

Международная научная конференция
Параллельные вычислительные технологии (ПавТ'2020)
27–29 мая 2020

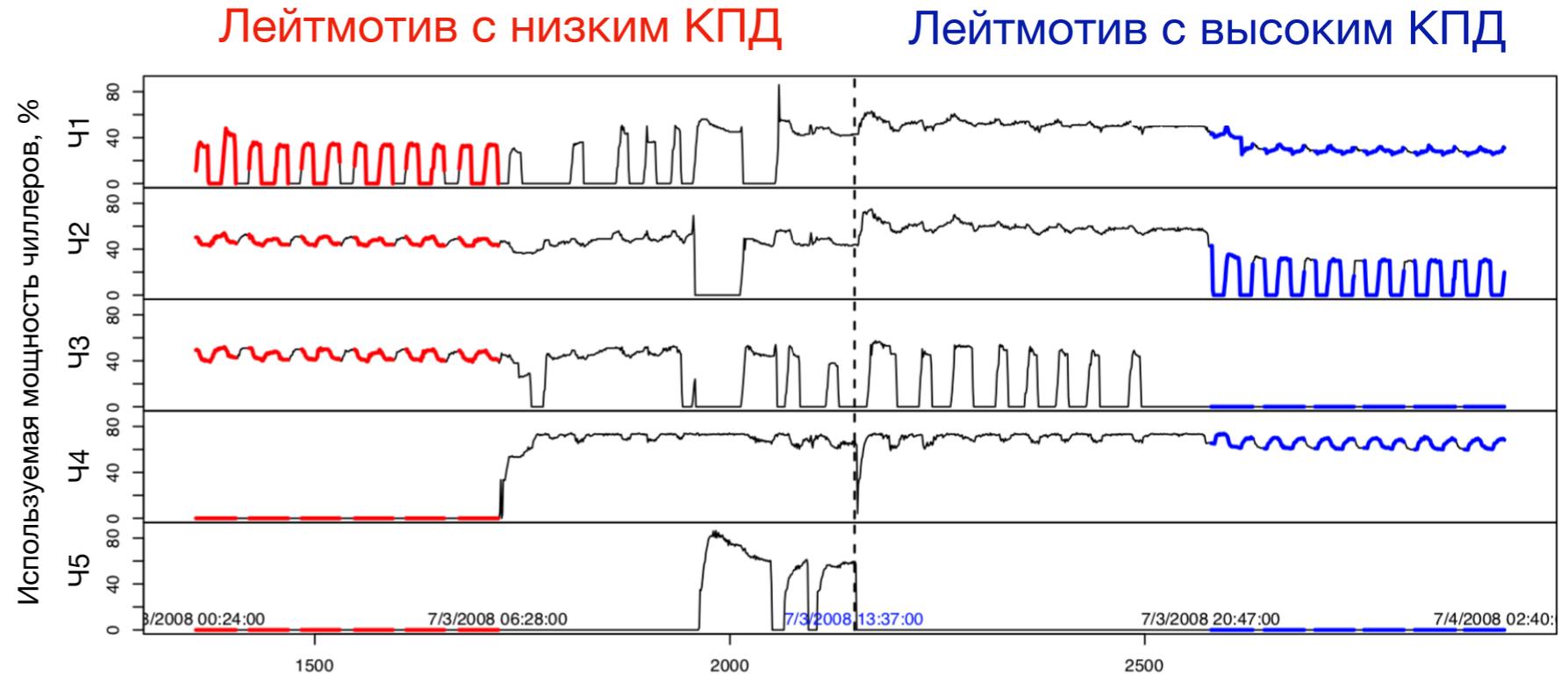
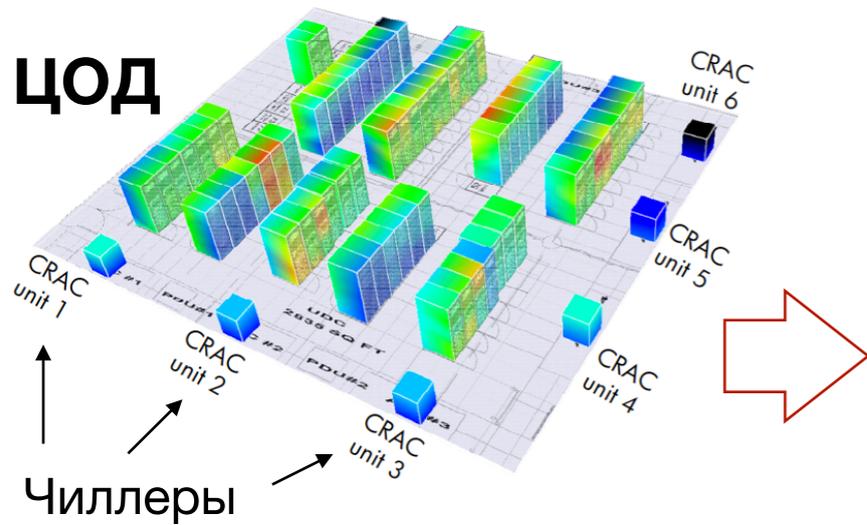
Параллельный алгоритм поиска лейтмотивов временного ряда для графического процессора

Я.А. Краева, М.Л. Цымблер

Южно-Уральский государственный университет (Челябинск)

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», уникальный идентификатор проекта RFMEFI57818X0265 (контракт № 075-15-2019-1339 (14.578.21.0265)) и Российского фонда фундаментальных исследований (проект № 20-07-00140-а).

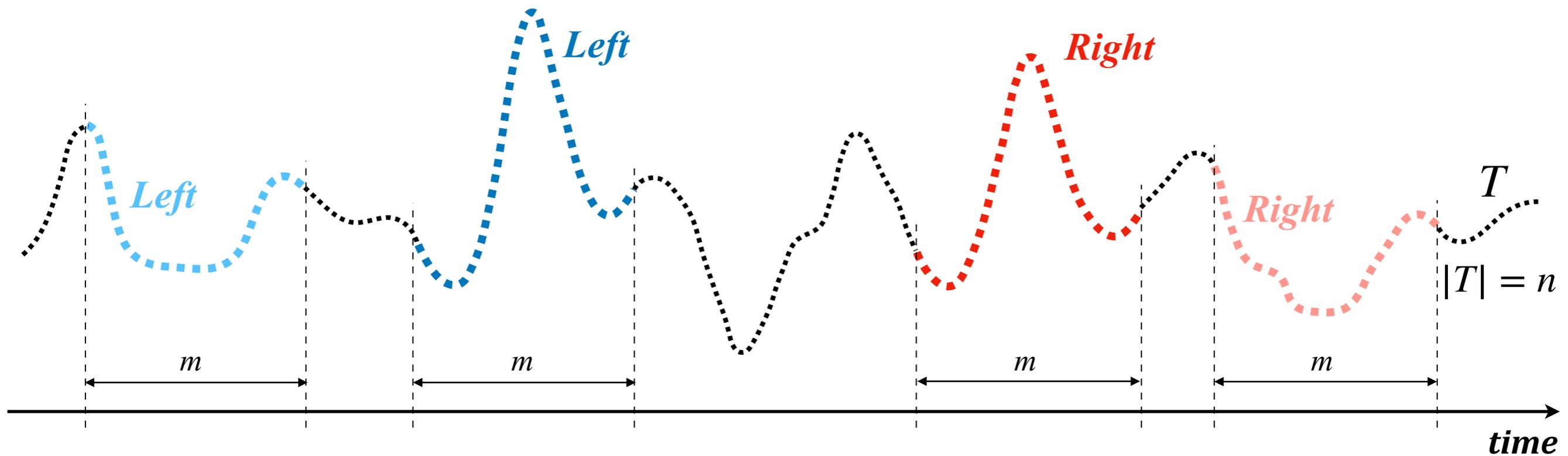
Лейтмотивы на практике



- Нахождение лейтмотивов во временных рядах показаний мощности чиллеров позволяет подобрать энергоэффективный режим работы системы охлаждения в ЦОД ¹⁾
- Экономический эффект **40 000\$ в год**

¹⁾ Patnaik D., *et al.* Sustainable operation and management of data center chillers using temporal data mining. SIGKDD 2009. pp. 1305–1314.

Лейтмотив временного ряда



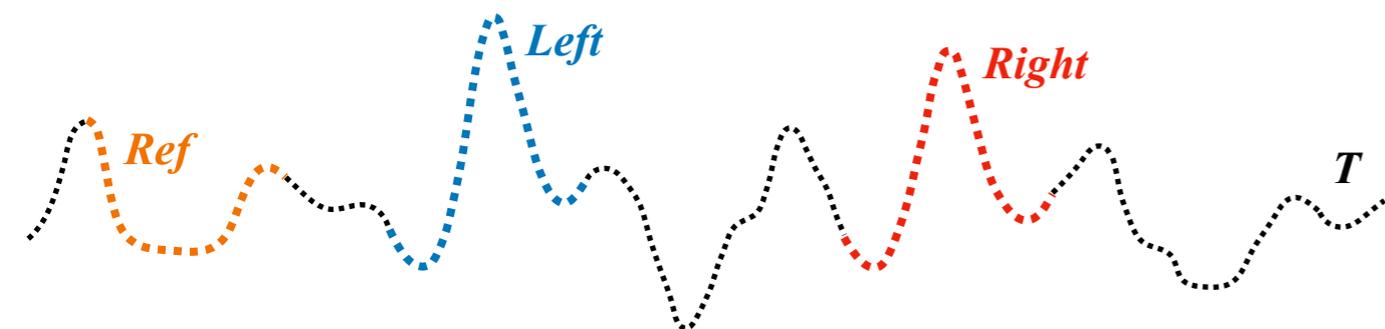
Лейтмотив $\{Left, Right\}$ – пара наиболее похожих подпоследовательностей временного ряда T

$$\forall Left, Right \quad ED(Left, Right) < ED(Left, Right)$$

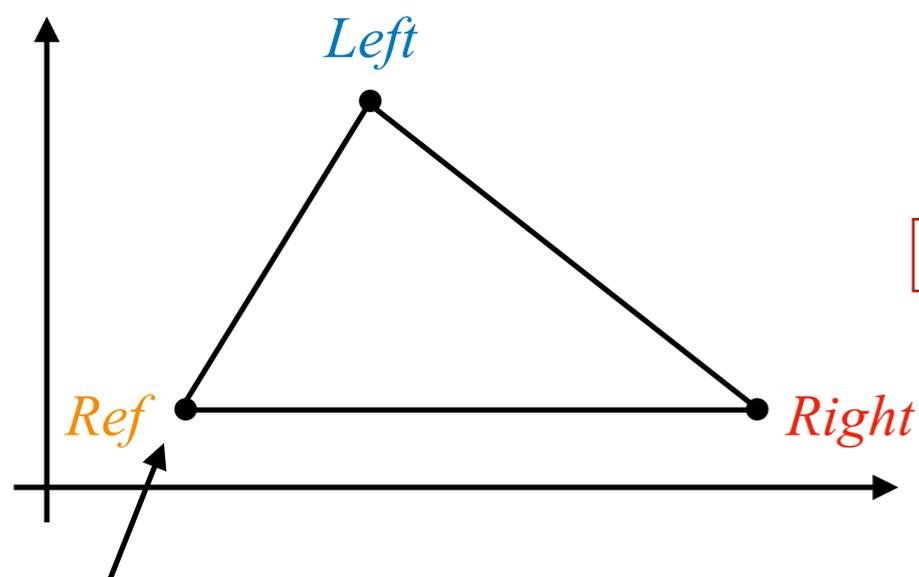
$$Left \cap Right = \emptyset,$$

$ED(\cdot, \cdot)$ – евклидово расстояние

Неравенство треугольника

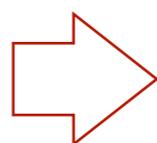


↓ Проецирование
в \mathbb{R}^2



опорная
подп-ть

Неравенство треугольника



$$ED(\text{Ref}, \text{Right}) + ED(\text{Left}, \text{Right}) \geq ED(\text{Ref}, \text{Left})$$



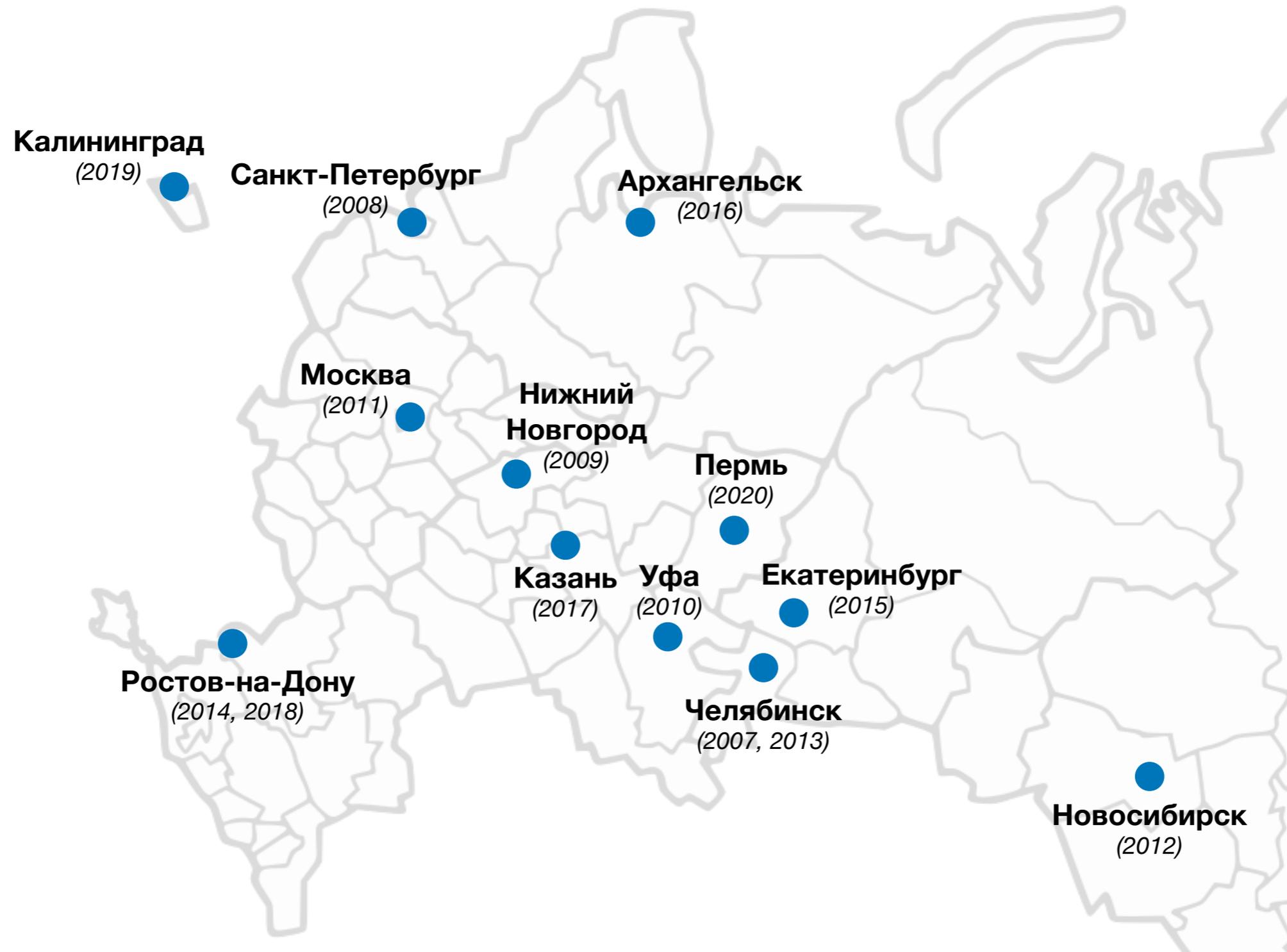
Обратное неравенство треугольника

$$\underbrace{ED(\text{Ref}, \text{Left}) - ED(\text{Ref}, \text{Right})}_{\text{Нижняя граница LB}} \leq \underbrace{ED(\text{Left}, \text{Right})}_{\text{Истинное расстояние}}$$

Нижняя граница LB

Истинное расстояние

Карта конференций ПаВТ



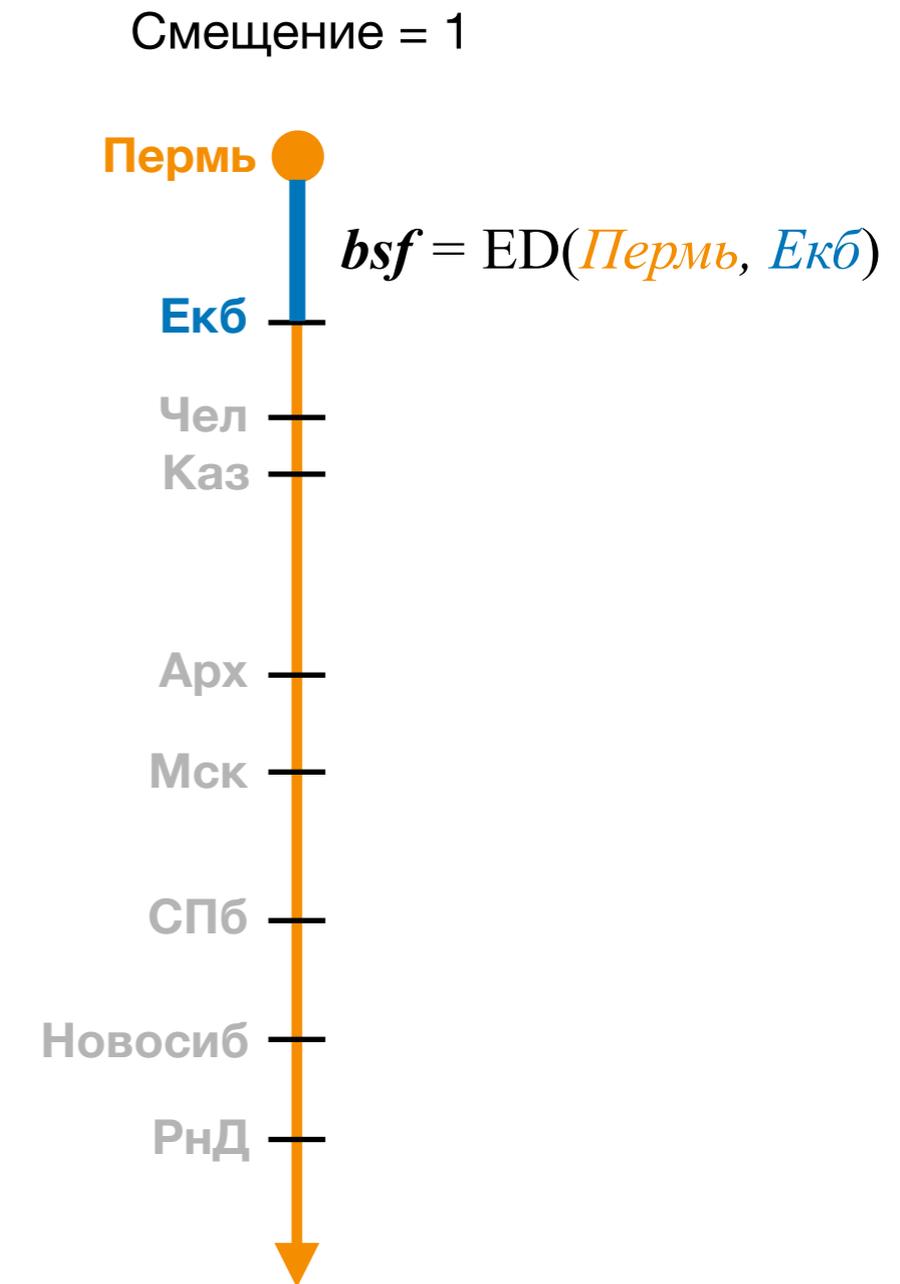
Выбор опорного города



Линейный порядок городов



Вычисление начального значения порога



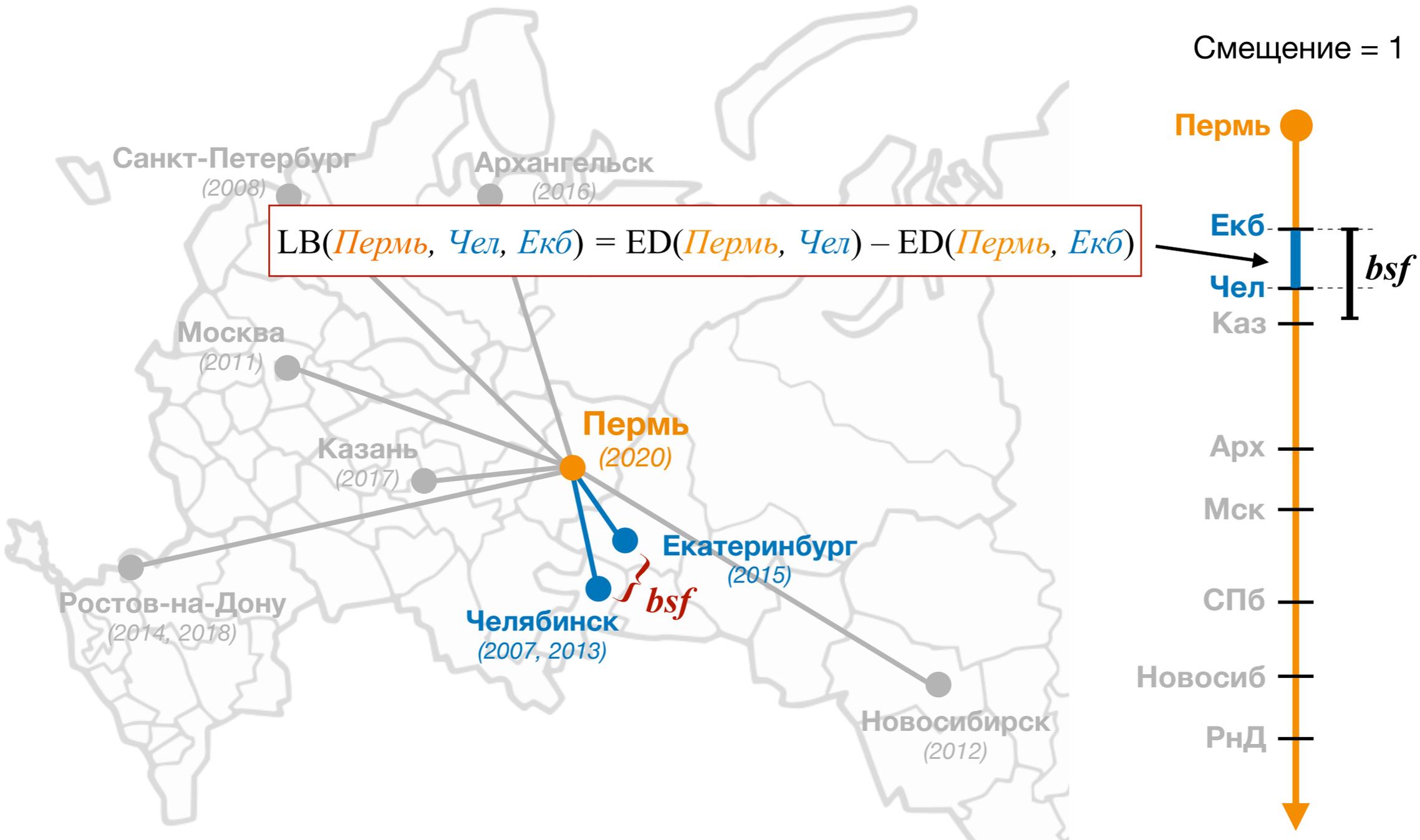
Вычисление нижней границы

$$LB(Ref, Left, Right) = ED(Ref, Left) - ED(Ref, Right)$$



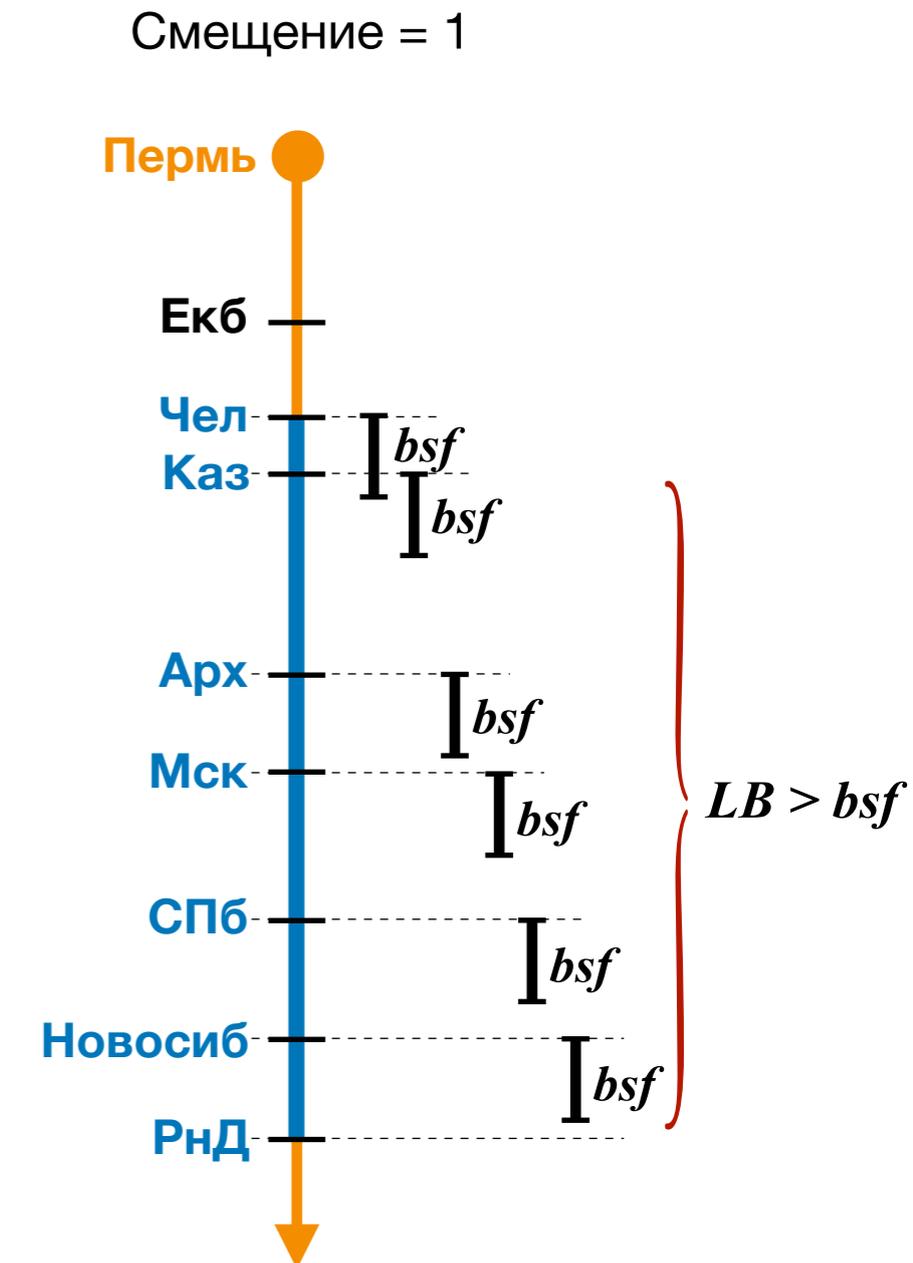
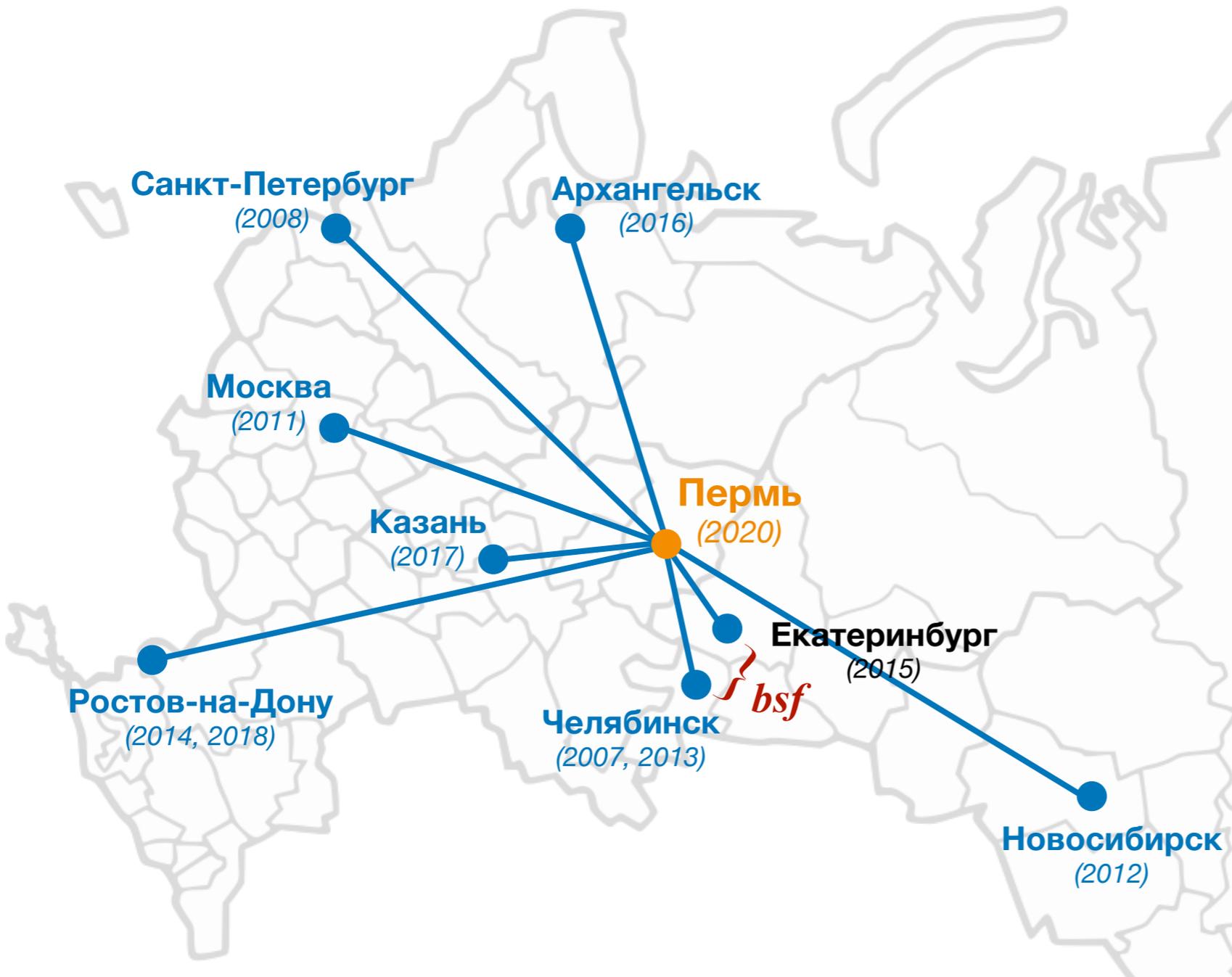
$$LB(Пермь, Чел, Екб) < bsf \Rightarrow bsf = \min(bsf, ED(Екб, Чел))$$

Обновление значения порога



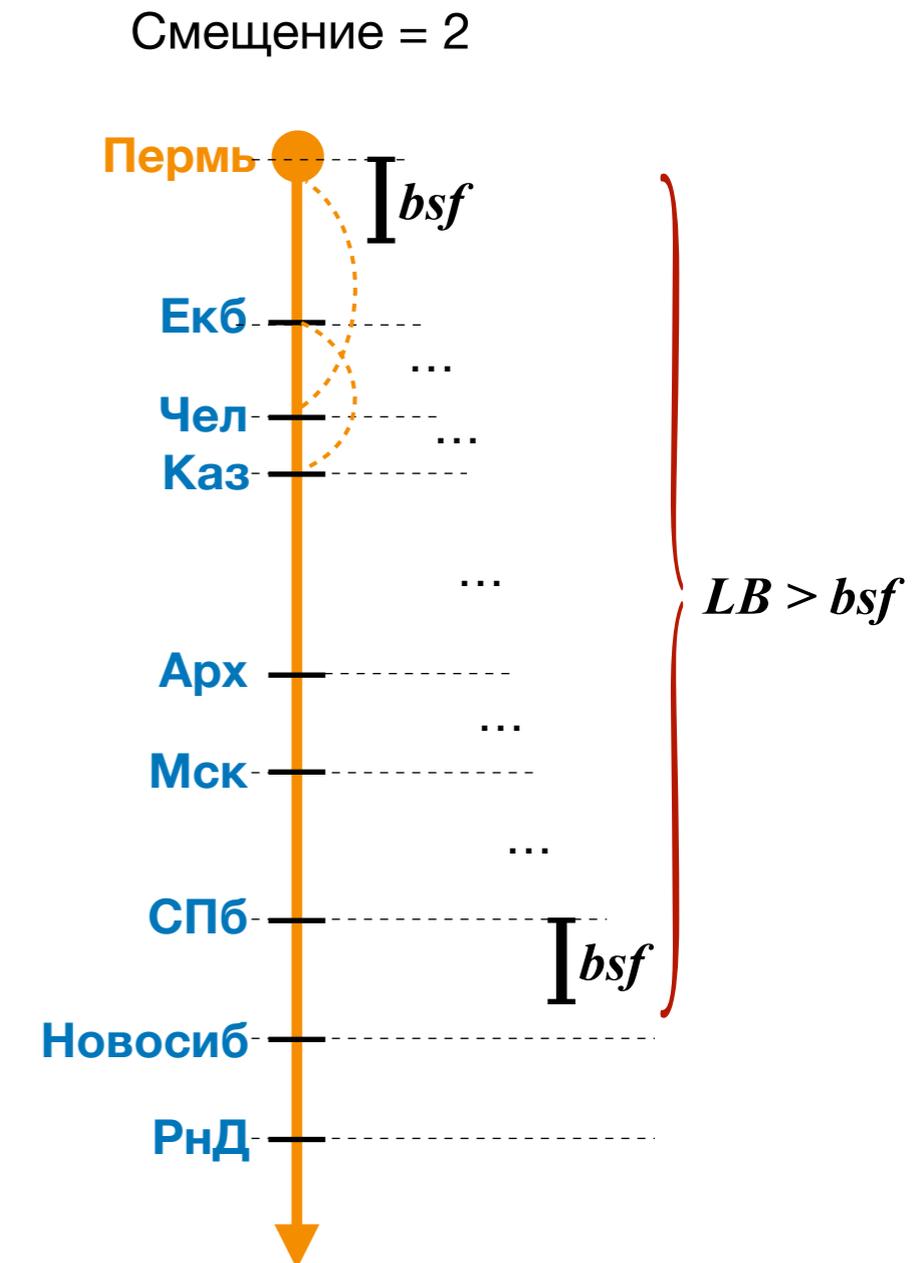
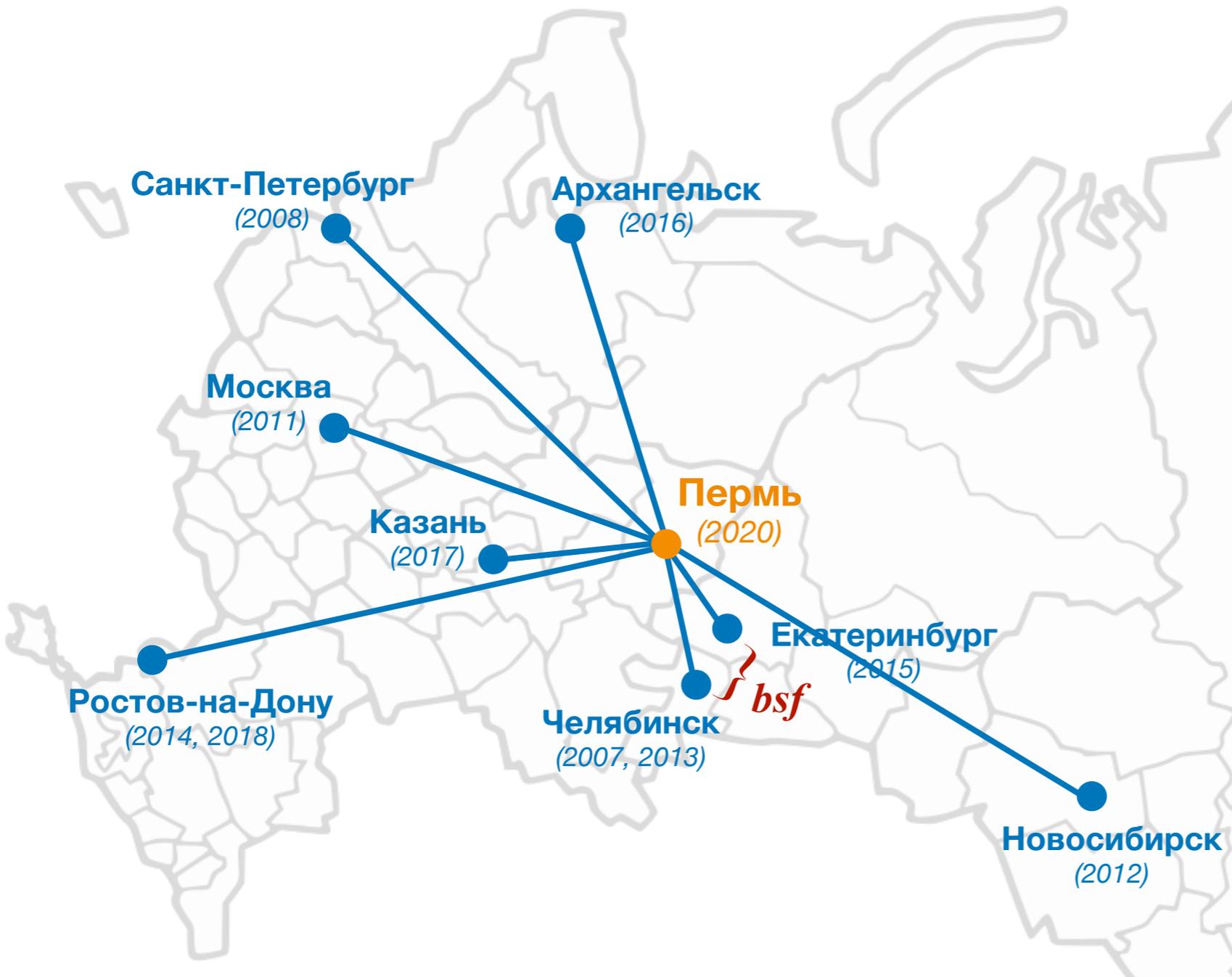
$$LB(\text{Пермь}, \text{Чел}, \text{Екб}) < bsf \Rightarrow bsf = \min(bsf, ED(\text{Екб}, \text{Чел}))$$

Отбрасывание кандидатов



$LB(Ref, Left, Right) > bsf \Rightarrow \{Left, Right\}$ отбрасывается без вычисления $ED(Left, Right)$

Отбрасывание кандидатов



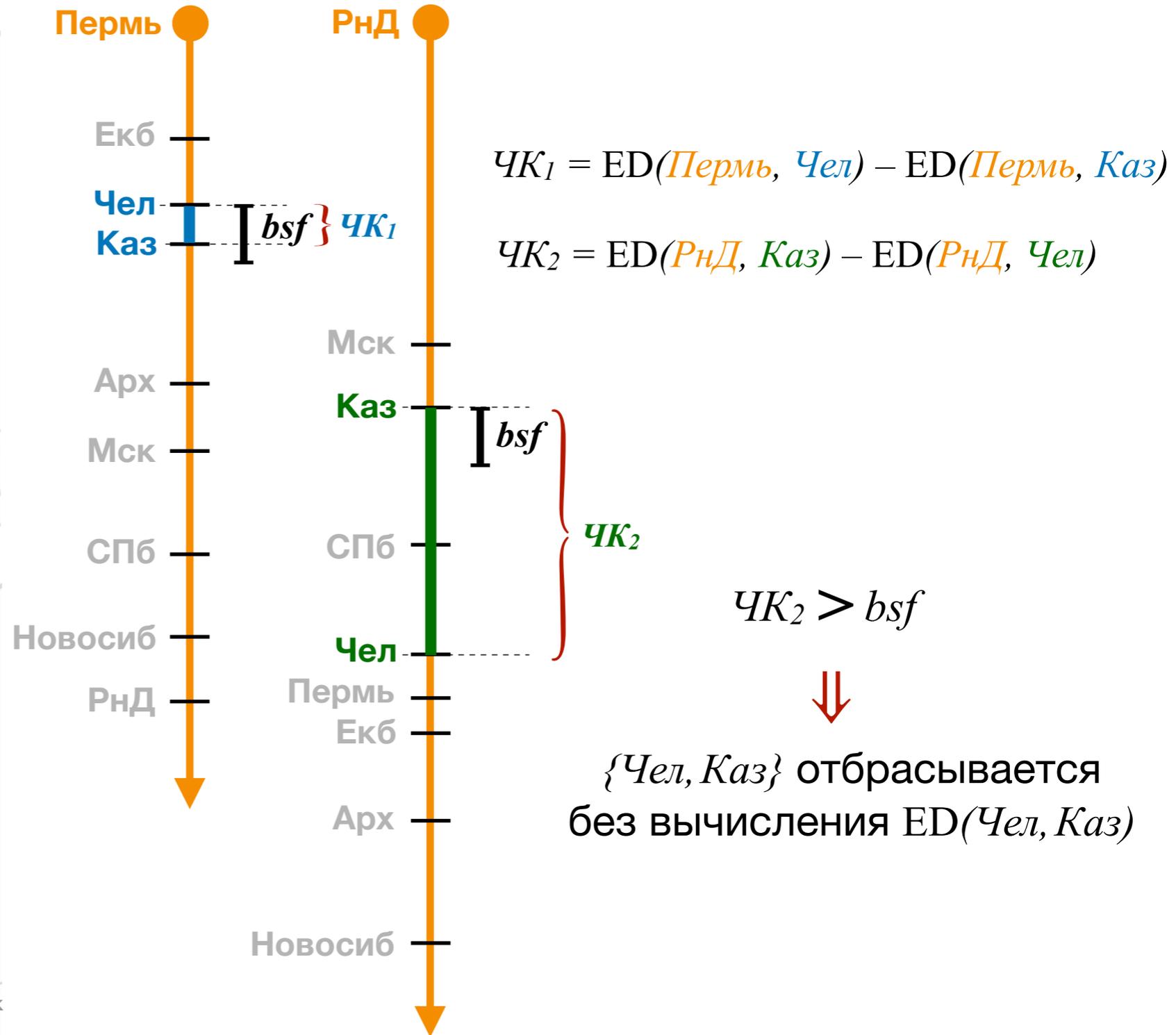
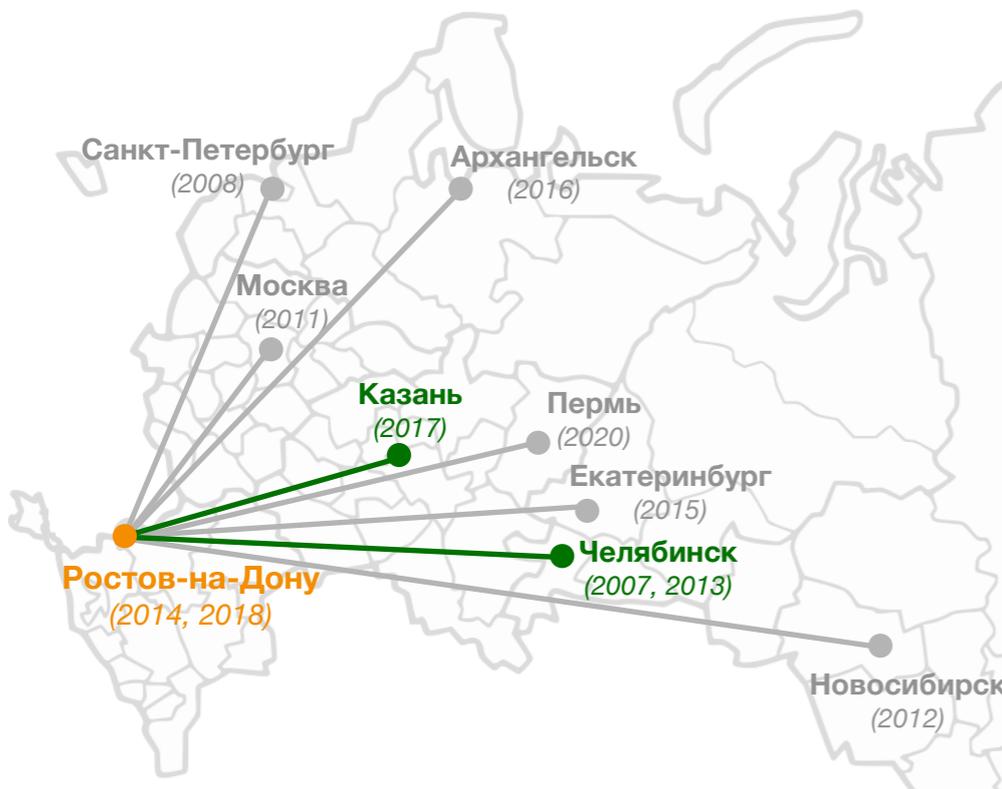
$\exists offset_{stop}$
 $\forall LB(Ref, Left, Right) > bsf \Rightarrow \forall offset_{next} \geq offset_{stop} \forall \{Left, Right\}$ отбрасывается
 без вычисления $ED(Left, Right)$

Карта проведения конференции ПаВТ

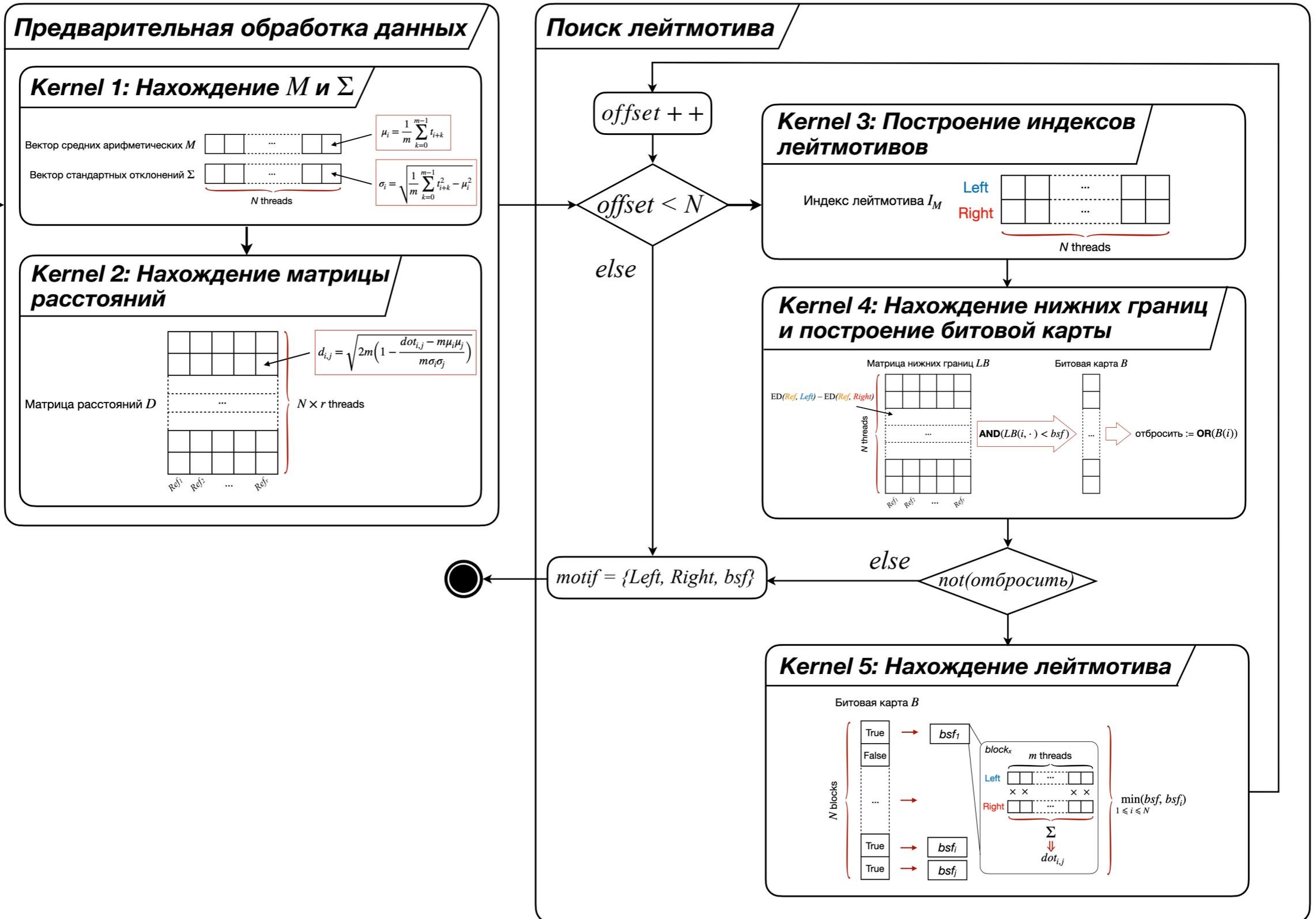


{Челябинск, Екатеринбург} – лейтмотив

Несколько опорных городов



Параллельный алгоритм для GPU



Предварительная обработка данных

$ED_{norm}(\cdot, \cdot)$ вместо $ED(\cdot, \cdot)$

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

Kernel 1: Нахождение M и Σ

Вектор средних арифметических M



$$\mu_i = \frac{1}{m} \sum_{k=0}^{m-1} t_{i+k}$$

Вектор стандартных отклонений Σ

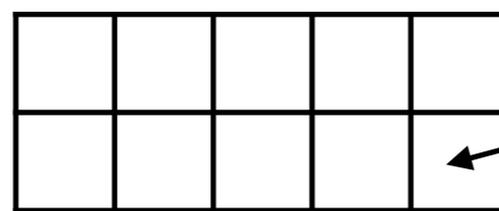


$$\sigma_i = \sqrt{\frac{1}{m} \sum_{k=0}^{m-1} t_{i+k}^2 - \mu_i^2}$$

N threads

Kernel 2: Нахождение матрицы расстояний

Матрица расстояний D



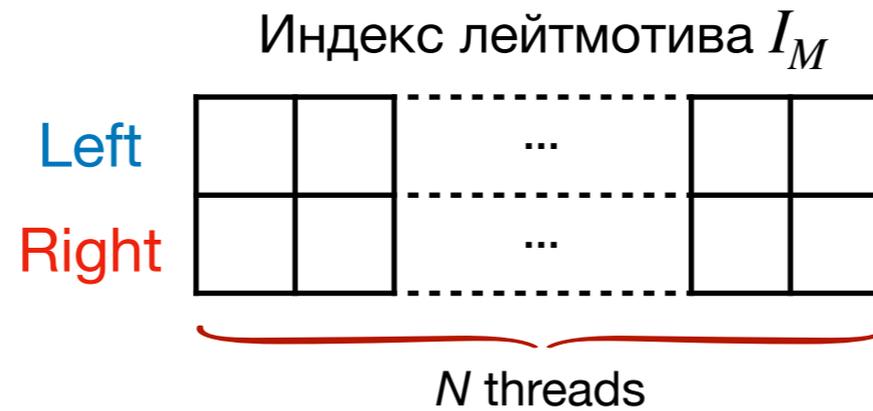
$$d_{i,j} = \sqrt{2m \left(1 - \frac{dot_{i,j} - m\mu_i\mu_j}{m\sigma_i\sigma_j} \right)}$$

$N \times r$ threads

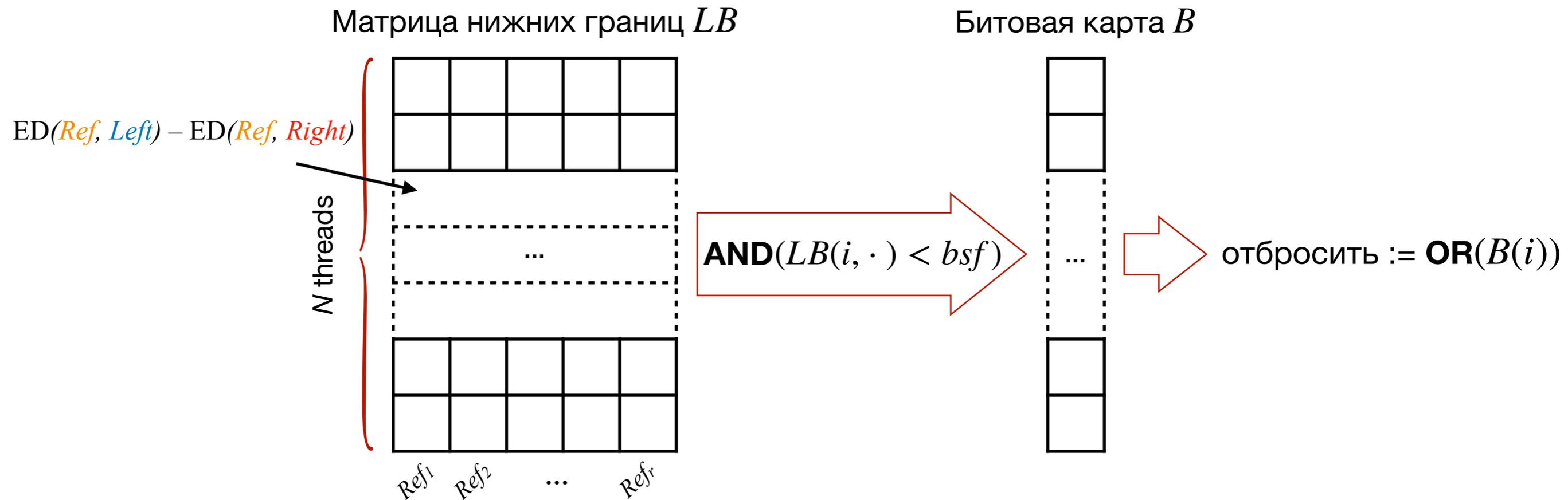
Ref₁ Ref₂ ... Ref_r

Поиск лейтмотива

Kernel 3: Построение индексов лейтмотива

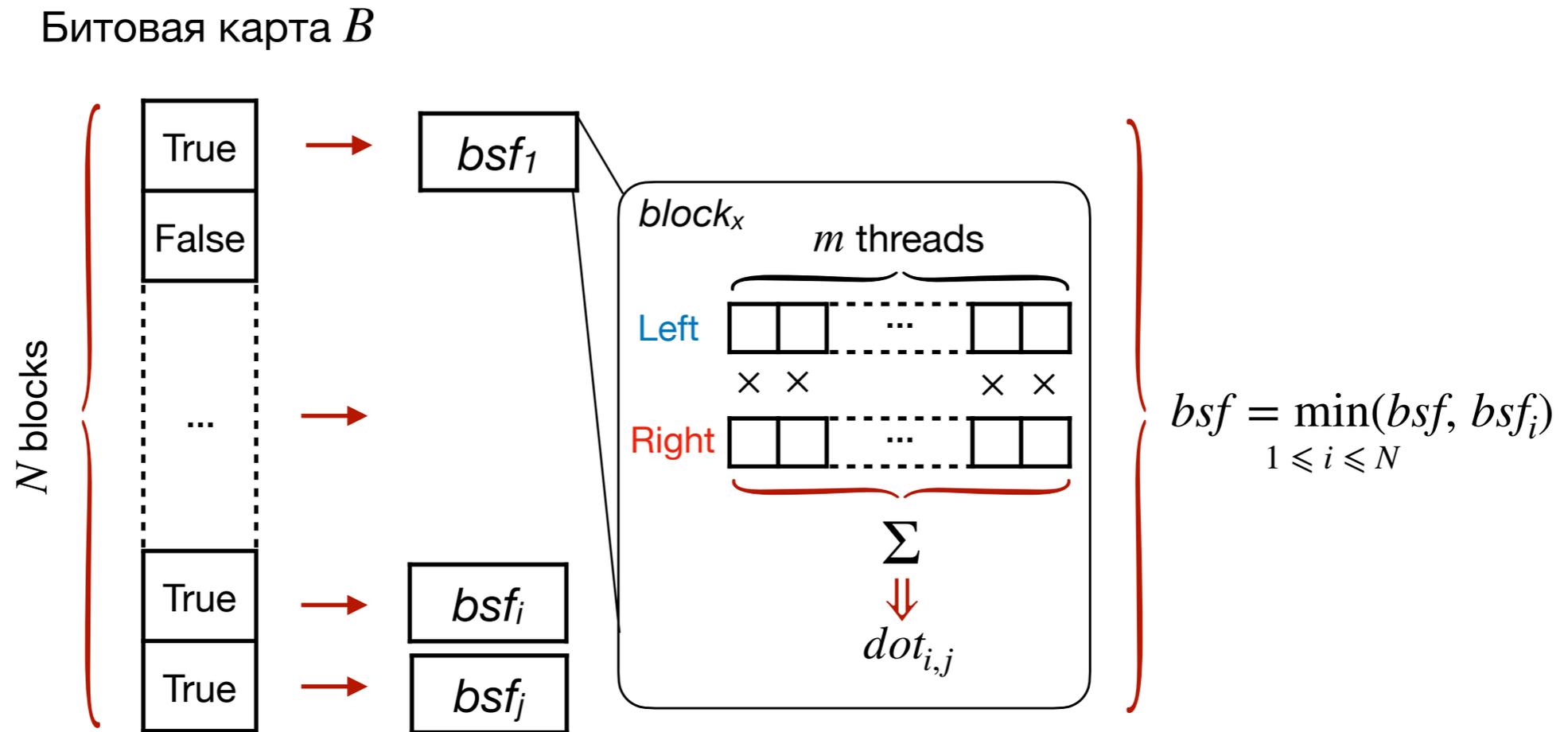


Kernel 4: Нахождение нижних границ и построение битовой карты



Поиск лейтмотива

Kernel 5: Нахождение лейтмотива



$$ED_{norm} = \sqrt{2m \left(1 - \frac{dot_{i,j} - m\mu_i\mu_j}{m\sigma_i\sigma_j} \right)}$$

- **Аппаратная платформа**

- Графический процессор: **NVIDIA GeForce GTX 2080Ti**
- Кол-во физ. ядер: **4 352 (68 мультипроцессоров)**
- Тактовая частота: **1.35 ГГц**
- Пиковая произв-ть: **11 TFLOPS**
- Память: **11 Гб**

- **Наборы данных**

- Random Walk ¹⁾: $|T| = 4 \cdot 10^5$
- ЭКГ ²⁾: $|T| = 4 \cdot 10^5$

- **Исследуемые показатели**

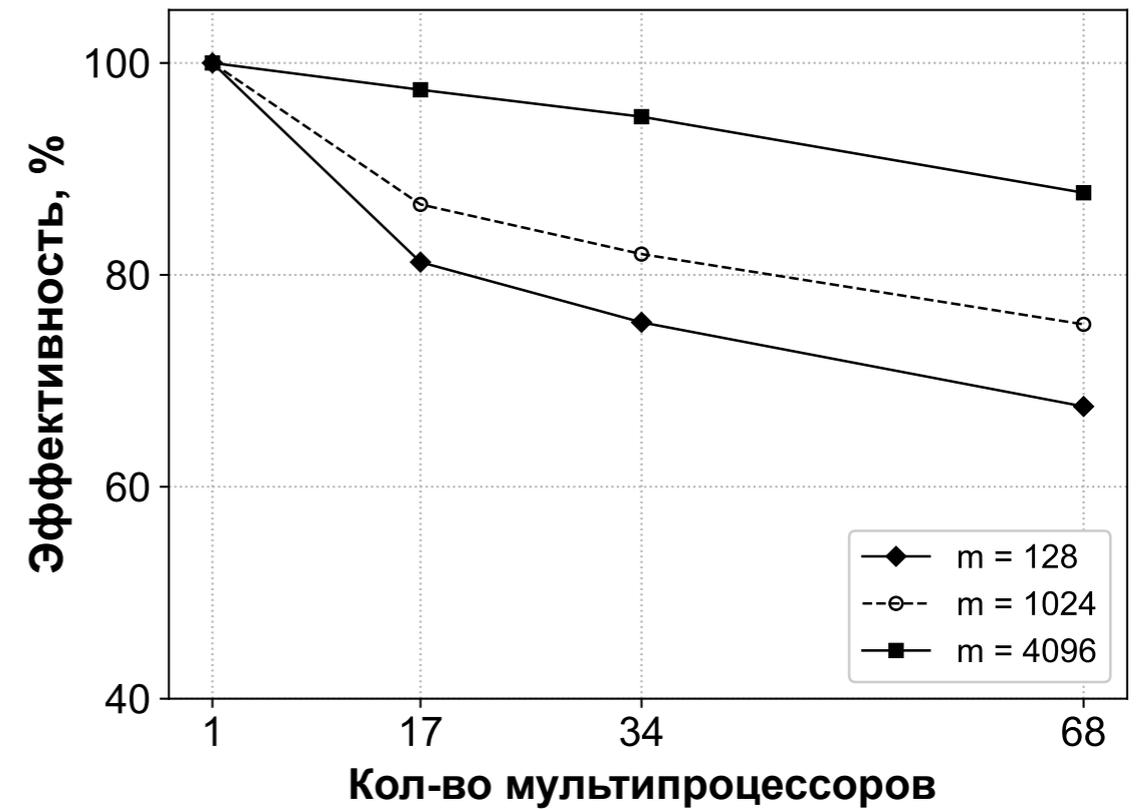
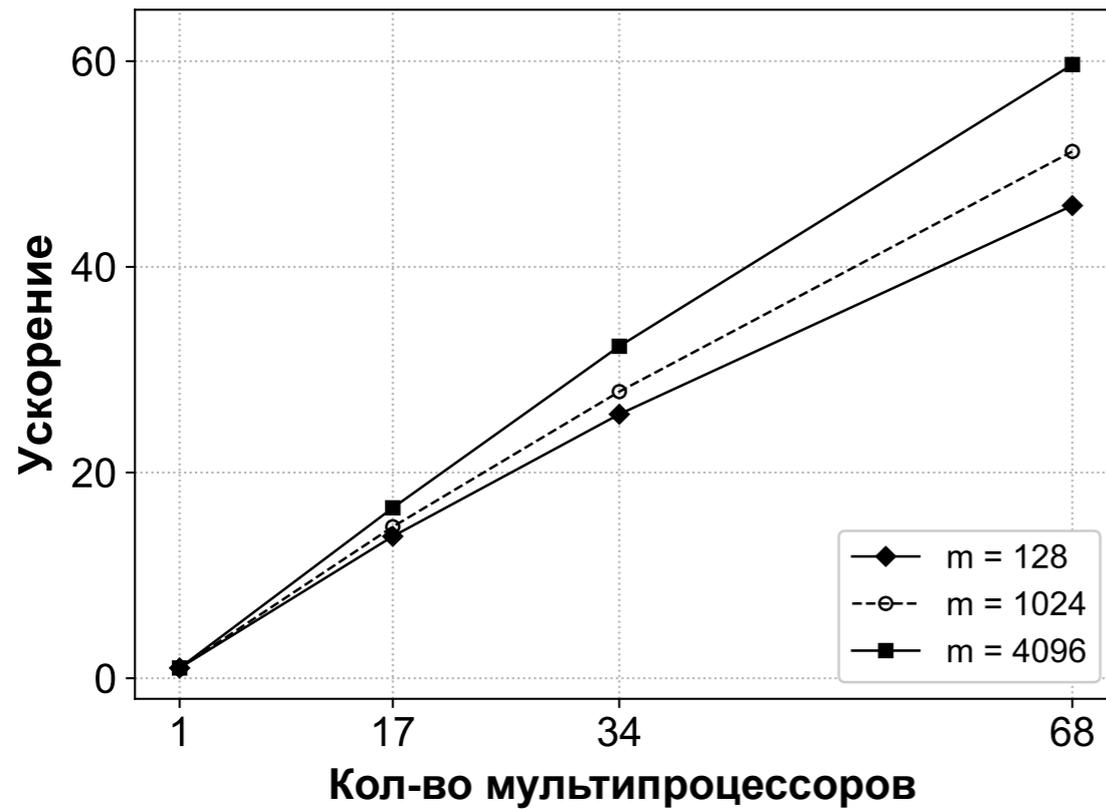
- Ускорение, $s(k) = \frac{t_1}{t_k}$, k – количество мультипроцессоров
- Эффективность, $e(k) = \frac{s(k)}{k}$

¹⁾ Mueen A., *et al.* Exact Discovery of Time Series Motifs. Proceedings of the SIAM Int. Conf. on Data Mining. 2009. pp. 473–484.

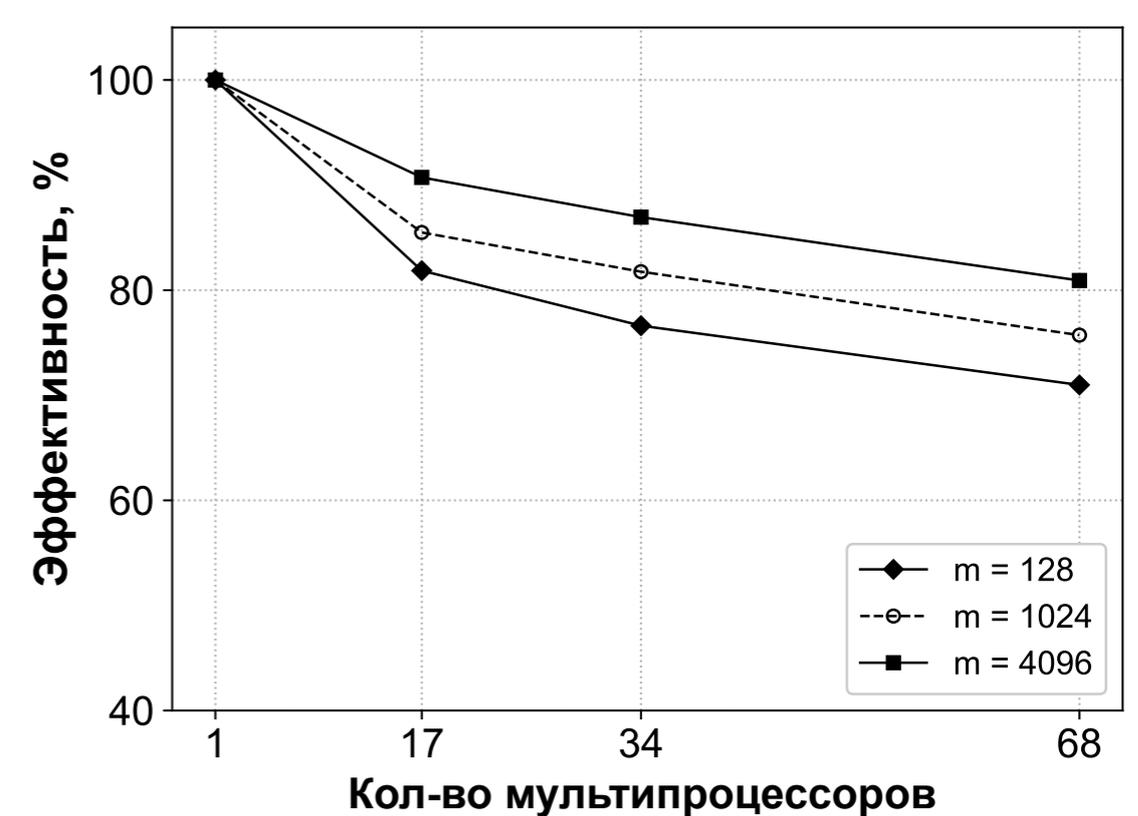
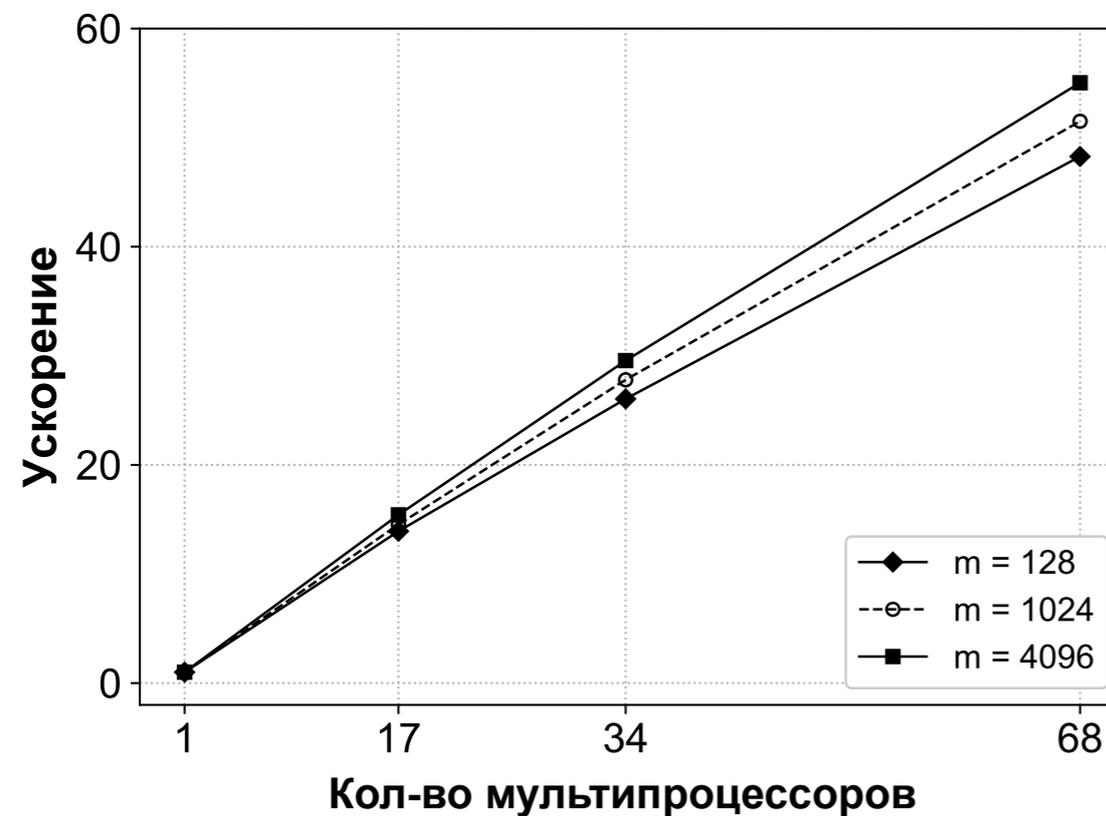
²⁾ Goldberger A.L., *et al.* PhysioBank, PhysioToolkit, and PhysioNet. Circulation. 2000. vol. 101. no. 23. pp. e215–e220.

Ускорение и эффективность

Random Walk



ЭКГ



Спасибо за внимание!

Вопросы?

Сравнение

Набор данных	Длина временного ряда	Длина лейтмотива	Время работы алгоритма на указанных платформах, с		
			МК 1)	PhiМК 2)	GPUMK
			Host (1 физ. ядро), 0.371 TFLOPS	Intel Xeon Phi, (244 нити), 1.076 TFLOPS	NVIDIA Turing, (400 ядер), 1.076 TFLOPS
Random Walk	$4 \cdot 10^5$	1024	13 275	662	245
ЭКГ			20 084	1 077	356

1) Mueen A., *et. al.* Exact discovery of time series motifs. Proceedings of the SIAM International Conference on Data Mining. 2009, pp. 473–484.

2) Zymbler M., Kraeva Ya. Discovery of Time Series Motifs on Intel Many-Core Systems. Lobachevskii Journal of Mathematics. 2019. vol. 40, no. 12. pp. 2124–2132.