

КДЗ. Экспериментальное
сравнение сортировок

Цель работы

Цель работы – экспериментальное определение временной сложности сортировок различных массивов разными методами и сравнительный анализ полученных результатов

<https://www.youtube.com/watch?v=kPRA0W1kECg> MergeSort vs QuickSort

<https://www.youtube.com/watch?v=kPRA0W1kECg> 15 сортировок

<https://www.youtube.com/watch?v=DSMCZZGbZo4> 50 сортировок

Задание

Провести эксперименты – измерить **время выполнения** алгоритмов сортировки

- 1) выбором
- 2) пузырьком
- 3) пузырьком с условием Айверсона 1
- 4) пузырьком с условием Айверсона 1+2
- 5) простыми вставками
- 6) бинарными вставками
- 7) подсчетом (устойчивая)
- 8) цифровой
- 9) слиянием
- 10) быстрой (разбиение Хоара и разбиение Ломута)
- 11) пирамидальной

Задание (продолжение)

1. Измерения провести для размеров массива (включительно)
 - от 50 до 300, шаг 10
 - от 100 до 4100 , шаг 100
2. Измерения выполнить для массивов, заполненных целыми неотрицательными числами:
 - случайными значениями от 0 до 5
 - случайными значениями от 0 до 4000
 - «Почти» отсортированными в требуемом порядке числами (например, в каждой тысяче элементов поменять местами N пар элементов отсортированного массива)
 - Отсортированными в обратном порядке (по убыванию) числами от 4100 до 1
3. Результаты измерений времени сортировки разных массивов разных размеров разными методами вывести в файл / файлы *.csv (разделители – точки с запятой «;»)

Задание (продолжение)

Сортировать разными методами надо один и тот же массив.

Не надо сортировать сортированный на предыдущем шаге!!!

В цикле для подсчета количества операций сгенерировать эталонный массив максимальной длины 4100 (4 раза – для 4-х видов массивов), затем перед каждой сортировкой копировать элементы эталонного массива в рабочий массив, который и сортировать.

Сначала копируете 50 элементов, потом 60 и т.д. для первого диапазона размерностей массивов

Сначала копируете 100 элементов, потом 100 и т.д. для второго диапазона размерностей массивов

Всего сортировок (4 вида массива * (26 + 41) длин массива * 12 методов)

Задание. Результаты

Представить результаты экспериментов в виде таблиц на листах в Excel, в которых указать: Метод сортировки, вид массива (например, «пузырек, случайные [0;5]» или «Простые вставки, обратно сортированный»)

Примерный вид таблиц (должно быть 2 таблицы для двух диапазонов размерностей):

Размер массива	Пузырек Случайные числа 0-5	Пузырек случ мак	Цифровая обратно сортир	Цифровая почти сортир
50						
...						
300						

Для этого выводите результаты измерений времени в файлы *.csv с разделителем «;» или другим.

Потом загрузите эти файлы в ЭТ (данные – импорт)

Задание. Результаты

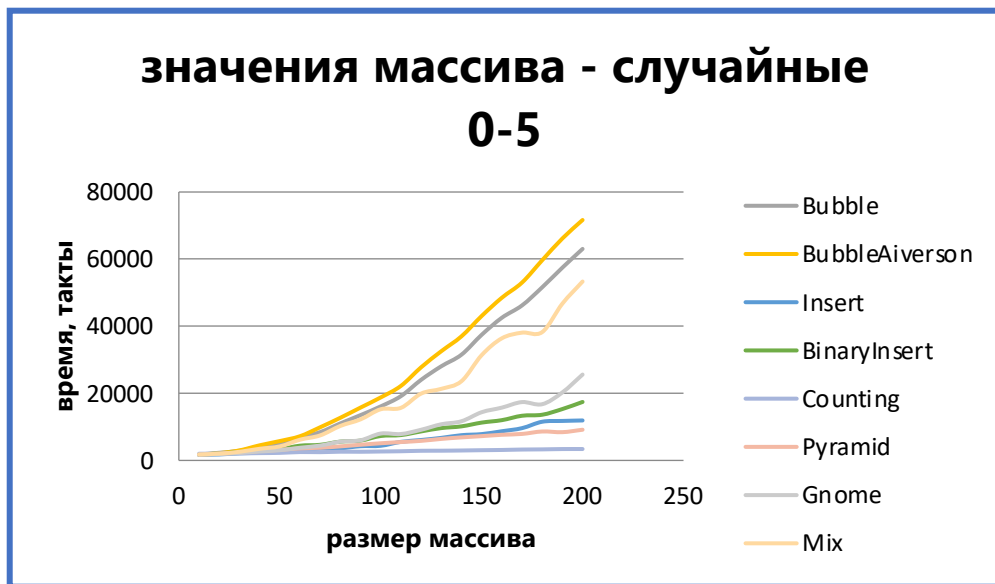
Представить результаты экспериментов в виде таблиц и графиков (в ЭТ)

«размер массива (ось x) – время сортировки (ось y), в нс или в мкс».

1. 4 + 4 графика, для каждого вида массива. Заголовок – вид массива (случайный 0–5, случайный 0–4000 и т.п.). Легенды – метод сортировки (4 графика, т.к 4 вида заполнения массива, на каждом по 12 кривых, соответствующих 12 методам сортировки)
2. 12 + 12 графиков, для каждого метода сортировки. Заголовок графика – метод сортировки. Легенды – вид массива (на каждом графике по 4 кривых, по количеству разных массивов)

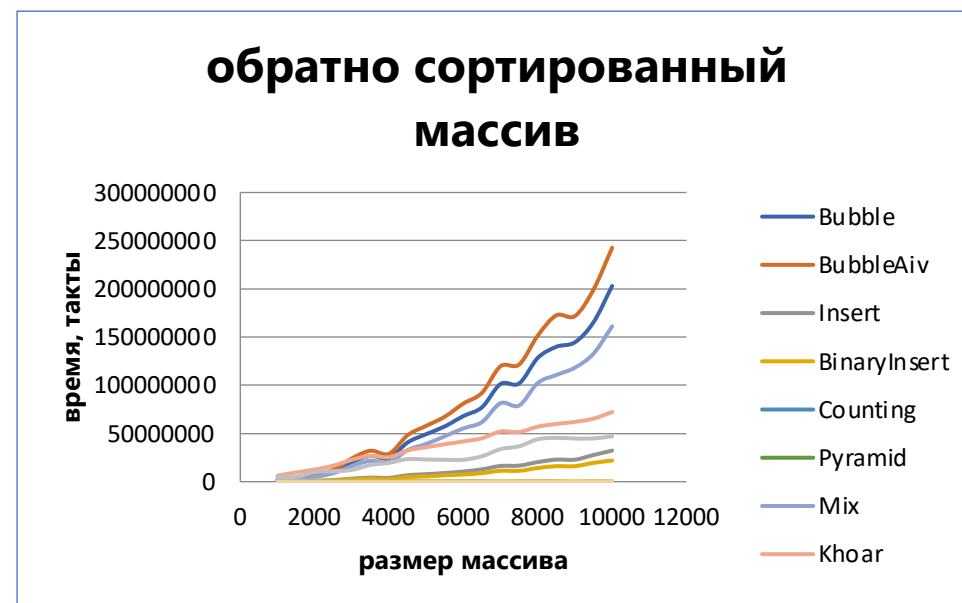
Обязательно проанализировать полученные результаты и сделать выводы – в отдельном текстовом файле

Примеры графиков 1

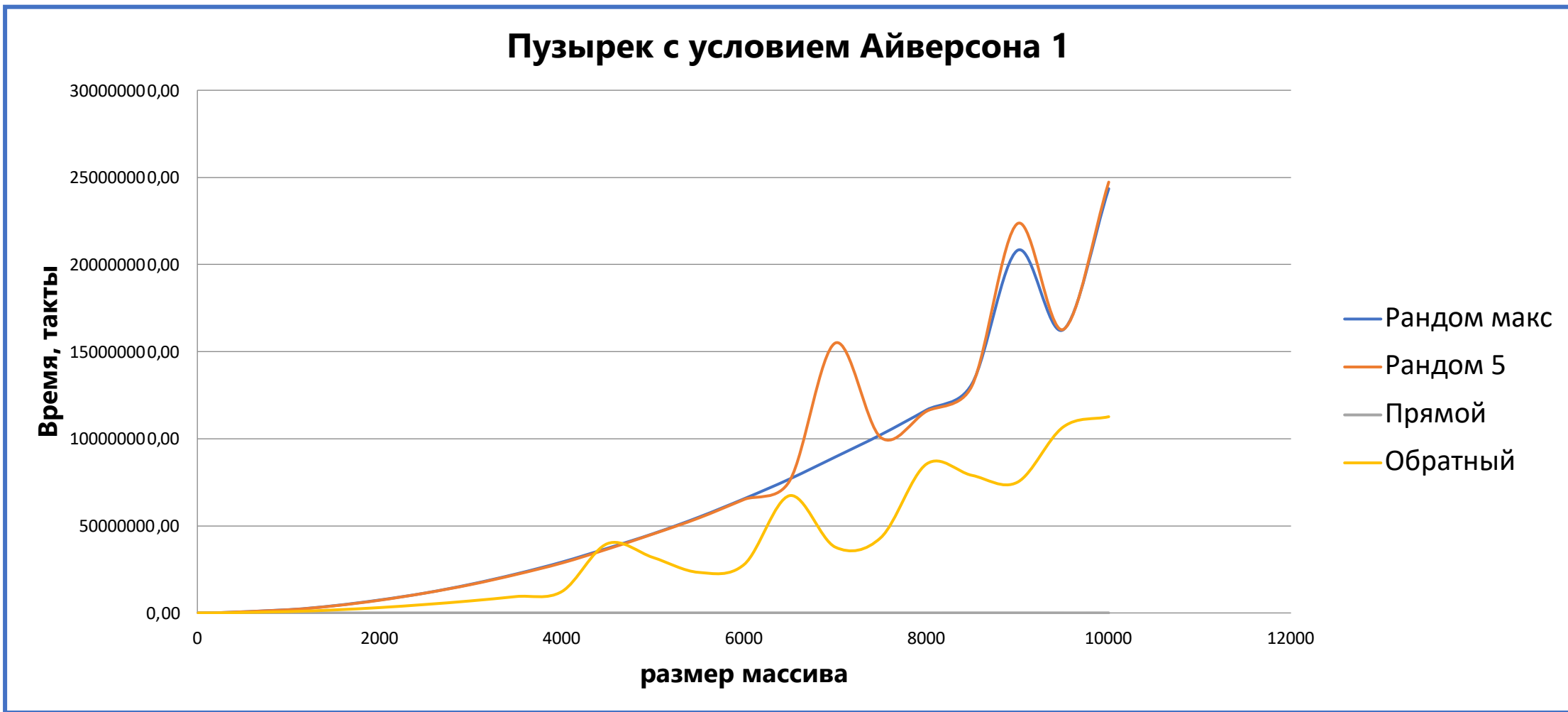


Оси на графиках должны быть
оцифрованы и обозначены

Если при размещении всех методов на
одном графике для одного вида исходного
массива невозможно их сравнить
(некоторые графики «лежат» на оси X),
можно разделить методы на две группы и
построить два графика)



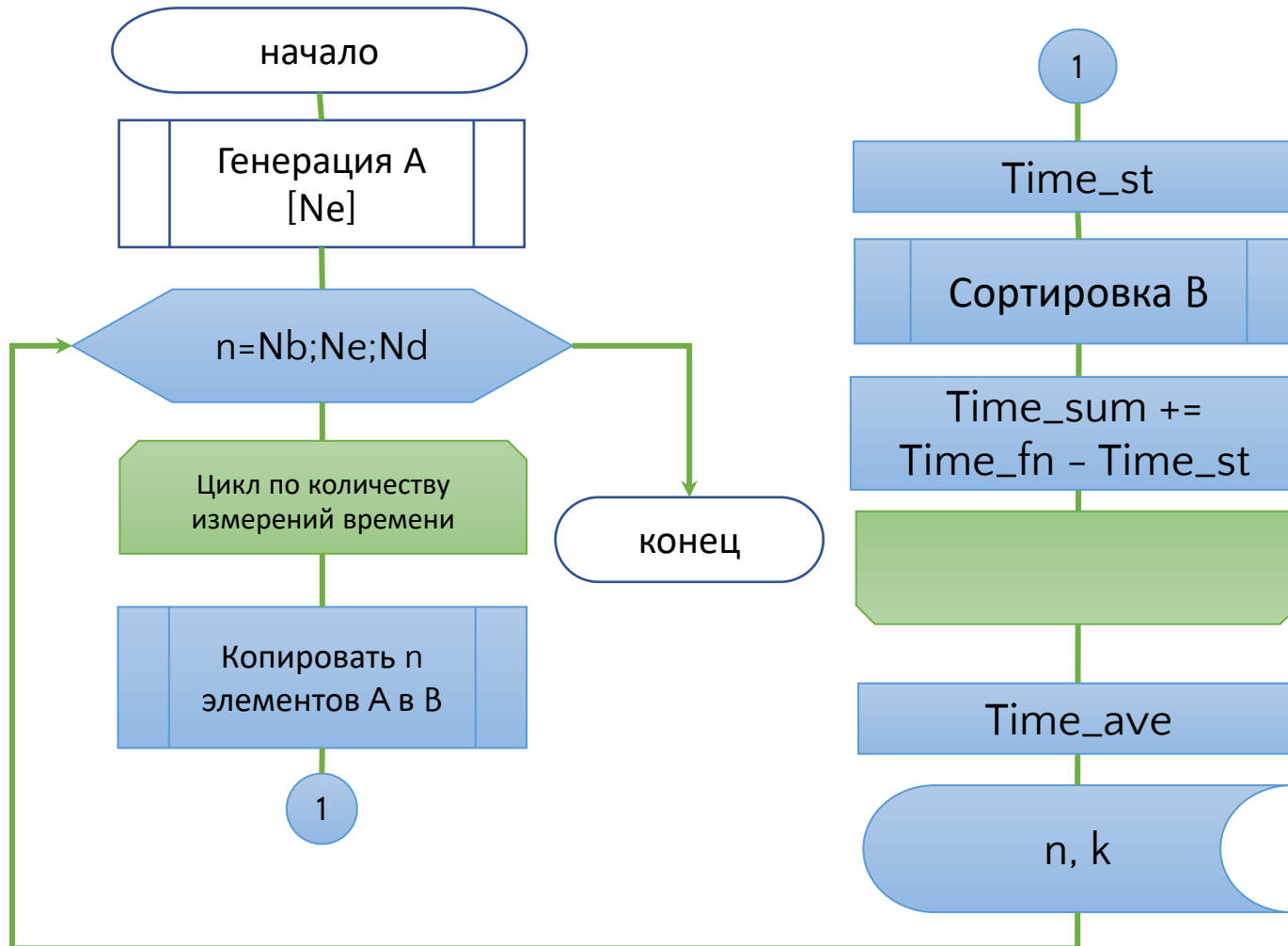
Пример графиков 2



Эксперимент попроще (один вид массива, один метод сортировки)

Nb = 50;
Ne = 300;
Nd = 50;

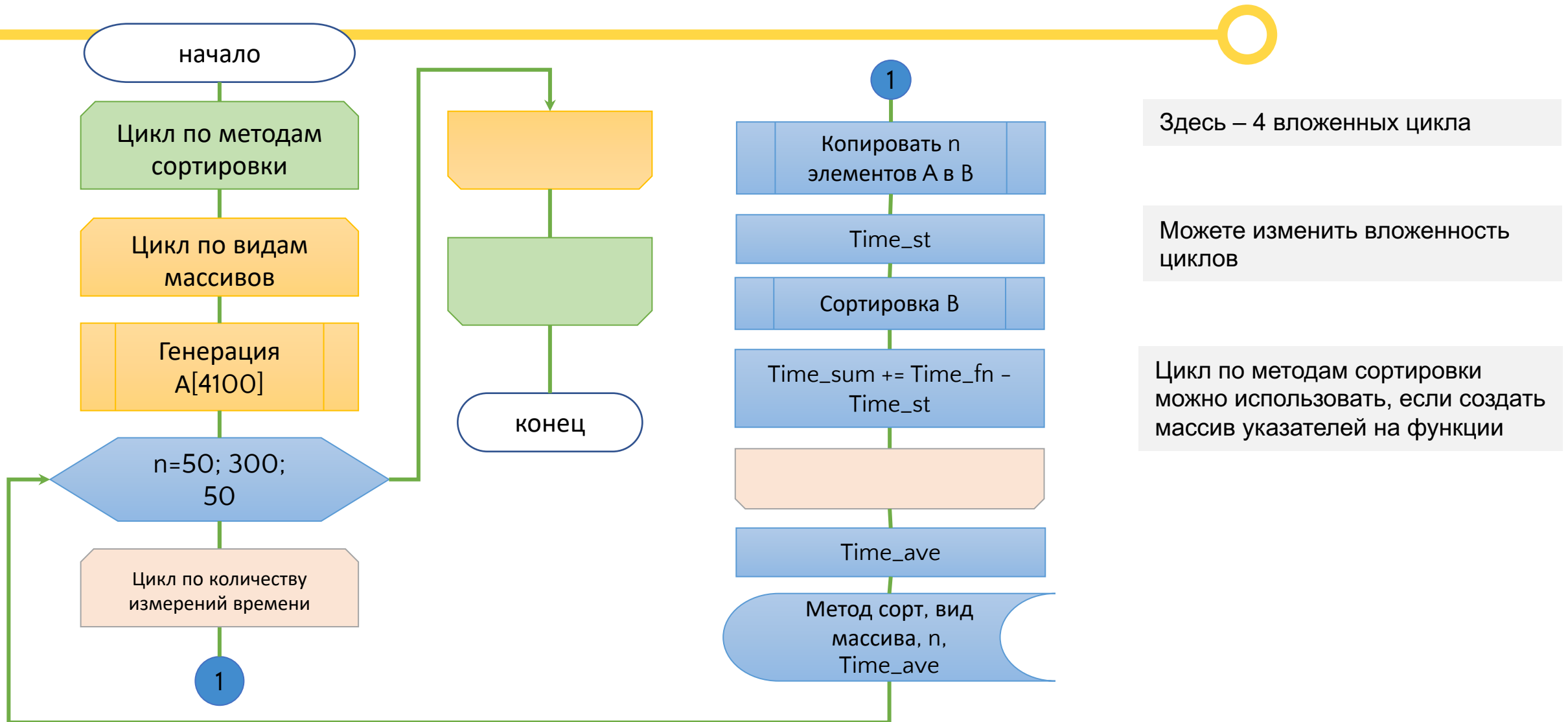
Nb = 100;
Ne = 4100;
Nd = 1000;



Здесь два цикла

- по количеству элементов массива n и
- по количеству измерений времени

Эксперимент сложнее (разные виды массива и методы сортировки)



Обозначения в блок-схемах



A – эталонный массив

B – рабочий массив

Time_* – время в нс или мкс

n – текущий размер массива для сортировки

Nb – начальный размер массива

Ne – конечный размер массива

Nd – шаг изменения размера массива

Измерение времени



Время надо измерять в наносекундах или в микросекундах.

Как это сделать – см. материал, подготовленный студентом гр. БПИ173 Кириллом Быковым

<https://darkydash.github.io/cpp-chrono/>

ВАЖНО!!!

- Первые прогоны программы выполняются медленнее, т.к. оптимизирующий компилятор еще не провел оптимизацию.
- Лучше исключить 2-3 первых замера времени сортировок из расчета среднего времени
- ДО измерений времени прогоните все сортировки «вхолостую»

Проверки



Напишите функцию проверки упорядоченности элементов массива. Она должна возвращать

Истину или 1, если массив сортирован,
Ложь или 0 – если массив не сортирован.

Предусмотрите вывод исходного и сортированного массивов в файлы `input.txt` и `output.txt` (для защиты проекта)

Комментарии в коде



В начале кода обязательно укажите

- 1) ПиАА 2022, ФИО, группа
- 2) ФИО и группа
- 3) Среда разработки
- 4) Перечислите, что сделано
- 5) Перечислите, что не сделано

Далее в коде пишите комментарии, не слишком много, но чтобы было понятно, что в этом фрагменте (цикле, вводе и проч.) выполняется

Загрузка результатов в SmartLMS



В Смарт ЛМС загружаем один или несколько архивов *.zip, содержащих

- Комментированный код (только файлы *.h и *.cpp)
- Результаты эксперимента в таблице - таблицы и графики.
- Отчет с анализом полученных результатов и выводами по каждому графику.
- Имена файлов таблиц и текста - <группа>_<FIO>.*

Возможные ошибки и недостатки



- нет таблиц и графиков;
- нет результатов и выводов или они неполные ;
- неправильно работают или не реализованы алгоритмы сортировки
- Изменяется эталонный массив или повторно подаётся на вход алгоритму отсортированный на предыдущем шаге массив;
- измерения для алгоритма отсутствуют или выполнены неправильно (например, время сортировки = 0),
- не создаётся CSV-файл с результатами измерений для дальнейшего создания и обработки таблиц;
- Используется дополнительная память там, где это не требуется;
- Наличие в функциях сортировок посторонних операций (копирование массивов, вывод и т.п.);
- Нет рандомной генерации или не меняется seed (не используется srand);
- Неправильно генерируется массив какого-то типа;
- Нет очистки памяти;
- Графики без названия, без обозначений осей и/или без легенд;
- Нет комментариев в коде;

Возможные достоинства работы



- + использование указателей на функции в качестве параметров функций;
- + полный и качественный анализ результатов;
- + наличие всех графиков;
- + отличные ответы на защите;
- + все минусы предыдущего слайда можно обратить в плюсы, если сделать все верно.

Консультации и защиты задания

Нужно будет защитить работу. Это примерно 5–7 минут с ответами на вопросы. О защитах будет сообщено дополнительно и выслано их расписание.

На защите будут вопросы по

- коду,
- алгоритмам,
- организации экспериментов,
- результатам экспериментов (таблицы и графики)
- анализу и выводам по полученным результатам.

