

Международная научная конференция
Параллельные вычислительные технологии (ПаВТ'2019)
Калининград, 2–4 апреля 2019

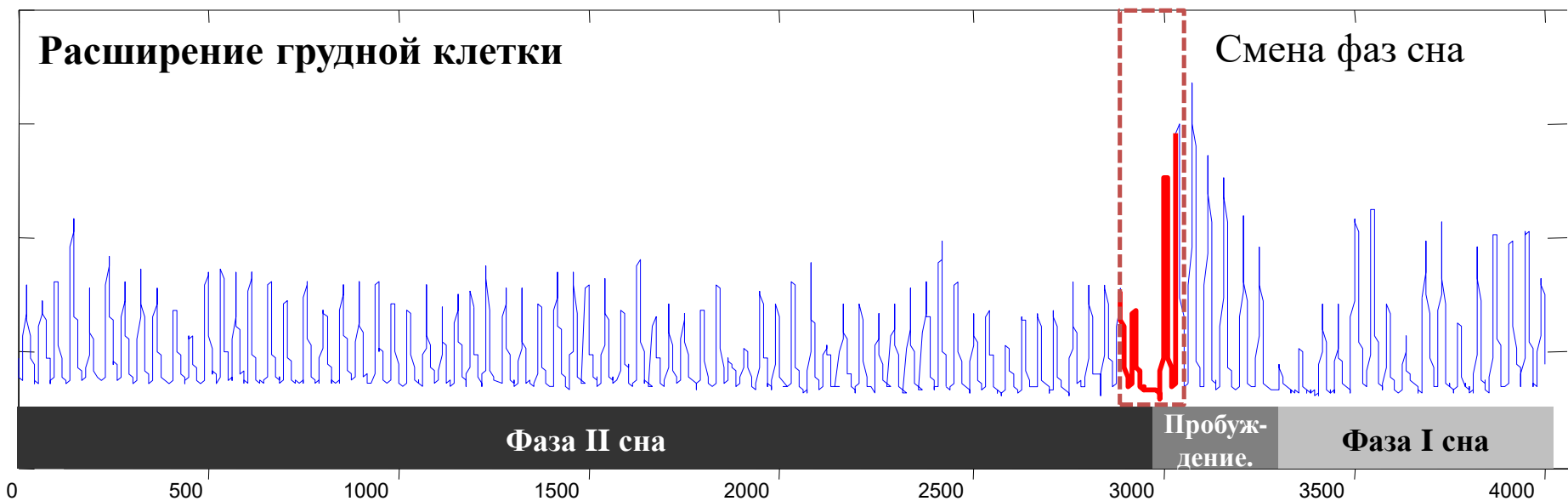
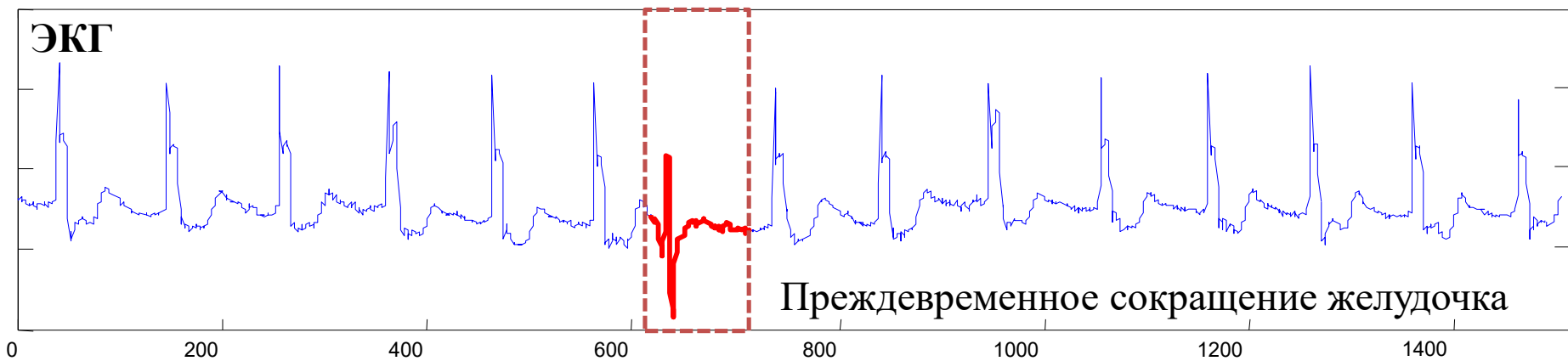
Параллельный алгоритм поиска диссонансов временного ряда для ускорителей Intel MIC

А.В. Поляков, М.Л. Цымблер

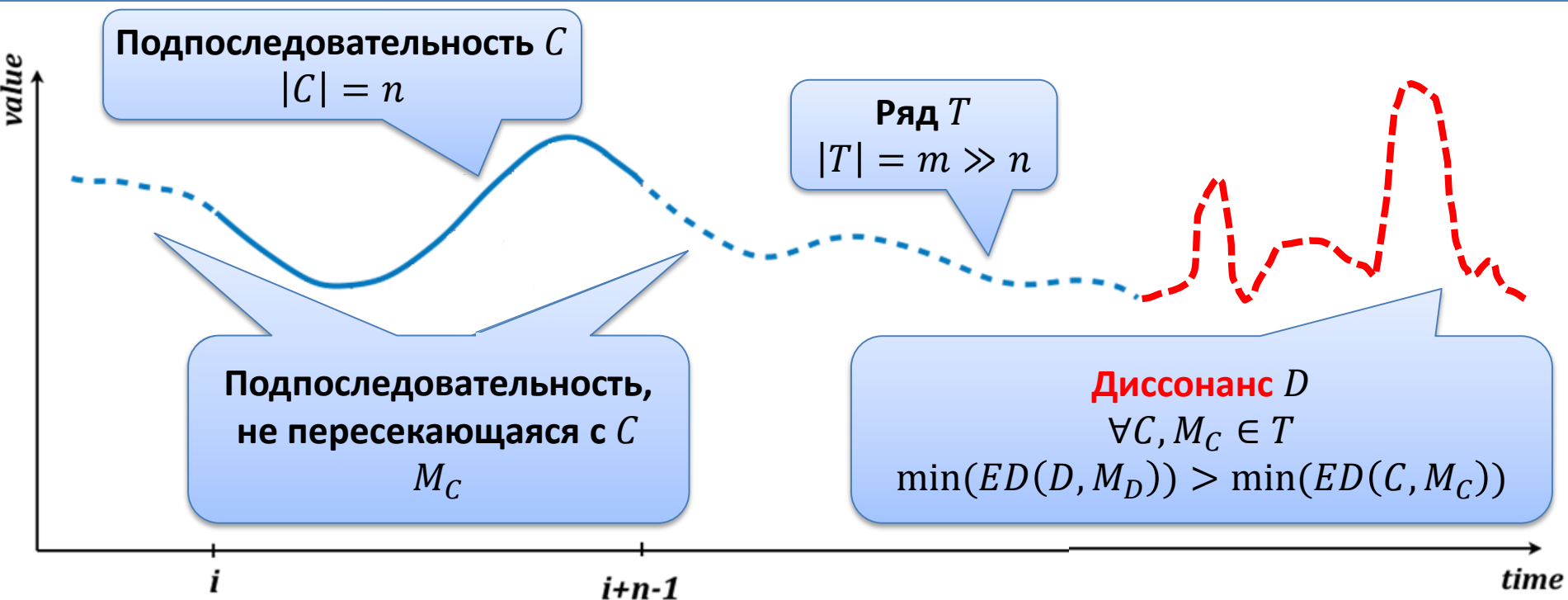
Южно-Уральский государственный университет (Челябинск)

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 17-07-00463), Правительства РФ в соответствии с Постановлением № 211 от 16.03.2013 (соглашение № 02.A03.21.0011) и Министерства образования и науки РФ (государственное задание 2.7905.2017/8.9).

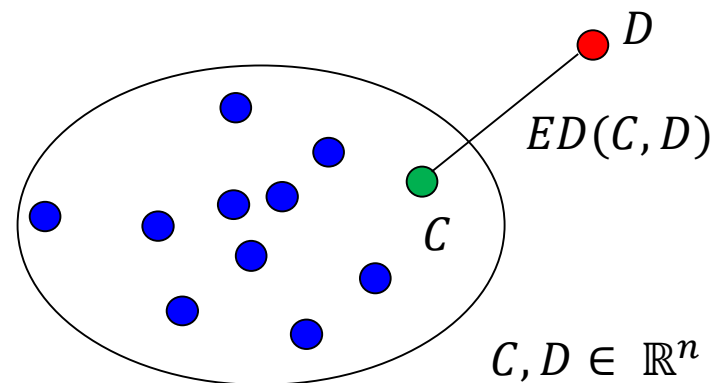
Аномалии во временных рядах



Диссонанс временного ряда

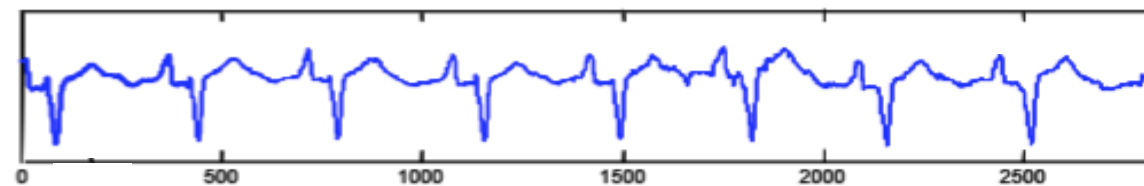


Количество подпоследовательностей
длины n в ряде T : $N = m - n + 1$



Keogh E.J., et al. **HOT SAX**: efficiently finding the most unusual time series subsequence. Proc. of the 5th IEEE Int. Conf. on Data Mining. P. 8.

Последовательный поиск диссонанса



z-нормализация

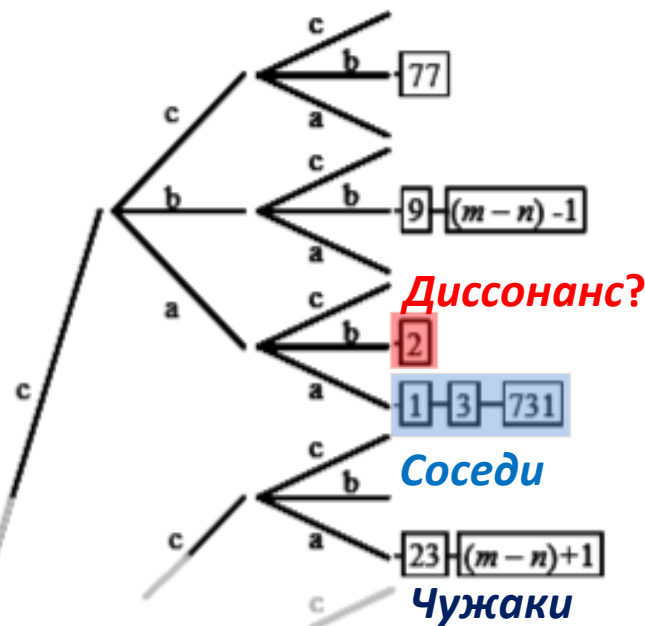
Сжатие (PAA, Piecewise Aggregate Approximation)

Кодирование (SAX, Symbolic Aggregate Approximation)

Частотный индекс слов

1	c	a	a	3
2	c	a	b	1
3	c	a	a	3
::	::	::	::	::
::	::	::	::	::
$(m-n)-1$	c	b	b	2
$(m-n)$	a	c	b	1
$(m-n)+1$	b	c	a	2

Префиксное дерево



Алгоритм HOTSAX

```

 $dist_{bsf} \leftarrow 0; dist_{min} \leftarrow \infty$ 
for  $C_i \in$  Диссонансы?, Остальные
  for  $C_j \in$  Соседи, Чужаки
     $d \leftarrow ED(C_i, C_j)$ 
    if  $d < dist_{bsf}$ 
      break
     $dist_{min} \leftarrow \min(d, dist_{min})$ 
   $dist_{bsf} \leftarrow \max(dist_{min}, dist_{bsf})$ 
   $pos_{bsf} \leftarrow i$ 
return  $\{pos_{bsf}, dist_{bsf}\}$ 

```

Параллельный поиск диссонанса

Цикл можно
разбить на два

Цикл можно
разбить на два

Можно
использовать
квадрат ED

```
 $dist_{bsf} \leftarrow 0; dist_{min} \leftarrow \infty$   
for  $C_i \in$  Диссонансы?, Остальные  
  for  $C_j \in$  Соседи, Чужаки  
     $d \leftarrow ED(C_i, C_j)$   
    if  $d < dist_{bsf}$   
      break  
     $dist_{min} \leftarrow \min(d, dist_{min})$   
   $dist_{bsf} \leftarrow \max(dist_{min}, dist_{bsf})$   
   $pos_{bsf} \leftarrow i$   
return  $\{pos_{bsf}, dist_{bsf}\}$ 
```

Параллельный алгоритм PhiDD

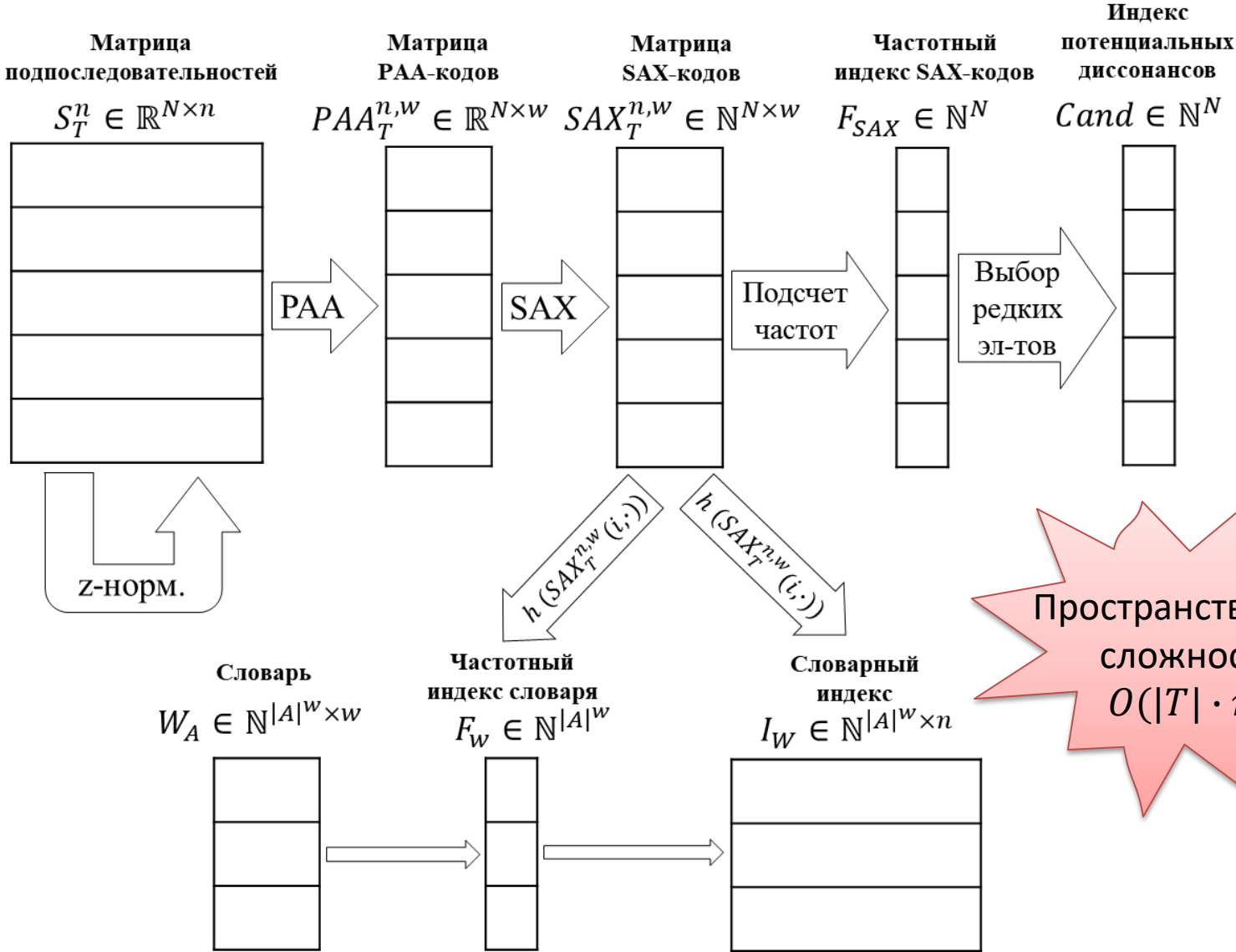
Поиск кандидатов

```
 $dist_{bsf} \leftarrow 0; dist_{min} \leftarrow \infty$   
for  $C_i \in$  Диссонансы?  
  PARALLEL for  $C_j \in$  Соседу  
     $d \leftarrow ED^2(C_i, C_j)$   
    if  $d < dist_{bsf}$   
      break  
     $dist_{min} \leftarrow \min(d, dist_{min})$   
  PARALLEL for  $C_j \in$  Чужаки  
     $d \leftarrow ED^2(C_i, C_j)$   
    if  $d < dist_{bsf}$   
      break  
     $dist_{min} \leftarrow \min(d, dist_{min})$   
   $dist_{bsf} \leftarrow \max(dist_{min}, dist_{bsf})$   
  if  $dist_{bsf} < dist_{min}$  then  $pos_{bsf} \leftarrow i$   
return  $\{pos_{bsf}, dist_{bsf}\}$ 
```

Уточнение

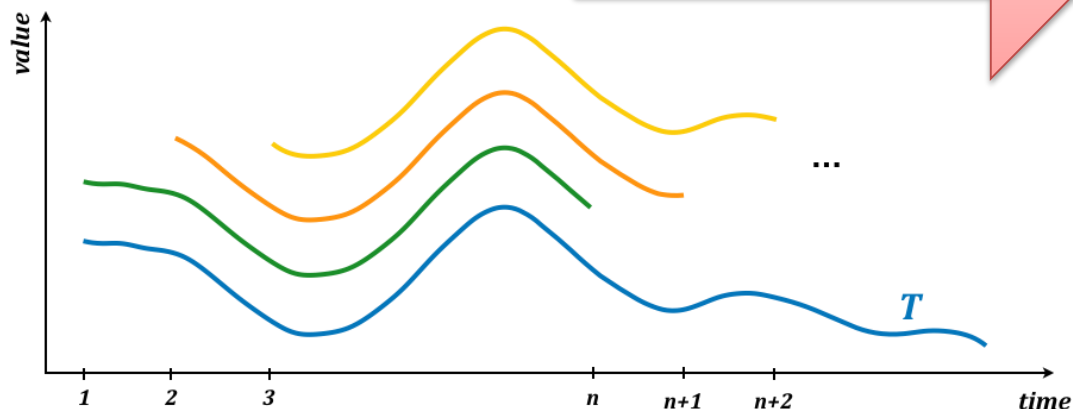
```
 $dist_{bsf} \leftarrow 0; dist_{min} \leftarrow \infty$   
PARALLEL for  $C_i \in$  Остальные  
  for  $C_j \in$  Соседу  
     $d \leftarrow ED^2(C_i, C_j)$   
    if  $d < dist_{bsf}$   
      break  
     $dist_{min} \leftarrow \min(d, dist_{min})$   
  for  $C_j \in$  Чужаки  
     $d \leftarrow ED^2(C_i, C_j)$   
    if  $d < dist_{bsf}$   
      break  
     $dist_{min} \leftarrow \min(d, dist_{min})$   
   $dist_{bsf} \leftarrow \max(dist_{min}, dist_{bsf})$   
  if  $dist_{bsf} < dist_{min}$  then  $pos_{bsf} \leftarrow i$   
return  $\{pos_{bsf}, \sqrt{dist_{bsf}}\}$ 
```

Матричное представление данных



Матрица подпоследовательностей

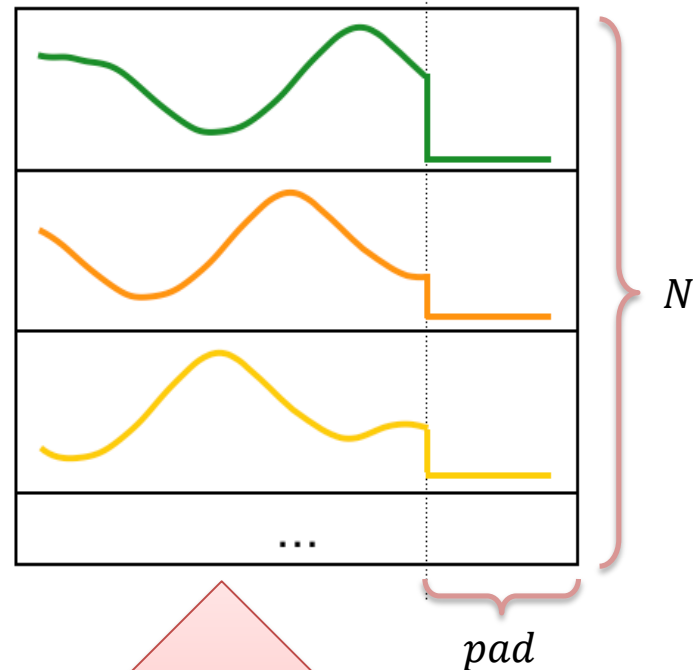
$$T \in \mathbb{R}^m$$



Выравнивание

$$(n + pad) : width_{VPU}$$

$$S \in \mathbb{R}^{N \times (n+pad)}$$



z-нормализация

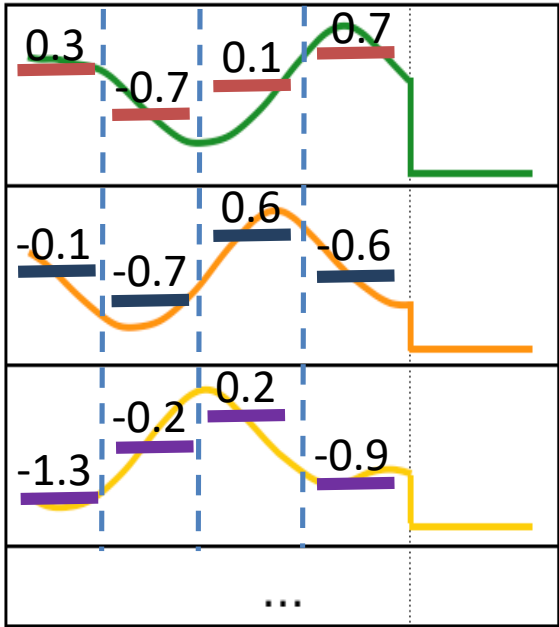
$$\hat{S} = (\hat{s}_1, \hat{s}_2, \dots, \hat{s}_n)$$

$$\hat{s}_i = \frac{s_i - \mu}{\sigma}, \quad \mu = \frac{1}{n} \sum_{i=1}^n s_i, \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n s_i^2 - \mu^2$$

РАА-сжатие

Матрица
подпоследовательностей

$$S \in \mathbb{R}^{N \times (n+pad)}$$



Piecewise Aggregate Approximation

$$PAA(i, k) = \frac{w}{n} \sum_{j=(\frac{n}{w})(i-1)+1}^{(\frac{n}{w})i} S(k, j)$$

Матрица
РАА-кодов

$$PAA \in \mathbb{R}^{N \times w}$$

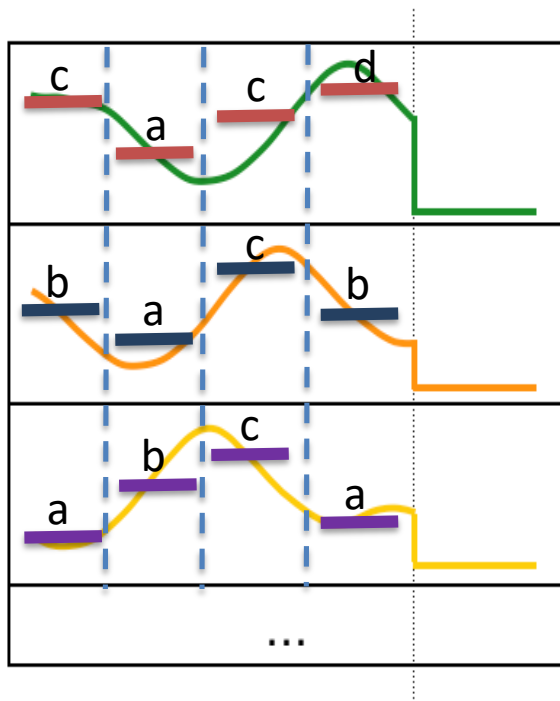
0.3	-0.7	0.1	0.7
-0.1	-0.7	0.6	-0.6
-1.3	-0.2	0.2	-0.9
...			

w – степень сжатия
(обычно 4)

SAX-кодирование

Матрица
подпоследовательностей

$$S \in \mathbb{R}^{N \times (n+pad)}$$



Матрица
SAX-кодов

$$SAX \in \mathbb{N}^{N \times w}$$

c	a	c	d
b	a	c	b
a	b	c	a
...			

Symbolic Aggregate Approximation

$(-\infty; -0.67)$	$[-0.67; 0)$	$[0; 0.67)$	$[0.67; \infty)$
a	b	c	d

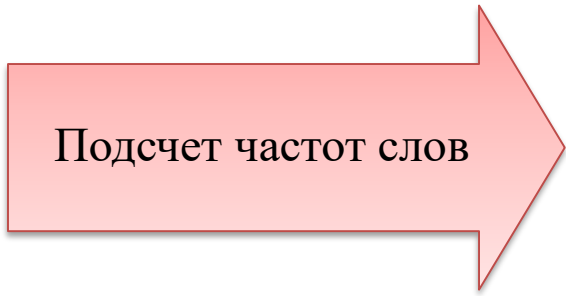
Таблица кодирования в алфавите A
(обычно $|A|=4$)

Частотный индекс и индекс потенциальных диссонансов

**Матрица
SAX-кодов**

$$SAX \in \mathbb{N}^{N \times w}$$

1	a	b	d	d
2	c	d	a	b
3	a	b	d	c
	...			
<i>j</i>	a	b	d	c
<i>l</i>	c	d	a	b
<i>u</i>	c	d	a	b
<i>k</i>	a	b	d	d



**Частотный
индекс
SAX-кодов**

$$F_{SAX} \in \mathbb{N}^N$$

2
3
2
...
2
3
3
2
...



**Индекс
потенциальных
диссонансов**

$$Cand \in \mathbb{N}^N$$

1
3
<i>j</i>
<i>k</i>
-
-
-
-

Построение словаря и его индексов

Словарь

$$W_A \in \mathbb{N}^{|A|^w \times w}$$

a	a	a	a
a	a	a	b
...			
a	b	d	c
a	b	d	d
...			
c	d	a	b
...			
d	d	d	d

Построение индексов

Матрица

SAX-кодов

$$SAX \in \mathbb{N}^{N \times w}$$

1	a	b	d	d
2	c	d	a	b
3	a	b	d	c
...				
j	a	b	d	c
ℓ	c	d	a	b
u	c	d	a	b
k	a	b	d	d
...				

Частотный индекс

$$F_W \in \mathbb{N}^{|A|^w}$$

0
0
...
2
2
...
3
...
0

Словарный индекс

$$I_W \in \mathbb{N}^{|A|^w \times N}$$

...				
...				
...				
3	j		...	
1	k		...	
...				
2	ℓ	u	...	
...				
...				

Размещения символов алфавита A по w символов с повторениями

Индексный доступ к подпоследовательностям

**Индекс
потенциальных
диссонансов**

**Матрица
SAX-кодов**

**Частотный
индекс**

**Словарный
индекс**

$Cand \in \mathbb{N}^N$

$SAX \in \mathbb{N}^{N \times w}$

$F_W \in \mathbb{N}^{|A|^w}$

$I_W \in \mathbb{N}^{|A|^w \times N}$

1	1	a	b	d	d
3	2	c	d	a	b
j	3	a	b	d	c
k		...			
-	j	a	b	d	c
-	ℓ	c	d	a	b
-	u	c	d	a	b
-	k	a	b	d	d
-					
...		...			

$h(a_1, a_2, \dots, a_w) = \sum_{j=1}^{w+1} a_j \cdot w^{w-j-1}$

0
0
...
2
2
...
3
...
0

			...	
			...	
...				
3	j		...	
1	k		...	
...				
2	ℓ	u	...	
...				
			...	

Индексный доступ к подпоследовательностям

Индекс
потенциальных
диссонансов

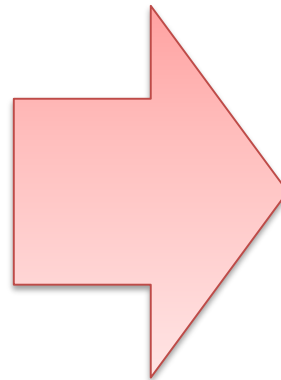
Словарный
индекс

Матрица
подпоследовательностей

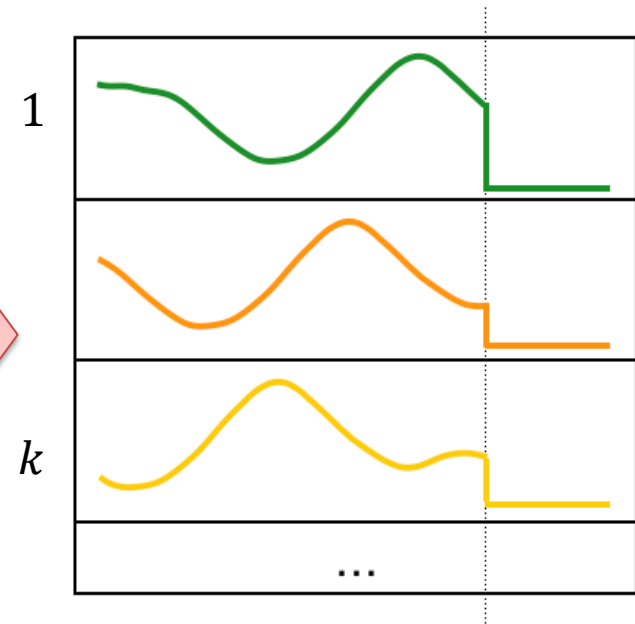
$$Cand \in \mathbb{N}^N \quad I_W \in \mathbb{N}^{|A|^w \times N}$$

1
3
j
k
-
-
-
-
-
...

			...	
			...	
...				
3	j		...	
1	k		...	
...				
2	ℓ	u	...	
...				
			...	



$$S \in \mathbb{R}^{N \times (n+pad)}$$



Поиск потенциального диссонанса

Словарный индекс
 $I_W \in \mathbb{N}^{|A|^w \times n}$

$I_W(I_{neighbor}(i), \cdot)$

$I_{neighbor}(i) = h(SAX_T^{n,w}(Cand(i), \cdot))$
 $I_{cands}(i, \cdot) = I_W(I_{neighbor}(i), \cdot)$

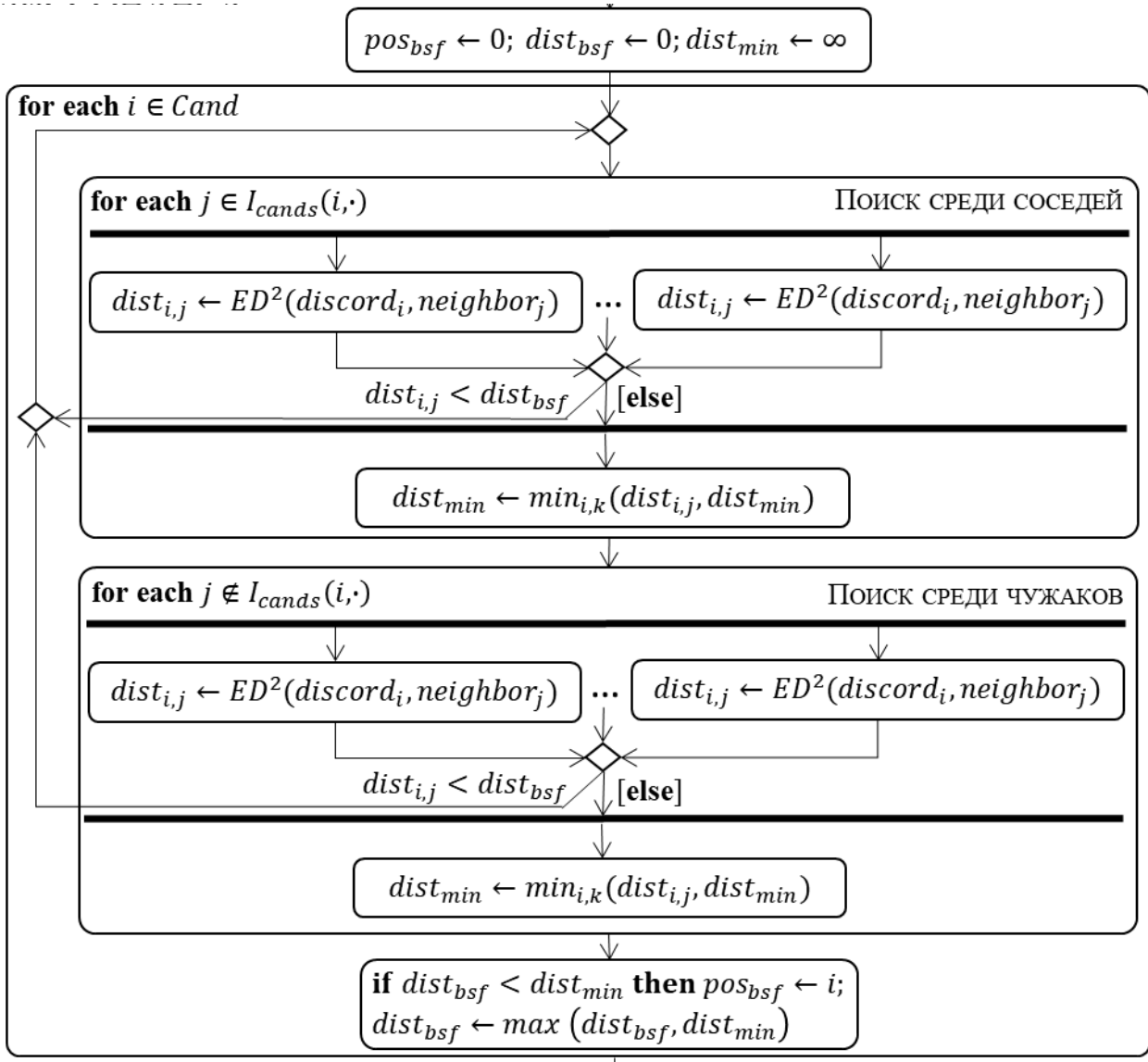
Матрица
 подпоследовательностей
 $S_T^n \in \mathbb{R}^{N \times n}$

$S_T^n(I_{cands}(i, j), \cdot)$
$S_T^n(i, \cdot), i \in Cand$
$S_T^n(k, \cdot), k \notin I_W$

«Сосед»
neighbor

Предположительный
 диссонанс
discord

«Чужак»
stranger



Уточнение диссонанса

Словарный индекс

$$I_W \in \mathbb{N}^{|A|^w \times n}$$

$I_W(I_{neighbor}(i), \cdot)$

$$I_{neighbor}(i) = h(SAX_T^{n,w}(Cand(i), \cdot))$$

$$I_{cands}(i, \cdot) = I_W(I_{neighbor}(i), \cdot)$$

Матрица подпоследовательностей

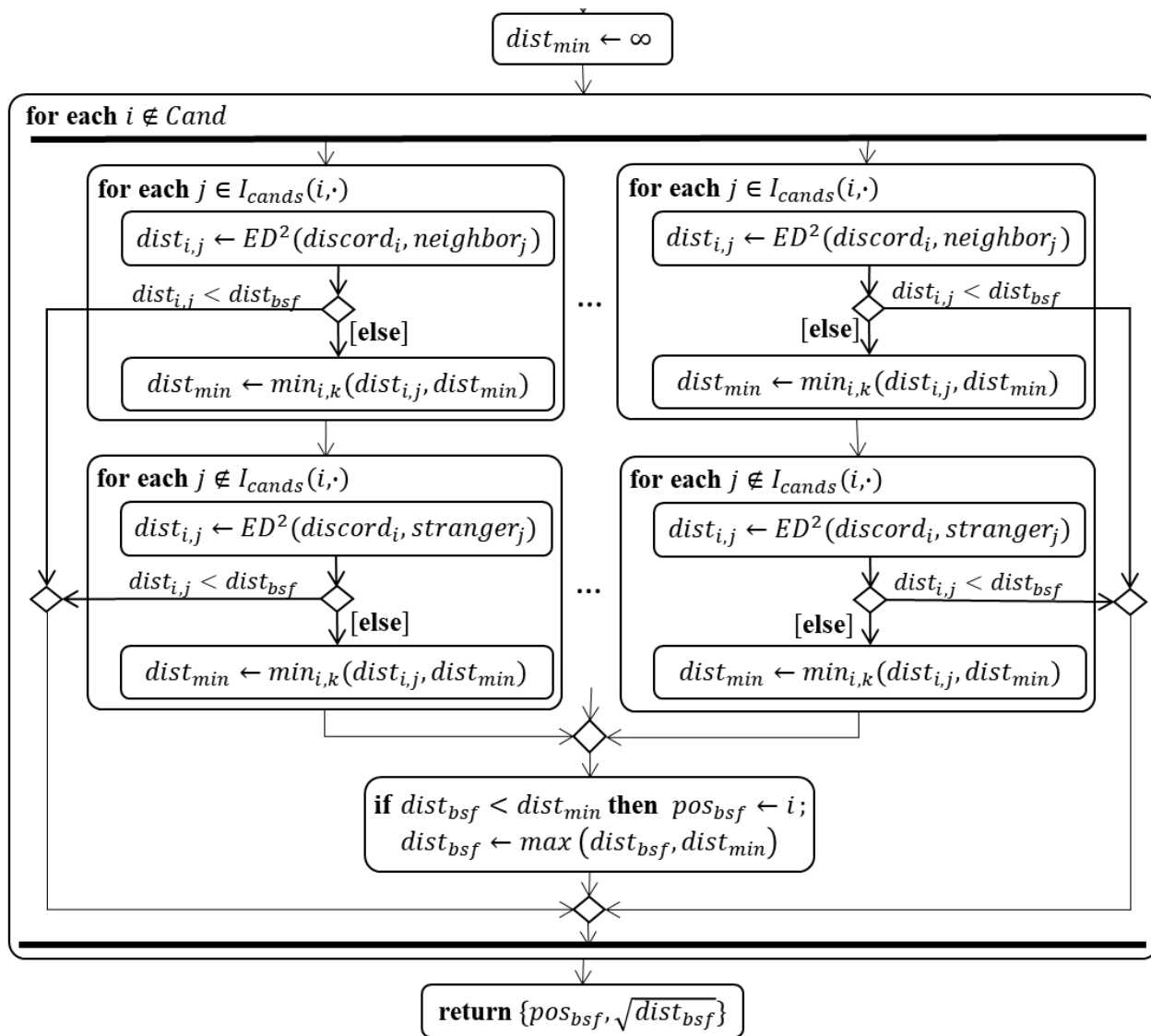
$$S_T^n \in \mathbb{R}^{N \times n}$$

$S_T^n(I_{cands}(i, j), \cdot)$
$S_T^n(i, \cdot), i \in Cand$
$S_T^n(k, \cdot), k \notin I_W$

«Сосед»
neighbor

Предположительный диссонанс
discord

«Чужак»
stranger



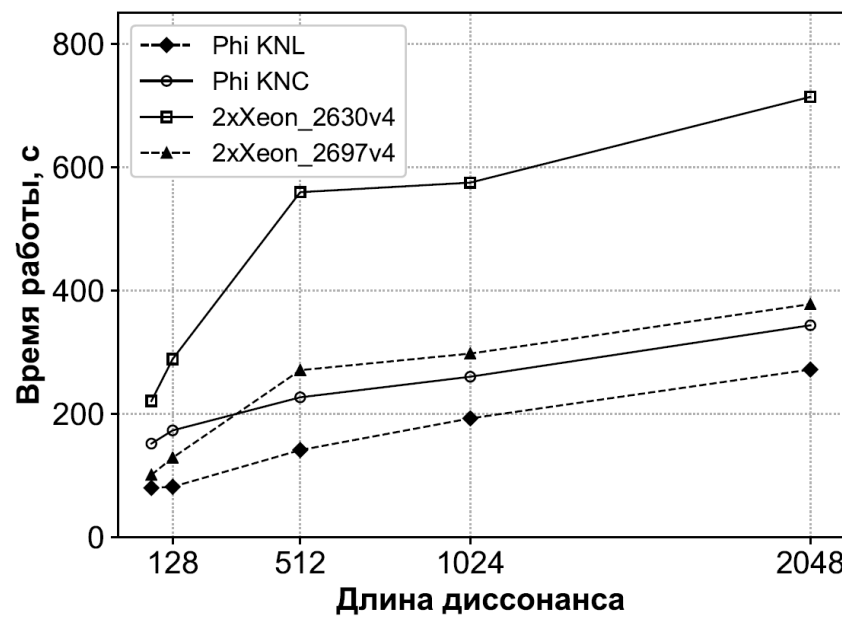
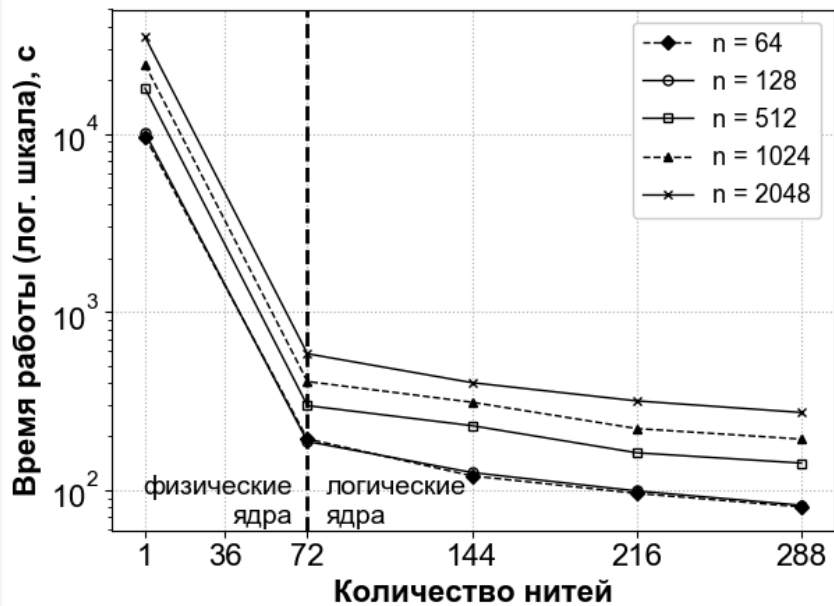
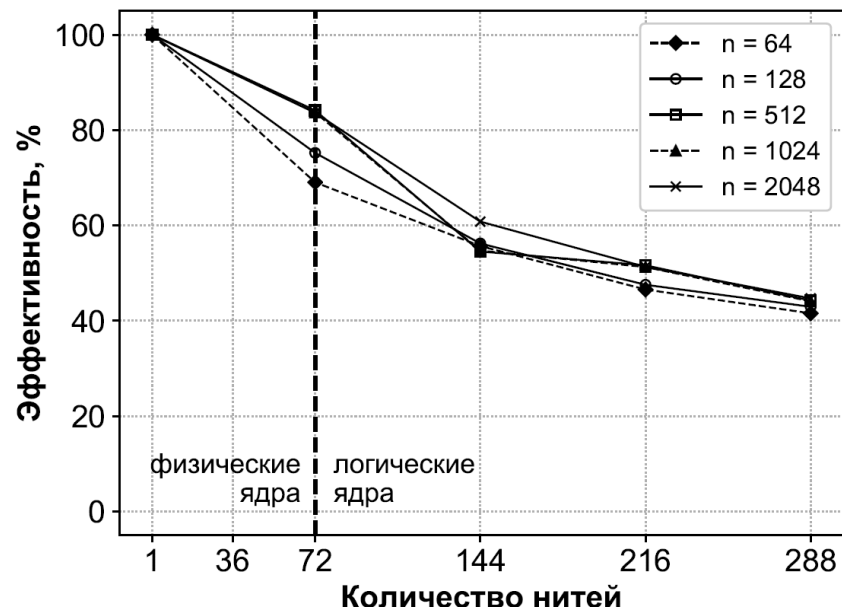
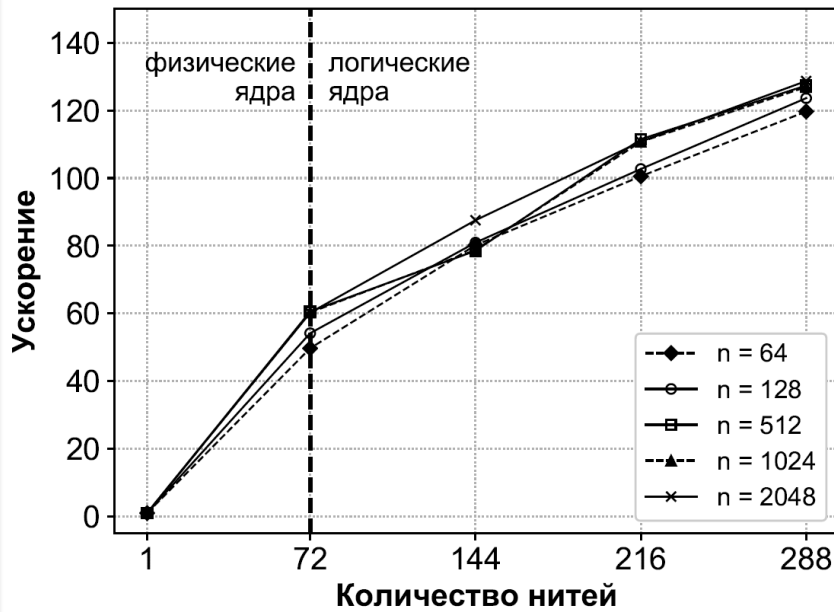
Эксперименты

- Аппаратная платформа

Характеристика \ Устройство	Торнадо ЮУрГУ	ЦКП СО РАН		
	Intel Xeon Phi SE10X (KNC)	Intel Xeon Phi 7290 (KNL)	2× Intel Xeon E5-2697v4	2× Intel Xeon E5-2630v4
К-во физ. ядер	61	72	2×16	2×10
Гиперпоточность	4×	4×	2×	2×
К-во лог. ядер	244	288	64	40
Частота, GHz	1.1	1.5	2.6	2.2
Память, Gb	8	16	128	128
Пик. произв-ть, TFLOPS	1.076	3.456	0.600	0,390

- Ряд: $|T| = 10^7$

Результаты



Сравнение с аналогами

Параметры	Время работы, с			Ускорение относительно <i>HOTSAX</i>			
	<i>PDD</i> ¹	<i>PhiDD</i>		<i>HOTSAX</i> ²	относительно <i>HOTSAX</i>		
<i>Ряд</i> $ T = 10^7$ <i>Диссонанс</i> $n=128$	10 CPU	Intel Xeon Phi KNL		Intel Core i5	<i>PDD</i>	<i>PhiDD</i>	
	1.2 ГГц	10 нитей	288 нитей	2.6 ГГц		10 CPU	10 нитей
	399 600	833.3	81.9	604 800	1.5×	7.2*10 ² ×	7.4*10 ³ ×

¹ Huang T., *et al.* **Parallel Discord Discovery**. Proc. of the Int. Conf. PAKDD 2016. pp. 233–244.

² Keogh E.J., *et al.* **HOT SAX**: efficiently finding the most unusual time series subsequence. Proc. of the 5th IEEE Int. Conf. on Data Mining. P. 8.

Заключение

- Предложен новый параллельный алгоритм *PhiDD* поиска диссонансов временного ряда для многоядерных ускорителей Intel MIC
- Проведены эксперименты, показавшие хорошую масштабируемость алгоритма
- Будущие исследования
 - Применение алгоритма *PhiDD* для поиска аномалий во временных рядах приложений Industry 4.0
 - Разработка распределенной версии алгоритма

Спасибо за внимание! Вопросы?

Михаил Леонидович Цымблер

mzym@susu.ru