

# Отчет о проделанной работе с использованием ресурсов ИВЦ НГУ

## Тема работы:

Сравнительный анализ алгоритмов вычислительной алгебры на различных компьютерных архитектурах

## Состав коллектива :

Нектягаев Константин Игоревич, студент НГУ, бакалавр (ММФ)

## Постановка задачи:

В рамках этой работы предлагается численное сравнение ключевого ядра для задач нефтефизики - матрично-векторного умножения в случае матрицы с разреженными данными (преимущественно нулевые элементы). В качестве основных архитектур для сравнения предлагается CPU и GPU, а так же сравнение с реализацией из высокоуровневых библиотек (MKL, cuSPARSE).

(Пример используемый в работе - умножение матрицы двумерного оператора лапласа на единичный вектор)

## Состояние проблемы:

Современный мир предлагает огромное количество вычислительных узлов - CPU и GPU, распараллеливание на общих и распределенных данных, векторизацию и равномерный доступ ко всему массиву памяти. Что делает выбор оптимальной архитектуры для произвольного алгоритма достаточно сложной задачей.

## Описание работы:

В данной работе главный вопрос звучит так *“Какая архитектура является более подходящей для этой задачи?”*.

### Преследуемые цели:

1. Реализовать матрично-векторное умножение для разреженных матриц для CPU и GPU (NVIDIA) как основное ядро итерационных алгоритмов и оценить качество полученной реализации.
2. Численно сравнить полученные реализации с аналогичными из

высокопроизводительных библиотек и друг с другом, чтобы сделать вывод об эффективности архитектур для итерационного решения задач с разреженной матрицей .

В работе кратко обозреваются CPU и GPU архитектуры, а также выделяются их сильные стороны.

Также приводится обзор используемых форматов хранения разреженных матриц:

- COO
- CSR/CRS
- DIAG

Использовались собственные алгоритмы умножения разреженной матрицы на вектор. Запускались библиотечные реализации данных алгоритмов (MKL, cuSPARSE).

В данной работе сравнение реализаций производится относительно предельной производительности полученной при помощи `stream benchmark`

**Используемые прогр. инструменты (предоставлены ИВЦ НГУ):**

- Cuda toolkit
- OpenMP
- MKL
- cuSPARSE

**Используемое аппаратное обеспечение (предоставлены ИВЦ НГУ):**

- блей д-сервер HP SL390s G7
  - Два 6-ядерных процессора Xeon X5670 с тактовой частотой 2933 МГц
  - Три карты NVIDIA Tesla M2090

**Обратить внимание:**

Следует отметить, в случае с ЦП (CPU) тест пропускной способности производится относительно оперативной памяти (RAM), т. к. основной интерес представляют матрицы большого размера, которые не помещаются в кэш-памяти (cache). Для удобства на всех графиках желтым цветом будут отмечены “пакетные” реализации. О других реализации на графиках, название соответствует формату хранения. Естественно, для каждого формата свой алгоритм.

Речь пойдет только о “параллельных” реализациях.

В вычислениях используется исключительно двойная точность (double).

Предельная производительность, относительно которой идет сравнение, рассчитана при помощи теста пропускной способности (`stream benchmark`) так называемой триады ('triad') -  $(a * x + y)$ , где  $a$  - константа.

# Результаты:

- Реализованы параллельные алгоритмы MV умножения (разреженная матрица) для архитектур ЦП(CPU) и графической карты(GPU)
- Подтверждена эффективность алгоритмов при сравнение с реализациями из высокопроизводительных библиотек (MKL, cuSPARSE)
- Подтверждена абсолютная эффективность как самих алгоритмов так и реализации из высокопроизводительных библиотек при сравнении с пиковыми значениями производительности.
- На основе проведенных сравнении - GPU оптимальная архитектура для определенных memory bandwidth задач.

# Полученные графики:

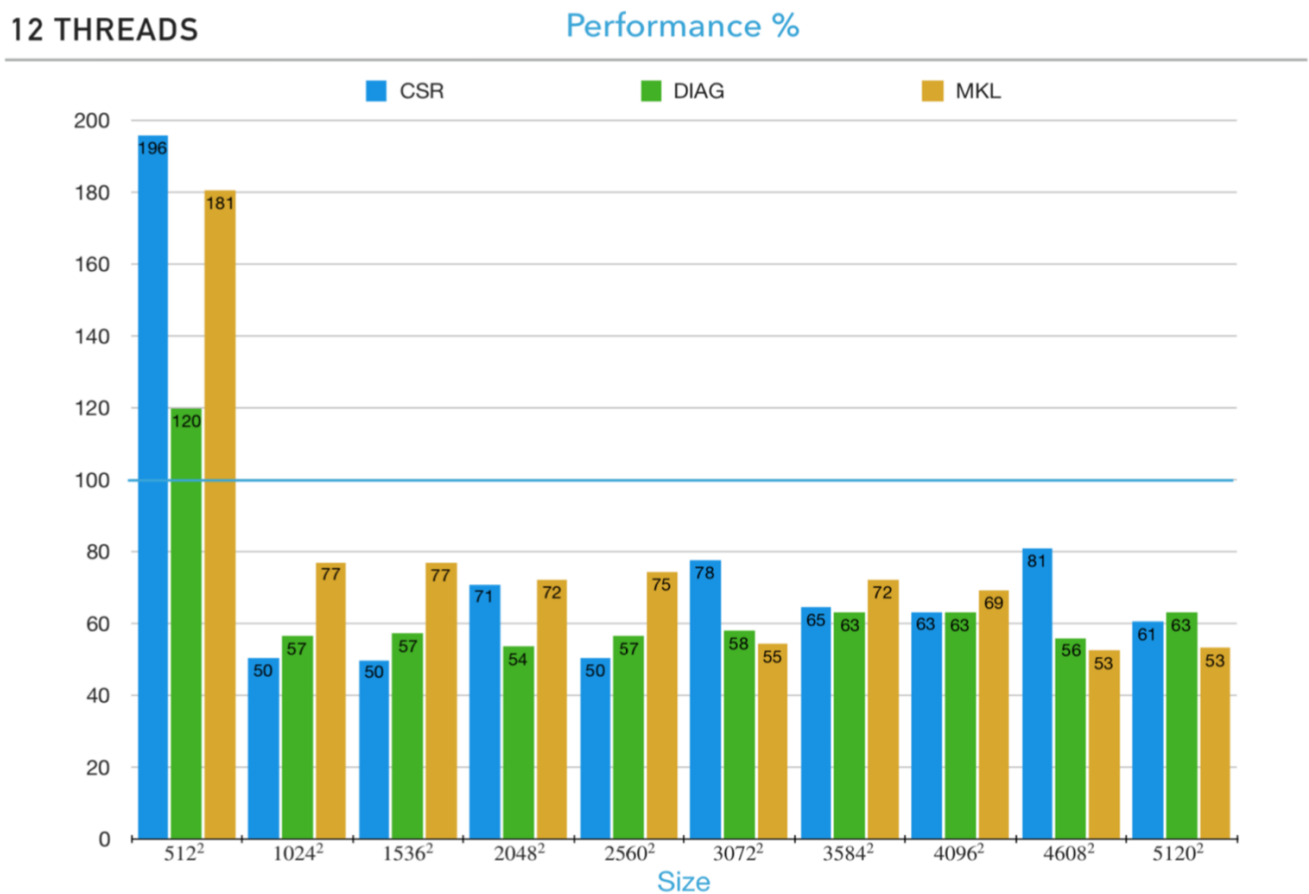


Figure 1: CPU 12 Threads

## 6 THREADS

## Performance %

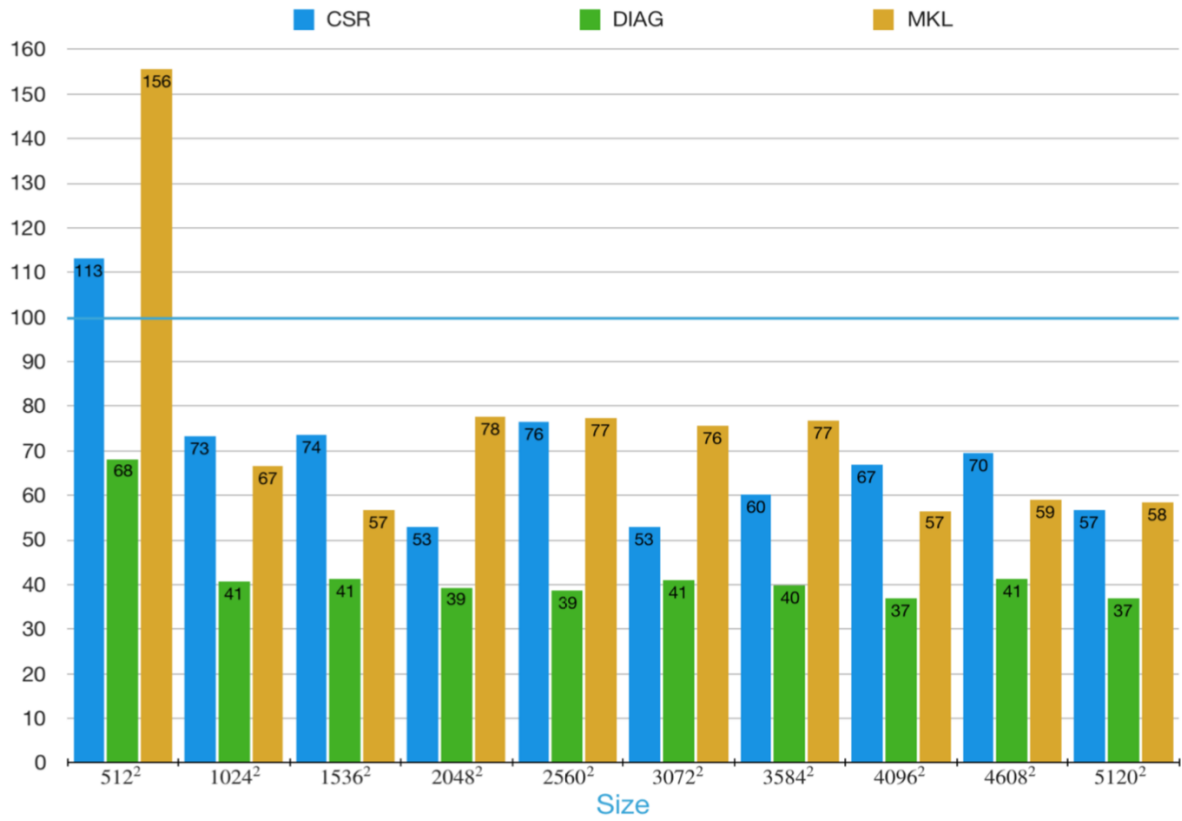


Figure 2: CPU 6 Threads

## GPU

## Performance %

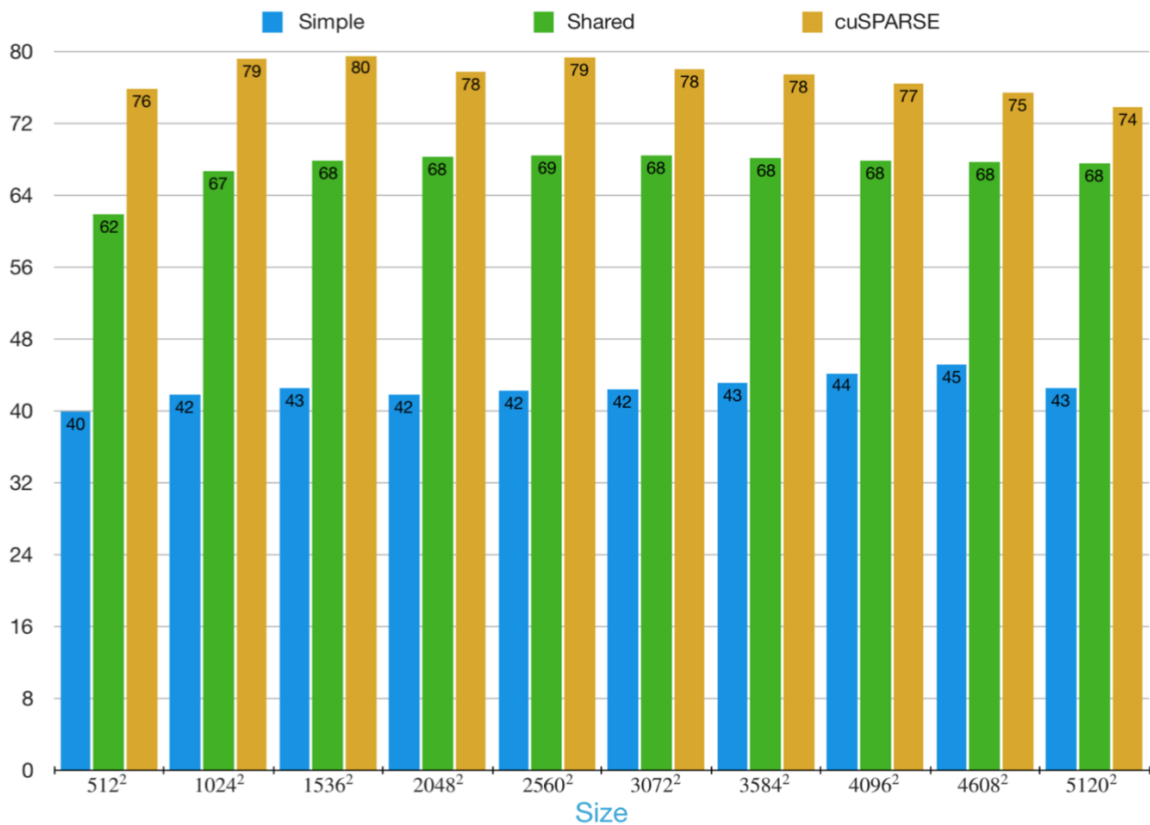


Figure 3: GPU

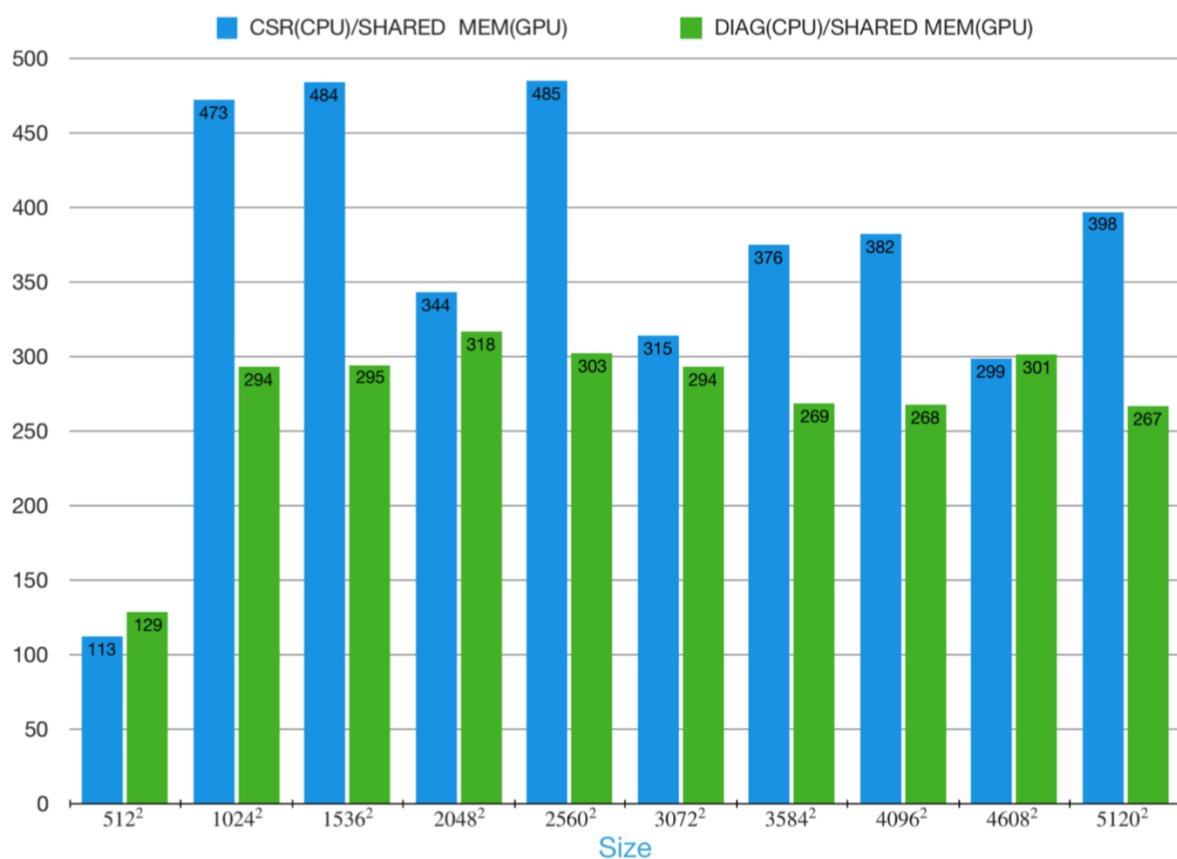


Figure 4: CPU vs GPU

## Эффект использования кластера:

Без кластера было бы трудно где-то найти подобного рода вычислительные ресурсы с такой доступностью. Можно считать доступ к кластеру дал возможность провести данное исследование.

## Впечатления от работы:

Хочу выразить благодарность Калюжному Владиславу Анатольевичу за крайне оперативную поддержку и подробные инструкции.