

Оптимизация расчетов биохимических процессов во внутренних водоемах суши на графических ускорителях

Гащук Е.М.^{1,2,3}, Ахтамьянов Р.А.^{3,5,6}, Ломов В.А.^{3,4,5}, Дебольский А.В.^{3,4},

Гладских Д.С.^{3,7}, Мортиков Е.В.^{3,2}

Мехмат МГУ им. М.В. Ломоносова, Москва¹

Институт вычислительной математики им. Г.И. Марчука РАН, Москва²

НИВЦ МГУ им. М.В. Ломоносова, Москва³

Институт физики атмосферы им. А.М. Обухова РАН, Москва⁴

Геофак МГУ им. М.В. Ломоносова, Москва⁵

Гидрометцентр РФ, Москва⁶

Институт прикладной физики им. А.В. Гапонова-Грехова РАН, Нижний Новгород⁷



Актуальность

- **Модели озер и водохранилищ** позволяют оценить ряд климатически значимых характеристик, включая **эмиссию парниковых газов**.
- **Использование трехмерных моделей является единственным способом получить полное распределение** термических и биохимических характеристик.
- **Прогноз погоды и климата** является вычислительно **сложной** задачей (нельзя использовать трехмерные модели озер в МЗС). Трехмерные модели могут использоваться как инструмент для калибровки и уточнения моделей меньшей пространственной детализации, используемых в МЗС.
- **Важной частью** современных моделей озер и водохранилищ является как можно более **детализированное описание переноса примесей**. Для численного решения задач переноса примесей часто требуется использовать схемы **высокого порядка** точности.

Необходимо разрабатывать вычислительно эффективные алгоритмы, учитывающие архитектуру суперкомпьютеров.

Разработка гидростатической модели внутренних водоемов суши и блока биохимии на GPU

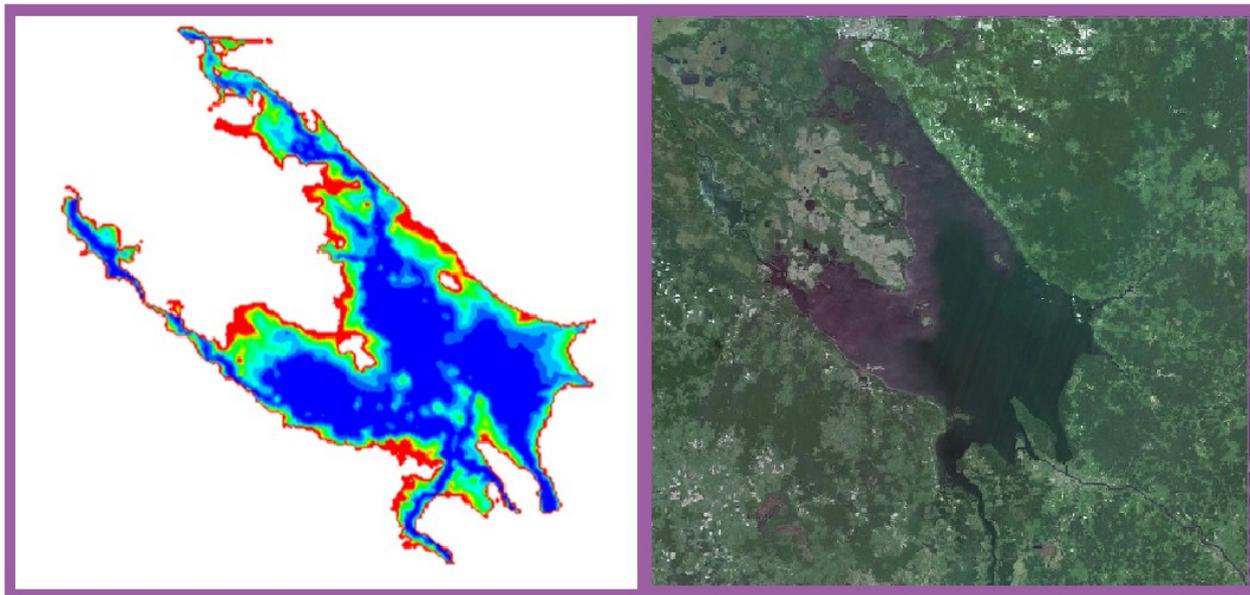


Рис. Мгновенное распределение потока метана (в моль/(м²·с)) по поверхности Рыбинского водохранилища (слева) и снимок Рыбинского водохранилища со спутника (справа).

Разработка гидростатической модели внутренних водоемов суши и блока биохимии на GPU

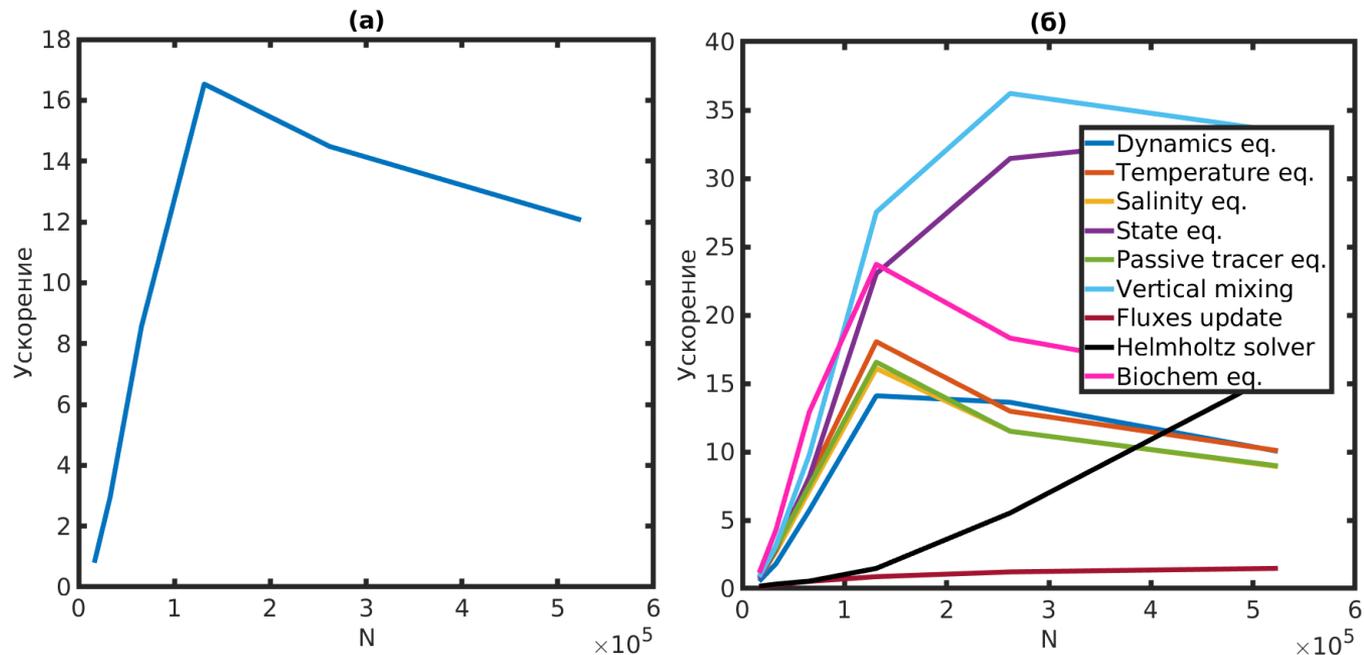


Рис. Ускорение расчетов на GPU реализации относительно CPU для всей модели (а)
и
отдельных блоков модели (б).

Разработка гидростатической модели внутренних водоемов суши и блока биохимии на GPU

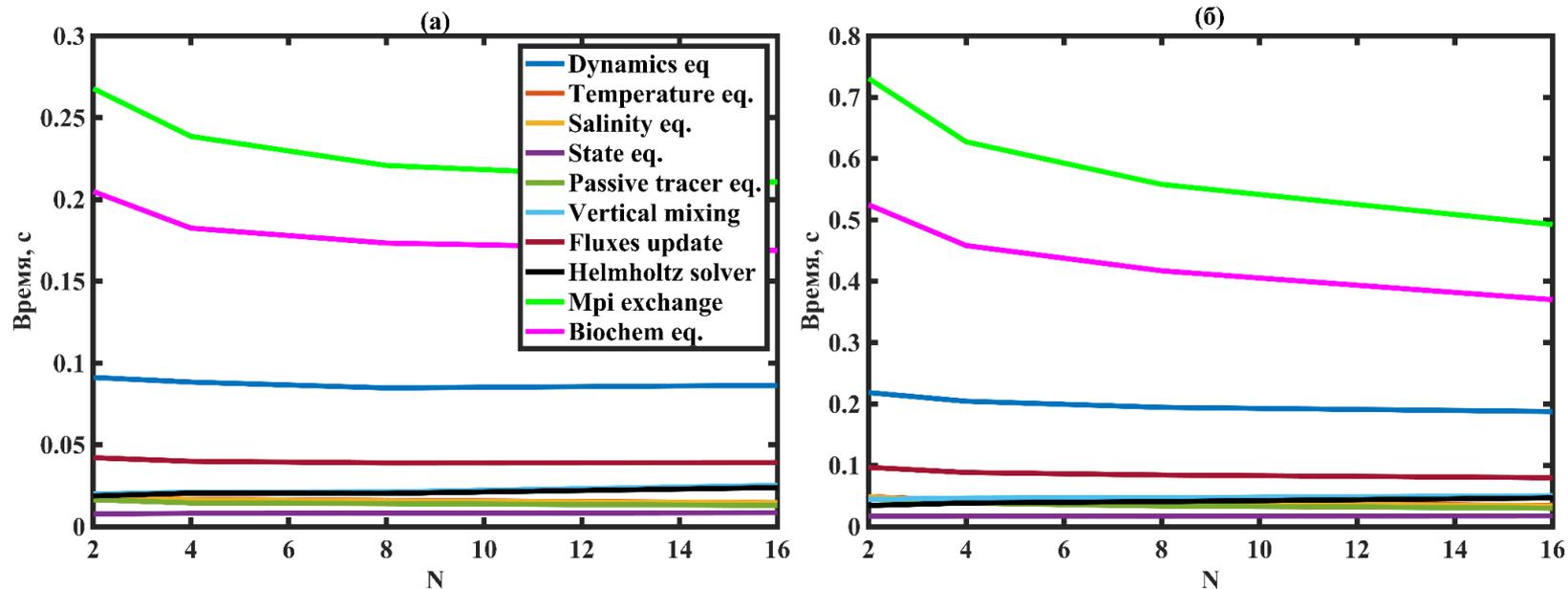


Рис. Время расчетов отдельных компонент модели на GPU на сетках размера $8 \times 32 \times 64$ (а) и $16 \times 64 \times 64$ (б) в зависимости от количества MPI-процессов (N).

Оптимизация алгоритмов схем адвективного переноса

Уравнение переноса пассивной примеси:

$$\frac{\partial C_k}{\partial t} + \frac{\partial u_i C_k}{\partial x_i} = 0$$

Схема	Условное обозначение
Схема с разностями против потока первого порядка точности	UW
Центрально-разностная схема второго порядка точности в дивергентной форме	DIV
Центрально-разностная схема второго порядка точности в кососимметричной форме	SKEW
Схема с разностями против потока третьего порядка точности	UW3
Схема Лакса-Вендроффа с ограничителем потока	LAX
Алгоритм Flux-Corrected Transport	FCT



Рис. Условные обозначения схем адвекции.

Оптимизация алгоритмов схем адвективного переноса

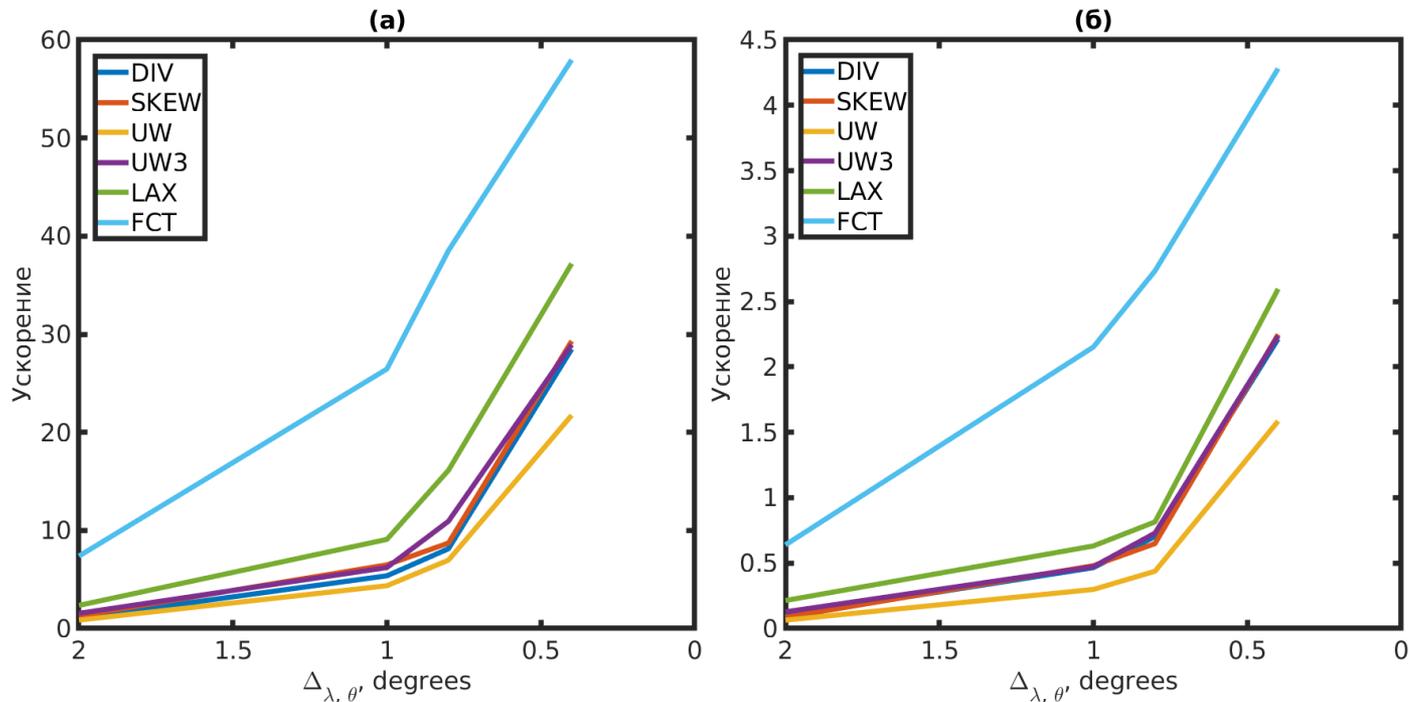


Рис. Результаты эффективности выполнения расчетов переноса примеси в двумерной модели: (а) ускорение на одном A100 по сравнению с одним ядром CPU и (б) ускорение на одном A100 по сравнению с одним узлом CPU.

Оптимизация алгоритмов схем адвективного переноса

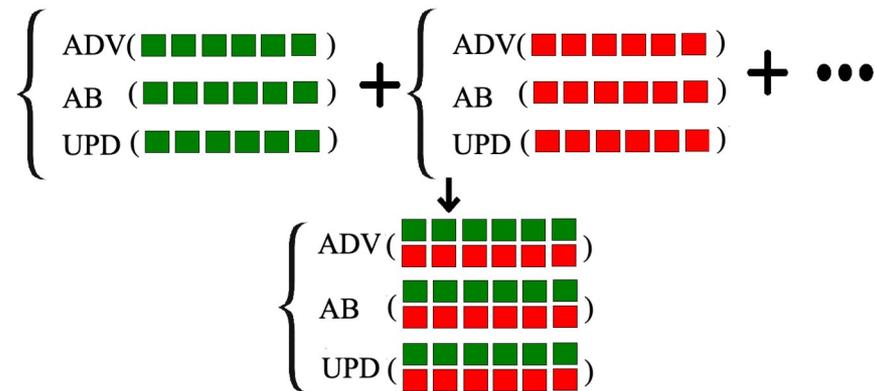


Рис. Базовая реализация численного решения задачи адвекции, различные цвета соответствуют различным концентрациям C_k

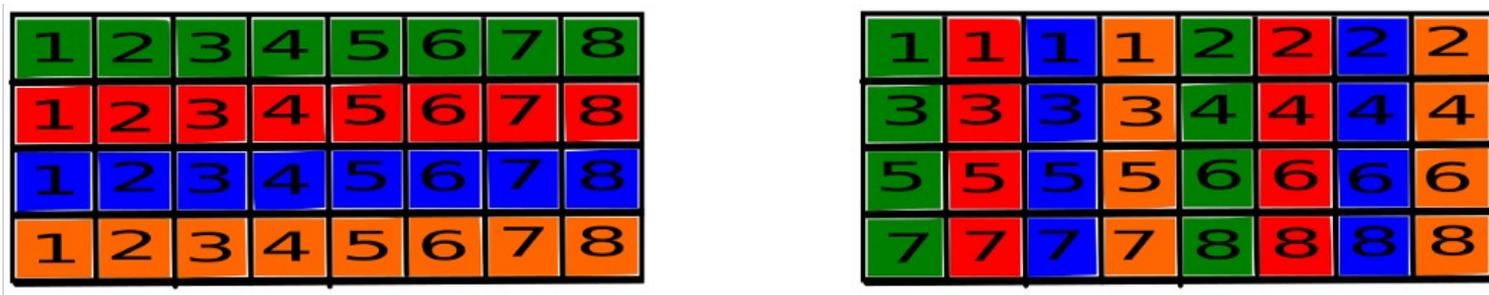


Рис. Размещение памяти примесей: прямой порядок (слева) и с переупорядочиванием (справа), число – индекс вычислительной сетки.

Оптимизация алгоритмов схем адвективного переноса

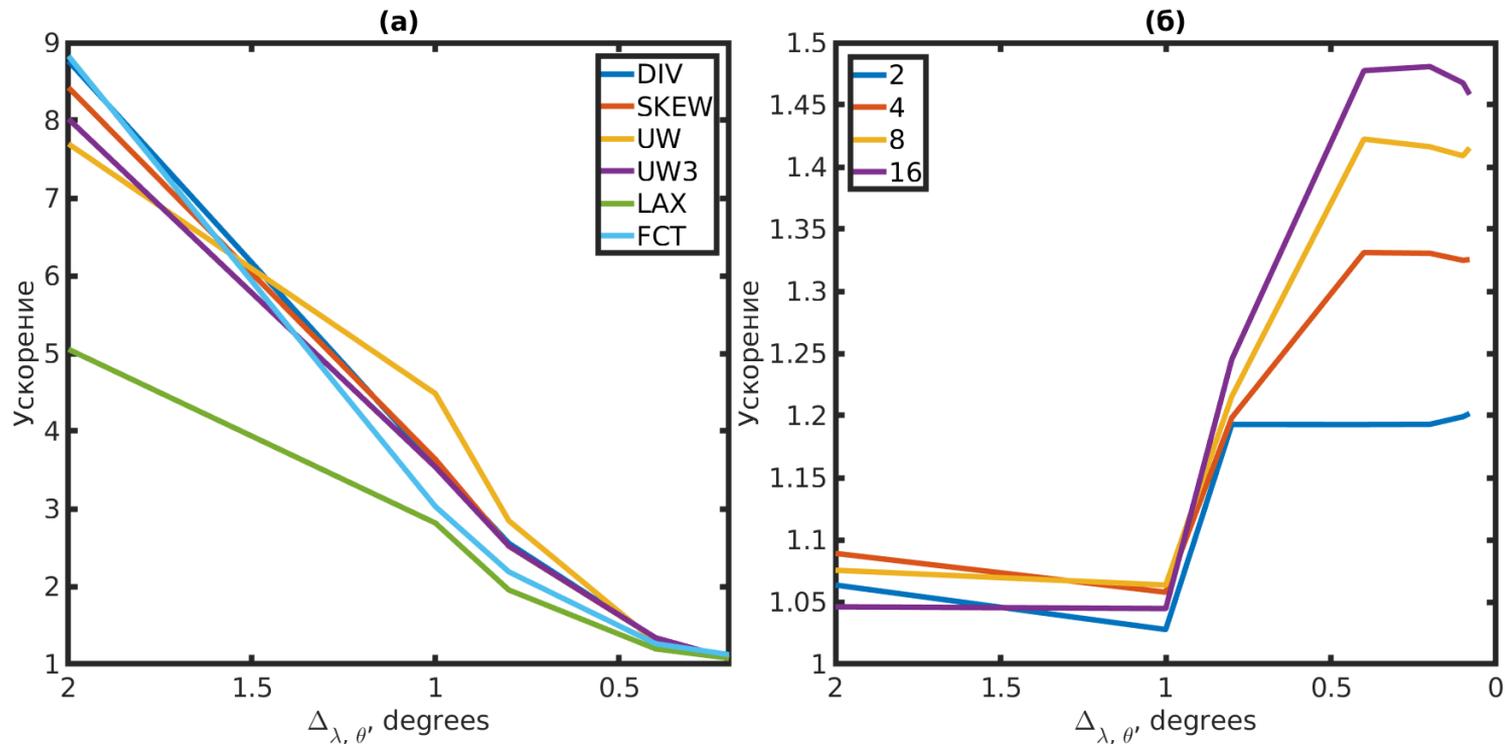


Рис. Ускорение метода scalar fusion с прямым порядком по отношению к базовой реализации для двумерной модели на A100 (а) и с переупорядочиванием по отношению к прямому порядку на V100 (б).

Оптимизация алгоритмов схем адвективного переноса

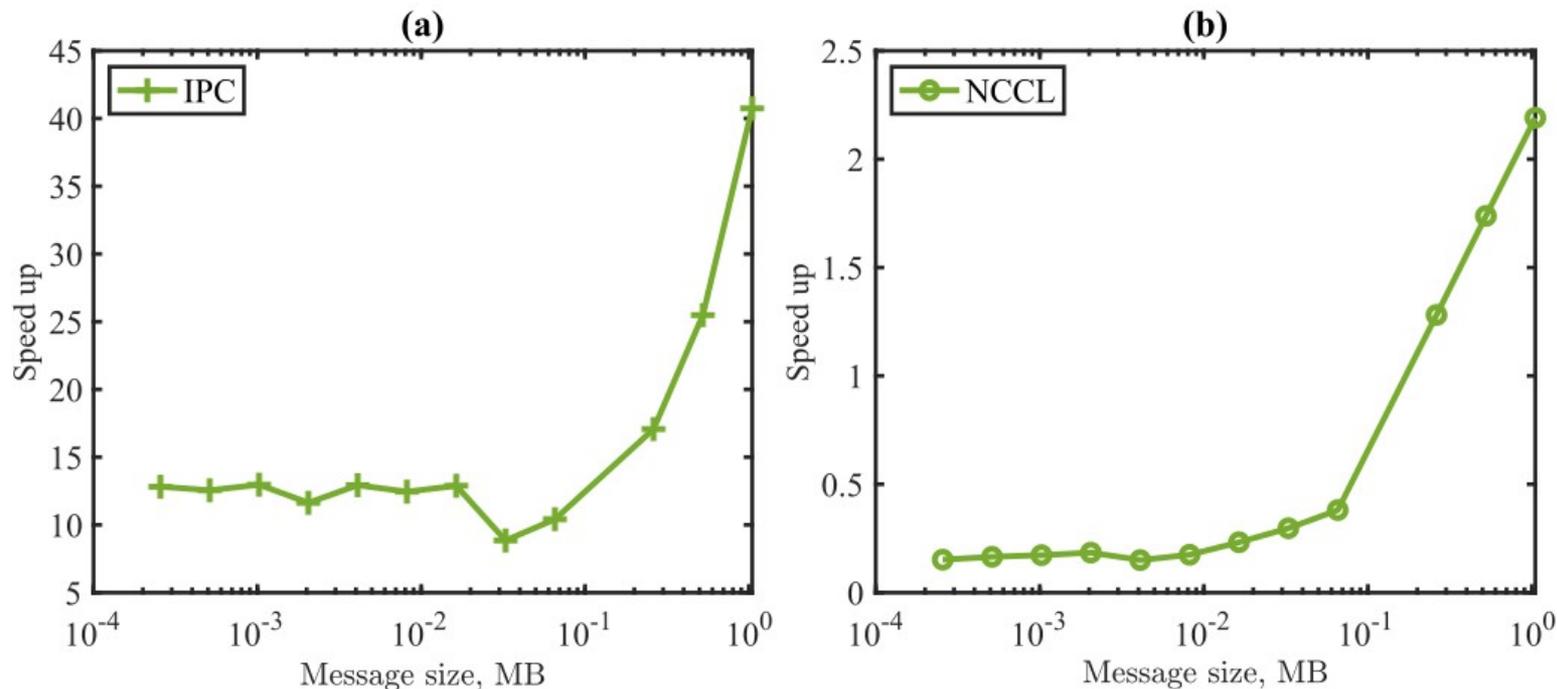


Рис. Ускорение реализации обмена с применением IPC (а) и NCCL (б) на V100 GPU в зависимости от размера пересылаемого сообщения

Разработка алгоритма переноса пассивной примеси в FP16

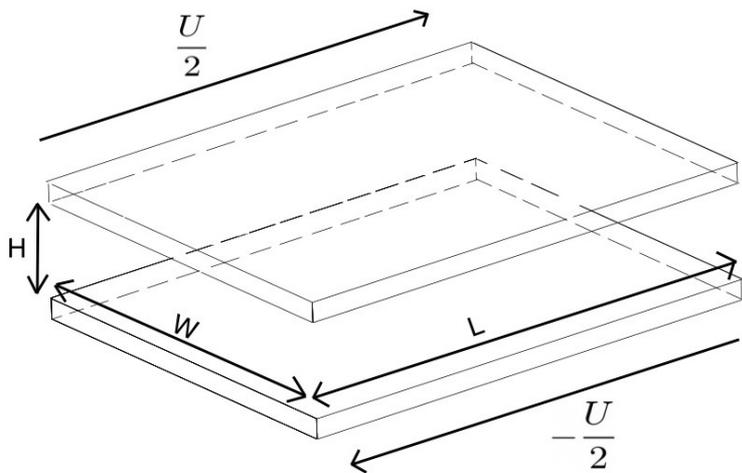


Рис. Течение Куэтта между движущимися стенками, L , W – продольный и поперечный размеры вычислительной области, H – высота канала, U – относительная скорость движения стенок.

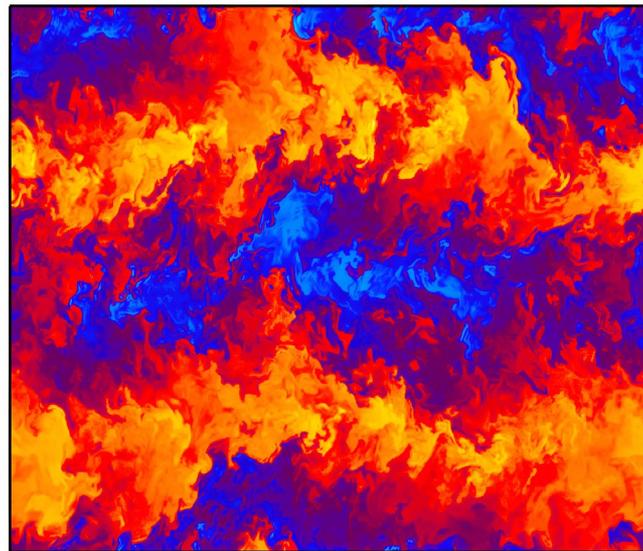


Рис. Продольная компонента скорости в турбулентном течении Куэтта.

Разработка алгоритма переноса пассивной примеси в FP16

Алгоритм интегрирования уравнения переноса в FP16:

- 1 $ADV^n = -\Delta t \left[\frac{\partial u_i C}{\partial x_i} \right]_h^n,$
- 2 $DIFF^n = \Delta t \left[\frac{1}{Re \cdot Sc} \frac{\partial^2 C}{\partial x_i \partial x_i} \right]_h^n,$
- 3 $RHS^n = \frac{3}{2}(ADV + DIFF)^n - \frac{1}{2}(ADV + DIFF)^{n-1} - \left[\Delta t T^{-1} C \right]_h^n,$
- 4 $C^{n+1} = C^n + RHS^n.$

Sc_k - число Шмидта для k -ой примеси, Re - число Рейнольдса, T_k - безразмерная величина, определяющая характерное время жизни для k -ой примеси.

Все данные и вся арифметика в FP16, шаг обновления концентрации 4 реализованы с использованием алгоритма компенсационного суммирования Кэхэна.

Разработка алгоритма переноса пассивной примеси в FP16

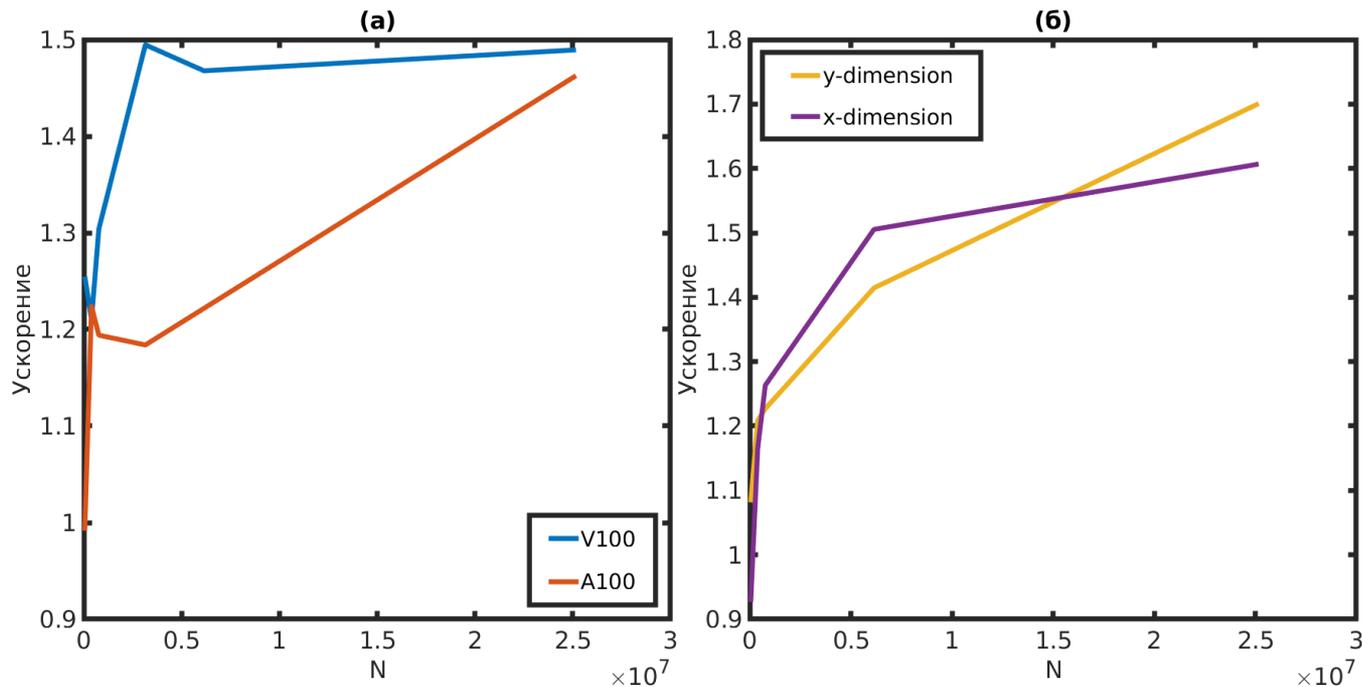


Рис. Ускорение моделирования переноса пассивной примеси на GPU (а) и ускорение MPI - обменов массива скаляра (б) в FP16 к FP32 в зависимости от размера сетки (N). Максимальный размер вычислительной сетки $N = 390 \times 262 \times 262$ соответствует $Re = 40000$, минимальный $N = 96 \times 64 \times 64 - Re = 5200$.

Заключение

Задача адвекции:

- **Прямой подход** техники **scalar fusion** **повысил эффективность** реализации **во всех** проведенных **тестах**, для двумерной модели метод обеспечил **ускорение** примерно в **9** раз на **A100 на самой грубой сетке**.
- **Scalar fusion** с переупорядочиванием обеспечил **ускорение** примерно в **1.5** раза на V100 по сравнению с прямым методом для двумерного случая при проведении экспериментов на **точных сетках**.

Ускорение обмена между GPU:

- Использование **IPC ускоряет обмен во всех экспериментах** (≈ 40 раз для сообщений размером 1МБ)
- Применение **NCCL** позволяет получить выигрыш в скорости проведения обмена только в случае **сообщений размером более 0.26МБ** (≈ 2.5 раза для сообщений размером 1МБ)

Задача адвекции-диффузии:

- Полная реализация блока переноса примеси в **FP16** с использованием алгоритма Кэхэна дает **достаточно точные** численные результаты
- **Ускорение** вычислений GPU - реализации в **FP16** до **1.5** раз в сравнении с исполнением в FP32 и **ускорение** вплоть до **1.6** раз выполнения **MPI-обменов** за счет уменьшения объема передаваемых данных

Модель внутренних водоемов:

- Разработана совместная модель термо-гидродинамики и биохимии внутренних водоемов суши на GPU. Максимальное достигнутое ускорение **16 раз** в сравнении с CPU ядром.