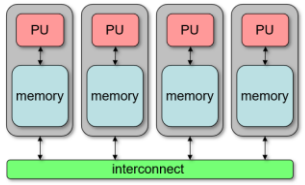
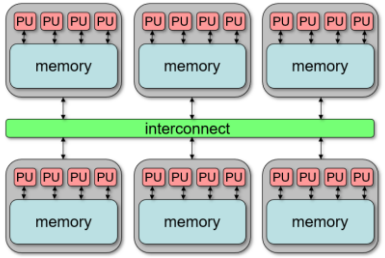
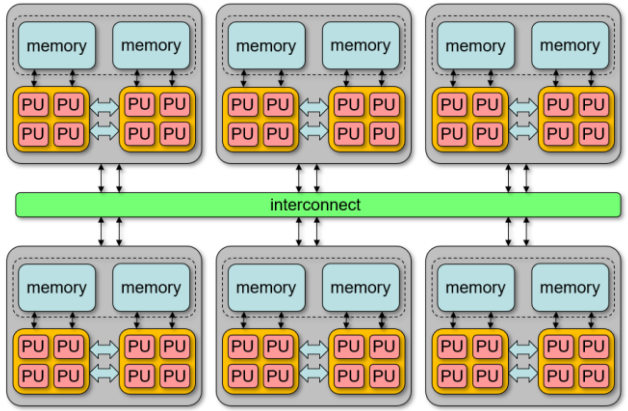
Dr. Strangebug Или: Как Я Перестал Беспокоиться И Полюбил Гетерогенные Вычисления

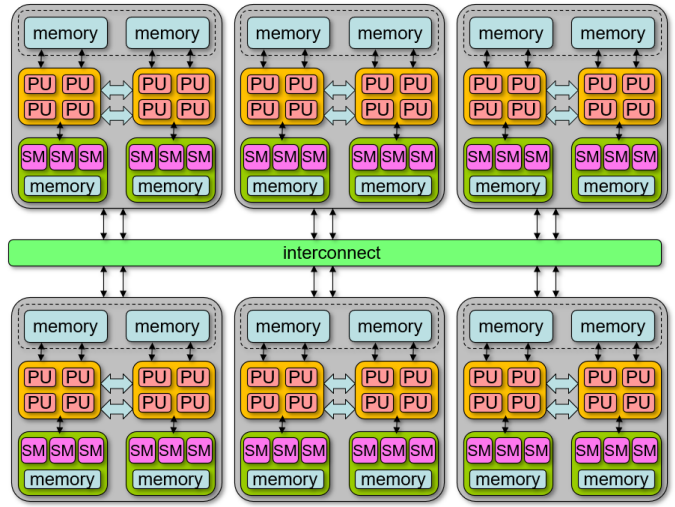
Сектор Вычислительной Аэродинамики и Аэроакустики
ИПМ им. М. В. Келдыша РАН

http://caa.imamod.ru

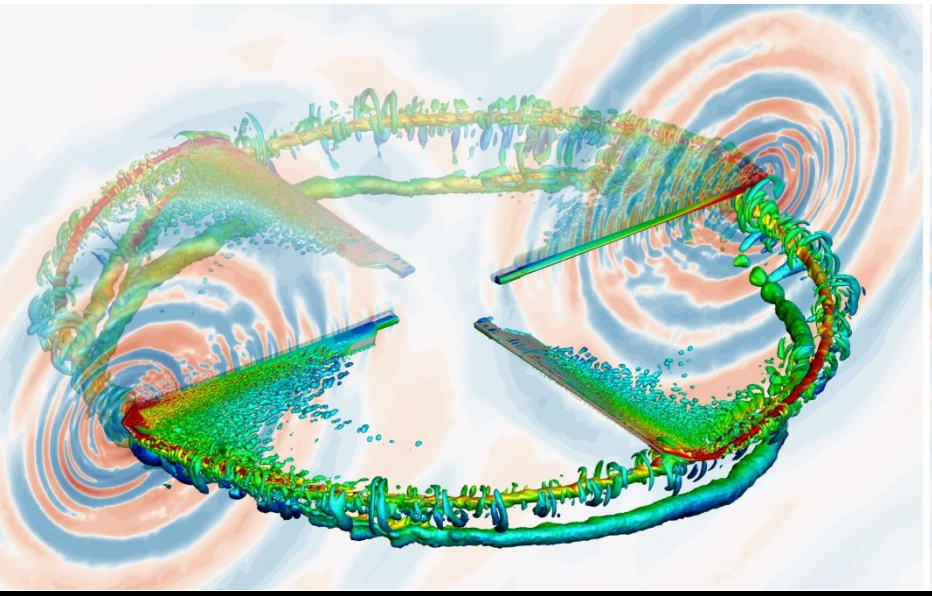


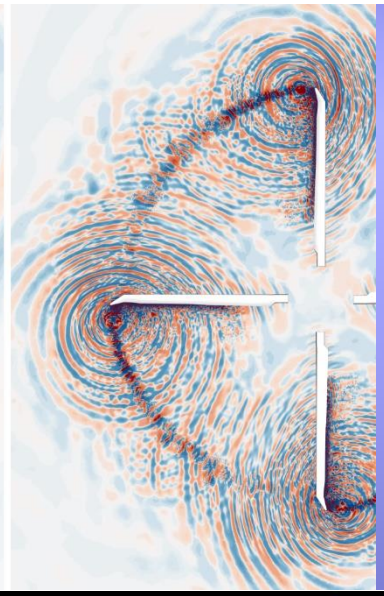


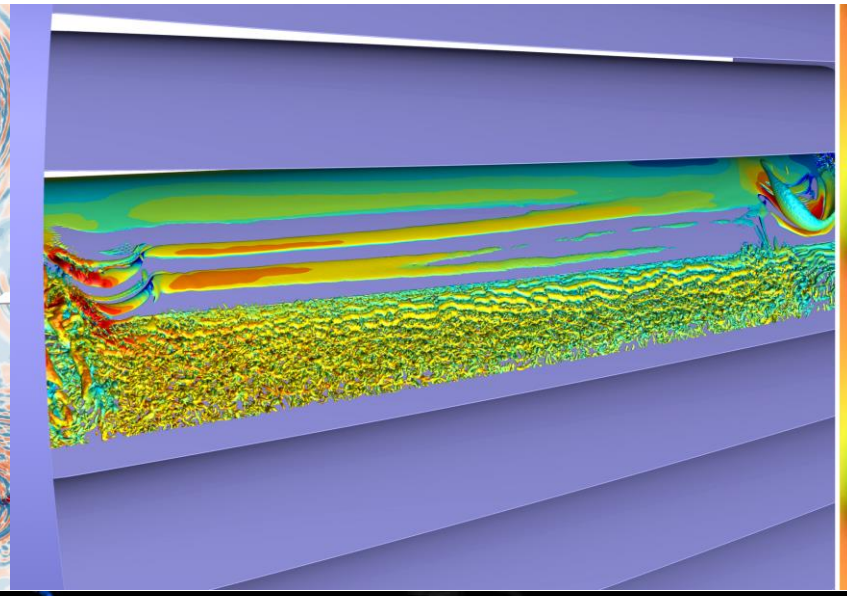


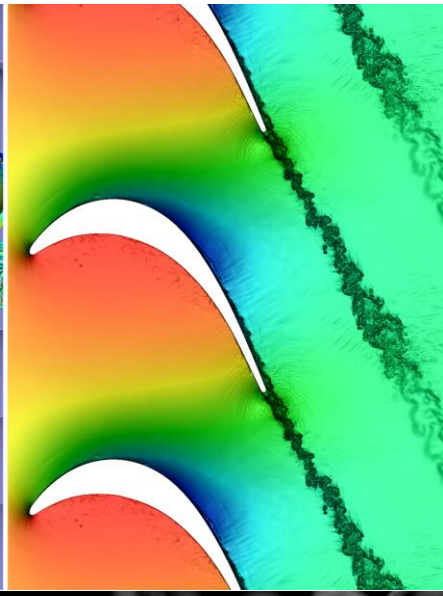


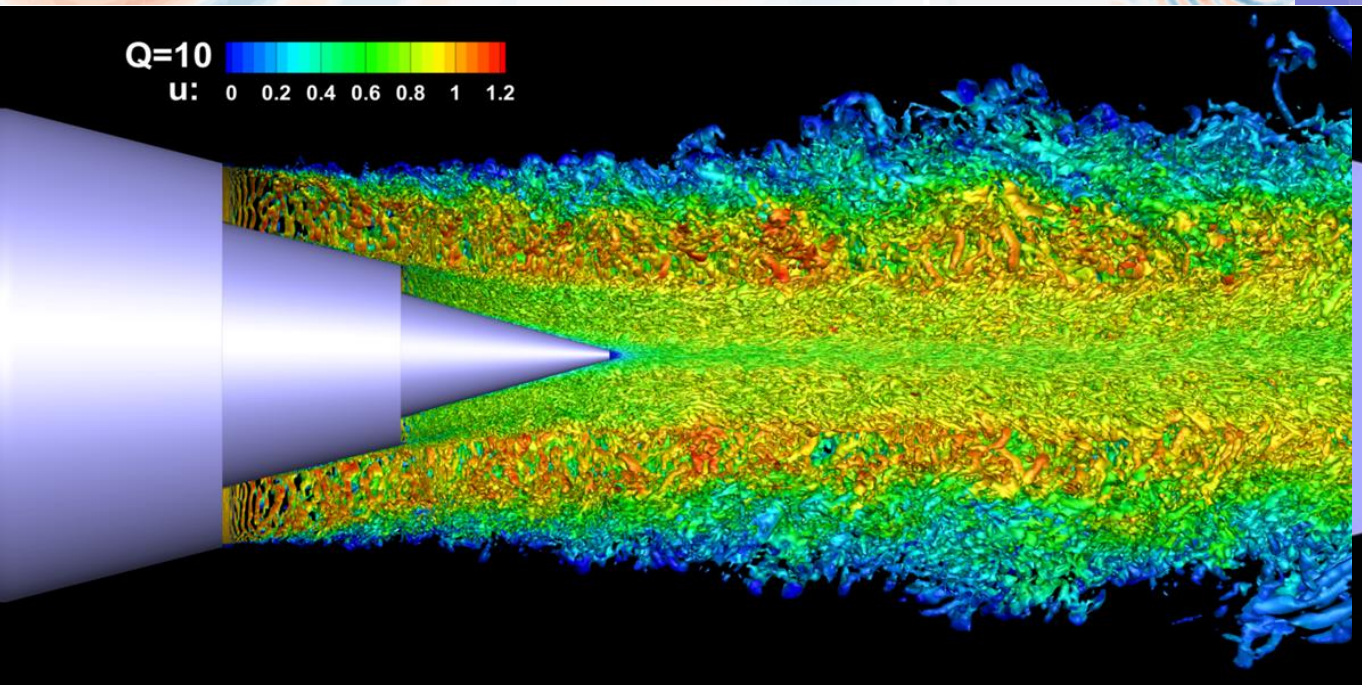
Вихреразрешающее моделирование

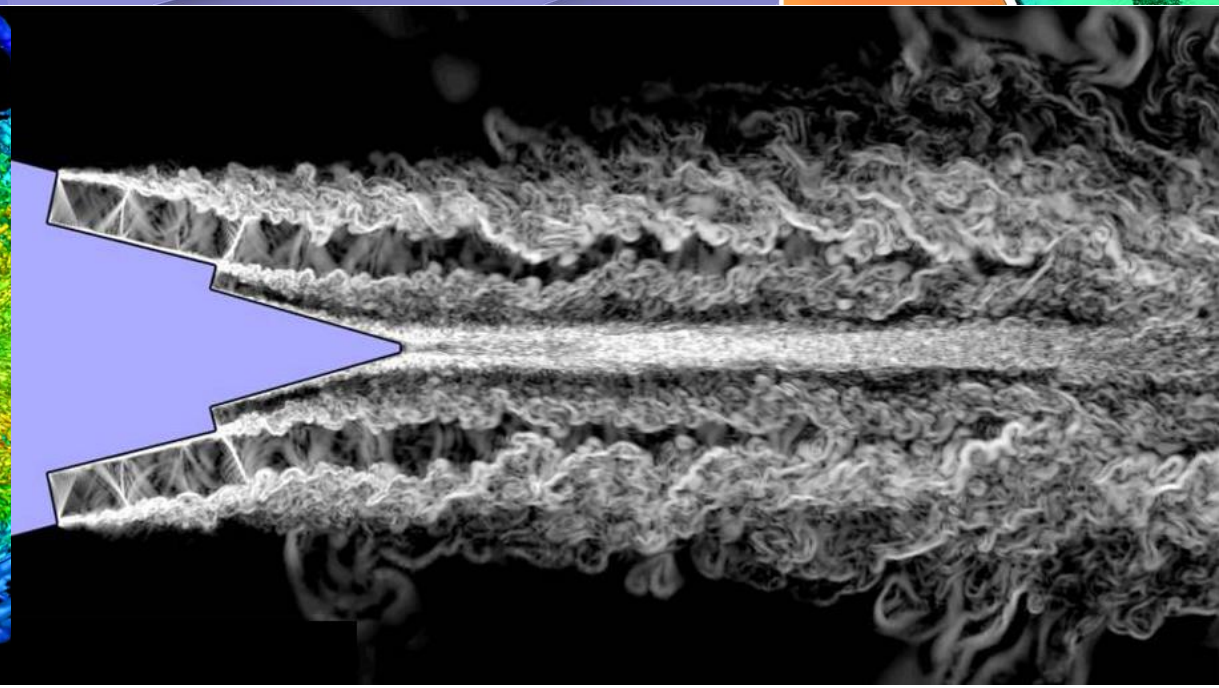












Гибридные супервычисления

CPU: Intel, AMD, IBM, ARM, Эльбрус

Manycore: Intel Xeon Phi, A64FX ARM

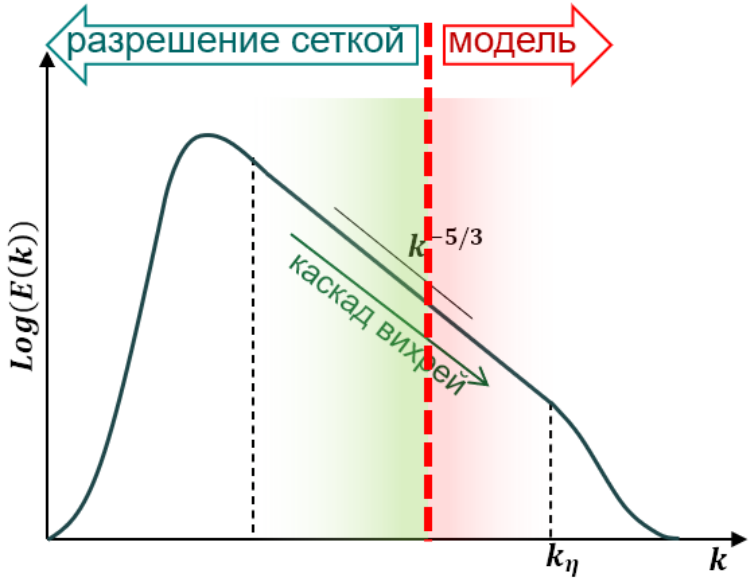
GPU: NVIDIA, AMD, Intel

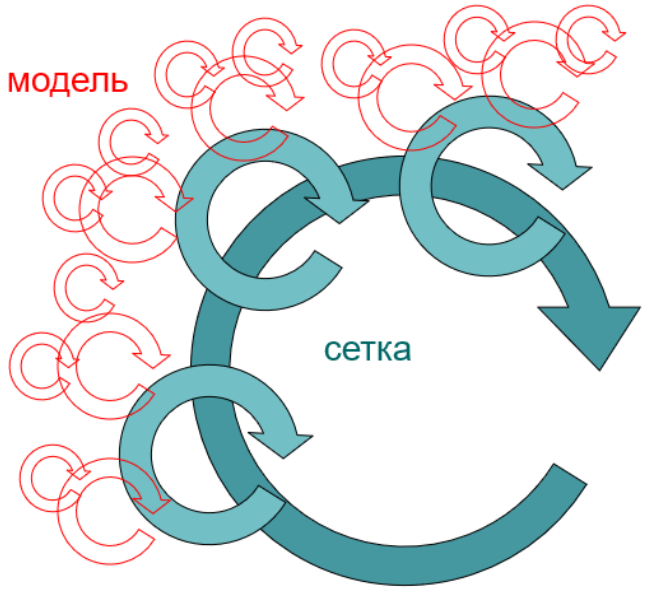
Основные понятия:

- **DM-MIMD**
- **SM-MIMD**
- **SIMD**
- **Stream processing**
- **FLOPS**
- **GB/s**
- **Latency**

Технология вихреразрешающего суперкомпьютерного моделирования

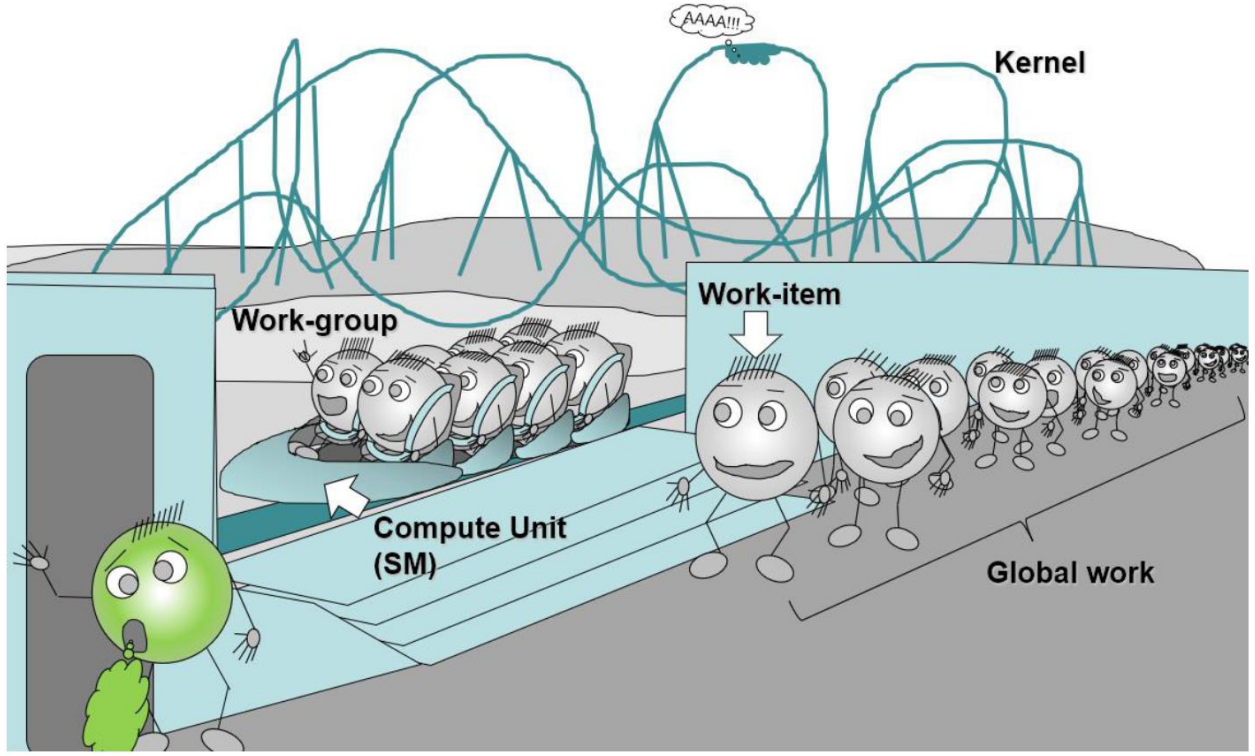
- **Схемы повышенной точности EBR**
  Неструктурированные смешанные сетки
- **Неявная схема по времени**
- **Гибридные RANS-LES методы**
  LES модель турбулентности
  Подсеточный масштаб
- **Параллельный алгоритм**
  DM-MIMD, SM-MIMD, Stream processing, SIMD
- **Переносимая гетерогенная реализация**
  MPI, OpenMP, OpenCL

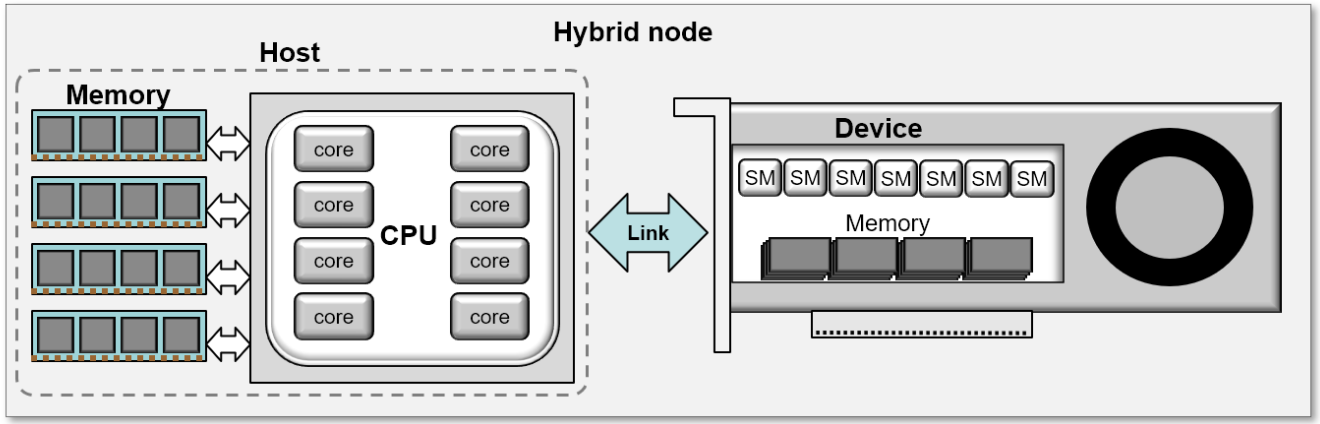


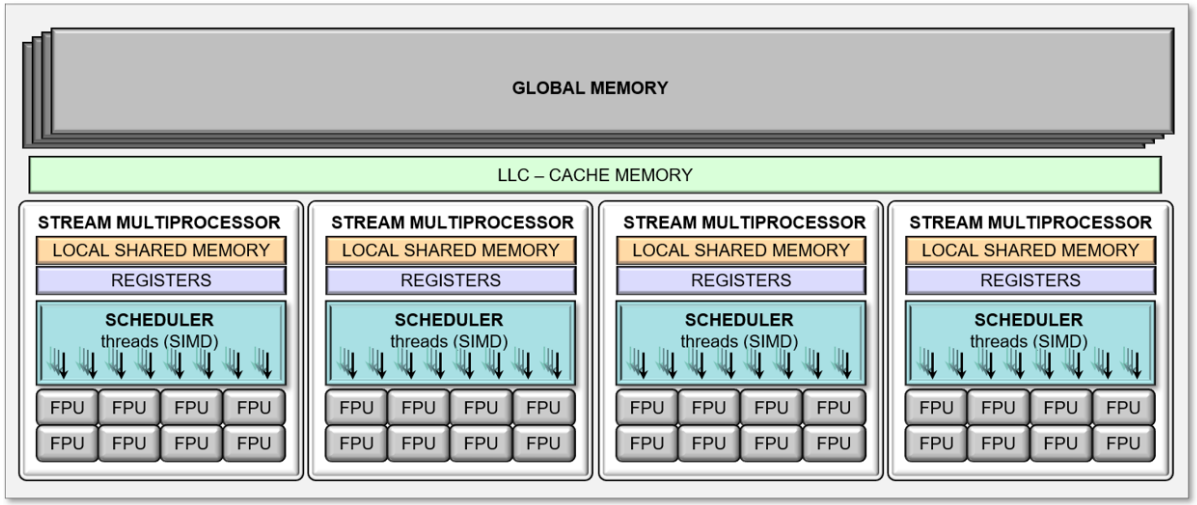


Критические проблемы в GPU-computing

- **Параллельная парадигма потоковой обработки**
  адаптация алгоритма к stream processing
- **Оперативная память**
  ее на порядок меньше, чем на CPU!
  необходимо сильно уменьшить аппетит
- **Организация обмена данными**
  Intranode + internode
  необходим асинхрон и overlap



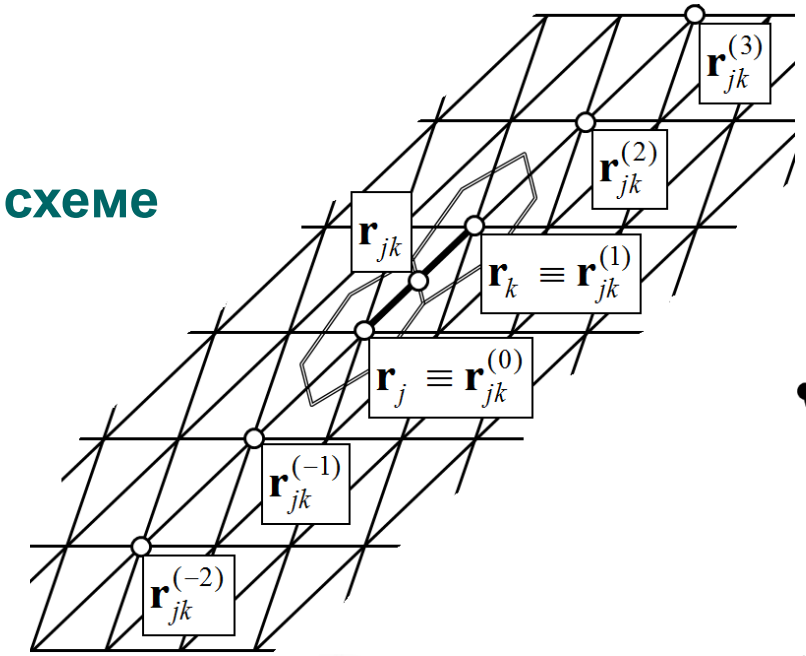


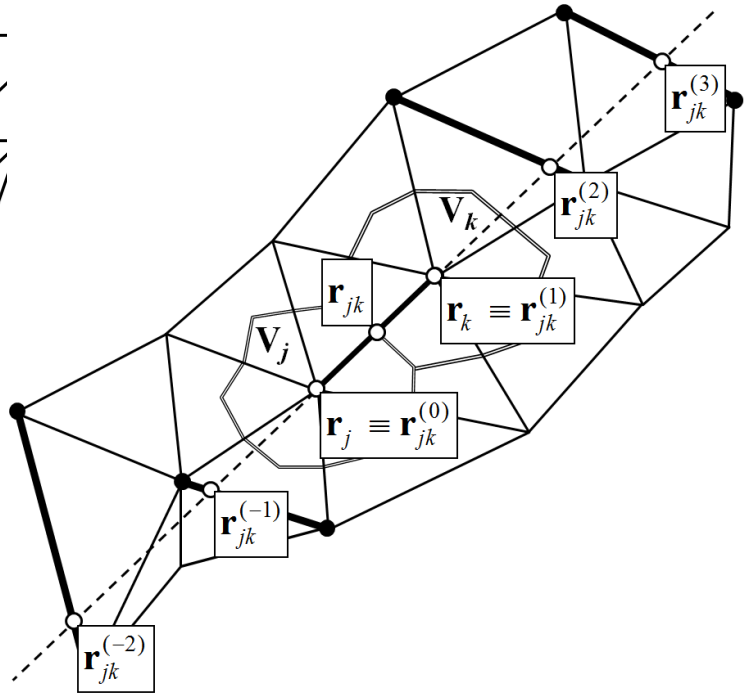


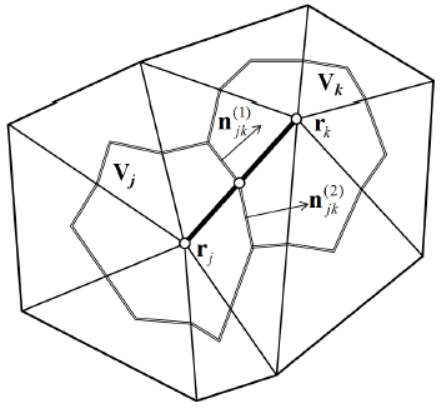
http://caa.imamod.ru/vmkvm

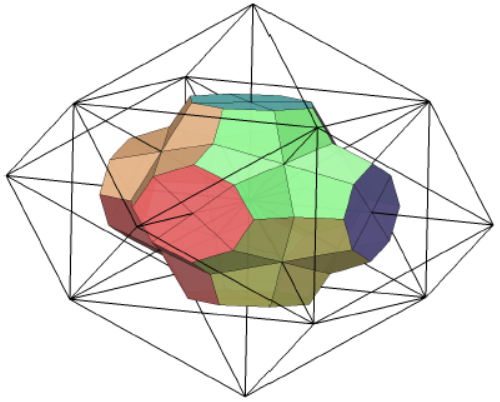
Численный метод – схема EBR

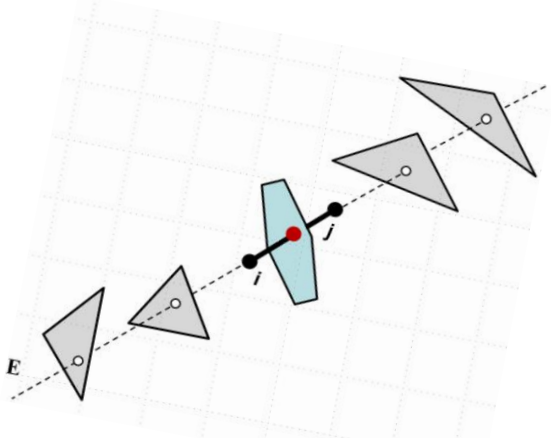
- **Сверхспособности на ТС-сетках**
- **Упрощенный якобиан в неявной схеме**
  как у схемы 1 порядка
- **Экономичность**
  - **по памяти**
  - **по вычислениям**
  - **по обменам**

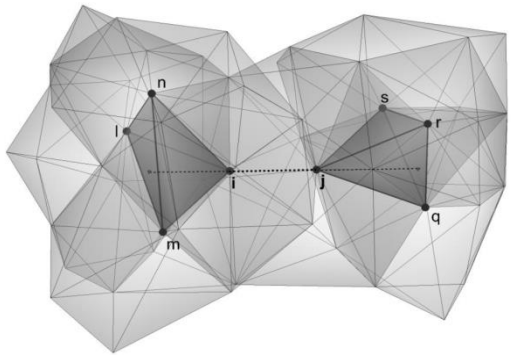












Bakhvalov Pavel, Abalakin Ilya, Kozubskaya Tatiana. Edge-based reconstruction schemes for unstructured tetrahedral meshes. International Journal for Numerical Methods in Fluids. 2016. Vol.81(6). P. 331–356. https://doi.org/10.1002/fld.4187

Bakhvalov Pavel, Kozubskaya Tatiana. EBR-WENO scheme for solving gas dynamics problems with discontinuities on unstructured meshes. Computers and Fluids. 2017. Vol. 157, p. 312-324. https://doi.org/10.1016/j.compfluid.2017.09.004

Численный метод – MLES дискретизация вязких членов H – C

P1 Галеркин для вязких членов дает 27 ненулей на гексах в портрет якобиану

Метод локальных разбиений

- является линейным
- совпадает с P1 Галёркиным на симплициальных сетках
- на 3D декартовых сетках аппроксимация лапласиана дает 7-точечную схема, а у Галёркина 27-точечный шаблон
- **можно без потери сходимости исключить из якобиана элементы, не входящие в 7-точечный шаблон**

Bakhvalov P. A. Method of local element splittings for diffusion terms discretization in edge-bases schemes. Keldysh Institute Preprints. 2020. Vol. 79. 1 – 43
https://doi.org/10.20948/prepr-2020-79-e

JCP – under review

Математическая модель – гибридный RANS-LES

Незонные гибридные методы семейства DES

- **DES, DDES, IDDES**

M. Shur, P. Spalart, M. Strelets, A. Travin. An Enhanced Version of DES with Rapid Transition form RANS to LES in Separated Flows. Flow, Turbulence and Combustion. 95 (2015) 709 – 737 https://doi.org/10.1007/s10494-015-9618-0

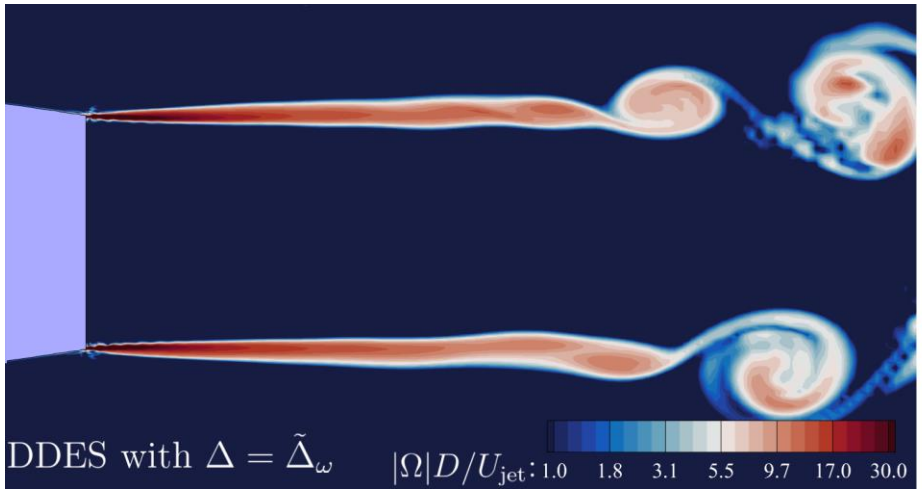
- **RANS SA, SST**

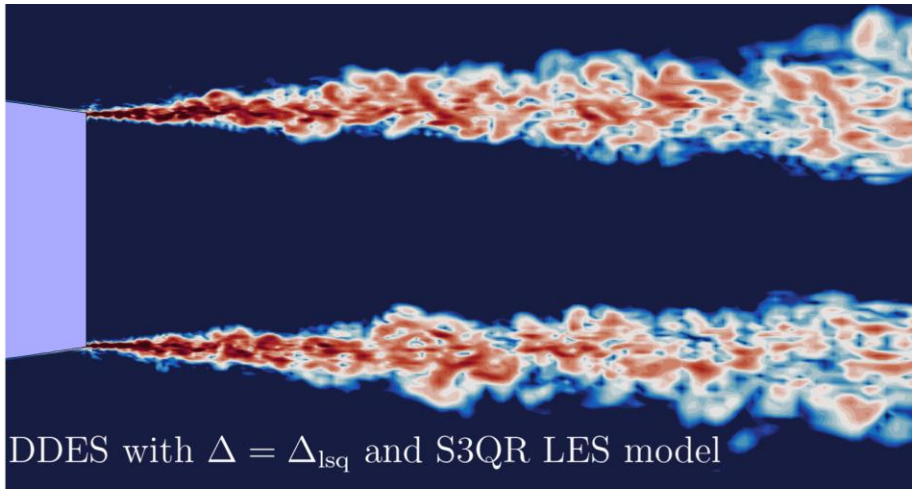
- **LES модели семейства S3**

F. X. Trias, D. Folch, A. Gorobets, A. Oliva. Building proper invariants for eddy-viscosity subgrid-scale models. Physics of Fluids 27 (2015) 065103 https://doi.org/10.1063/1.4921817

- **Подсеточный масштаб Delta-LSQ**

F.X.Trias, A.Gorobets, M.H.Silvis, R.W.C.P.Verstappen, and A.Oliva. A new subgrid characteristic length for turbulence simulations on anisotropic grids. Physics of Fluids.29 (2017) 115109 https://doi.org/10.1063/1.5012546

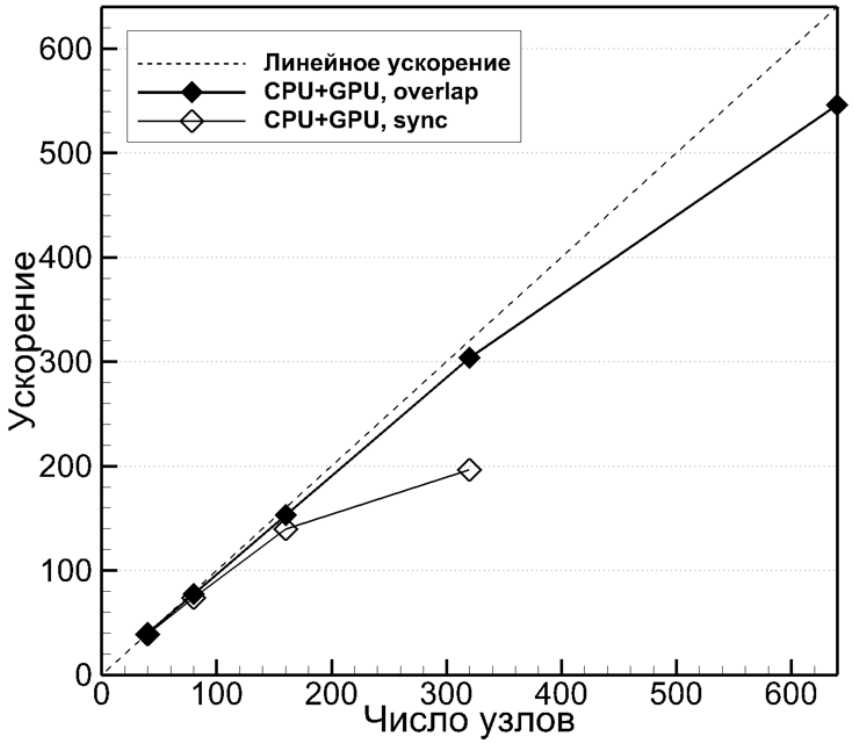


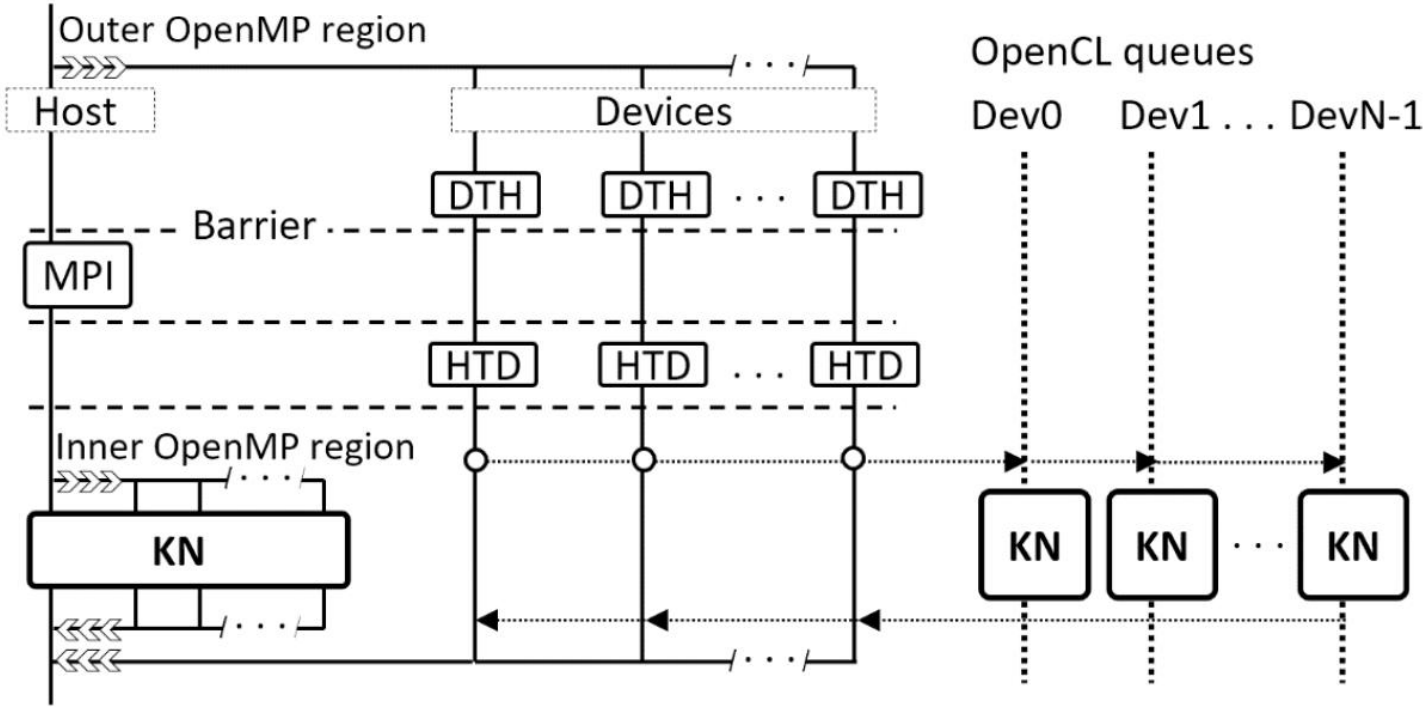


Предыдущие гетерогенные реализации

Гетерогенный код Тапир

Cell-center, полиномиальная реконструкция, явная схема



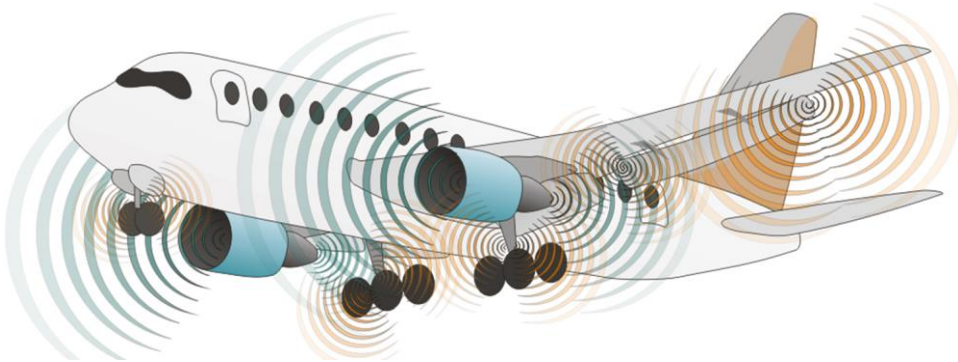


A.Gorobets, S.Soukov, P.Bogdanov. Multilevel parallelization for simulating turbulent flows on most kinds of hybrid supercomputers. Computers and Fluids. Vol. 173, 2018, pp. 171–177.

S. A. Soukov, A. V. Gorobets. Heterogeneous Computing in Resource-Intensive CFD Simulations. Doklady Mathematics. Vol. 98, No. 2, 2018, pp. 1–3.

NOISETTE – игровая площадка и код для приложений

- **C++, MPI, OpenMP, OpenCL**
- **Схемы повышенной точности EBR**
  неструктурированные смешанные сетки
- **Вихреразрешающее моделирование**
  сжимаемые течения,
  аэродинамика и аэроакустика
  гибридные RANS-LES методы
- **HPC реализация**
  масштабируемые параллельные алгоритмы
  гетерогенные вычисления, …
- **Вычислительные технологии**
  IBC, адаптация сетки, STG,
  параллельный постпроцессор,
  FW/H, beamforming,
  скользящие сетки, ротор-статор,
  ускорение расчета, …



http://caa.imamod.ru/noisette


РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ
№ 2013619362


РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ
№ 2016662466


РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ
№ 2013619363


РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ
№ 2012614096


РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО
о государственной регистрации программы для ЭВМ
№ 2015661493

Подготовка к вычислениям на GPU

Проблема: код слишком сложно отлаживать, сложно контролировать качество

Решение:

- **Автоматическое тестирование с расширенным QA покрытием**
- **Конфигурация безопасного режима со встроенным стек-трейсером**
- **Контейнеры с контролем доступа по двум индексам, по типу, именованные**
- **Встроенный мониторчик выделения памяти для отчетов и ловли утечек**
- **Встроенное инструментальное профилирование**

A. Gorobets, P. Bakhvalov. Improving Reliability of Supercomputer CFD Codes on Unstructured Meshes. Supercomputing Frontiers and Innovations. 2019. Vol. 6, No. 4, pp. 44-56. http://dx.doi.org/10.14529/jsfi190403

Конфигурация SAFE MODE

Программные уровни

LL – Low Level – высокочастотные функции
performance-critical, обработка одного сеточного объекта
ифы, свичи, ассёрты, вызовы функций – создают оверхэд

HL – High Level – низкочастотные функции
обработка наборов сеточных объектов
ифы, свичи, ассёрты, вызовы функций – пренебрежимы

- **Low-level проверки включены в конфиге SAFE MODE, выключены в релизе**
- **QA процедура – в сейфмоде**
- **Проверка доступа к контейнерам – массивам V[i], блочным массивам V[i][j] по обоим индексам**
- **Сейфмод медленнее релиза раза в полтора из-за множества высокочастотных проверок**

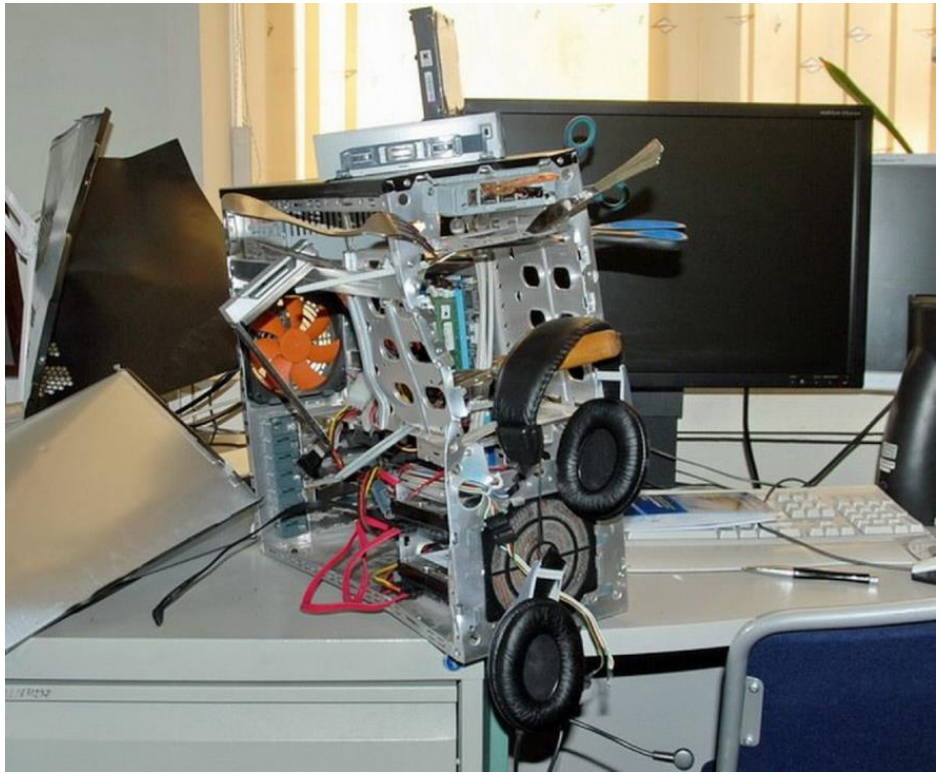


Иллюстрация воздействия нестабильных программных ошибок вида **Strange bug** на разработчика и сопутствующий матчасти

A. Gorobets, P. Bakhvalov. Improving Reliability of Supercomputer CFD Codes on Unstructured Meshes. Supercomputing Frontiers and Innovations. 2019. Vol. 6, No. 4, pp. 44-56. http://dx.doi.org/10.14529/jsfi190403

Подготовка к вычислениям на GPU

Проблема: код слишком прожорлив по памяти, а на GPU ее совсем мало

Решение:

- **Новый метод расчета вязких потоков MLES в разы сократил потребление памяти**
  Bakhvalov P. A. Method of local element splittings for diffusion terms discretization in edge-bases schemes. Keldysh Institute Preprints. 2020. Vol. 79. 1 – 43

- **Упрощенный метод MLES еще сильно сократил потребление памяти**
  Галёркин тоже упрощается по коэффициентам, тоже записывается в порёберной форме,
  но якобиан не упрощабелен

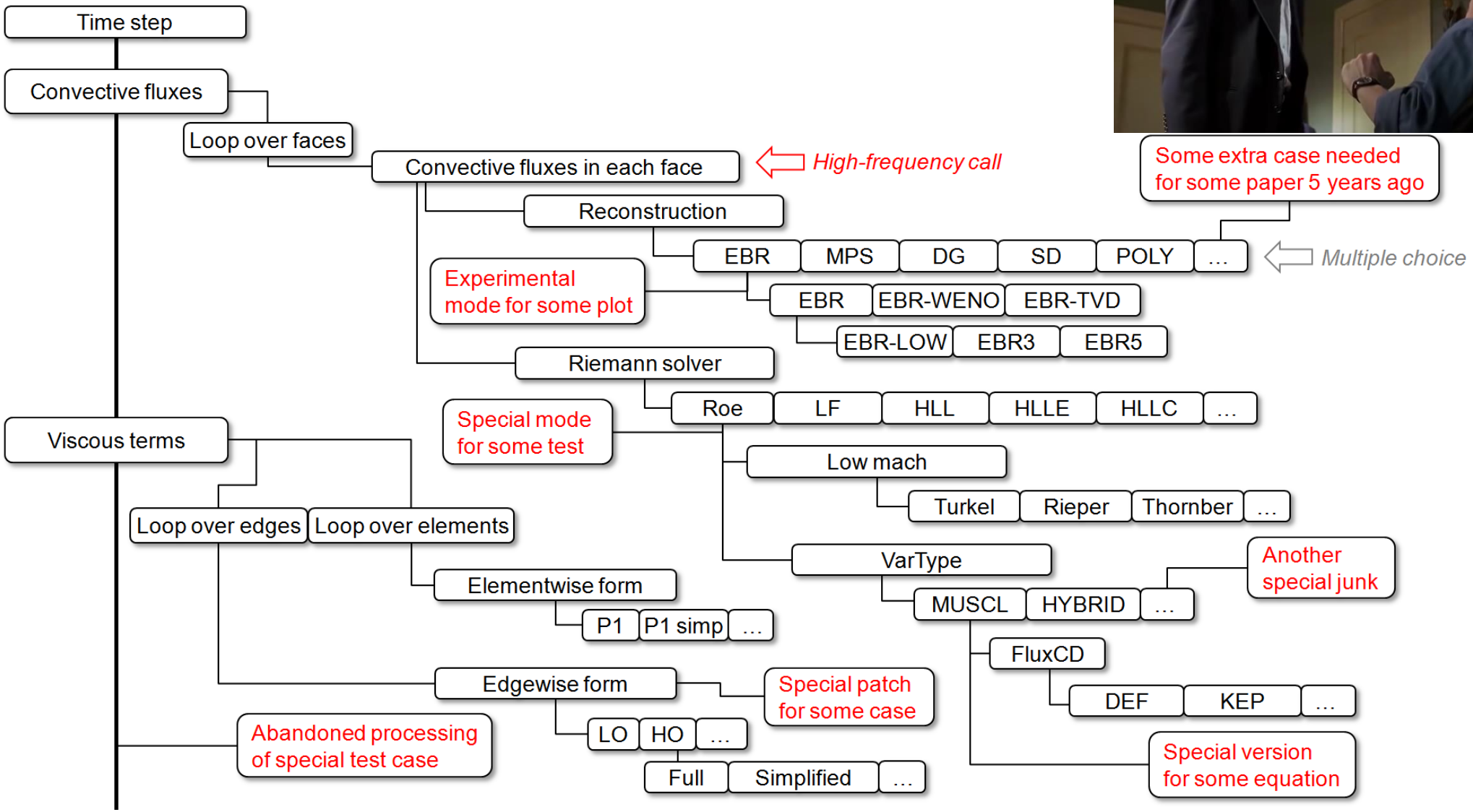
- **Смешанная точность FP64/FP32 – еще почти двукратное сокращение потребления и трафика**
  все в двойной точности кроме тяжелых массивов коэффициентов дискретных операторов для вязких потоков и якобиана неявной схемы

Точность результатов не пострадала, как было показано тут:

A. Gorobets, P. Bakhvalov, A. Duben, P. Rodionov. Acceleration of NOISEtte Code for Scale-resolving Supercomputer Simulations of Turbulent Flows. Lobachevskii Journal of Mathematics. Vol. 41, No. 8, 2020, pp. 1463–1474.
http://dx.doi.org/10.1134/S1995080220080077

Подготовка к вычислениям на GPU

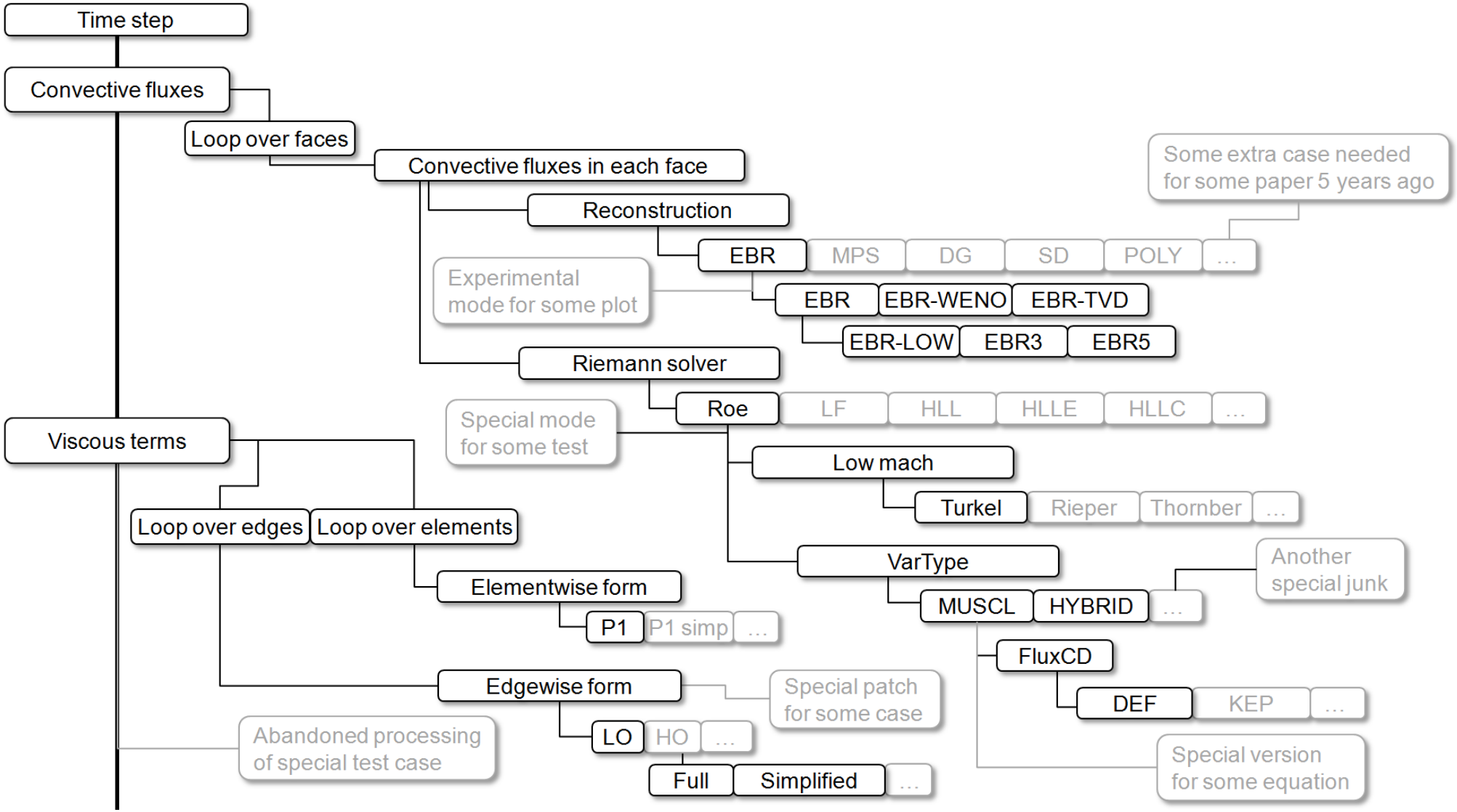
Проблема: код слишком сложный



Подготовка к вычислениям на GPU

Проблема: код слишком сложный

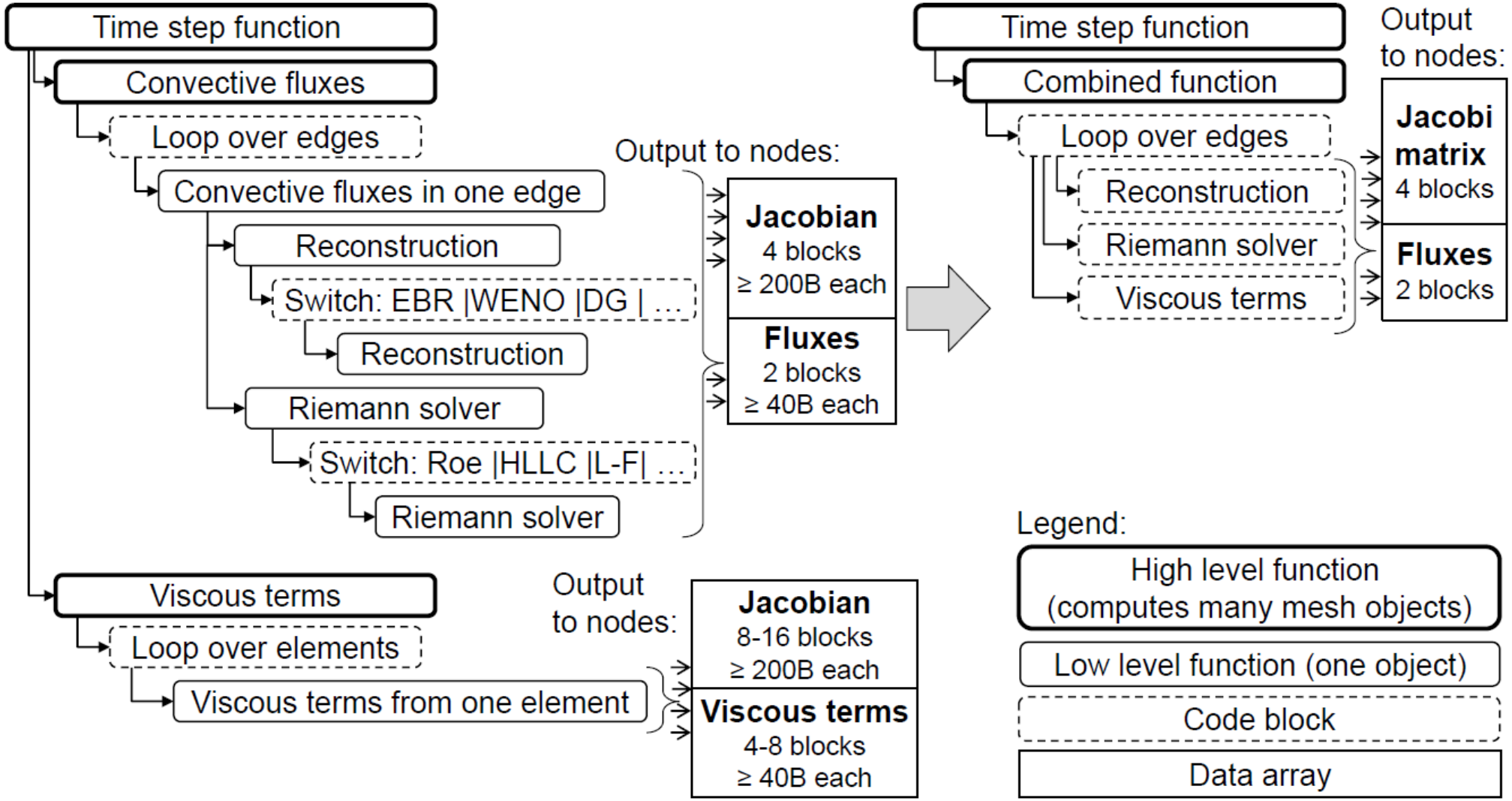
Решение: уборка мусора под ifdef, прочистка свитчей и ветвлений на нижнем уровне



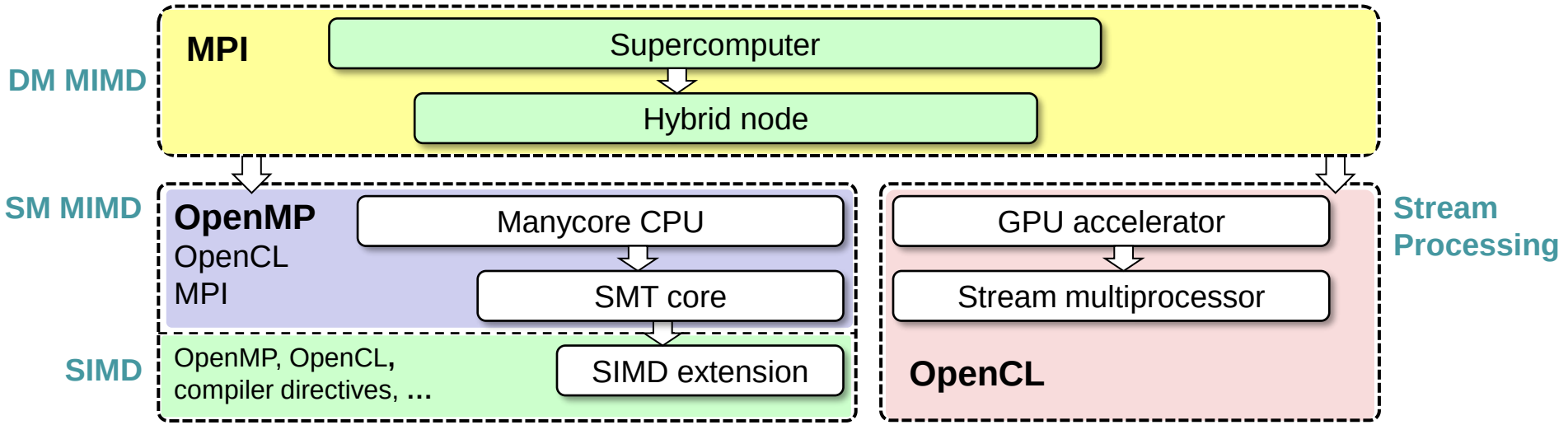
Подготовка к вычислениям на GPU

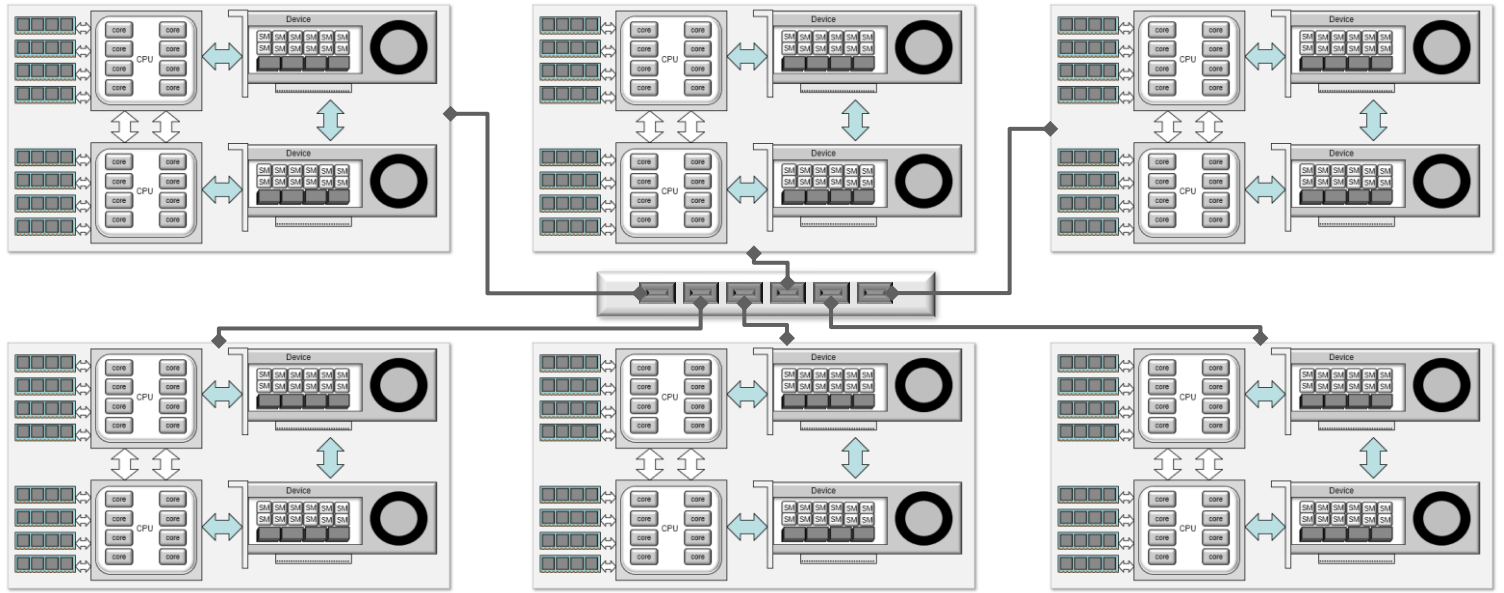
Проблема: много вызовов функций на нижнем уровне, доступ в якобиан не по разу

Решение: комби-функции для продуктового режима – “бизнес-ланч”
устранение ветвлений, внутренних вызовов функций, сокращение memory-трафика



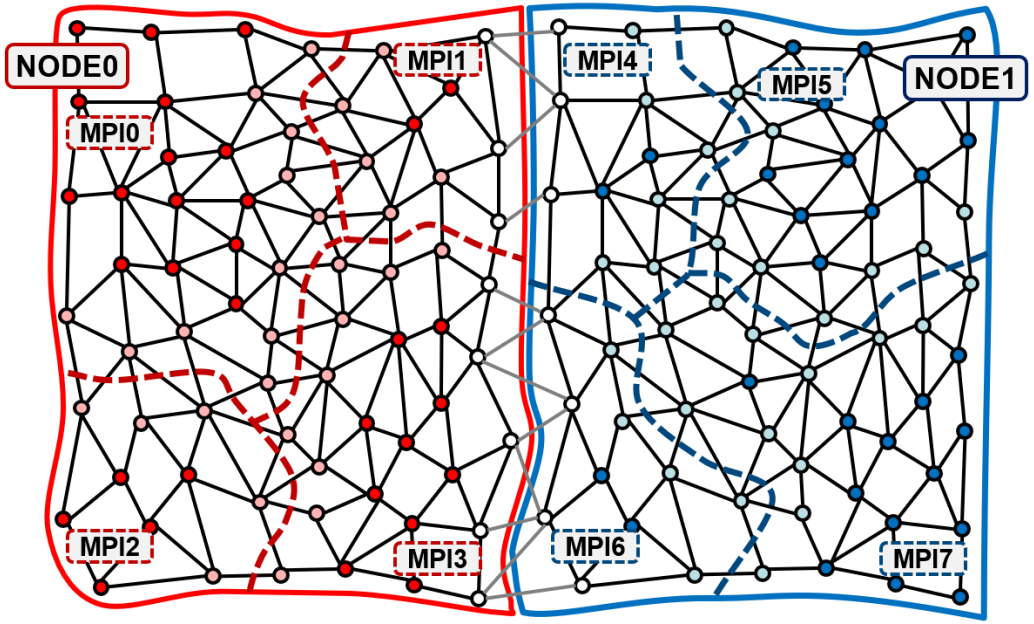
Многоуровневое распараллеливание MPI + OpenMP + OpenCL

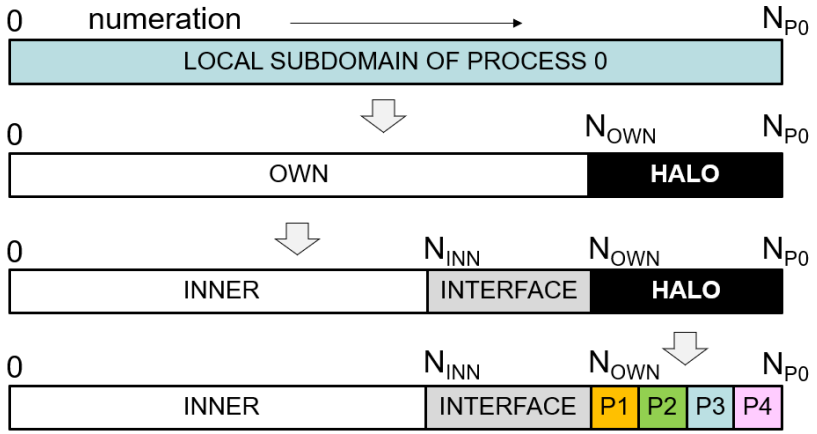


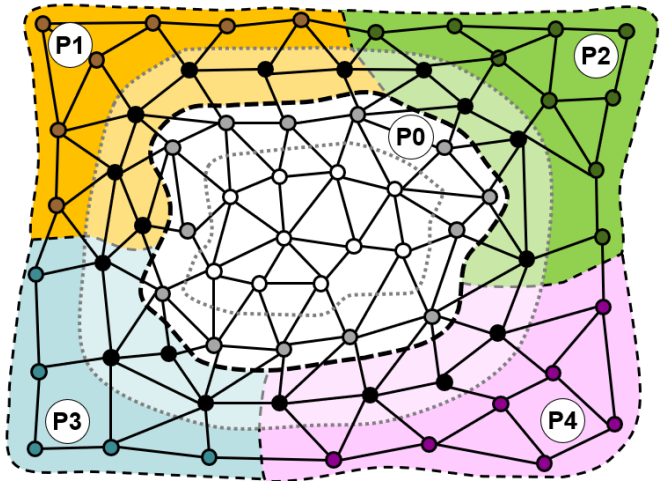


MPI распараллеливание

- **Многоуровневая декомпозиция**
  уменьшение сетевого трафика
- **Переупорядочивание Inner-Interface-Halo**
- **Overlap для вычислений и обменов**
  сокрытие накладных на передачу данных
- **OpenMP-параллельная обработка MPI сообщений**
  для ускорения обновления гало
- **Улучшенная реализация обменов**
  один список узлов по всем соседям
  переупорядочивание гало
- **Динамическая схема обмена**
  для скользящих сеток



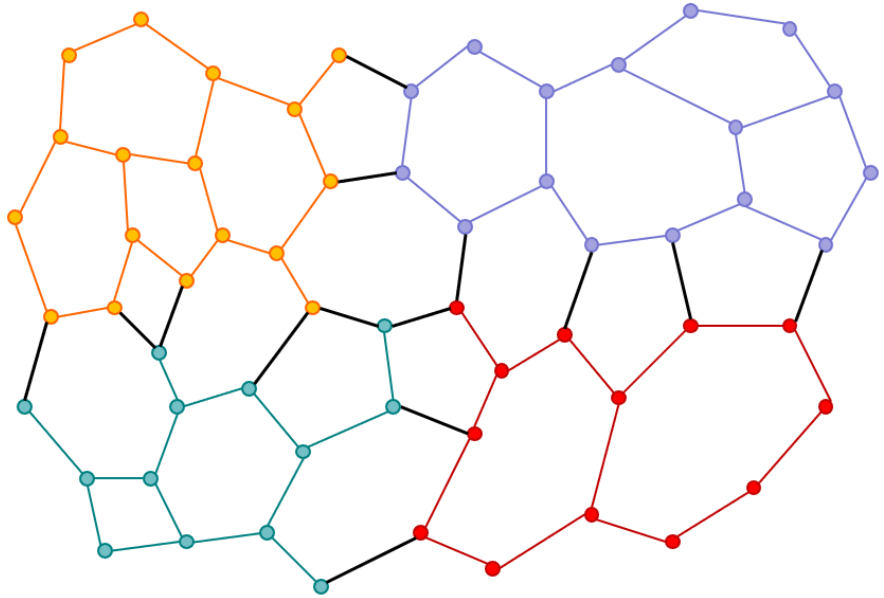




A.Gorobets. Parallel Algorithm of the NOISEtte Code for CFD and CAA Simulations. Lobachevskii Journal of Mathematics. 2018, Vol. 39, No. 4, pp. 524–532. https://doi.org/10.1134/S1995080218040078

OpenMP распараллеливание

- **Переход с параллелизма циклов на декомпозицию**
- **Переупорядочивание по подобластям нитей**
- **RCM для улучшения локальности доступа к памяти**
- **Многопоточная обработка MPI обменов**



EBR5, неявная BDF2, IDDES, сетка >1M узлов

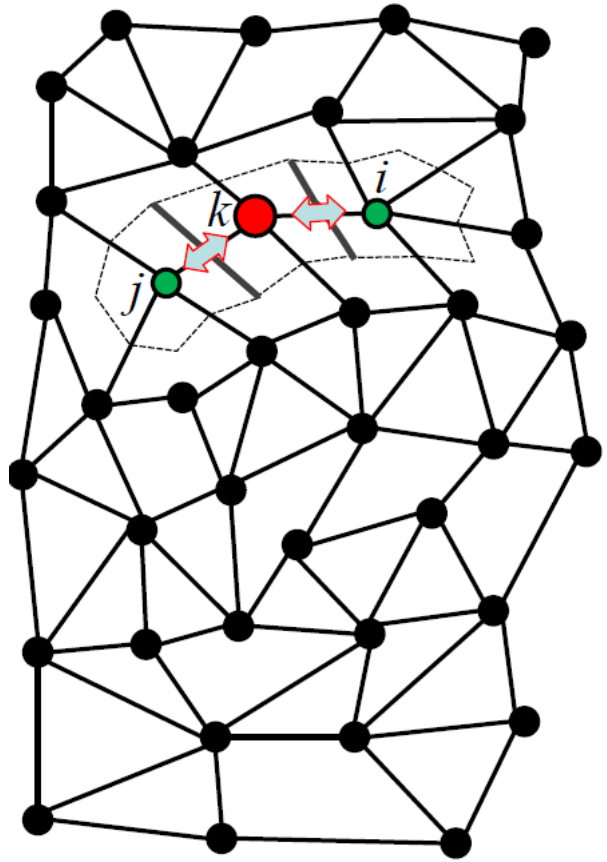
Intel Xeon	Ядер	Ускорение
E5-2683 v4	16	**11**
Gold 6142	16	**13.5**
Phi 7250	68	**92**

A.Gorobets. Parallel Algorithm of the NOISEtte Code for CFD and CAA Simulations. Lobachevskii Journal of Mathematics. 2018, Vol. 39, No. 4, pp. 524–532. https://doi.org/10.1134/S1995080218040078

Адаптация к потоковой обработке на GPU

Устранение зависимостей и гонки в циклах по ребрам

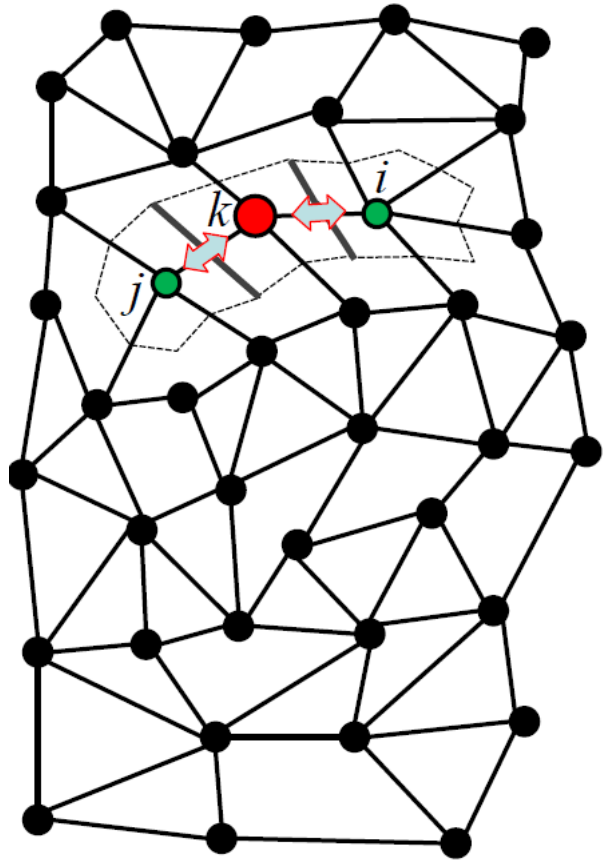
- **атомики**
- **дублирование вычислений**
- **раскраска графа**
- **разделение на 2 этапа с промежуточным массивом:**
  1. считаем результат в цикле по ребрам
  2. суммируем в узлы в цикле по узлам



Адаптация к потоковой обработке на GPU

Устранение зависимостей и гонки в циклах по ребрам

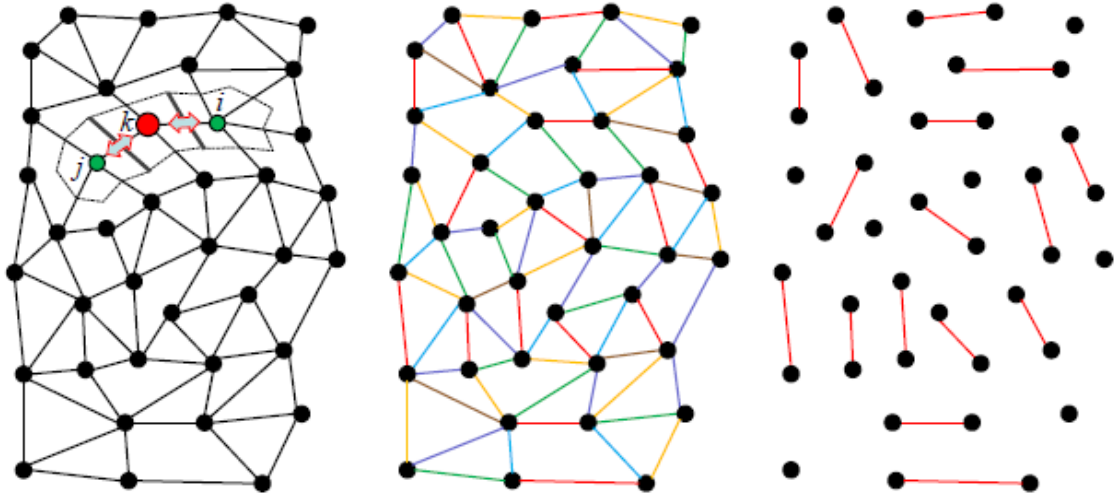
- ~~атомики~~
- дублирование вычислений
- раскраска графа
- разделение на 2 этапа
  с промежуточным массивом:
  1. считаем результат в цикле по ребрам
  2. суммируем в узлы в цикле по узлам



Адаптация к потоковой обработке на GPU

Устранение зависимостей и гонки в циклах по ребрам

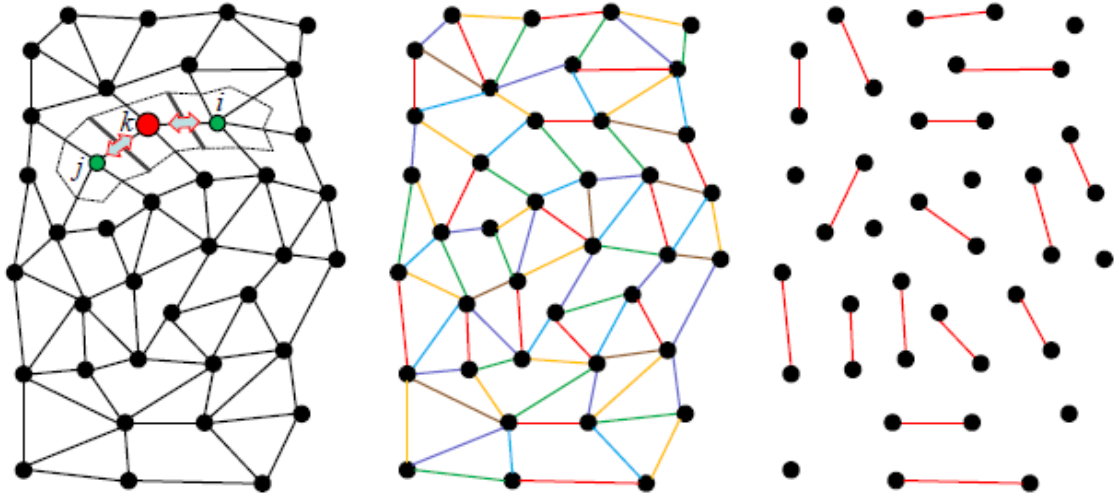
- ~~атомики~~
- ~~дублирование вычислений~~
- раскраска графа
- разделение на 2 этапа
  с промежуточным массивом:
  1. считаем результат в цикле по ребрам
  2. суммируем в узлы в цикле по узлам



Адаптация к потоковой обработке на GPU

Устранение зависимостей и гонки в циклах по ребрам

- ~~**атомики**~~
- ~~**дублирование вычислений**~~
- ~~**раскраска графа**~~
- **разделение на 2 этапа с промежуточным массивом:**
  1. считаем результат в цикле по ребрам
  2. суммируем в узлы в цикле по узлам



Адаптация к потоковой обработке на GPU

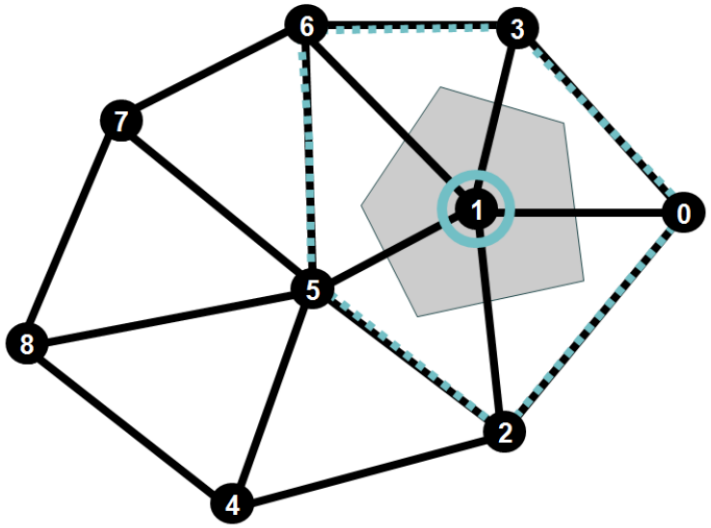
Устранение зависимостей и гонки в циклах по ребрам

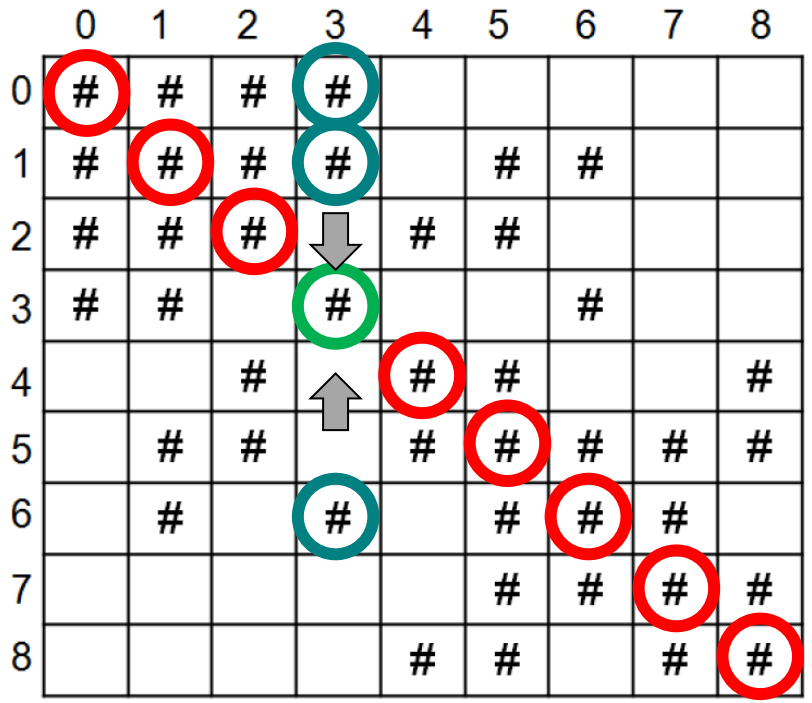
- ~~атомики~~
- ~~дублирование вычислений~~
- ~~раскраска графа~~
- ~~разделение на 2 этапа с промежуточным массивом:~~
  1. считаем результат в цикле по ребрам
  2. суммируем в узлы в цикле по узлам
- **Что же делать!?**
  Все пропало! Все пропало!

Адаптация к потоковой обработке на GPU

Устранение зависимостей и гонки в циклах по ребрам

- ~~атомики~~
- ~~дублирование вычислений~~
- ~~раскраска графа~~
- **разделение на 2 этапа**
  с промежуточным массивом:
  1. считаем результат в цикле по ребрам
  2. суммируем в узлы в цикле по узлам
- **Zero column sum**
  пишем в якобиан только внедиагональные блоки
  собираем диагональ суммированием по столбцам





Вычисления на OpenCL

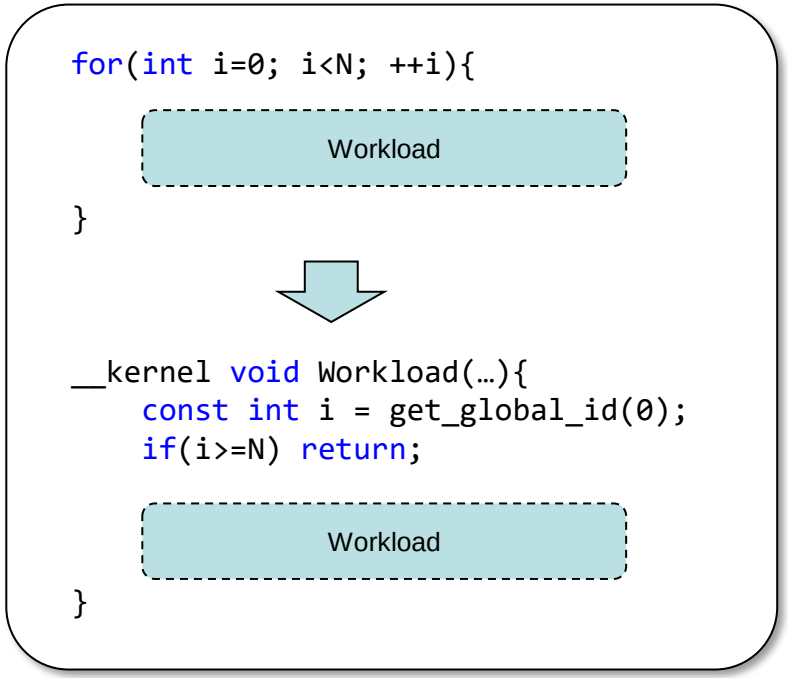
- **Максимальное подобие CPU и OpenCL версий, общий формат данных**
  для упрощения изменений и поддержания кода в форме

- **Генерация кернел кода на рантайме**
  чтобы избавиться от if-else-й и switch-ей

- **Работа с множественными файлами исходного кода**
  обработка лога компилятора для подстановки файлов и строк

- **Overlap – перекрытие обменов и вычисления**
  сокрытие расходов на передачу данных

- **Минимизация заданий work-item-ов**
  для повышения загрузки потоковых мультипроцессоров



- **Смешанная точность FP64/FP32**
  одинарная для якобиана и некоторых прожорливых дискретных операторов –
  чтобы сильно уменьшить потребление памяти и memory-трафик.
  конечно, без потери точности результата

- **Всяческое переупорядочивание узлов**
  блочный RCM, лексикографическая сортировка ребер – для повышения локальности

Гетерогенная реализация

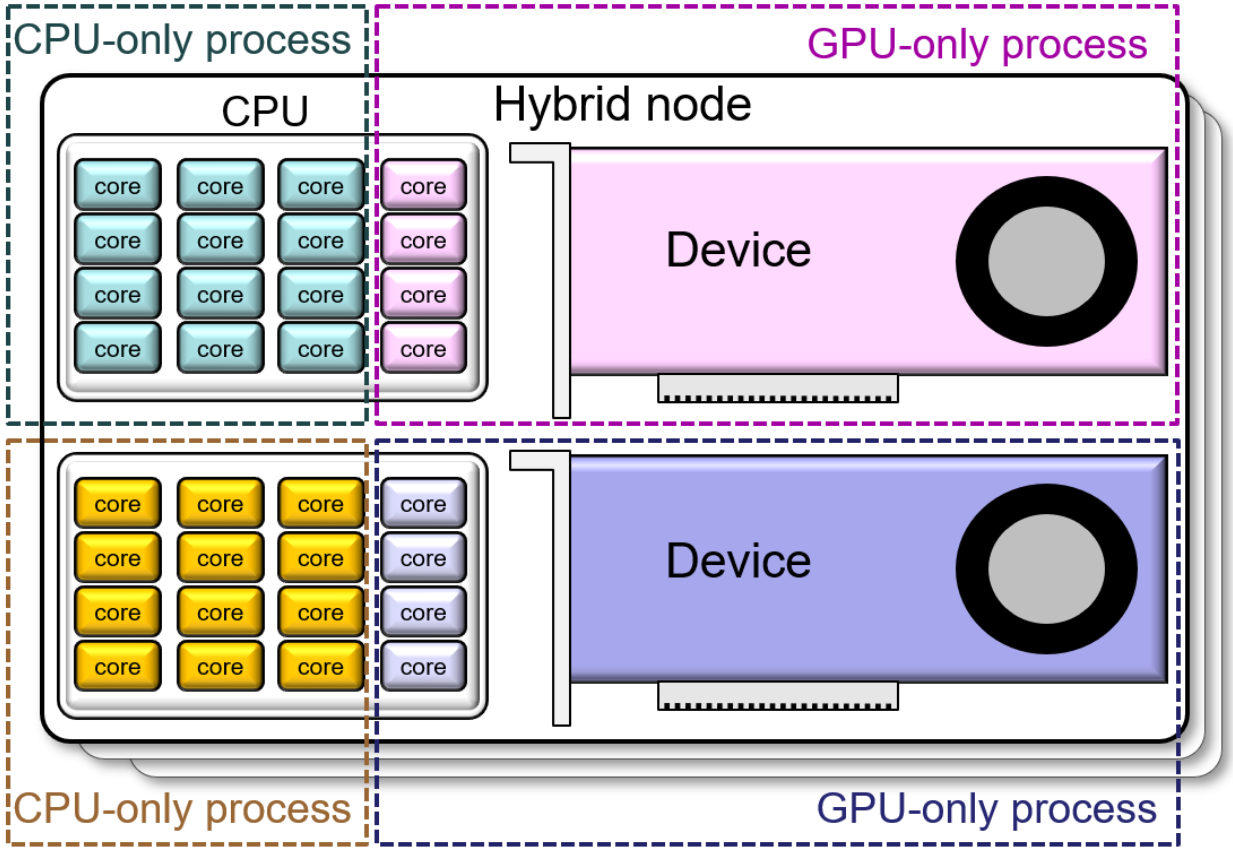
Гетерогенные вычисления CPU & GPU

Подход: один процесс – одно устройство

MPI процессы делятся на CPU и GPU роли

Двухуровневая декомпозиция
с эмпирической балансировкой

Полная согласованность CPU и GPU версий



Защищенность от внесения ошибок

Автоматизированная QA система с расширенным покрытием
сравнение с эталонными решениями последовательной CPU версии

Автоконтроль согласованности версий OpenCL vs CPU
чтобы CPU и GPU делали ровно одно и то же

- **проверка по каждому кернелу всех выходных данных**
- **проверка полных шагов по времени**

```
          INIT absS DIFF REL NORM = 1.94547e-16
   INIT viscousTurb DIFF REL NORM = 6.73523e-19
       FLUX CE TURB DIFF REL NORM = 2.37454e-21
            INIT UA DIFF REL NORM = 0
            INIT UB DIFF REL NORM = 0
            INIT CE DIFF REL NORM = 2.37454e-21
       INIT UA_Phys DIFF REL NORM = 0
            FLUX CE DIFF REL NORM = 2.73317e-15
       UB before BC DIFF REL NORM = 0
         BC FLUX CE DIFF REL NORM = 2.73393e-15
        NPREP2 CE DIFF REL NORM = 2.73393e-15
               resC DIFF NORM 9.9592e-16
               resL DIFF NORM 5.18985e-17
         Solver DUA DIFF REL NORM = 2.33053e-14
         RenewUA UA DIFF REL NORM = 1.22693e-17
         BCForce UA DIFF REL NORM = 1.22693e-17
    CellSchemeFlags DIFF REL NORM = 0
       UA after CHK DIFF REL NORM = 1.22693e-17
       INIT UA_Phys DIFF REL NORM = 2.63178e-18
FINAL ACCUMULATED ERROR res QA = 1.28782e-13
Doing CPU vs GPU whole timestep test
-----------------------------------------------------------------
    FULL TEST CHECK DIFF REL NORM = 1.83438e-17
```

Производительность: только CPU

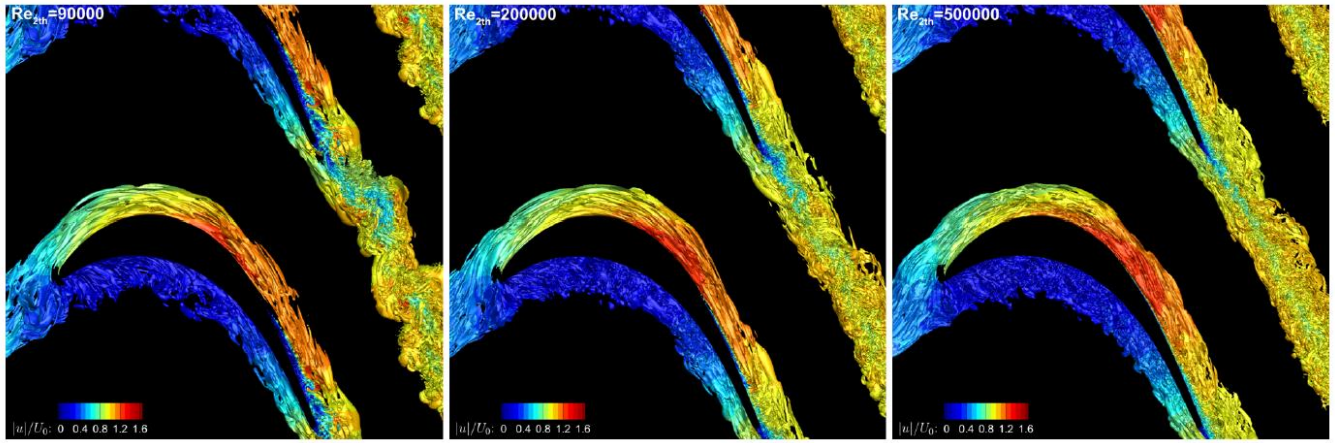
Численная конфигурация:

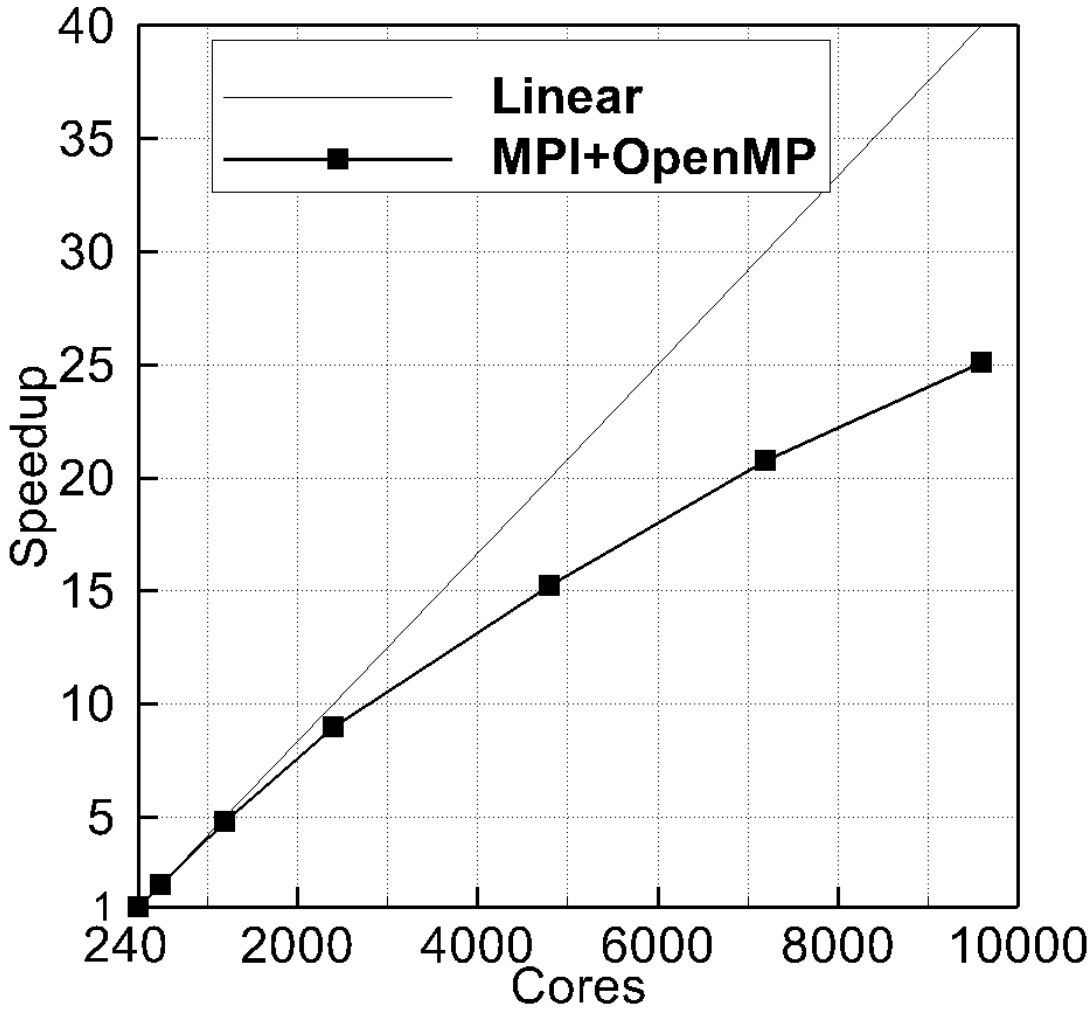
- Схема EBR5
- Неявка BDF2 с солвером BiCGSTAB
- Гибридный RANS-LES подход IDDES
- Смешанная точность FP64/32

Параллельная конфигурация: MPI+OpenMP, NT=24

Сетка: 80M узлов (обтекание лопатки ТНД)

Узлы: 2 x 24C Intel Xeon Platinum





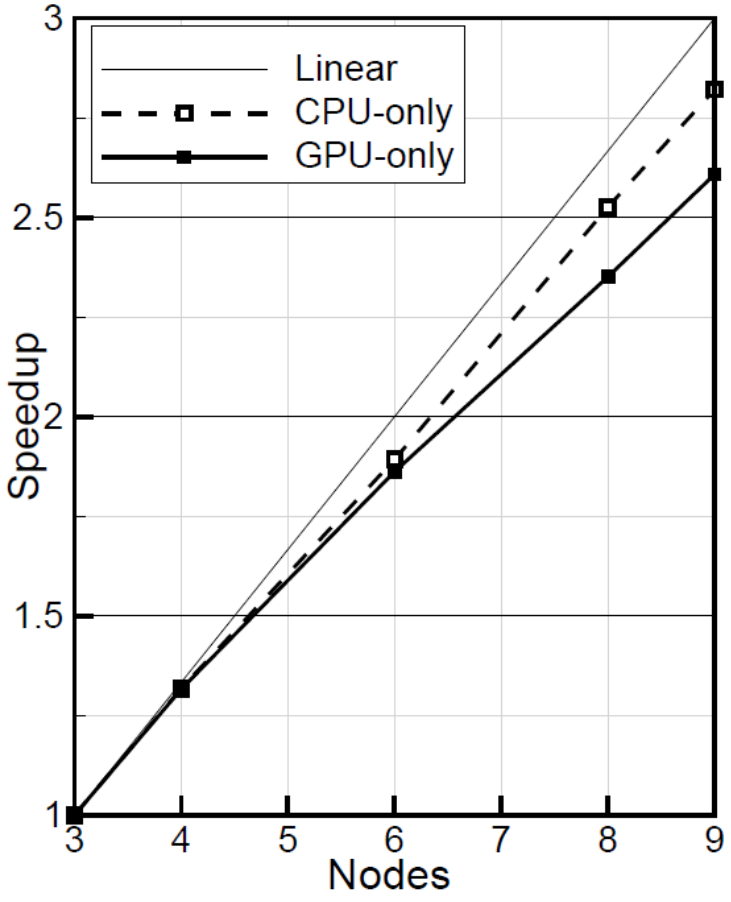
Производительность: только GPU

Численная конфигурация: EBR5, BDF2, IDDES

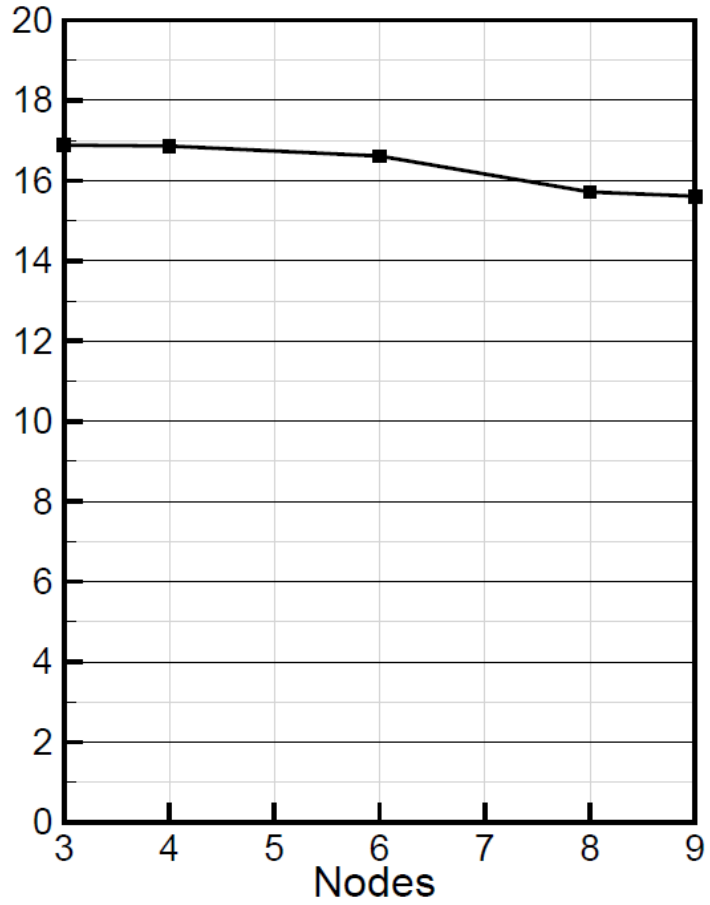
Параллельная конфигурация: MPI+OpenMP+OpenCL, NT=16

Сетка: 80M узлов (обтекание лопатки ТНД)

Параллельное ускорение



Соотношение (4 GPU vs 2 CPU)



Гибридный кластер K60-GPU

Узлы:

2 x16C Intel Xeon **Gold 6142** (120 GB/s)

4 x GPU NVIDIA **V100** (32 GB, 900GB/s)

Сеть: InfiniBand FDR

Соотношение производительности CPU vs GPU

- **1 V100 GPU = 8 16-core Intel Xeon Gold CPU = 140 Gold ядер**
  соотношение пропускной способности памяти 7:1

  1 GeForce RTX 3070 = 1 AMD RX 5700 = 80 Gold ядер

- **36 V100 GPU > 4,500** ядер 16C Intel Xeon Gold CPU
  это если сравнивать с 288 ядрами

- **36 V100 GPU = 10,000** ядер MareNostrum 4 - 24C Intel Xeon Platinum CPU
  прямое сравнение (время 0.27 с на шаг)

Выводы:

GPU с HBM2 и даже игровые с GDDR6
рвут как грелку процессоры с DDR4

Гетерогенный режим CPU + GPU уже (пока еще) не очень имеет смысл

Гетерогенный режим CPU + GPU

Численная конфигурация: EBR5, BDF2, IDDES

Параллельная конфигурация: MPI+OpenMP+OpenCL, NT=14/4

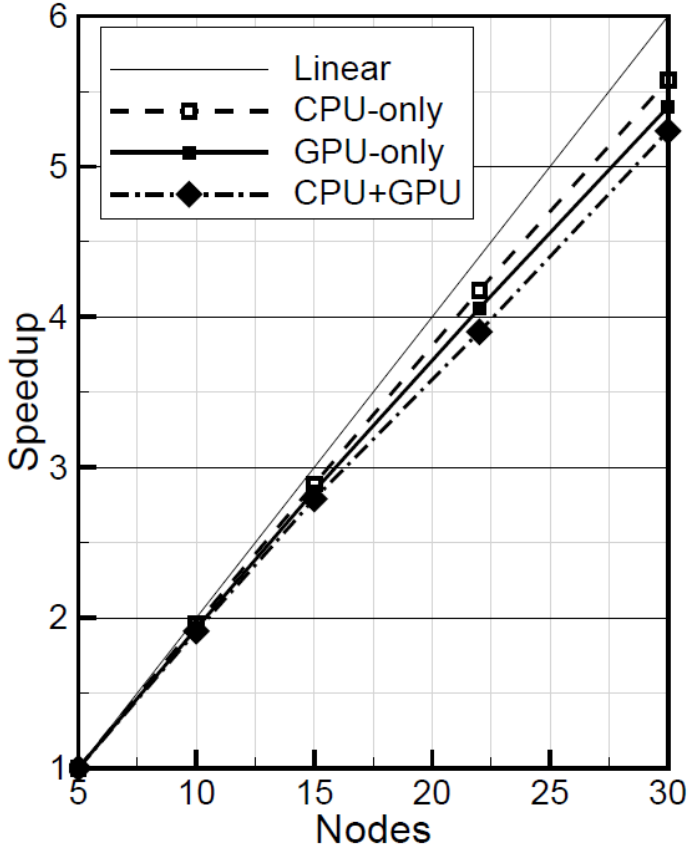
Сетка: 12M узлов (обтекание круглого цилиндра)

Lomonosov-2
14C Intel Xeon E5-2697v3 & NVIDIA K40

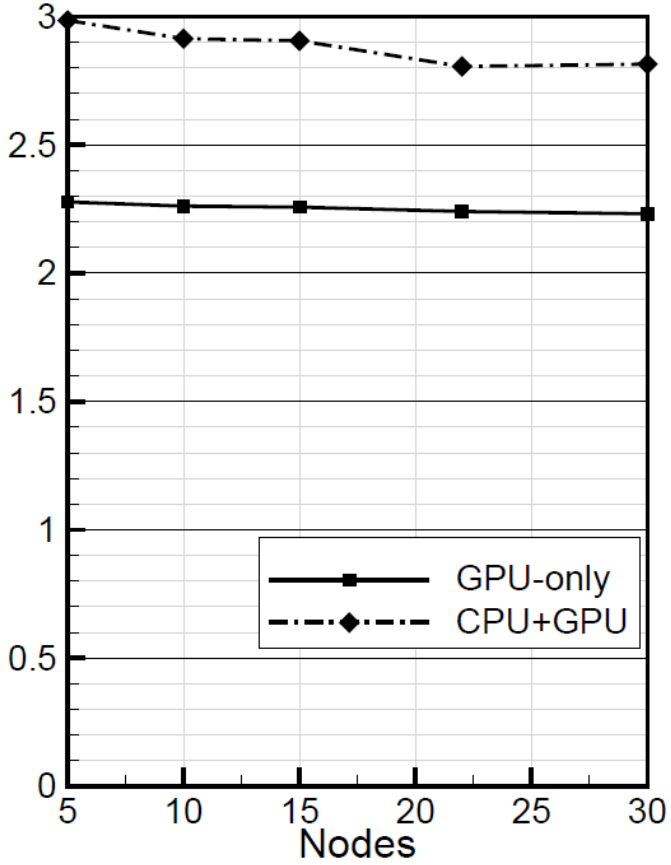
Умеренное соотношение GPU vs CPU **2:1**

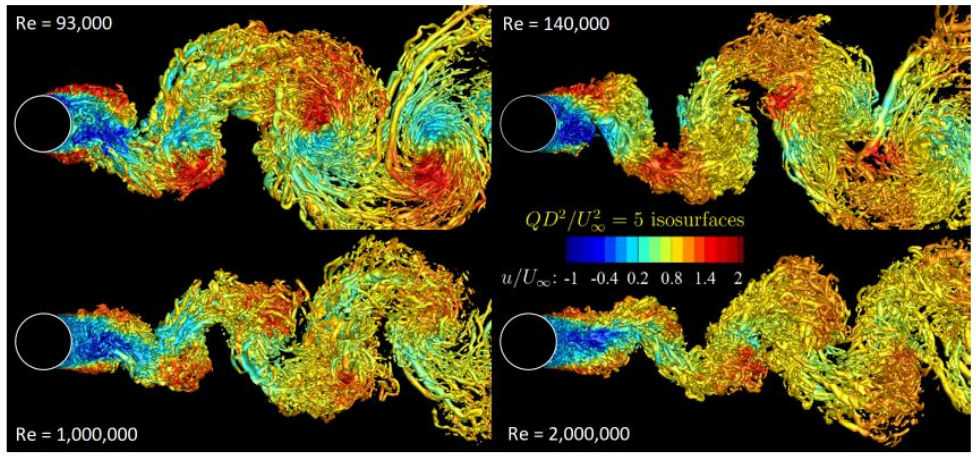
http://hpc.msu.ru/node/159

Параллельное ускорение



Соотношение с CPU





Практическое использование OpenCL версии

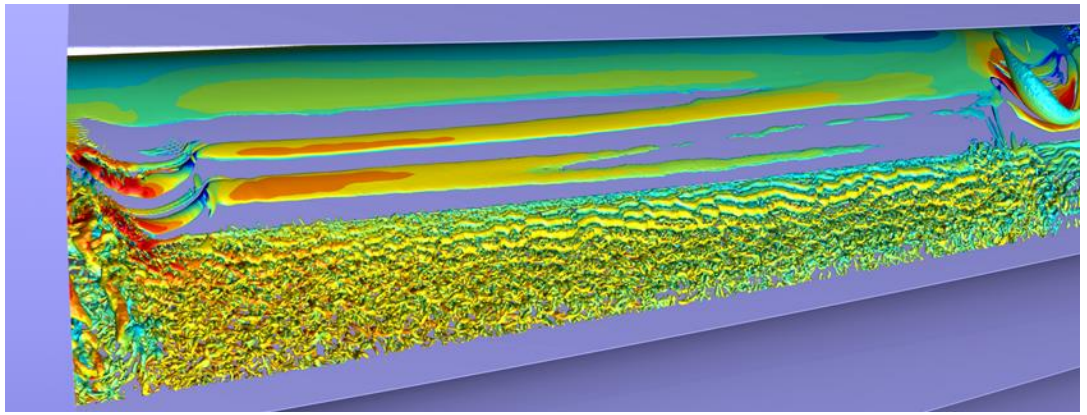
T106C ТНД модельная задачка

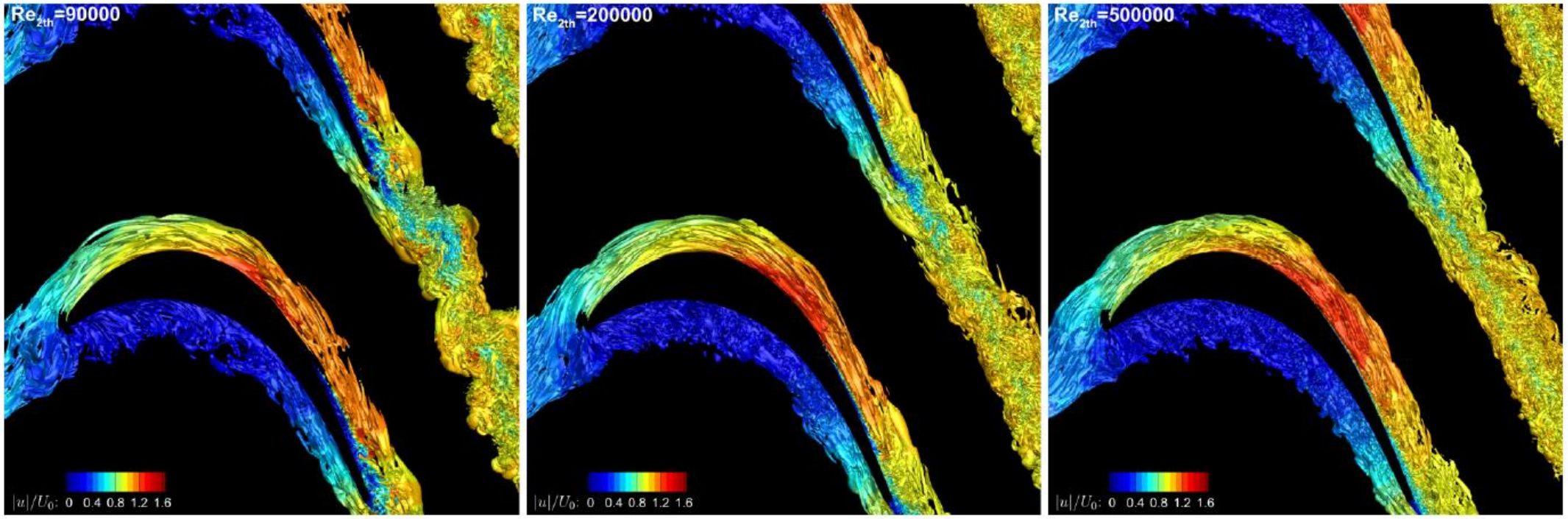
EBR5, BRD2, IDDES, 80M узлов

Главная проблема на GPU – объем памяти.
В 1 GPU с 32 GB впихивается только 6 – 7 M узлов.

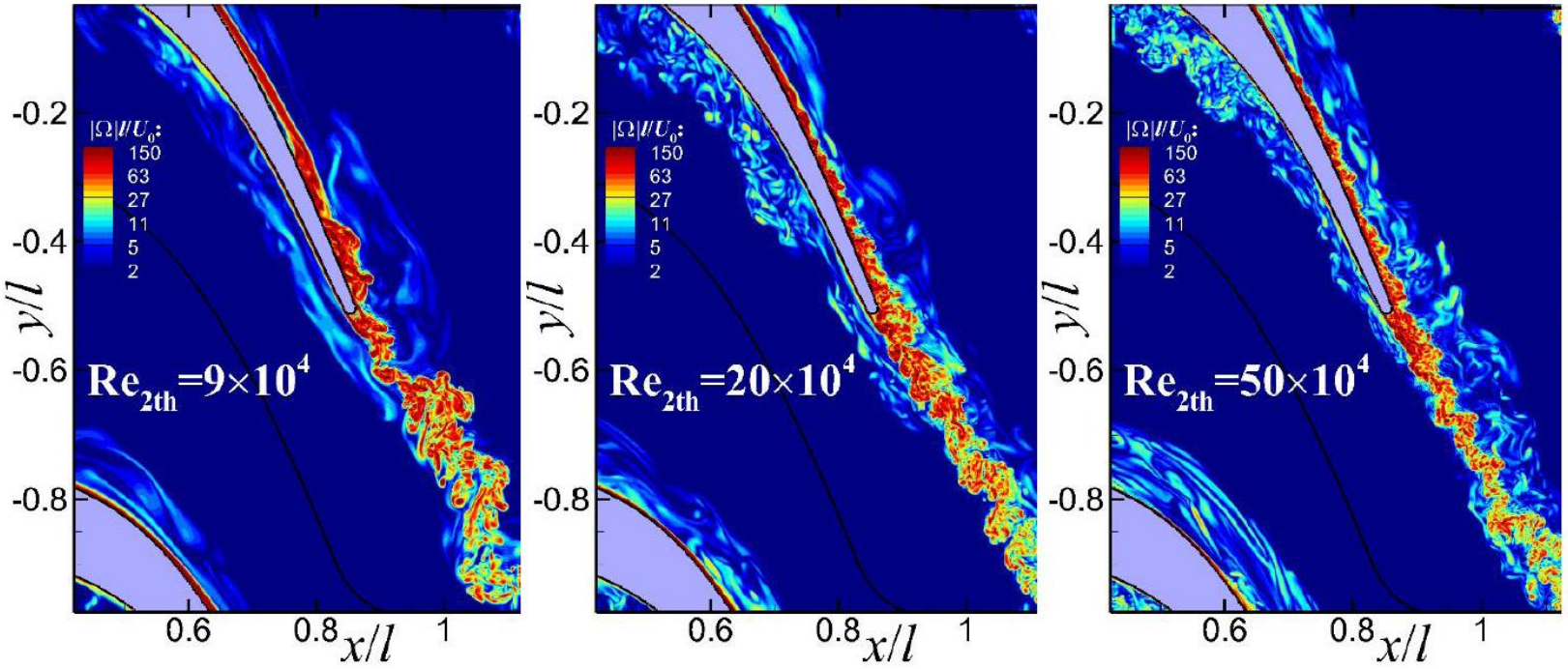
Стоимость примерно 2.7K CPUh на TU

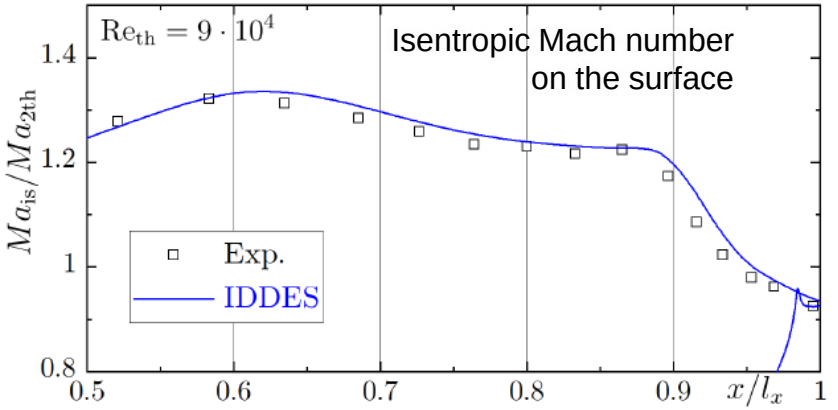
Промышленные расчеты ТНД
Полные лопатки, сетки **150M** узлов

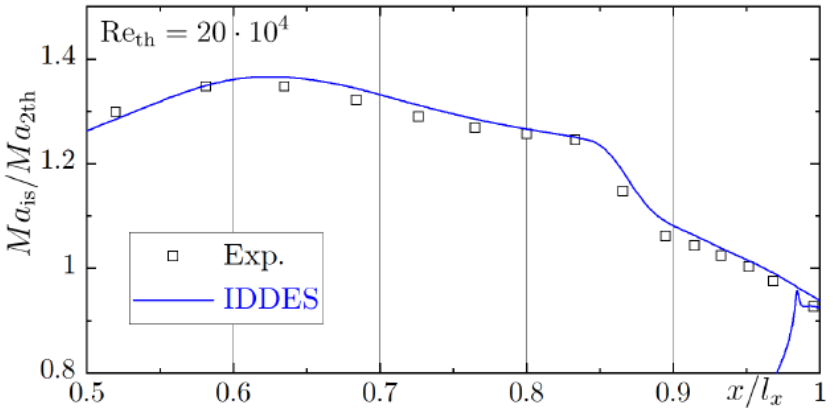


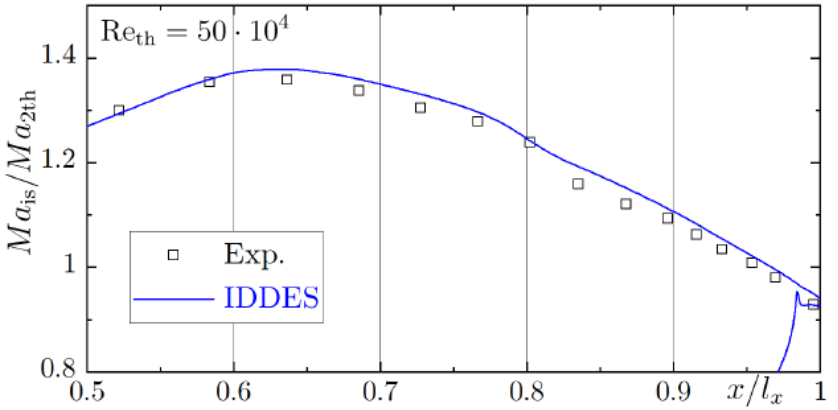


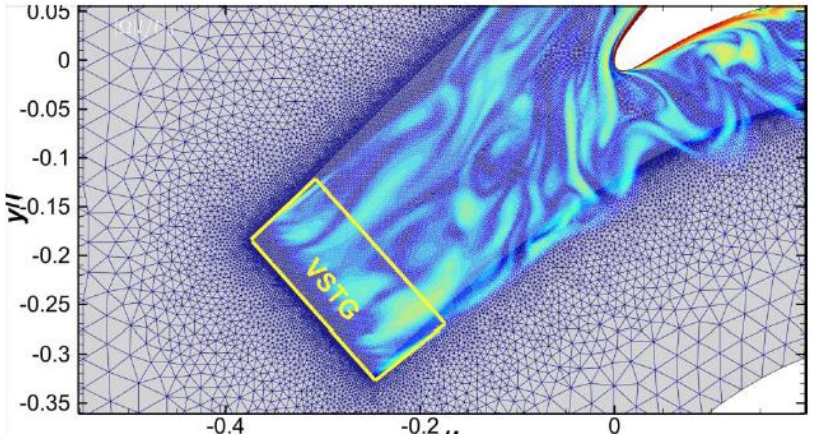
Практическое использование OpenCL версии

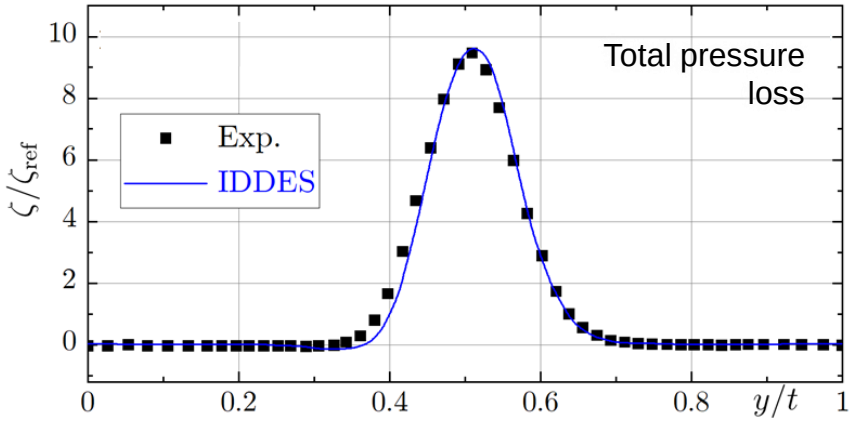


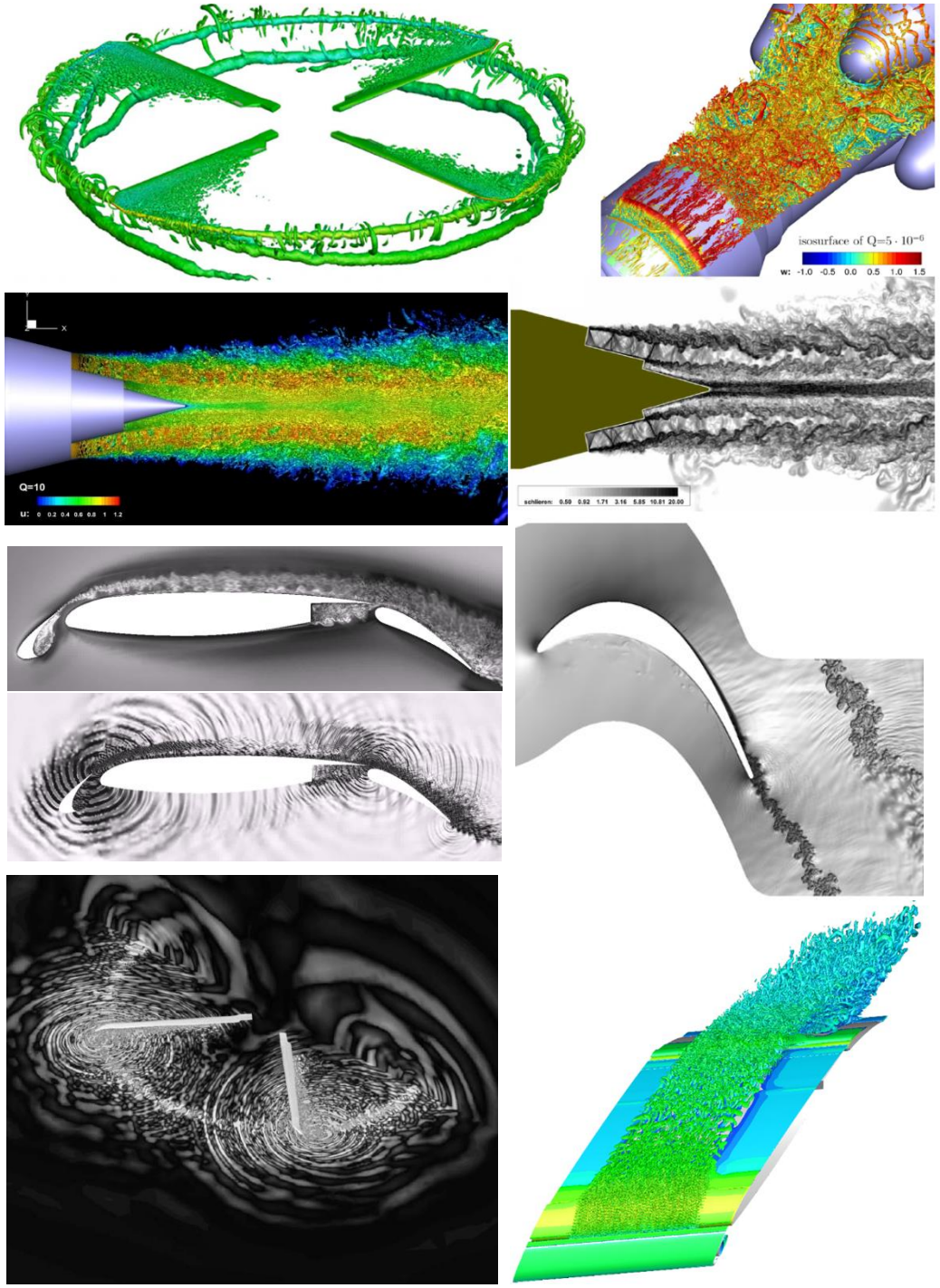












Выводы

- **CPU – слишком медленные**
  на DDR4 вообще никаких шансов, да и на DDR5 тоже

- **GPU – слишком мало памяти**
  если 1 GPU = 150 ядер, то это всего 200MB на ядро

- **OpenCL версия:**
  код сложнее модифицировать и поддерживать
  но эти проблемы можно заметно мягчить
  ускорение против CPU того стоит

- **Гетерогенный режим CPU+GPU 1 процесс – 1 девайс**
  сильно проще, чем окучивать гибридный узел из-под OpenMP

  Пока в co-execution нет смысла, GPU слишком быстрые
  Но мы ждем, что CPU нанесут ответный удар и восстановят паритет с GPU

Спасибо за внимание!

Внеклассное чтение

Материалы к учебному курсу магистратуры ВМК МГУ

http://caa.imamod.ru/vmkvm

Базовое распараллеливание NOISETTE для CPU

A. Gorobets. Parallel Algorithm of the NOISEtte Code for CFD and CAA Simulations. Lobachevskii Journal of Mathematics. Vol. 39, No. 4, 2018, pp. 524–532.

Подготовка NOISETTE к гетерогенным вычислениям

A. Gorobets, P. Bakhvalov, A. Duben, P. Rodionov. Acceleration of NOISEtte Code for Scale-resolving Supercomputer Simulations of Turbulent Flows. Lobachevskii Journal of Mathematics. Vol. 41, No. 8, 2020, pp. 1463–1474.

A. Gorobets, P. Bakhvalov. Improving Reliability of Supercomputer CFD Codes on Unstructured Meshes. Supercomputing Frontiers and Innovations. Vol. 6, No. 4, 2019, pp. 44–56.

Предыдущие гетерогенные реализации

A.Gorobets, S.Soukov, P.Bogdanov. Multilevel parallelization for simulating turbulent flows on most kinds of hybrid supercomputers. Computers and Fluids. Vol. 173, 2018, pp. 171–177.

S. A. Soukov, A. V. Gorobets. Heterogeneous Computing in Resource-Intensive CFD Simulations. Doklady Mathematics. Vol. 98, No. 2, 2018, pp. 1–3.

EBR схемы

I. Abalakin, P. Bakhvalov, T. Kozubskaya, Edge-based reconstruction schemes for unstructured tetrahedral meshes. International Journal for Numerical Methods in Fluids, Vol. 81, No. 6, 2016, pp. 331–356.

P. Bakhvalov, T. Kozubskaya, EBR-WENO scheme for solving gas dynamics problems with discontinuities on unstructured meshes. Computers and Fluids, Vol. 157, 2017, pp. 312–324.