

Параллельные вычислительные технологии (ПаВТ) 2015
(30 марта - 3 апреля 2015 г.)

Исследование
эффективности
применения сжатия при
передаче данных из
основной памяти в память
графического ускорителя

Приказчиков С.О., Костенецкий П.С

Южно-Уральский государственный университет

Работа выполнена при финансовой поддержке Минобрнауки РФ в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы» (Соглашение № 14.574.21.0035).

АКТУАЛЬНОСТЬ

- Современные графические ускорители обладают очень высокой производительностью.
- Графические ускорители могут применяться для обработки баз данных.
- В то же время скорость передачи данных по шине PCI-Express в десятки раз меньше скорости передачи данных внутри ГПУ.



Параметры эксперимента

Размер базы данных (без сжатия)	1 500 МБ
Количество кортежей	187 500 000

- База данных сжимается в оперативной памяти компьютера, после чего передается на ГПУ;
- Запрос выполняется на ГПУ.

Моделируемый запрос

```
SELECT SUM(attribute) FROM Relation
```

Используемое оборудование

Оборудование	Характеристики	
Процессор	Intel Xeon E5-2687W v2 – 2 шт.	
ОЗУ	128 Гб DDR3 ECC Registered	
Системная шина	PCI Express 3.0	
Графический ускоритель	Модель	NVIDIA Tesla K40m
	Ядер CUDA	2880
	Тактовая частота ядра	745 МГц
	Тактовая частота памяти	6000 МГц
	Пропускная способность памяти	288 Гб/с
	Разрядность памяти	384-bit

Сжатие

Выбраны 2 метода сжатия:

- Кодирование длин серий (Run-length encoding);
- Удаление нулей (Null suppression).

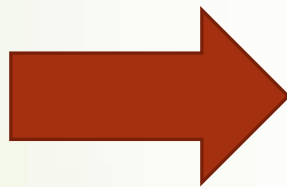
Оба этих метода позволяют применять агрегатные функции без предварительной распаковки

Кодирование длин серий

Данные

5
5
5
5
7
7
7
7
7
7
4

Длины серий Элементы

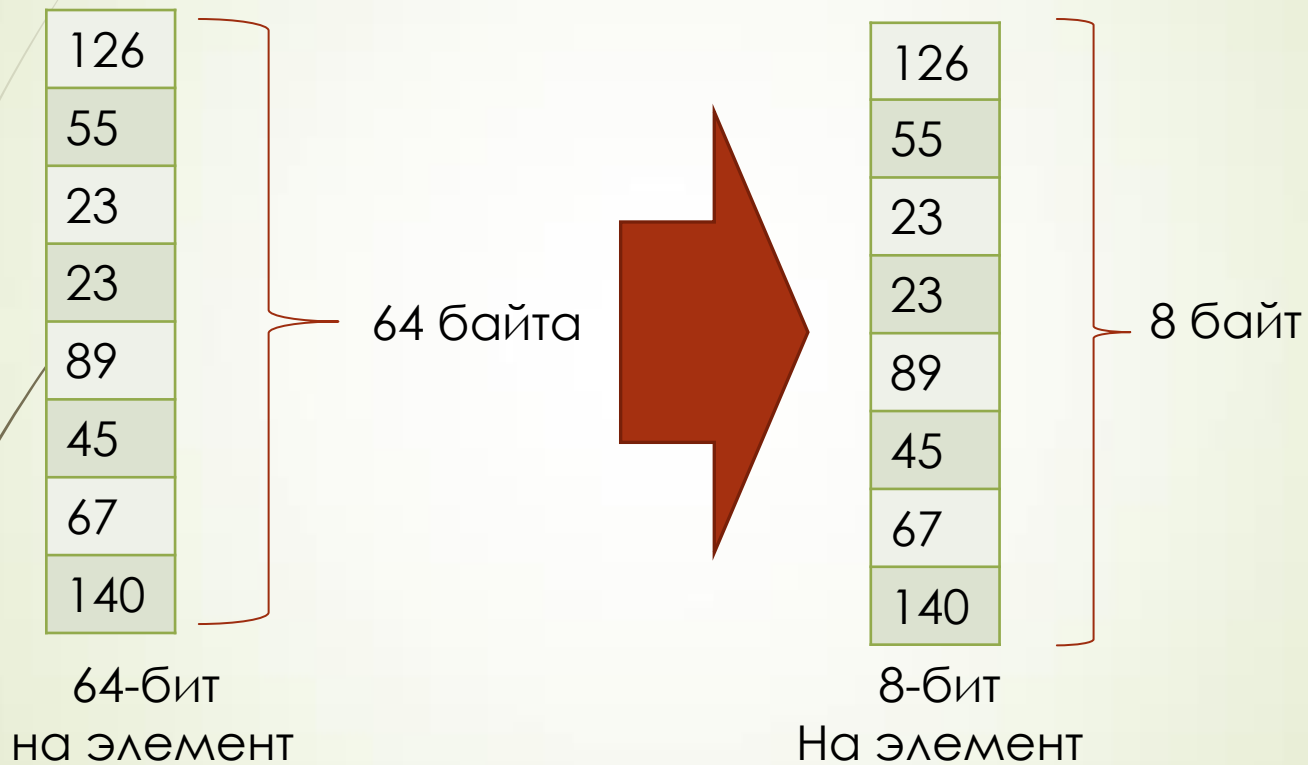


4
5
1

+

5
7
4

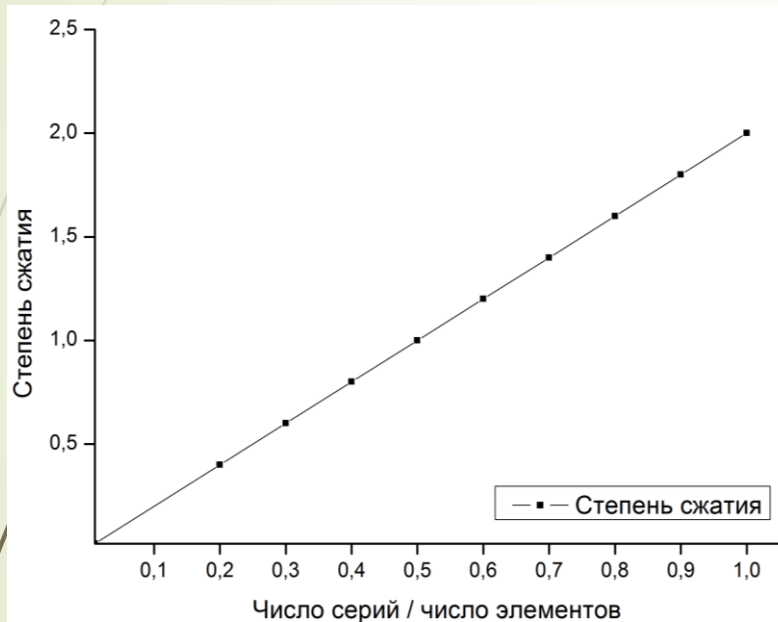
Удаление нулей



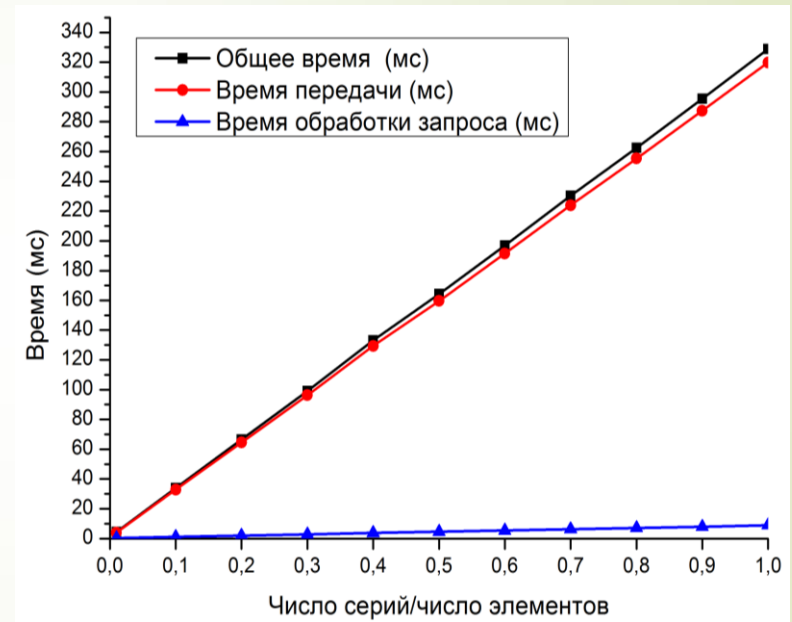
Выбранные параметры СЖИМАЕМЫХ ДАННЫХ

- ▶ Алгоритм кодирования длин серий:
Отношение количества серий к общему количеству элементов данных
- ▶ Алгоритм удаления нулей:
Оставшийся после сжатия объем элементов

Результаты экспериментов

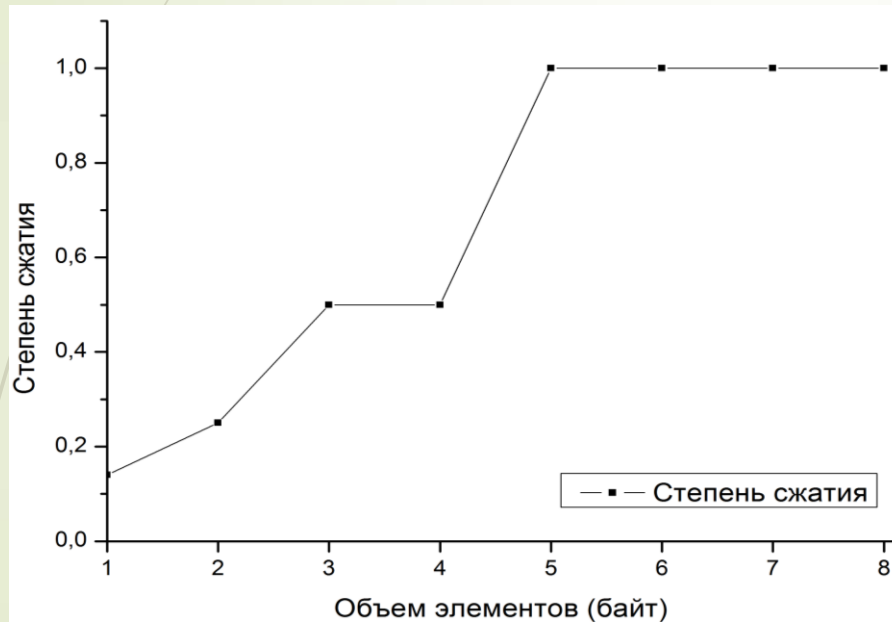


Степень сжатия методом кодирования длин серий

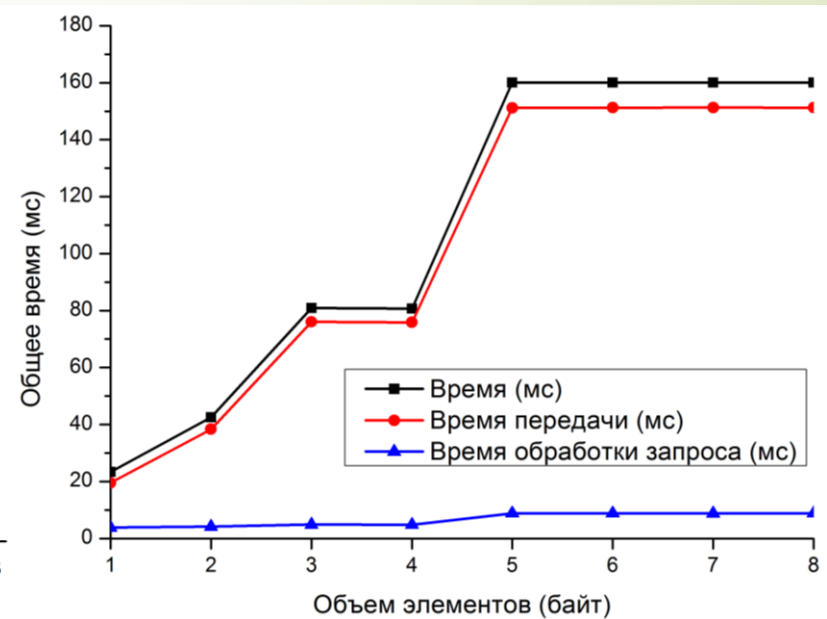


Время обработки данных, сжатых методом кодирования длин серий

Результаты экспериментов



Степень сжатия методом
удаления нулей



Время обработки данных,
сжатых методом
удаления нулей

Заключение

Предложенный подход позволяет:

- быстрее передавать данные в память ГПУ;
- экономить память ГПУ;
- повысить скорость выполнения запросов к БД.