

Рациональная декомпозиция сеток

Якововский М.В.

lira@imamod.ru

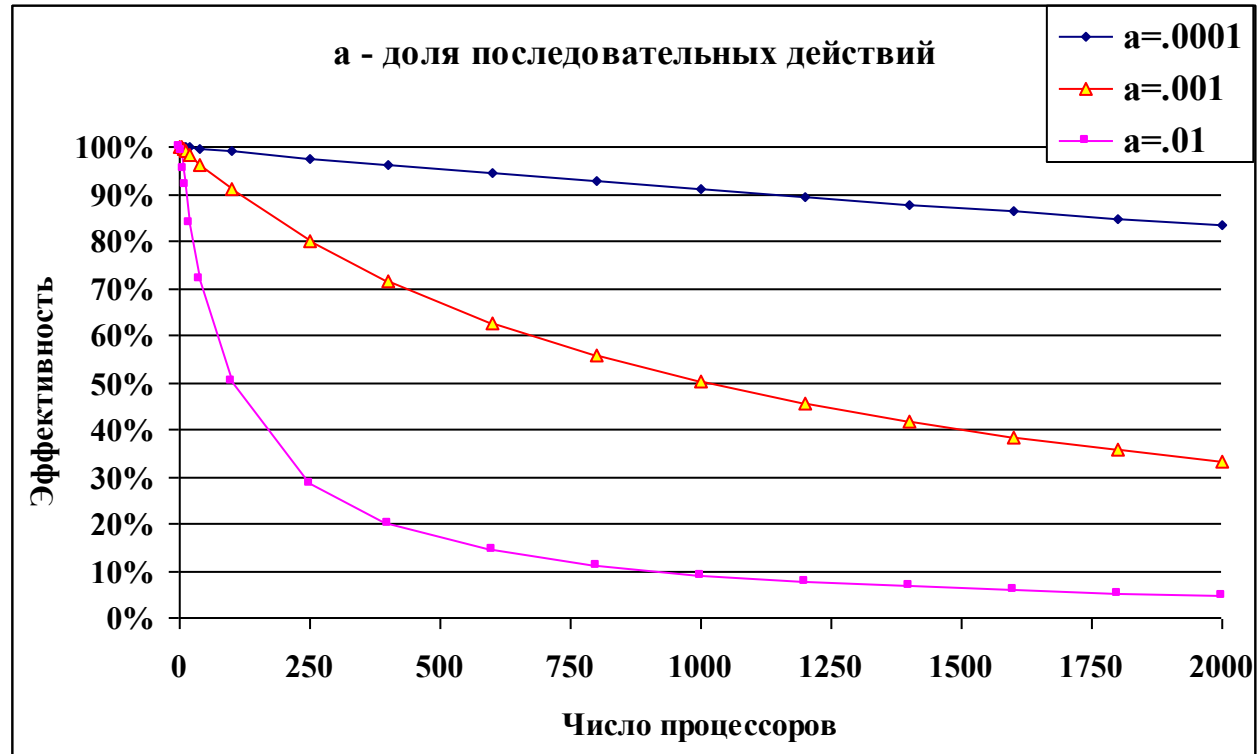
[www://lira.imamod.ru](http://www.lira.imamod.ru)

Ограничения

- Закон Амдаля

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

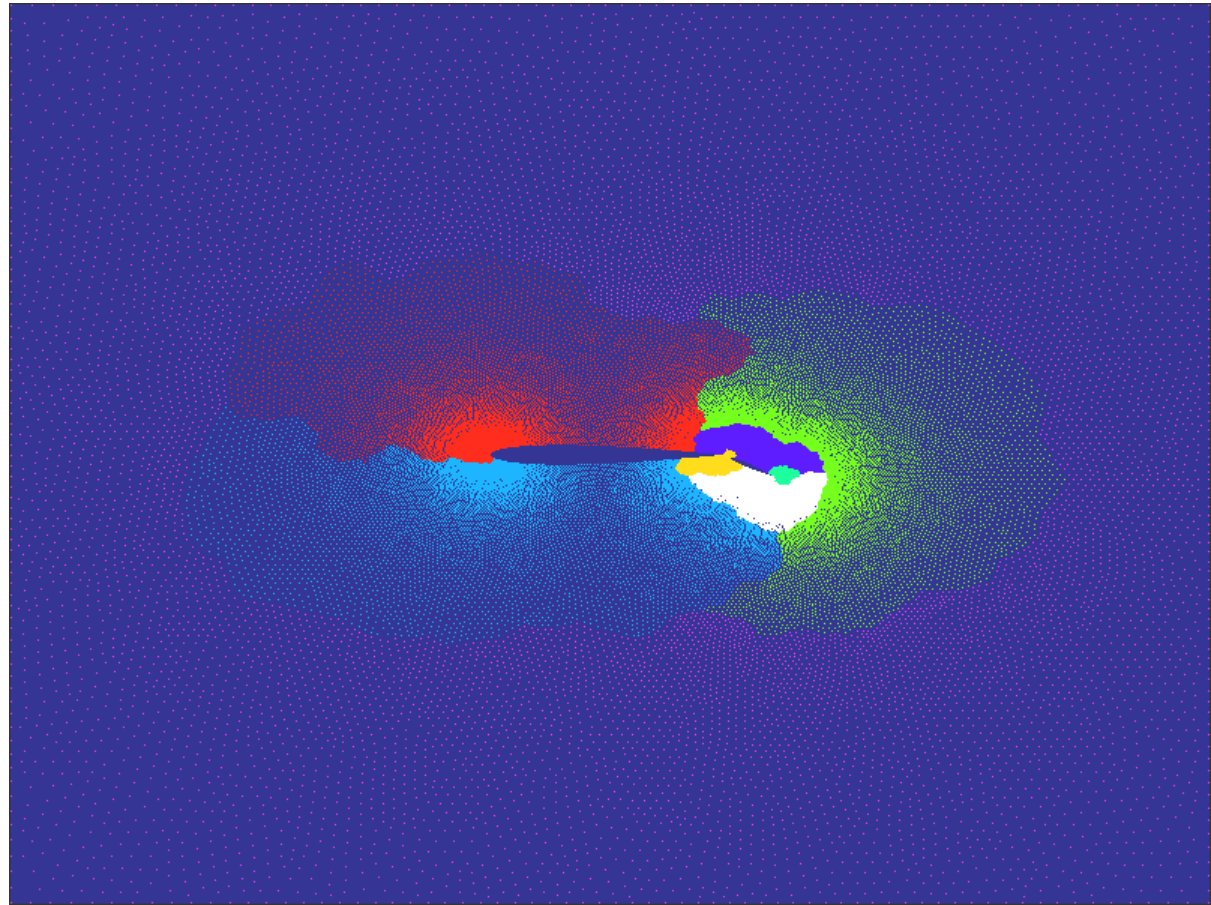
$$E(p) = \frac{1}{1 + a(p-1)}$$

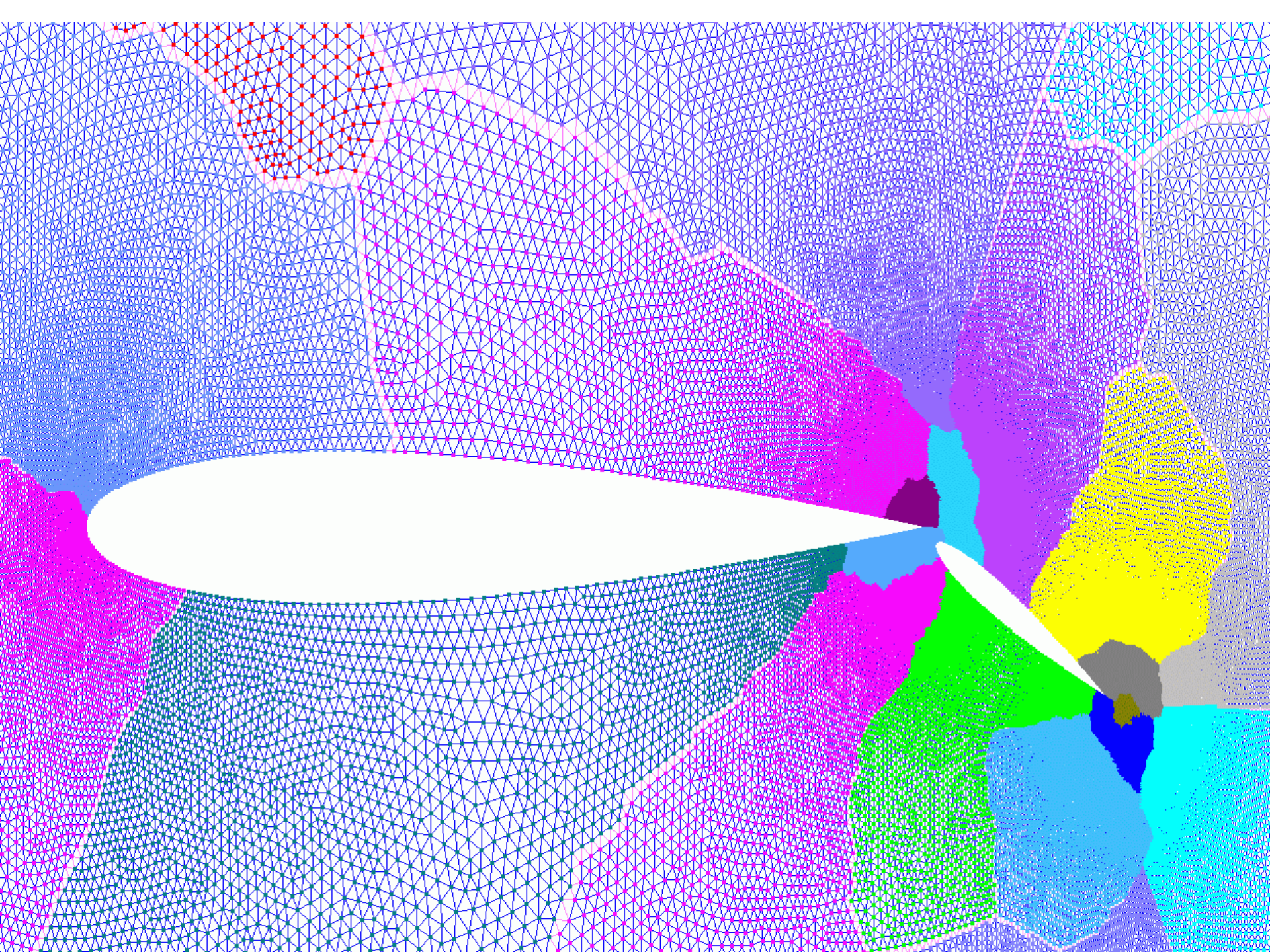


- Пакетный режим исполнения и отладки приложений
- Процедуры авторизованного доступа к удаленным системам
- Высокая динамика изменения конфигурации суперкомпьютеров
- Несоизмеримость ресурсов рабочей станции пользователя и суперкомпьютера

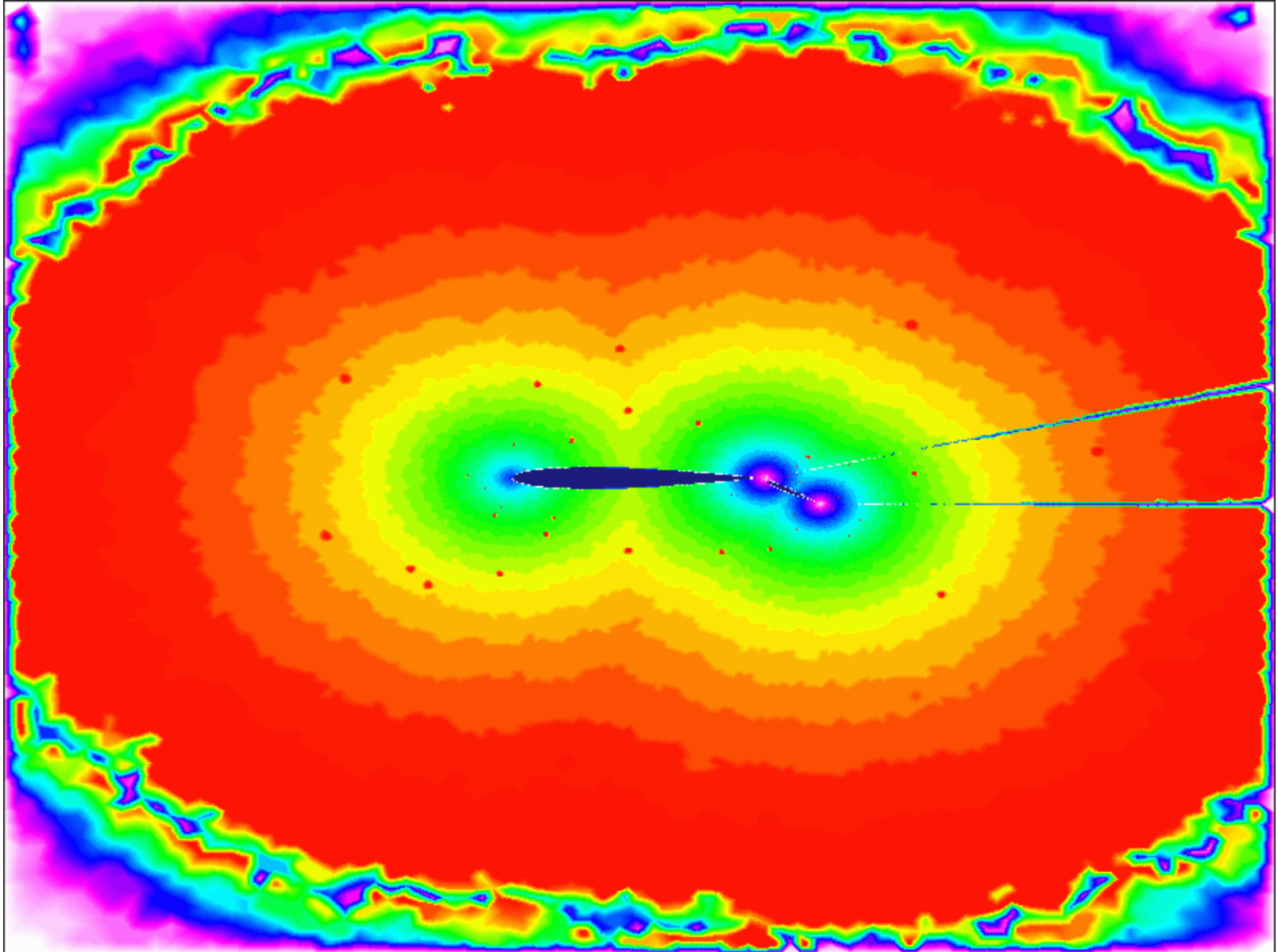
Статическая балансировка загрузки

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

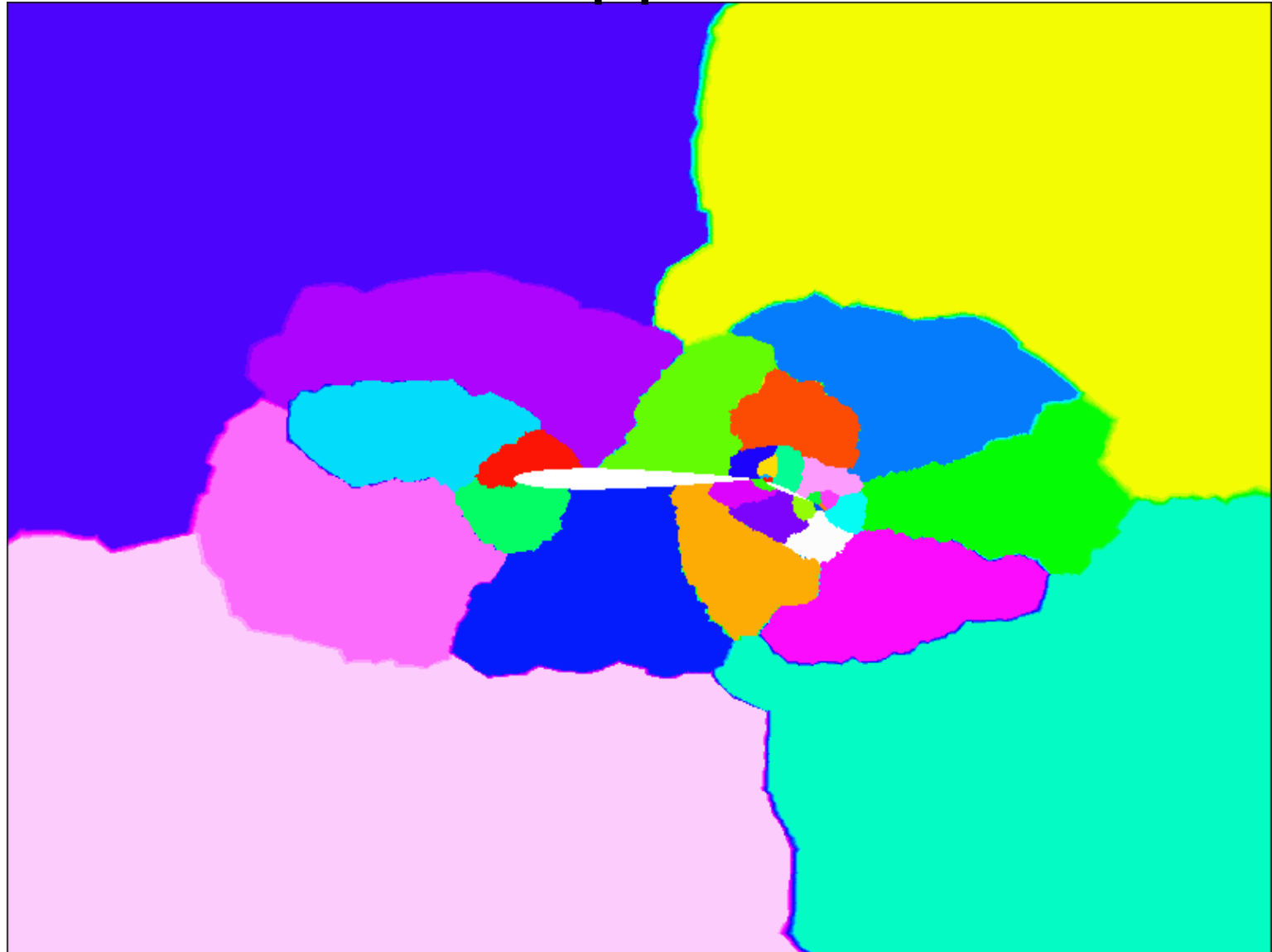




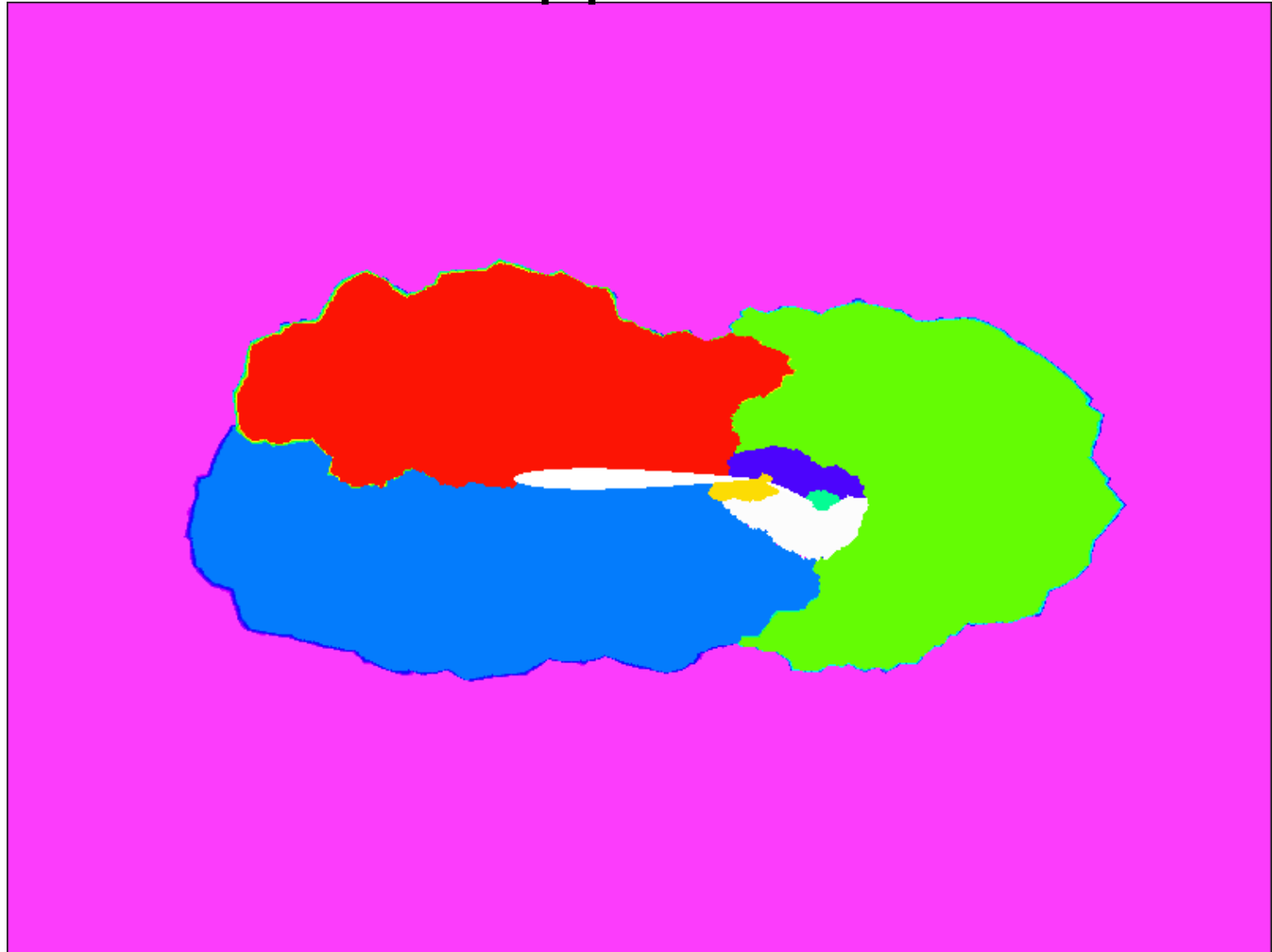
Простое разбиение на 32 домена



Рациональное разбиение на 32 домена



Рациональное разбиение на 8 доменов



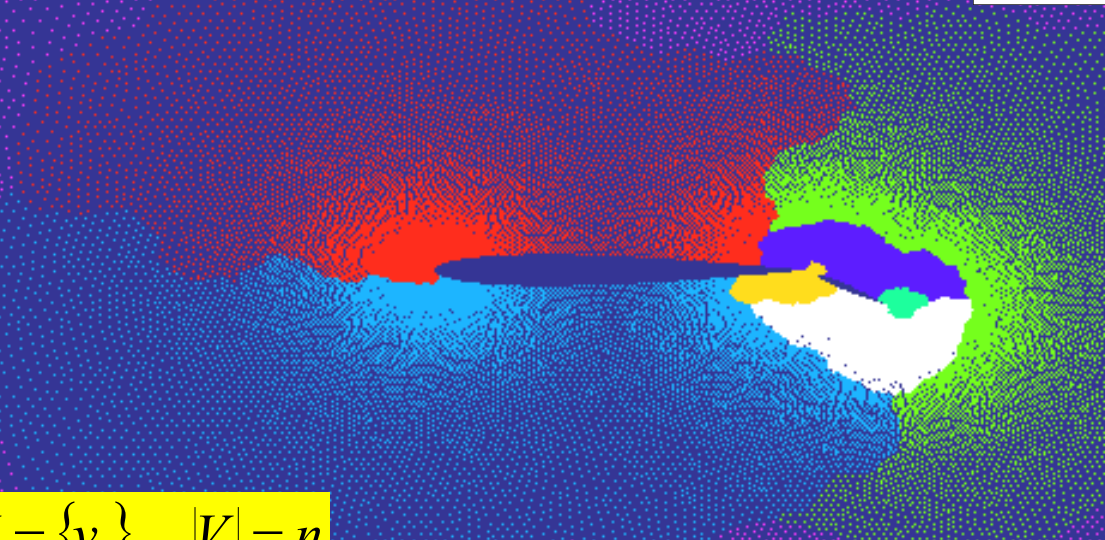
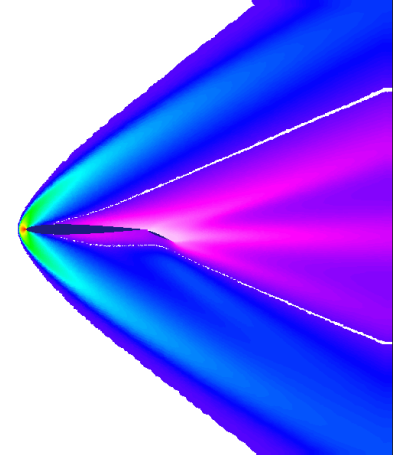
- Статическая балансировка загрузки

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

$$R(V) = (V_1, \dots, V_p)$$

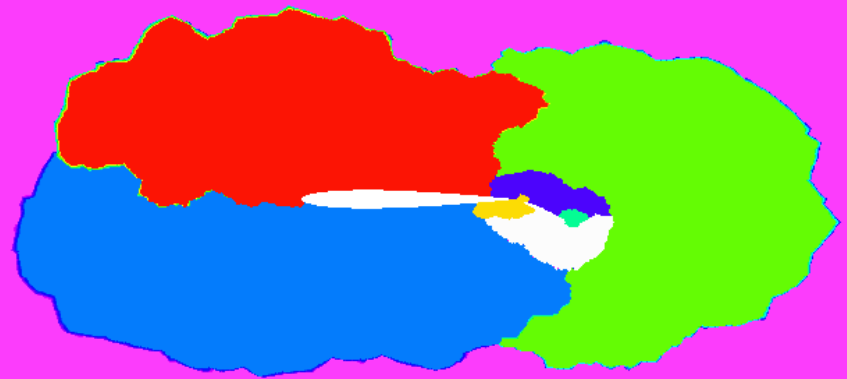
$$V = \bigcup_{k=1}^p V_k, \quad V_i \cap V_j = \emptyset, \quad i \neq j$$

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



Критерии декомпозиции графов

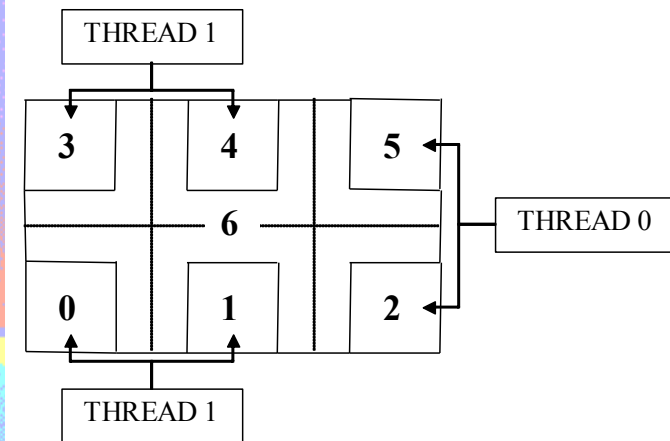
- Равномерное распределение суммарного веса узлов/рёбер
- Минимизация максимального веса исходящих из домена ребер / Минимизация суммарного веса разрезанных ребер



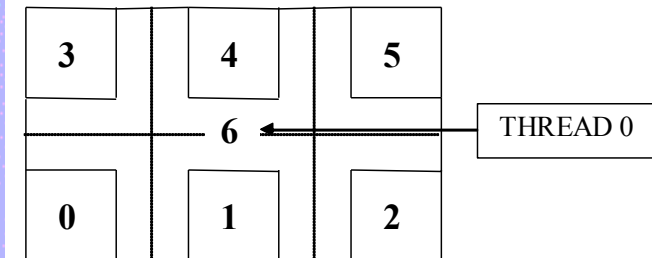
Критерии декомпозиции графов

- Исключение связей между доменами
- Минимизация суммарного веса разрезанных ребер

ПАРАЛЛЕЛЬНЫЙ РЕЖИМ

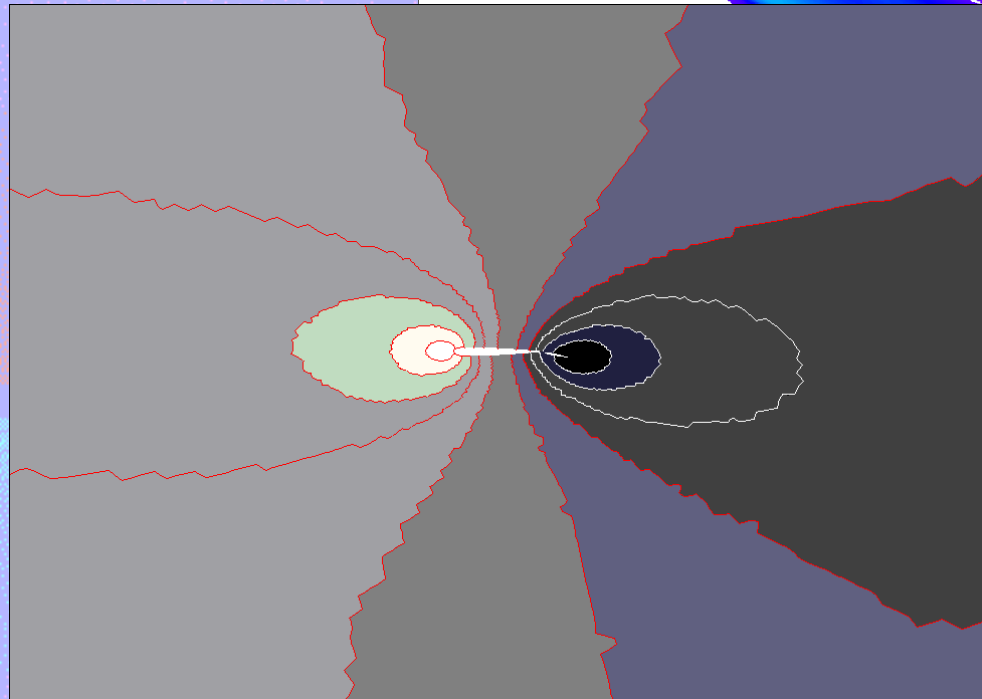


ПОСЛЕДОВАТЕЛЬНЫЙ РЕЖИМ



Критерии декомпозиции графов

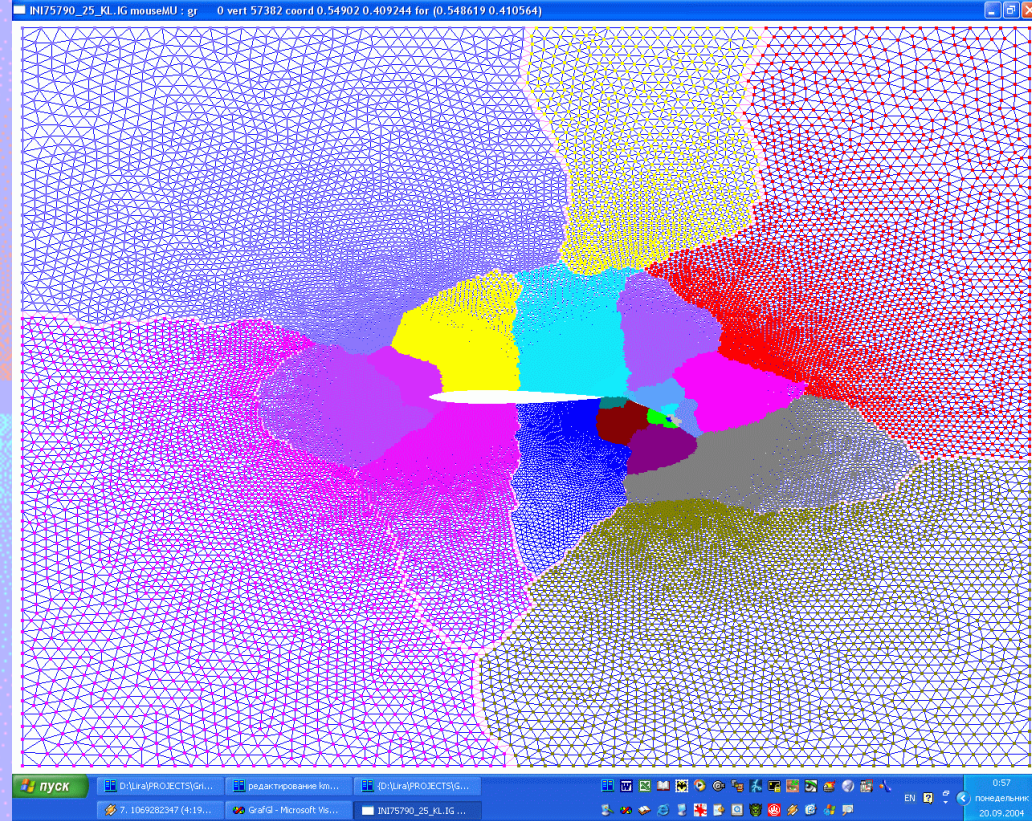
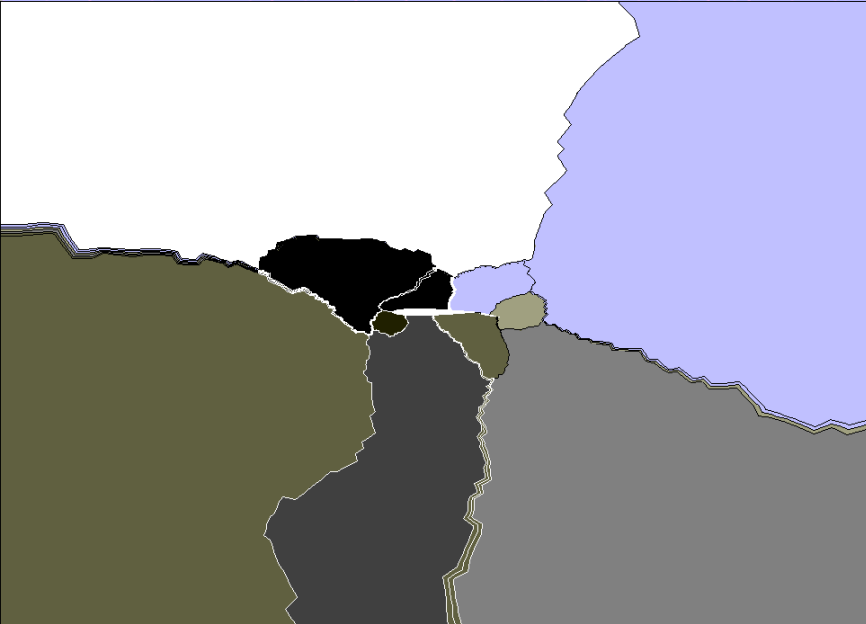
- Минимизация максимальной степени доменов



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

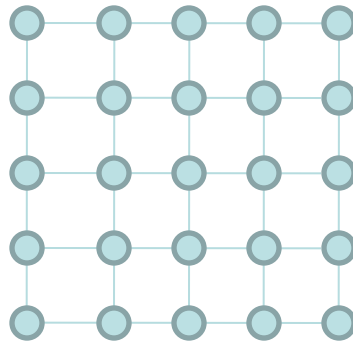
Критерии декомпозиции графов



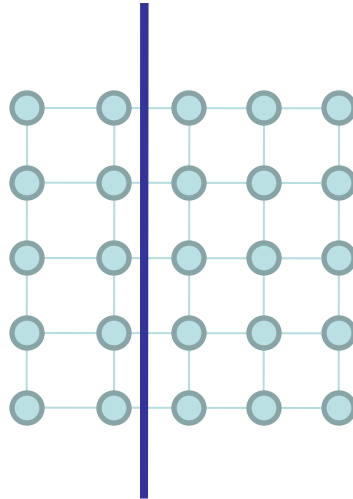
- Обеспечение связности доменов
- Обеспечение связности множества внутренних узлов доменов

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

- Разрезать решетку 5 x 5 на 4 части

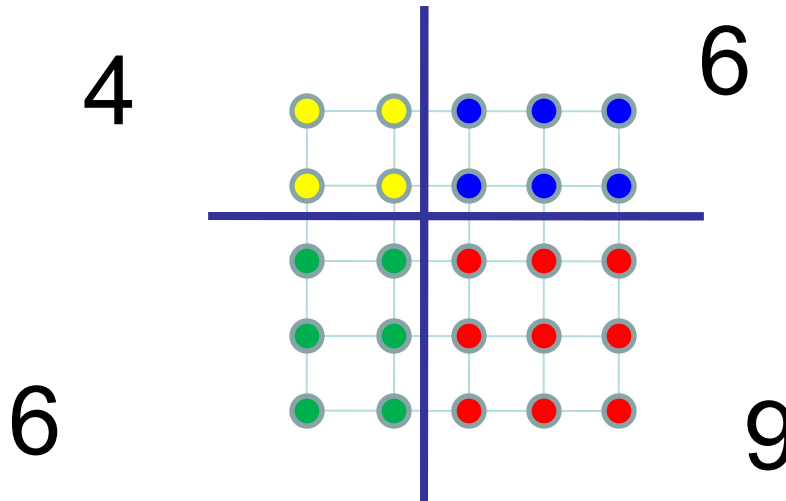


Декомпозиция сетки из 25 узлов на 4 части



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

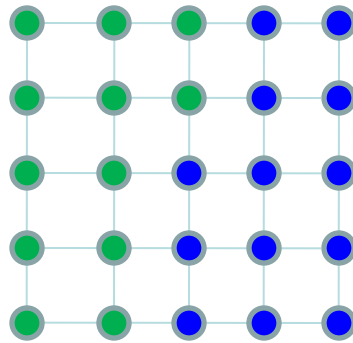
- Декомпозиция решетки 5 x 5 на 4 домена



- Дисбаланс $9/4=2.25$

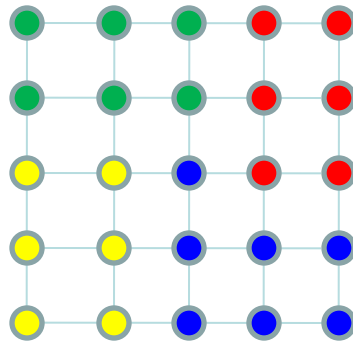
$$25/4 = 4 ? 6 ? 9$$

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



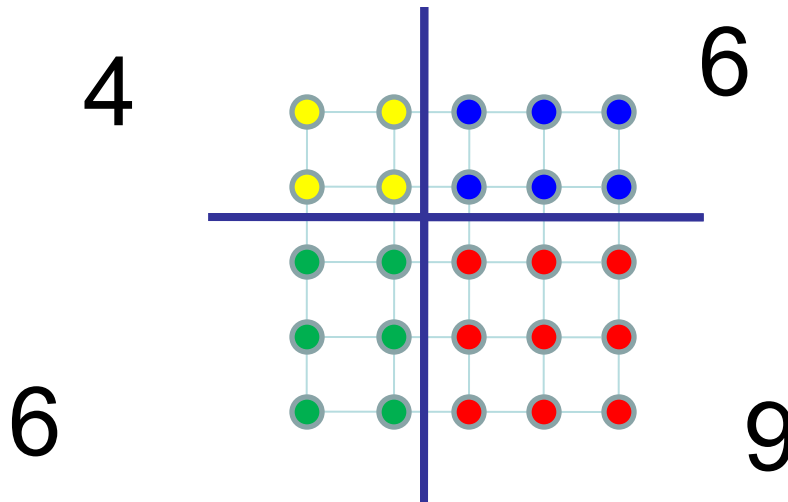
$$25/4 = 4 ? 6 ? 9$$

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



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

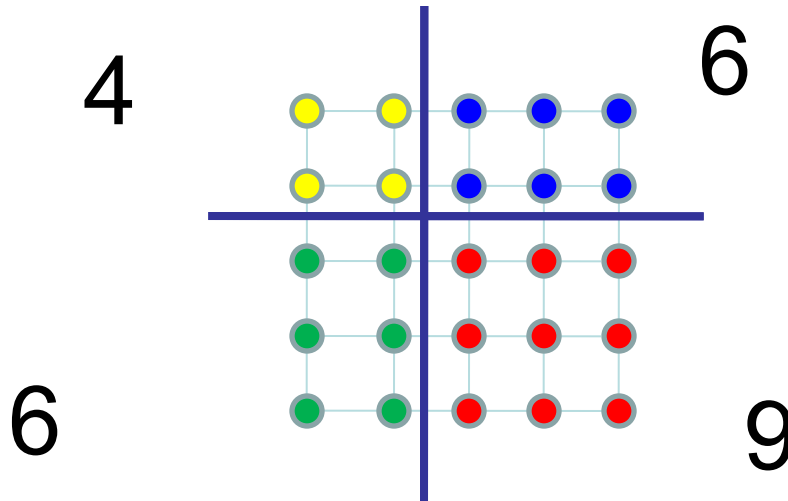
- Декомпозиция решетки 5 x 5 на 4 домена



- Дисбаланс $9/4=2.25$

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

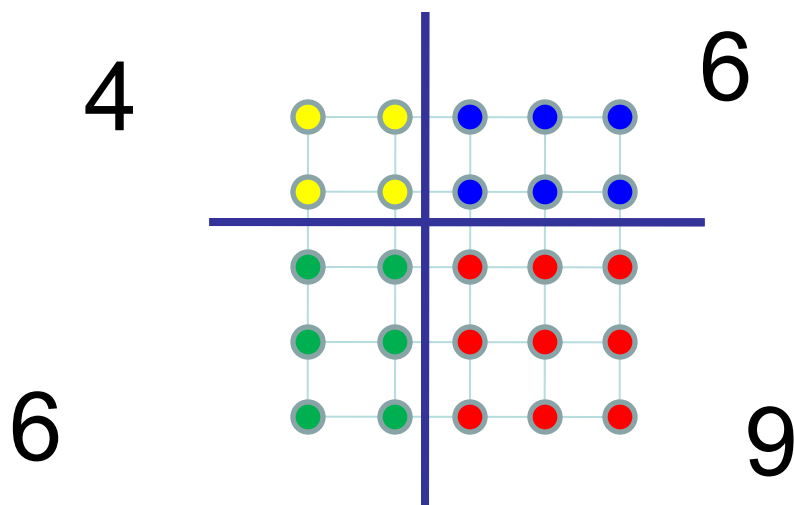
- Декомпозиция решетки 5 x 5 на 4 домена



- ~~• Дисбаланс $9/4=2.25$~~

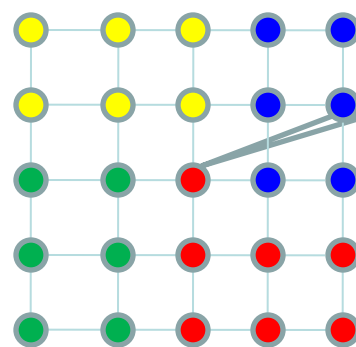
$$25/4 = 4 ? 6 ? 9$$

- Декомпозиция решетки 5 x 5 на 4 домена



Потери

$$9/6.25=1.44$$

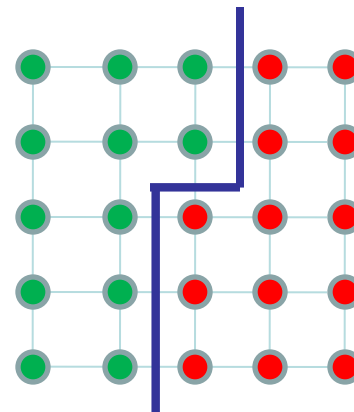
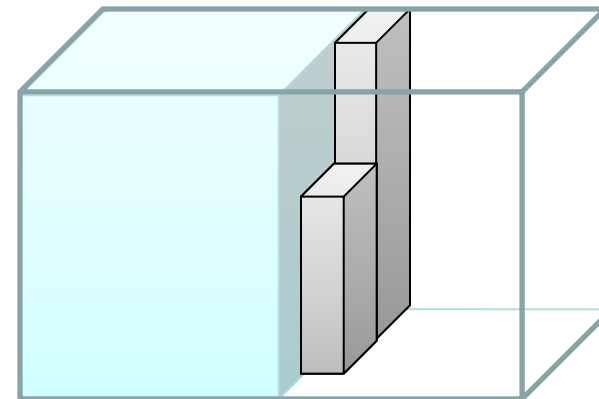
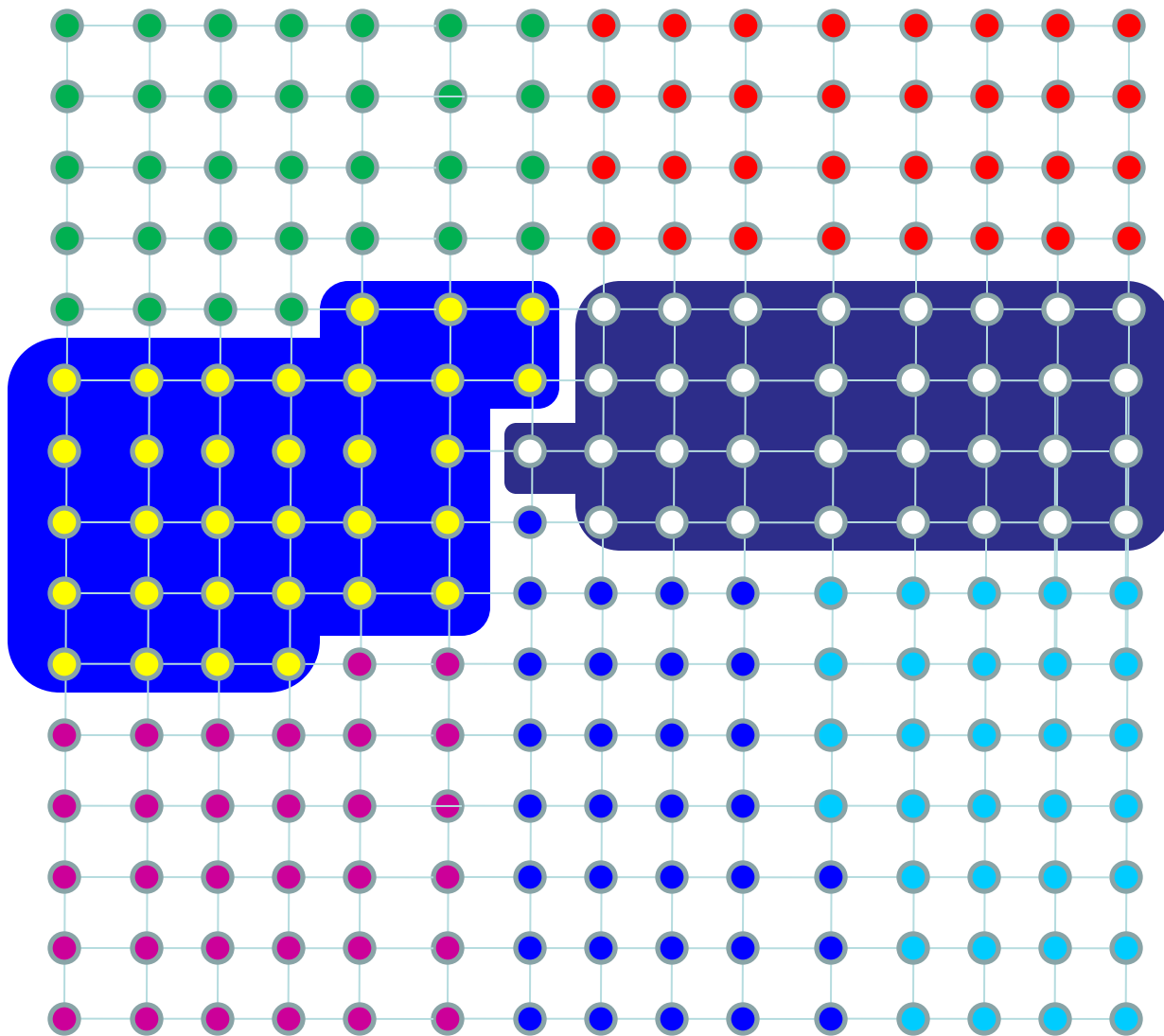


Потери

$$9/7=1.29$$

- ~~Дисбаланс $9/4=2.25$~~

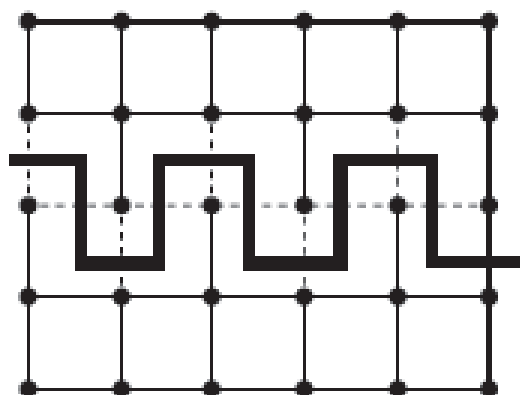
Декомпозиция сетки 25x25 на 7 частей



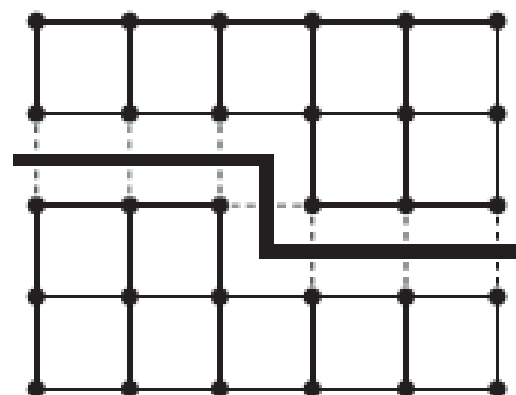
Разрезано ребер

10

7



$$y[i-1] \leq y[i]$$



$$(y[i-1] < y[i])$$

МММ

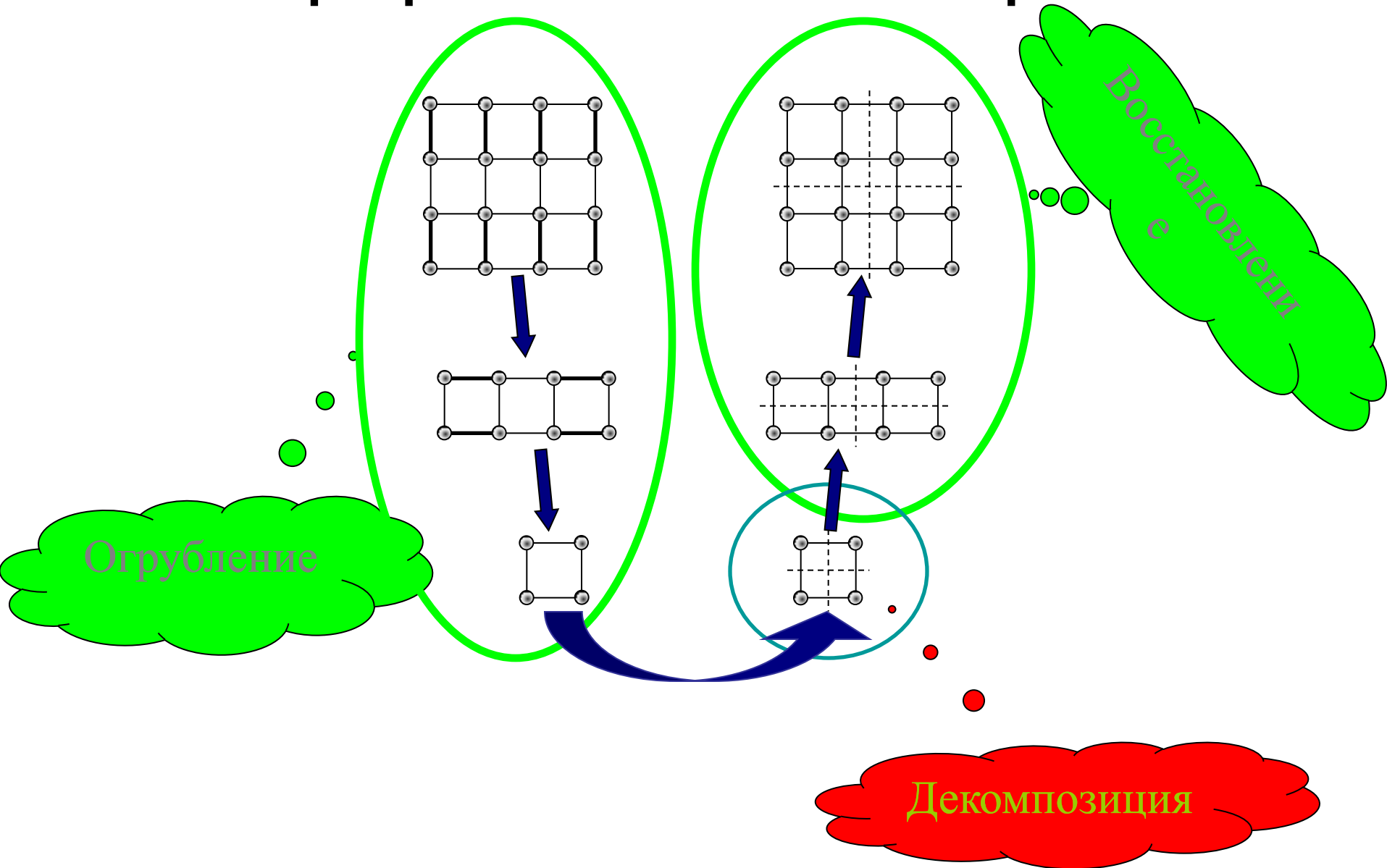
$$((y[i-1]==y[i]) \vee (x[i-1]<x[i]))$$

Рис. 76. Координатная бисекция

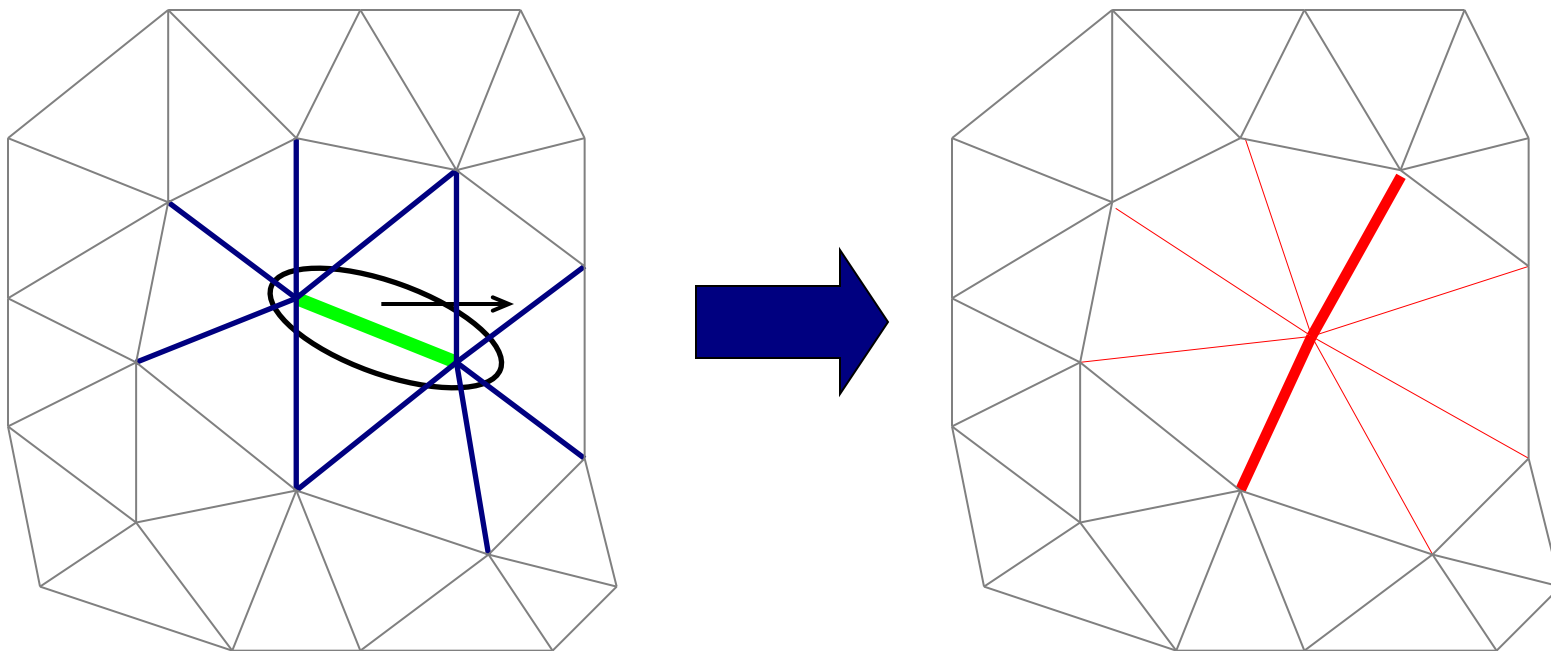
Пакеты декомпозиции графов

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

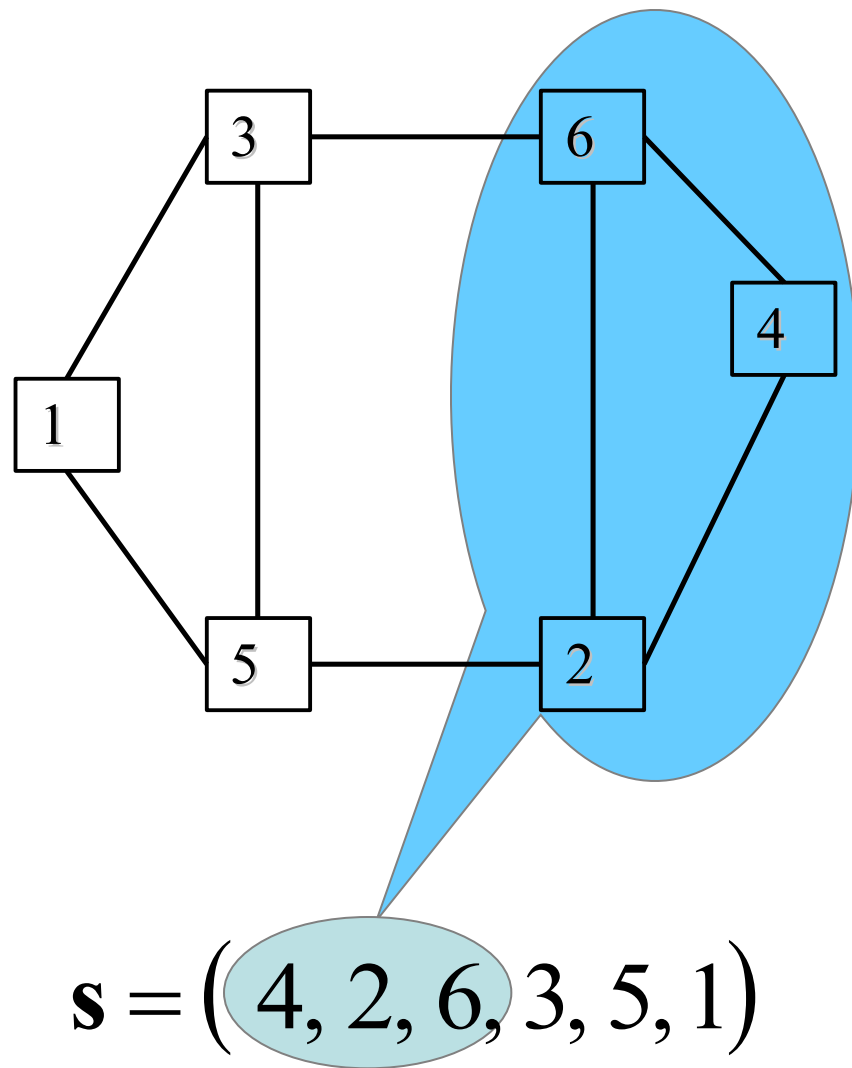


Спектральный метод

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

$$\lambda_2 = -1$$

$$x_2 = (2, -1, 1, -2, 1, -1)$$



Спектральная бисекция

Пусть

$$q[i] = \{-1, 1\}$$

потребуем, что бы $\sum_{i=1}^{|V|} q_i = 0$

Тогда число разрезанных ребер

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

Спектральный метод

Нормировка $\sum_{i=1}^{|V|} q_i^2 = q^T q = |V|$

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

$Lq = \lambda q$ 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

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

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

$\{-1\}$ и $\{1\}$ содержат одинаковое число вершин

Рекурсивная бисекция

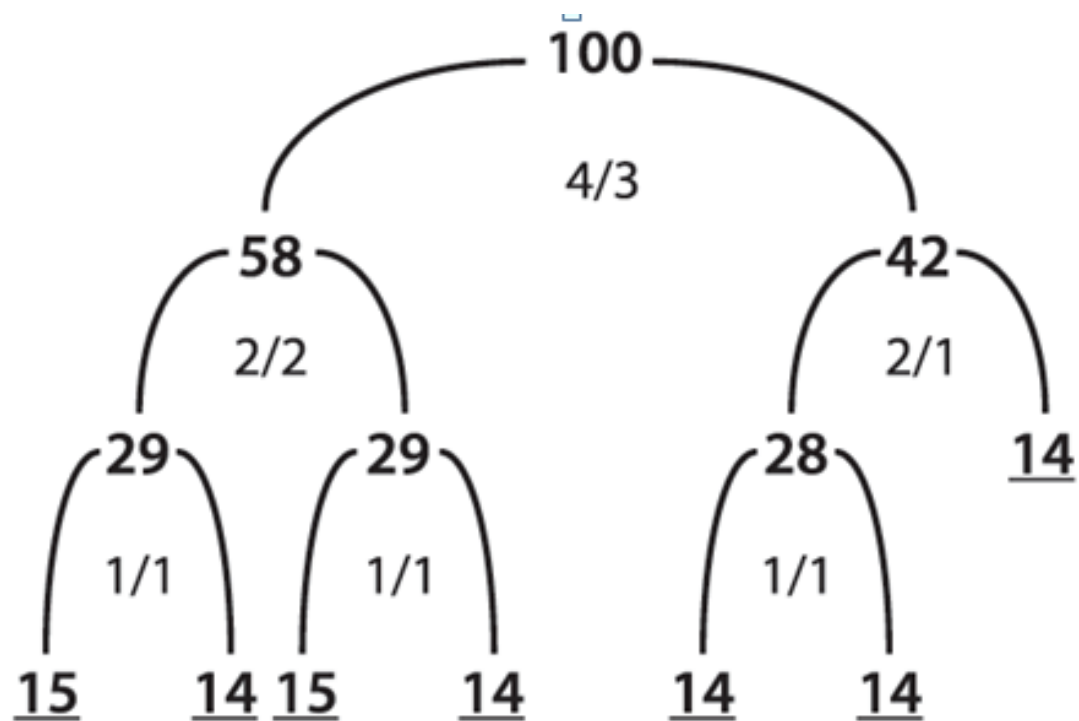
Рекурсивная бисекция

```
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)

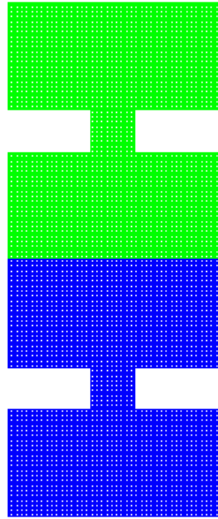
  if (k1 > 1) RecursiveBisect (subgraf1, n1, k1)
  if (k2 > 1) RecursiveBisect (subgraf2, n2, k2)
}
```

Бинарное дерево разрезов

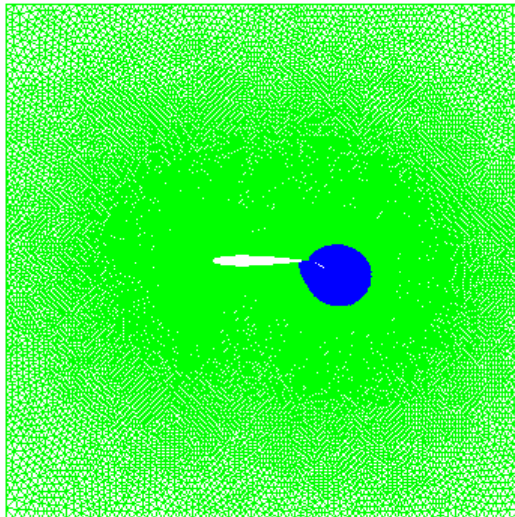


Метод спектральной бисекции

Spectral Partition

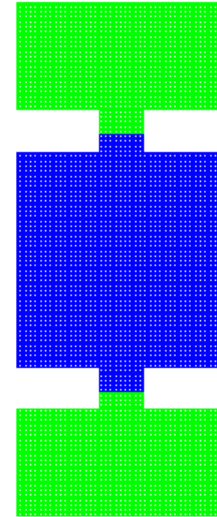


100 cut edges

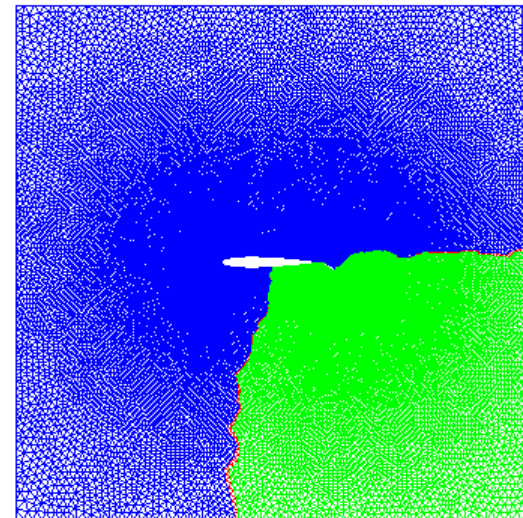


655 cut edges

Metis Partition

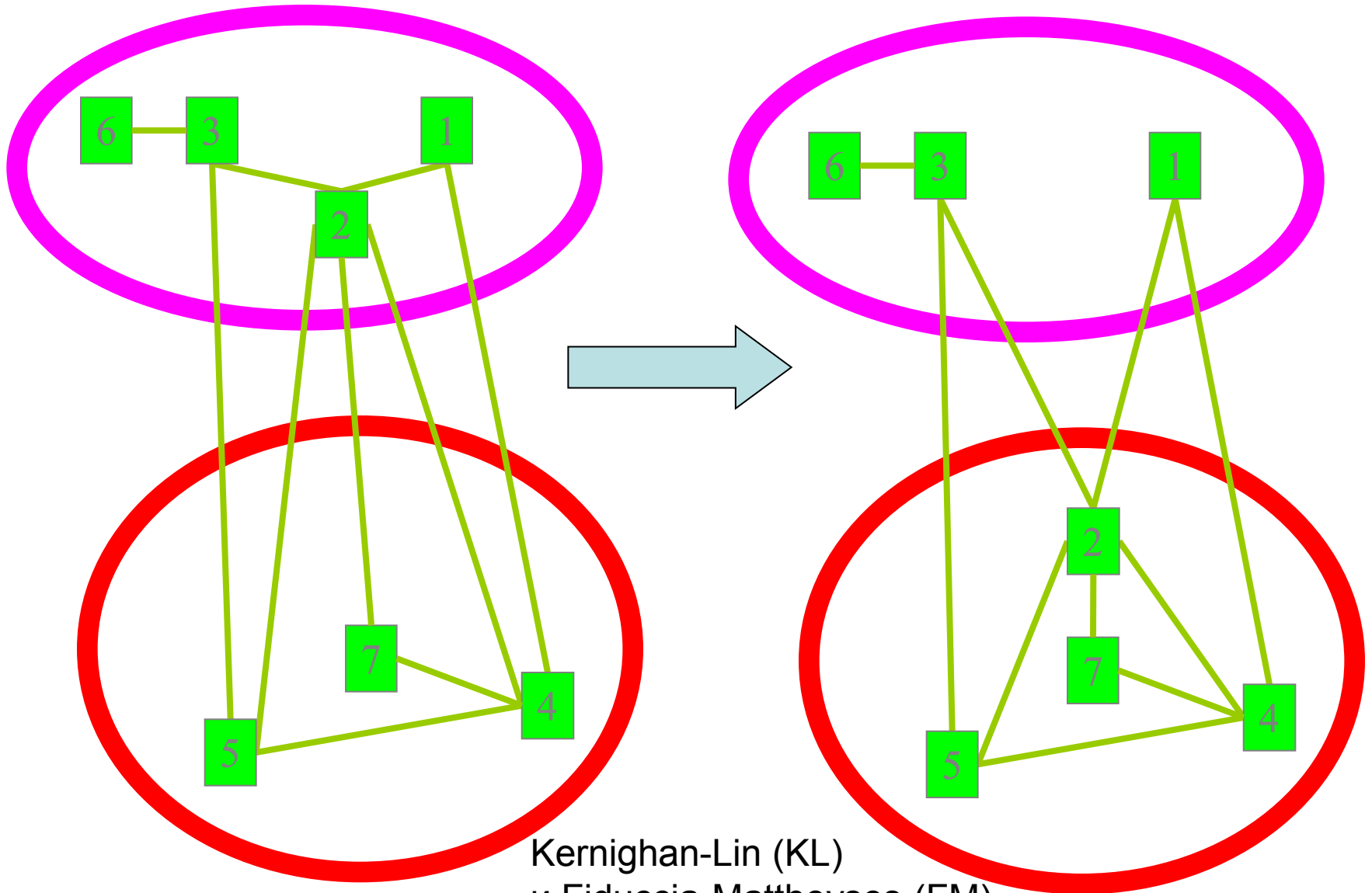


42 cut edges



524 cut edges

Локальное уточнение

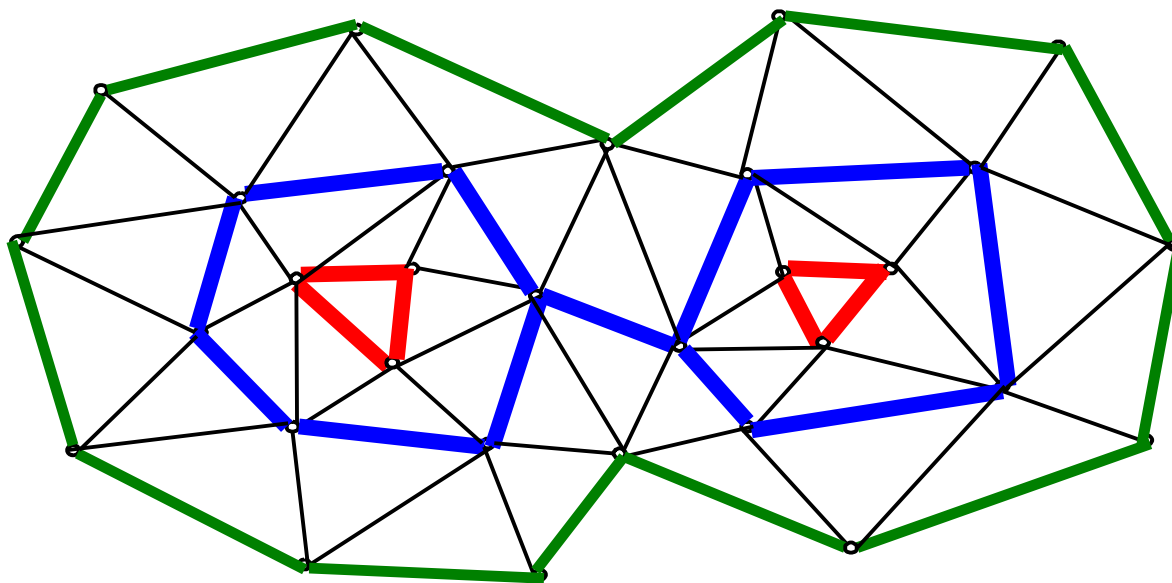


Kernighan-Lin (KL)
и Fiduccia-Mattheyses (FM)

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.

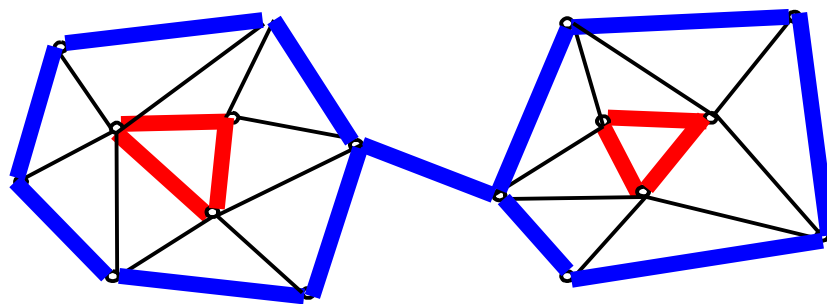
Связность ядер доменов



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

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

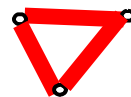
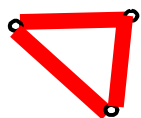
Связность ядер доменов



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

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

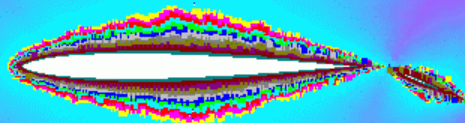
Связность ядер доменов



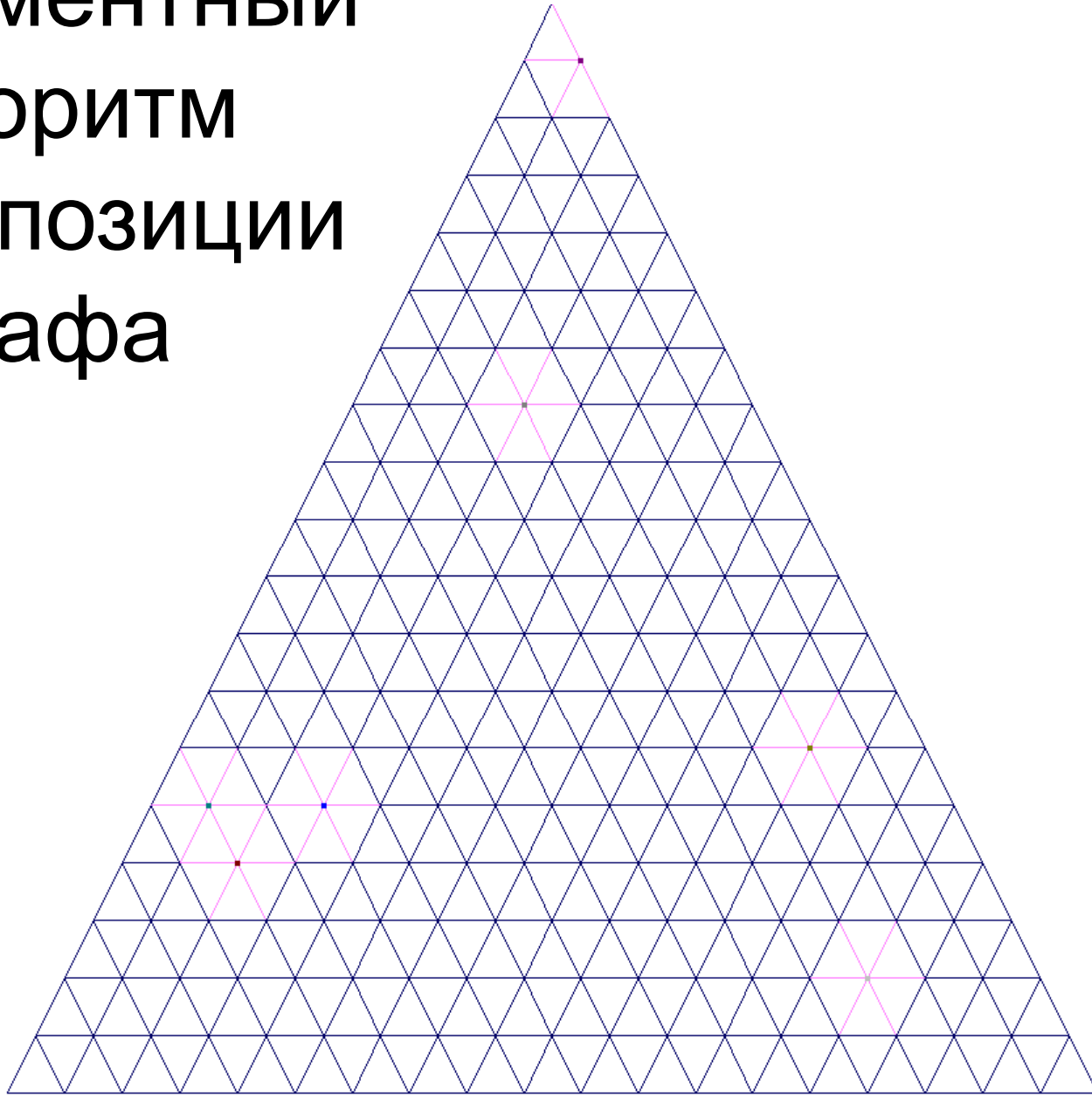
$$d(i) = \min_k : \mathbf{A}^k v_i \cap B \neq \emptyset$$

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

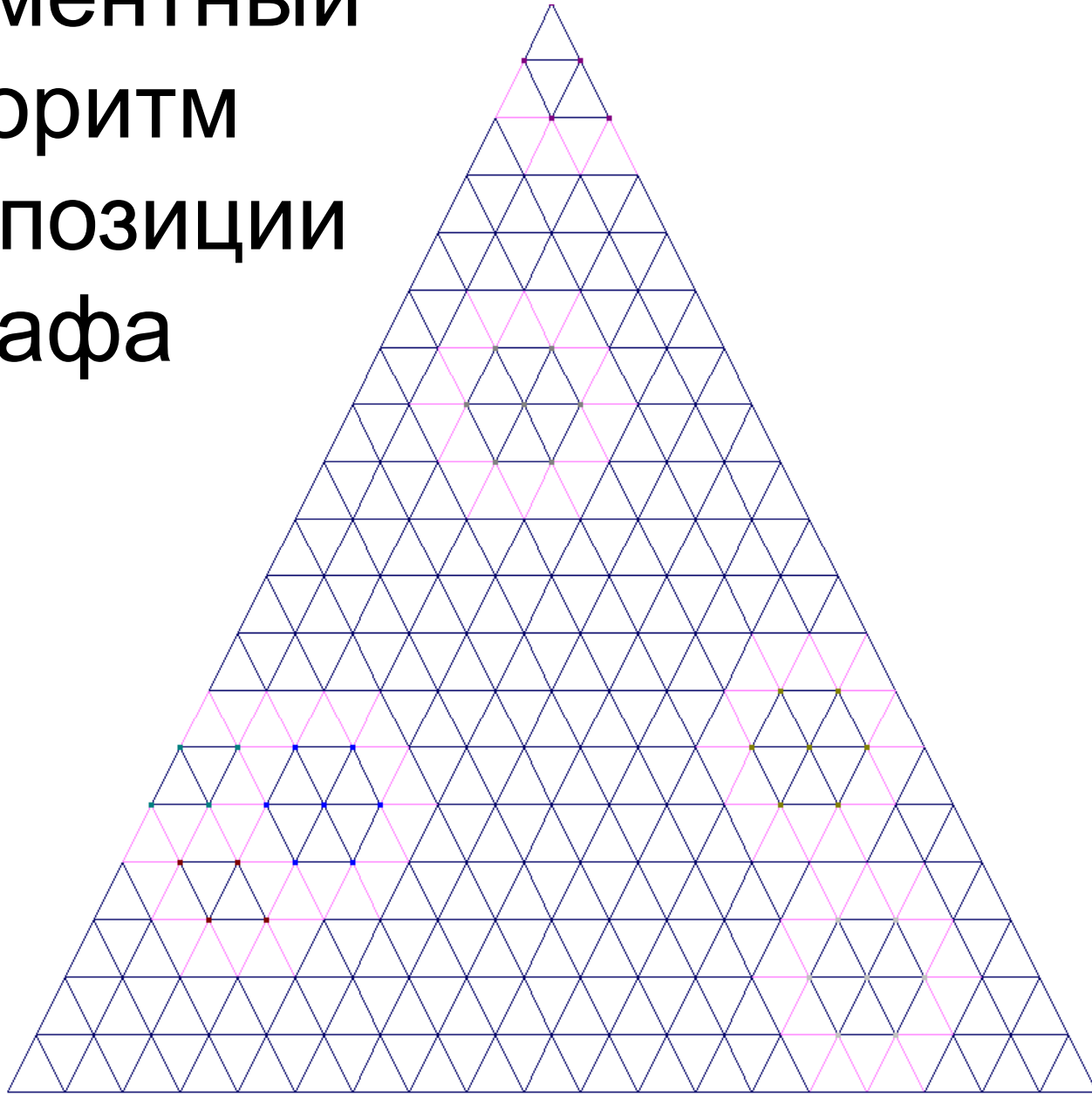
Оболочки



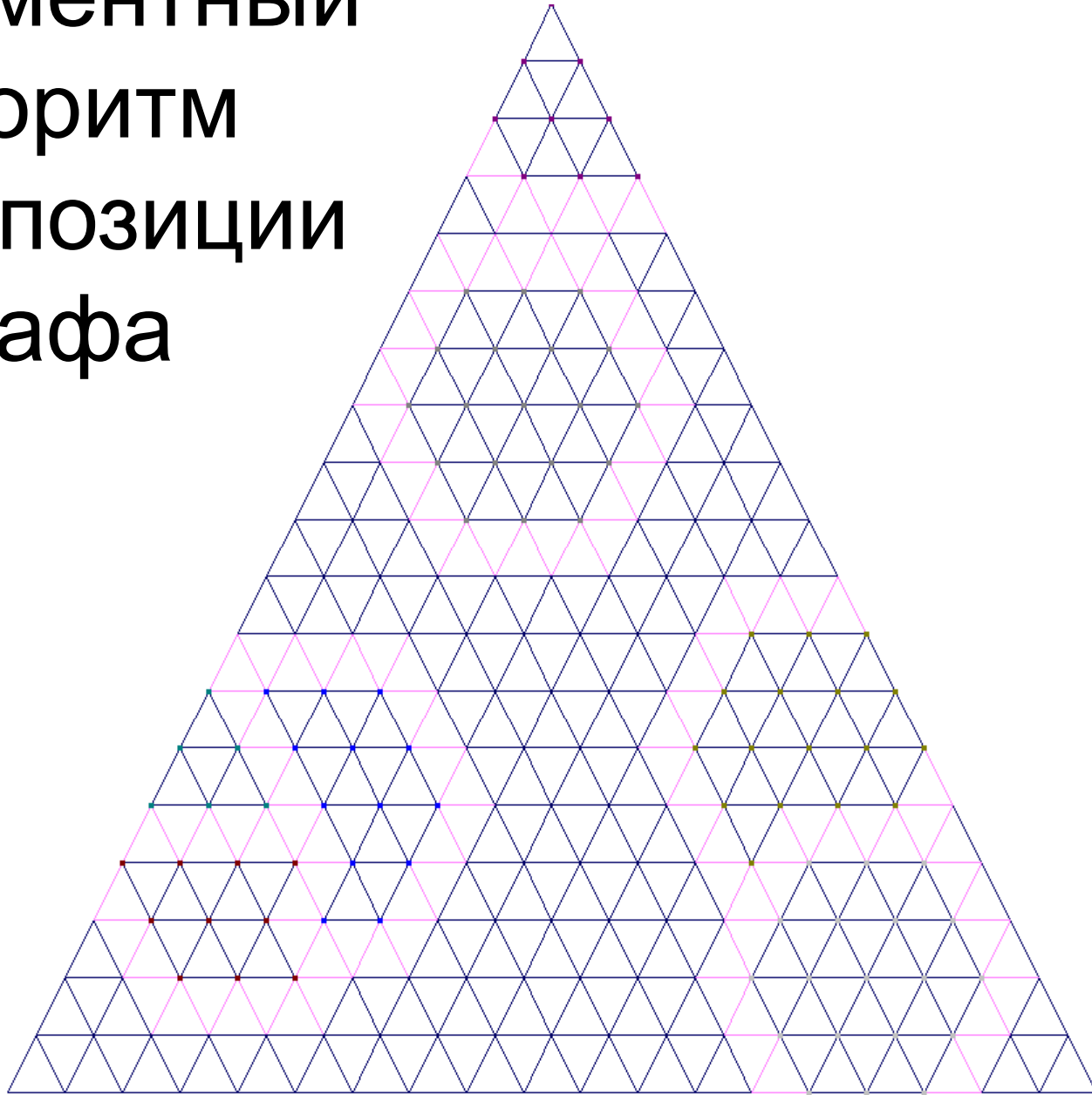
Инкрементный алгоритм декомпозиции графа



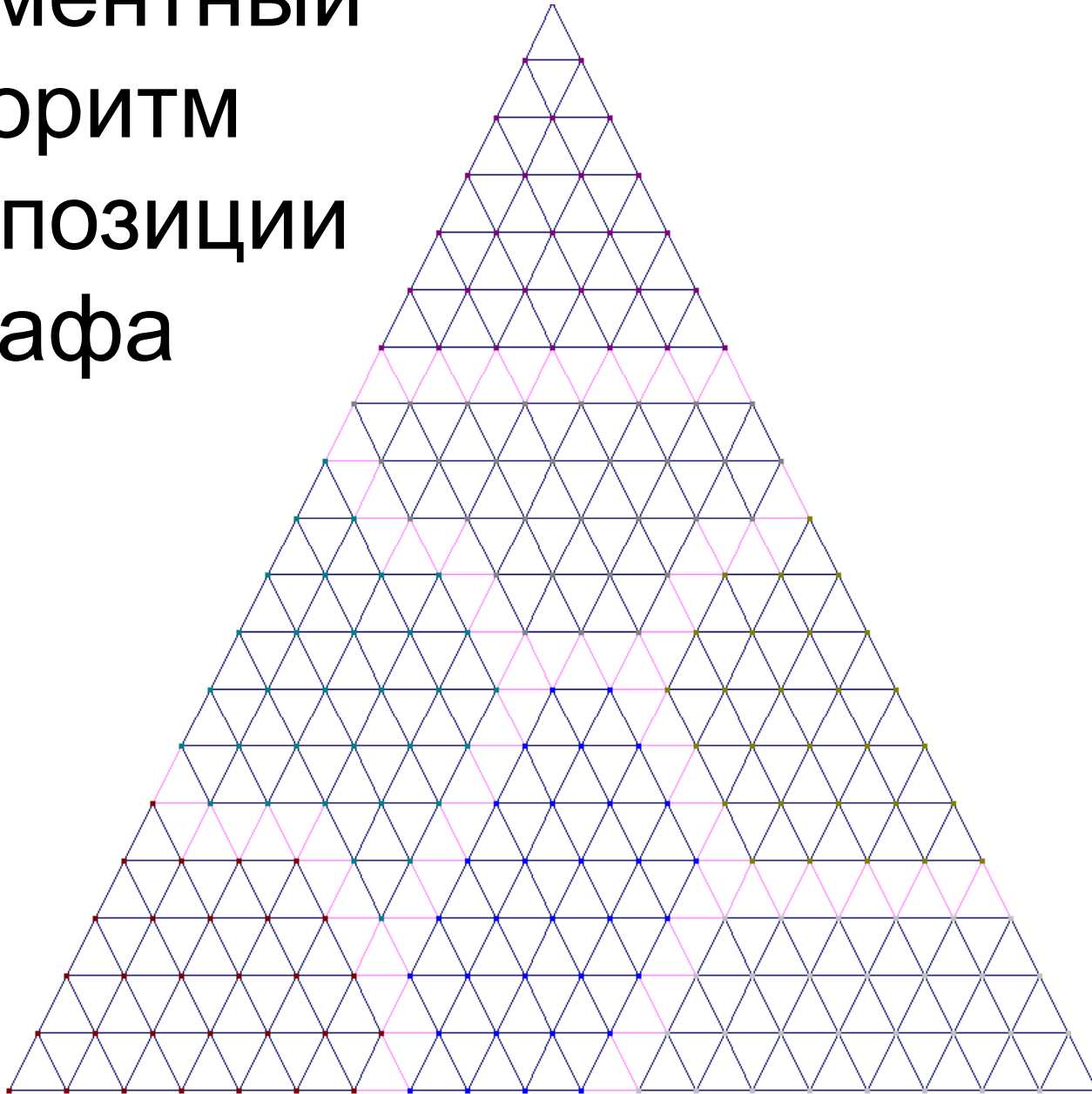
Инкрементный алгоритм декомпозиции графа



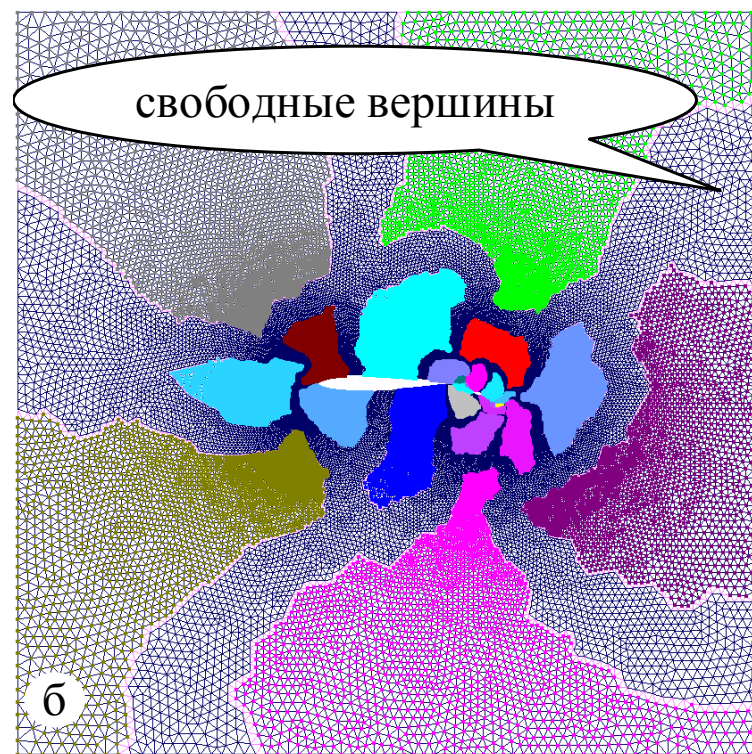
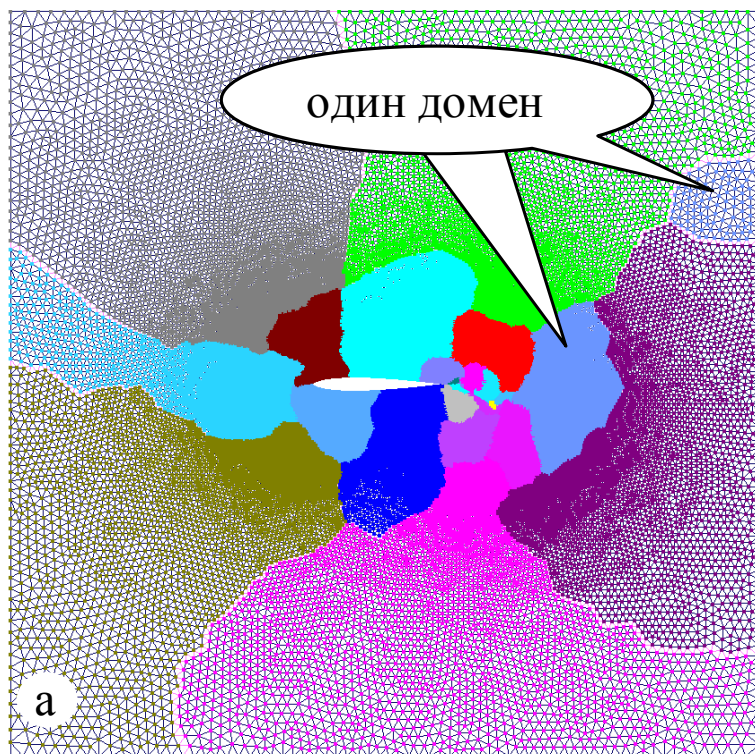
Инкрементный алгоритм декомпозиции графа



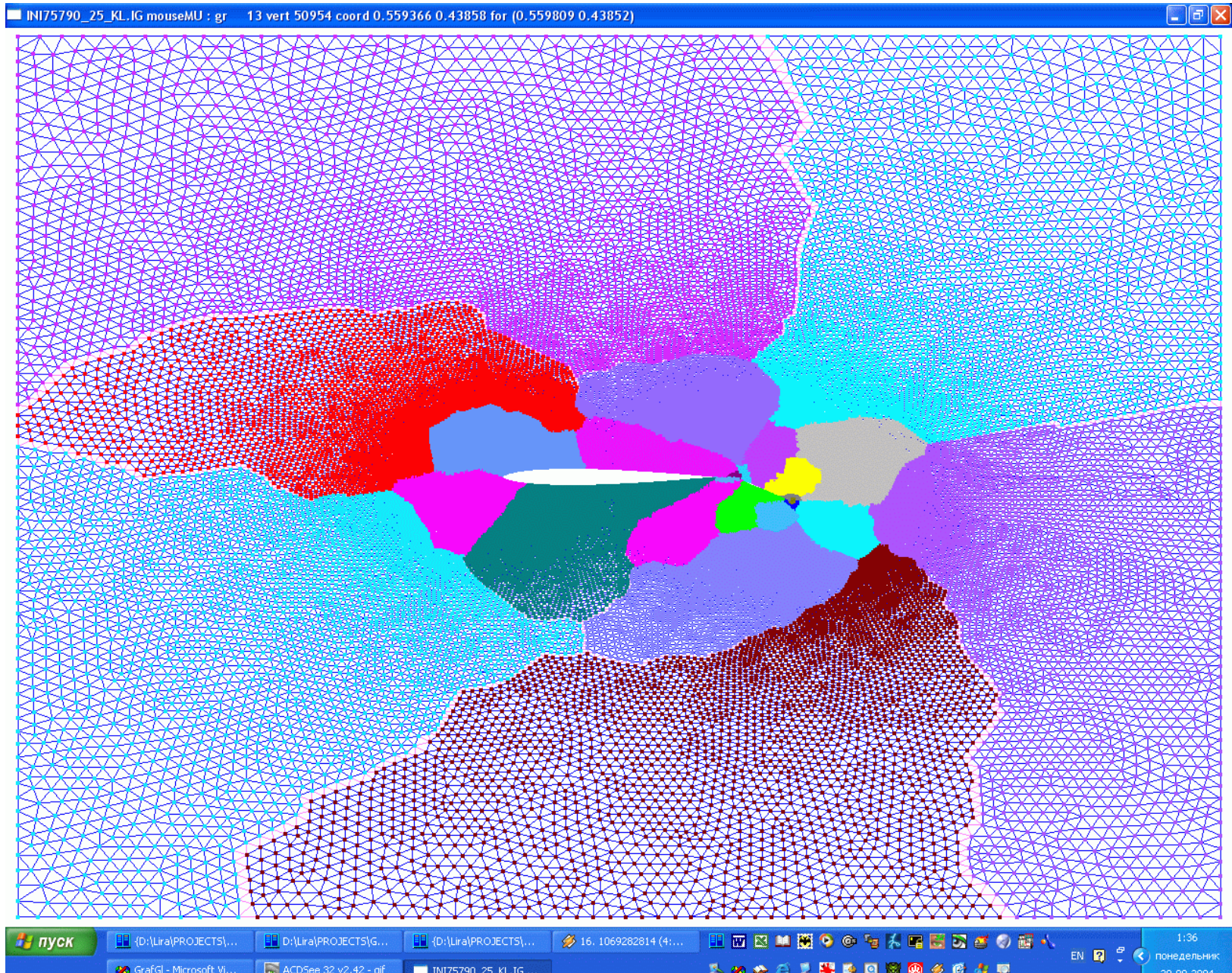
Инкрементный алгоритм декомпозиции графа



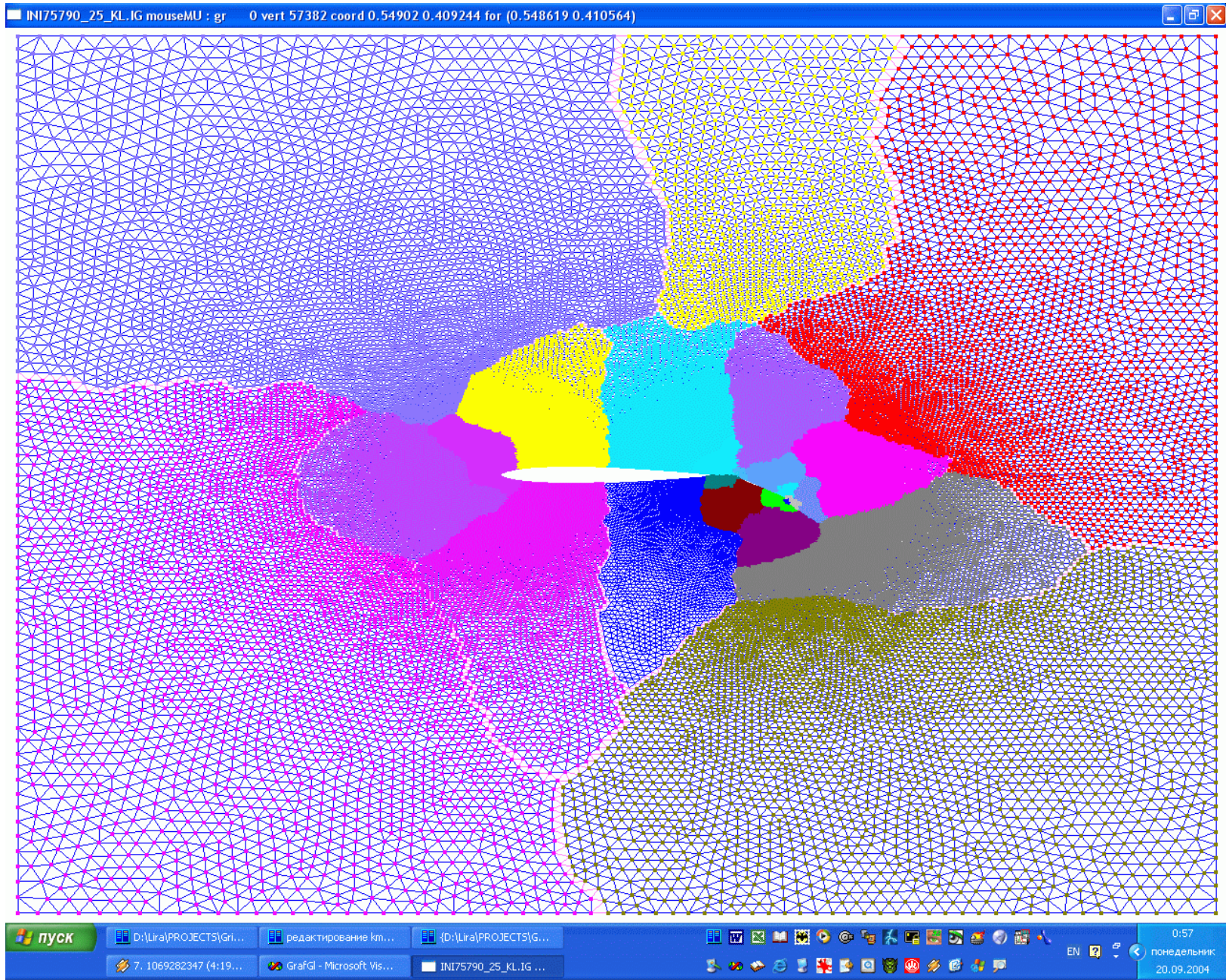
Редуцирование доменов



Инкрементный алгоритм, $Dm=25$



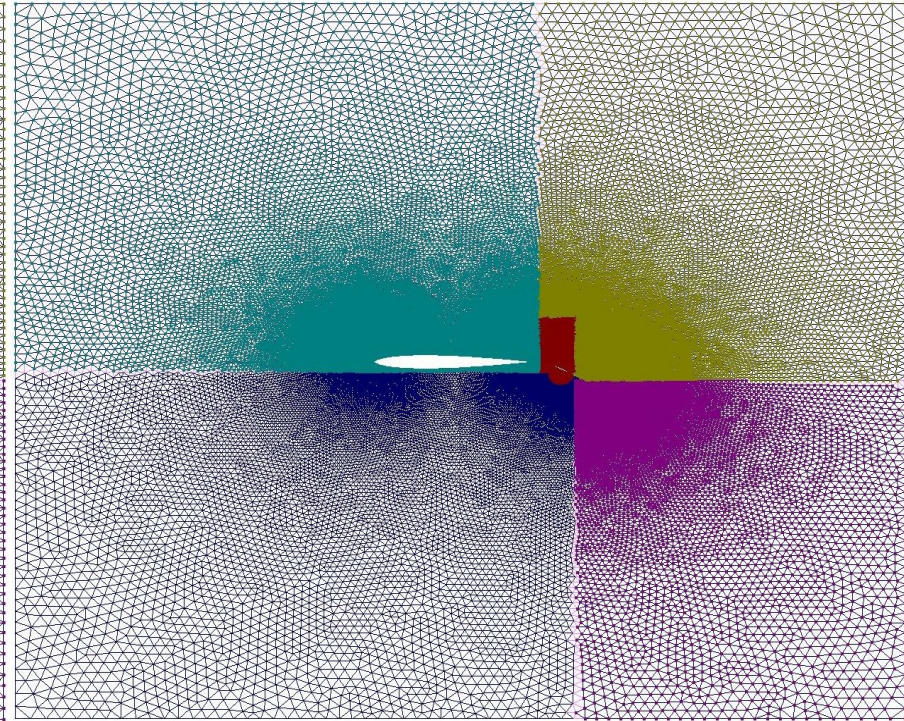
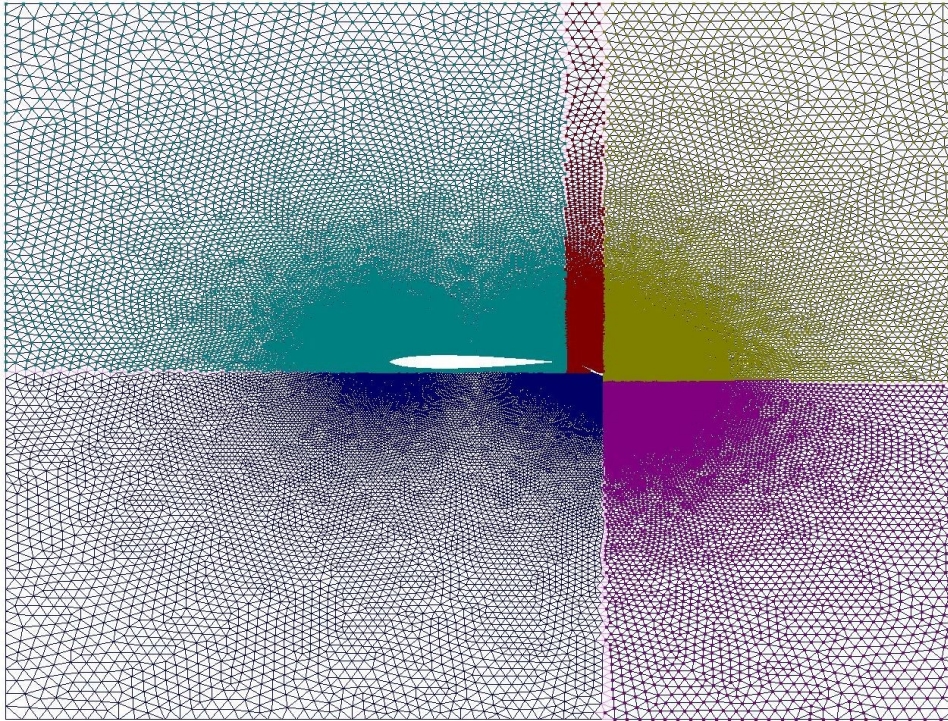
Kmetis, $Dm=25$



Треугольная сетка из 75790 вершин (пространство вокруг крыла)

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

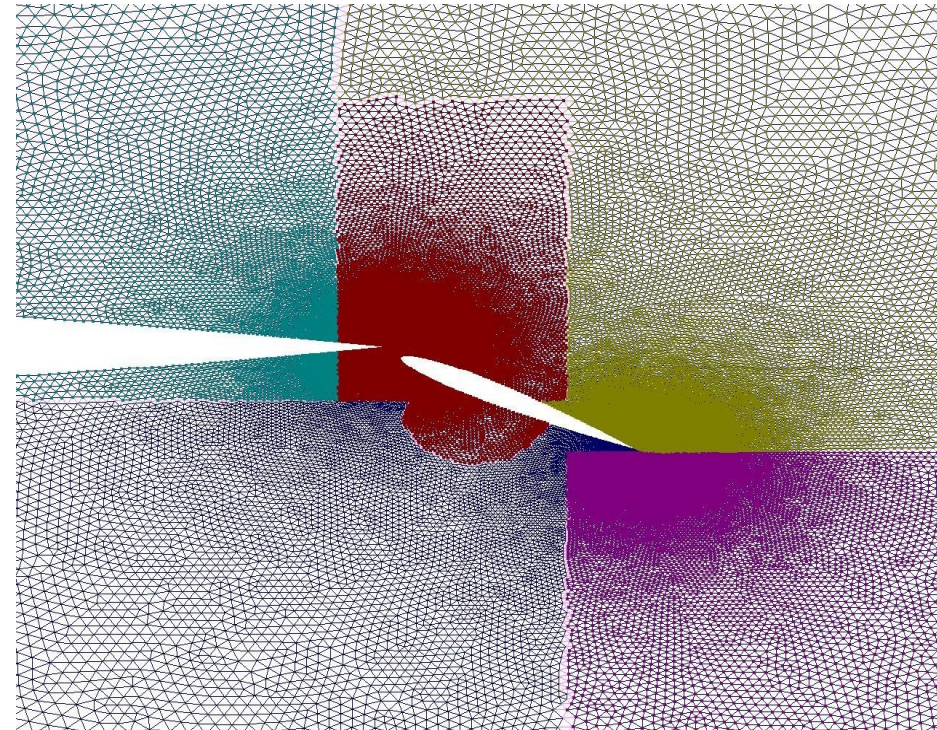
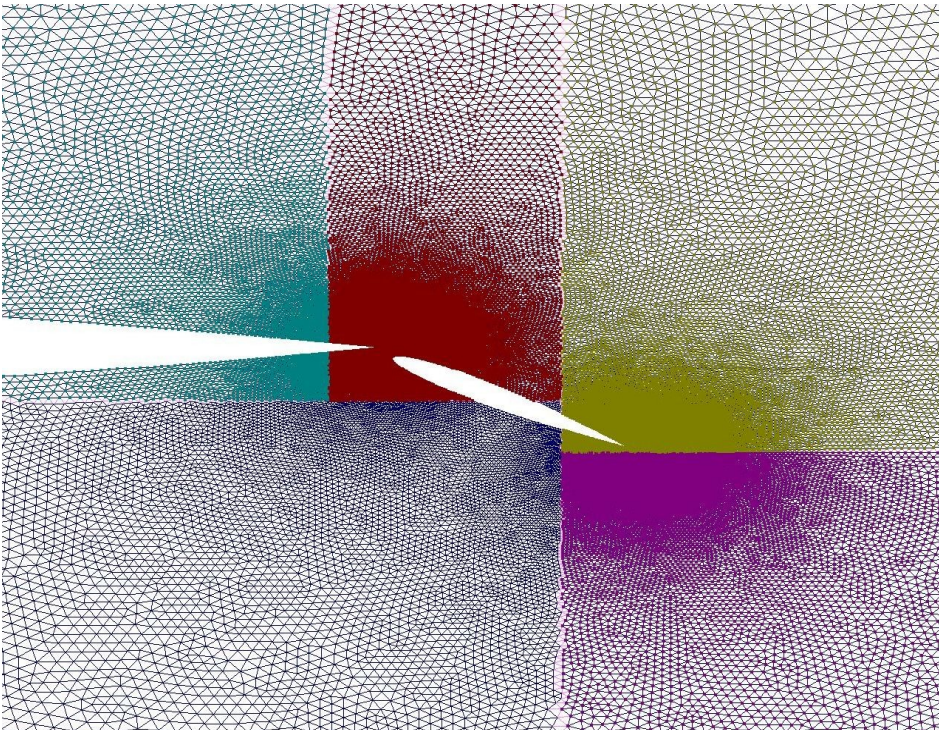
результат перераспределения
малых блоков вершин



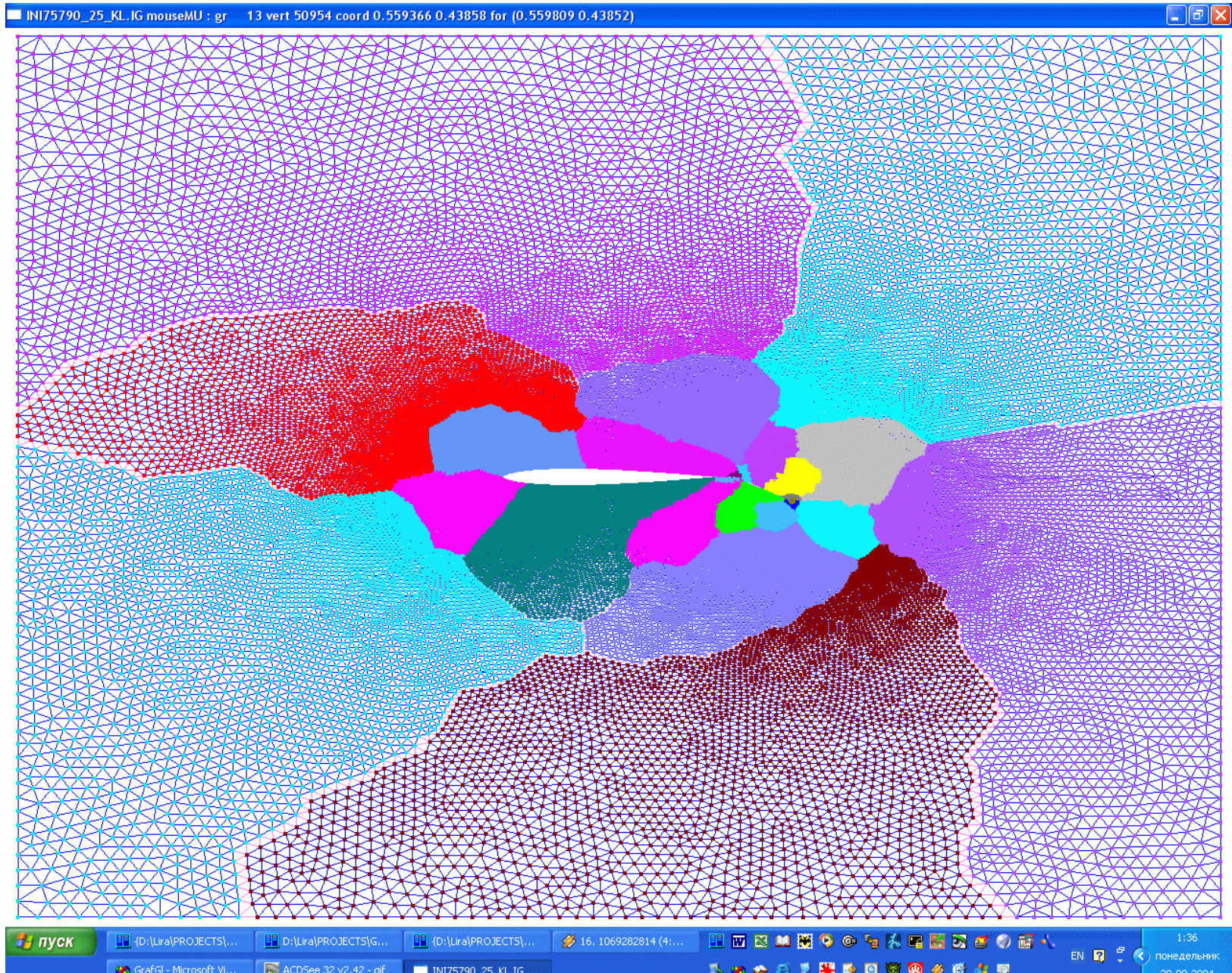
Фрагмент треугольной сетки из 75790 вершин

результат геометрической
декомпозиции

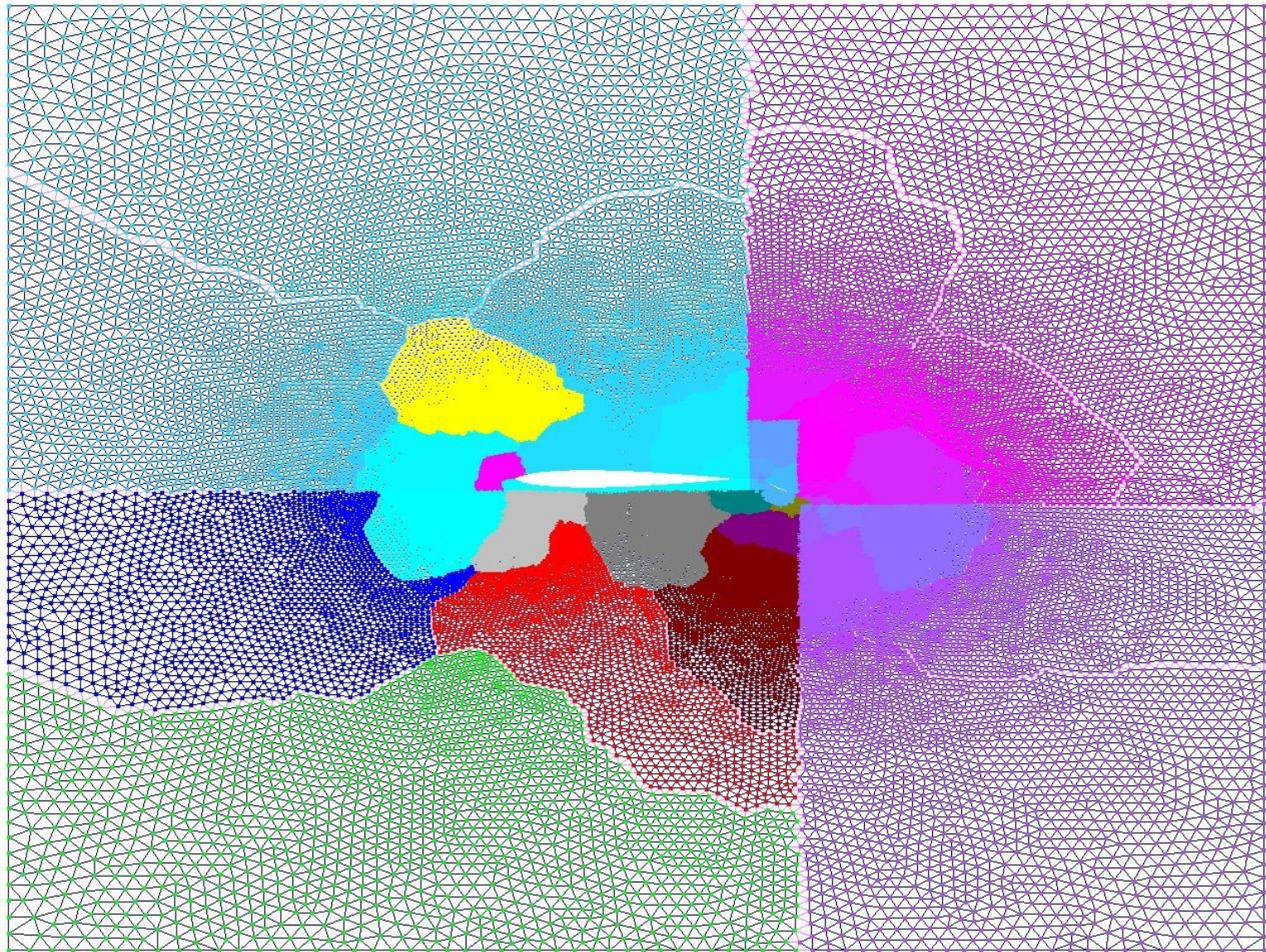
результат перераспределения
малых блоков вершин



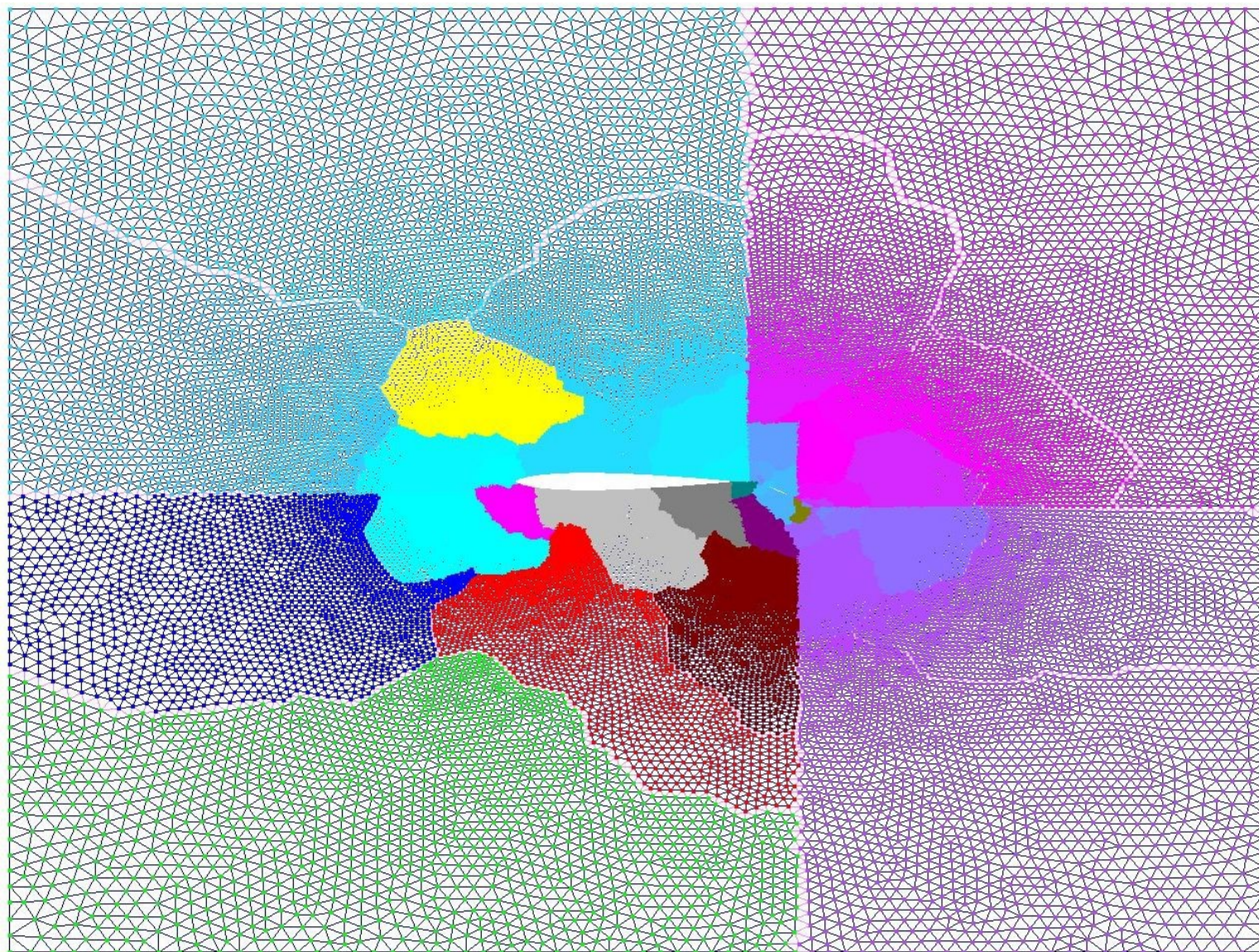
Инкрементный алгоритм, $Dm=25$



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



Результат сбора плохих групп доменов и их повторного разбиения

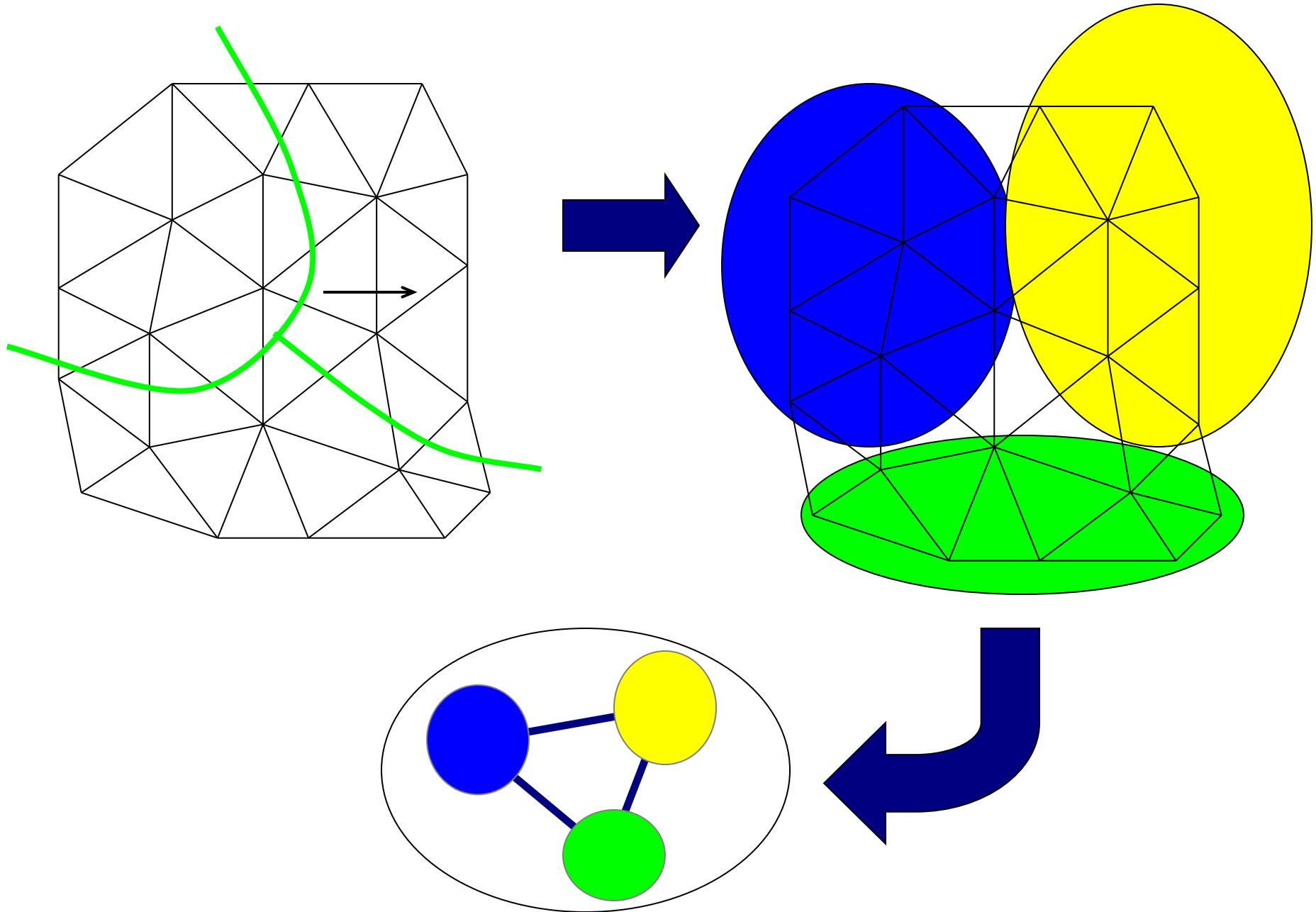


Разбиение тетраэдральной сетки, содержащей $2 \cdot 10^8$ узлов, на 125 процессорах

- вычисления производились на кластере СКИФ МГУ (1250 4-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	

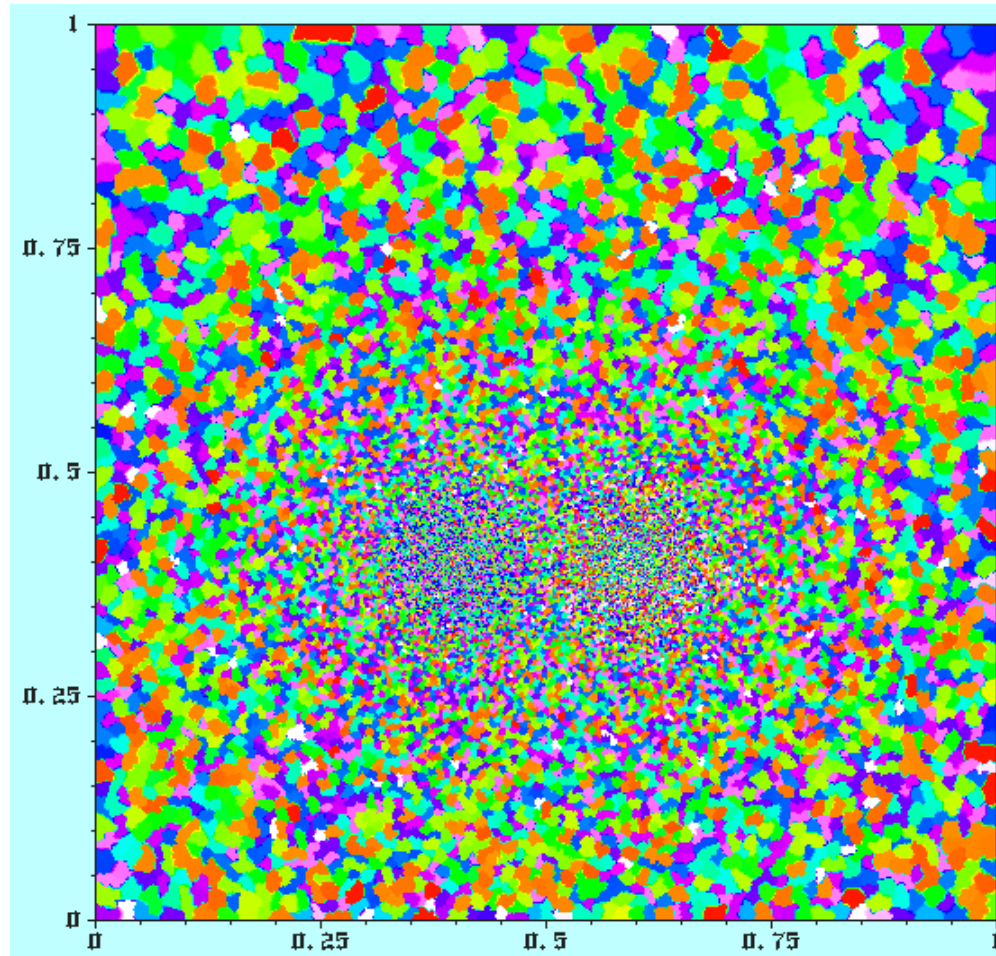
Формирование макрографа



Сетка микродоменов

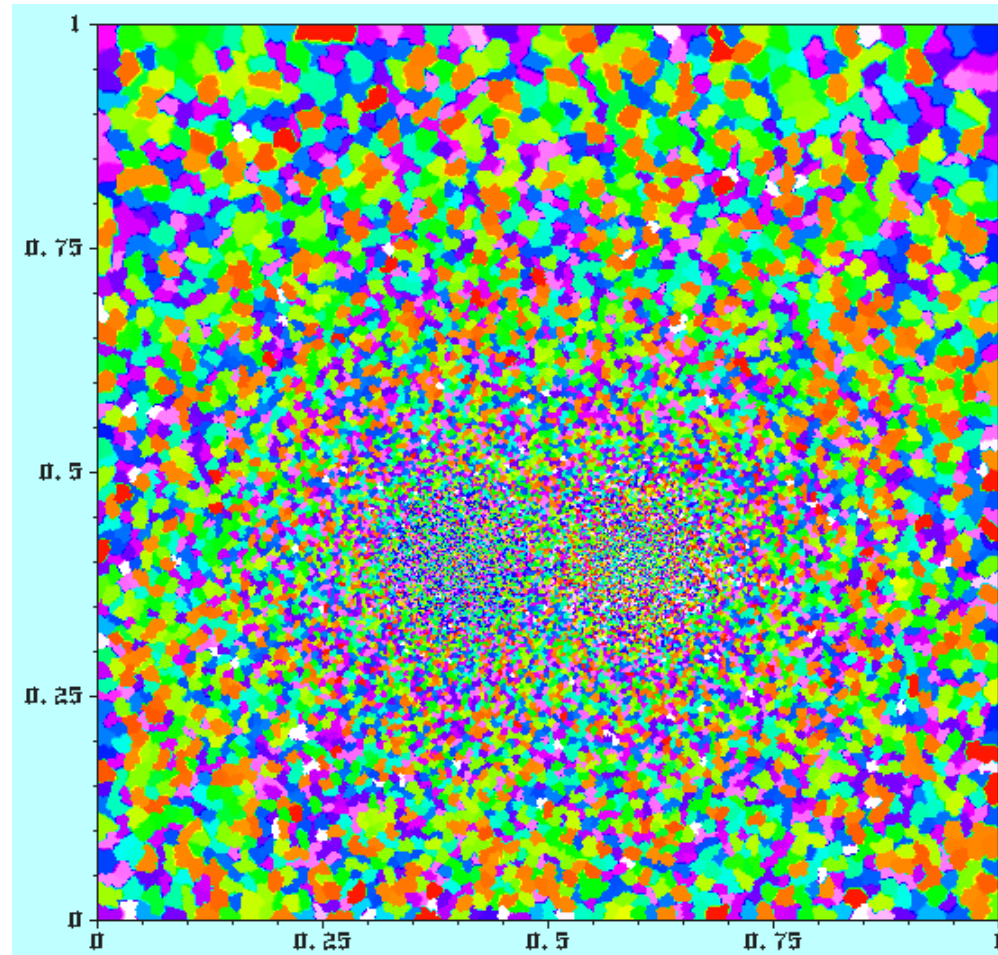
51 538 микродоменов

в каждом около
20 узлов



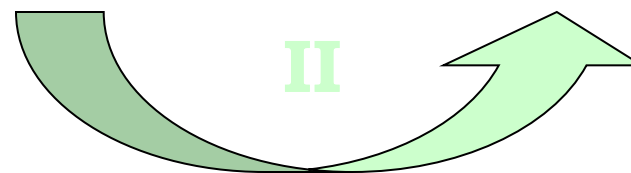
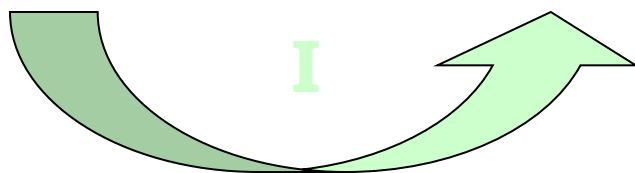
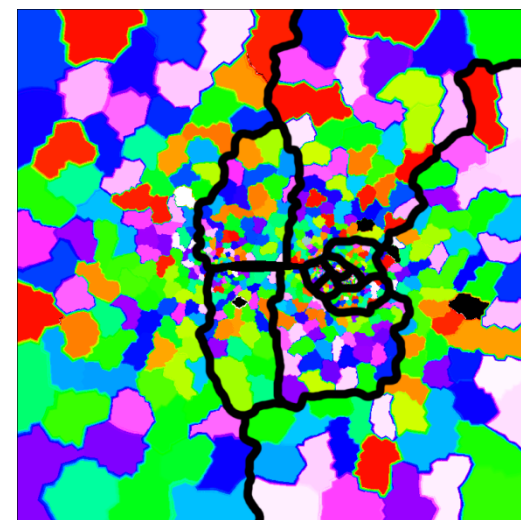
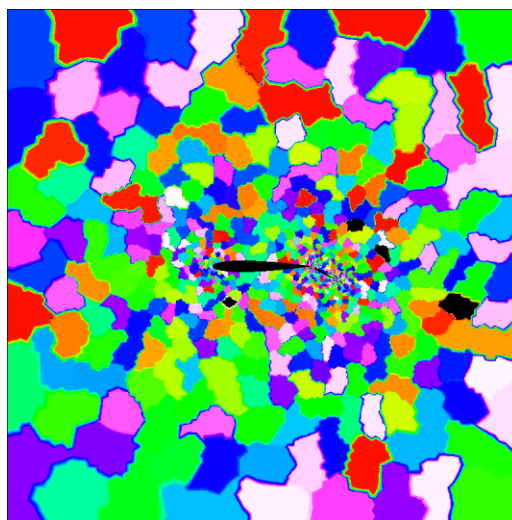
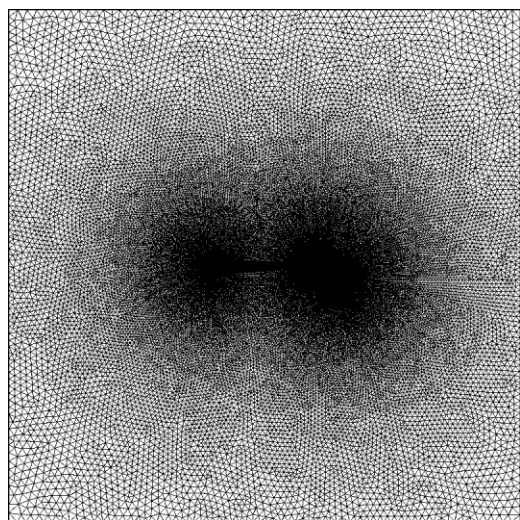
Сетка микродоменов

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 микродоменов

Двухуровневое разбиение



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

Вершины *макрографа*
распределяются по процессорам

Метод эффективен для сверхбольших сеток

Литература

1. Якобовский М.В. [Введение в параллельные методы решения задач](#): Учебное пособие / Предисл.: В. А. Садовничий. – М.: [Издательство Московского университета](#), 2013. – 328 с., илл. – (Серия «Суперкомпьютерное образование») ISBN 978-5-211-06382-2
2. *Fiedler M.* Eigenvectors of acyclic matrices // Czechoslovak Mathematical Journal, 25(100) – Praha, 1975. – Pp. 607–618.
3. *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.
4. *George Karypis* Family of Graph and Hypergraph Partitioning Software URL: <http://glaros.dtc.umn.edu/gkhome/views/metis/>
5. *Bruce Hendrickson, Rob Leland* Chaco: Software for Partitioning Graphs. URL: <http://www.cs.sandia.gov/~bahendr/chaco.html>
6. *Chris Walshaw* JOSTLE — graph partitioning software. URL: <http://staffweb.cms.gre.ac.uk/~c.walshaw/jostle/>
7. *François Pellegrini* Software package and libraries for sequential and parallel graph partitioning, static mapping, and sparse matrix block ordering, and sequential mesh and hypergraph partitioning. URL: <http://www.labri.u-bordeaux.fr/~pelegrin/scotch/>
8. *Головченко Е.Н., Якобовский М.В.* Библиотека параллельной декомпозиции больших сеток. 2013. URL: <http://lira.imamod.ru/FondProgramm/Decomposition/>

Литература

9. *Евстигнеев В.А.* Применение теории графов в программировании / Под ред. А.П. Ершова. – М.: Наука, Главная редакция физико-математической литературы, 1985. – 352 с.
10. *B. Hendrickson, R. Leland.* Multilevel Algorithm for Partitioning Graphs // Supercomputing '95 Proceedings. – San Diego, CA, 1995.
11. *Hendrickson B. and Leland R.* A Multi-Level Algorithm for Partitioning Graphs, Tech. Rep. SAND93-1301, Sandia National Laboratories, Albuquerque, October 1993.
12. *Hendrickson B. and Leland R.* An Improved Spectral Partitioning Algorithm for Mapping Parallel Computations. // SIAM J. Comput. Phys. –March 1995. –Vol. 16. – № 2. – P. 452–469.
13. *G.Karypis, V.Kumar.* Multilevel k-way partitioning scheme for irregular graphs, Journal of Parallel and Distributed Computing, 48 (1998), pp. 96-129.
14. *K. Devine, E. Boman, R. Heaphy, B. Hendrickson, and C. Vaughan* Zoltan Data Management Services for Parallel Dynamic Applications. Computing in Science and Engineering, Vol. 4, No. 2, March/April 2002, pp. 90-97.