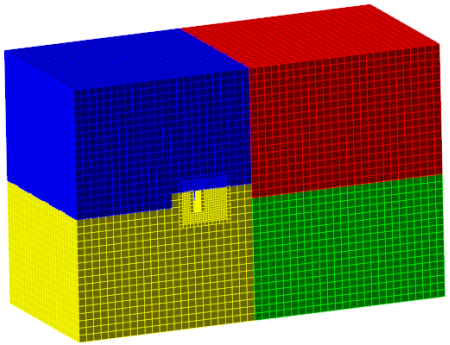
Тестирование производительности ПК FlowVision на вычислительном кластере с высокоскоростной коммуникационной сетью Альфа



Фишер Ю.В., *начальник группы проектного отдела*

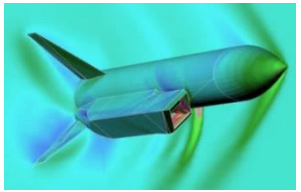
Дуплянкин Р.А., *инженер 1 кат.*

Москва, 2026

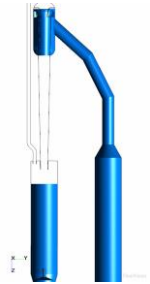
FlowVision – это инструмент



- решения промышленных, научных и учебных задач в условиях дефицита времени без потери качества получаемого решения
- моделирования сложных многодисциплинарных процессов удобными простыми методиками
- решения на сложной геометрической модели, в том числе сборках
- расчетов на современной вычислительной технике, начиная от ноутбука и заканчивая суперкомпьютером
- многодисциплинарных рабочих цепочек взаимодействия и оптимизации с другими CAE-пакетами
- автоматизации собственной работы для решения типовых задач



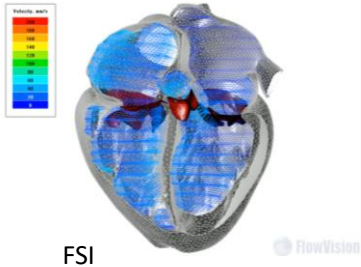
Внешнее обтекание



Многофазные среды



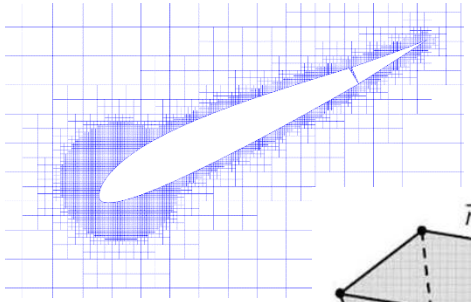
Движение спортсмена

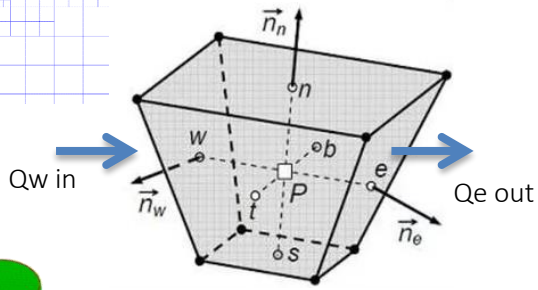


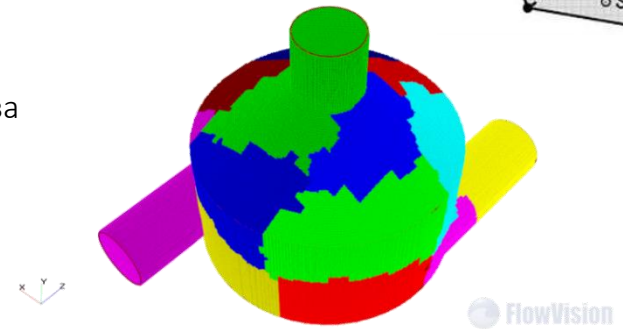
FSI

Производительность CFD

- Пространство делится на ячейки
- В ячейках через потоки на гранях вычисляются значения в центрах
- В итоге необходимо решить СЛАУ (большие матрицы)
- Для распараллеливания расчетная сетка делится по процессорам (происходит декомпозиция вычислительной задачи)
- Масштабируемость задачи при увеличении количества решателей ухудшается из-за обменных операций

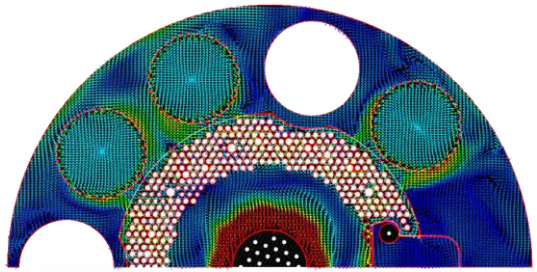


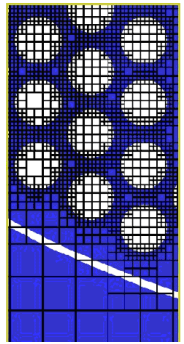


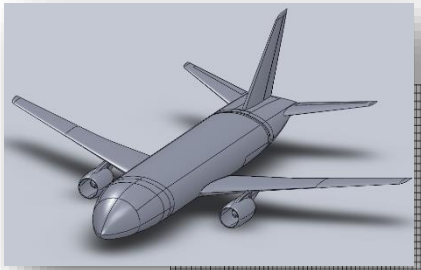


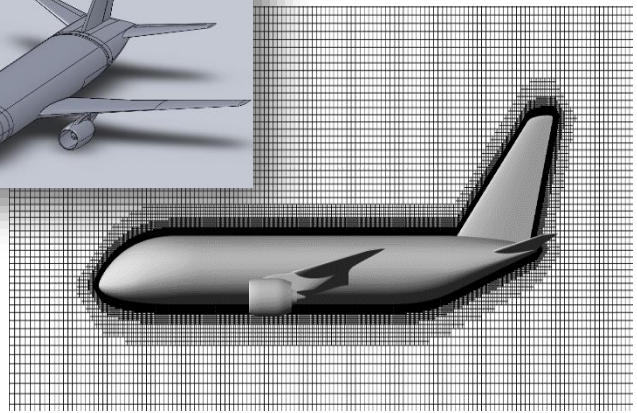
Зачем исследовать производительность?

- Определять необходимый ресурс для решения конкретной задачи
- Знать, как посчитать быстрее и дешевле
- Понять, как эффективно строить расчетную сетку
- Прогнозировать длительность расчетов на проекте
- Для разработчиков (улучшение алгоритмов)





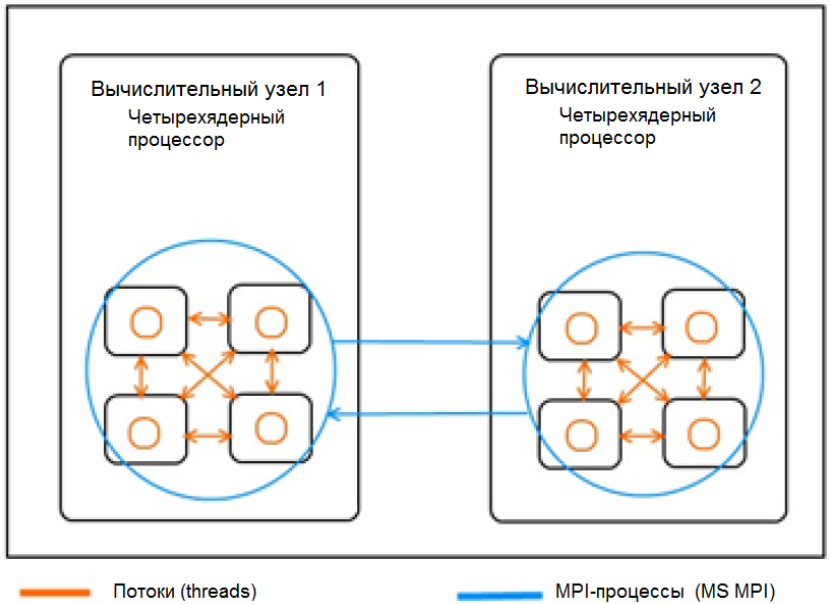




FlowVision

Комбинированная организация параллельных вычислений

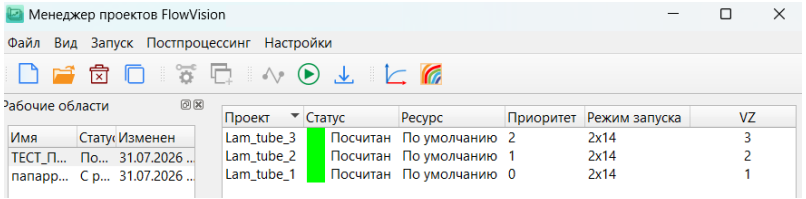
- С помощью библиотеки Intel TBB: распараллеливание внутри процессора с общим доступом к памяти по потокам (нитям, threads)
- С помощью библиотеки MPI: для распараллеливания между процессорами с раздельной памятью (с неравномерным доступом к памяти



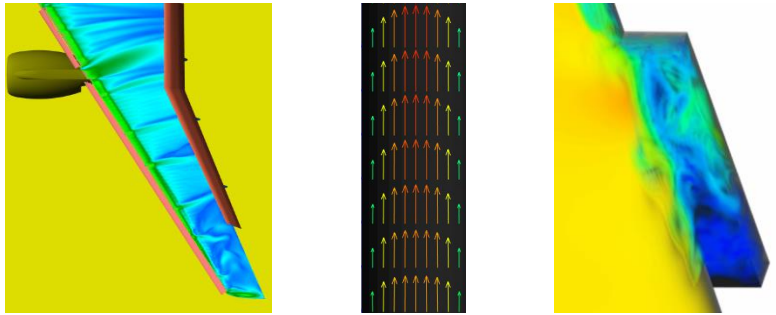
НИР по тестированию производительности

Этапы:

1. **Автоматизация методики** и тестирование производительности и масштабируемости ПАК с высокоскоростной коммуникационной сетью «Альфа» и ПК FlowVision



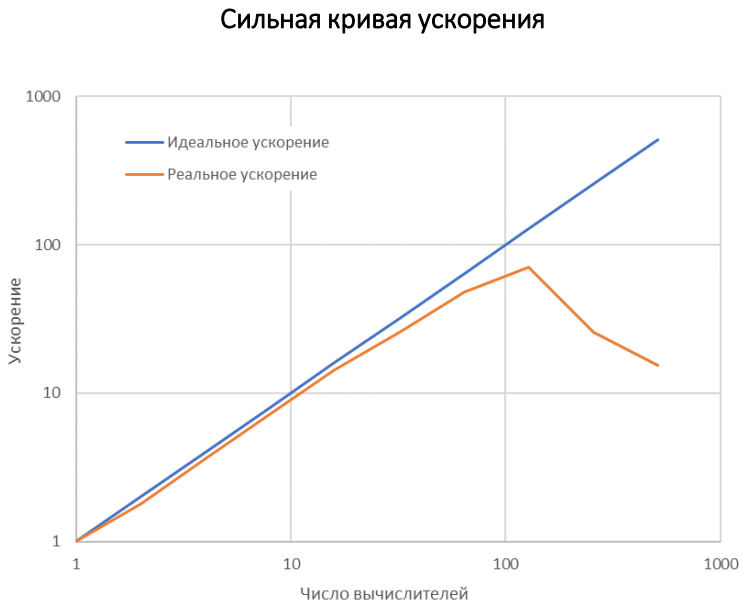
2. Исследования производительности на наборе **различных задач**:
   - Сравнение с другими кластерами[2]



1. M. J. C. Henshaw, “M219 Cavity Case: Verification and Validation Data for Computational Unsteady Aerodynamics,” Tech. Rep. RTO-TR-26, AC/323(AVT)TP/19, QinetiQ, UK, 453–472 (2002).
2. Акимов В.С., Силаев Д.П., Симонов А.С., Семенов А.С. Исследование масштабируемости FlowVision на кластере с интерконнектом Ангара //Вычислительные методы и программирование: новые вычислительные технологии. 2017. Т. 18. № 4. С. 406-415.

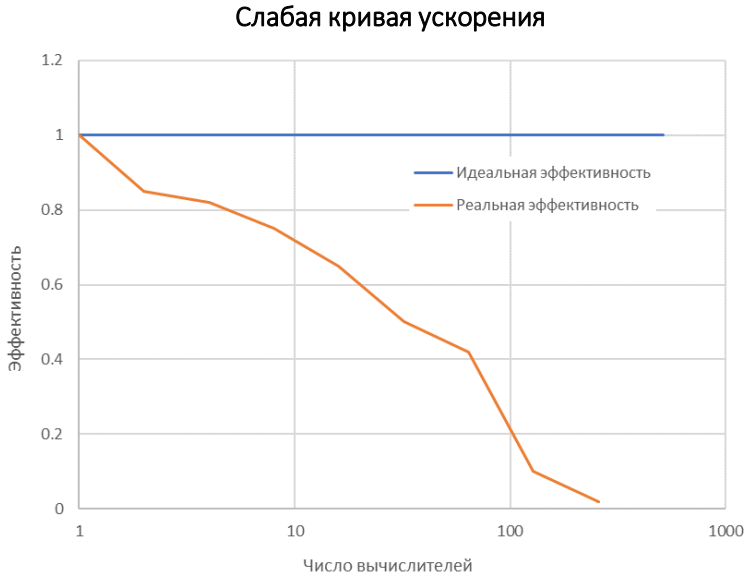
Сильная кривая ускорения

- Ускорение – отношение времени счета задачи на одном вычислителе ко времени счета на N вычислителях для задачи **фиксированного размера** сетки. В качестве N рассматривают ядра процессора или узлы кластера
- Зависимость ускорения от числа вычислителей – **сильная кривая ускорения**
- Кривая отклоняется от линейной, когда приходится мало ячеек на вычислитель



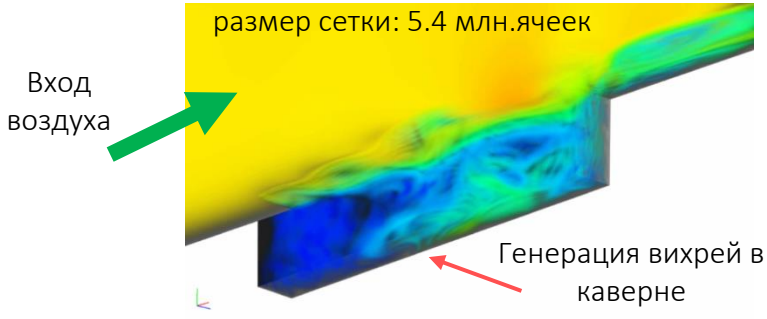
Слабая кривая ускорения

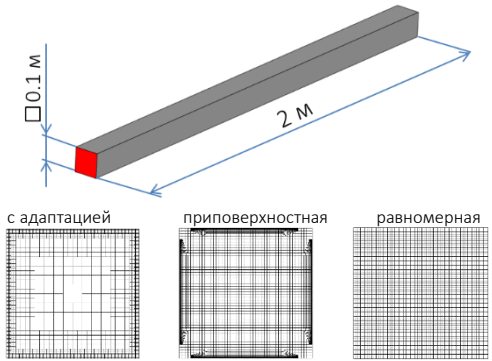
- **Эффективность** - отношение времени счета задачи на одном вычислителе ко времени счета на N вычислителях при **фиксированном количестве ячеек на ядро**
- Зависимость эффективности от числа вычислителей – **слабая кривая ускорения**
- Эффективность снижается, когда начинают доминировать затраты на обмены и обращения в память (много вычислителей)

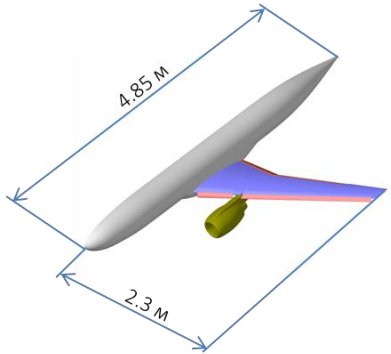


Задачи для тестирования

- Сверхзвуковое течение в каверне M219
- Течение в квадратной трубе
- Дозвуковое обтекание самолета HLPW
- Влияние настроек:
  - MPIxThreads
  - Приповерхностная сетка
  - Скорость интерконнекта
  - Hyper-threading
  - Numa affinity…







Режимы запуска

M219, 18 сеток: от 1.2 до 120 млн. ячеек

№	MPI на узел	MPI на проц	threads на MPI
1	48	24	1
2	24	12	2
3	16	8	3
4	12	6	4
5	8	4	6
6	6	3	8
7	4	2	12
8	2	1	24

Всего 760 запусков

	Число узлов	1	2	3	4	5
№ сетки	M, млн.яч	n, тыс.яч/ядро				
1	1.2	25.0	12.5	8.3	6.3	5.0
2	4	83.3	41.7	27.8	20.8	16.7
3	8	166.7	83.3	55.6	41.7	33.3
4	12	250.0	125.0	83.3	62.5	50.0
5	24	500.0	250.0	166.7	125.0	100.0
6	2	41.7	-	-	-	-
7	6	125	-	41.7	-	-
8	10	-	-	-	-	41.7
9	16	-	-	-	83.3	-
10	18	-	-	125	-	-
11	20	-	-	-	-	83.3
12	30	-	-	-	-	125
13	36	-	-	250	-	-
14	48	-	500	-	250	-
15	60	-	-	-	-	250
16	72	-	-	500	-	-
17	96	-	-	-	500	-
18	120	-	-	-	-	500

Мини-кластер «Скворец»

Характеристики

Управляющий узел

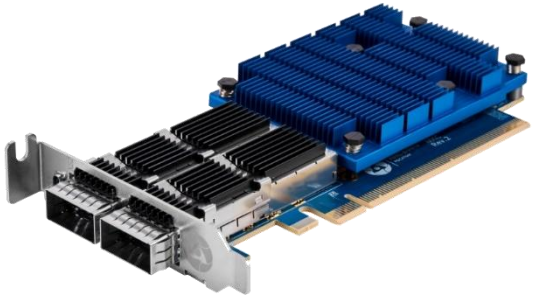
5 вычислительных узлов, каждый:

- Сервер KRAFTWAY TS3100 1U
- 2 процессора Intel Xeon Gold 6336Y 2.4 ГГц, всего 48 ядер (96 в режиме Hyper-Threading)
- Оперативная память: 512 ГБ
- 2-портовый сетевой адаптер Альфа



Сеть Альфа

- Российский аналог Infiniband
- 32-портовый коммутатор
- порт: 80 Гбит/с
- Операционная система: Alt Linux (ALT SP Server 10)
- Компилятор C/C++: gcc 10.3.1
- Библиотека MPI: MPICH 4.3.0, поддержка OpenMP
- Планировщик задач: Slurm 25.05.0



Характеристики исследуемых суперкомпьютеров

Суперкомпьютер	«Скворец»	«Ангара-К1» (раздел А)	«Ломоносов-2»
Процессор	Intel Xeon Gold 6336Y, 2.4 GHz	Intel Xeon E5- 2630, 2.30 GHz	Intel Xeon E5-2697v3, 2.6 GHz
Количество физических ядер процессора	24	6	14
Количество логических ядер при использовании Hyper-Threading (HT)	48	12	28
Количество процессоров на узле	2	2	1
Максимальная пропускная способность памяти, ГБ/с	?	42.6	68
Количество оперативной памяти на узел, ГБ	512	64	64
Топология, интерконнект	Альфа (80 Гбит/с)	Ангара, 3D-тор 4x3x3	Mellanox FDR InfiniBand (56 Гбит/с)
Реализация MPI	MPICH 4.3.0 (2 версии)	MPICH 3.0.4	OpenMPI 1.8.4
Операционная система	ALT Linux		

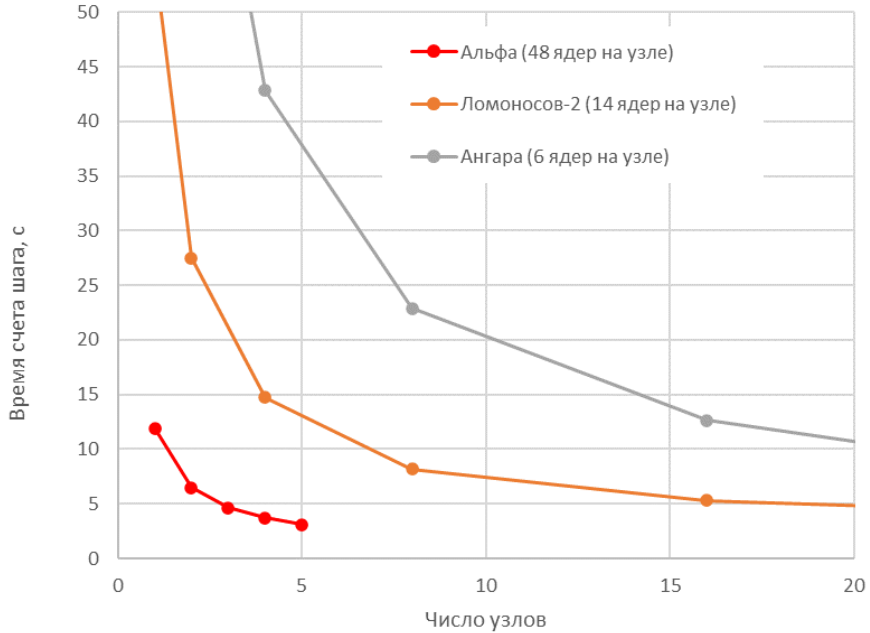
Исследуется в настоящей работе

Результаты расчетов: M219, сетка 5.44 млн.яч

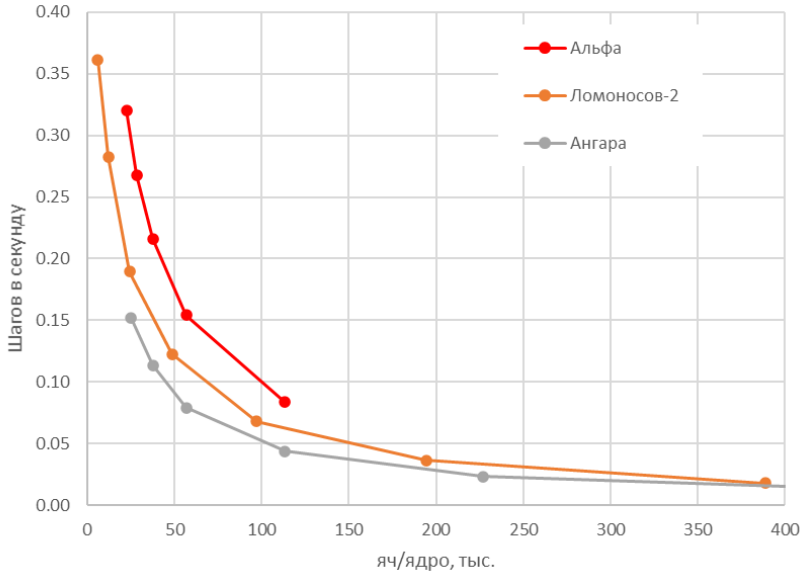
- Показан лучший режим на «Скворце»:

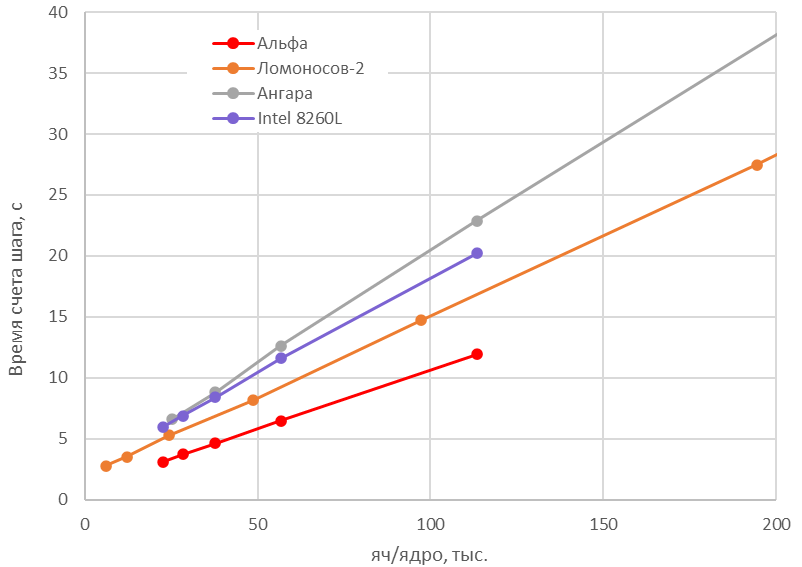
12 MPI на процессор x 2 потока

- Среднее время вычисления шага на современном кластере «Скворец» заметно ниже при том же количестве узлов



Результаты расчетов: M219, сетка 5.44 млн.яч





- При одинаковых количествах ячеек на ядро видно опережение «Скворца» относительно других кластеров

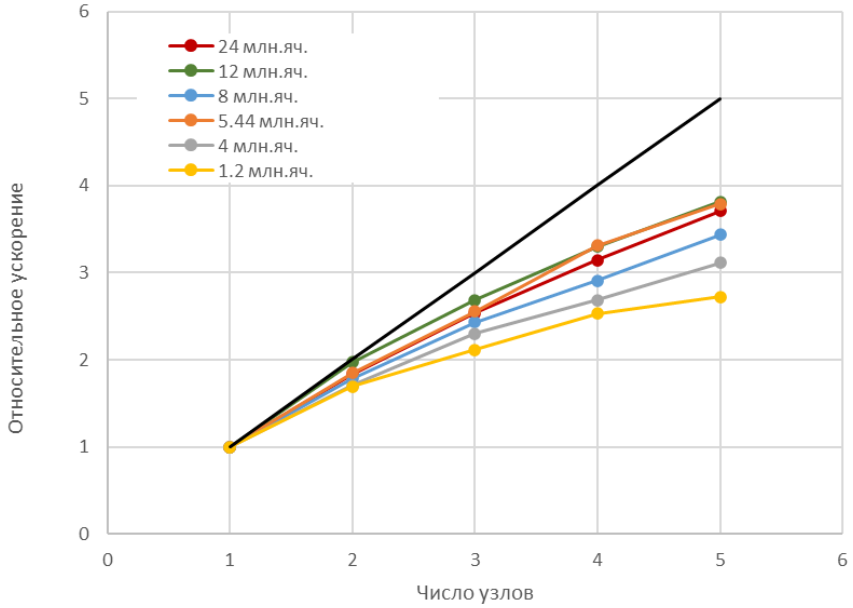
Сильные кривые ускорения

- Показан режим на «Скворце»:

12 MPI на процессор x 2 потока

- На сетках с меньшим количеством ячеек на ядро масштабируемость задачи хуже, что связано с увеличением влияния непараллельной части алгоритма и обменов между узлами

- На сетках, где число ячеек на ядро высокое, масштабируемость улучшается



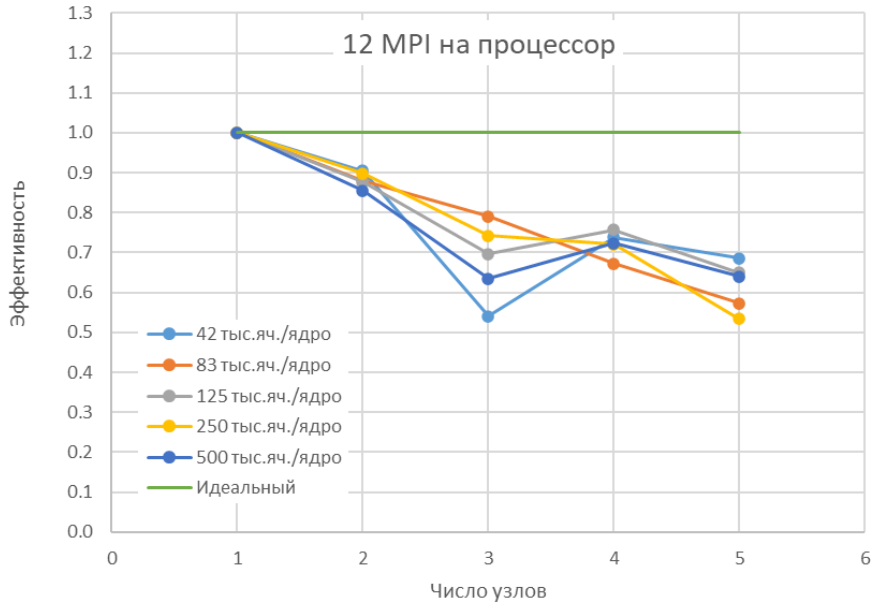
Слабые кривые ускорения

- Показан режим на «Скворце»:

12 MPI на процессор x 2 потока

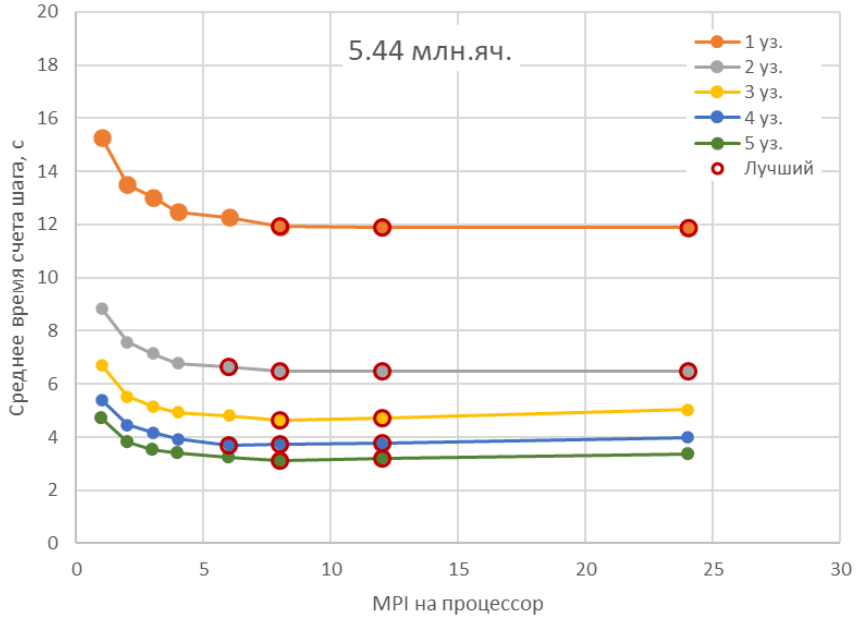
- При чётном количестве узлов эффективность распараллеливания выше, чем при нечётном

- Эффективность распараллеливания снижается на 5-10% с каждым добавленным узлом



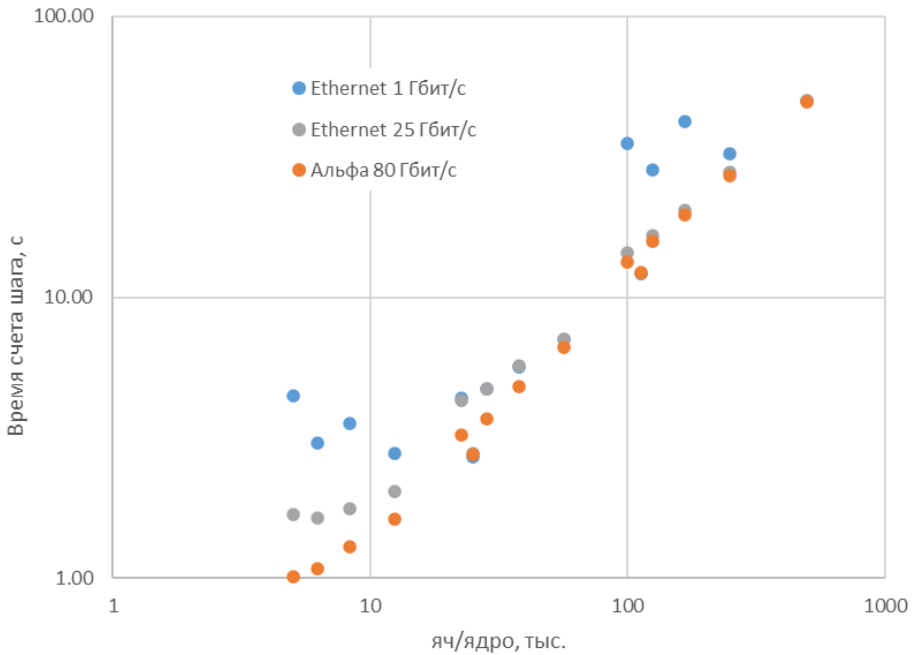
Влияние режима запуска: M219

- Для всех рассмотренных запусков минимальное время счета шага достигается в режимах **4-12 MPI (2-6 потока) на процессор**



Влияние используемого соединения узлов: M219

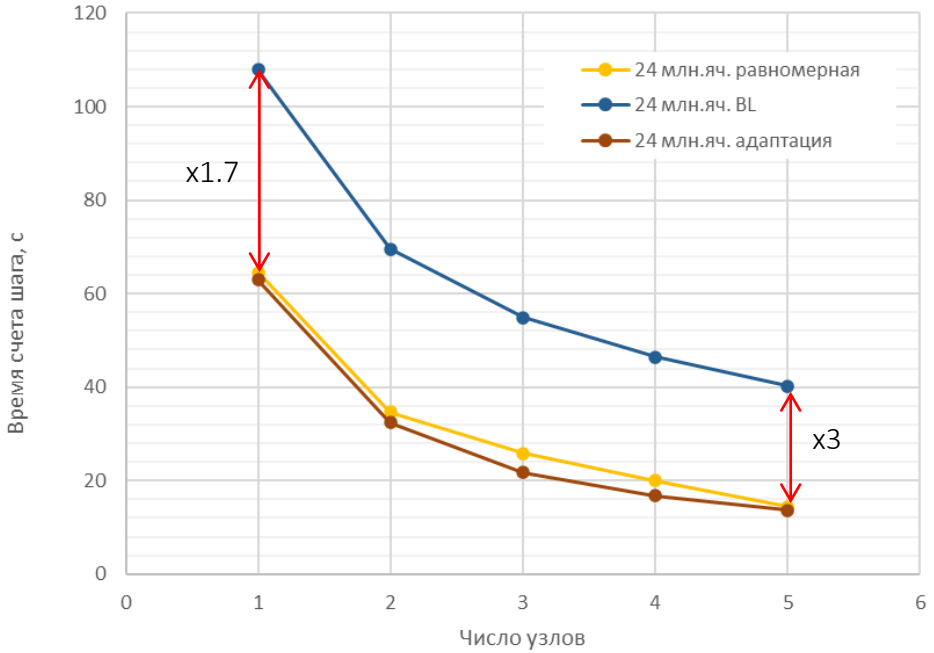
- Запуски задач 1.2 млн.яч. и 24 млн.яч
- На сетках с маленьким количеством ячеек на ядро (5÷25 тыс.яч/ядро) наблюдается значительный рост времени счета при уменьшении скорости связи межу узлами
- На сетках с большим количеством ячеек на ядро (100÷500 тыс.яч/ядро), эффект скорости связи наблюдается только при скорости 1 Гбит/с



Время счета шага на различных видах сеток: труба

- Скорость счета задачи с использованием приповерхностной сетки повышается в 1.5-3 раза по сравнению с другими рассмотренными вариантами

- Скорость счета задачи с использованием равномерной сетки и адаптаций для данной задачи не отличаются



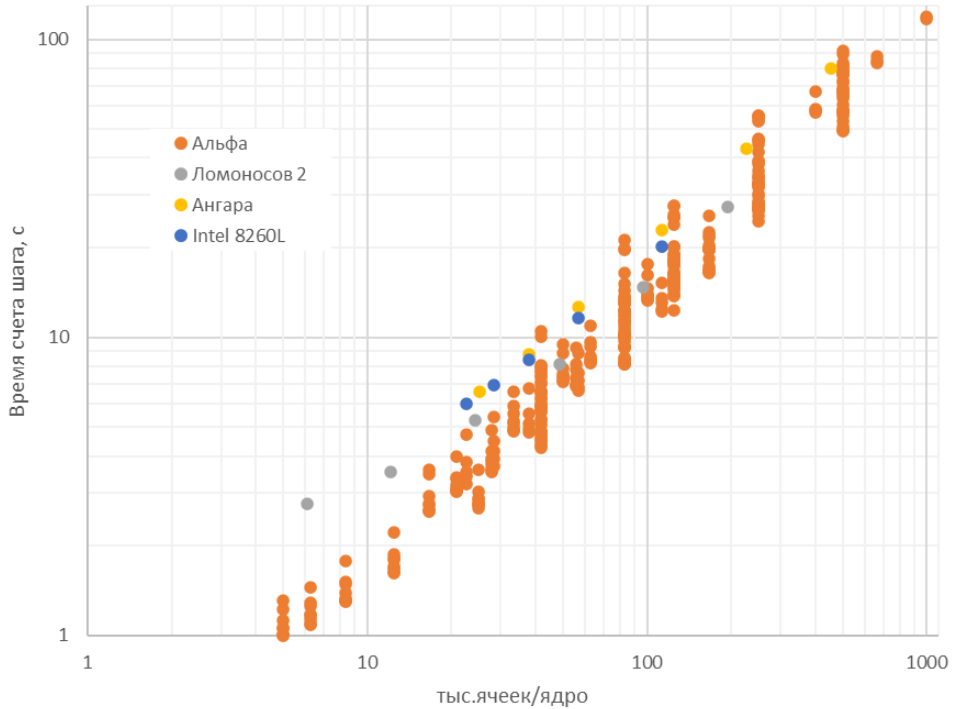
Абсолютные времена счёта по всем расчетным точкам

Не влияет:

- адаптация
- сгущение сетки
- сложность геометрии

Влияет:

- количество ячеек на ядро
- режим запуска



Выводы

Исследованы:

- Различные по физике и по опциям задачи
- Расчетные сетки от 1.2 до 120 млн. ячеек (от 5 тыс. до 1 млн. ячеек на ядро)

Получены рекомендации:

- MPI x threads → лучшие результаты при 4-12 MPI на процессор, худшие результаты 1 MPI на процессор
- Hyper-threading → добавок к ускорению <5%
- приповерхностная сетки приводит к замедлению в 1.5-3 раза → данная опция FlowVision на доработке

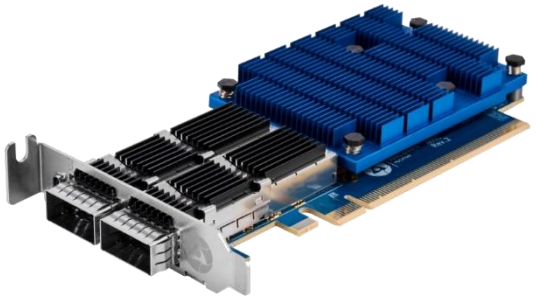
Интерконнект:

- «Альфа» – при 1.2 млн.ячеек ускорение <2.7, при 24 млн.ячеек ускорение <3.7

Кривые ускорения:

- Сильная, 5 узлов→ ускорение от 2.5 до 4.3
- Слабая → эффективность снижается на 5-10% за каждый узел





Получен набор расчетных точек для создания прогнозов времени счета произвольной задачи на кластере