Спецкурс кафедры «Вычислительной математки»

Параллельные алгоритмы вычислительной алгебры

Александр Калинкин

Сергей Гололобов

Часть 2: Современные компьютеры

История развития компьютеров.

Особенности современных ЦПУ и графических ускорителей.

История развития компьютеров

Это не есть исторический экскурс в прошлое в классическом понимании:

1. **Человек**
2. **Счёты**
3. **Механические счёты**
4. **Компьютер**
5. **Кластер**
6. **Многоядерный компьютер**
7. **Многоядерный кластер**
8. **Неоднородный многоядерный компьютер**
9. **Неоднородный многоядерный кластер**

История развития компьютеров

1. Человек

Особенности: требует еду, питьё, жильё, необходим отдых, неоднородно обучаем, изобретателен, ...

Пример: Напишите число «пи» до куда сможете...

3,14159265358979323846264338327795

Дополнительные данные (окружение):

100 000 лет назад?
5000 лет назад?
600 лет назад?
100 лет назад?
50 лет назад?
Сейчас? Дорогой арифмометр
Завтра? Очень дорогой арифмометр
В Китае?
В России?

Вопросы устойчивости возникли именно благодаря желанию человека посчитать!

Элементарная вычислительная математика родилась здесь.

История развития компьютеров

2. (Механические) Счёты

Особенности: конечная арфиметика, требуют не сильно квалифицированного оператора, ускоряют процесс вычислений, ...

В России исчезли лет 20 назад. Первая атомная бомба была рассчитана на счётах. Схема Годунова возникла на счётах.

Главное: позволяют делать вычисления параллельно!

Не имели широкого применения до появления сильной нужды в моделировании. Дополнительно: появилась нужда в оптимизации вычислений на основе аналитических рассуждений.

История развития компьютеров

4. **Компьютер**

Особенности: электрический привод, бинарное предсталение о мире, повторябельность, умение выполнять программы, умение хранить биты, надёжность?...

Острая нужда в оптимизации вычислений в связи с дороговизной машины поначалу.

Именно в этот момент и родилась классическая вычислительная математика, которую вы учите в нашем университете.

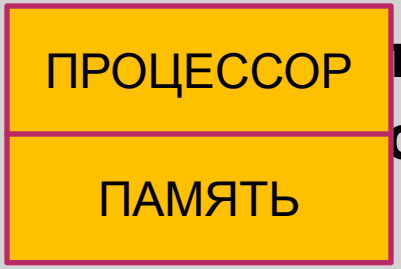
Основная проблема вычислительной математики: минимизировать вычисления (число операций в алгоритме) и минимизировать используемую память (сопряжённые градиенты, например)

Всего 25 лет назад компьютер с 40МБ памяти на диске можно было обменять на автомобиль.

История развития компьютеров

4. Компьютер

Итак, компьютер это



ть выполнения операций («флопсы»)
ство требуемой памяти (байты)

Дополнительно: какие именно операции выполняются, например, схема единственного деления в методе Гаусса

Почему? Деление – это операция, которая не может быть выполнена за ~1 такт в отличие от сложения, умножения, вычитания

История развития компьютеров

4. Компьютер

Тактовая частота – основная характеристика процессора. Именно она определяет сколько операций (тех операций, что процессор в состоянии исполнять) процессор в состоянии выполнить за секунду.

С точки зрения вычислительной линейной алгебры вторая основная характеристика – количество операций с плавающей точкой (с вещественными числами), которые может выполнить процессор за 1 такт

Итог:

Flops(Флопс) = floating point operations per second.

Объединение основных характеристик даёт нам главную характеристику с точки зрения вычислительной линейной алгебры – количество операций с плавающей точкой, которые может выполнить процессор.

История развития компьютеров

4. Компьютер

Дополнительно: относительно современный процессор может выполнять несколько операций с плавающей точкой за 1 такт параллельно (несколько вычислительных блоков могут работать одновременно и несколько чисел могут обрабатываться на 1 блоке [векторизация]) . Но...

с точки зрения нашей вычислительной математики это не столь важно, об этом заботиться либо компилятор, либо библиотека высокопроизводительных программ.

Тем не менее, вы должны знать о выравнивании данных:

__declspec(align(128)) /4096 и др. степени 2/

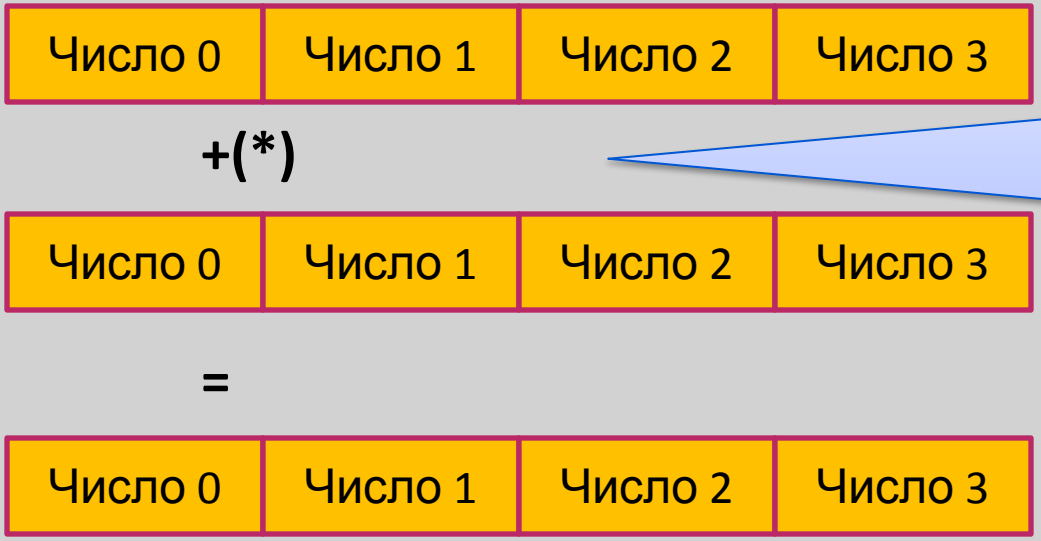
malloc и иже с ним – даёт не выровненные данные

Массивы нужно было выравнивать всегда (до последнего времени)

История развития компьютеров

4. Компьютер

Выравнивание (связано с векторизацией, т.е., с SSE, AVX и прочими подобными вещами):



Невыровненные данные увеличивают время вычислений в разы, но только в начале и конце серии однотипных вычислений

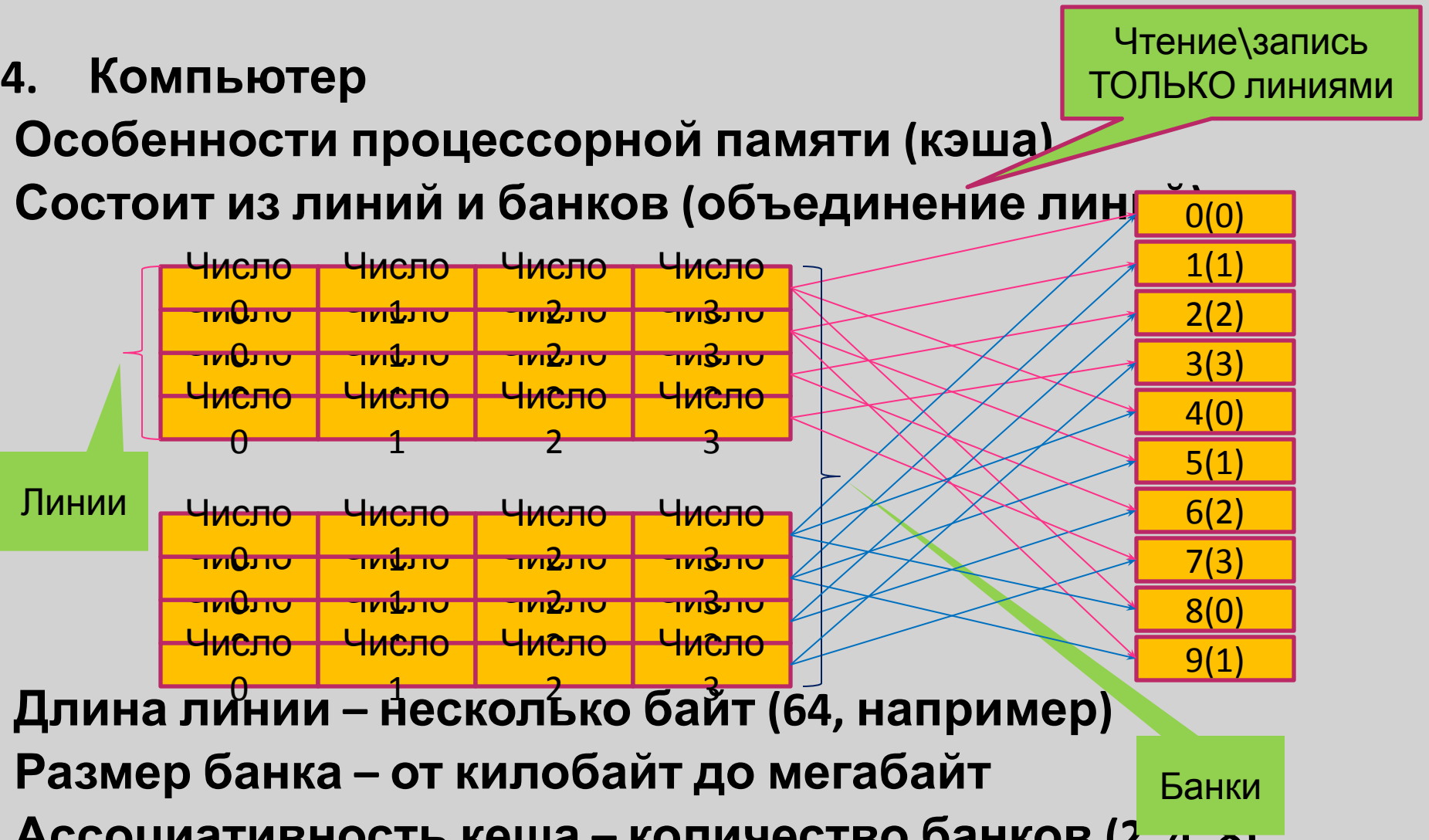
Чтобы выполнять такие операции, адрес «Число 0» должен быть кратен некоторому числу байт равному степени 2 (выровнен на это число байт) – это ускоряет программу и даёт стабильный результат по производительности на нынче устаревающих процессорах (например, процессоры Интел до Нехалема (Nehalem))

История развития компьютеров

4. Компьютер

Особенности процессорной памяти (кэша)

Состоит из линий и банков (объединение лин



Длина линии – несколько байт (64, например)

Размер банка – от килобайт до мегабайт

Ассоциативность кеша – количество банков (2, 4, 8)

<u>Следствие 1</u>: Избегать НЕ непрерывных обращений к памяти

Следствие 2: Избегать обращений к памяти кратных

История развития компьютеров

4. **Компьютер**

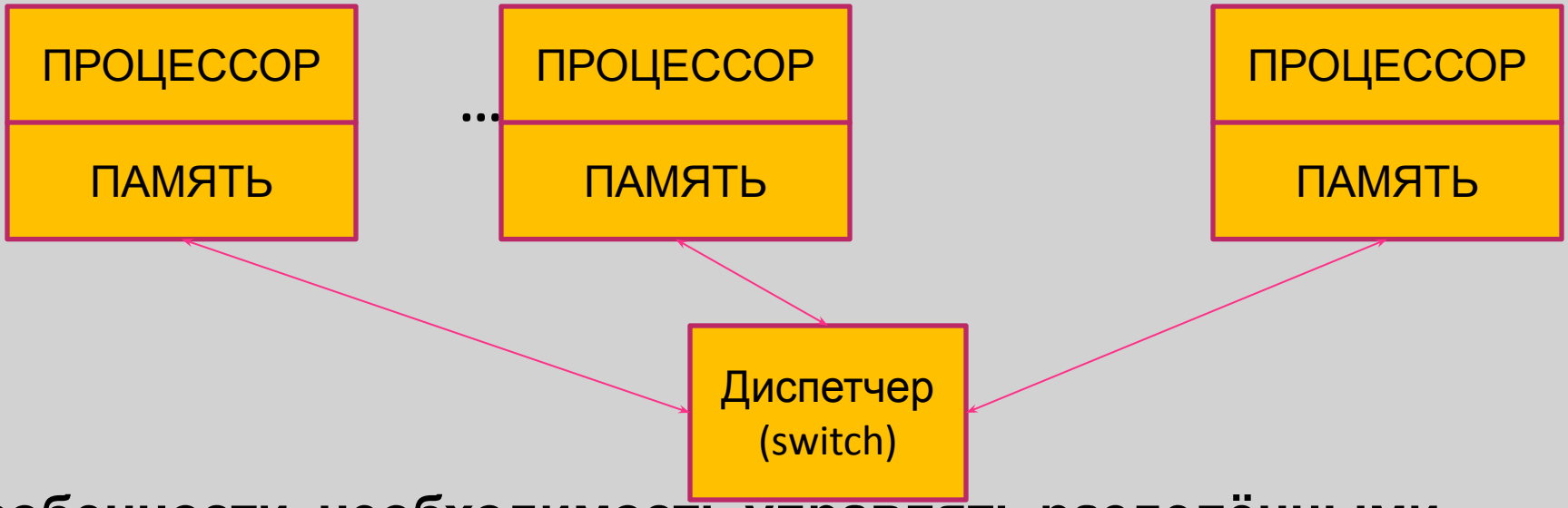
Дополнительно:

Скорость доставки данных растёт на порядок медленнее, чем скорость обработки оных данных (флопсы)

Появляется необходимость минимизировать пересылки из\в память и эксплуатировать параллелизм вычислений и (более медленной!) доставки данных

История развития компьютеров

5. Кластер

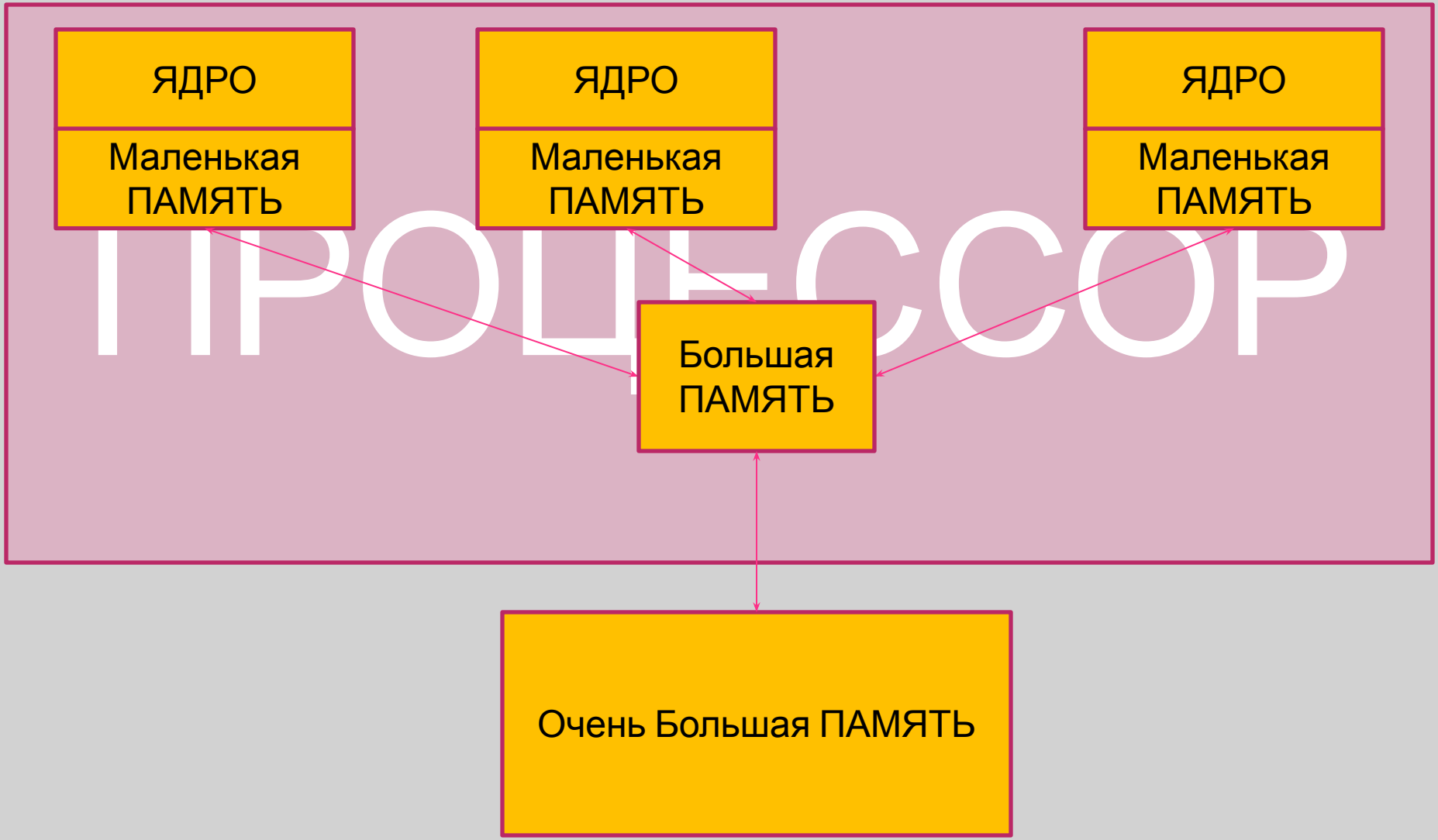


Особенности: необходимость управлять разделёнными данными, учёт скорости обмена данными, …

Та же самая проблема, что и в отдельном процессоре – скорость передачи данных мала по сравнению с вычислительными возможностями

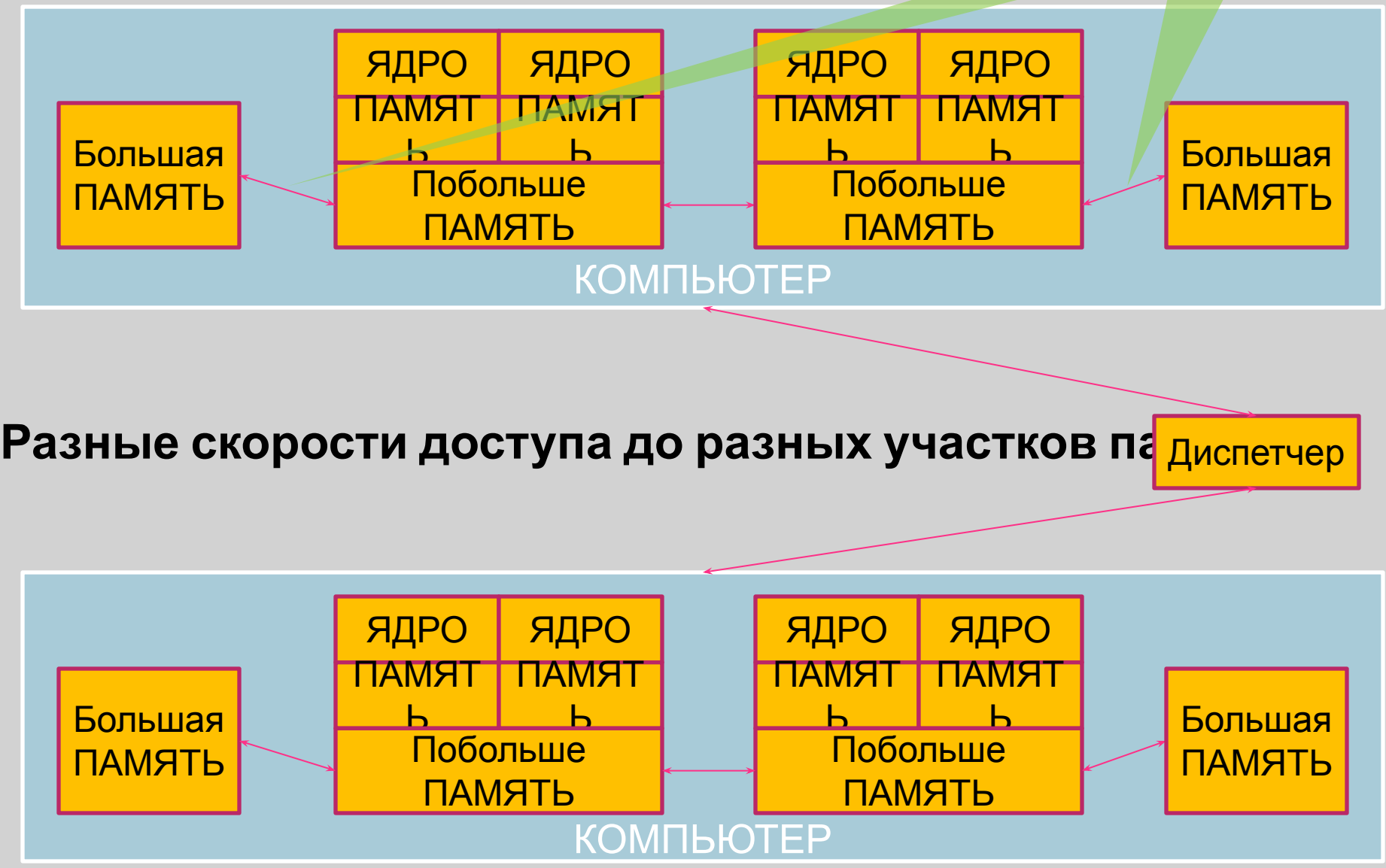
История развития компьютеров

6. Многоядерный компьютер



История развития компьютеров

7. Многоядерный кластер



Разные скорости доступа до разных участков па

История развития компьютеров

8. **Неоднородный многоядерный компьютер**
9. **Неоднородный многоядерный кластер**

То же самое, что и обычный многоядерный компьютер\кластер, только ядра могут иметь разные характеристики в пределах одного компьютера

История развития компьютеров

Итог:

Компьютер – это иерархия вычислительных модулей, иерархия памяти и связи между ними, работающие с существенно разными скоростями.

Дополнительно: графические карты – это тот же процессор с памятью. Сейчас в состоянии работать независимо от существования ЦПУ, в противном случае порождают неоднородный компьютер, где характеристики ЦПУ (мало ядер, но они много умеют) отличаются от характеристик карты (много ядер, но они мало что умеют)

Резюме

Вычислительные метдоы существуют с тех пор, как человек научился считать

Компьютеров много и разных – сначала нужно узнать, что за компьютер вам достался

Компьютер = процессор (Флопсы) + память (байты) + провода (биты в секунду)

Выравнивайте данные (массивы)

Избегайте последовательных обращений в массивах кратных размеру кэша (степень 2!)

Избегать НЕ непрерывных обращений к памяти

Задания на понимание

1. Найдите и проанализируйте схему ЦПУ на предмет особенностей
2. Найдите и проанализируйте схему графической карты на предмет особенностей
3. Посчитайте, сколько операций использующих У чисел каждая и занимающих 1 такт нужно проделать над К числами находящимися в памяти на процессоре, работающем на частоте М ГГц, чтобы за это время успеть загрузить в ту же память ещё К чисел по каналу, работающему со скоростью С бит в секунду.
4. Решите задачу 3 с двумя каналами в память идущими последовательно друг за другом, первый из которых работает со скоростью С1 бит\секунду, а второй С2 бит\секунду.
5. Пусть из процессора с частотой М ГГц идут два канала в память, последовательно друг за другом, первый из которых работает со скоростью С1 бит\секунду, а второй С2 бит\секунду. Между этими каналами есть память размером в Т чисел. Сколько времени число должно находиться в промежуточной памяти, чтобы второй канал не простаивал.