

Оптимизация производительности алгоритма избыточного кодирования данных

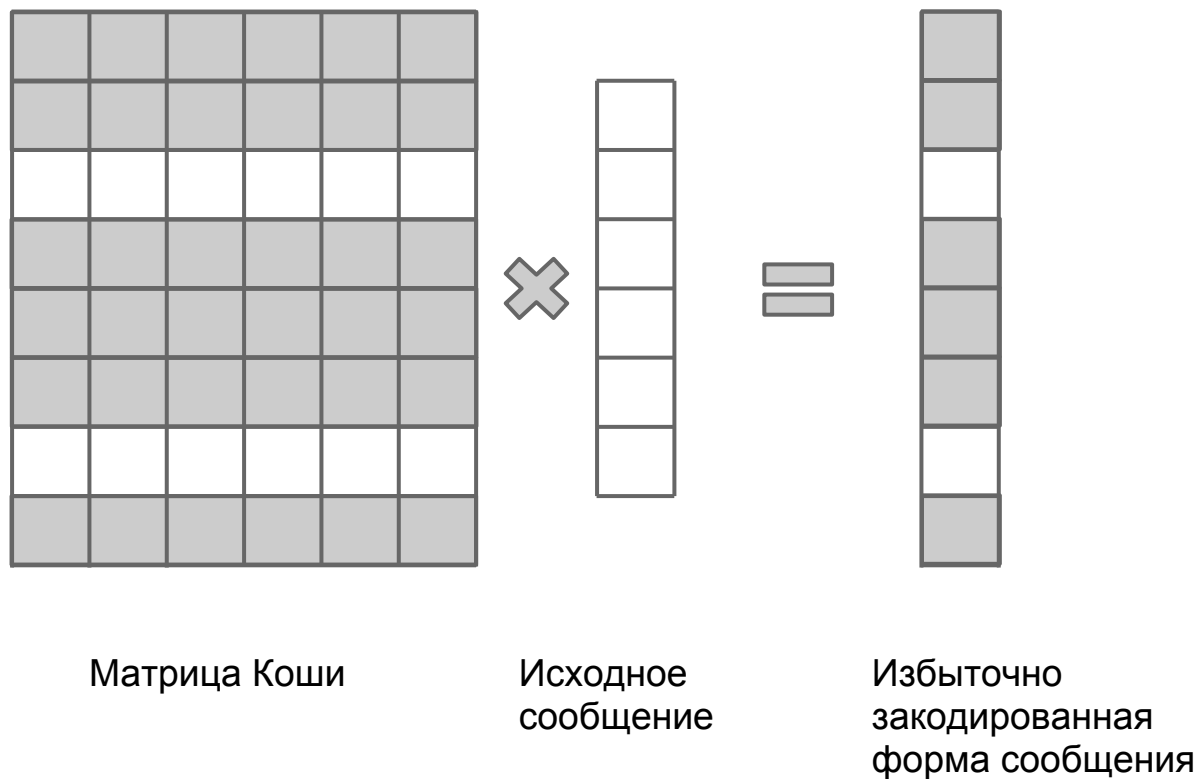
Сумбатянц Илья
Научный руководитель: Д. В. Иртегов

Исходные данные

Система *распределенного* резервного копирования на основе алгоритма *избыточного кодирования* данных.

- *Распределенность*: Distributed hash table
- *Избыточность кодирования*: алгоритм на основе матриц Коши над полем Галуа

Алгоритм избыточного кодирования



Алгоритм избыточного кодирования

■	■	■	■	■	■
■	■	■	■	■	■
□	□	□	□	□	□
■	■	■	■	■	■
■	■	■	■	■	■
■	■	■	■	■	■
□	□	□	□	□	□
■	■	■	■	■	■

Матрица Коши

×

□	□	□	□
□	□	□	□
□	□	□	□
□	□	□	□
□	□	□	□
□	□	□	□
□	□	□	□
□	□	□	□

Исходные
данные

=

■	■	■	■
■	■	■	■
□	□	□	□
■	■	■	■
■	■	■	■
■	■	■	■
□	□	□	□
■	■	■	■

Избыточно
закодированная
форма данных

Цели и идеи

Цель: оптимизация алгоритма избыточного кодирования

Идеи:

1. Нативный код
2. Аппаратное ускорение:
 - a. SIMD
 - b. GPGPU

Задачи

1. Тестирование
2. Рефакторинг
3. **Реализация и оптимизация**
4. Интеграция
5. Сбор результатов

Реализация и оптимизация

1. Реализация
2. Интеграция
3. Тестирование
4. Измерение времени исполнения
5. Оптимизация (изменение реализации)



Реализация на С

- Основа дальнейших оптимизаций
- Реализация арифметики поля Галуа
- Определение интерфейса

SSE

Векторизация:

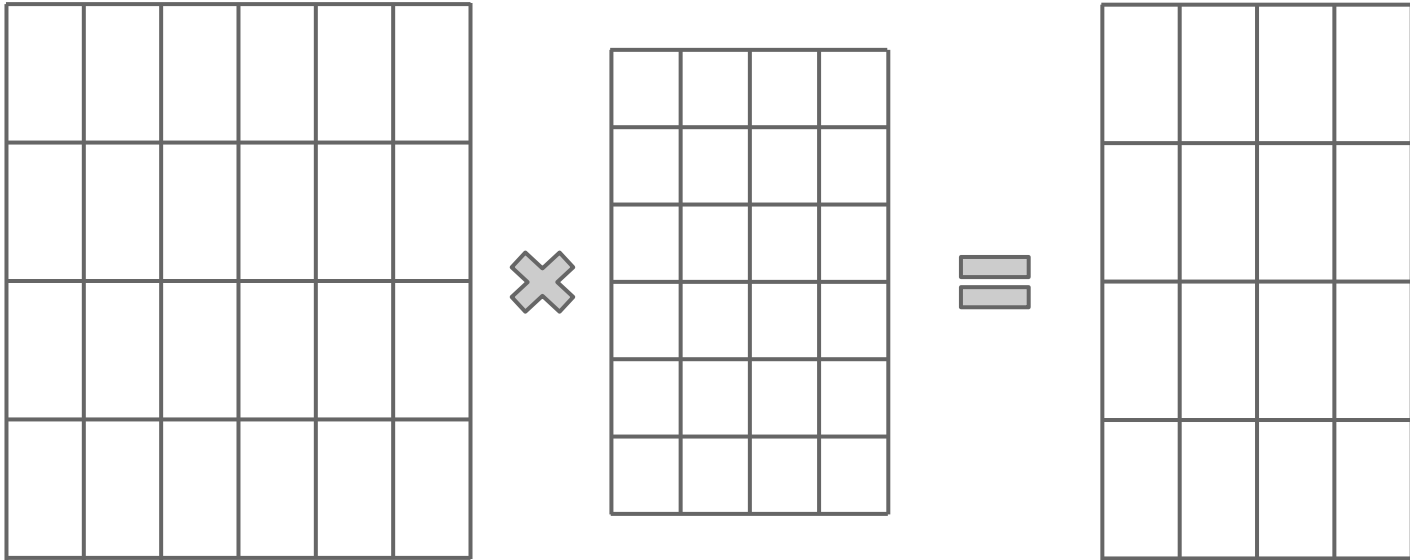
- Сложение
- Умножение

128-битные регистры: 16 элементов поля
Галуа

Выравнивание данных:

- Расширение матрицы Коши и выходных данных

SSE



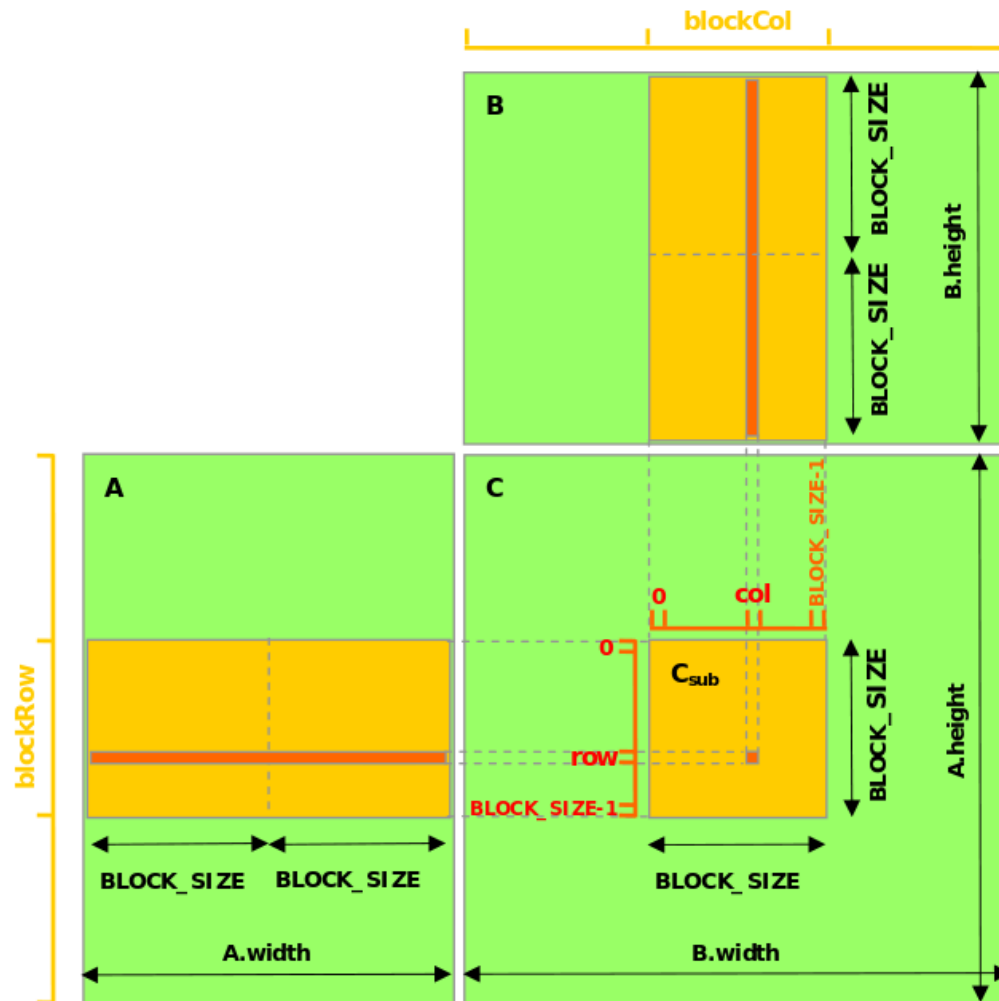
GPGPU

Фреймворк: OpenCL

Типы памяти (в терминах OpenCL):

1. Глобальная
2. Локальная

GPGPU



Результаты

Сравнение с Java-реализацией

ЦПУ:

Производительнее в 2.5 - 4 раза

Видеоадаптер:

Производительнее в 3 - 18 (10 - 18) раз.

Дальнейшее развитие

В рамках оптимизации алгоритма:

1. Параллельная реализация существующих решений
2. Комбинированная реализация (GPU + CPU)
3. Вычислительная сеть (разделение задачи разбиения между участниками сети)

В рамках системы:

1. Упрощение прототипа до состояния распределенного хранилища (без графического интерфейса)
2. Создание интерфейса управления сервером (например, tcr)
3. Реализация оберток над распределенным хранилищем:
 - a. Резервное копирование
 - b. Передача в сеть многотомные архивов

Спасибо за внимание