

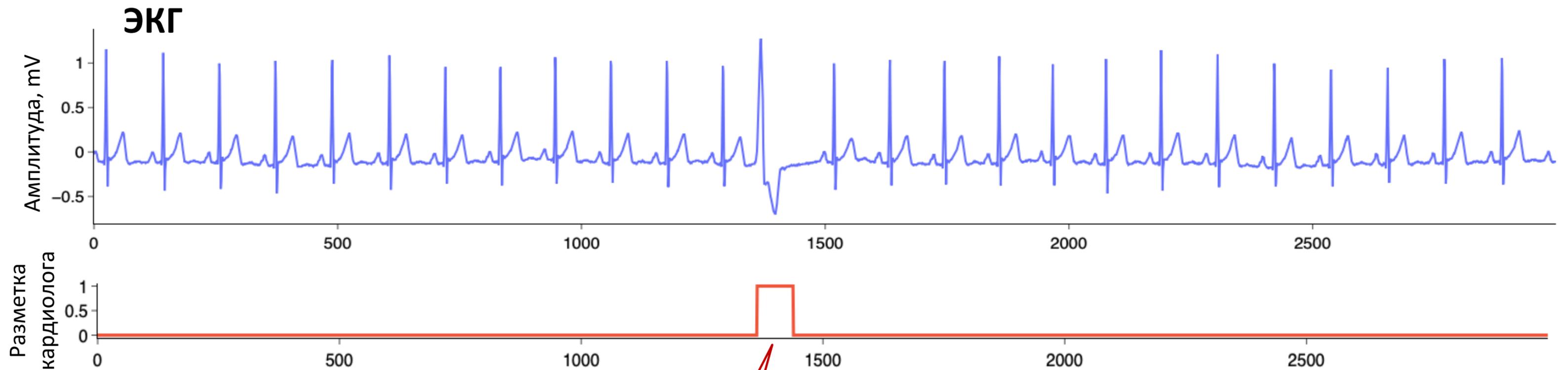
Международная научная конференция
Суперкомпьютерные дни в России 2023
Москва, 25–26 сентября 2023 г.

Поиск аномалий в больших временных рядах на кластере с GPU узлами

Я.А. Краева, М.Л. Цымблер

Южно-Уральский государственный университет (Челябинск)

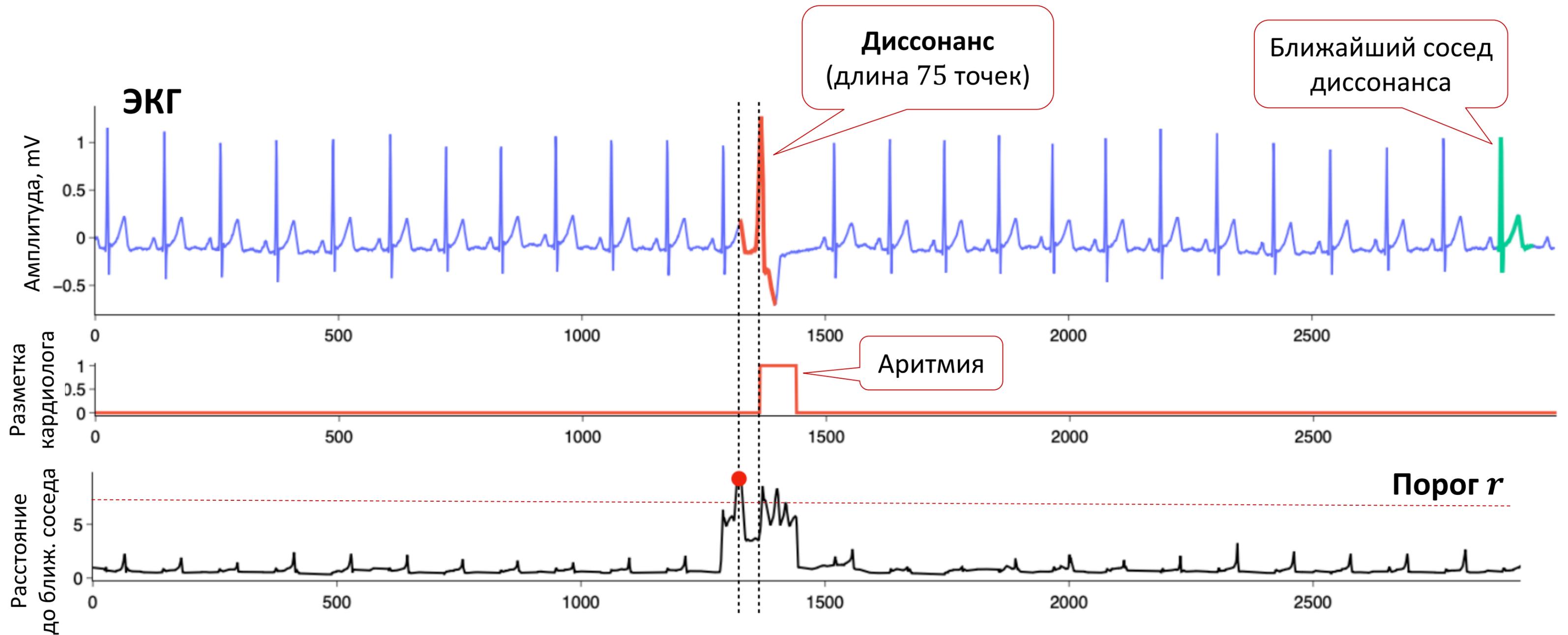
Диссонанс отражает аномалию ряда



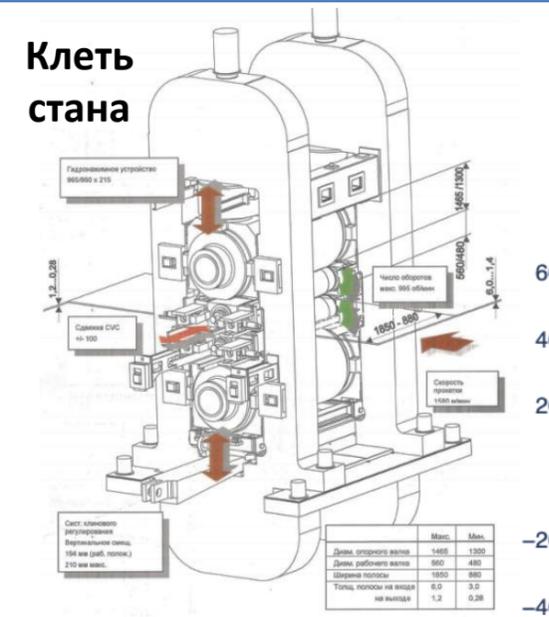
Аритмия



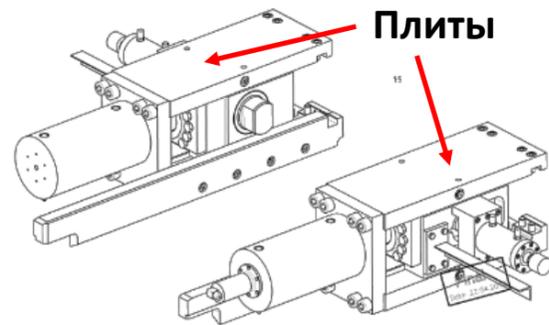
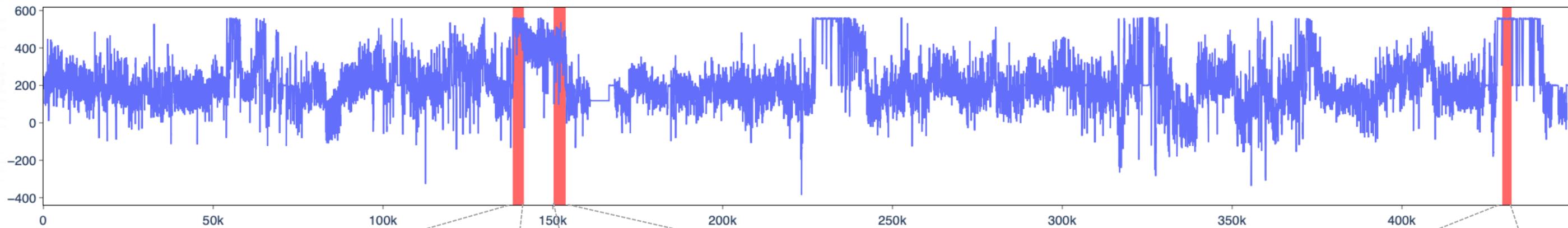
Диссонанс отражает аномалию ряда



Диссонансы всевозможных длин в больших временных рядах

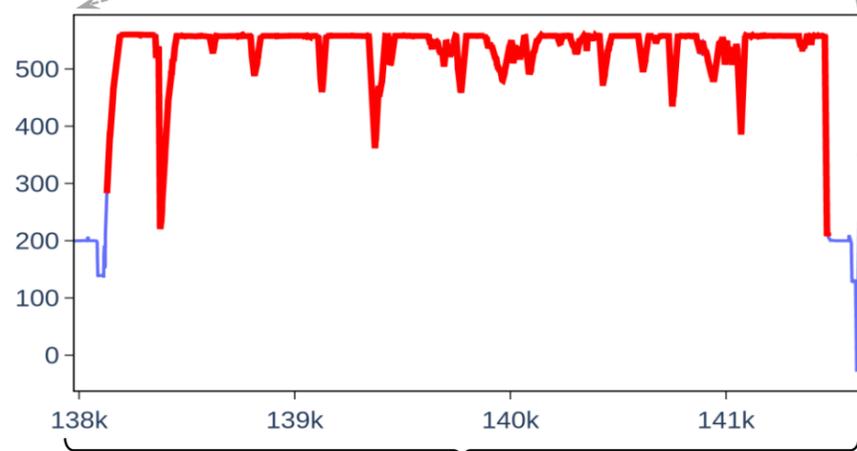


Временной ряд показаний тензOMETрического датчика (изгибающее усилие прокатки)

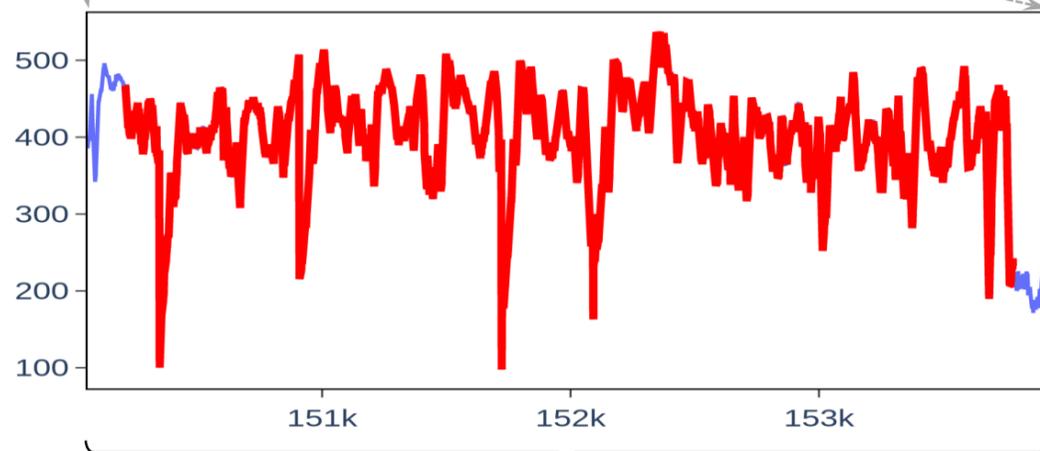


Система валков с непрерывно изменяемой кривизной

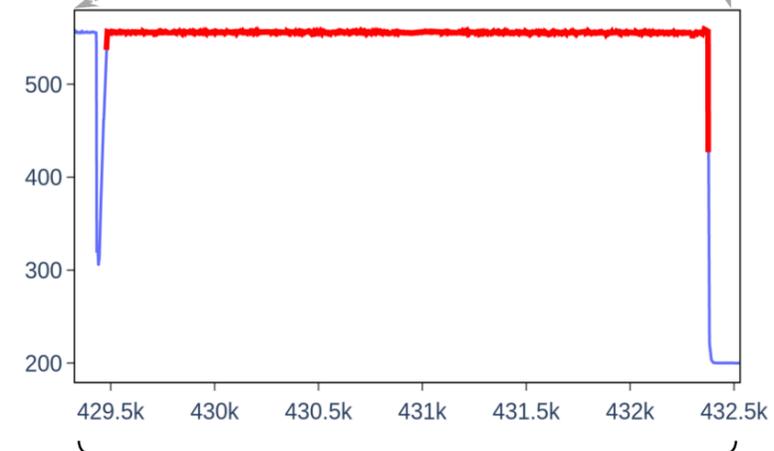
Диссонансы (критическое напряжение плиты)



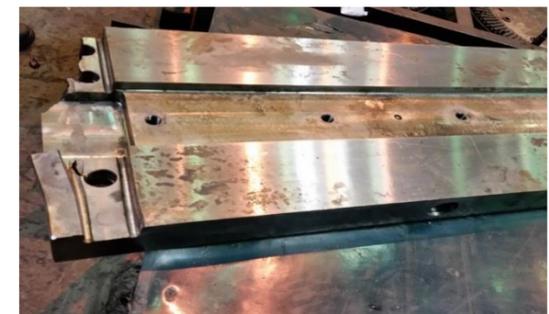
55 минут



1 час



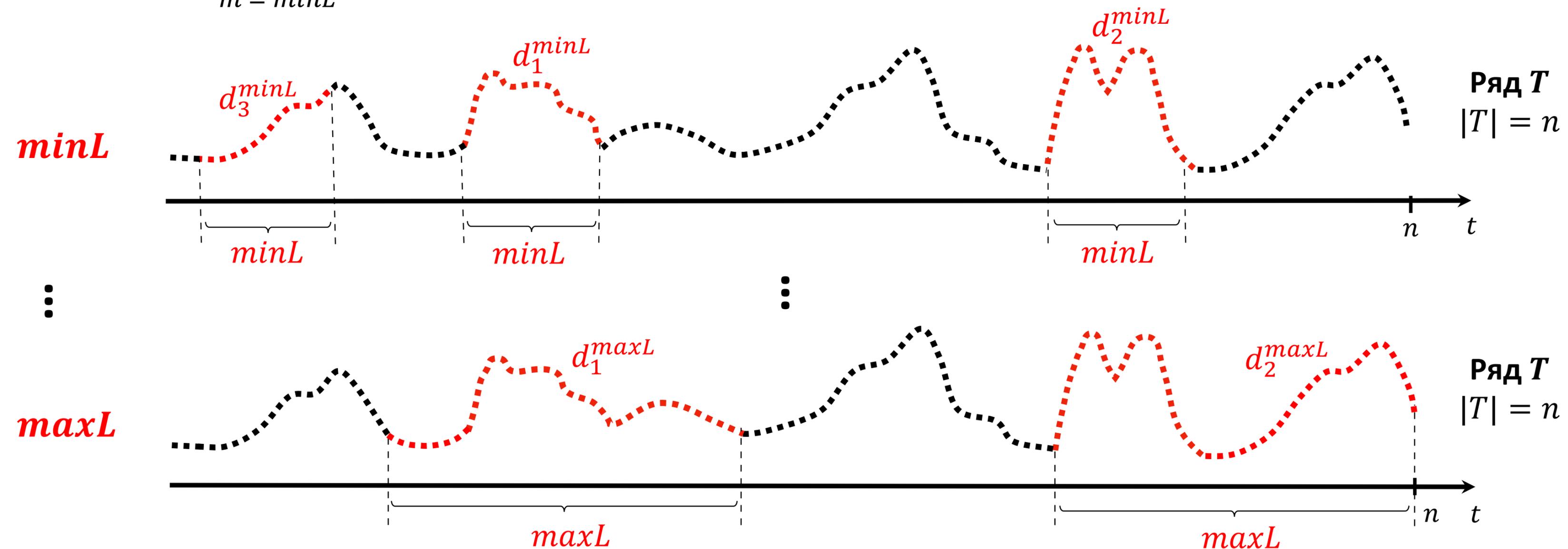
48 минут



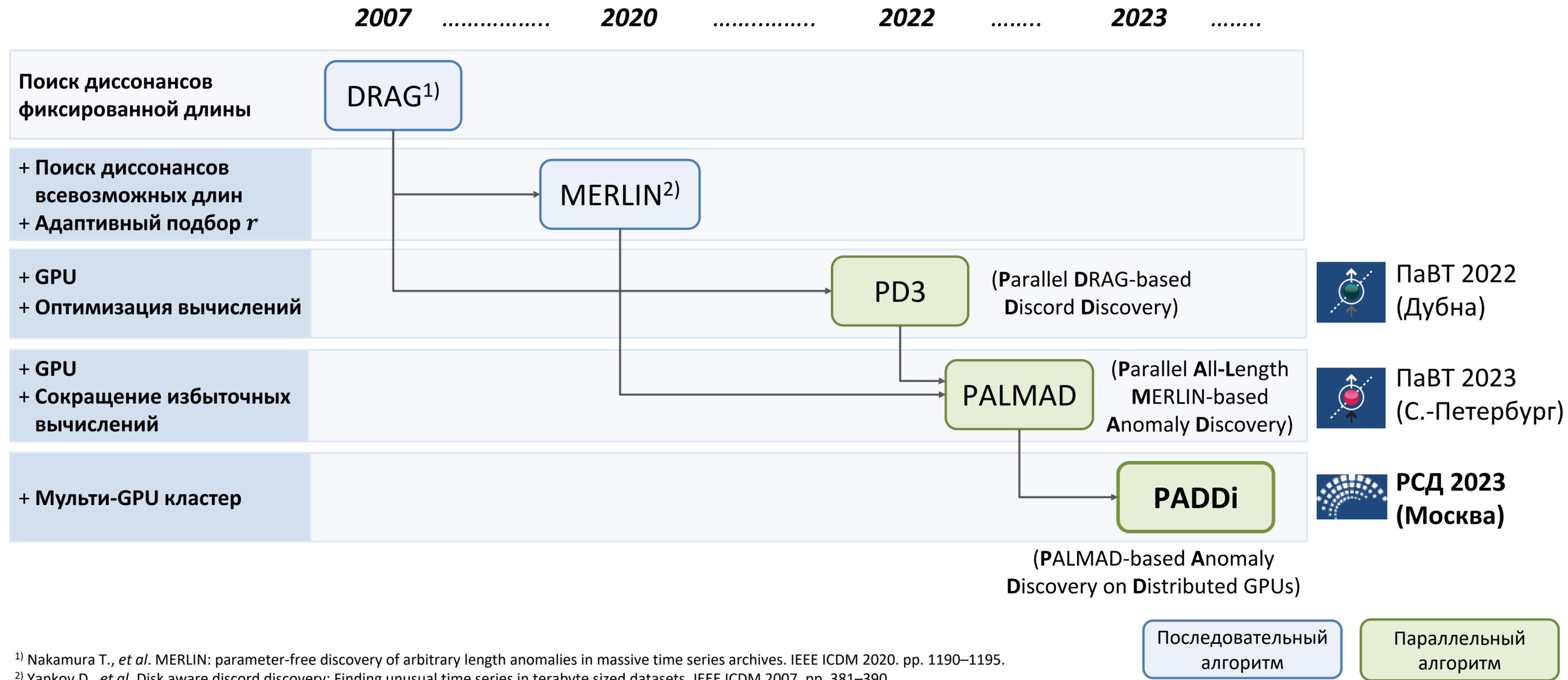
Разрушение плит

Задача поиска диссонансов всевозможных длин

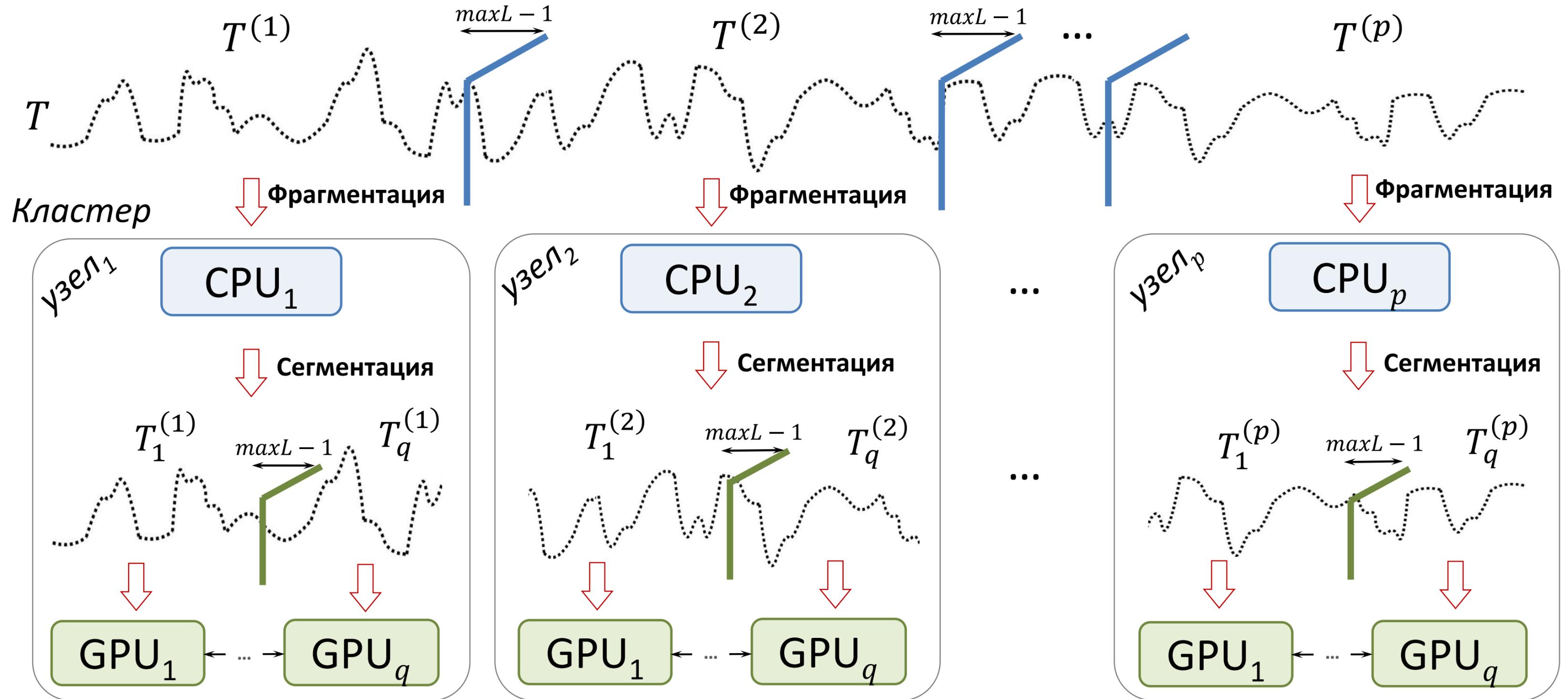
- **Диссонанс** – подпоследовательность ряда, отдаленная от ближайшего соседа не менее чем на порог r
- **Дано:** временной ряд T , диапазон длин диссонансов $minL, \dots, maxL$
- **Найти:** $\mathcal{D} = \bigcup_{m=minL}^{maxL} D_m$, $D_m = \{d_1^m, d_2^m, \dots\}$, d_i^m – диссонансы



Последовательные и параллельные алгоритмы поиска диссонансов



Фрагментация и сегментация ряда

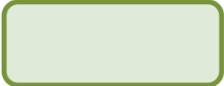


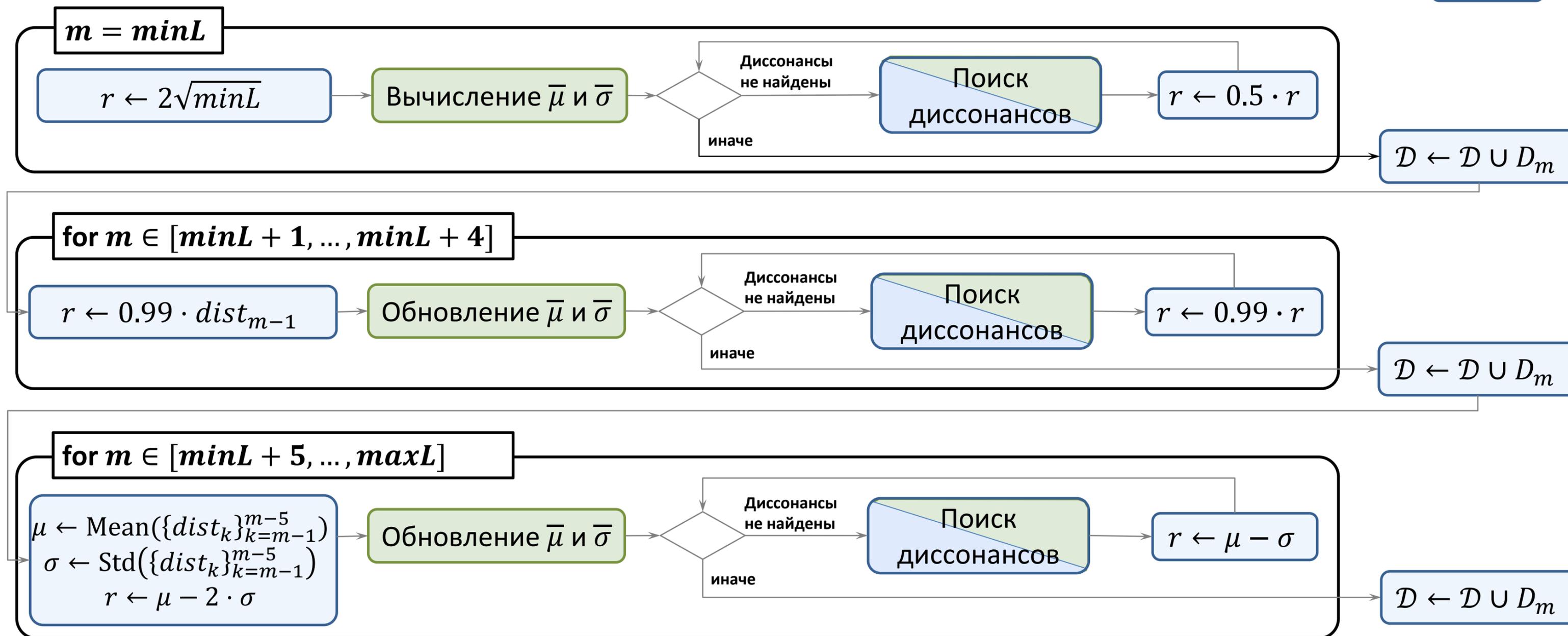
Общая схема вычислений

Инициализация порога r

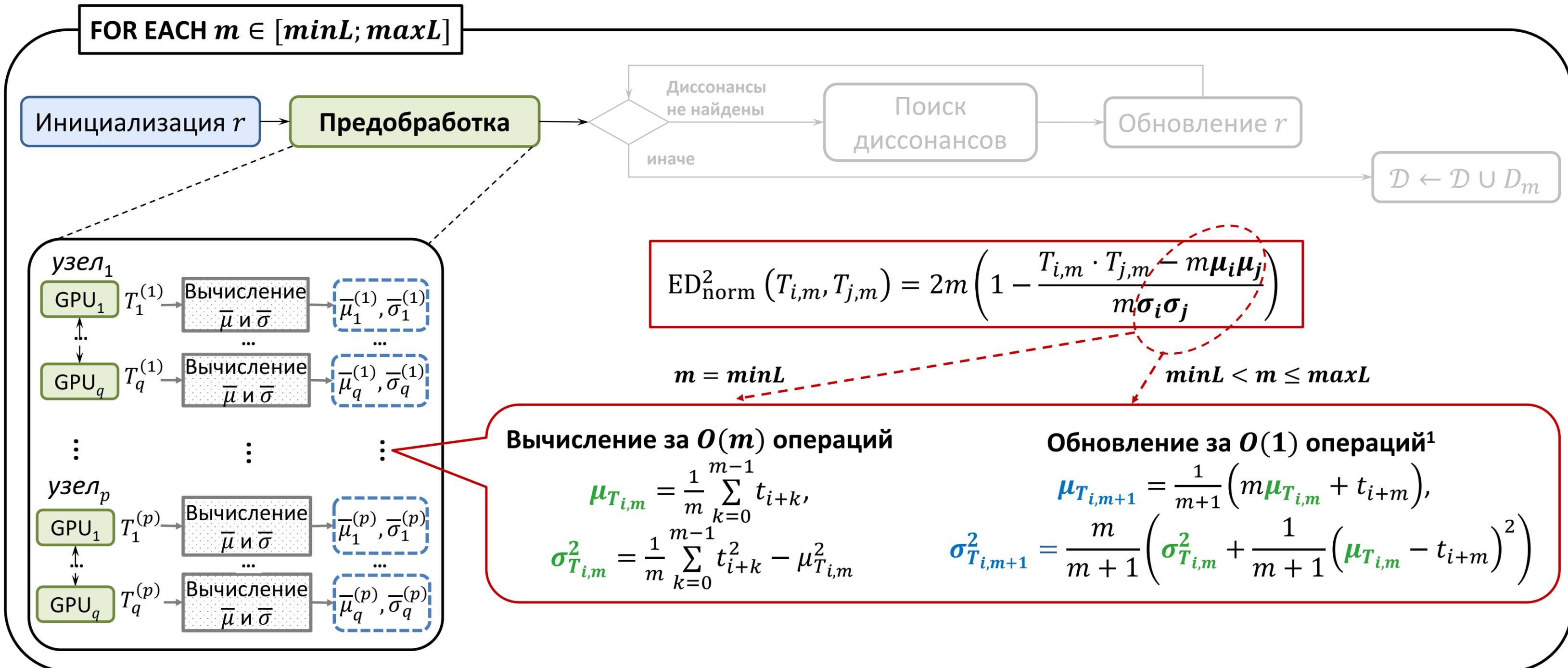
Предобработка

Подбор порога r

 – GPU
 – CPU



Предобработка



¹ Zymbler M., Kraeva Y. High-Performance Time Series Anomaly Discovery on Graphics Processors. Mathematics 11(14). Article 3193. 2023. DOI: 10.3390/math11143193 (Top10% WoS)

Поиск диссонансов

FOR EACH $m \in [minL; maxL]$

Инициализация r

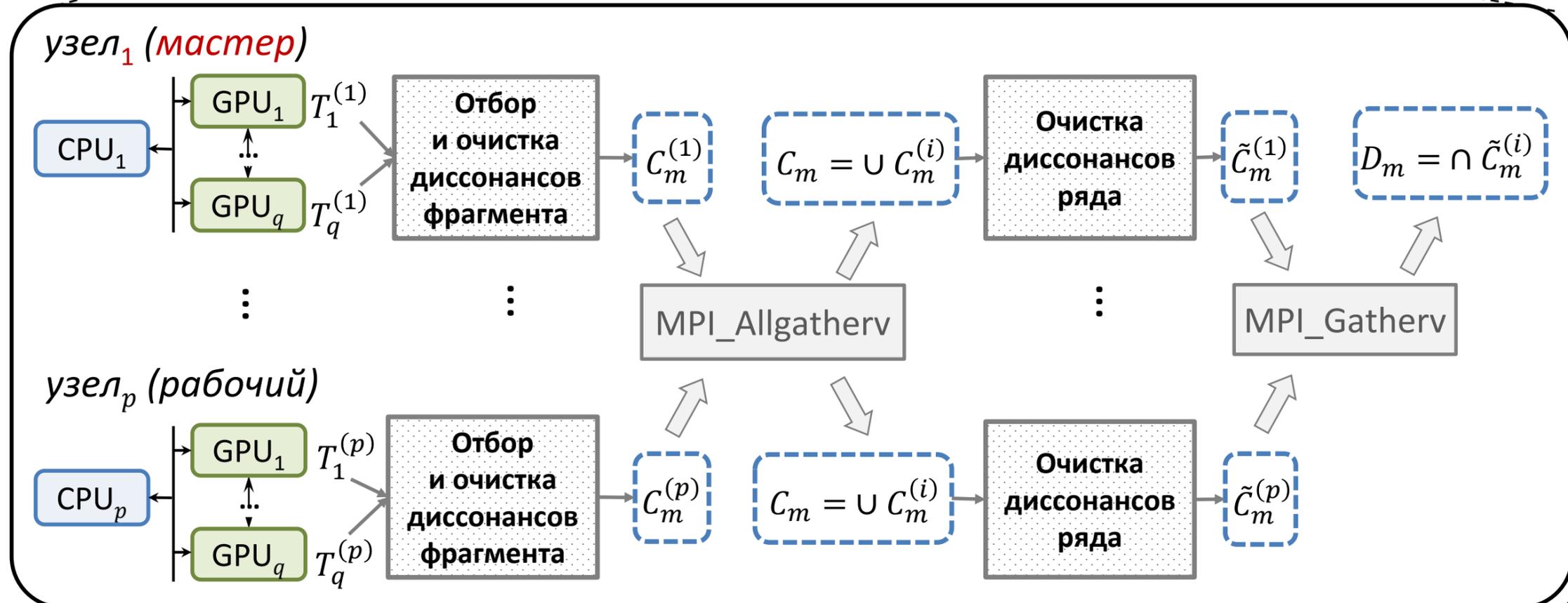
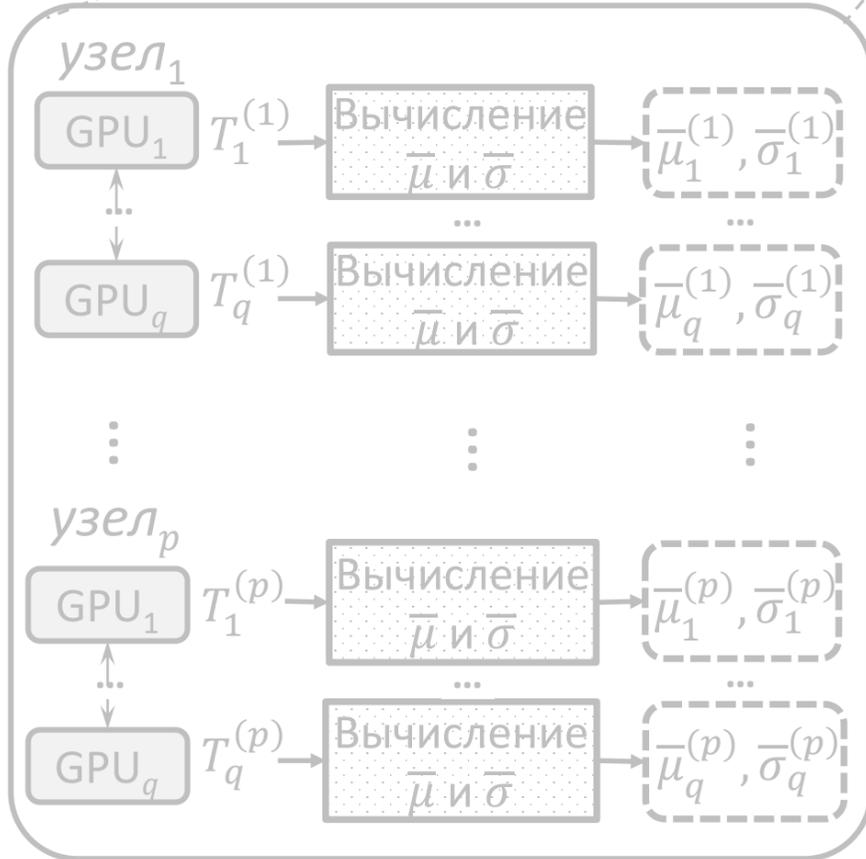
Предобработка

Диссонансы
не найдены
иначе

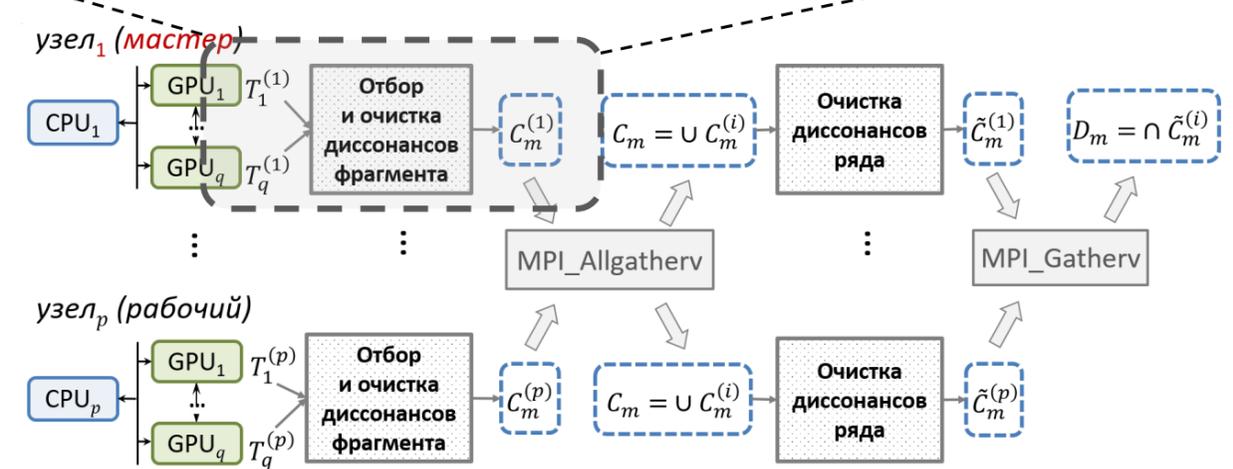
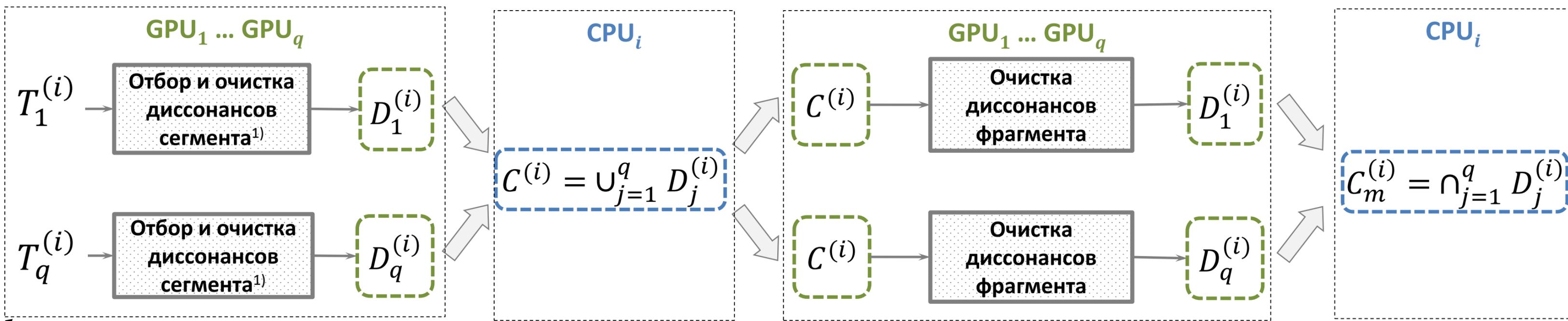
Поиск
диссонансов

Обновление r

$D \leftarrow D \cup D_m$

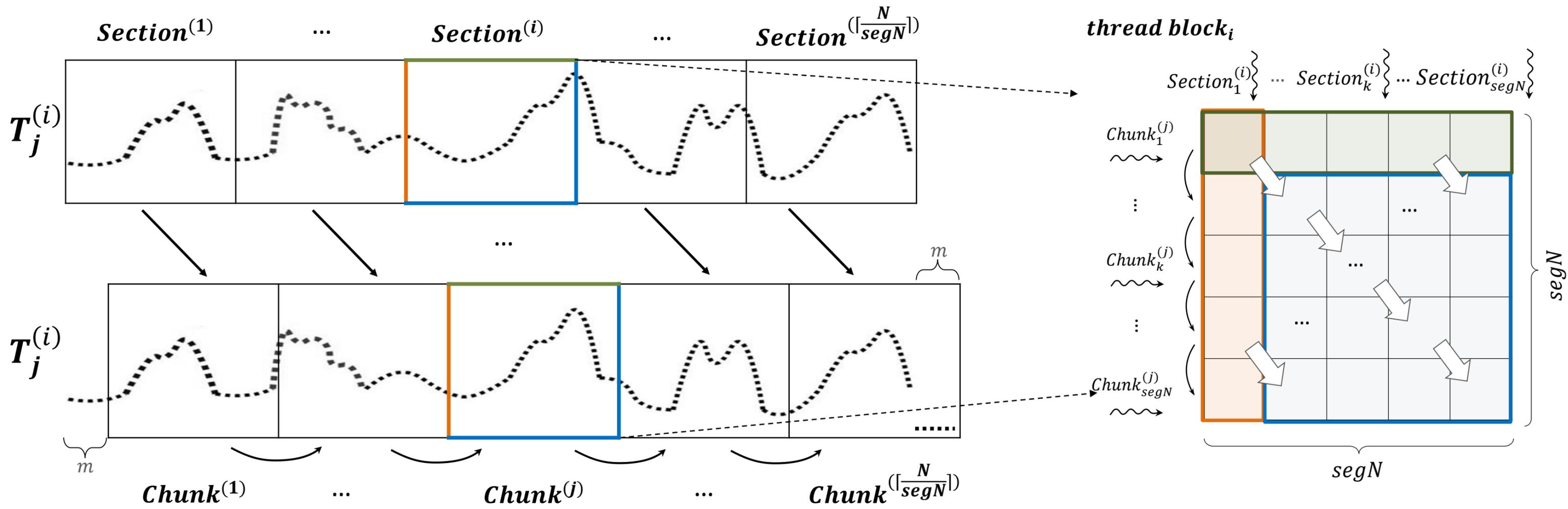


Отбор и очистка диссонансов фрагмента



¹⁾ Kraeva Y., Zymbler M. A Parallel Discord Discovery Algorithm for a Graphics Processor. Pattern Recognition and Image Analysis. 2023. Vol. 33, No. 2. P. 101–112. DOI: 10.1134/S1054661823020062.

Отбор и очистка диссонансов в сегменте



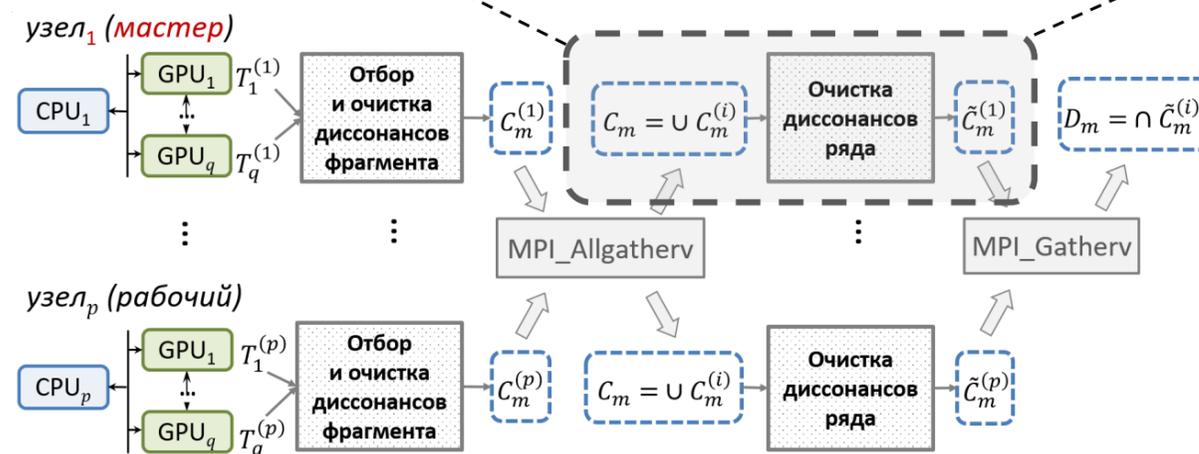
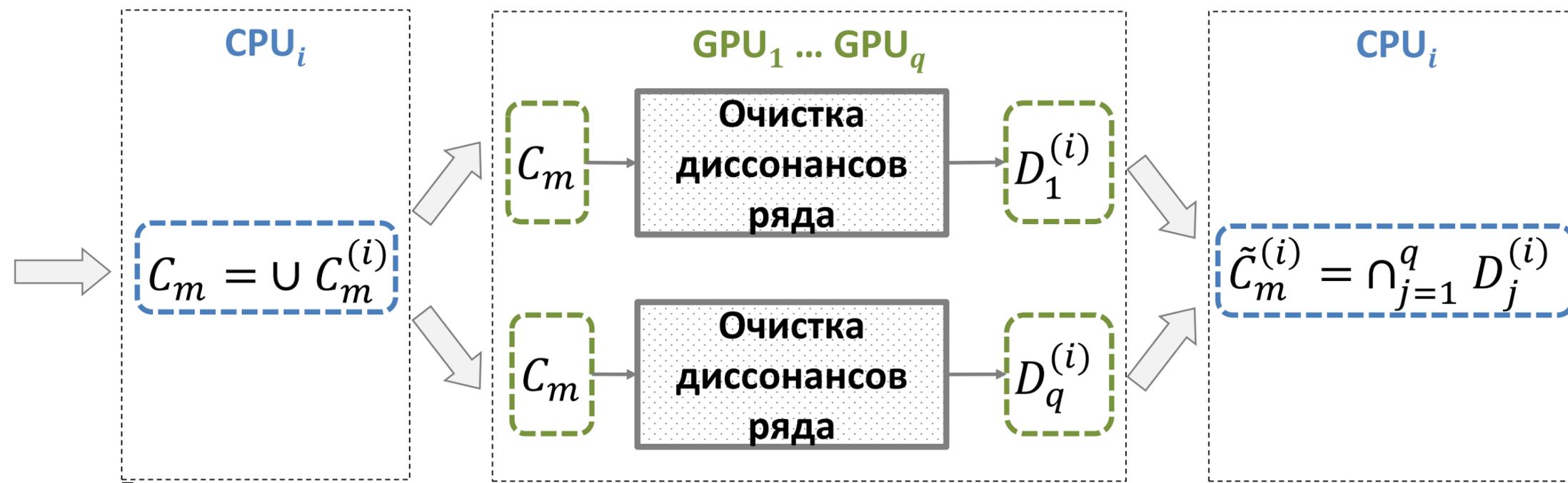
$$ED_{\text{norm}}^2(T_{i,m}, T_{j,m}) = 2m \left(1 - \frac{T_{i,m} \cdot T_{j,m} - m\mu_i\mu_j}{m\sigma_i\sigma_j} \right)$$

$$\text{orange box } QTrow^{(i)}(tid) = \sum_{k=1}^m T_{tid}^{(i)}(k) \cdot Chunk_1^{(j)}(k) \quad \text{green box } QTcol^{(i)}(tid) = \sum_{k=1}^m T_1^{(i)}(k) \cdot Chunk_{tid}^{(j)}(k)$$

$O(1)$ вместо $O(m)$!

$$\text{blue box } QTrow^{(i)}(tid) = QTrow^{(i)}(tid - 1) - T_{tid-1}^{(i)}(1) \cdot Chunk_{tid-1}^{(j)}(1) + T_{tid}^{(i)}(m) \cdot Chunk_{tid}^{(j)}(m)$$

Очистка диссонансов ряда

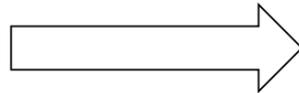
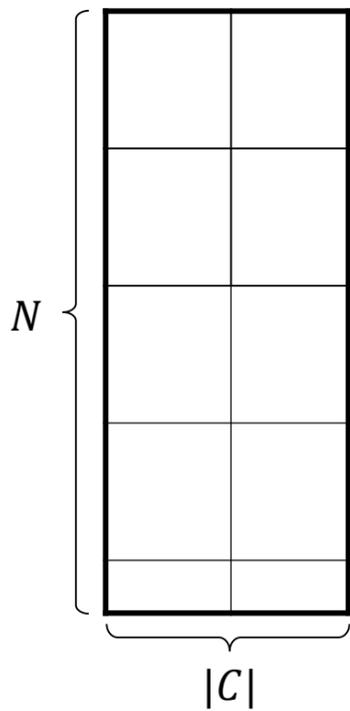


Очистка диссонансов

Матрица расстояний

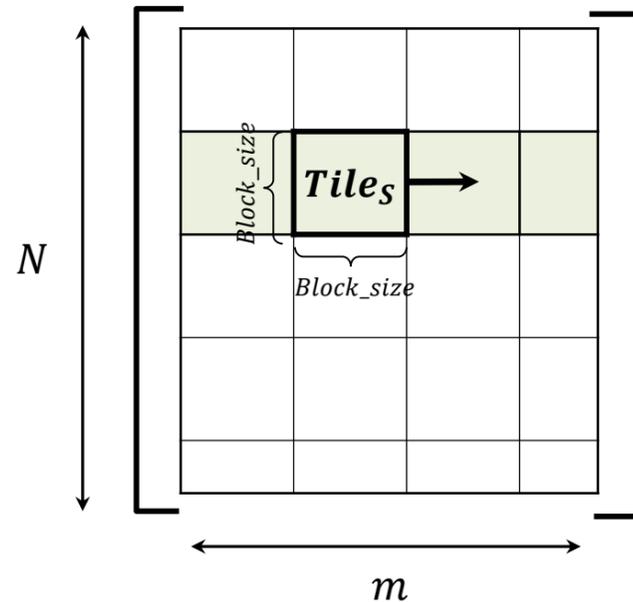
$$Dist \in \mathbb{R}^{N \times |C|}$$

$$Dist_{i,j} = 2m \left(1 - \frac{T_{i,m} \cdot T_{j,m} - m\mu_i\mu_j}{m\sigma_i\sigma_j} \right)$$



Подпоследовательности сегмента

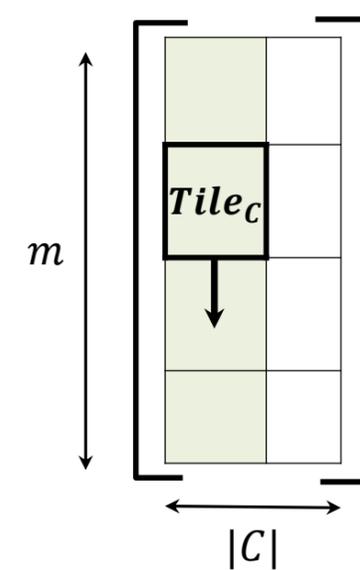
$$S \in \mathbb{R}^{N \times m}$$



\times

Кандидаты в диссонансы

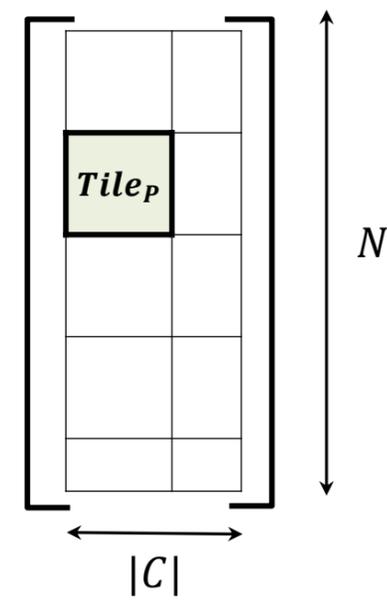
$$C^T \in \mathbb{R}^{|C| \times m}$$



$=$

Матрица скалярных произведений

$$P \in \mathbb{R}^{N \times |C|}$$

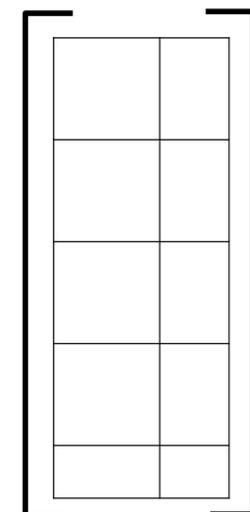


Битовая карта диссонансов



Отбрасывание кандидата C_j , если $\min_{1 < i < N} (Dist_{i,j}) < r^2$

$$2m \left(1 - \frac{P - m\mu_i\mu_j}{m\sigma_i\sigma_j} \right) =$$



Матрица расстояний $Dist \in \mathbb{R}^{N \times |C|}$

Эксперименты

- **Аппаратные платформы**

- Лобачевский, ННГУ

- 1 узел: 2xSandy Bridge E5-2660 с 3xNVIDIA Kepler K20X (2 688 CUDA-ядер, @0.732 GHz, 1.3 TFLOPS)

- Ломоносов-2, МГУ

- 1 узел: 1xXeon Gold 6126 с 2xNVIDIA Tesla P100 (3 584 CUDA-ядер, @1.19 GHz, 4 TFLOPS)

- **Данные**

Ряд	Длина ряда, n	Диапазон длин диссонансов, $minL..maxL$	Предметная область
ECG ¹⁾	$2 \cdot 10^6$	64..128	ЭКГ взрослого человека
GAP ²⁾			Энергопотребление частного дома во Франции
MGAB ³⁾			Синтетический ряд на основе уравнения Макки–Гласса

- **Показатели**

- Быстродействие: усредненное время работы алгоритма по 10 запускам без I/O

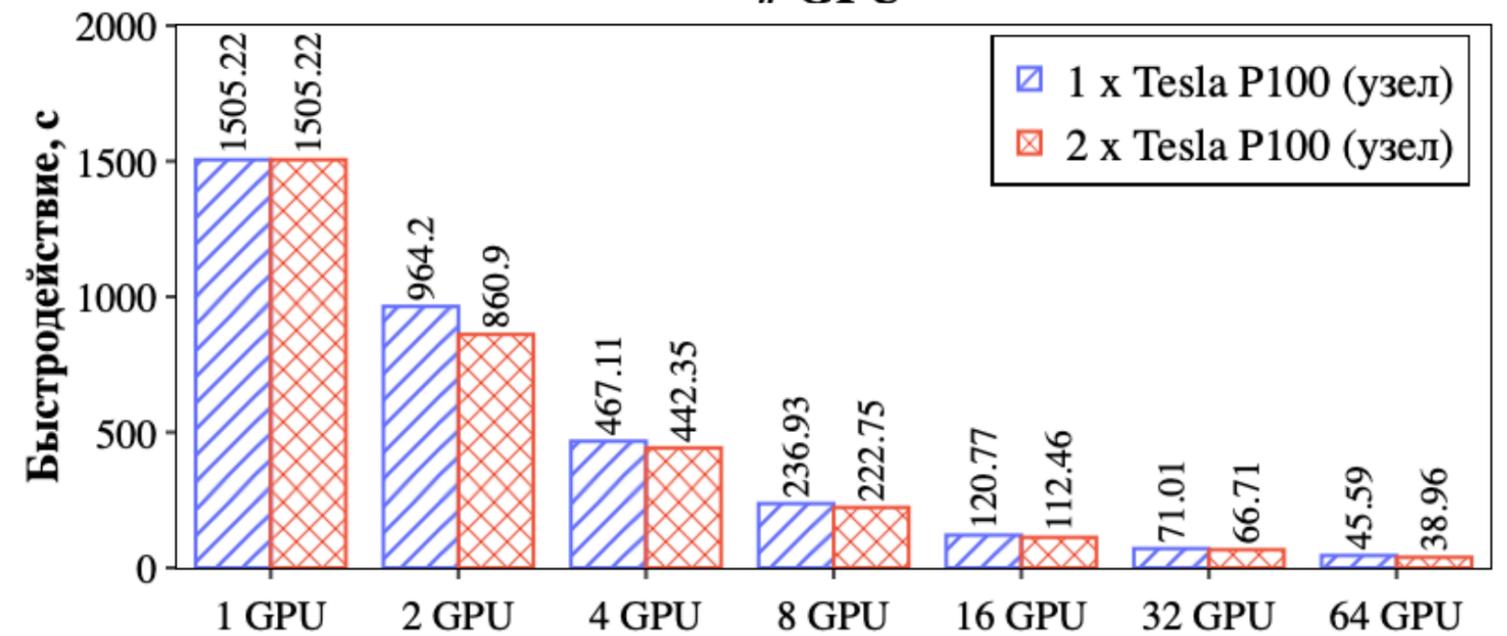
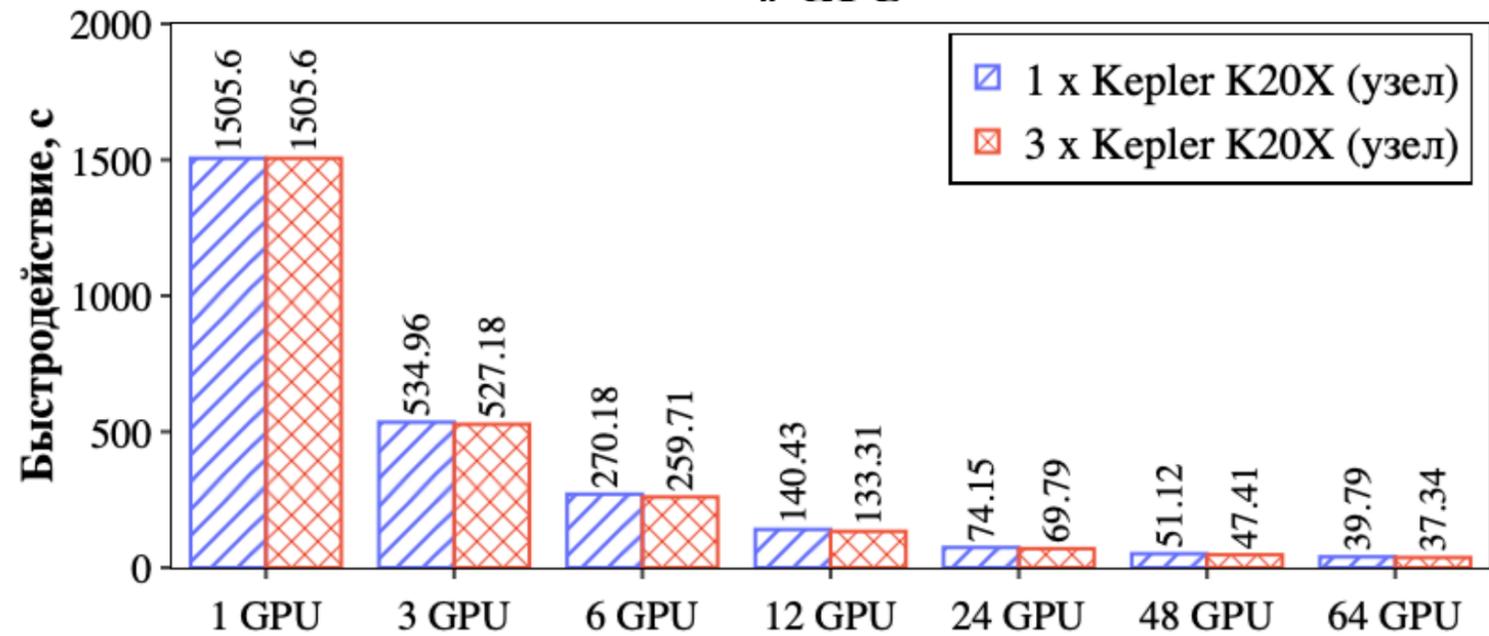
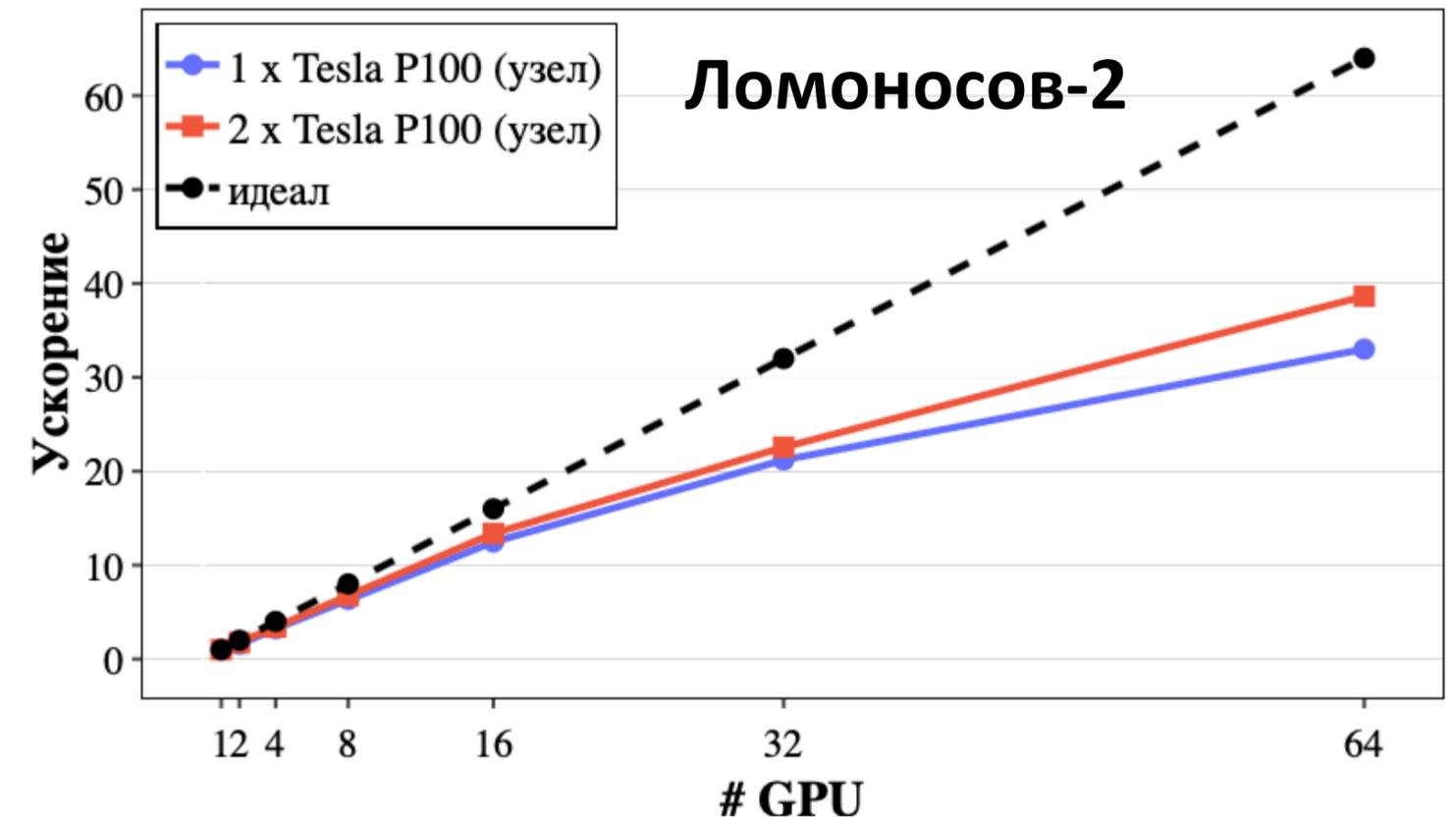
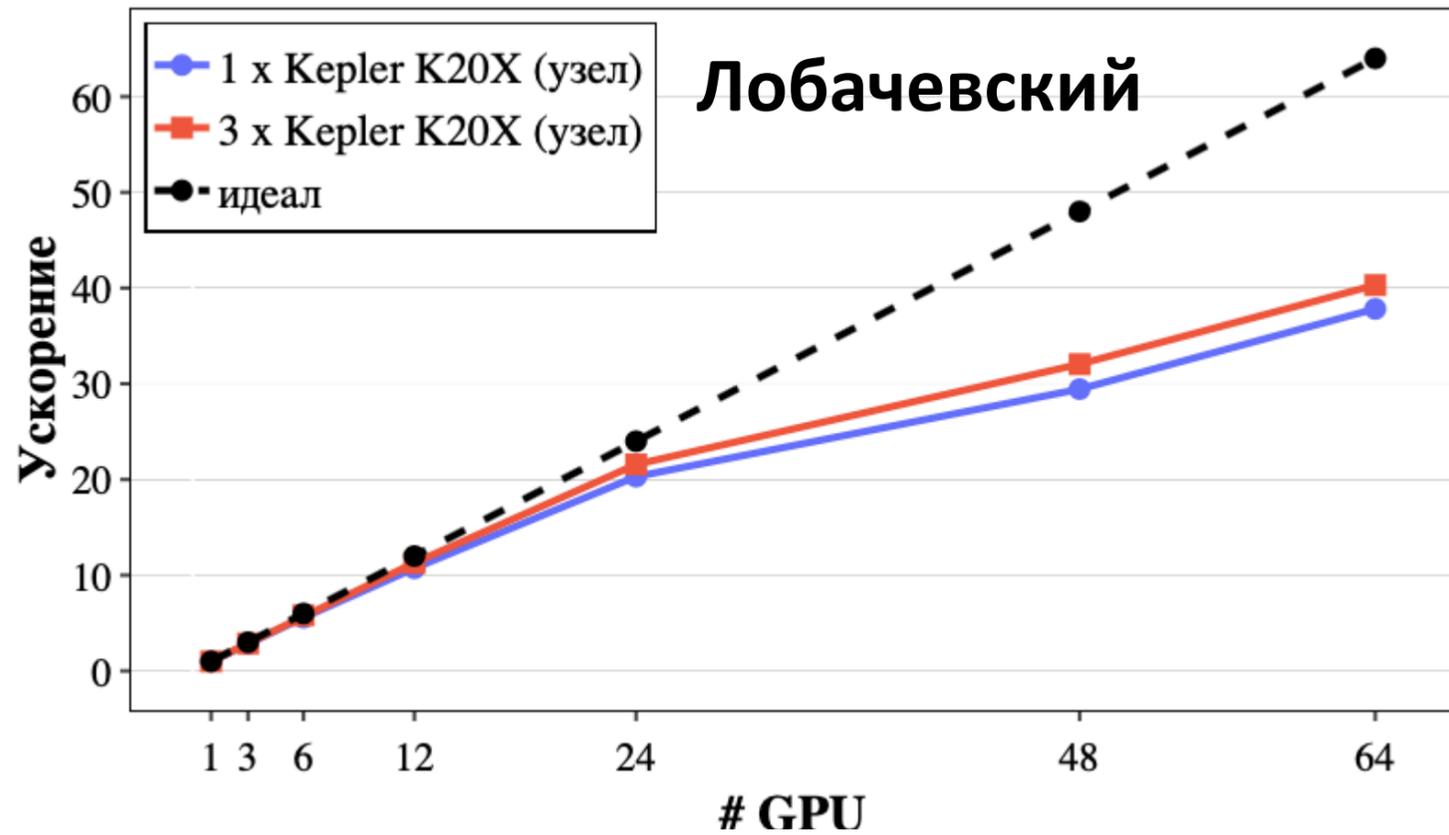
- Ускорение: $s(p) = t_1/t_p$, где t_1 – время на одном GPU, t_p – время на p GPU

¹⁾ Lu Y. et al. Matrix Profile XXIV: Scaling Time Series Anomaly Detection to Trillions of Datapoints and Ultra-fast Arriving Data Streams. ACM SIGKDD 2022. pp. 1173-1182.

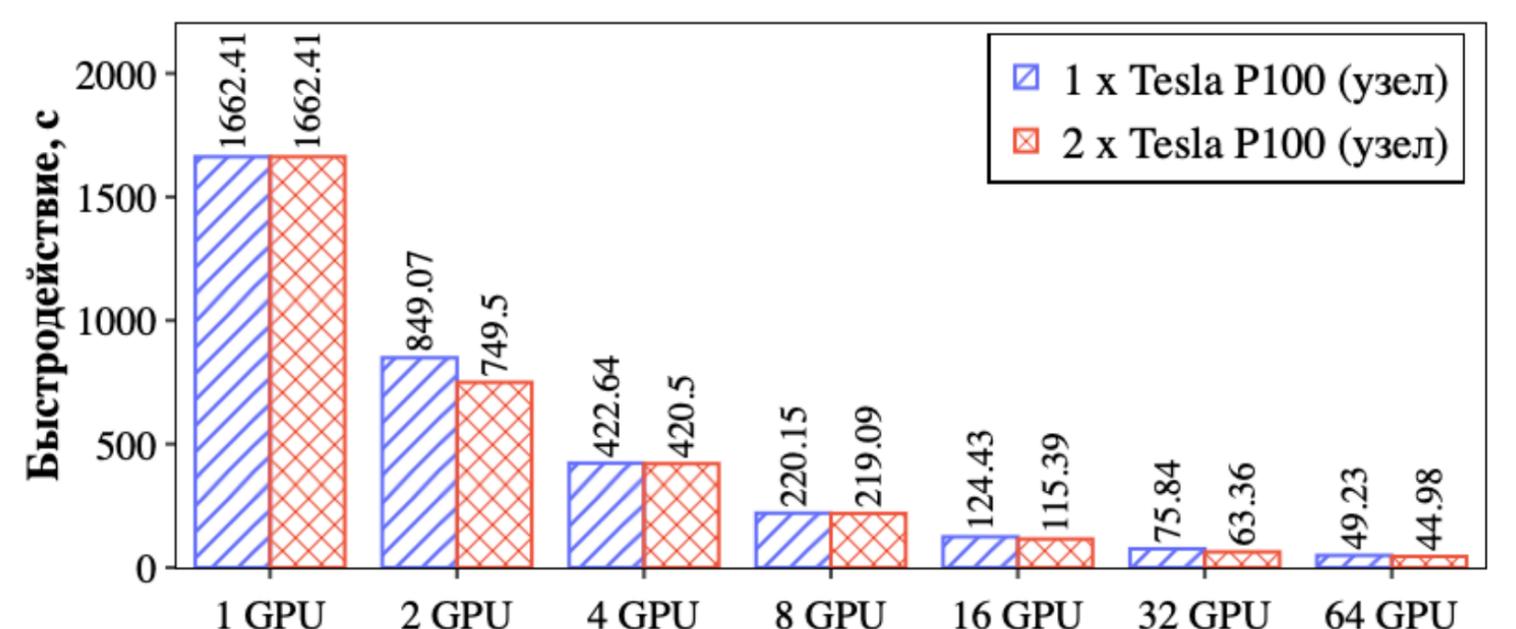
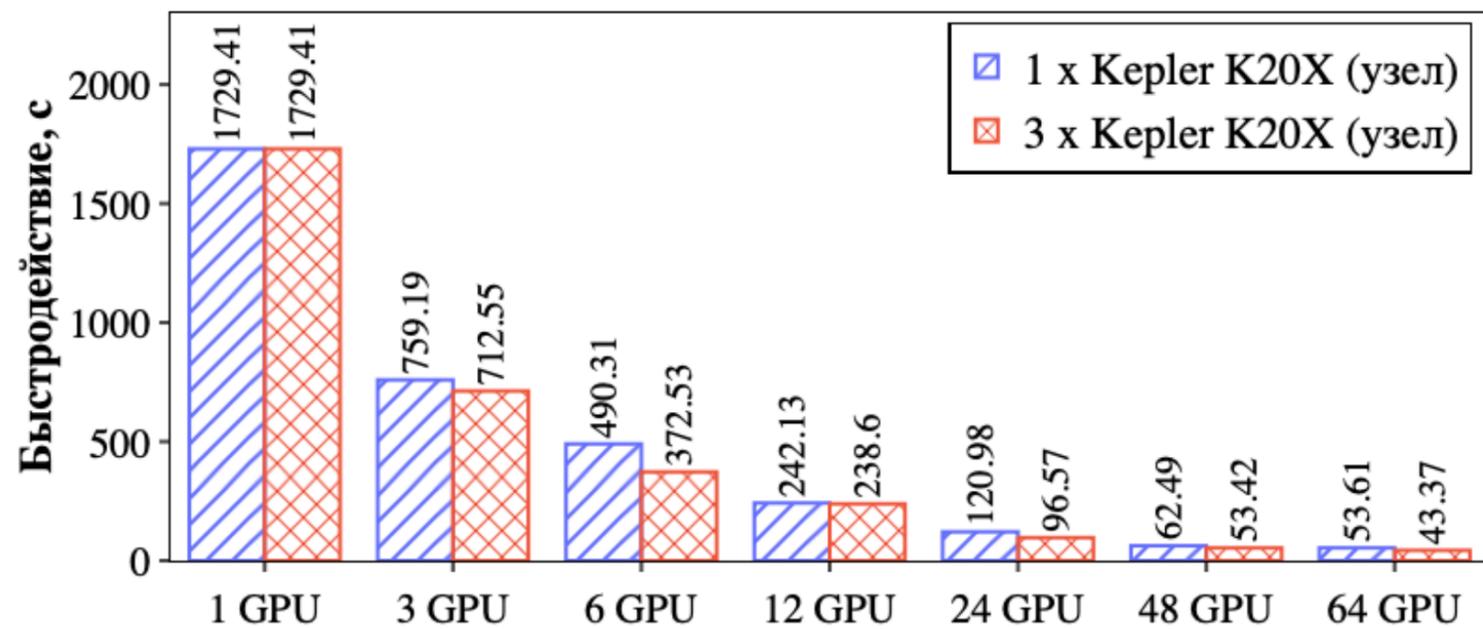
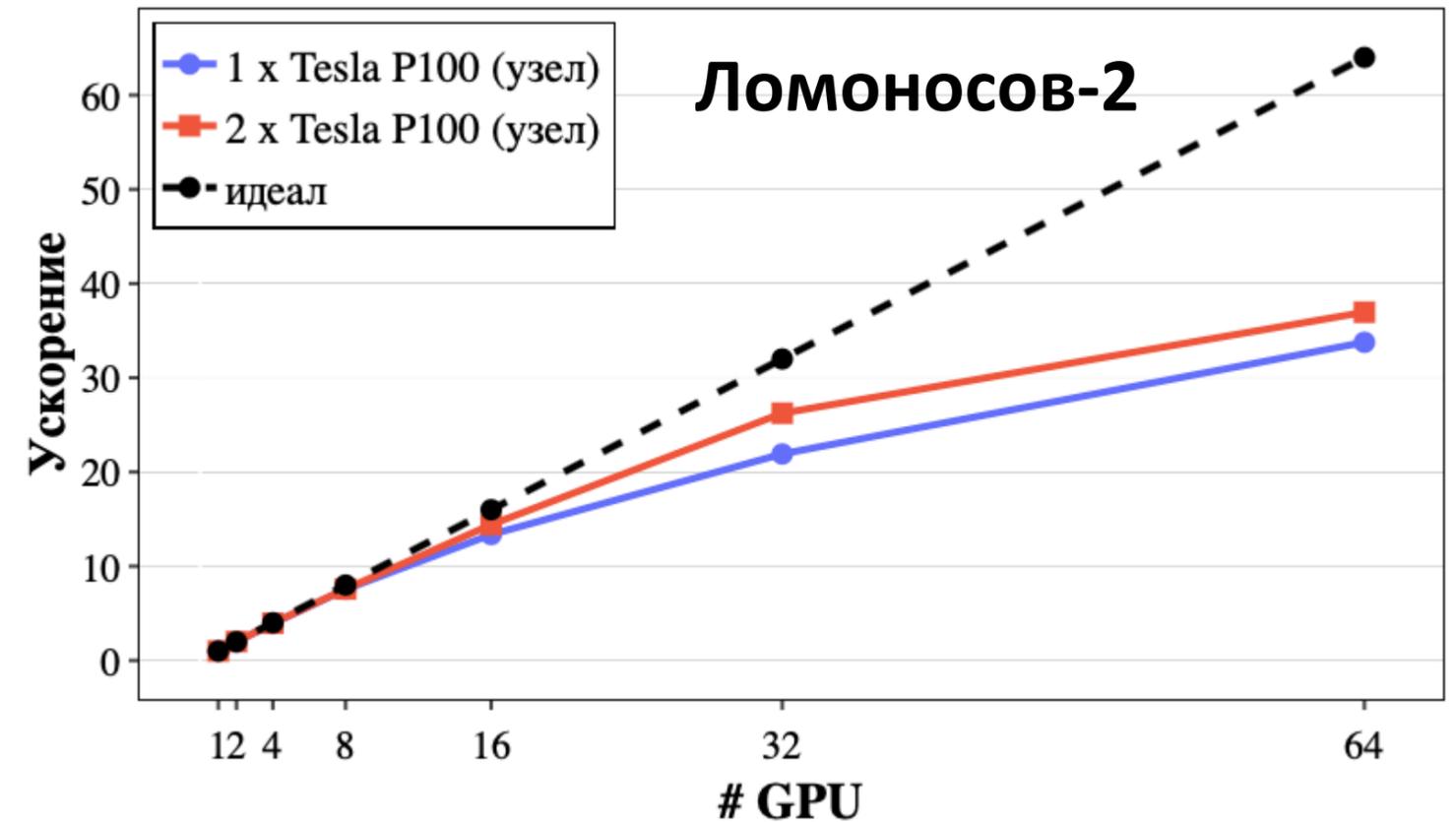
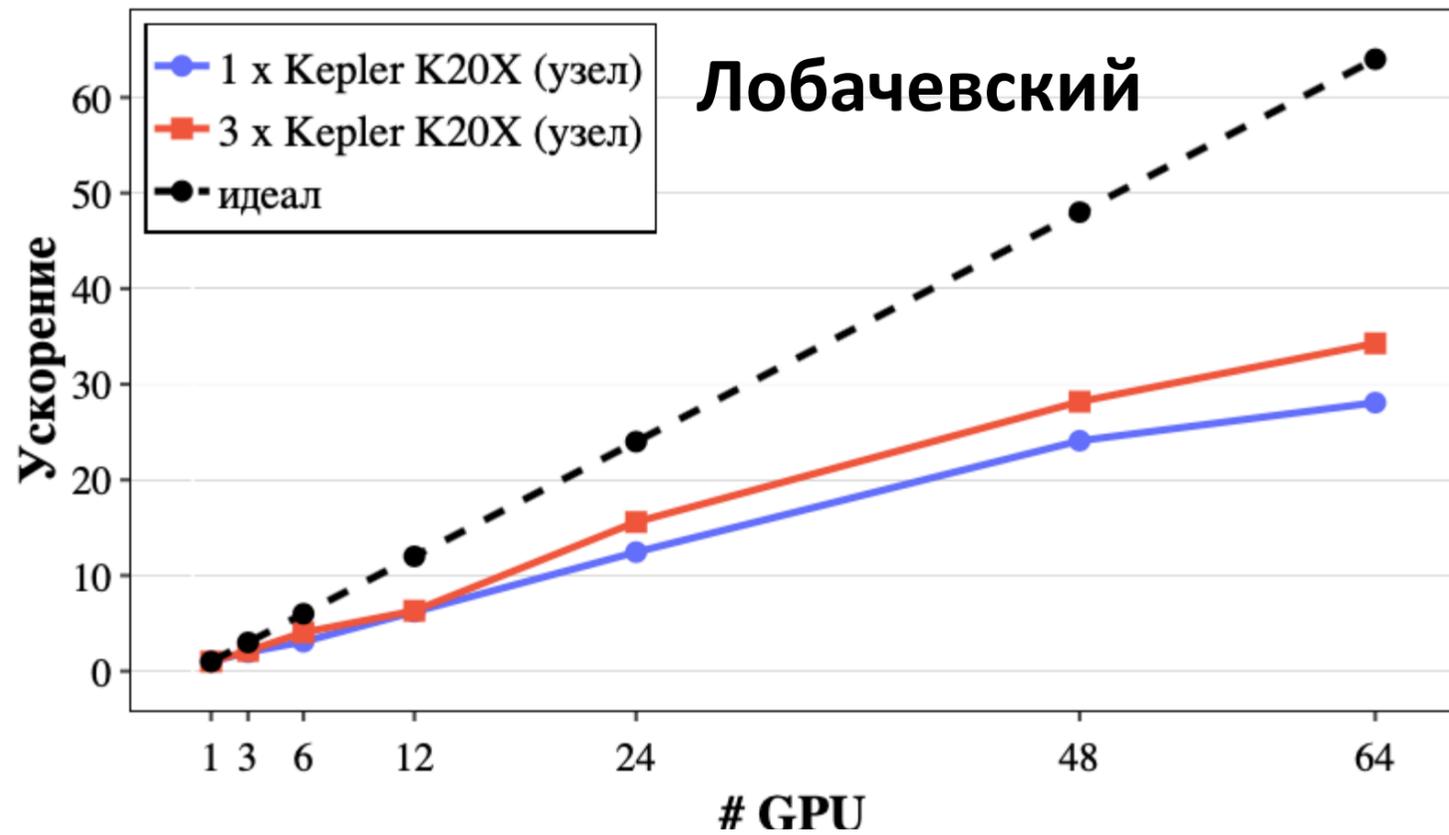
²⁾ Linardi M. et al. Matrix Profile X: VALMOD - Scalable Discovery of Variable-Length Motifs in Data Series. SIGMOD 2018. pp. 1053-1066.

³⁾ Mackey M.C. et al. Oscillation and chaos in physiological control systems. Science 197 (4300), 287–289 (1977).

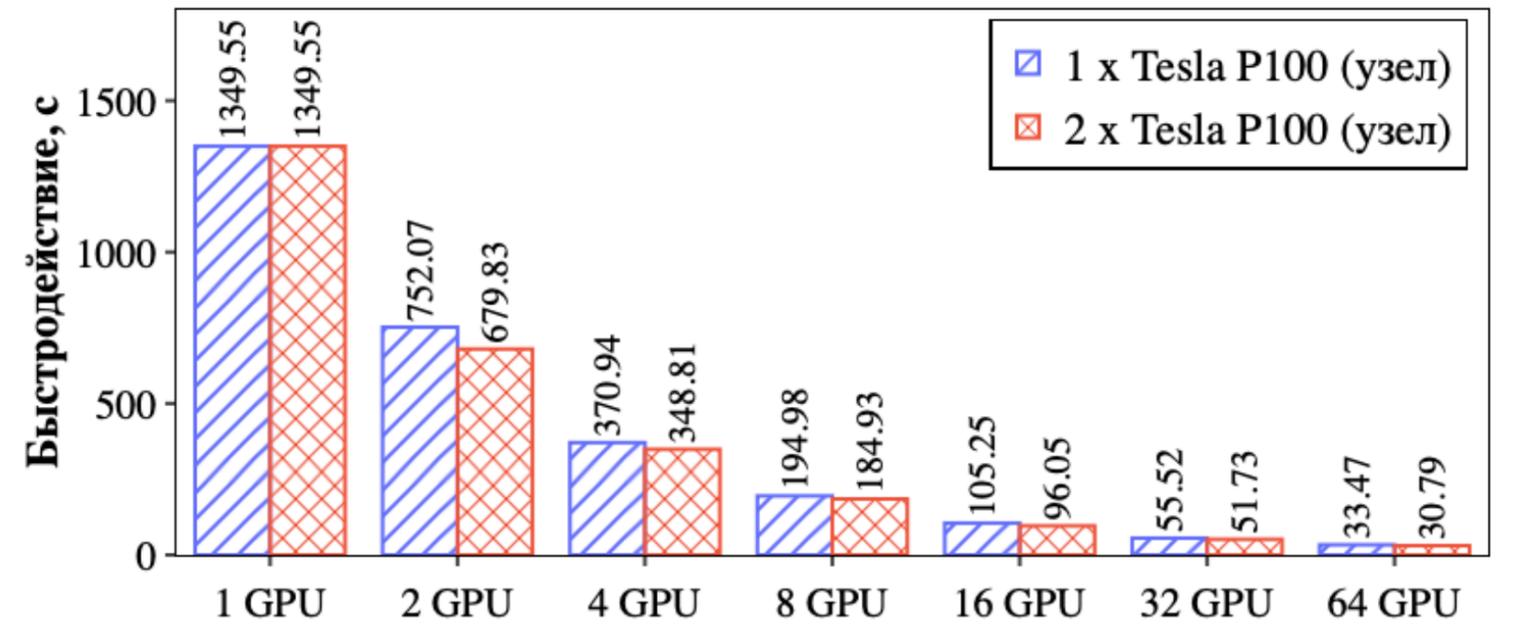
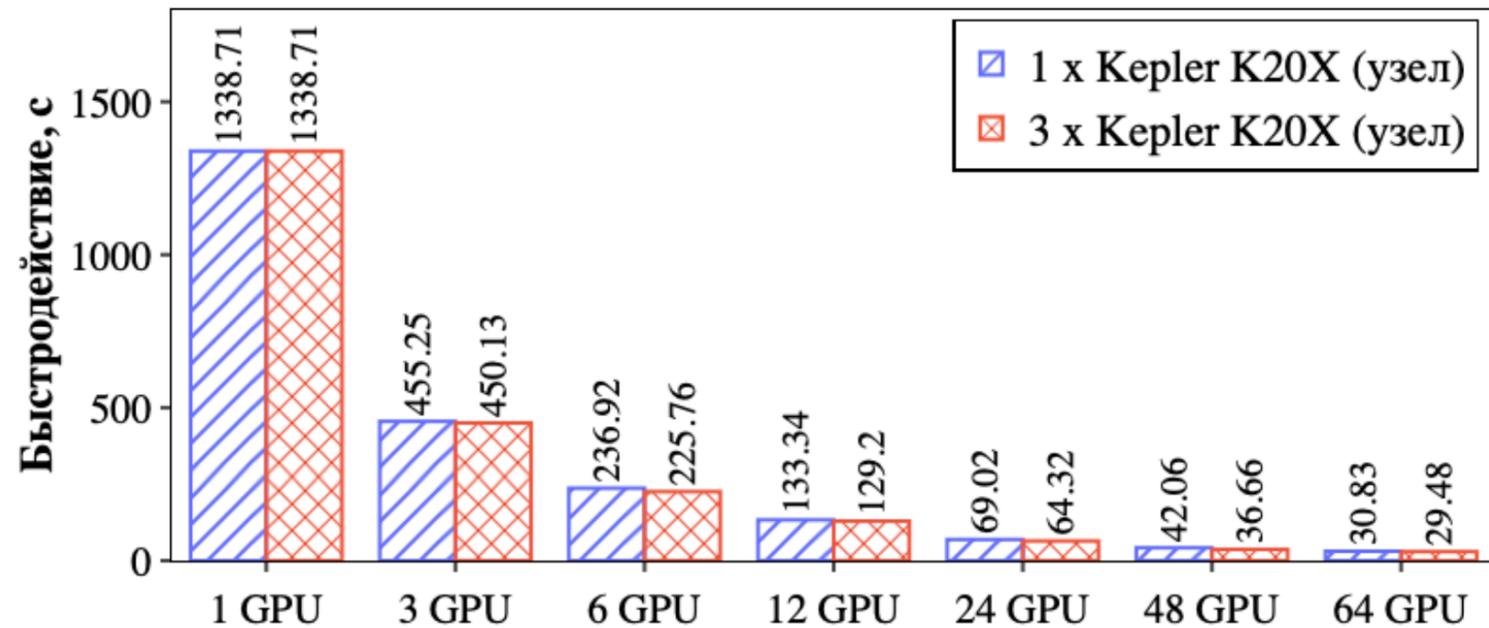
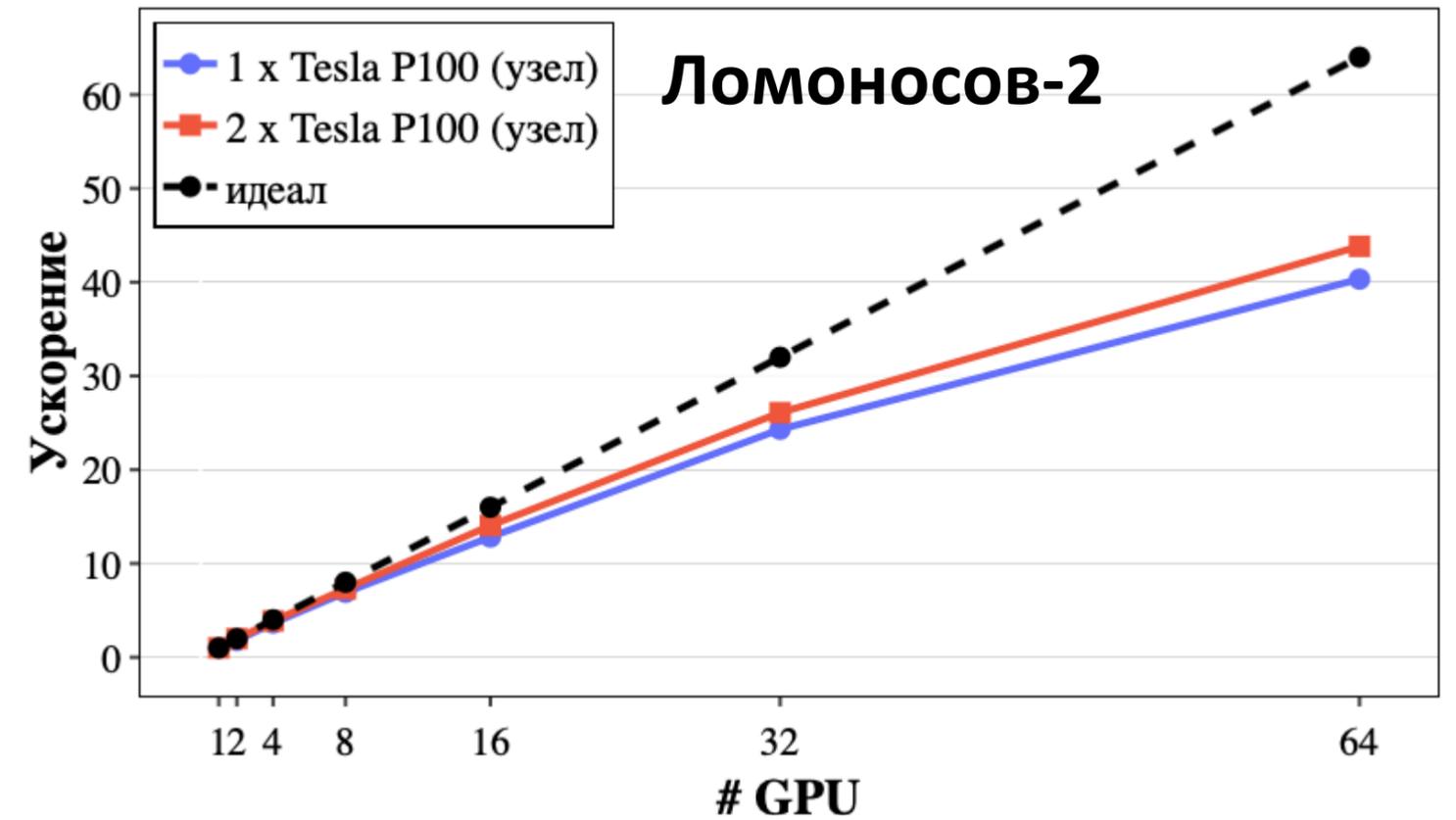
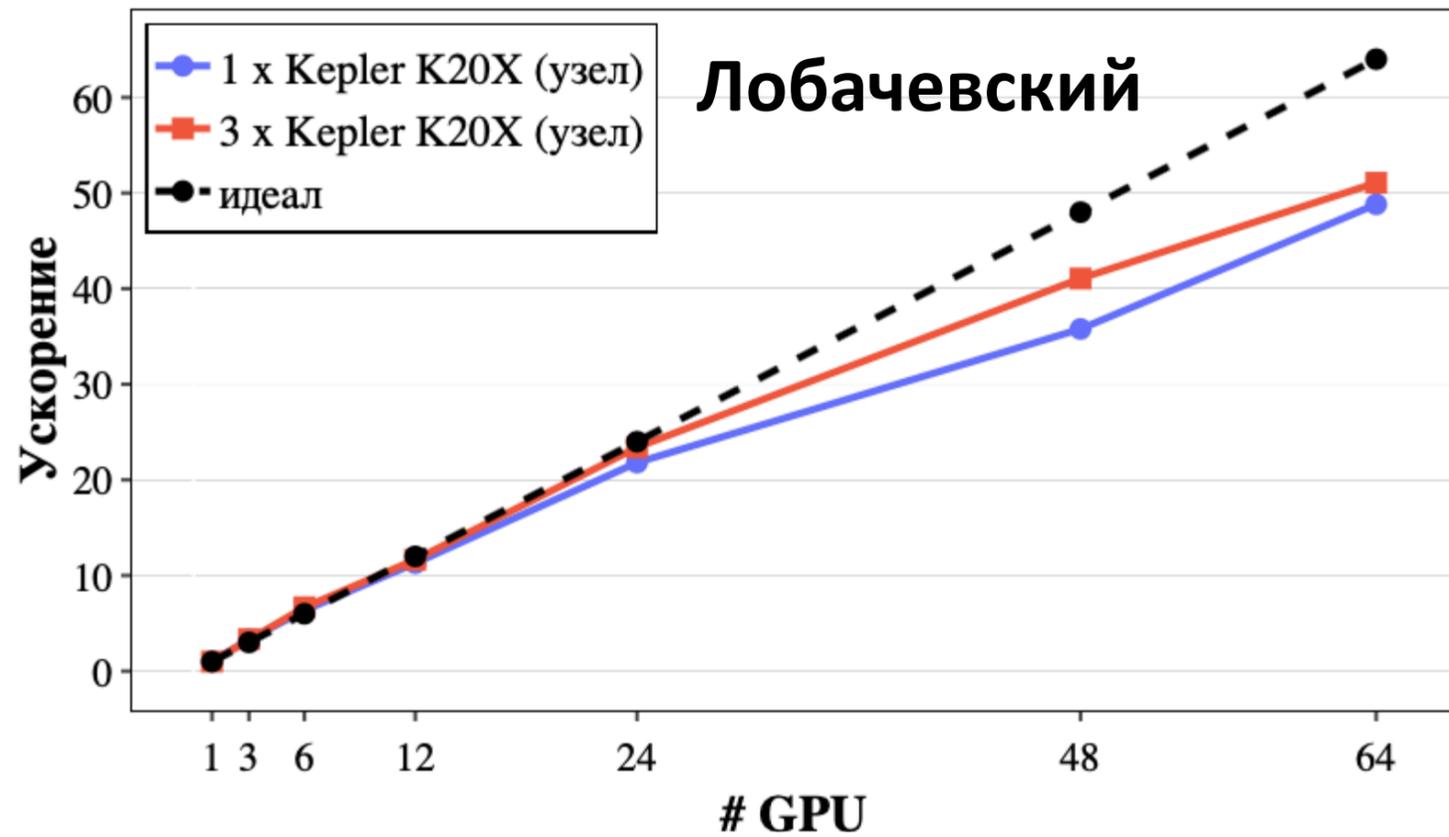
Ускорение и быстродействие: ряд ESG



Ускорение и быстродействие: ряд GPR



Ускорение и быстродействие: ряд MGAB



Заключение

- Предложен новый параллельный алгоритм поиска диссонансов в больших временных рядах на вычислительном кластере с GPU-узлами
- Проведены вычислительные эксперименты на реальных и синтетических рядах, показывающие высокую масштабируемость алгоритма
- Будущие исследования: применение алгоритма в нейросетевой модели для поиска аномалий в режиме онлайн

Спасибо за внимание! Вопросы?

Яна Александровна Краева

kraevaya@susu.ru