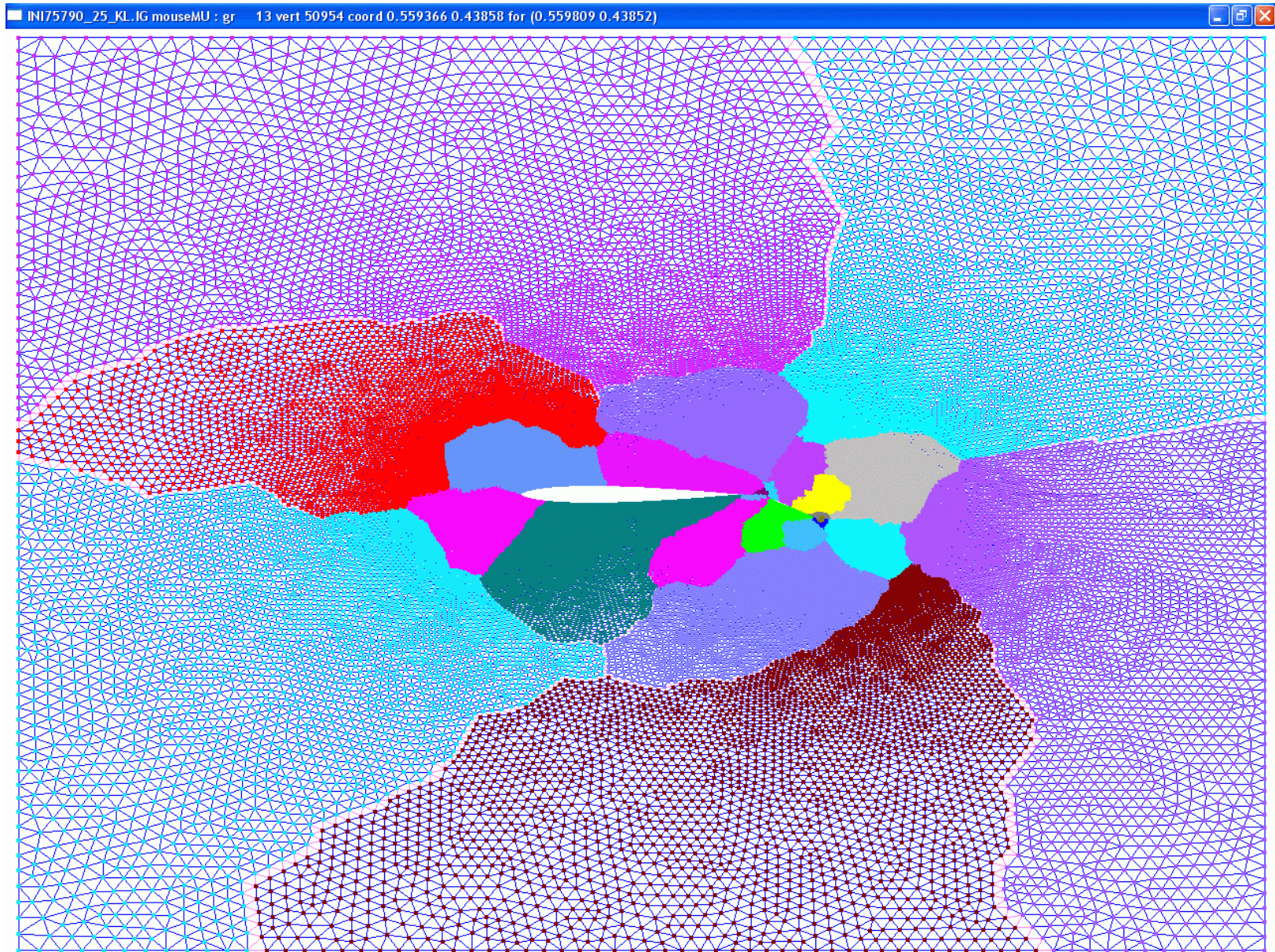
# Фрагмент треугольной сетки из 75790 вершин

результат геометрической  
декомпозиции

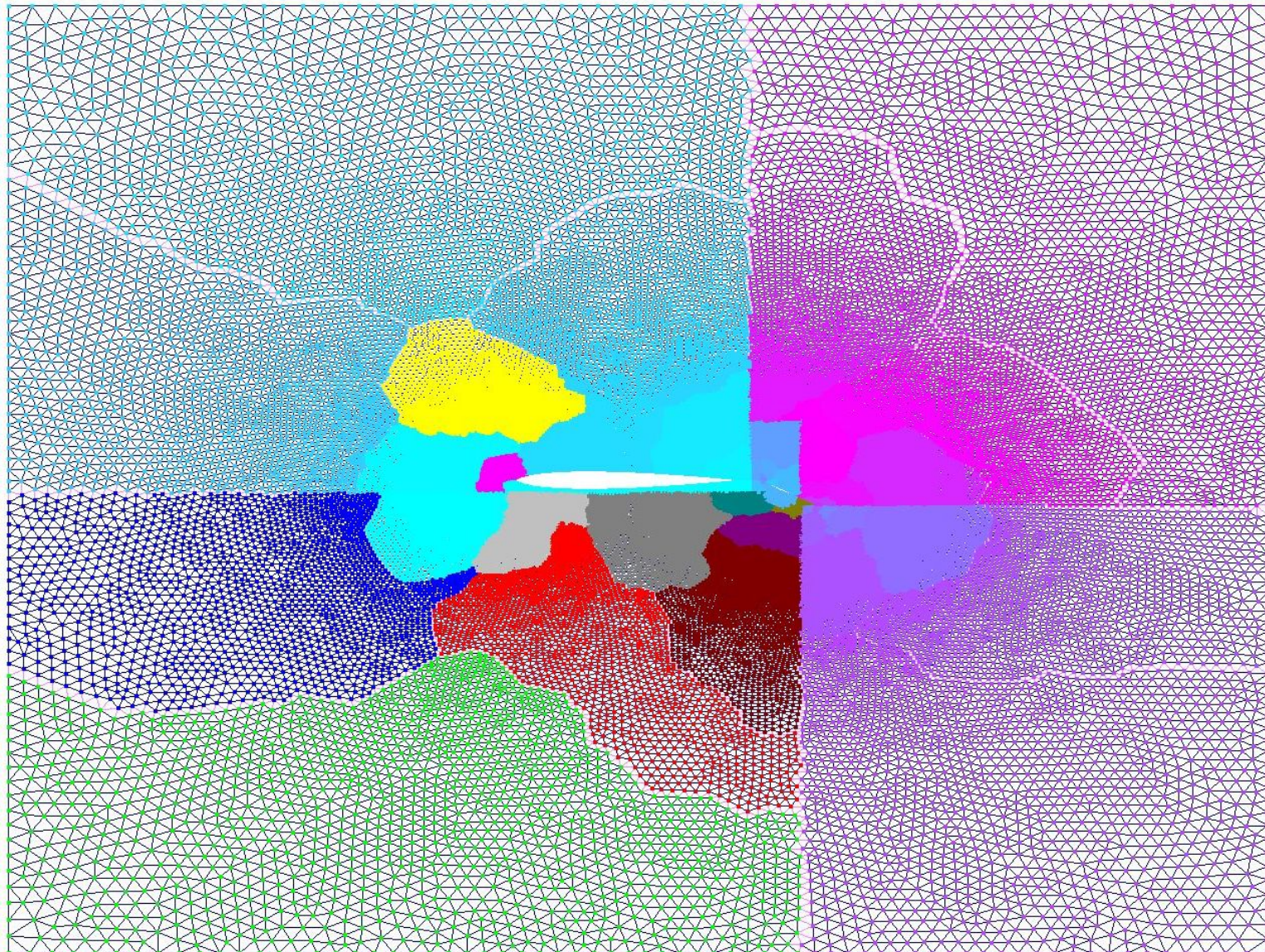
результат перераспределения  
малых блоков вершин



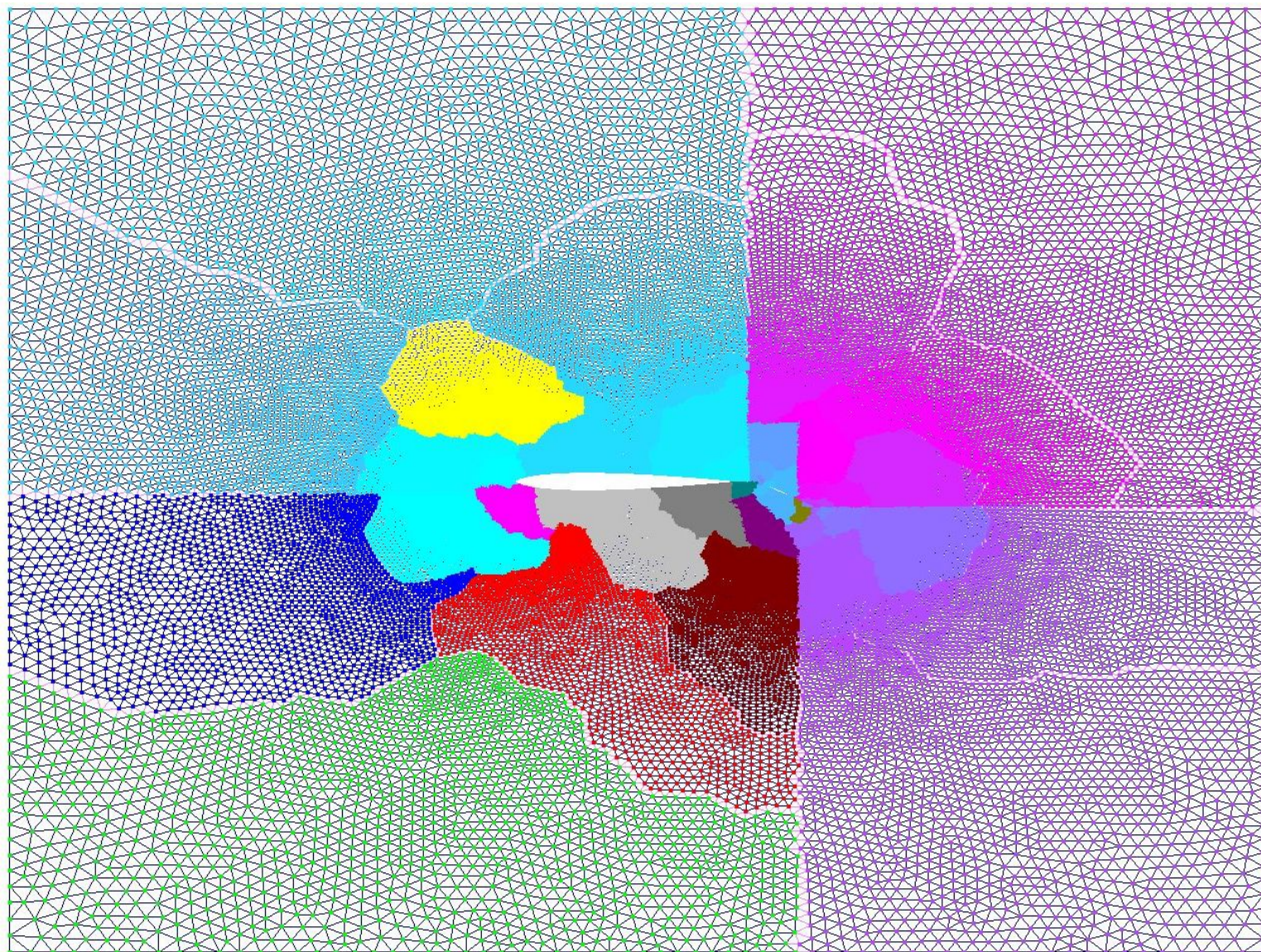
# Инкрементный алгоритм, $Dm=25$



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



## Результат сбора плохих групп доменов и их повторного разбиения

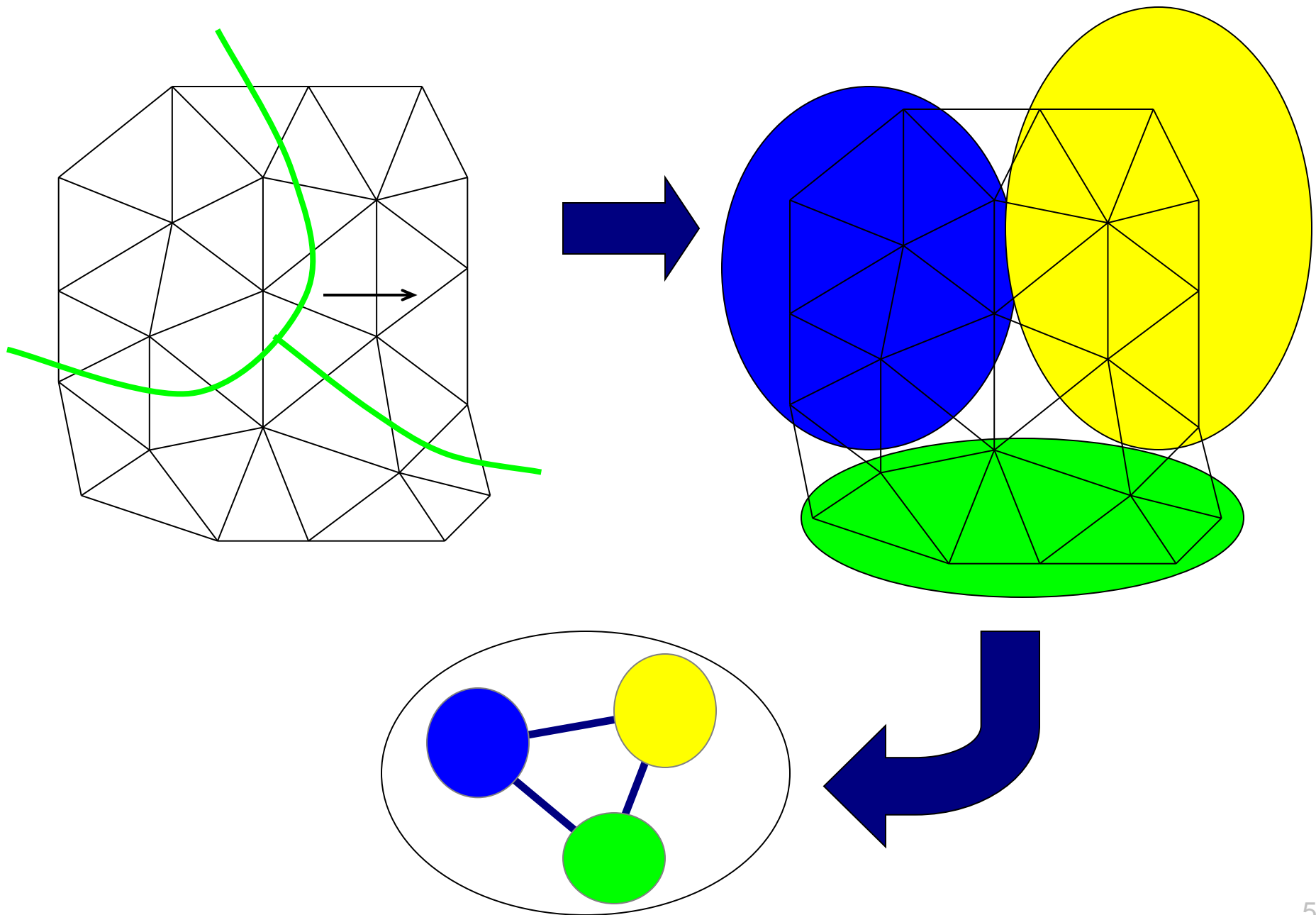


# Разбиение тетраэдральной сетки, содержащей $2 \cdot 10^8$ узлов, на 125 процессорах

- вычисления производились на кластере СКИФ МГУ (1250 4-ядерных 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	

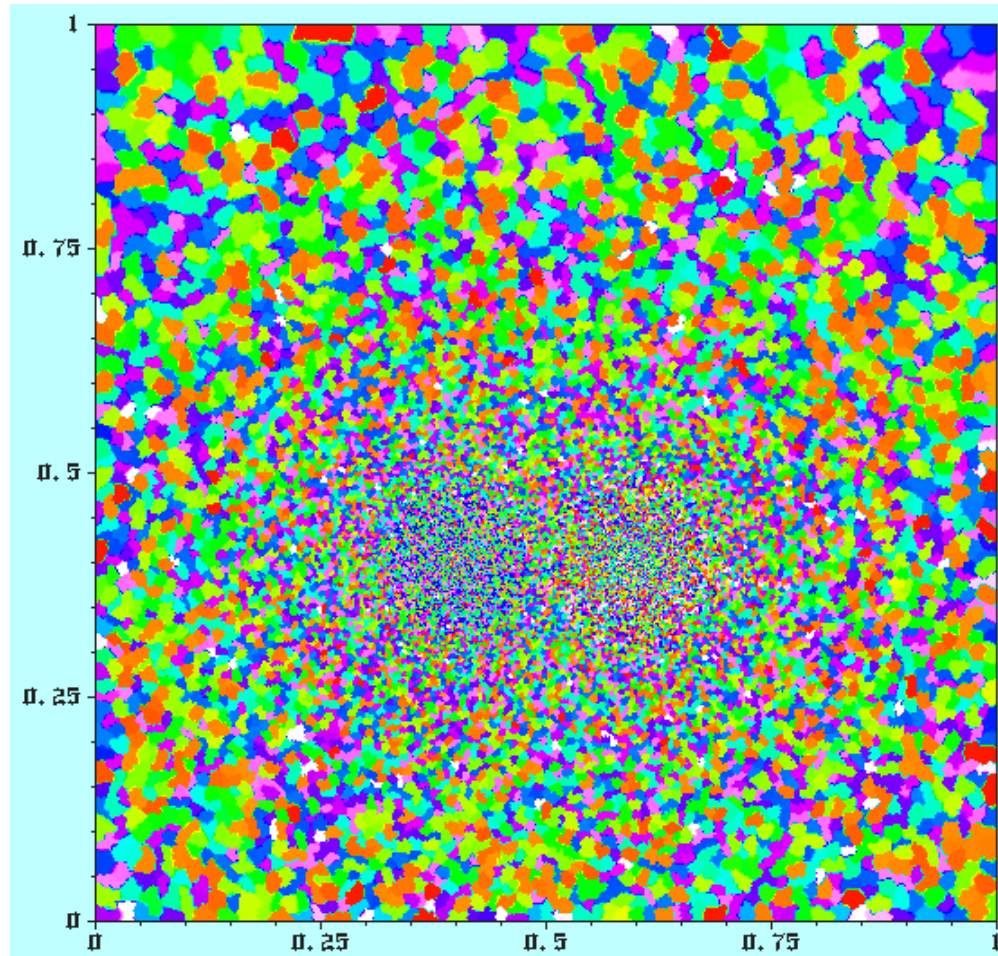
# Формирование макрографа



# Сетка микродоменов

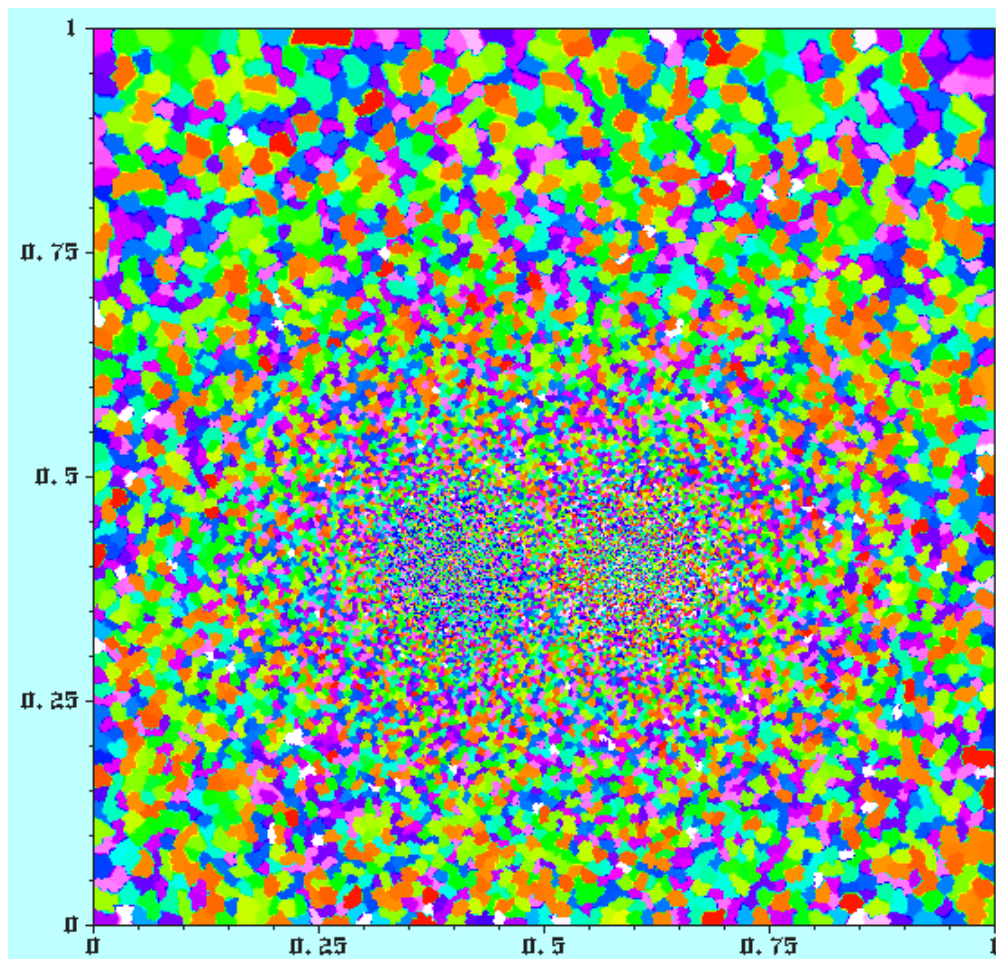
**51 538 микродоменов**

**в каждом около  
20 узлов**



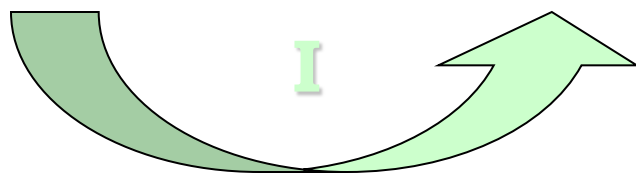
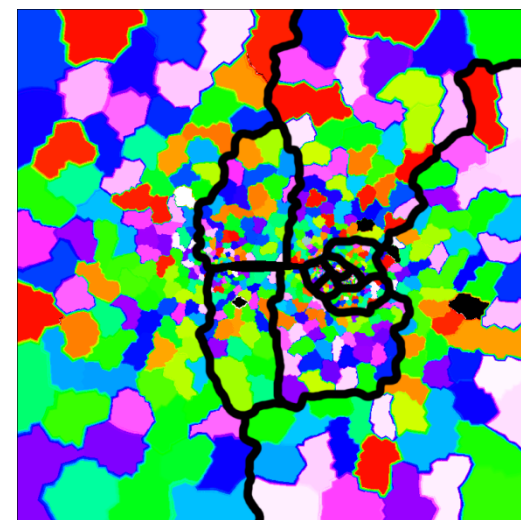
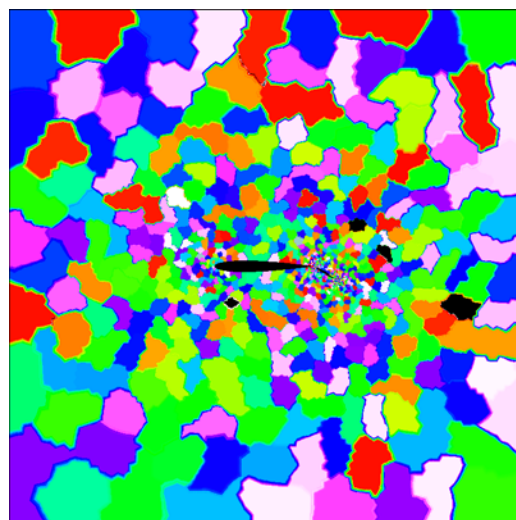
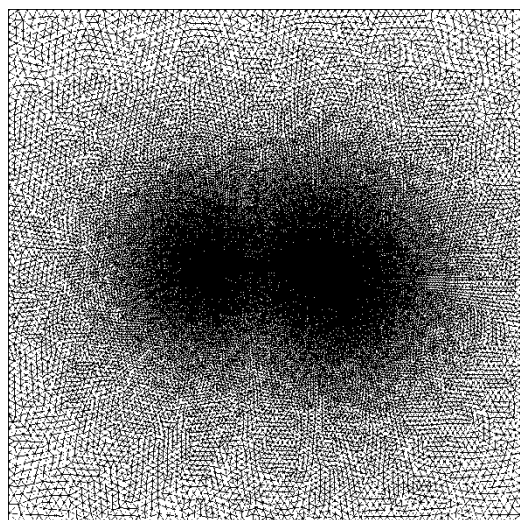
# Сетка микродоменов

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



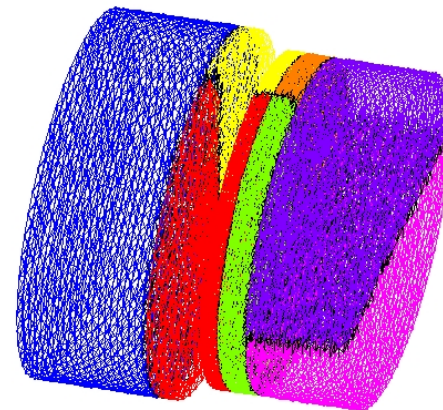
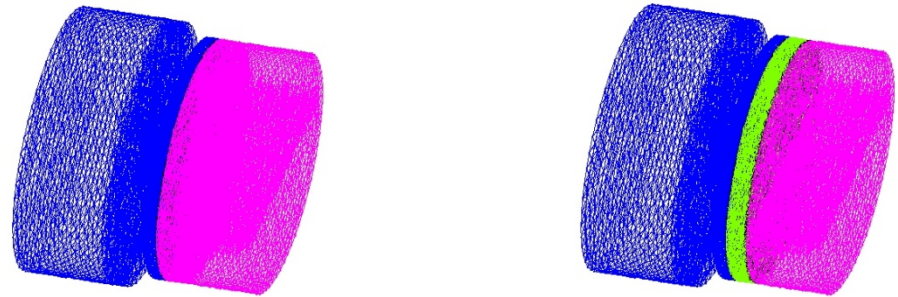
Сетка предварительно  
разбивается на большое число  
*микродоменов*,  
образующих *макрограф*

Вершины *макрографа*  
распределяются по процессорам

Метод эффективен для сверхбольших сеток

# Параллельный алгоритм геометрической декомпозиции сеток

- рекурсивная  
координатная  
бисекция

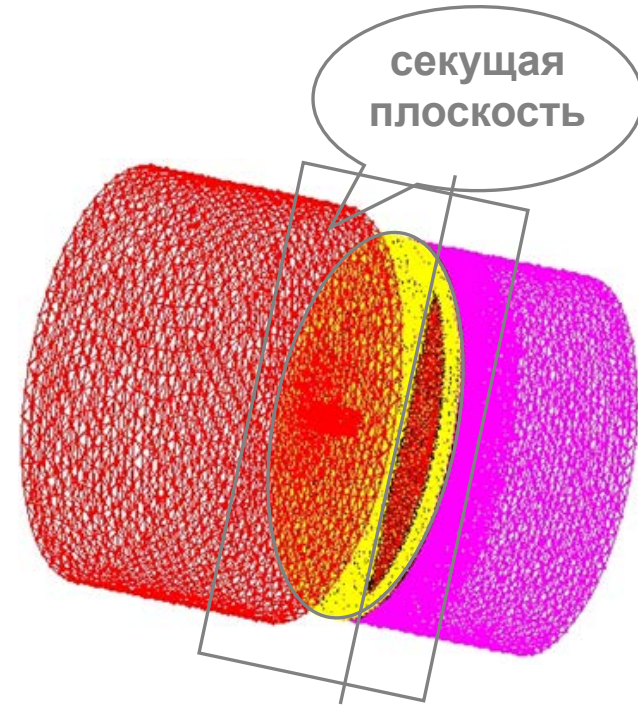


используется распределенная сортировка

# Параллельный алгоритм геометрической декомпозиции сеток:

## Отличия

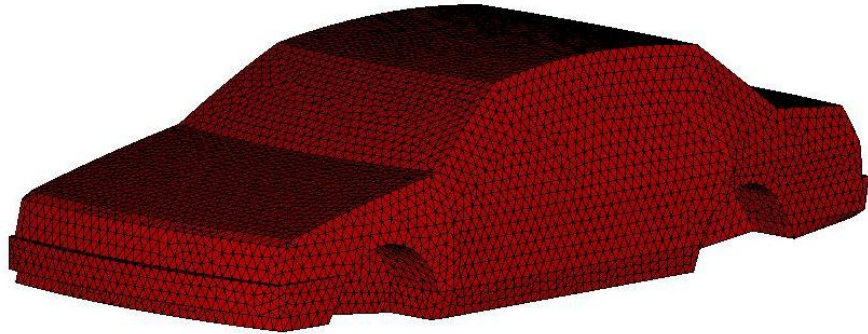
- **секущая плоскость (медиана) разрезается по нескольким координатам**
- **в локальной рекурсивной бисекции сортируются только координаты вершин вблизи медианы**



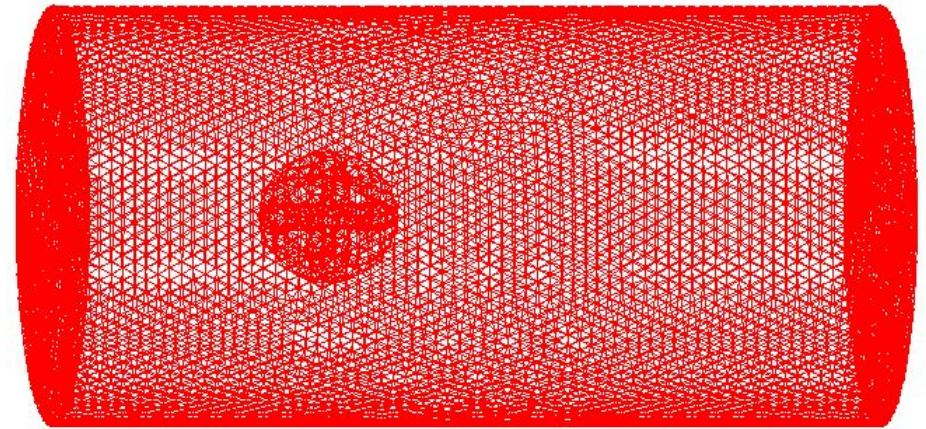
## Достоинство

- **разница числа вершин в доменах не превышает единицу**

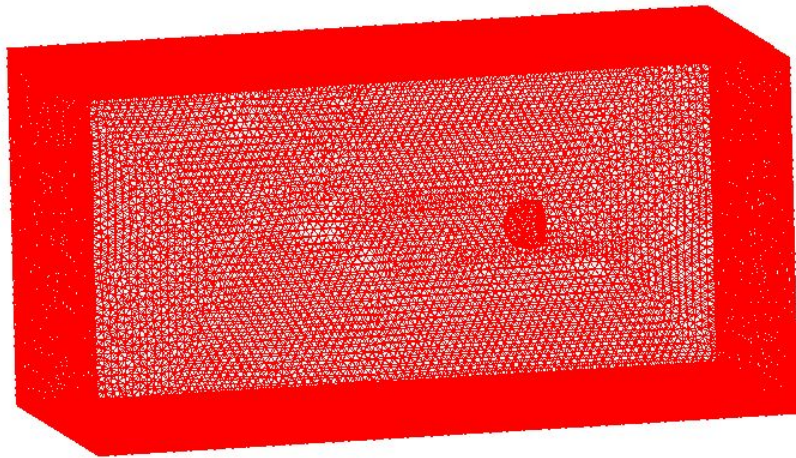
# Тетраэдральные сетки



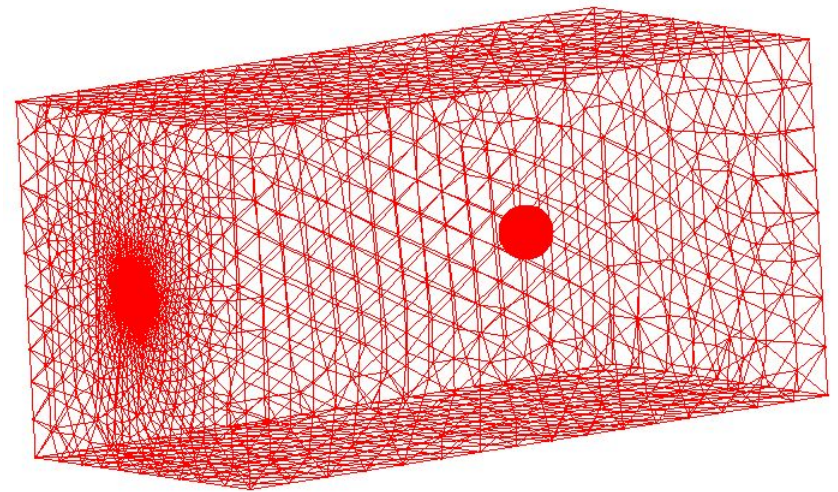
$2 \cdot 10^8$  вершин,  $1.46 \cdot 10^9$  ребер



$2.6 \cdot 10^8$  вершин,  $1.8 \cdot 10^9$  ребер



$10^8$  вершин,  $7.7 \cdot 10^8$  ребер



$2.7 \cdot 10^8$  вершин,  $1.9 \cdot 10^9$  ребер

# Разбиения на микродомены

дисбаланс числа вершин в 25600 микродоменах, %

Методы	Сетка 1	Сетка 2	Сетка 3	Сетка 4
разбиение графов				
IncrDecomp	3,5	0,1	0,3	0,2
PartKway	53,4	59,8	58,6	64,3
PartGeomKway	48,7	50,4	62,4	56,5
PT-Scotch	8,3	8,3	8,3	8,3
геометрические методы				
GeomDecomp	0,01	0,01	0,02	0,01
RCB	0,01	0,01	0,02	0,01

# Разбиения на микродомены

## число несвязных микродоменов из 25600

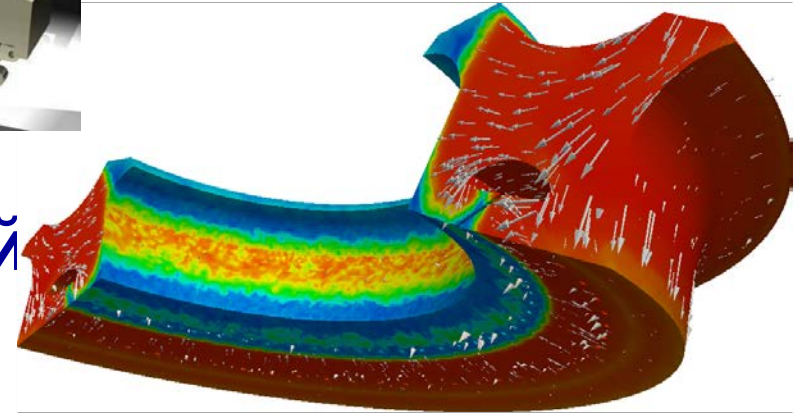
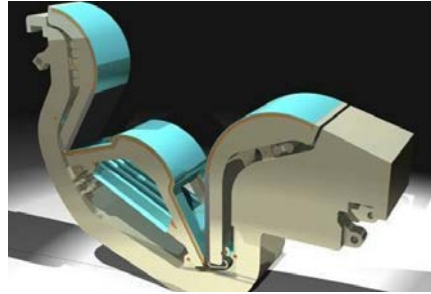
Методы	Сетка 1	Сетка 2	Сетка 3	Сетка 4
разбиение графов				
IncrDecomp	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
PartKway	<b>69</b>	<b>35</b>	<b>37</b>	<b>29</b>
PartGeomKway	<b>67</b>	<b>34</b>	<b>28</b>	<b>37</b>
PT-Scotch	<b>7</b>	<b>0</b>	<b>2</b>	<b>4</b>
геометрические методы				
GeomDecomp	<b>62</b>	<b>38</b>	<b>16</b>	<b>33</b>
RCB	<b>64</b>	<b>43</b>	<b>14</b>	<b>44</b>

# Пакет MARPLE3D

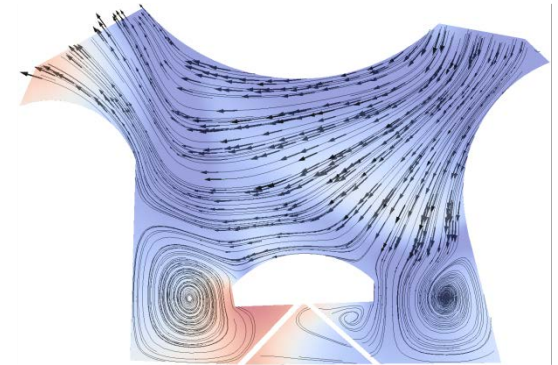
(ИПМ им. М.В.Келдыша РАН)

- предметная область –  
двухтемпературная радиационная  
магнитная гидродинамика
- На задачах газовой динамики проведено тестирование разбиений, полученных методами пакетов GridSpiderPar, ParMETIS, Zoltan и PT-Scotch
- Сравнивалась эффективность параллельного счета физических задач пакетом MARPLE3D при распределении сеток по ядрам в соответствии с различными разбиениями

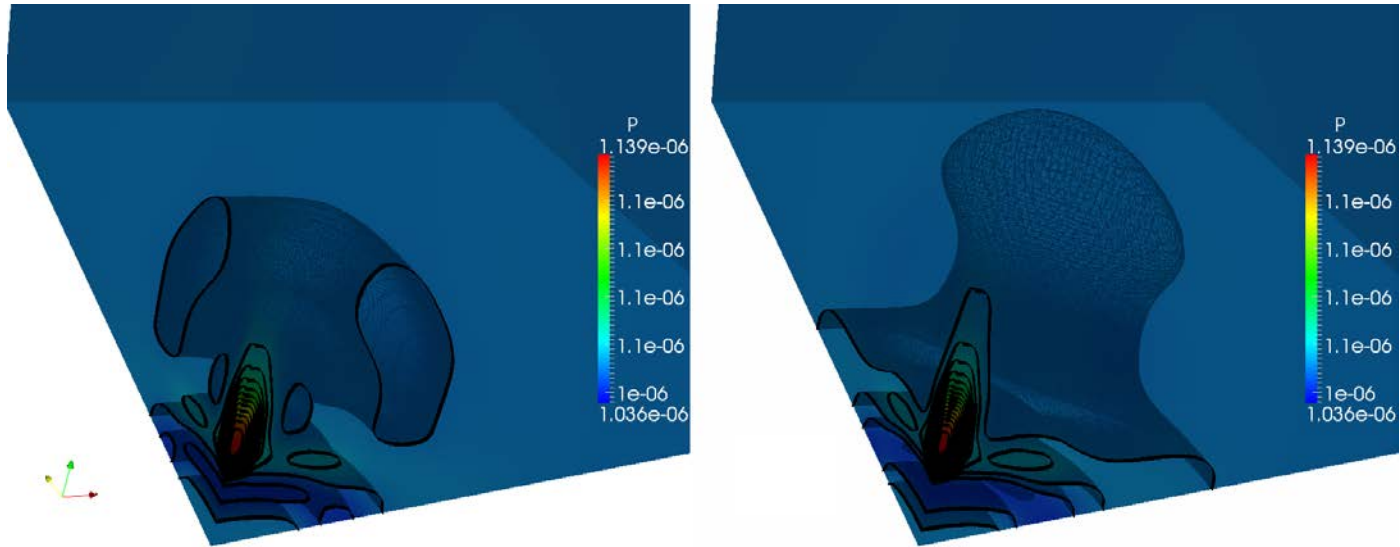
# Моделирование газоплазменных потоков в диверторе токамака ITER



- Полная система уравнений газовой динамики
- Турбулентная вязкость
- Радиационный и кондуктивный теплоперенос
- Явные и неявные схемы



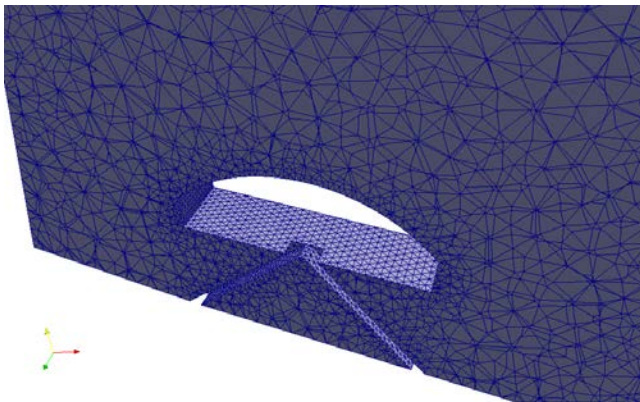
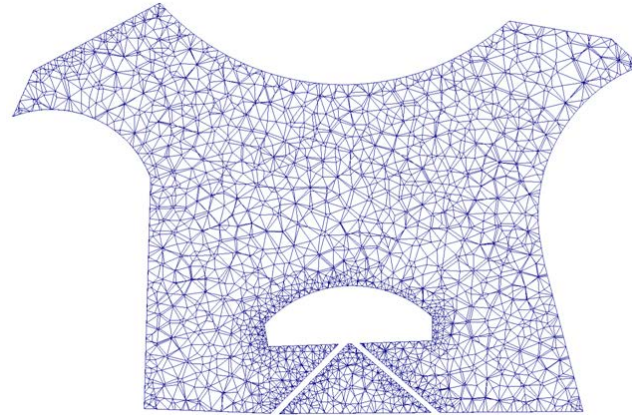
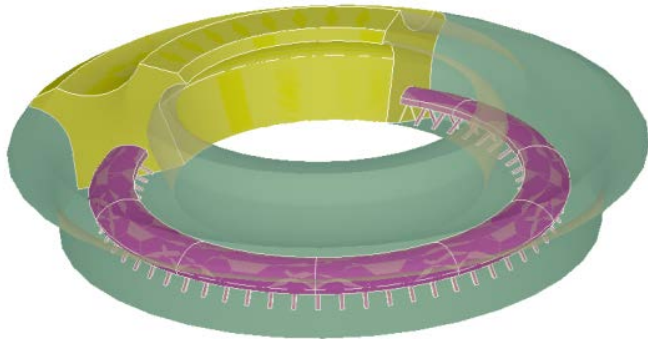
# Моделирование распространения ударной волны от приземного источника энергии взрывного типа



- Полная система уравнений газовой динамики
- Кондуктивный теплоперенос
- Явные и неявные схемы

# Расчетные сетки

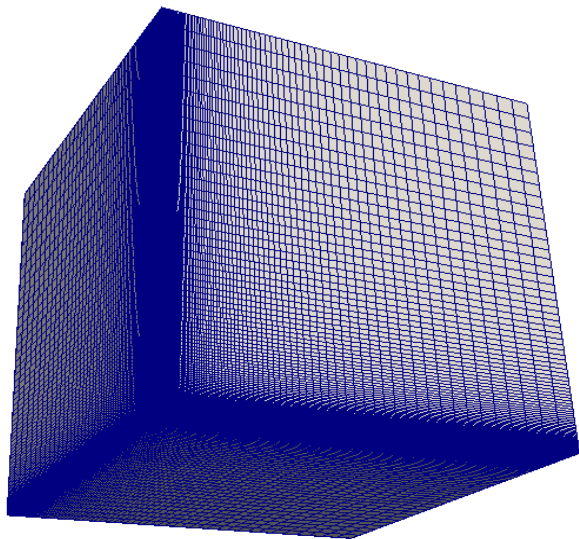
## Дивертор токамака (divertor)



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

# Расчетные сетки

## Приземный взрыв (boom и boomL)



Гексаэдральные сетки

61 миллион ячеек для “boom”

116 миллионов ячеек для “boomL”

Параллелепипеды с  
различными aspect ratio

boom:

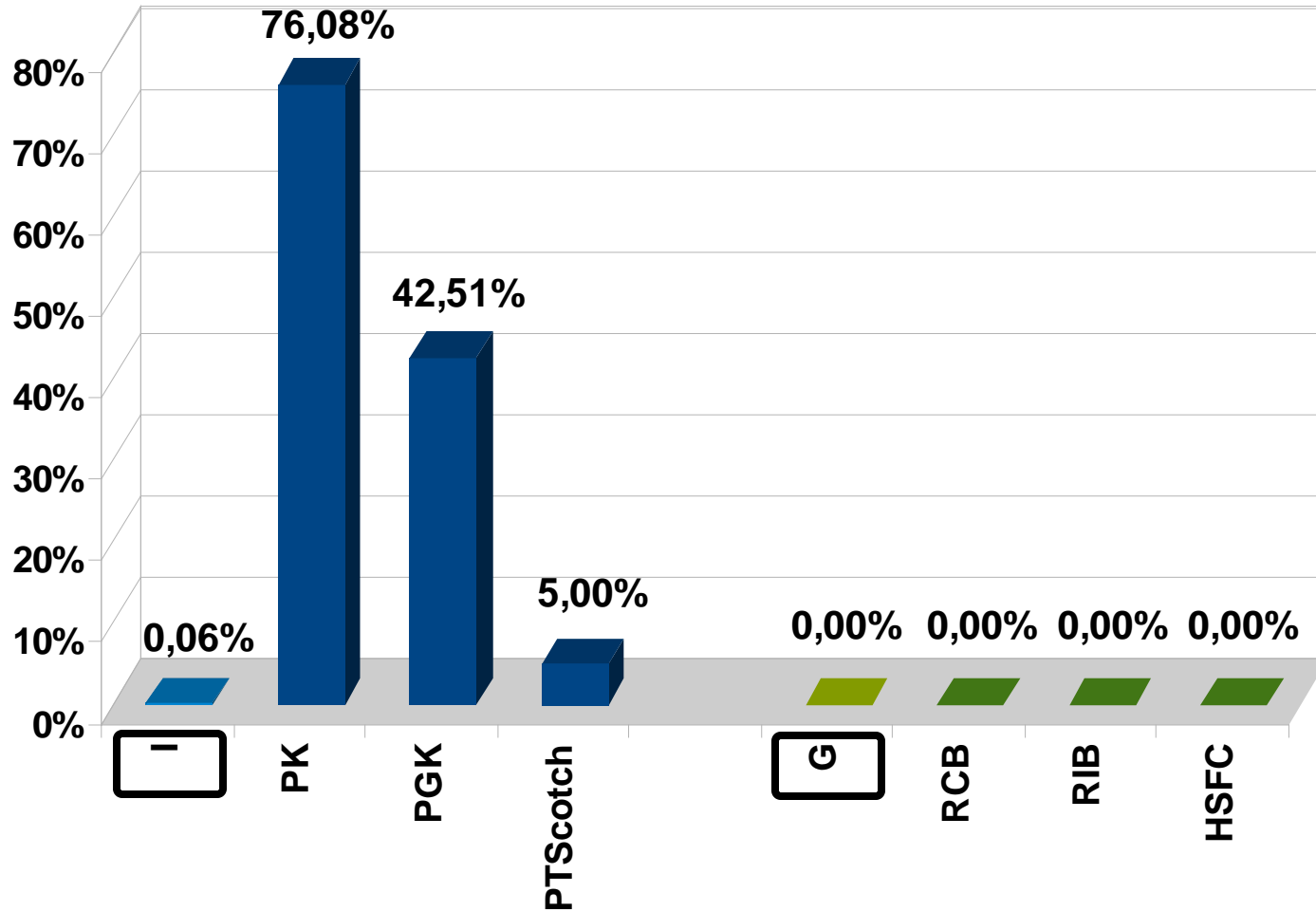
- 4096 доменов

boomL:

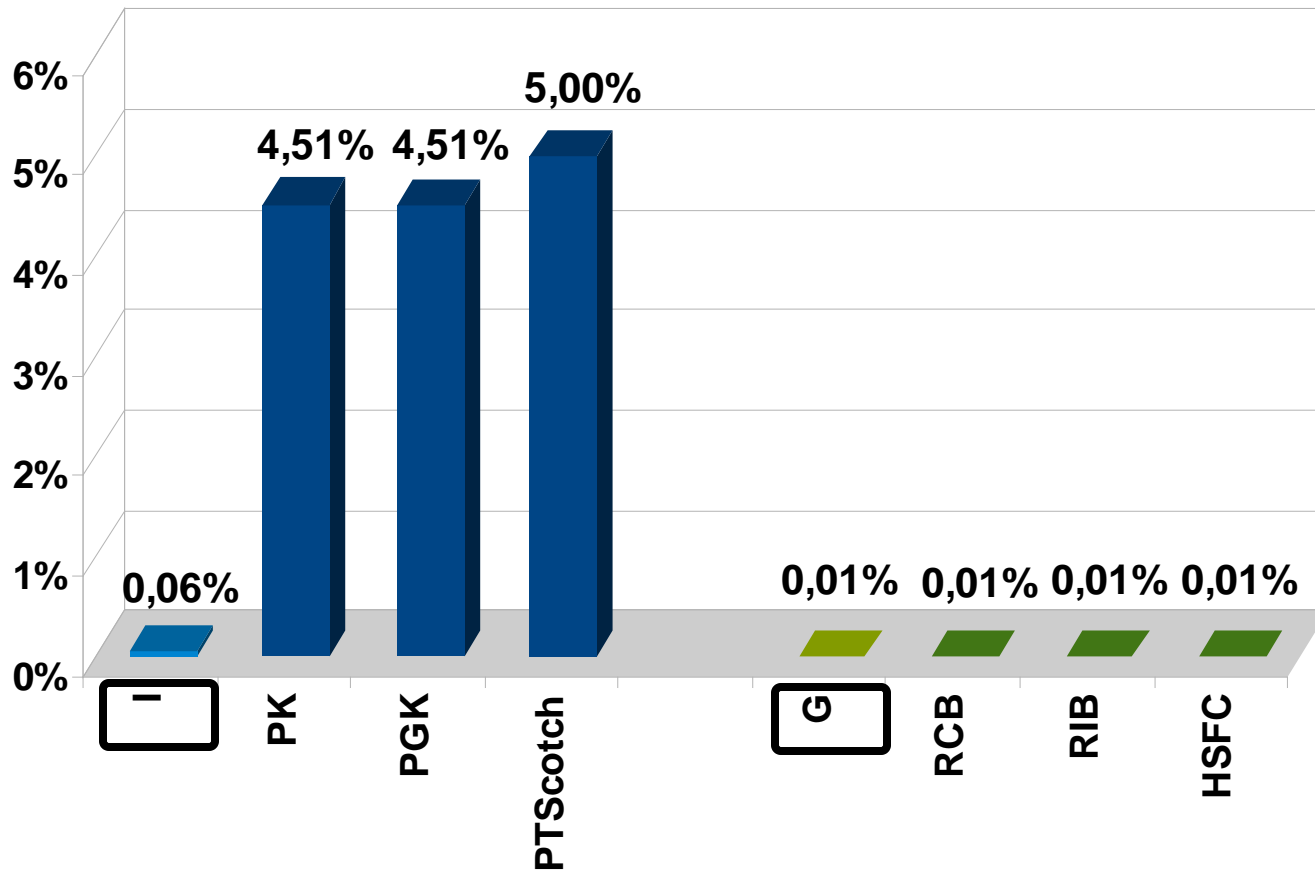
- 10080 доменов

- 
- Для всех расчетных сеток были построены дуальные графы с числом вершин  $2.8 \cdot 10^6$  -  $1.2 \cdot 10^8$  и числом ребер  $2.3 \cdot 10^7$  -  $1.0 \cdot 10^9$
  - Вычисления проводились на кластерах MBC-100K (227,94 TFlop/s), "Ломоносов" (1700 TFlops) и «Helios» (1524.1 TFlop/s)

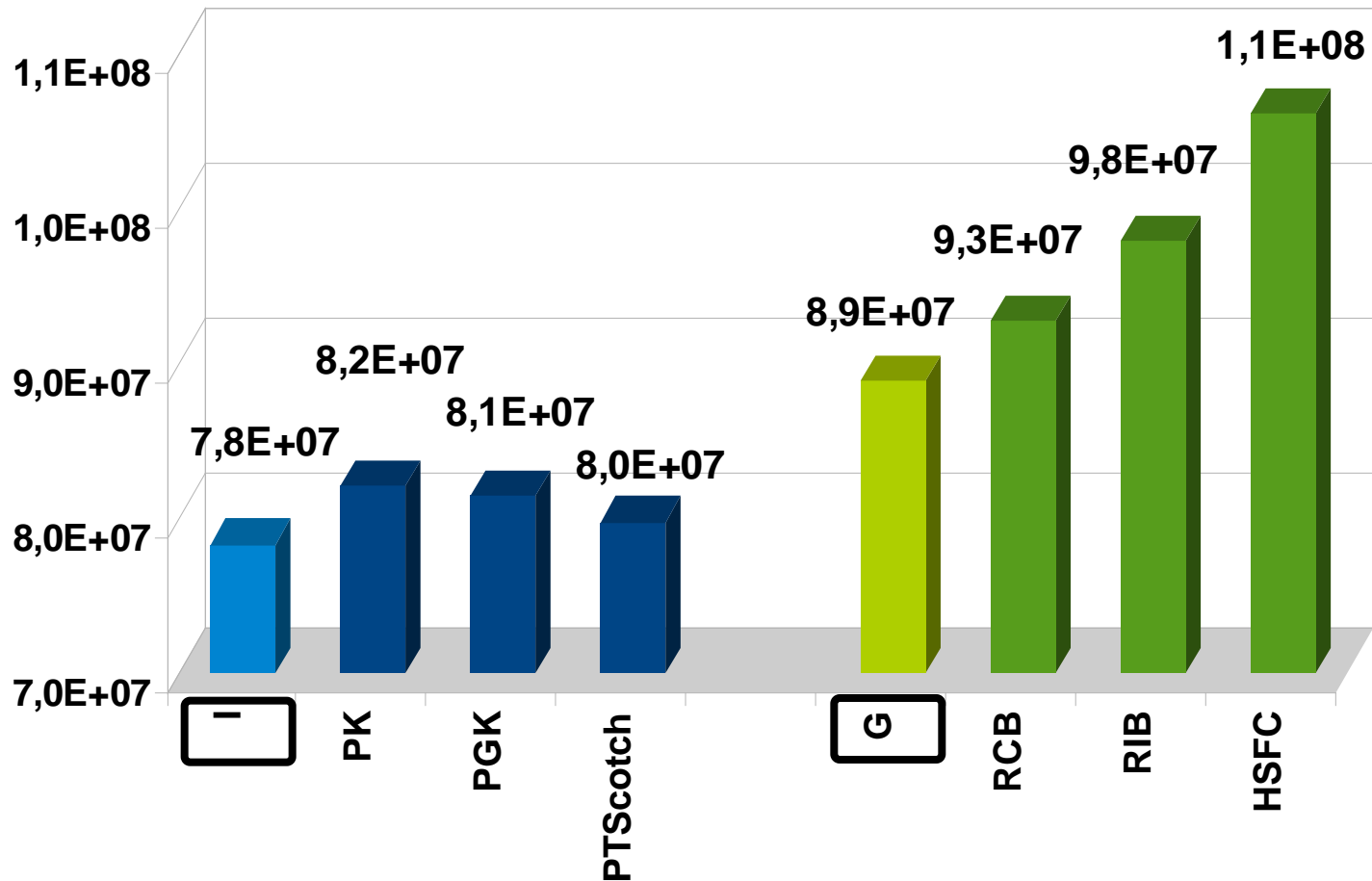
# Дисбаланс числа вершин в доменах: недостаток вершин (boom)



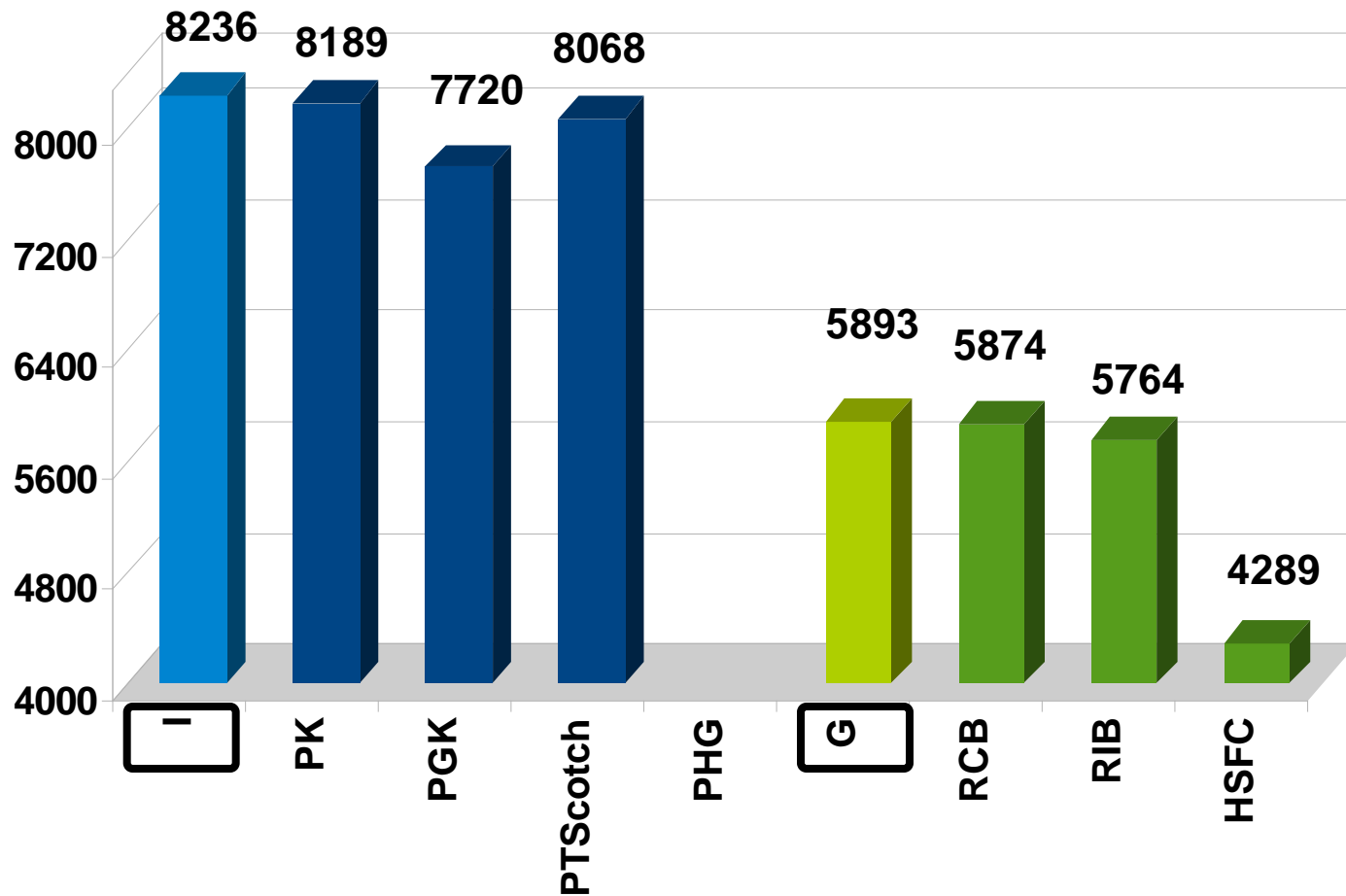
# Дисбаланс числа вершин в доменах: избыток вершин (boom)



# Число разрезанных ребер (boomL)

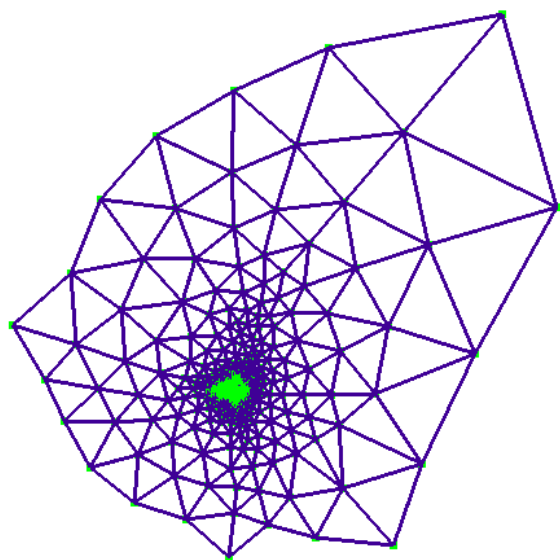
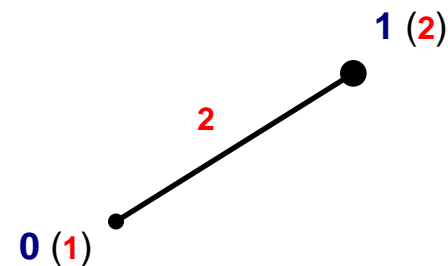
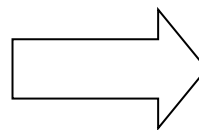
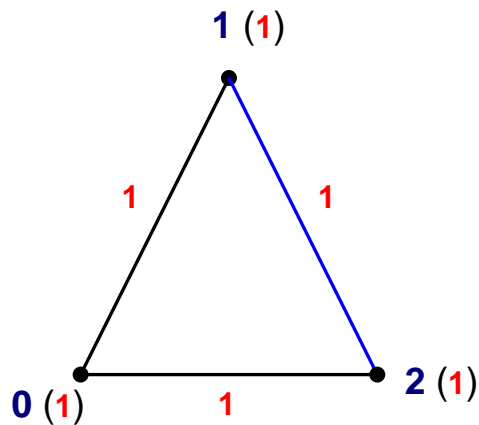


# Число шагов по времени (divertor)

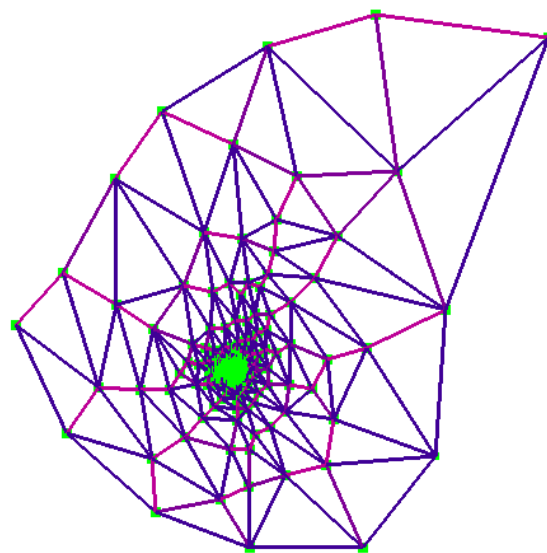


# Огрубление на этапе преддекомпозиции

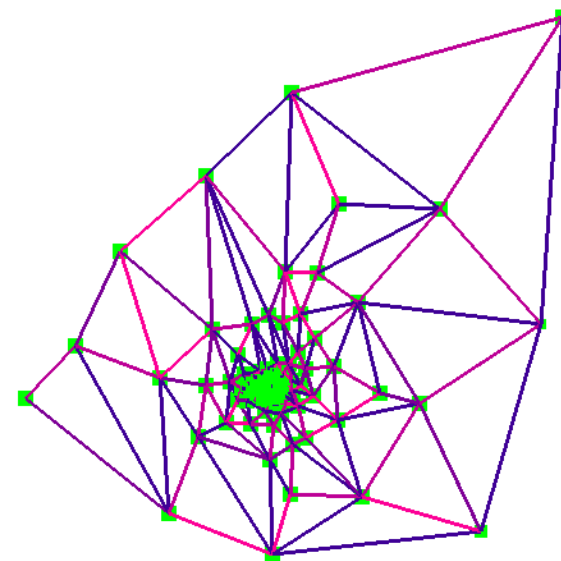
огрубление



исходная сетка



1-ое огрубление



2-ое огрубление

# Литература

1. *Fiedler M.* A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory // Czechoslovak Mathematical Journal. – 25(100) – Praha, 1975. – Pp. 619–633.  
URL: [http://www3.math.tu-berlin.de/Vorlesungen/SS14/MatricesGraphsPDEs/paper\\_for\\_students/CZMathJ-25-1975-Fiedler%20copy.pdf](http://www3.math.tu-berlin.de/Vorlesungen/SS14/MatricesGraphsPDEs/paper_for_students/CZMathJ-25-1975-Fiedler%20copy.pdf)
2. *B. Hendrickson, R. Leland.* A Multilevel Algorithm for Partitioning Graphs // Supercomputing '95 Proceedings. – San Diego, CA, 1995.  
<http://www.leonidzhukov.net/hse/2016/networks/papers/MultilevelAlgorithmPartitioningGraphs.pdf>
3. *George Karypis* Family of Graph and Hypergraph Partitioning Software URL: <http://glaros.dtc.umn.edu/gkhome/views/metis/>
4. *Alex Pothen, Horst D. Simon, and Kang-Pu Paul Liu.* Partitioning Sparse Matrices with Eigenvectors of Graphs. Report RNR-89-009, July 1989. URL: <http://snap.stanford.edu/class/cs224w-readings/Pothen89Partition.pdf>
5. *Е. Н. Головченко, Е. Ю. Дорофеева, В. А. Гасилов, М. В. Якововский.* Вычислительный эксперимент по оценке качества алгоритмов параллельной декомпозиции больших сеток // *Препринты ИПМ им. М.В. Келдыша.* — 2013. — № 7. — С. 1–32. URL: [http://keldysh.ru/papers/2013/prep2013\\_07.pdf](http://keldysh.ru/papers/2013/prep2013_07.pdf)

# Контакты

**Якобовский М.В.**

Член.-корр. РАН, проф., д.ф.-м.н.,  
зам. директора по научной работе  
Института прикладной математики им.  
М.В.Келдыша Российской академии наук

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

web: <http://lira.imamod.ru>