

Джосан О.В., Попова Н.Н.

31.01.2010

oxana@cs.msu.su

popova@cs.msu.su

**Параллельный метод сжатия изображений для
визуализации данных на массивно-параллельных
вычислительных
системах**



Летняя школа для студентов, аспирантов и молодых ученых «Суперкомпьютерное моделирование и визуализация в научных исследованиях».

**Московский государственный университет имени М.В.Ломоносова
4 - 14 июля 2010 г.**

Московский государственный университет имени М.В.Ломоносова приглашает студентов, аспирантов и молодых ученых с 4 по 14 июля 2010 г. пройти обучение в летней школе «Суперкомпьютерное моделирование и визуализация в научных исследованиях». Школа проводится при поддержке Суперкомпьютерного консорциума университетов России на базе факультета Вычислительной математики и кибернетики совместно с Нижегородским государственным университетом имени Н.И.Лобачевского.

Тематики школы:

- Теория и технологии суперкомпьютерного моделирования в междисциплинарных научных исследованиях
- Технологии анализа и визуализации данных больших размерностей
- Управление междисциплинарными исследовательскими проектами, включающими суперкомпьютерное моделирование
- Методики обучения суперкомпьютерным технологиям для научных исследований

exascale.ru

Размер данных



Размер данных в задачах крупномасштабного моделирования

- При моделировании турбулентного горения на вычислительной системе Blue Gene/P для проекта FLASH с использованием 8000 четырехядерных вычислительных узлов генерируется 16Gb данных каждые 10-15 минут, общий объем данных эксперимента составляет 300Tb[1].
- Объем данных, получаемых в задаче климатического моделирования, составляет 345Tb[2].
- Объем данных, получаемых в результате проведения экспериментов на системе Blue Gene/P, установленной в МГУ, измеряется сотнями гигабайт[3]

[1] T. Peterka, R. B. Ross, H.-W. Shen, K.-L. Ma, W. Kendall, and H. Yu, "Parallel Visualization on Leadership Computing Resources," Preprint ANL/MCS-P1656-0709, July 2009.

[2] T. Peterka, R. Ross, H. Yu, K.-L. Ma, W. Kendall, and J. Huang, "Assessing and Improving Large Scale Parallel Volume Rendering on the IBM Blue Gene/P," Preprint ANL/MCS-P1554-1008, October 2008

[3] Джосан О.В., Попова Н.Н., Шумкин Г.Н. Методы визуальной поддержки для задач молекулярного моделирования на суперкомпьютере Blue Gene /P // тезисы конференции «Научный сервис в сети Интернет», Россия, Новороссийск, 2009, сс. 426-428, М.: Изд-во МГУ, 2009.

Сжатие данных



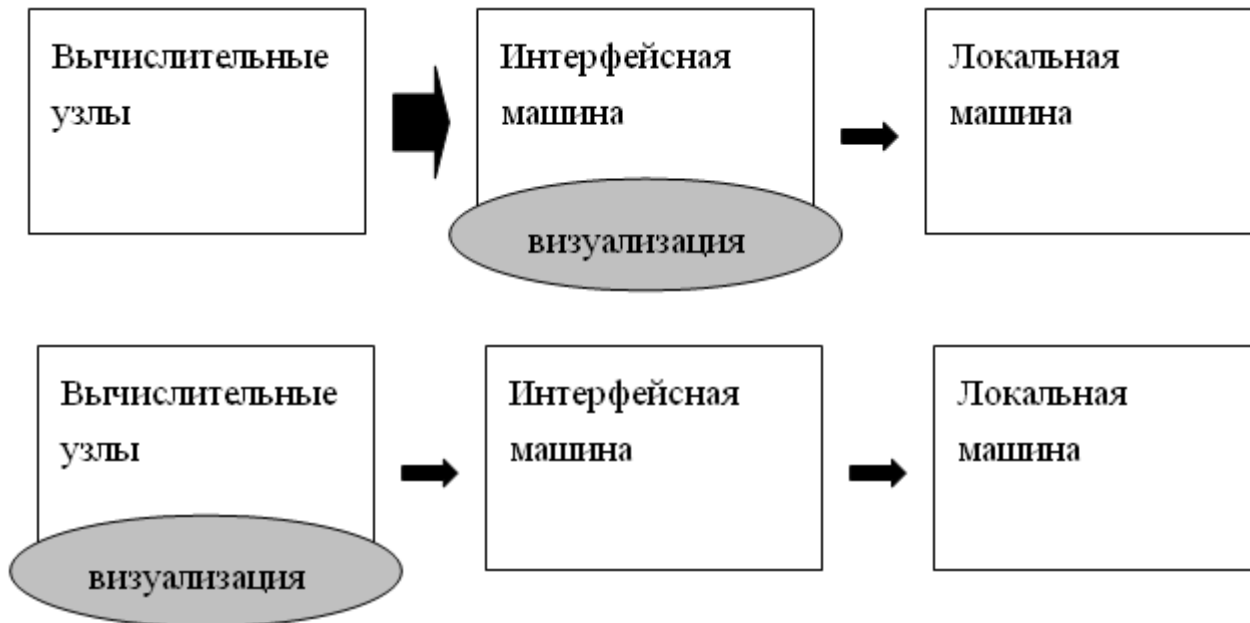
Алгоритмы сжатия при обработке данных на суперкомпьютере

- Сжатие данных – уменьшение размера передаваемых данных
- В работе рассмотрен метод параллельного сжатия изображений
- **Области применения** в суперкомпьютерных вычислениях:
 - 1. Визуализация данных при суперкомпьютерном моделировании
 - 2. Обработка изображений больших размеров на суперкомпьютерах
 - 3. ...

Требования к алгоритму сжатия



Требования, накладываемые на алгоритм сжатия при данных постановках задач

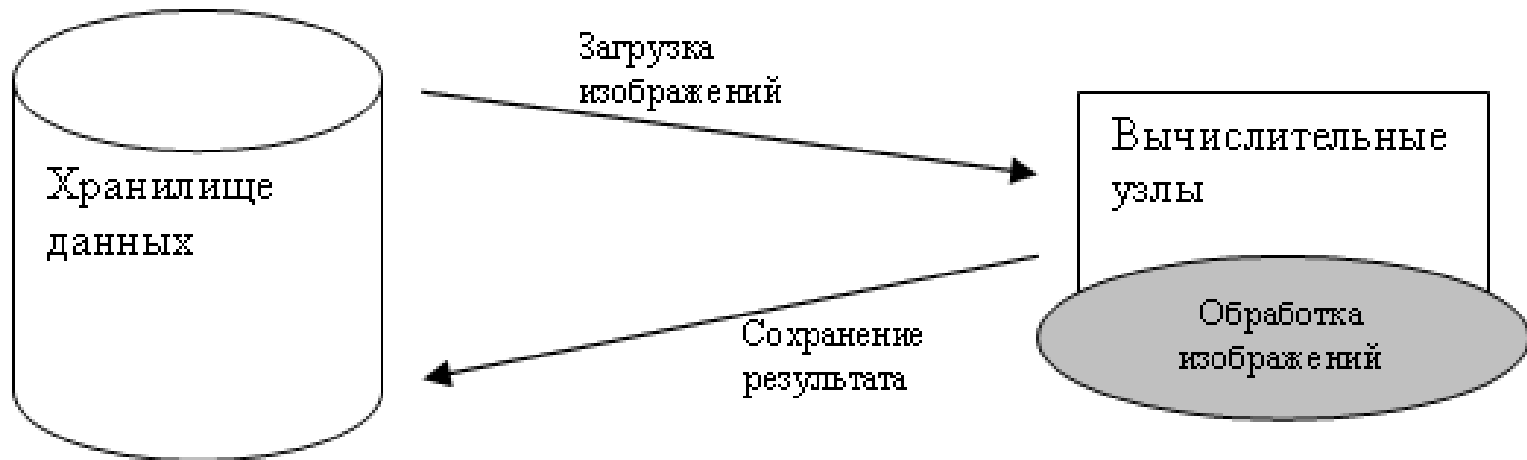


- 1. Малая вычислительная сложность алгоритма
- 2. Эффективная работа на небольшом количестве ядер
- 3. Блочный метод кодирования – без пересылки данных

Требования к алгоритму сжатия



Требования, накладываемые на алгоритм сжатия при данных постановках задач



- 1. Быстрый метод декодирования → использование префиксного кодирования
- 2. Использование локальных областей для кодирования

Требования к алгоритму сжатия



Требования, накладываемые на алгоритм сжатия при данных постановках задач

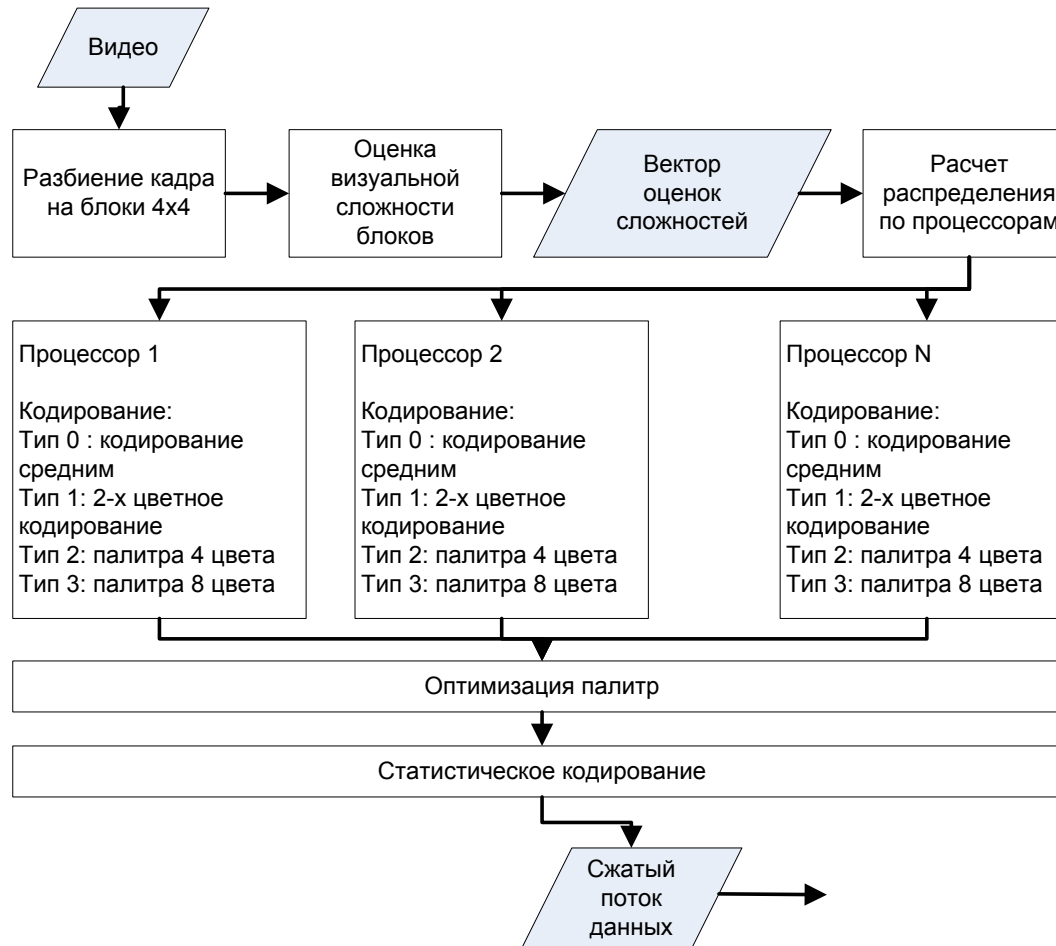
Итого:

- **Малая вычислительная сложность алгоритма**
- **Быстрый метод декодирования → использование префиксного кодирования**
- **Использование локальных областей для кодирования**
- **Блочный метод кодирования – без пересылки данных**
- **Эффективная работа на небольшом количестве ядер**
- **Критерий степени сжатия – качество без визуальных потерь или фиксированная степень сжатия**
- **Выбор метода кодирования по априорным данным**

Схема алгоритма кодирования



Общая схема предлагаемого алгоритма сжатия изображений



Априорная оценка



Оценка визуальной сложности блока

Предложена следующая формула для оценки визуальной сложности блока:

$$F = \frac{\sigma}{12} \left(\frac{D_x}{4} + \frac{D_Y}{4} + \frac{D_{d1}}{3} + \frac{D_{d2}}{3} \right) + \frac{\sigma}{2} \sum_{T \in \{Y, Cb, Cr\}} A_T + B_T$$

Предложен следующий метод расчета коэффициентов:

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i, \quad \sigma = \frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})^2$$



$$B_T = \frac{1}{4} (|I_T(i, j-) - I_T(i- j)| + |I_T(i- j) - I_T(i, j+)| + |I_T(i, j+) - I_T(i+ j)| + |I_T(i+ j) - I_T(i, j-)|)$$

Априорная оценка



Оценка визуальной сложности блока

Предложен следующий метод расчета коэффициентов:

$$D_X = \sum_{\substack{k=-1..1 \\ l=-1..2}} |I_Y(i+k+1, j+l) - I_Y(i+k, j+l)|$$

$$D_Y = \sum_{\substack{k=-1..2 \\ l=-1..1}} |I_Y(i+k, j+l+1) - I_Y(i+k, j+l)|$$

$$D_{d1} = \sum_{\substack{k=-1..1 \\ l=-1..1}} |I_Y(i+k+1, j+l+1) - I_Y(i+k, j+l)|$$

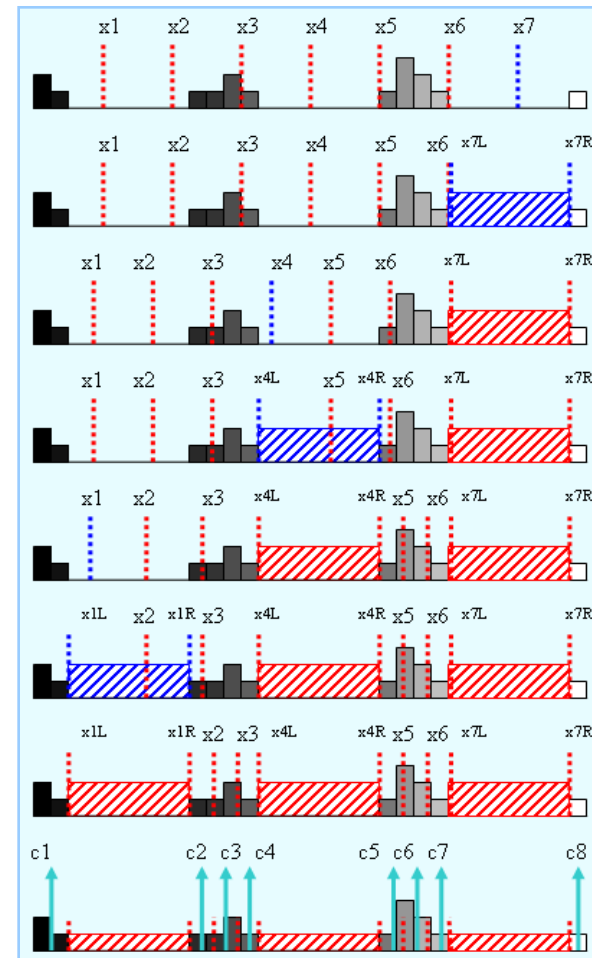
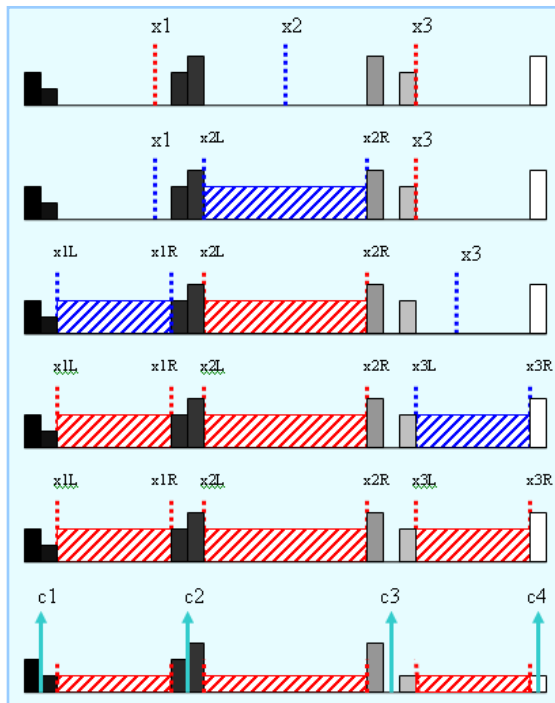
$$D_{d2} = \sum_{\substack{k=-1..1 \\ l=-1..1}} |I_Y(i+k+1, j+l) - I_Y(i+k, j+l+1)|$$

Кодирование



Расчет оптимальной палитры

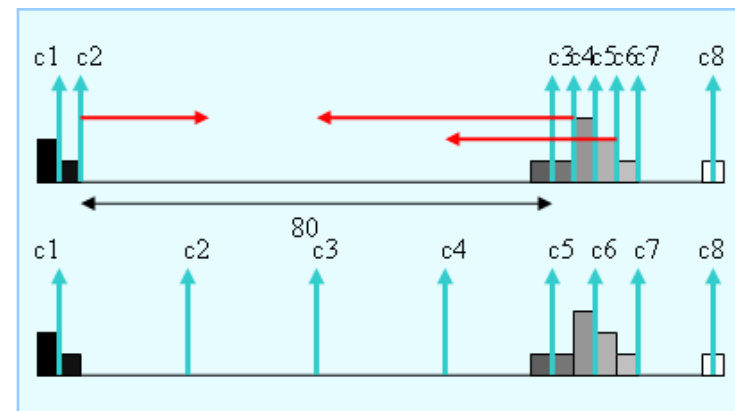
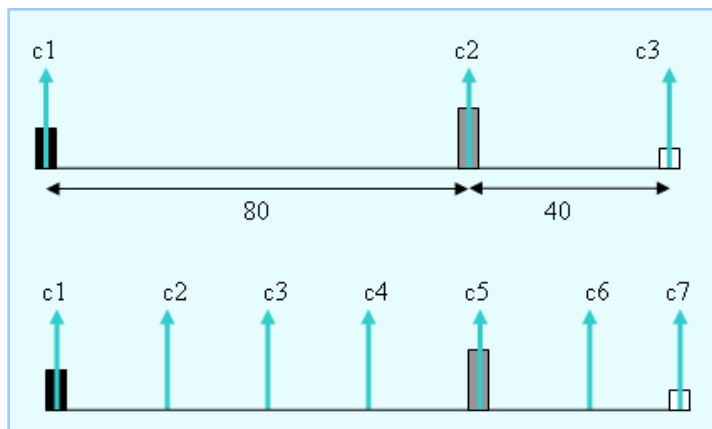
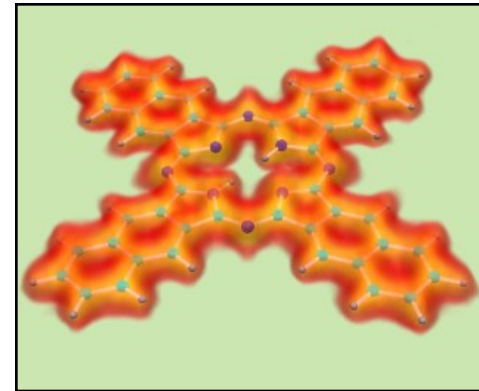
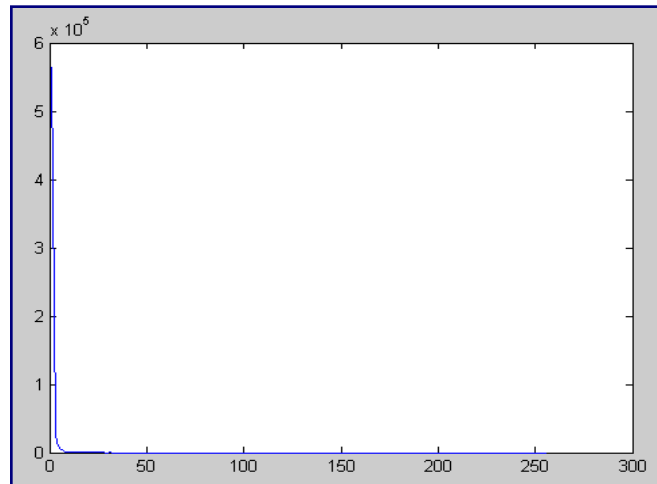
На слайде проиллюстрирован предложенный метод построения палитры для кодируемого блока:



Метод сжатия видео



Расчет оптимальной палитры

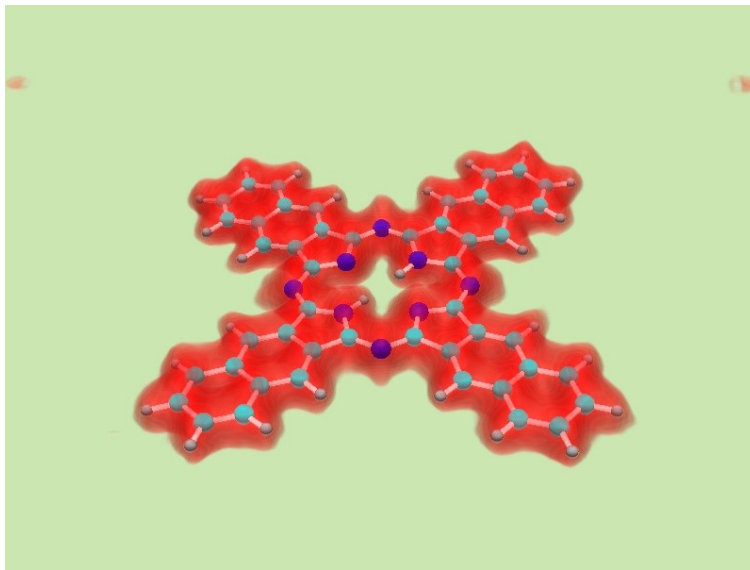


Исследование эффективности



Исследование эффективности на примере визуализации

Моделирование молекулярных переключателей с помощью кода CPMD
(совместная работа кафедры АНИ ВМК МГУ и IBM Zurich Research Lab)



Пример модели: 82 атома, 129 орбиталей,
временной шаг = 0.1 fs, решетка
320x320x192

Размер данных одного шага — 30Mb
Общий размер данных — 200Gb

- Для эксперимента использовались первые 1024 файла исходных данных, представленных в наборе файлов формата .cube, каждый из которых описывает положение молекулы в момент времени.
- Размер получаемого при визуализации изображения по одному файлу до применения алгоритма сжатия составлял от 120К для изображения 200x200 до порядка 2М для изображения 800x800.

Исследование эффективности



Исследование эффективности на примере визуализации



График показывает степень сжатия для изображения размером 200x200 для качества без визуальных потерь с использованием предложенного многопоточного метода сжатия.

В среднем степень сжатия равна 10. Таким образом, исходный размер данных вычислительного эксперимента, используемых при данном модельном эксперименте, составлял порядка 31 гигабайта. Размер полученной в этом эксперименте видеопоследовательности до применения сжатия составил 123М, размер полученной видео последовательности после применения алгоритма сжатия составил около 12М.

Исследование эффективности



Исследование эффективности на примере обработки изображений

**Обработка спутниковых изображений на суперкомпьютере Blue Gene /P
(Использование метода сжатия в рамках библиотеки ParImPr, разрабатываемой на кафедре АСВК МГУ. ParImPr – библиотека обработки изображений на суперкомпьютерах с большим количеством вычислительных узлов)**



Пример модели:

Размер одного изображения ~8000x8000
пикселей

Формат представления – RGB

Размер данных одного шага — 30Mb

Общий размер данных — 30Gb

Исследование эффективности



Исследование эффективности на примере обработки изображений



График показывает степень сжатия для изображения размером 8000x8000 для качества без визуальных потерь с использованием предложенного многопоточного метода сжатия.

В среднем, степень сжатия составляет 8.5 раза.

Дальнейшие планы



Дальнейшие планы по использованию методов сжатия в суперкомпьютерном моделировании

- Уменьшение вычислительной сложности алгоритма сжатия, использование неитеративных методов кодирования, использование большего набора методов для кодирования
- Разработка методов сжатия для типов данных, отличных от изображений
- Интеграция разработанных методов сжатия в библиотеку ParImPr



Спасибо за внимание



Летняя школа для студентов, аспирантов и молодых ученых «Суперкомпьютерное моделирование и визуализация в научных исследованиях».

**Московский государственный университет имени М.В.Ломоносова
4 - 14 июля 2010 г.**

Московский государственный университет имени М.В.Ломоносова приглашает студентов, аспирантов и молодых ученых с 4 по 14 июля 2010 г. пройти обучение в летней школе «Суперкомпьютерное моделирование и визуализация в научных исследованиях». Школа проводится при поддержке Суперкомпьютерного консорциума университетов России на базе факультета Вычислительной математики и кибернетики совместно с Нижегородским государственным университетом имени Н.И.Лобачевского.

Тематики школы:

- Теория и технологии суперкомпьютерного моделирования в междисциплинарных научных исследованиях
- Технологии анализа и визуализации данных больших размерностей
- Управление междисциплинарными исследовательскими проектами, включающими суперкомпьютерное моделирование
- Методики обучения суперкомпьютерным технологиям для научных исследований

exascale.ru