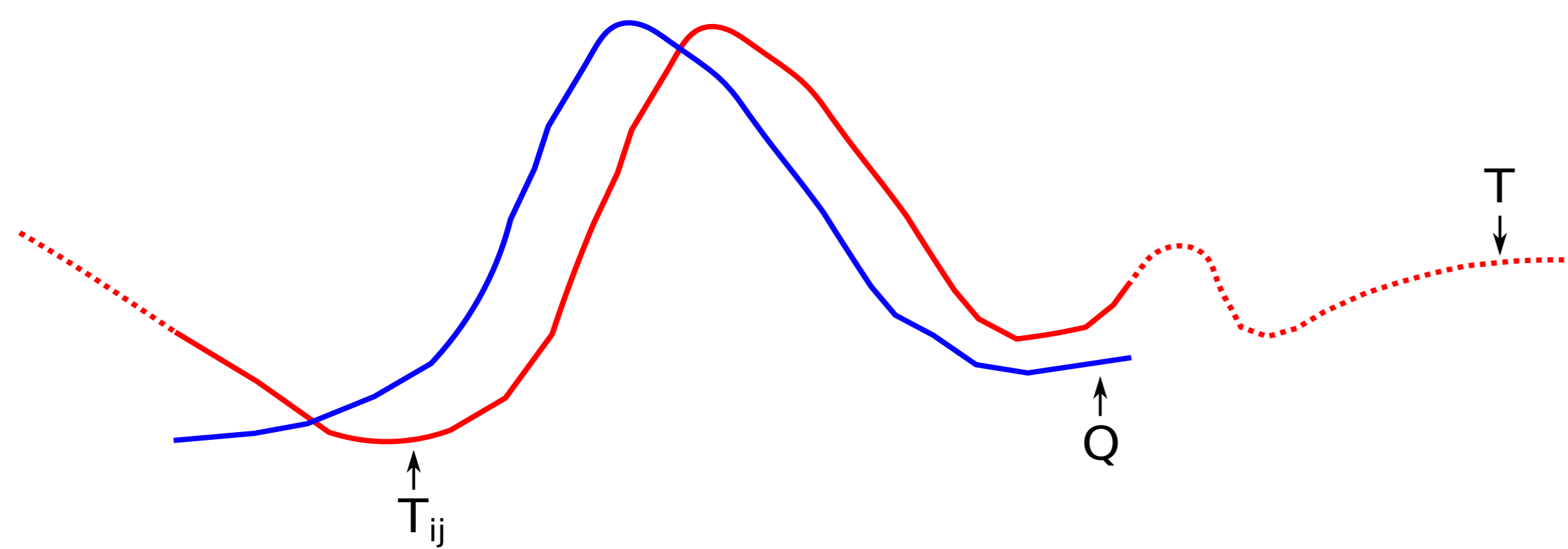


Разработка параллельного алгоритма поиска похожих подпоследовательностей временного ряда для сопроцессора Intel Xeon Phi*

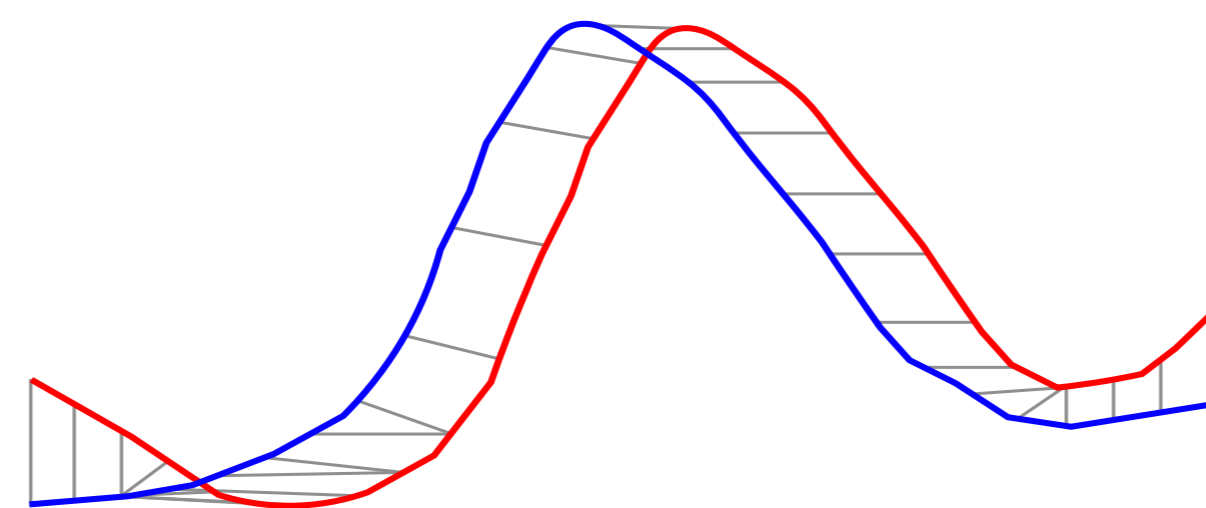
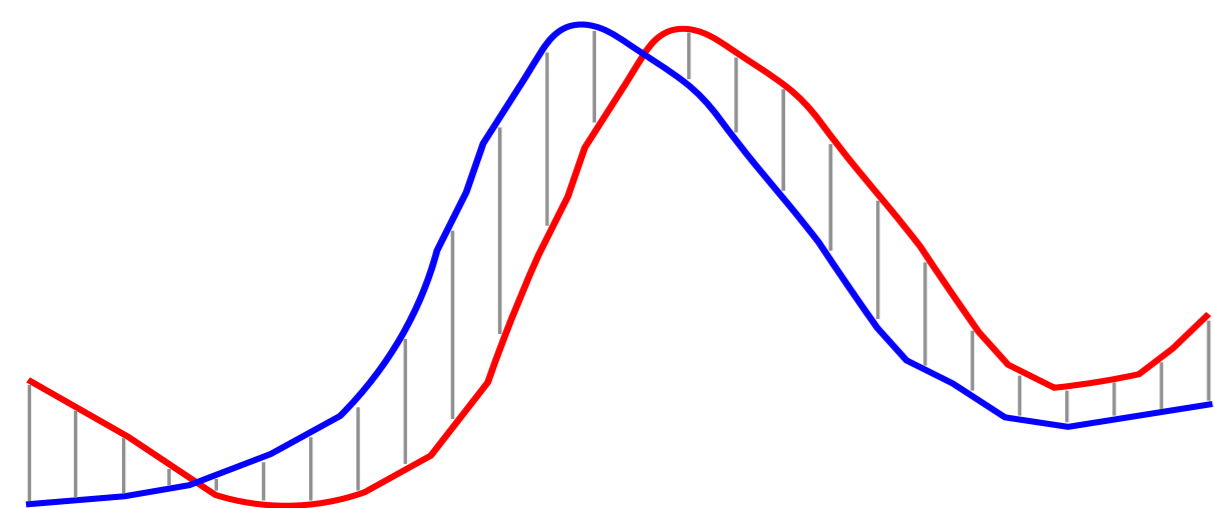
А.В. Мовчан, М.Л. Цымблер
Южно-Уральский государственный университет

Поиск похожих подпоследовательностей
поиск подпоследовательностей T_{ij} временного ряда T , которые больше всего похожи на другой, значительно меньший временной ряд Q , называемый запросом.



Евклидова метрика

Динамическая трансформация времени (DTW)



Последовательный алгоритм

Точный последовательный поиск с оптимизациями:

1. Оценки LB_Kim , LB_Keogh

$$L_i = \max(Q_{i-R}:Q_{i+R}) \quad U_i = \max(Q_{i-R}:Q_{i+R})$$

$$LB_{keogh}(Q, C) = \sqrt{\sum_{i=1}^n \begin{cases} (C_i - U_i)^2, & \text{если } C_i > U_i \\ (C_i - L_i)^2, & \text{если } C_i < L_i \\ 0, & \text{в остальных случаях} \end{cases}}$$

2. Заблаговременная отмена расчета LB_Keogh

3. Заблаговременная отмена расчета DTW

4. Заблаговременная отмена расчета Z-нормализации

$$\mu = \frac{1}{m} \sum x_i \quad \sigma^2 = \frac{1}{m} \sum x_i^2 - \mu^2 \quad x_{norm} = \frac{x - \mu}{\sigma}$$

5. Переупорядочивание расчета LB_Keogh

6. Обмен ролей запроса и данных в LB_Keogh ($LB_KeoghEC$)

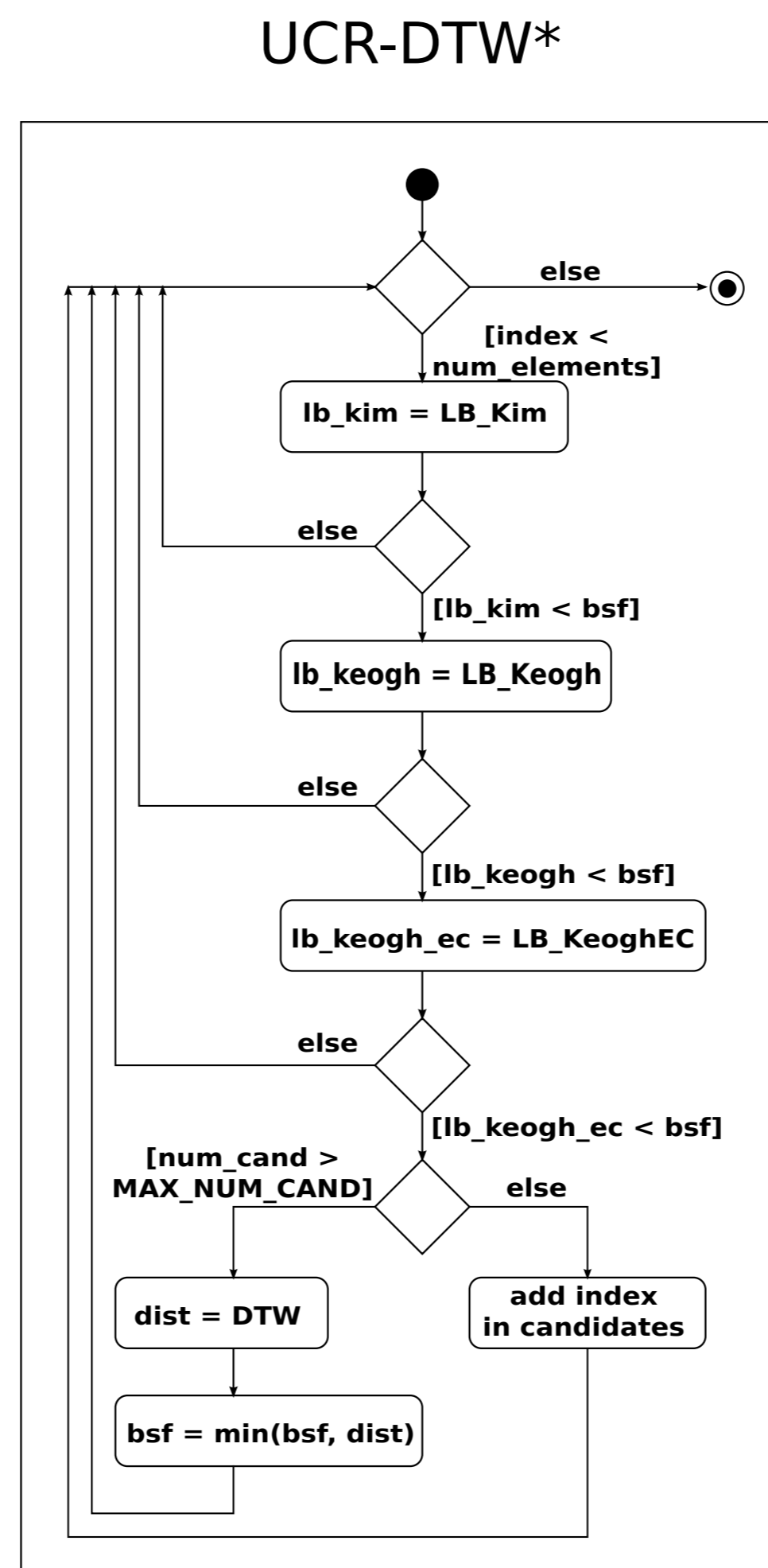
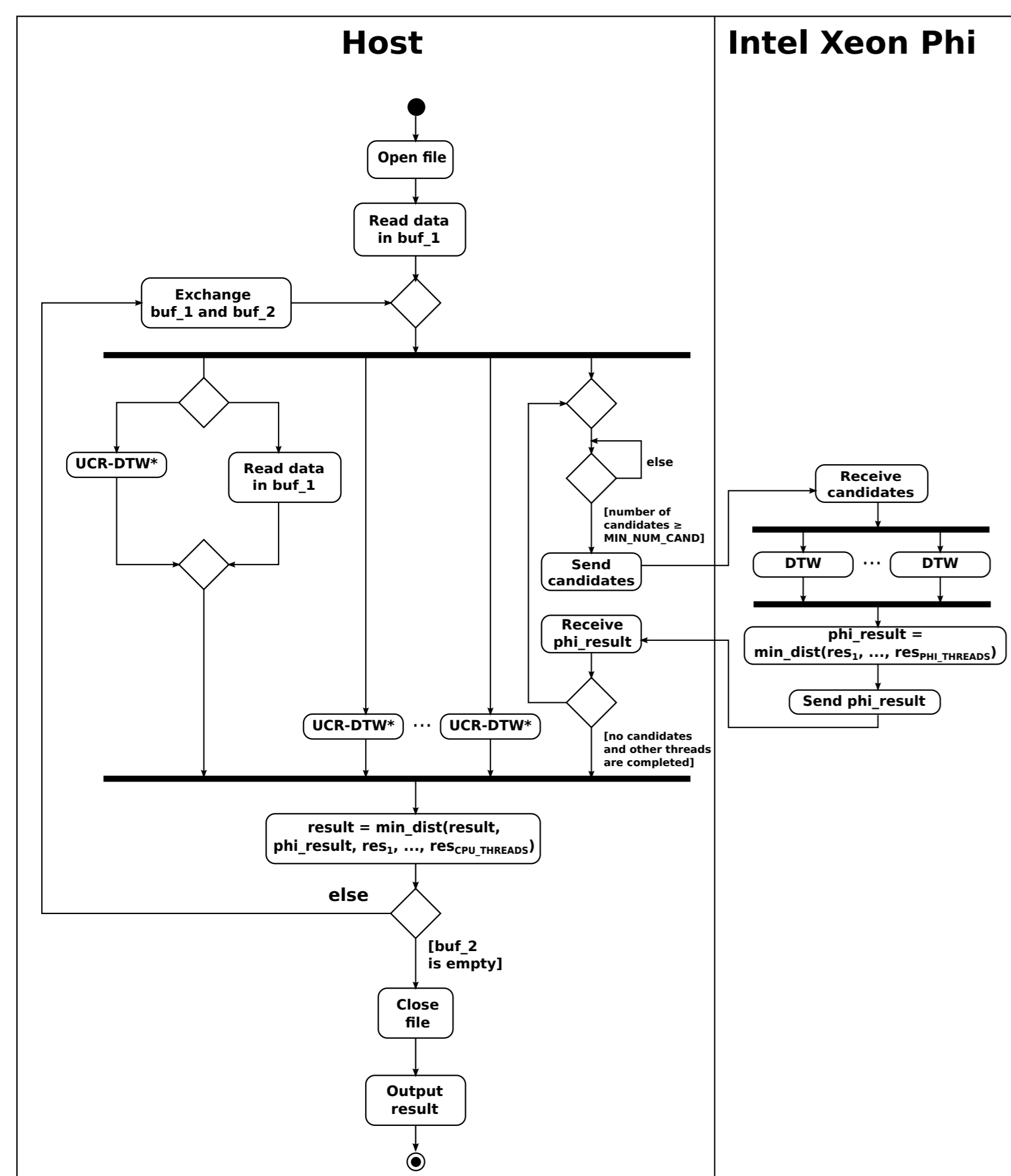
$$L_i = \max(C_{i-R}:C_{i+R}) \quad U_i = \max(C_{i-R}:C_{i+R})$$

$$LB_{keoghEC}(Q, C) = \sqrt{\sum_{i=1}^n \begin{cases} (Q_i - U_i)^2, & \text{если } Q_i > U_i \\ (Q_i - L_i)^2, & \text{если } Q_i < L_i \\ 0, & \text{в остальных случаях} \end{cases}}$$

7. Каскадное применение оценок

- 1) LB_Kim
- 2) LB_Keogh
- 3) $LB_KeoghEC$
- 4) DTW

Параллельный алгоритм



Эксперименты

Процессор: Intel Xeon X5680 (6 ядер по 3.33 GHz), 0,371 Тфлопс
Сопроцессор: Intel Xeon Phi SE10X (61 ядро по 1.1 GHz) 1,076 Тфлопс



* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 12-07-00443-а.