



LOBACHEVSKY  
UNIVERSITY

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. Н.И. ЛОБАЧЕВСКОГО  
ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МАТЕМАТИКИ И МЕХАНИКИ  
КАФЕДРА ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ И  
СИСТЕМНОГО ПРОГРАММИРОВАНИЯ

# **Сравнение производительности операции SpMV для различных форматов хранения разреженной матрицы на процессорах архитектуры RISC-V**

А.Ю. Пирова, А.А. Воденеева,  
К.И. Ковалев, А.В. Устинов,  
М.А. Загрядсков, Д.Д. Литвяков,  
А.И. Кулик, Е.А. Козинков,  
В.Д. Волокитин, И.Б. Мееров



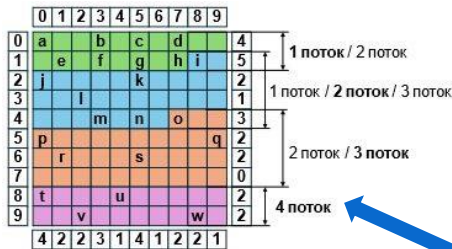
«Суперкомпьютерные дни в России»  
Москва, 28-29 сентября 2026 г.

# Цель работы

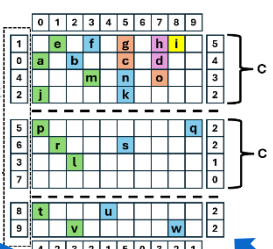
- ❑ **Цель работы:** Сравнение реализаций алгоритма умножения матрицы на вектор ( $\text{SpMV}$ ,  $y := \alpha Ax + \beta y$ ) для различных форматов хранения разреженных матриц на устройствах RISC-V.
- ❑ **Проблема:** производительность сильно зависит от формата хранения и алгоритма, заранее не понятно, какой формат использовать для конкретной матрицы.
- ❑ **Шаги по решению:**
  - Реализована операция  $\text{SpMV}$  для 9 форматов хранения разреженных матриц: **CSR, Sell-C-Sigma, CSR2, CSR5, CVR, VHCC, VNEC, LAV, HCSR**.
  - Проанализированы характеристики матриц, влияющие на выбор формата для наилучшей производительности  $\text{SPMV}$ .
  - Разработана программа для автоматического подбора формата.

# Форматы хранения разреженной матрицы

VNEC



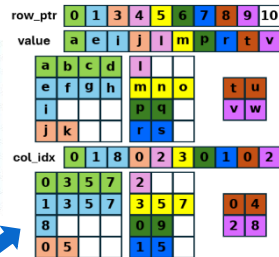
Sell-C-σ



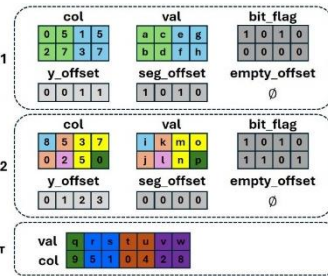
LAV



CRS2



CRS5



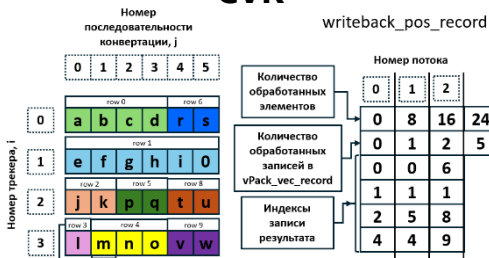
CRS



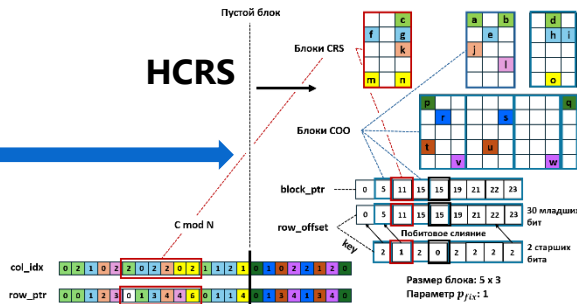
VHCC



CVR



HCRS



# Программная реализация

- ❑ **Библиотека RVVLASparse** – открытая библиотека операций с разреженными матрицами для процессоров RISC-V с использованием интринсиков RVV 1.0.
  - Операция SpMV для 9 форматов хранения разреженных матриц.
  - В разработке: SpMM, маскированный SpMM, TRSM, TRSV.
- ❑ **Программа автоматического подбора формата**
  - Вычисление 67 признаков матриц – программа на C++, результат в формате CSV.
  - Обучение моделей и подбор формата – программа на Python (используются библиотеки pandas, scikit-learn).
  - Лучшие методы: Random Forest Classifier для double (проигрыш оптимальному формату в среднем 2%, в худшем случае 22%) и Random Forest Regressor для float (6% и 32%).

# Тестовая инфраструктура

## ❑ Тестовое окружение:

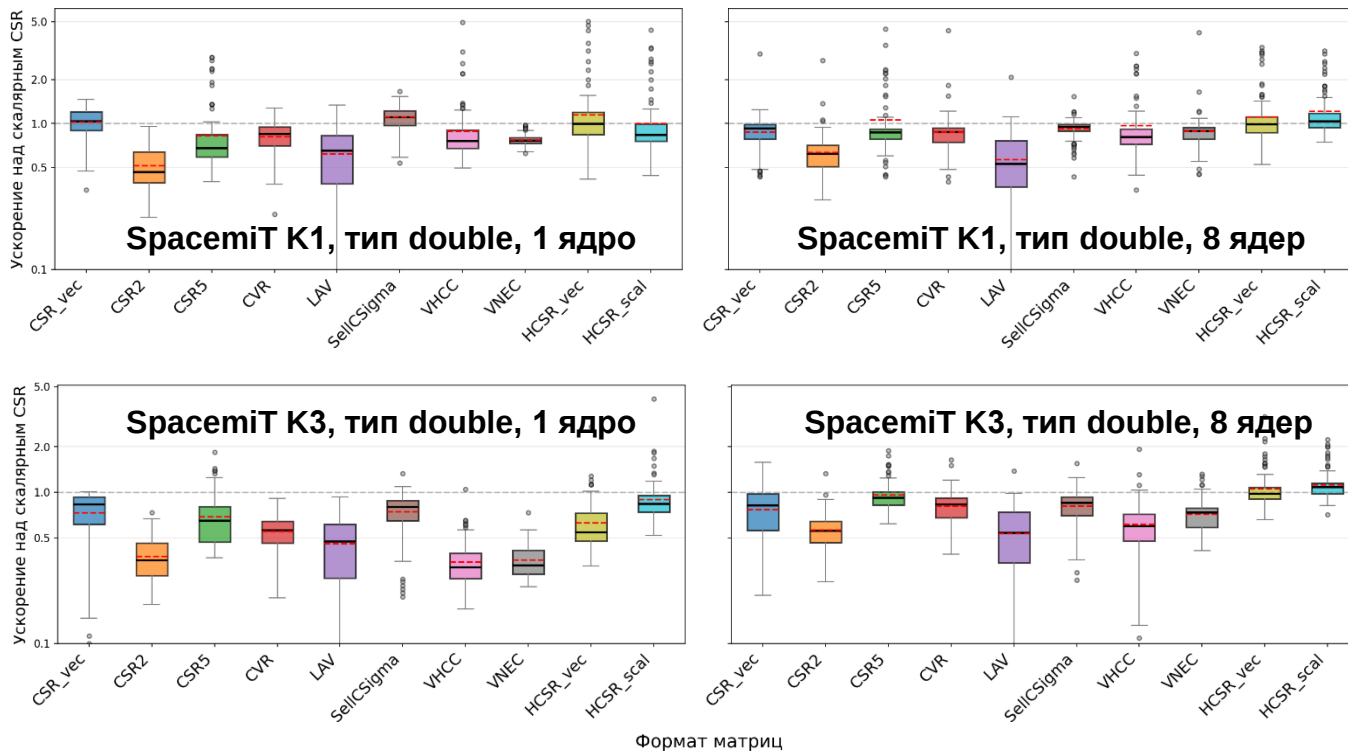
- 1) Banana Pi BPI-SM10, процессор SpacemiT Keystone K3 (x100), Out-of-Order, 256-bit RVV 1.0, 8 GB, ОС Bianbu 4.0.1.
- 2) Banana Pi BPI-F3, процессор SpacemiT Keystone K1 (x60), In-Order, 256-bit RVV 1.0, 16 GB, ОС Bianbu 2.2.1.

## ❑ Кросс-компилятор GCC RISC-V 14.2.0.

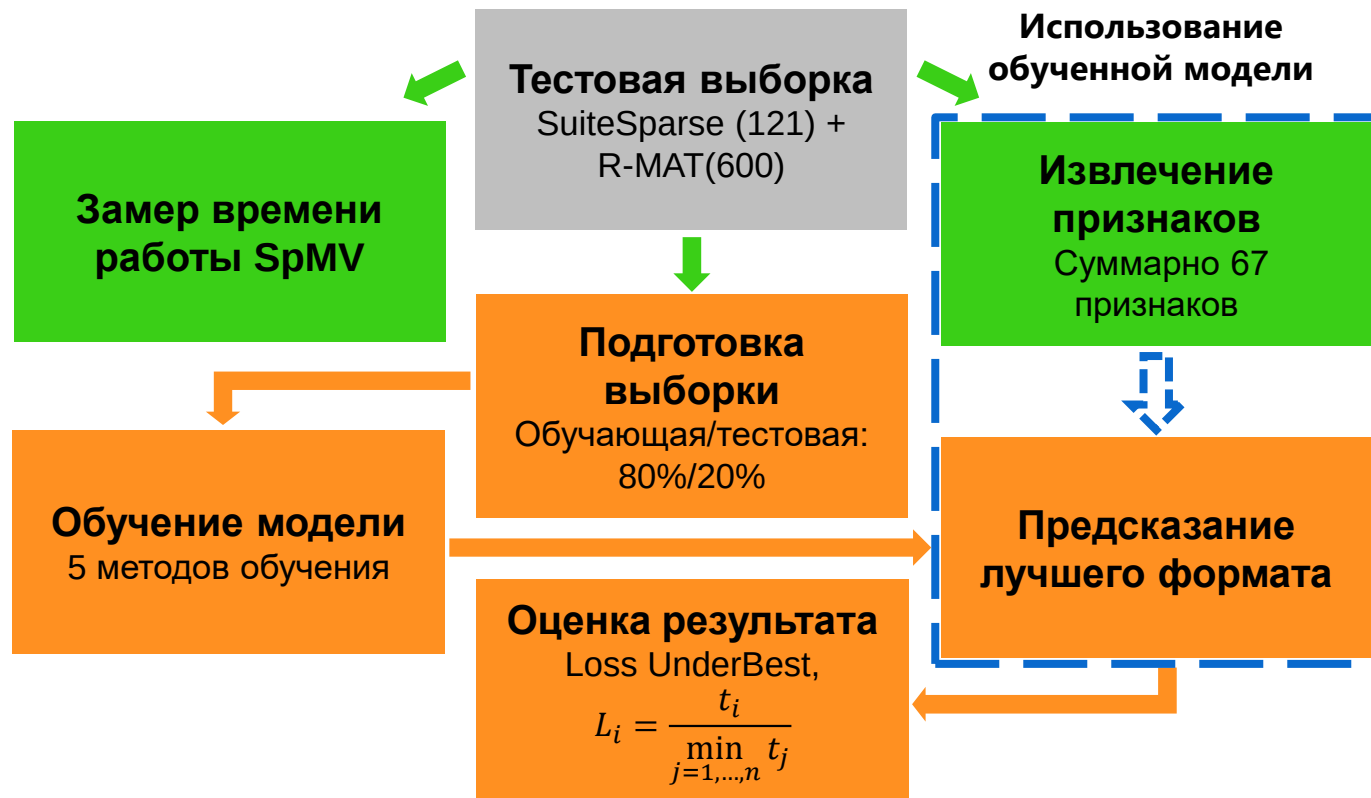
## ❑ Тестовая выборка: 121 матрицы из коллекции SuiteSparse размером от $0.5 \cdot 10^3 \times 400 \cdot 10^3$ до $8.39 \cdot 10^6 \times 8.39 \cdot 10^6$ и заполненностью от $3.1 \cdot 10^{-7}$ до $1.4 \cdot 10^{-2}$ . Плотный вектор заполнялся случайными числами из интервала $[0, 1)$ , значения $\alpha = 1.0$ , $\beta = 0.0$ фиксированы.

# Вычислительные эксперименты

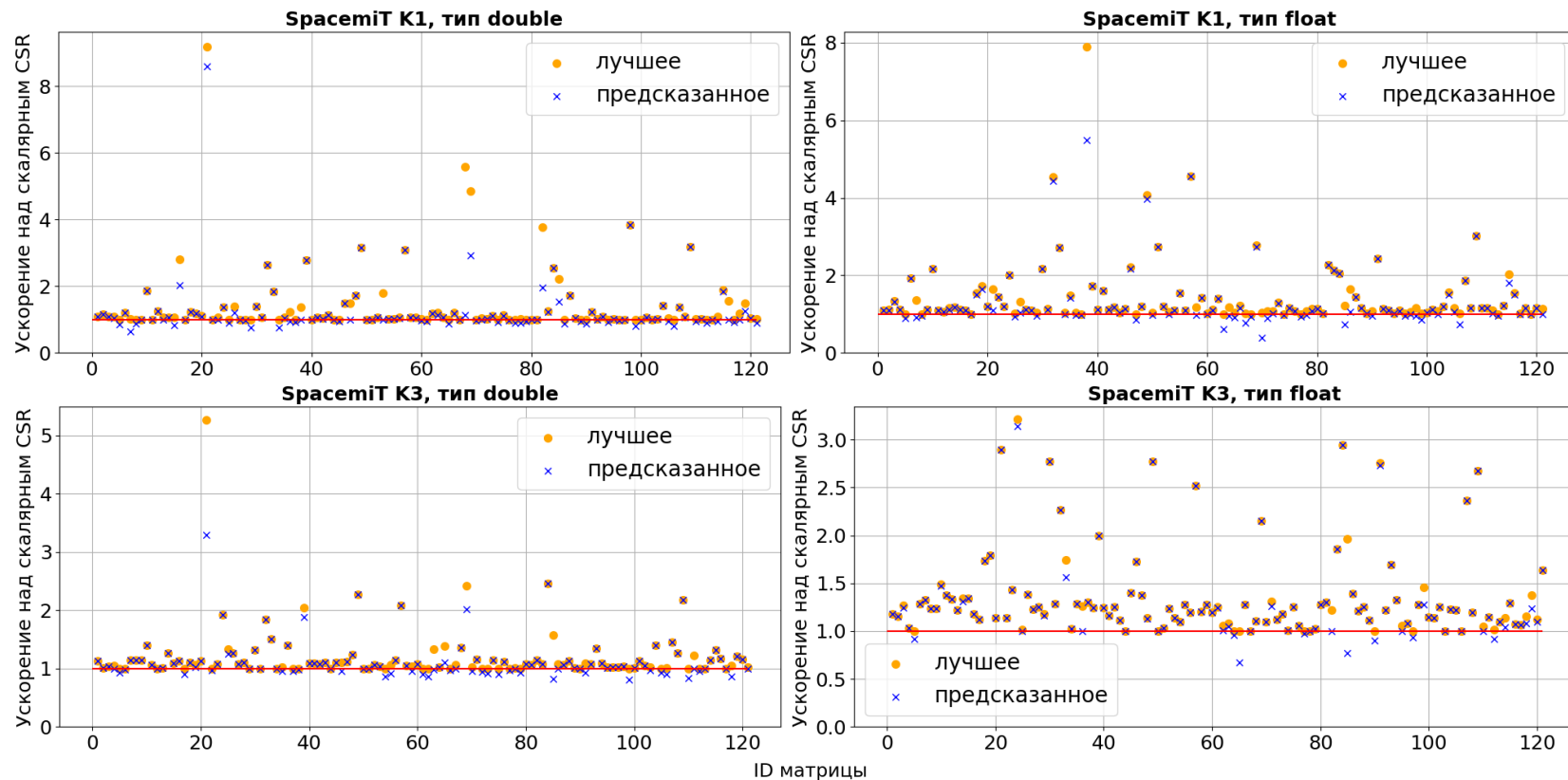
## Распределение ускорения в различных форматах



# Схема обучения и тестирования модели автоматического подбора формата



# Результаты тестирования модели автоматического подбора формата



# Заключение

- ❑ Лучшие результаты получены в форматах **CSR**, **Sell-C-Sigma**, векторном и скалярном **HCSR**. Наибольшее среднее ускорение даёт скалярный HCSR на 8 ядрах: в **1,22** (double) и **1,33** (float) раза на **SpacemiT K1**, в **1,12** и **1,34** раза на **SpacemiT K3**.
- ❑ На **SpacemiT K1** векторизация и параллелизм ускоряют вычисления: векторный **HCSR** ускоряет SpMV в **1,15** (double) и **1,35** (float) раза на 1 ядре и в **1,11** и **1,30** раза на 8 ядрах.
- ❑ На **SpacemiT K3** выигрыш от векторизации есть только в многопоточной версии: на 8 ядрах векторный **HCSR** ускоряет в **1,05** и **1,17** раза, а на 1 ядре даже лучший векторный формат (**Sell-C-Sigma**) в среднем медленнее скалярного **CSR** (**0,74** и **0,84**).
- ❑ Проанализировано, какие особенности портрета матрицы влияют на эффективность **SpMV** значительнее всего.
- ❑ Реализация доступна на GitHub: <https://github.com/UNN-ITMM-Software/RVVLASparse>

\* Работа выполнена при поддержке программы академического лидерства ННГУ "Приоритет-2030".

