

Интернет Университет Суперкомпьютерных технологий

Лекция 5

Сортировка данных с точки зрения МВС (часть 1)

Учебный курс

Введение в параллельные алгоритмы

Якобовский Михаил Владимирович
чл.-кор. РАН, проф., д.ф.-м.н.
Институт прикладной математики
им. М.В.Келдыша РАН, Москва

**Расположить в порядке
неубывания
 N элементов массива
чисел,
используя p процессоров**

К вопросу о

- ❑ Наилучшем последовательном алгоритме
- ❑ Медленном последовательном алгоритме
- ❑ Высокой степени внутреннего параллелизма

Две задачи сортировки массива чисел

- A. Объём оперативной памяти одного процессорного узла **достаточен** для одновременного размещения в ней всех элементов массива

- B. Объём оперативной памяти одного процессорного узла **мал** для одновременного размещения в ней всех элементов массива

Задача А

- Расположить N элементов массива a таким образом, чтобы для любого

$$i = 0, \dots, N - 2$$

выполнялось неравенство

$$a_i \leq a_{i+1}$$

Задача В

- Части массива хранятся на нескольких процессорах
 - Каждая часть массива должна быть упорядочена
 - На процессорах с большими номерами должны быть размещены элементы массива с большими значениями

• Правильно

$\langle 1,2,3,5 \rangle$ $\langle 5,6,7,7 \rangle$ $\langle 8,8,9 \rangle$

• Ошибка

$\langle 1,2,3,5 \rangle$ $\langle 5,7,6,7 \rangle$ $\langle 8,8,9 \rangle$

• Ошибка

$\langle 1,2,3,5 \rangle$ $\langle 5,6,7,8 \rangle$ $\langle 7,8,9 \rangle$

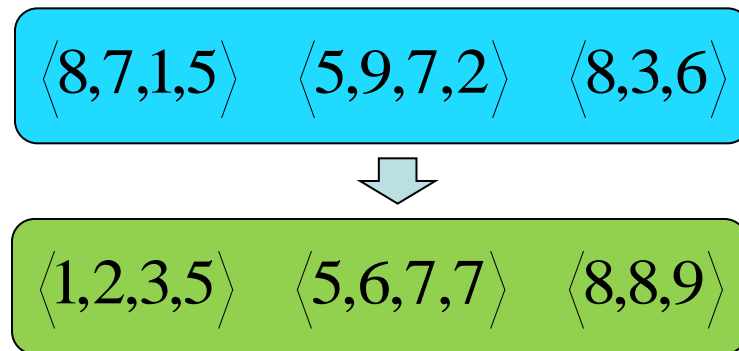
$N = 11$

$p = 3$

Задача В

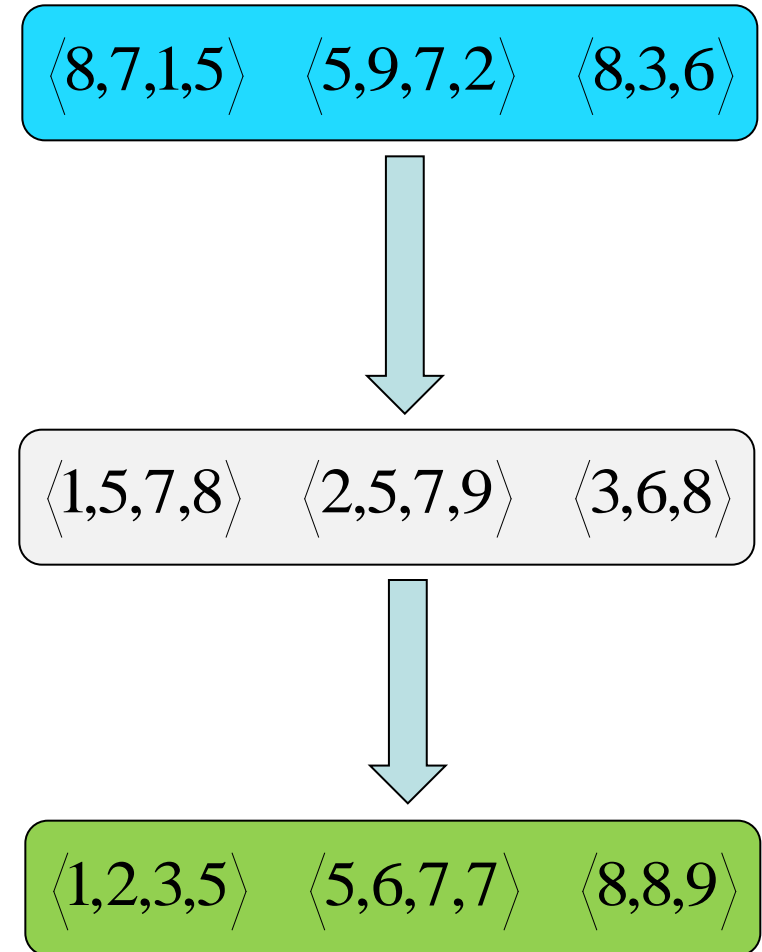
□ Будем рассматривать только процесс упорядочивания элементов:

- Перед началом сортировки на каждом из процессоров уже есть часть элементов массива
- После окончания сортировки на каждом из процессоров должно остаться столько элементов, сколько их было в начале (но, это уже другие элементы, ранее расположенные на других процессорах)



Предлагаемая стратегия: Этапы сортировки

- Упорядочивание фрагментов массива на каждом из процессоров
- Перераспределение элементов массива между процессорами



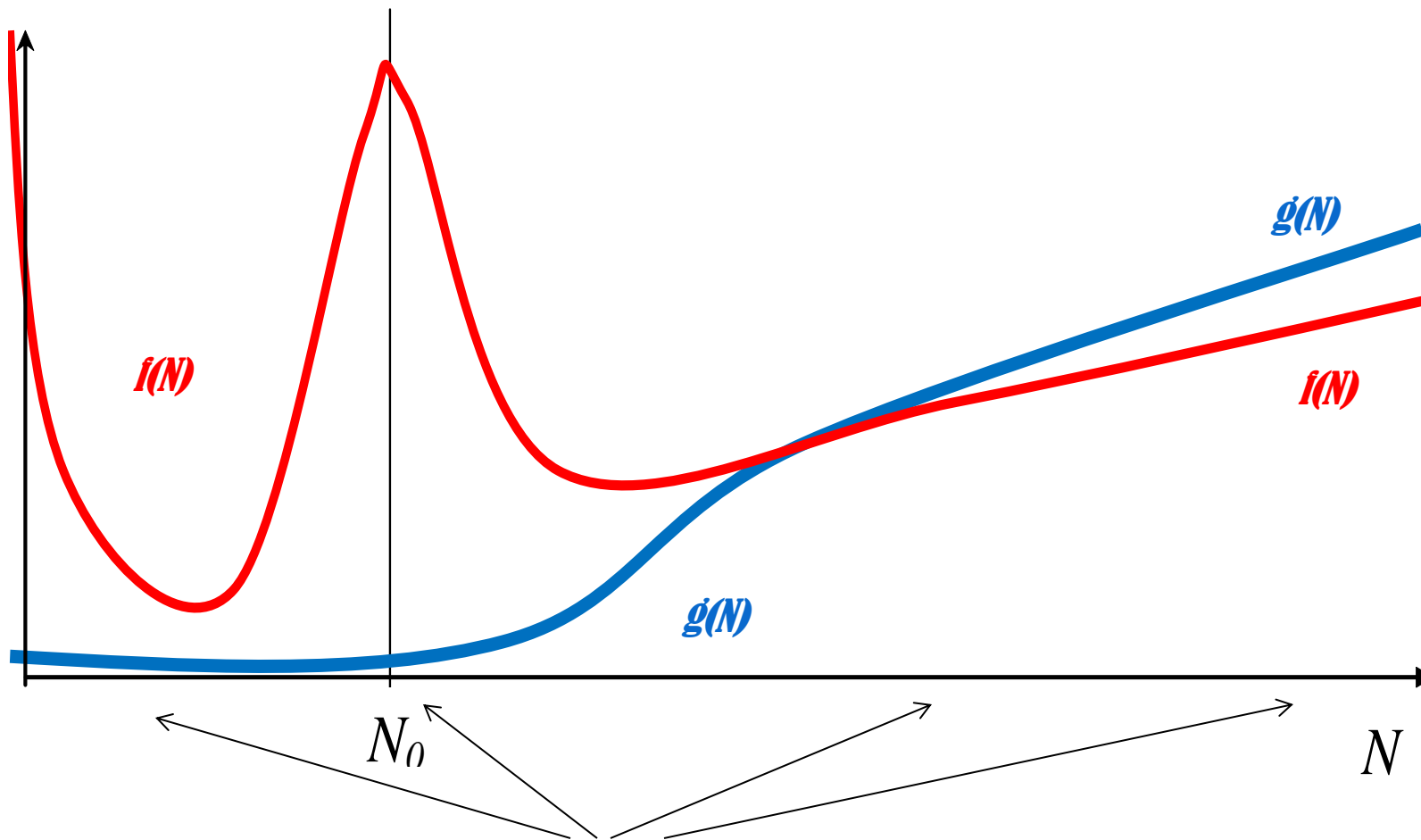
Конструирование наилучшего последовательного алгоритма

Сравнение последовательных алгоритмов сортировки

$$M(n) < Cn^2 \leftarrow$$

Алгоритм сортировки	Среднее число операций	Максимальное число операций
Быстрая (<i>qsort</i>)	$11.7 n \log_2 n$	$O(n^2)$
Пирамидальная (<i>hsort</i>)	$16 n \log_2 n$	$18 n \log_2 n + 38n$
Слияние списков (<i>lsort</i>)	$10 n \log_2 n$	$O(n \log_2 n)$

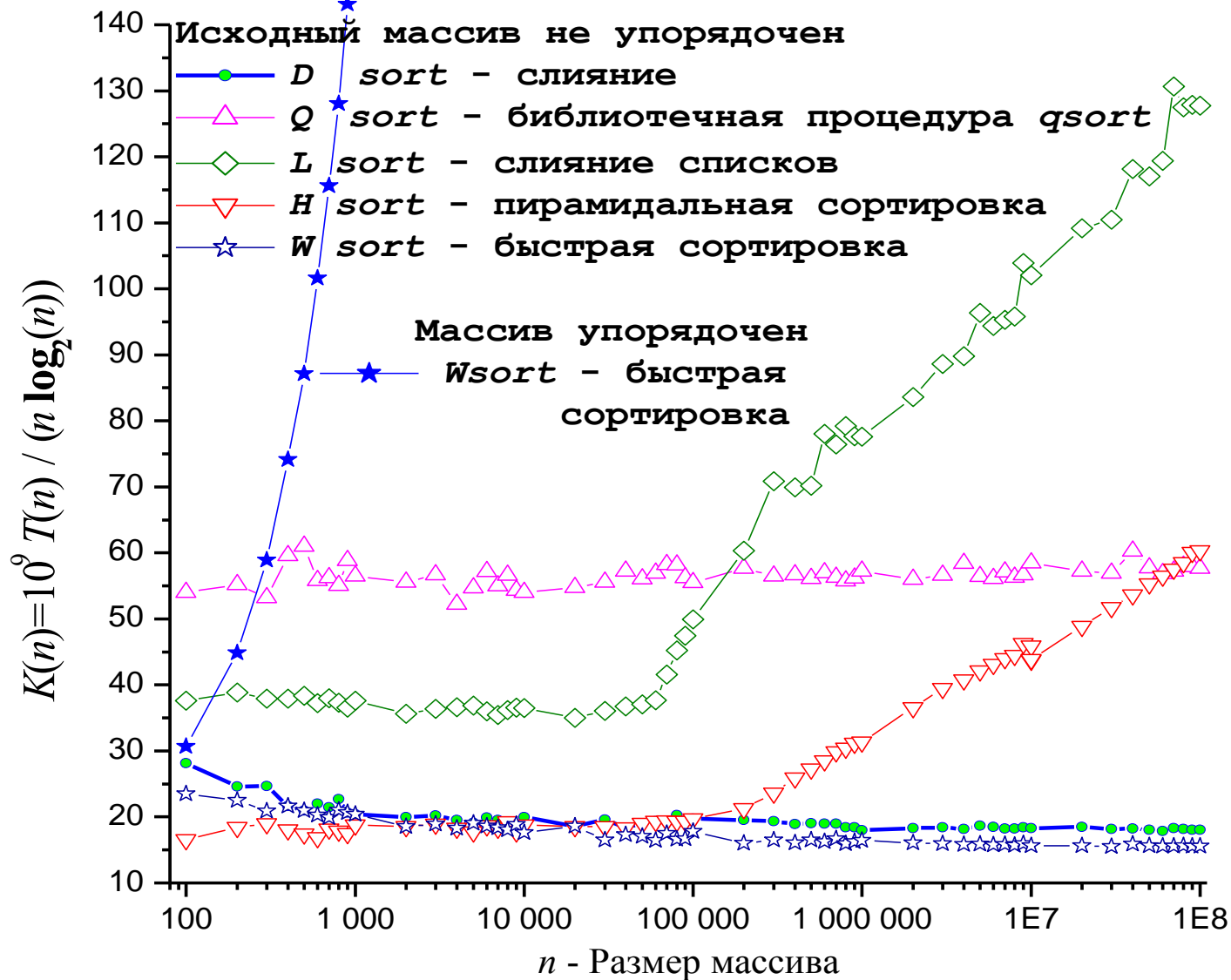
Пусть $f(N) < C \cdot g(N)$, ну и что?



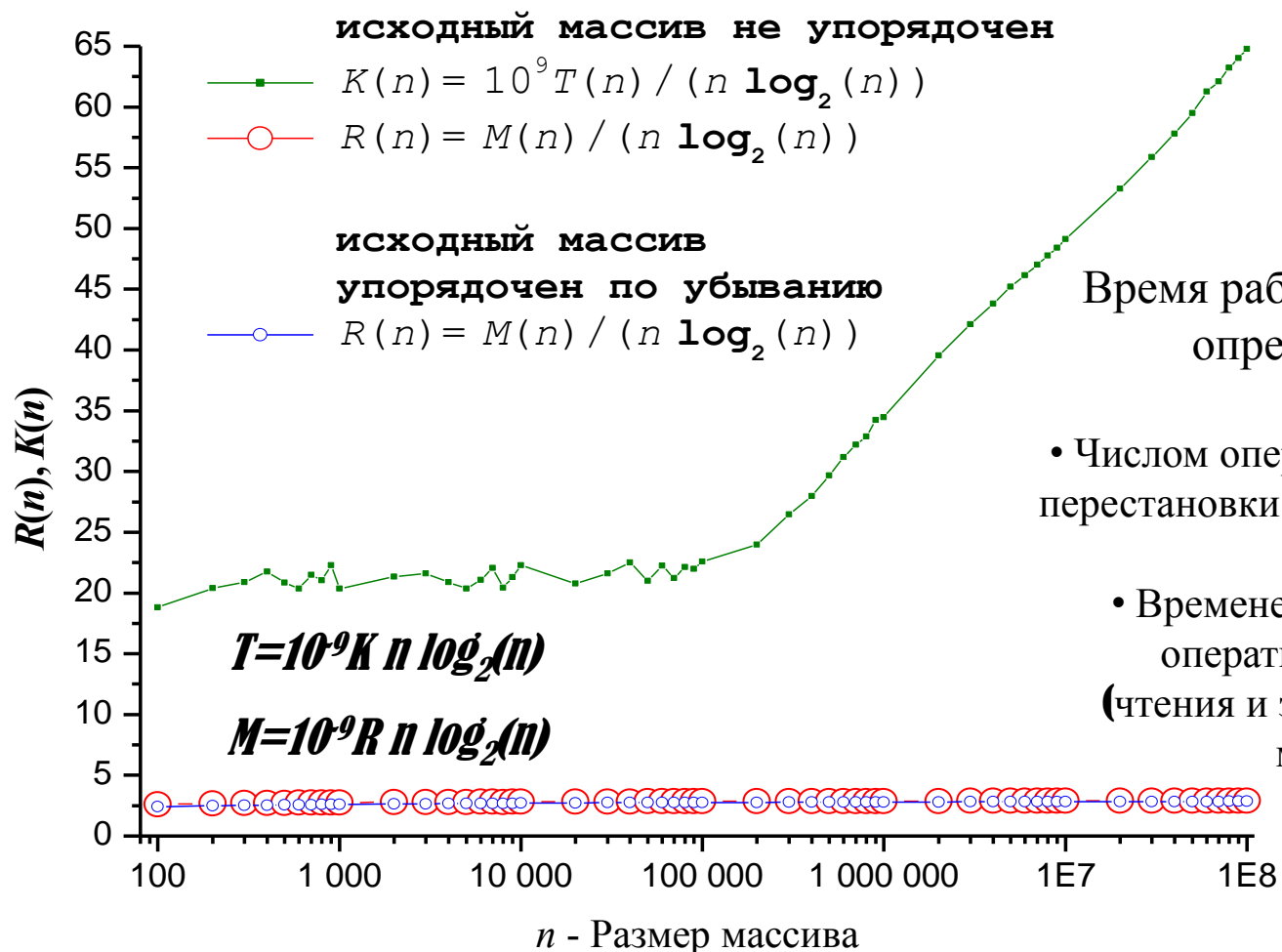
□ Где тут наши 2 Гигбайта оперативной памяти???

Константа времени сортировки

$$T=10^9 K N \log_2(N)$$

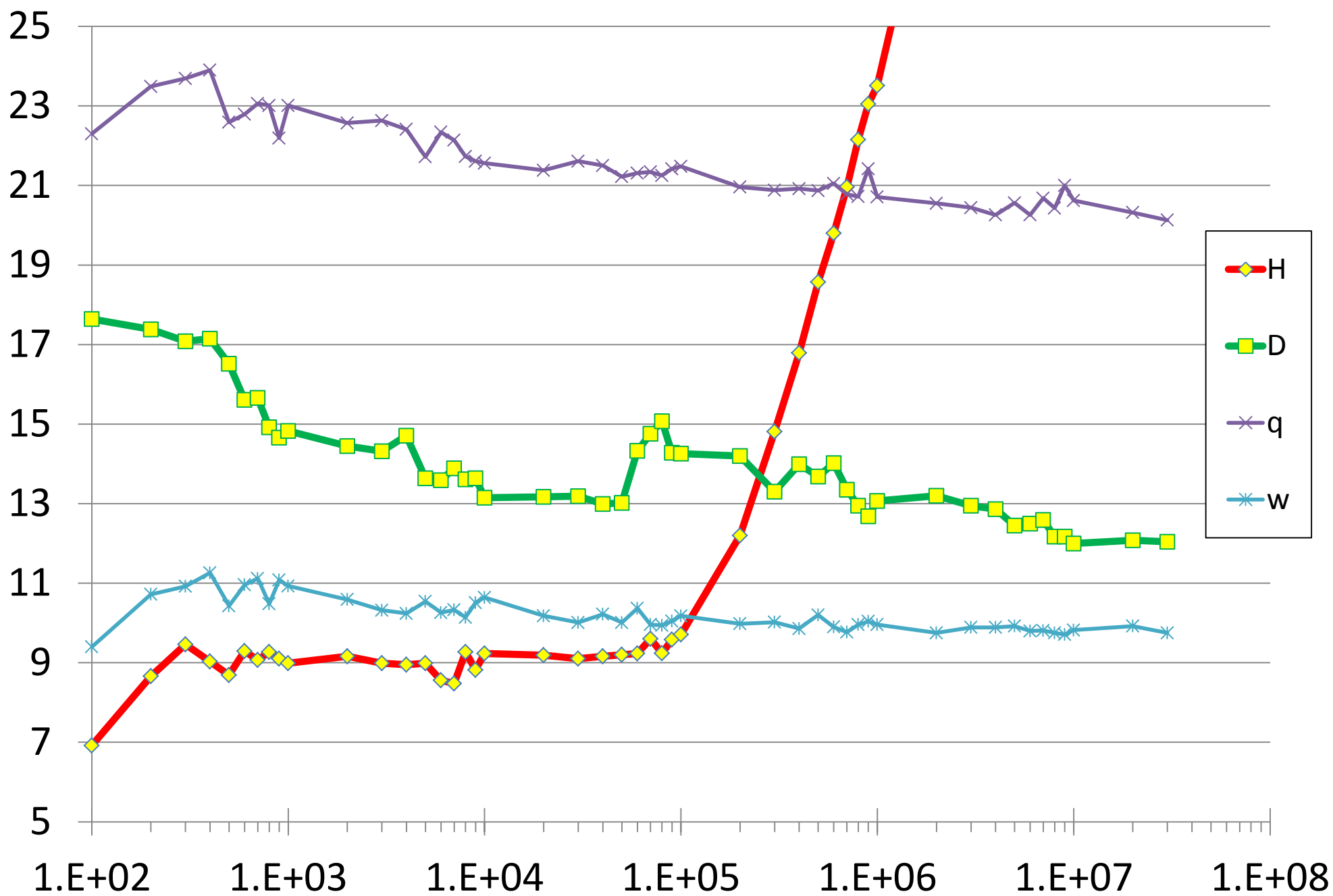


Пирамидальная сортировка: константы времени и числа операций

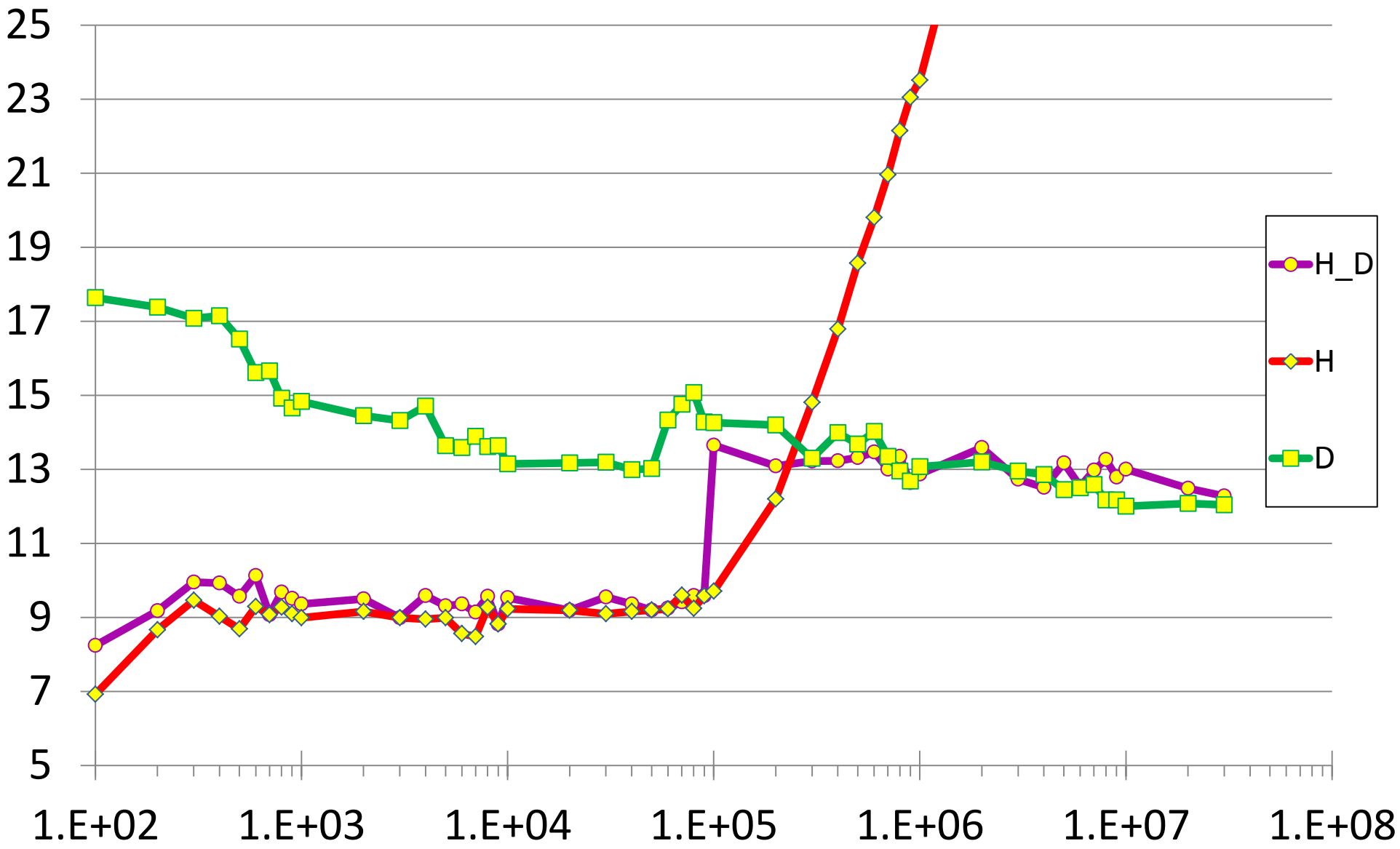


Время работы алгоритма определяется:

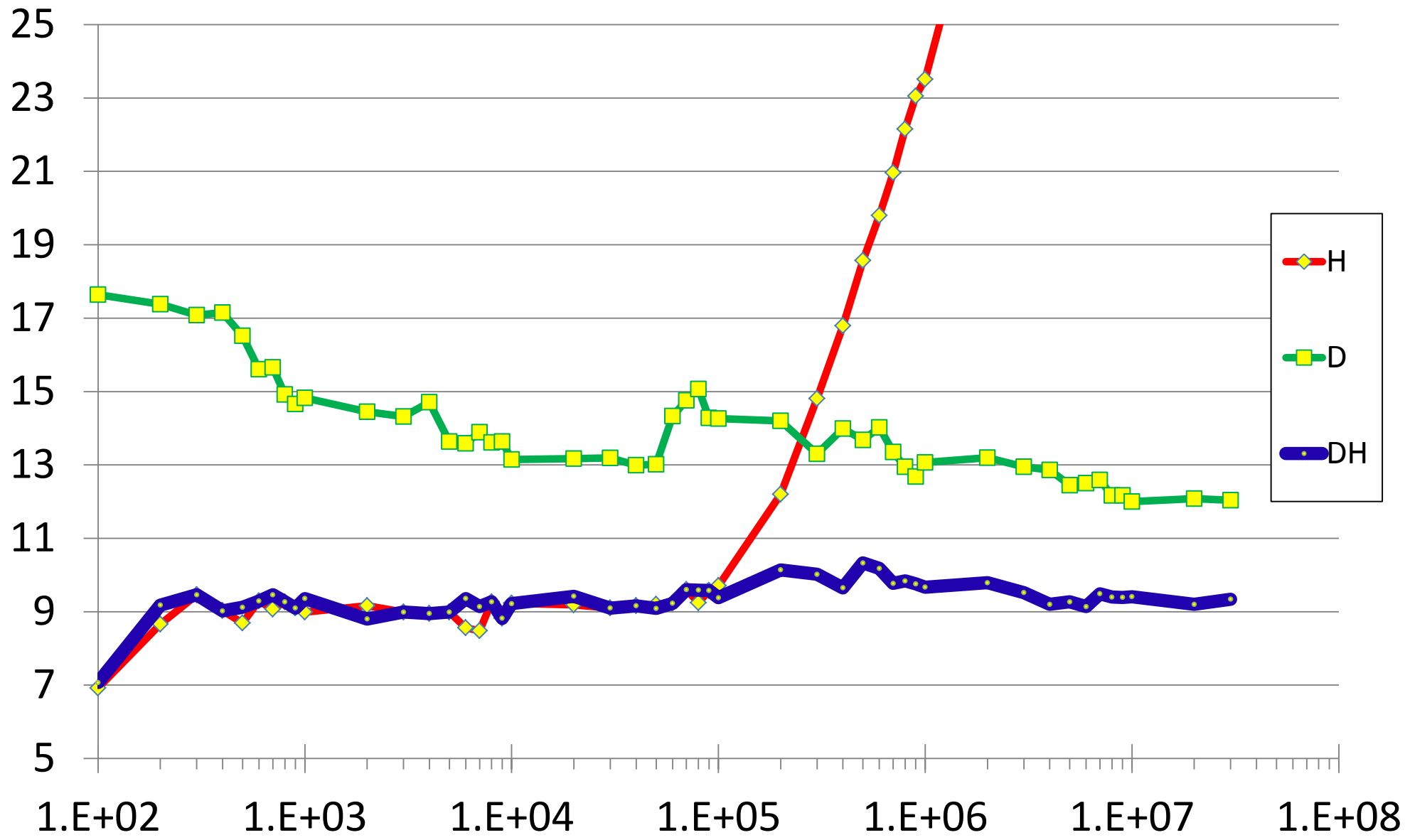
- Числом операций сравнения и перестановки элементов массива
- Временем обращения к оперативной памяти (чтения и записи элементов массива)

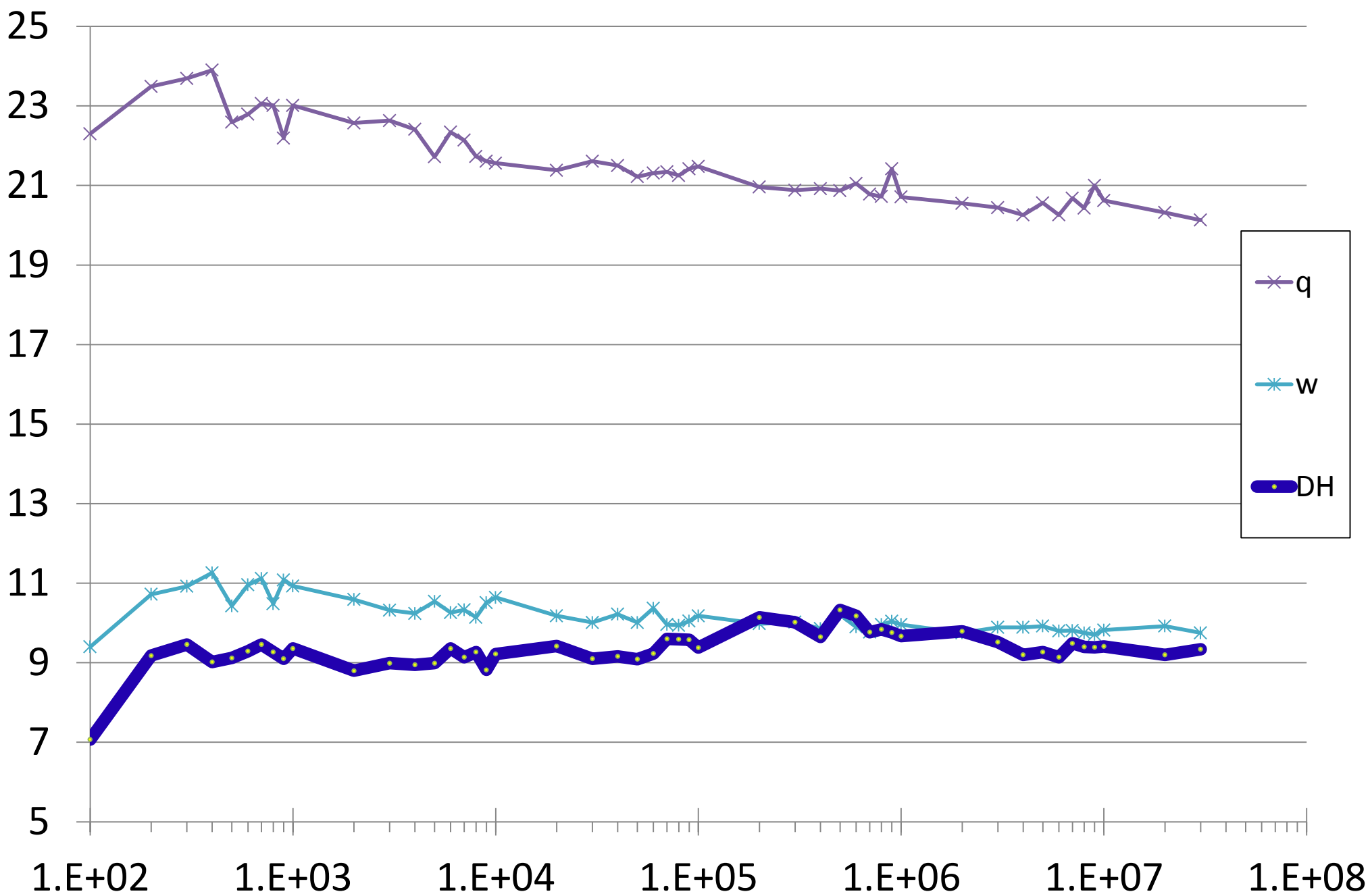


Меньше 10^5 - пирамидальная, больше - слияние



Меньше 100 000 элементов – пирамидальная,
иначе – (пирамидальная над фрагментами + слияние упорядоченных фрагментов)



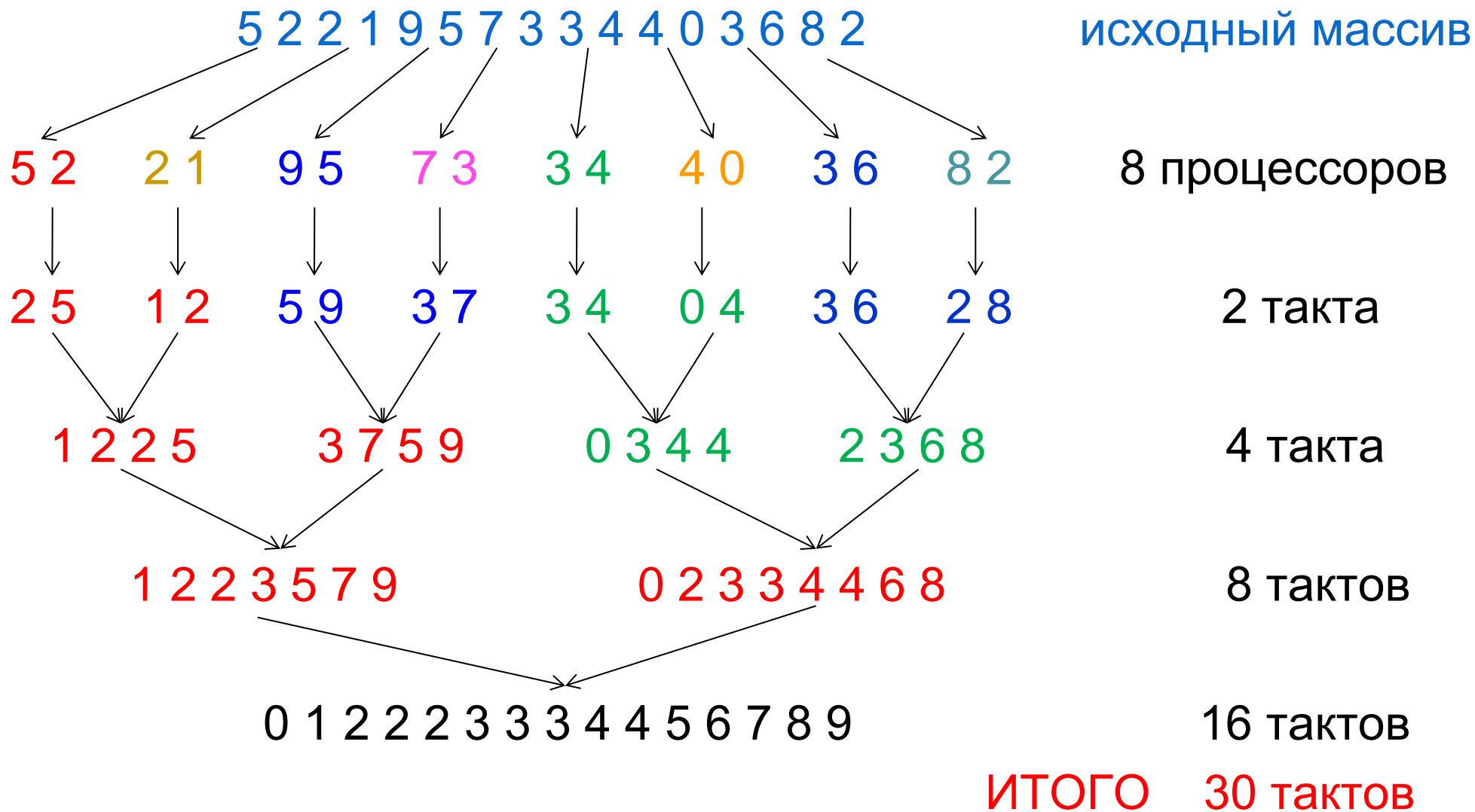


Какую сортировку следует использовать?

Алгоритм сортировки	Среднее число операций	Максимальное число операций
Быстрая (<i>qsort</i>)	$11.7 n \log_2 n$	$O(n^2)$
Пирамидальная (<i>hsort</i>)	$16 n \log_2 n$	$18 n \log_2 n + 38n$
Слияние списков (<i>lsort</i>)	$10 n \log_2 n$	$O(n \log_2 n)$
Простые вставки	$O(n^2)$	$O(n^2)$

Сортировка слиянием методом сдваивания

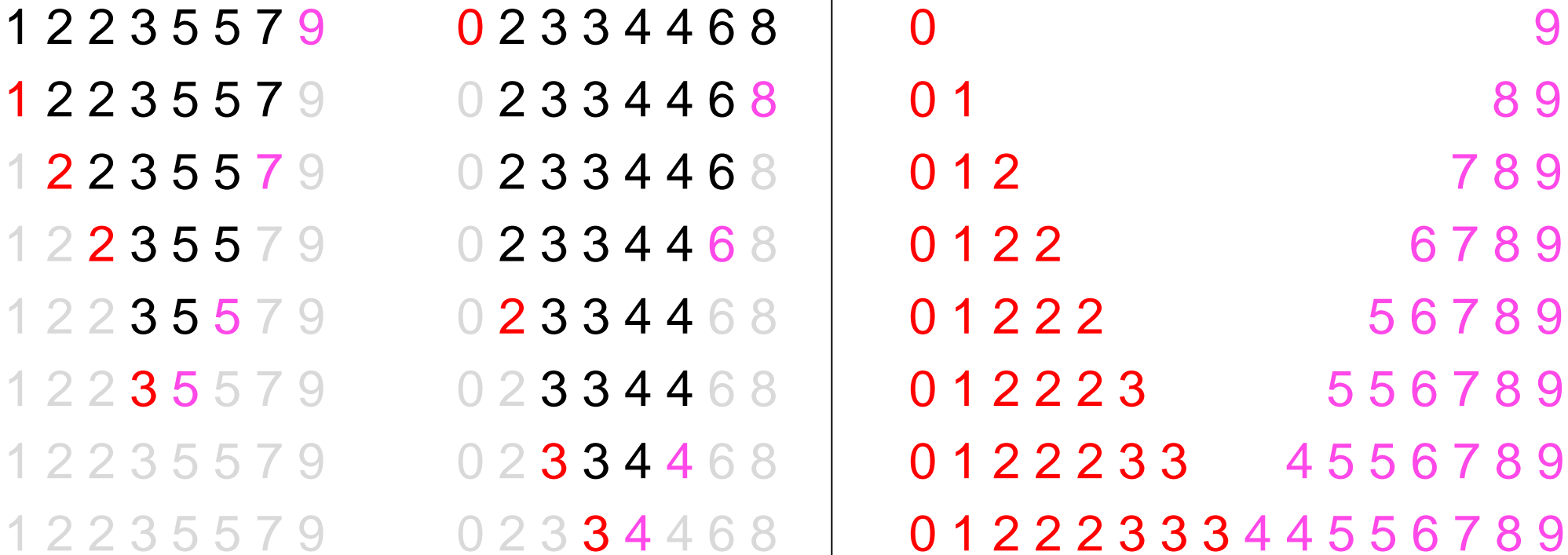
Требуется $2 + 4 + 8 + 16 = 30$ тактов (8 процессоров)



Слияние одним процессором. Требуется 16 тактов

(1 2 2 3 5 5 7 9)	(0 2 3 3 4 4 6 8)	
1 2 2 3 5 5 7 9	0 2 3 3 4 4 6 8	0
1 2 2 3 5 5 7 9	0 2 3 3 4 4 6 8	0 1
1 2 2 3 5 5 7 9	0 2 3 3 4 4 6 8	0 1 2
1 2 2 3 5 5 7 9	0 2 3 3 4 4 6 8	0 1 2 2
1 2 2 3 5 5 7 9	0 2 3 3 4 4 6 8	0 1 2 2 3
1 2 2 3 5 5 7 9	0 2 3 3 4 4 6 8	0 1 2 2 3 3
1 2 2 3 5 5 7 9	0 2 3 3 4 4 6 8	0 1 2 2 3 3 3
...		

Слияние двумя процессорами. Требуется 8 тактов



Ускорение при методе сдваивания

k_1 – сортировка, k_2 – передача данных

$$S(n, p) = \frac{T(n, 1)}{T(n, p)} = \frac{k_1 n \log_2 n}{\frac{n}{p} \left[k_1 \left(\log_2 \frac{n}{p} + 2p - 1 \right) + k_2 (p - 1) \right]}$$

$$S(10^9, 4) \approx \frac{4}{1.13 + \frac{1}{30} \frac{k_2}{k_1}} < 3.5$$

$$S(10^9, 32) = \frac{32}{1 + \frac{1}{30} \left(56 + 31 \frac{k_2}{k_1} \right)} \approx \frac{32}{3 + \frac{k_2}{k_1}} < 11$$

Контакты

Якобовский М.В.

чл.-кор. РАН, проф., д.ф.-м.н.,

зав. сектором

«Программного обеспечения многопроцессорных систем и вычислительных сетей»

Института прикладной математики им.
М.В.Келдыша Российской академии наук

[mail: lira@imamod.ru](mailto:lira@imamod.ru)

[web: http://lira.imamod.ru](http://lira.imamod.ru)