

Исследование масштабируемости библиотеки XAMG при использовании GPU для решения систем уравнений

Р.М. Куприй^{1,2}, Б.И. Краснопольский¹, К.А. Жуков²
r.kuprii@imec.msu.ru

¹Институт механики МГУ имени М.В. Ломоносова,

²Факультет ВМК МГУ имени М.В. Ломоносова

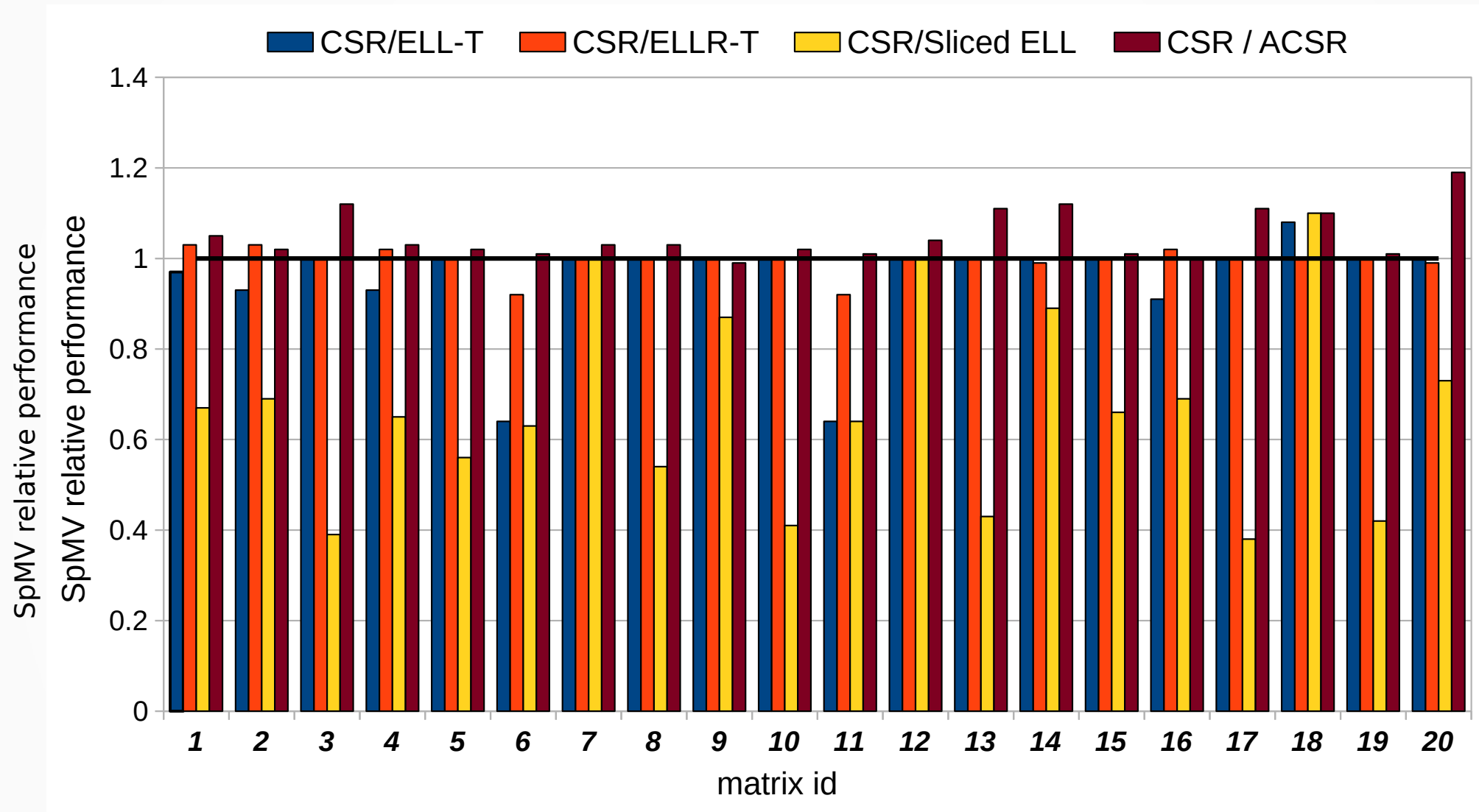
Введение

- Нахождение решения систем линейных алгебраических уравнений (СЛАУ)
 - типичная задача в численном моделировании
 - большие разреженные матрицы
 - до 95% от всего времени расчёта
- Для решения СЛАУ могут применяться итерационные численные методы
 - операции с матрицами и векторами
 - узкое место — пропускная способность шины памяти
- Использование GPU для ускорения вычислений
 - адаптация всех операций для GPU

Реализация в XAMG

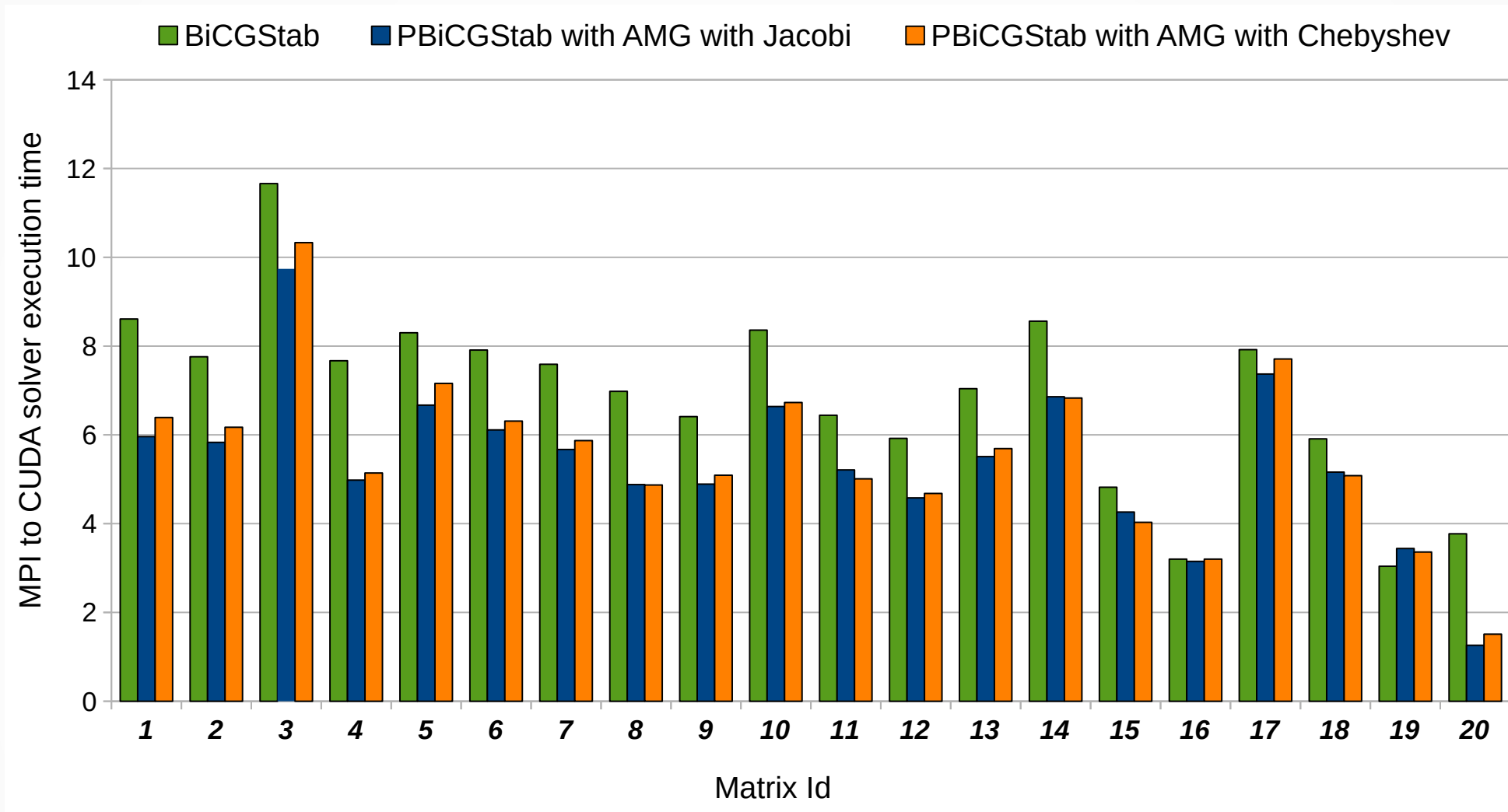
- Исследование проводилось в рамках развития библиотеки численных методов XAMG (gitlab.com/xamg/xamg)
 - Для хранения матриц на GPU реализованы различные форматы хранения: ELL, ELLR, Sliced ELL, CSR, Adaptive CSR
 - Реализованы соответствующие версии алгоритма SpMV с использованием CUDA
- Для тестирования использовался репрезентативный набор из 20 матриц из SuiteSparse Matrix Collection
- Вычислительные эксперименты проводились на узле кластера
 - CPU: 2x AMD EPYC 7742; GPU: 8x NVIDIA A100
 - $BW_{GPU} / BW_{CPU} \sim 7$ раз

SpMV: сравнение различных форматов с CSR



- Форматы на основе ELL в среднем менее эффективны
 - Формат Adaptive CSR наиболее эффективен

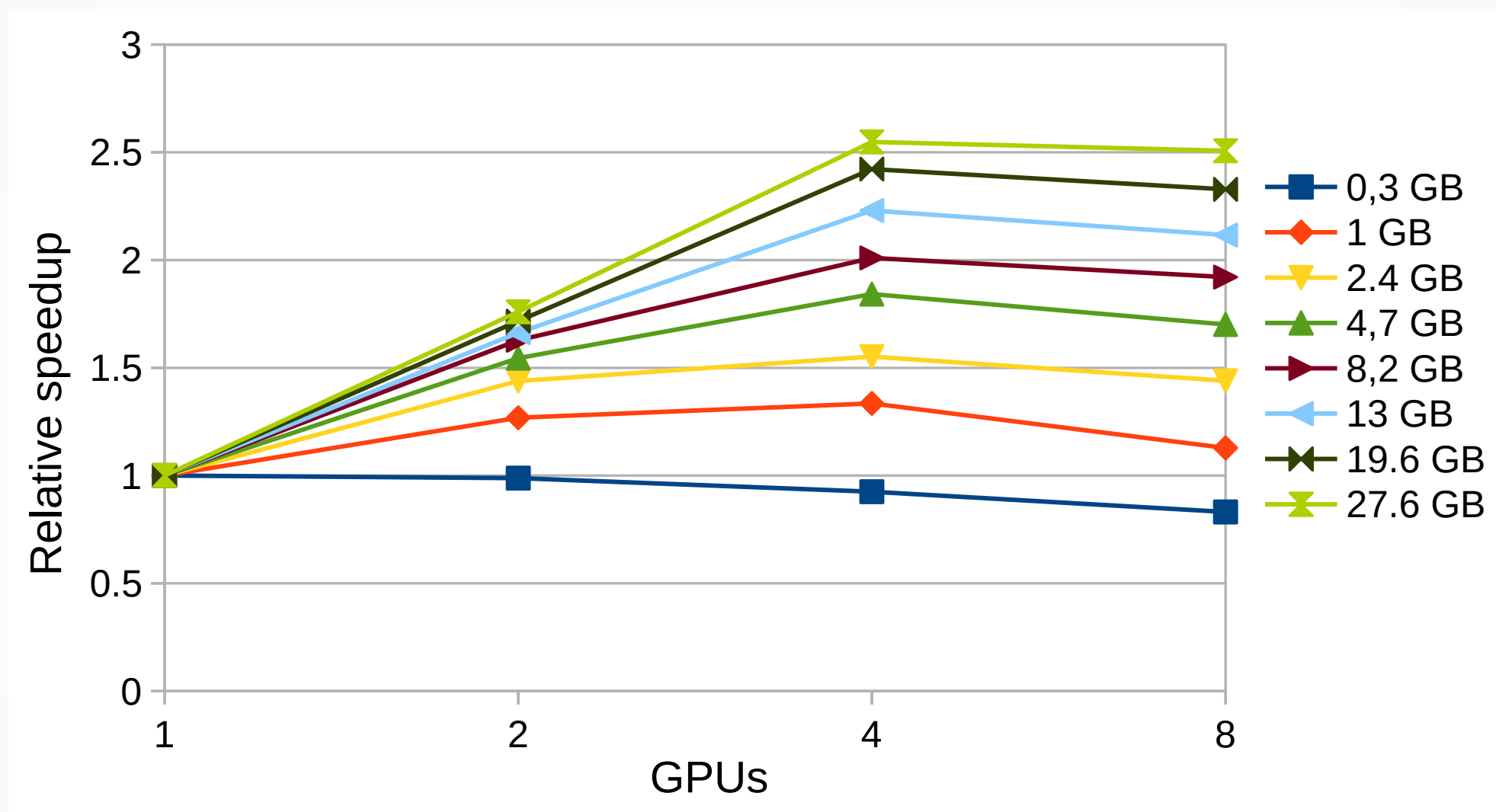
Решение СЛАУ: CPU vs GPU



Достагаемое ускорение варьируется от 2 до 12 раз
и зависит от размера матрицы

В среднем ускорение составляет 7 раз для метода без предобуславливателя
и 5.5 раз для метода с предобуславливателем

Масштабируемость на нескольких GPU



Эффективность масштабируемости ограничена 4 GPU

Максимальное достигаемое ускорение составляет 2.5 раза

Заключение

- Проведено исследование форматов хранения разреженных матриц: CSR, Adaptive CSR, ELL, ELL-R, Sliced ELL
- Форматы и соответствующие реализации операции SpMV на языке CUDA интегрированы в библиотеку XAMG
- Выполнена серия экспериментов, установлено что:
 - Наиболее оптимальным является формат: Adaptive CSR
 - Использование GPU позволяет ускорить решение СЛАУ до 12 раз в зависимости от объема данных требуемых для хранения матрицы
 - Масштабируемость при использовании нескольких GPU позволяет укорить расчёты до 2.5 раз при использовании 4 GPU

Спасибо за внимание!