



Параллельный алгоритм вычисления матричного профиля потокowego временного ряда на системе с несколькими графическими процессорами

А.А. Валиулин, М.Л. Цымблер

{valiulinaa, mzym}@susu.ru

Южно-Уральский государственный университет, Челябинск, Россия

Применение потоковых временных рядов



**Энергосистемы
и «умные» сети**



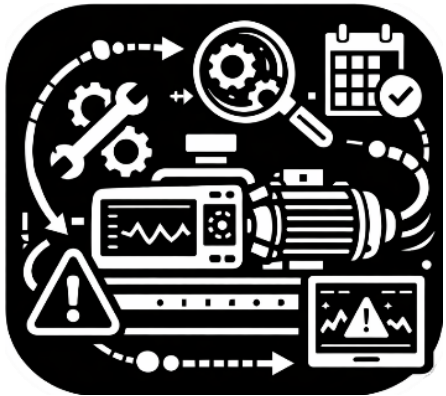
**Экологический
мониторинг**



**Предсказание
катаклизмов**



**Персональное
здравоохранение**



**Предиктивное
техобслуживание**



**Кибер-
безопасность**



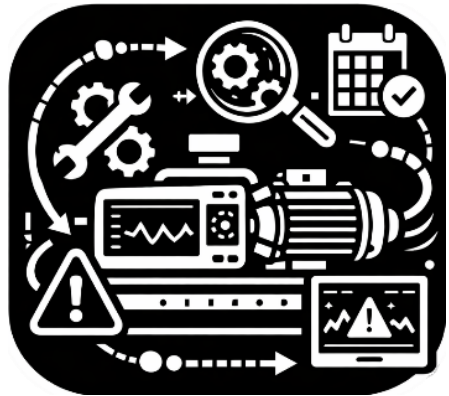
**Мониторинг
погоды и климата**



**Интернет
вещей**

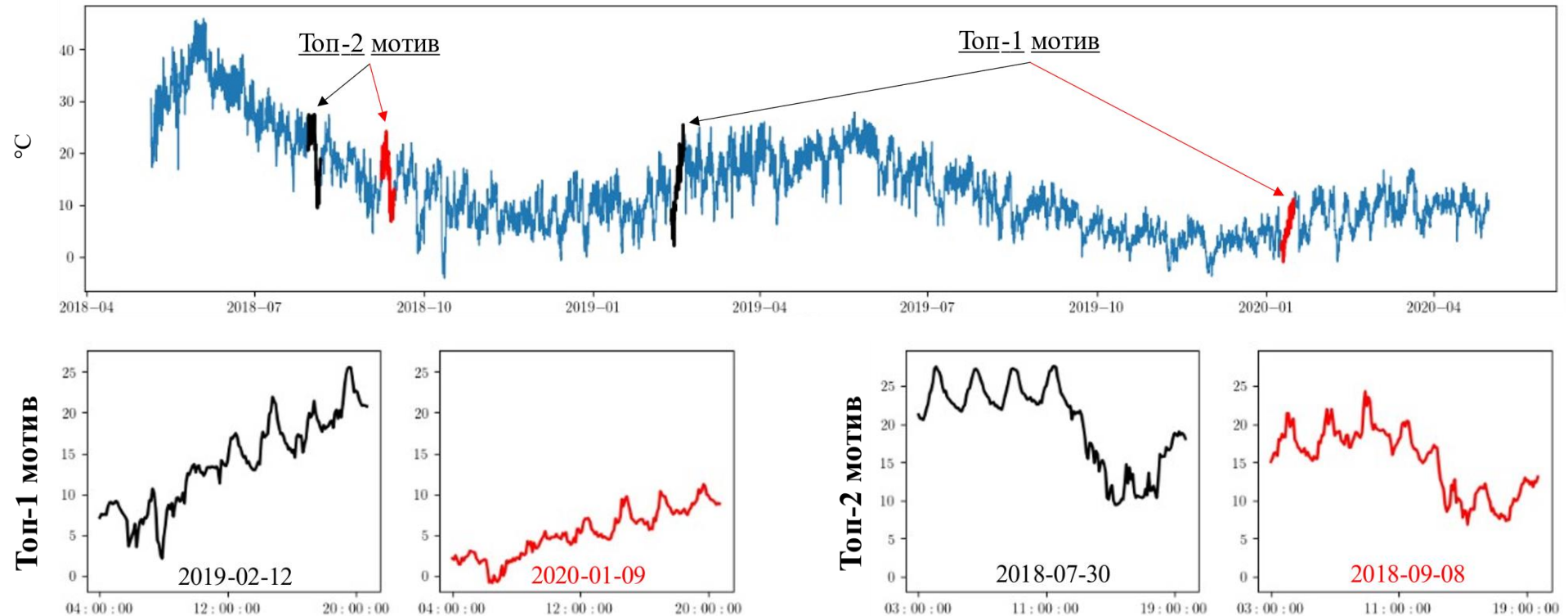


Энергосистемы и
«умные» сети



Промышленное
оборудование

Температура силового трансформатора (Китай)*



* Zhou H. *et al.* Informer: beyond efficient transformer for long sequence time-series forecasting. AAAI 2021: 11106-11115. DOI: [10.1609/aaai.v35i12.17325](https://doi.org/10.1609/aaai.v35i12.17325).

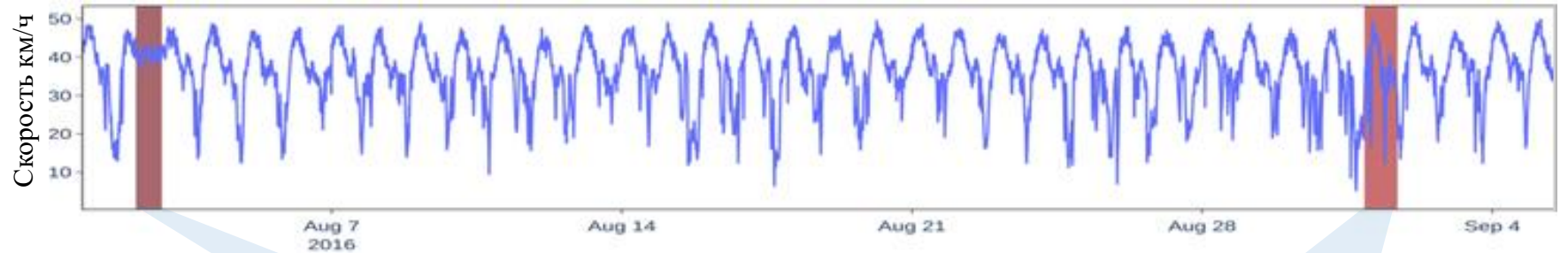


Мониторинг
погоды и климата

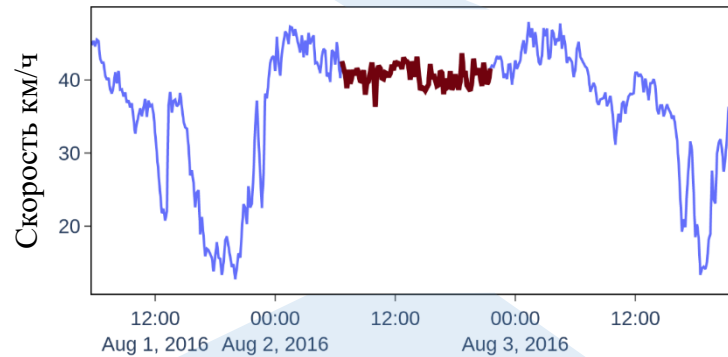


Системы
умного дома

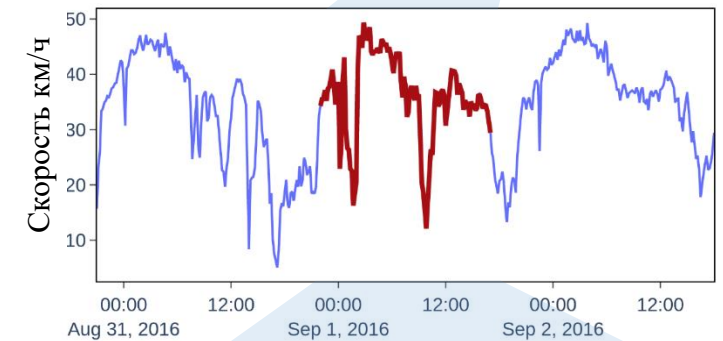
Скорость городского транспорта Гуанчжоу, Китай*



Топ-1 диссонанс

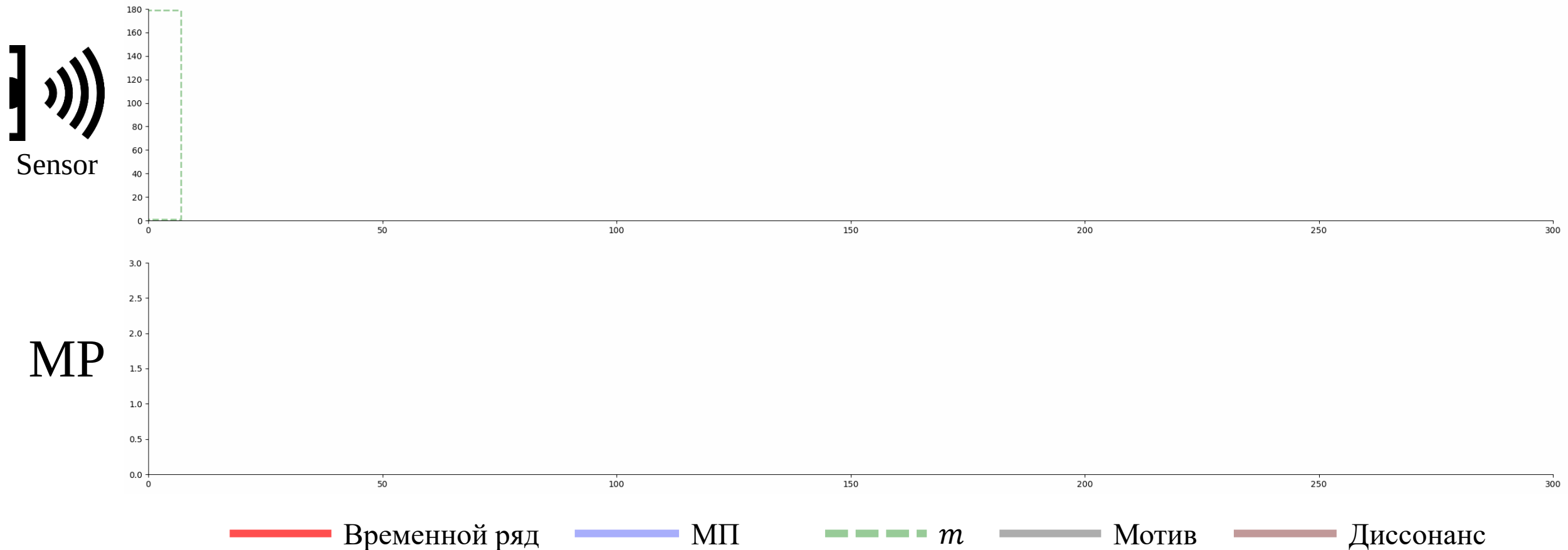


Топ-2 диссонанс



* Chen X, Chen Y, He Z. Urban traffic speed dataset of Guangzhou, China. 2018. DOI: [10.5281/zenodo.1205229](https://doi.org/10.5281/zenodo.1205229).

Матричный профиль – ключ к поиску мотивов, аномалий и др.



$$P = \{d_i\}_{i=1}^{n-m+1}, d_i = \arg \min_{T_i \cap T_j = \emptyset} \text{ED}(T_i, T_j)$$

Алгоритмы вычисления матричного профиля

#	Алгоритм	Платформа			Потоковый режим	Литература
		CPU	GPU	Cluster		
1	STAMP	✓				Yeh C.-C.M. <i>et al.</i> Matrix Profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. ICDM 2016. 1317-1322 (2016). https://doi.org/10.1109/ICDM.2016.0179
2	STAMPI	✓			✓	
3	STOMP	✓				
4	SCRIMP++	✓		✓		Zhu Y. <i>et al.</i> Matrix profile XI: SCRIMP++: time series motif discovery at interactive speeds. ICDM 2018. 837-846 (2018). https://doi.org/10.1109/ICDM.2018.00099
5	STUMP	✓	✓			Zhu Y. <i>et al.</i> Exploiting a novel algorithm and GPUs to break the ten quadrillion pairwise comparisons barrier for time series motifs and joins. Knowl. Inf. Syst. 54, 203-236 (2018). https://doi.org/10.1007/s10115-017-1138-x
9	SCAMP	✓	✓	✓		Zimmerman Z. <i>et al.</i> Matrix Profile XIV: Scaling time series motif discovery with GPUs to break a quintillion pairwise comparisons a day and beyond. SoCC 2019. 74-86 (2019). https://doi.org/10.1145/3357223.3362721

Новый алгоритм StreaMP++

- Использование множества GPU
- Минимизация операций чтения и записи с GPU
- **Новый подход** к вычислению матричного профиля



2025

StreaMP



2026

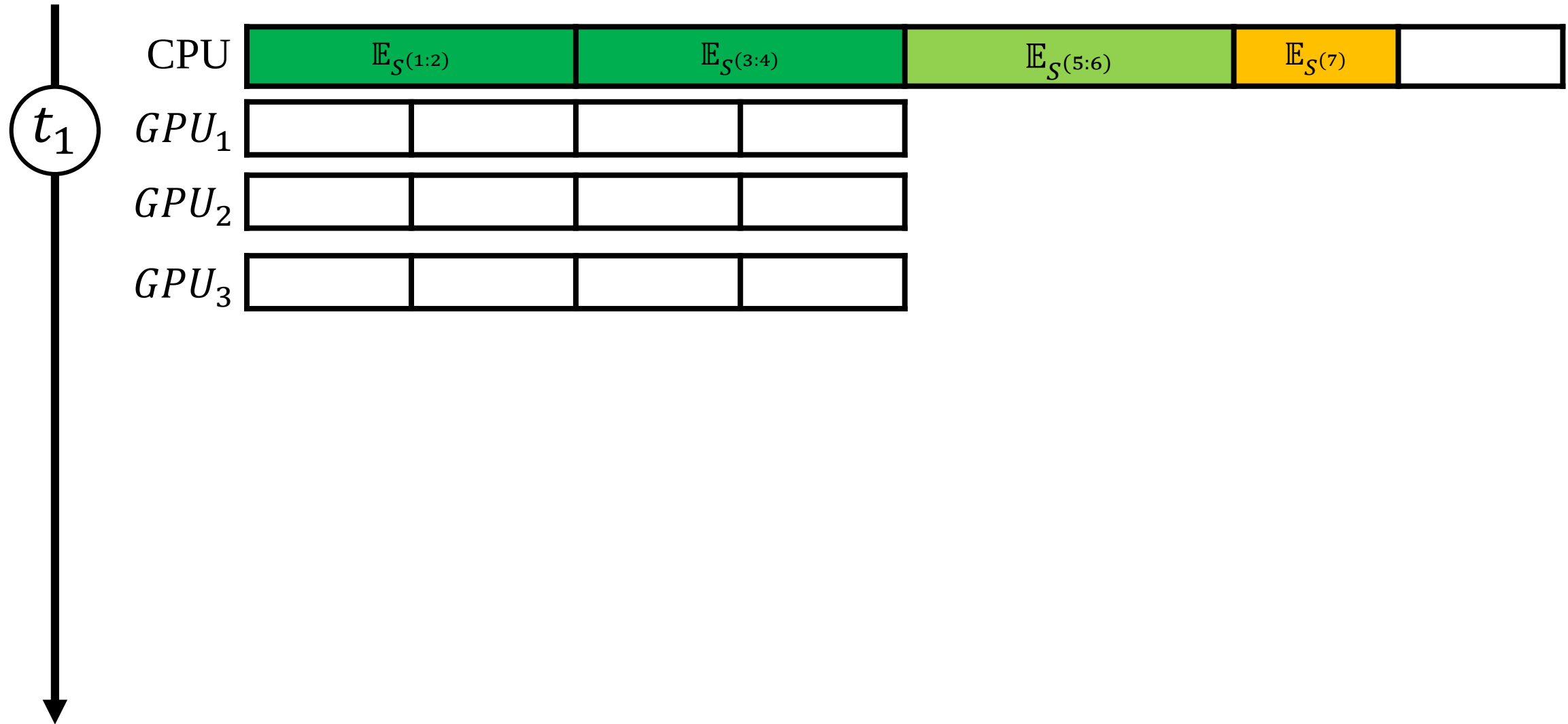
StreaMP++



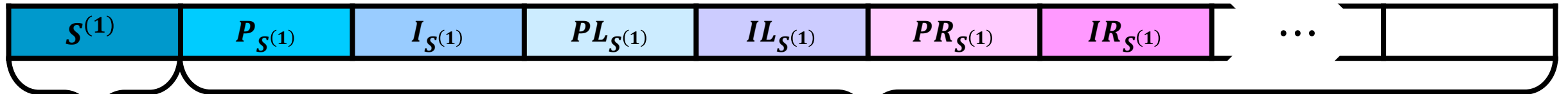
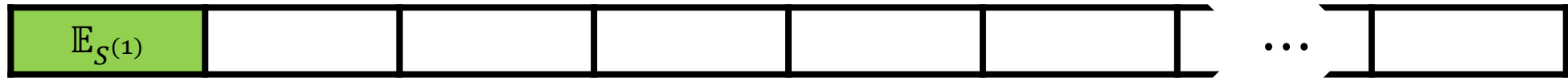
2027

?

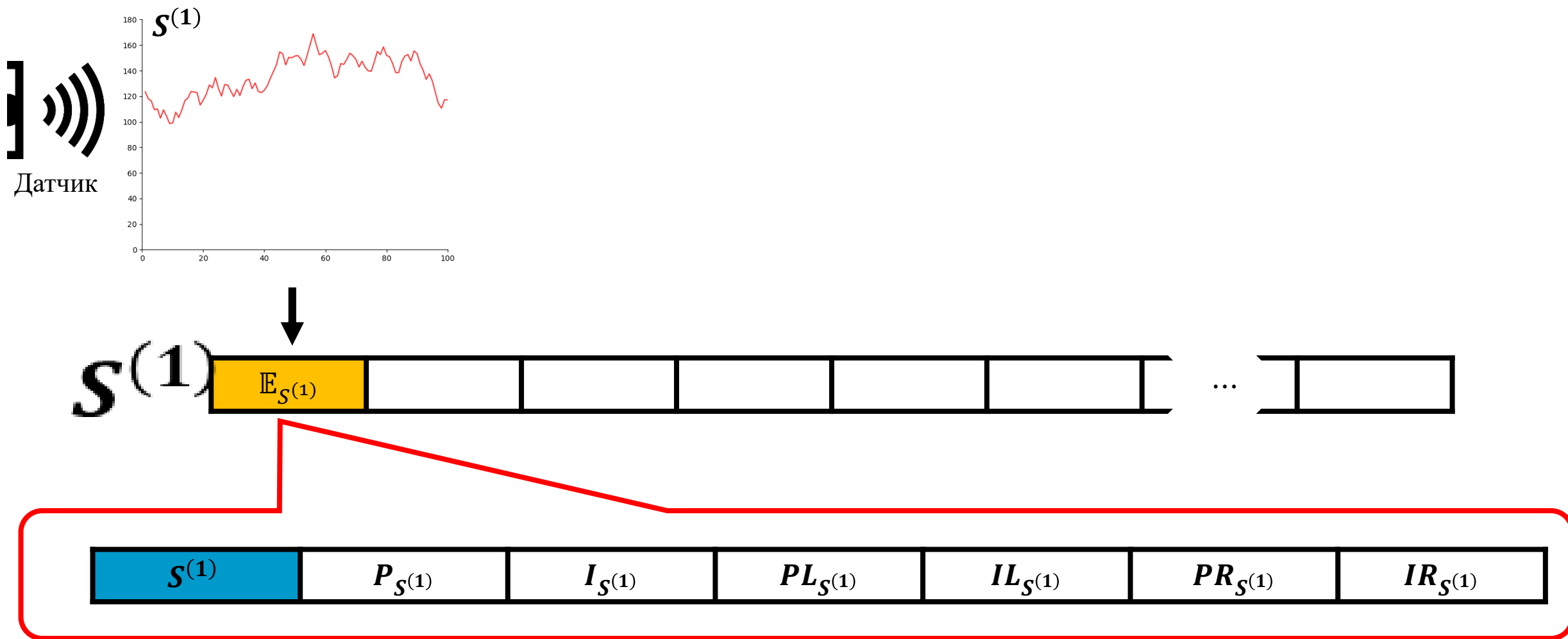
Основная логика алгоритма StreaMP++



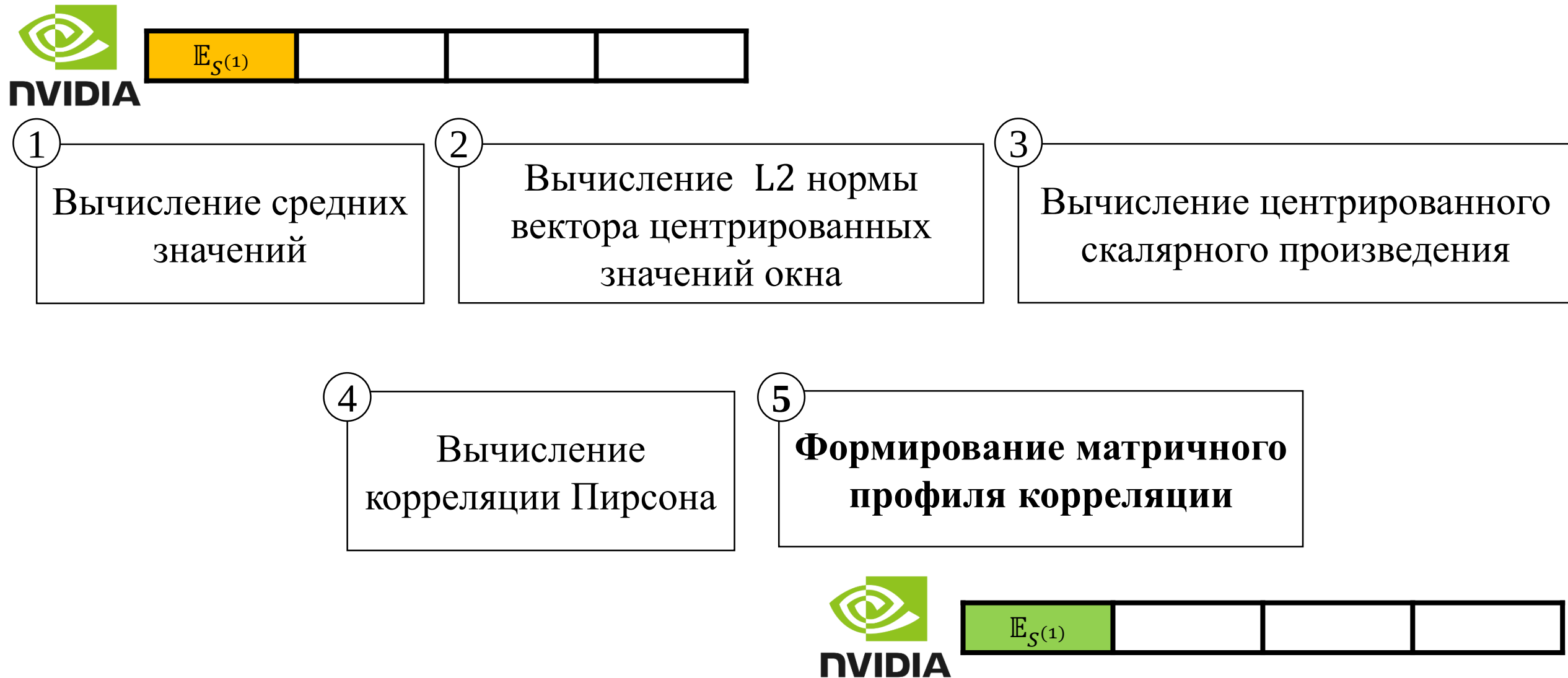
Структура сегмента $\mathbb{E}_{S^{(1)}}$



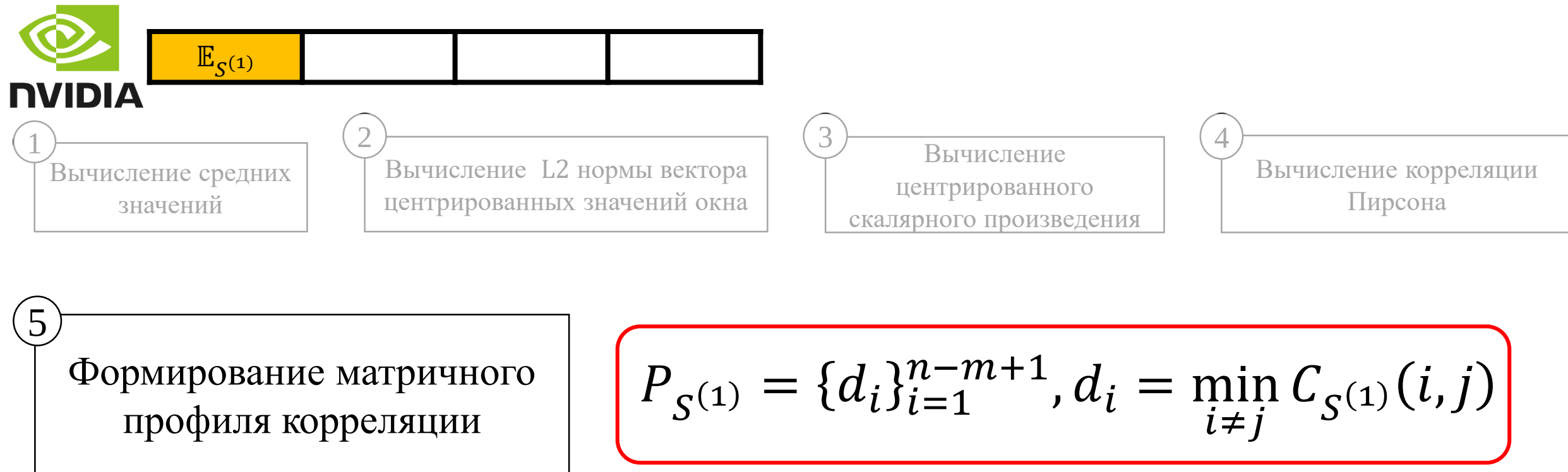
Поступление нового сегмента



Вычисление расширенного МП



Шаг 5: Формирование МП



Формирование блока $\mathbb{E}_S^{(1:2)}$

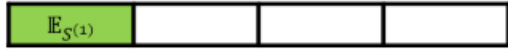


1
Вычисление
МП $\mathbb{E}_S^{(2)}$

Вычисление расширенного МП



- 1
Вычисление средних значений
- 2
Вычисление L2 нормы вектора центрированных значений окна
- 3
Вычисление центрированного скалярного произведения
- 4
Вычисление корреляции Пирсона
- 5
Формирование матричного профиля корреляции



Формирование блока



1

Вычисление
МП $E_{S(2)}$

2

Вычисление
конкатенации РМП $E_{S(1:2)}$



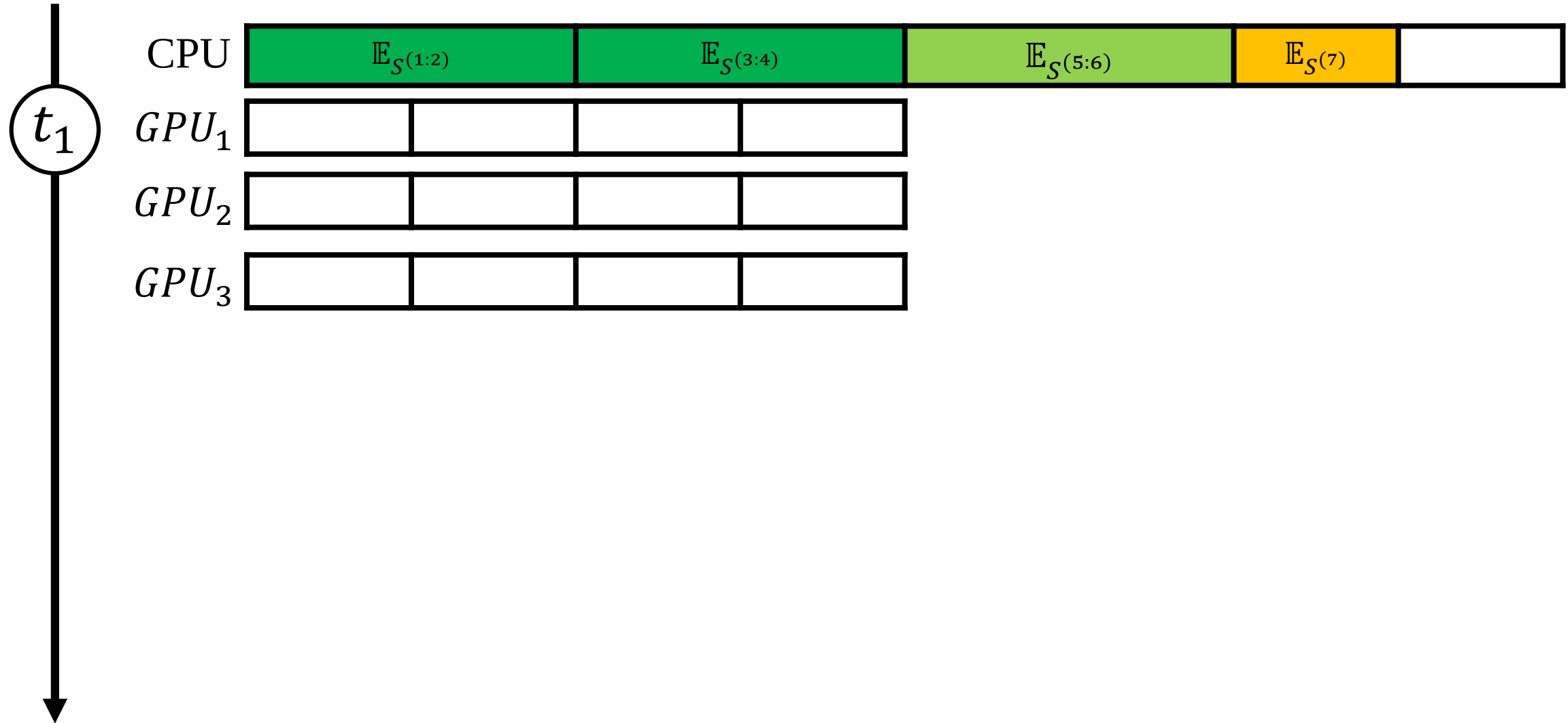
for $i = 1$ **to** N **do**
 for $j = 1$ **to** N **do**
 if $i = j$ **then continue**
 Вычислить $P_{S(i)} P_{S(j)}$

$$P_{S(1:N)} = \{d_i\}_{i=1}^{N(n-m+1)}$$

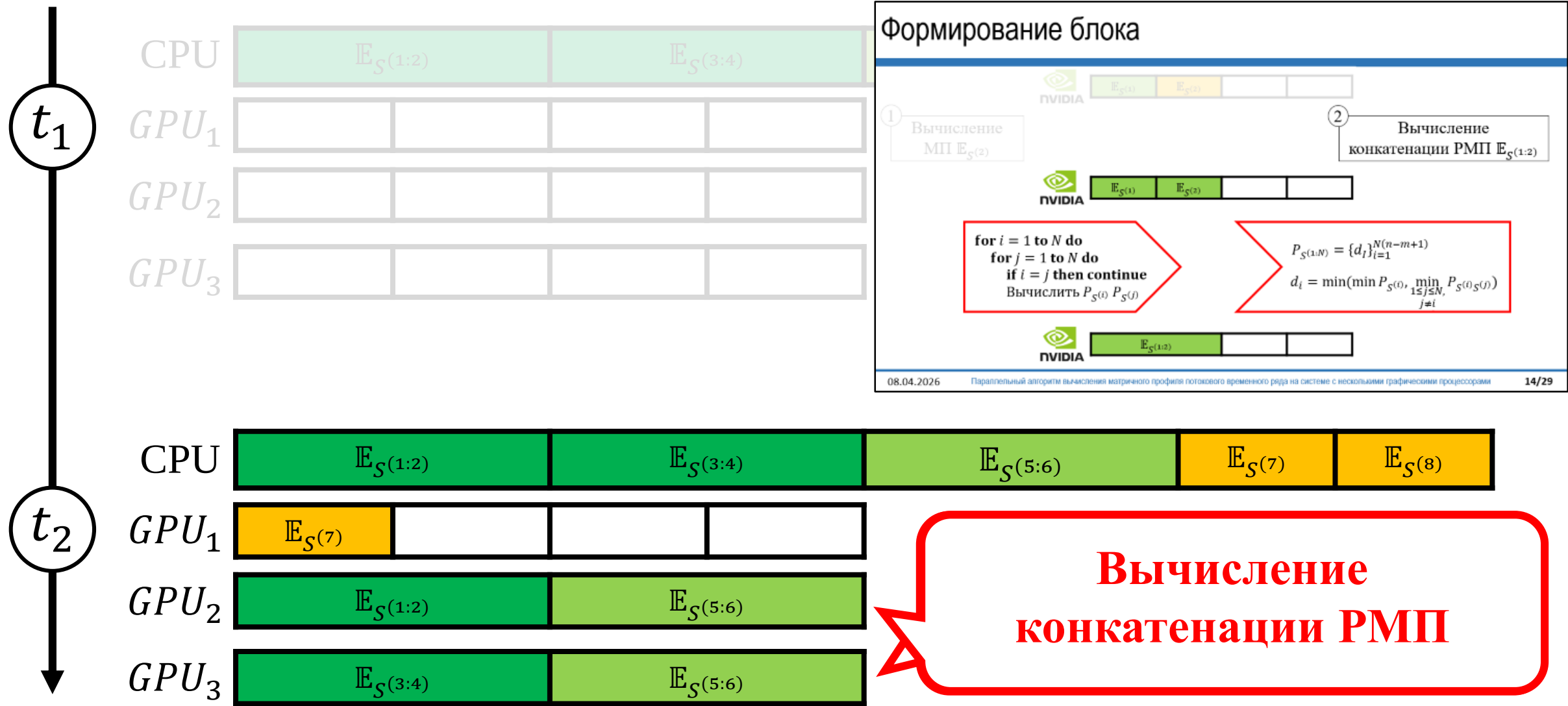
$$d_i = \min(\min P_{S(i)}, \min_{\substack{1 \leq j \leq N, \\ j \neq i}} P_{S(i)S(j)})$$



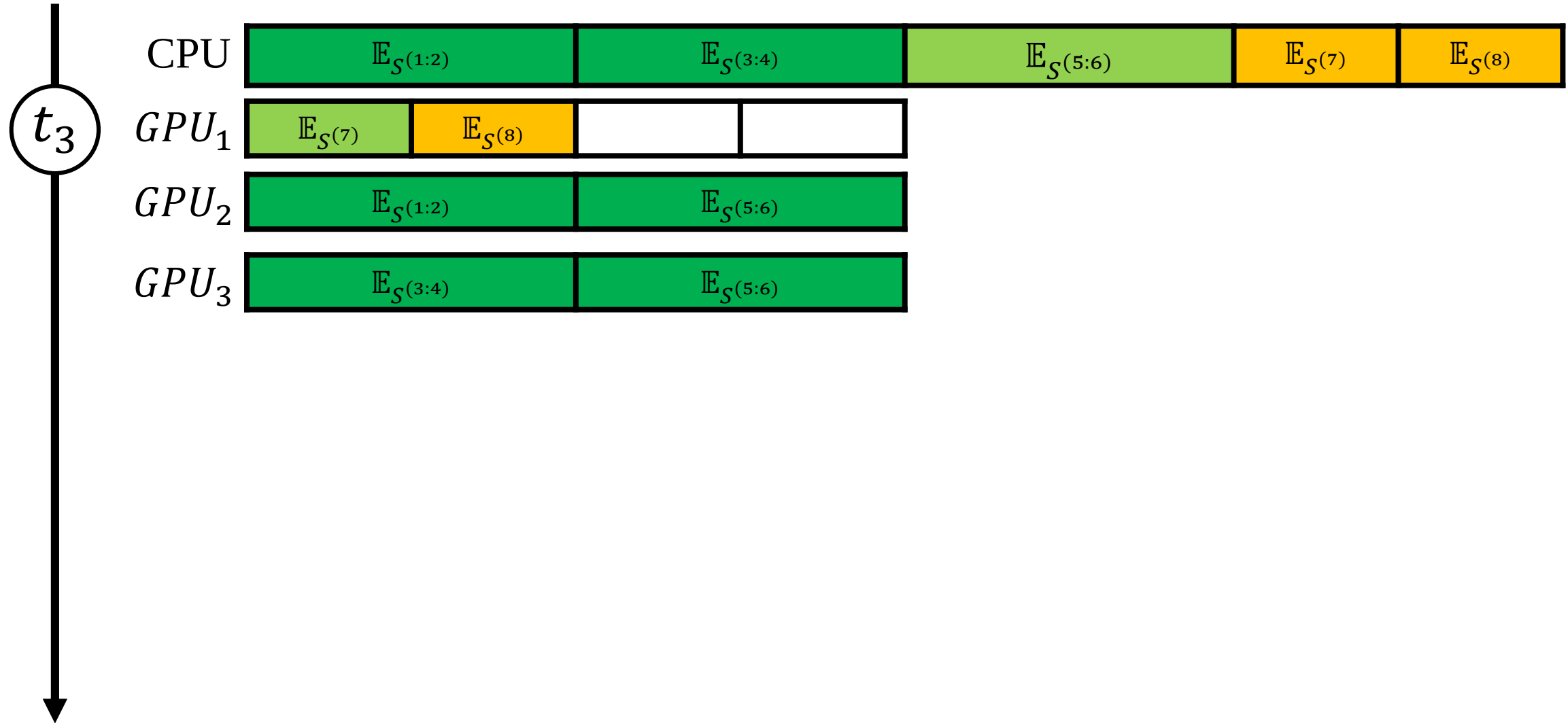
Использование нескольких GPU



Использование нескольких GPU



Использование нескольких GPU



Использование нескольких GPU



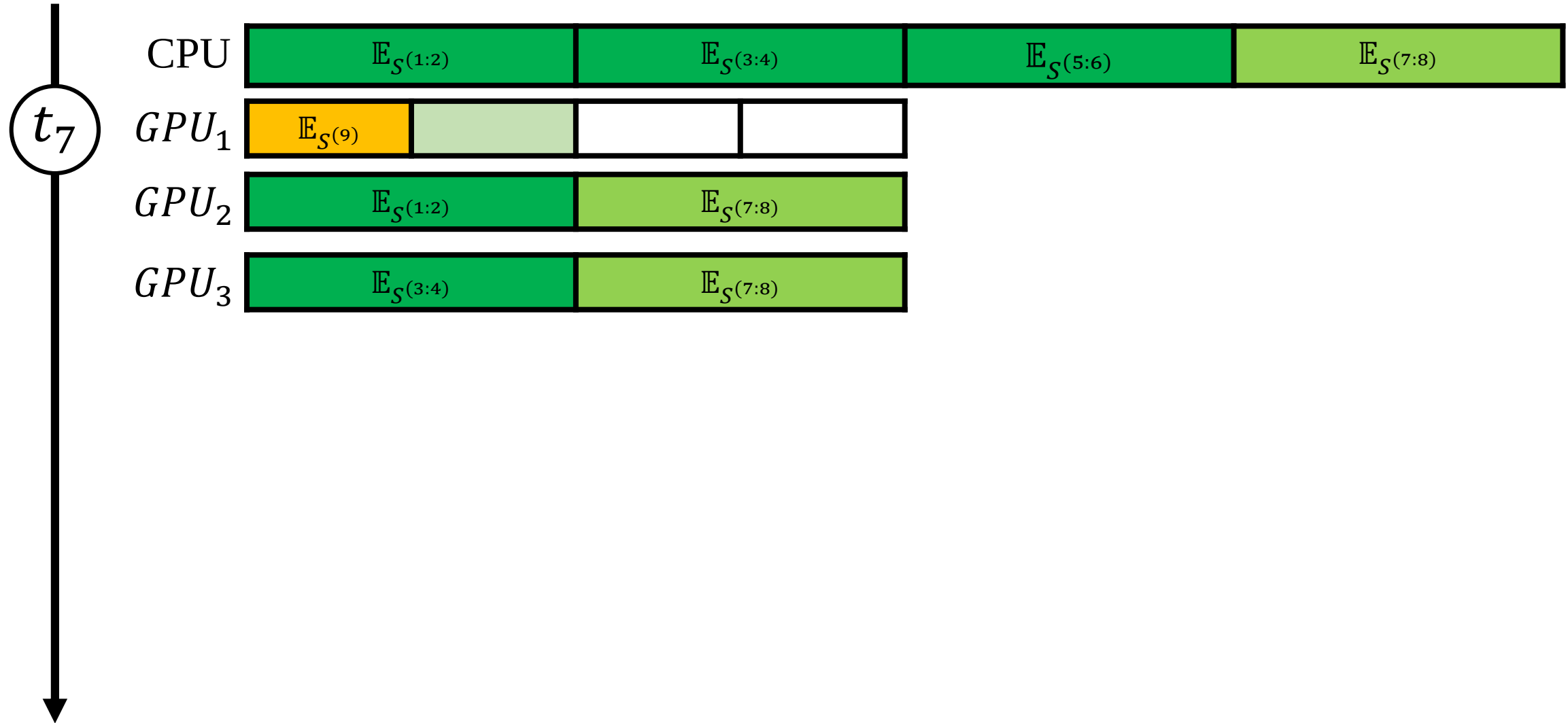
Использование нескольких GPU



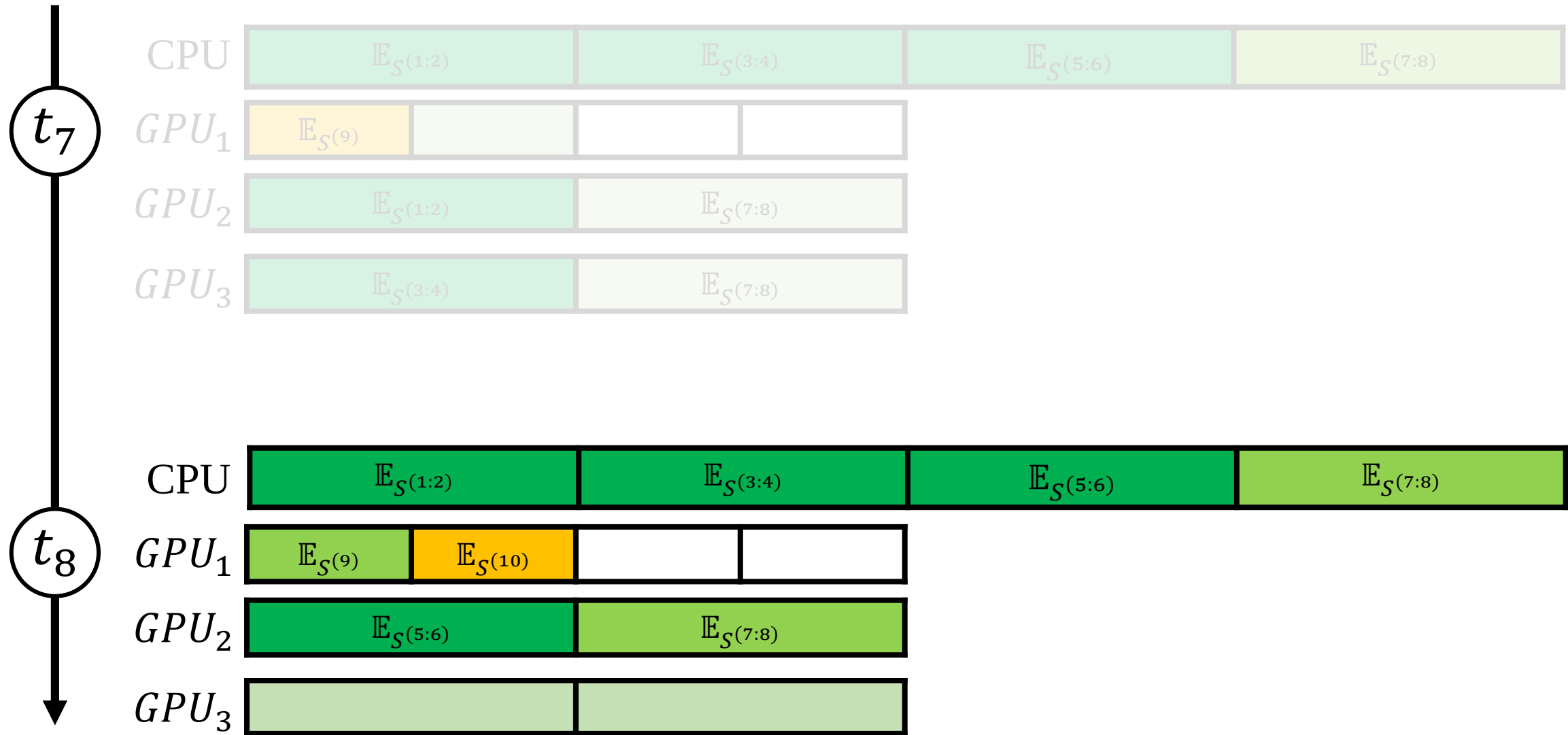
Использование нескольких GPU



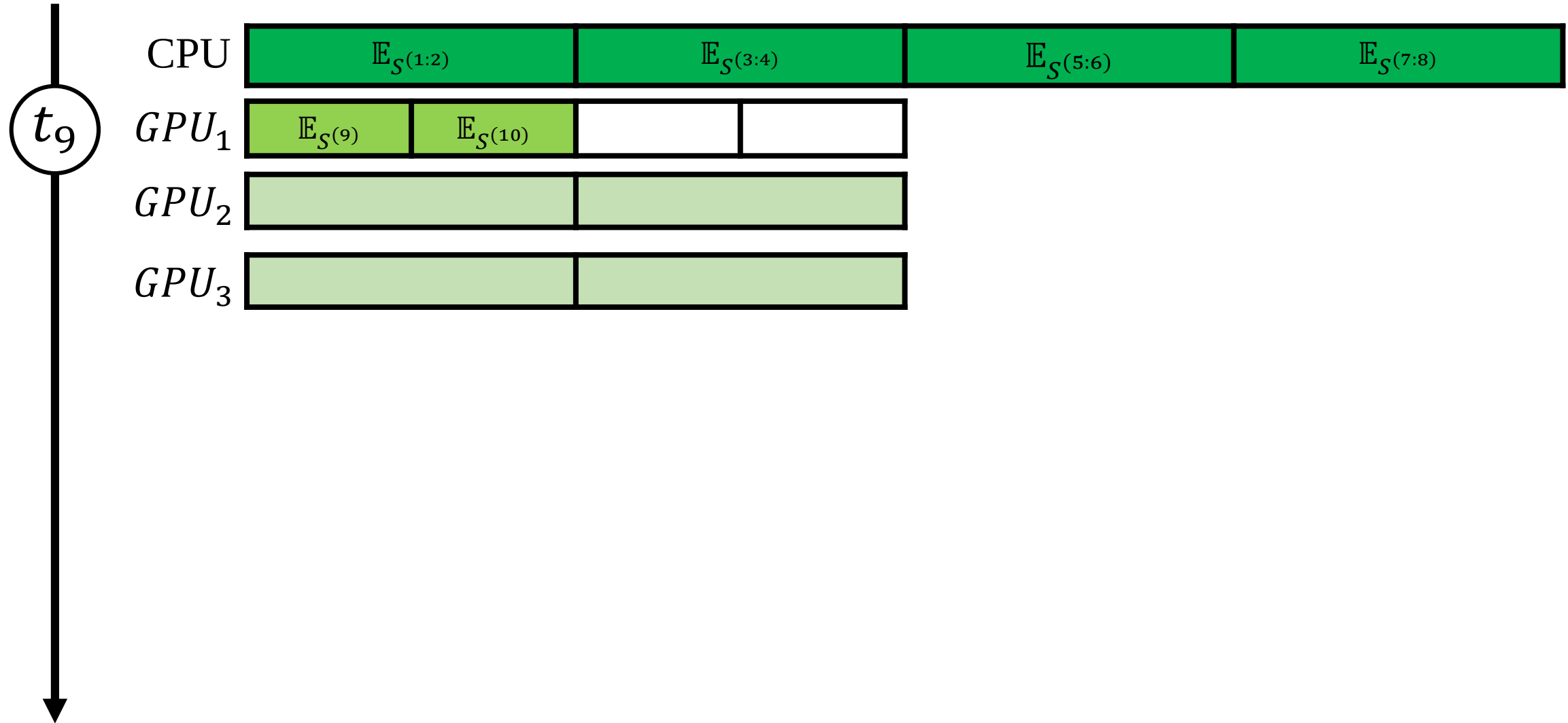
Использование нескольких GPU



Использование нескольких GPU



Использование нескольких GPU



Использование нескольких GPU



Вычислительные эксперименты

- **Аппаратная платформа**

- CPU: AMD Intel Xeon Gold 6254
- RAM: 192 GB
- GPU: NVIDIA Tesla V100 SXM2 × 4, 32 Gb

- **Данные**

- Временной ряд: $T = 2^{16}, 2^{17}, 2^{18}, 2^{19}, 2^{20}, 2^{21}, 2^{22}, 2^{23}, 2^{24}, 2^{25}, 2^{26}$
- Длина подпоследовательности: $m = 265$
- Длина сегмента: $|S| = 2^{16}$

Исследования выполнены с использованием суперкомпьютерных ресурсов ЮУрГУ

Имитируемые датчики-источники потоковых данных

Длина ряда, 2^x	Частота, KHz	
16	240	
17	230	
18	210	
19	153	
20	50	
21	50	
22	25	
23	10	
24	2	
25	0.6	



Бесконтактный мониторинг дыхания



Акустика в движущихся средах



Дистанционное зондирование морского дна



Выявление дефектов подшипников и ротора



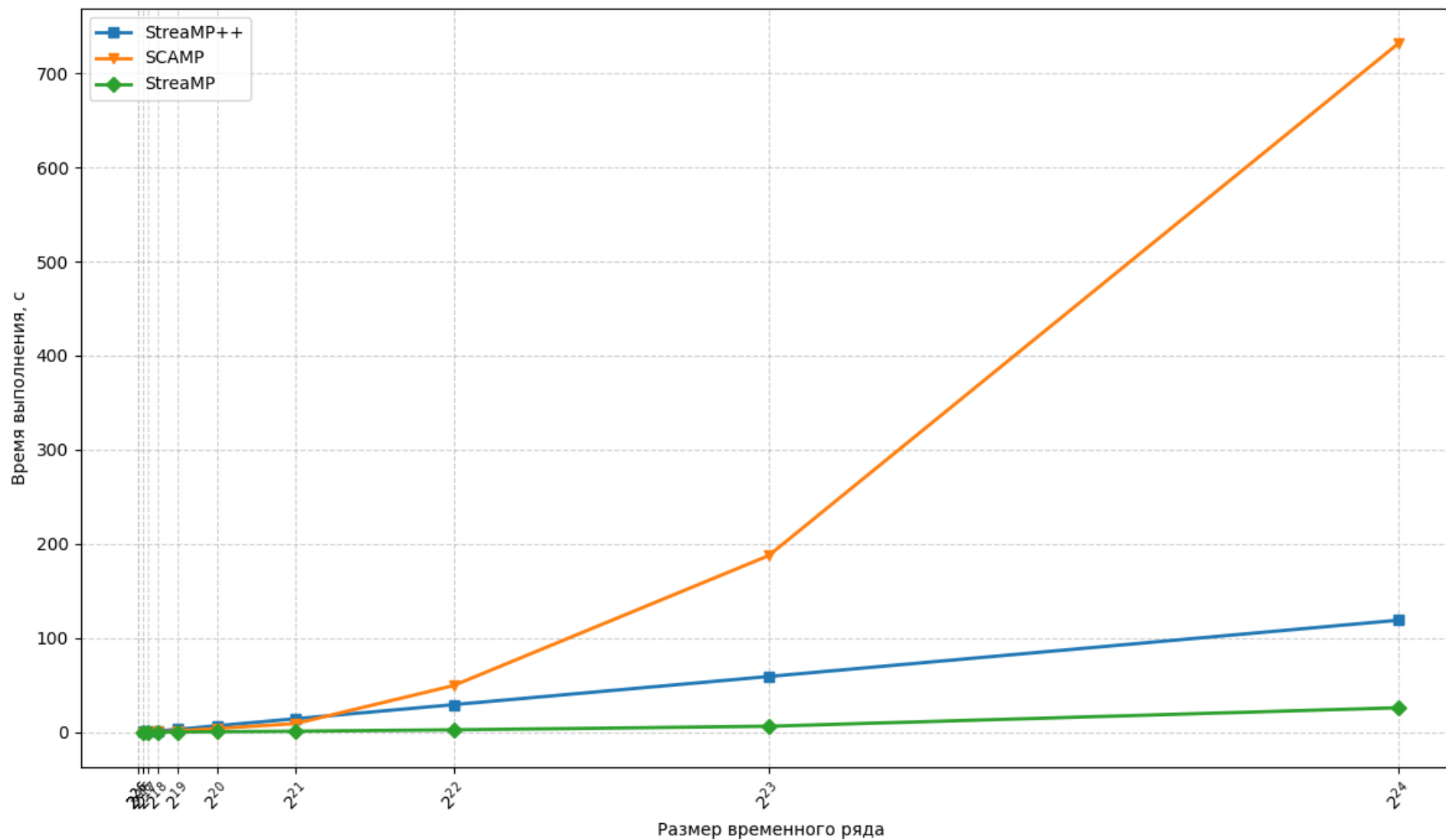
Положение беспилотного летательного аппарата



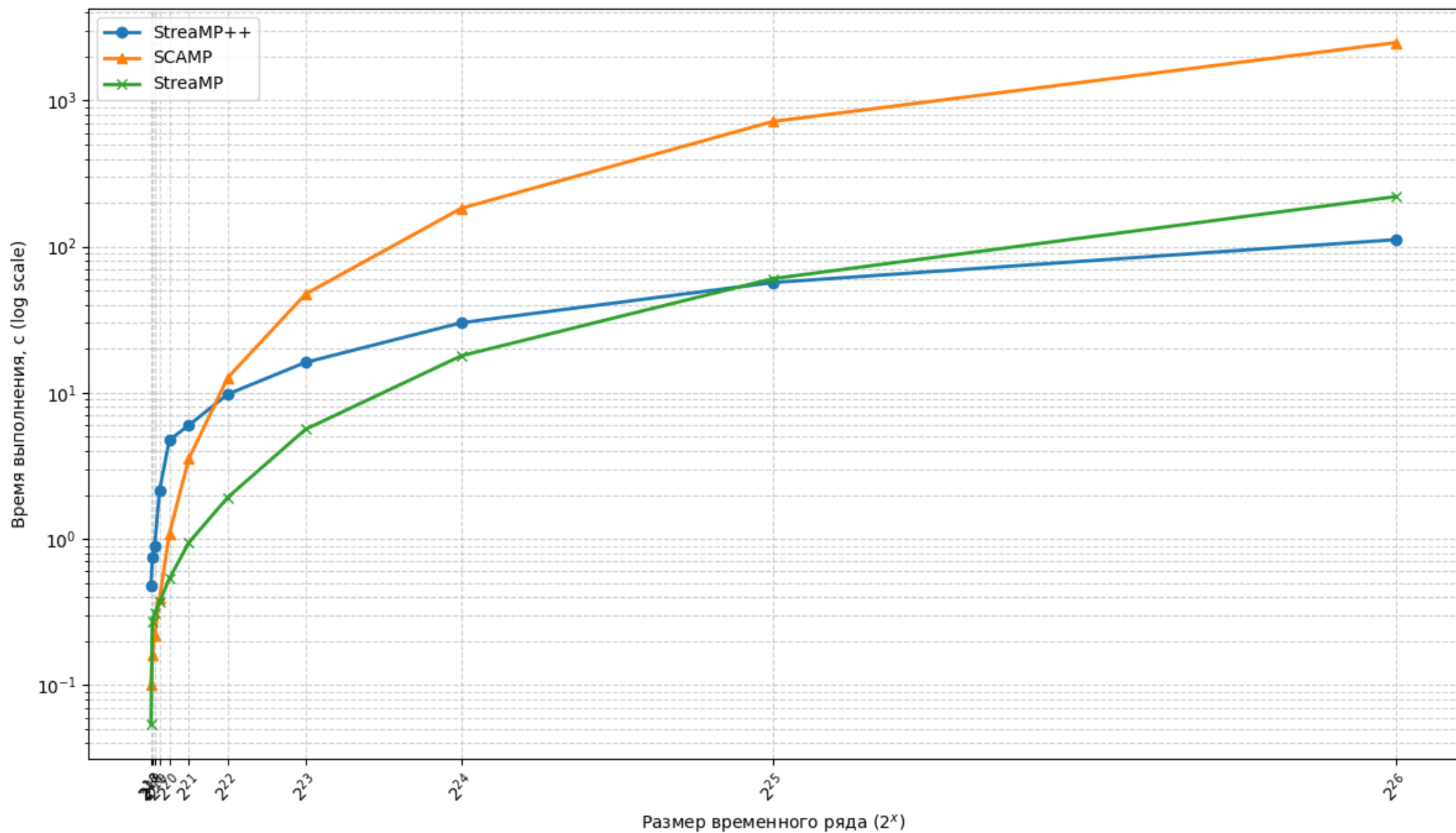
Мониторинг сейсмических волн



Результаты экспериментов (1 GPU)



Результаты экспериментов (4 GPU)



Заключение

- **StreaMP++**, новый параллельный алгоритм вычисления матричного профиля (МП) потокового временного ряда для кластера с GPU-узлами
- StreaMP++ превосходит передовые аналоги по быстродействию более чем в 10 раз
- МП, вычисленный с помощью StreaMP++, обеспечивает в потоковом режиме поиск мотивов, диссонансов и др. паттернов за линейное время