



iSCALARE



Лаборатория суперкомпьютерных технологий для биомедицины, фармакологии и малоразмерных структур

# Моделирование сверхоперативной памяти

Евгений Юлюгин

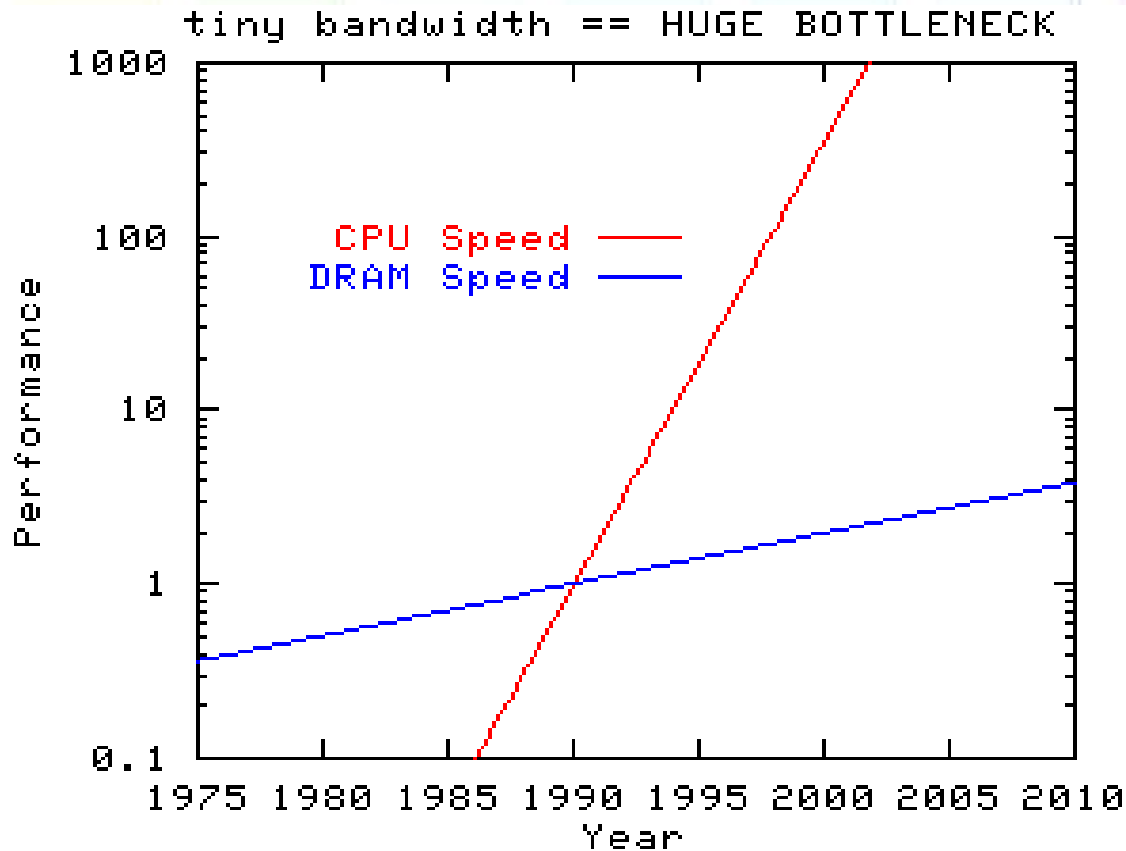
[yulyugin@gmail.com](mailto:yulyugin@gmail.com)



- Назначение
- Устройство и принципы работы
- Моделирование



# «Стена памяти» (memory wall)



Источник: <http://www.cs.virginia.edu/stream>

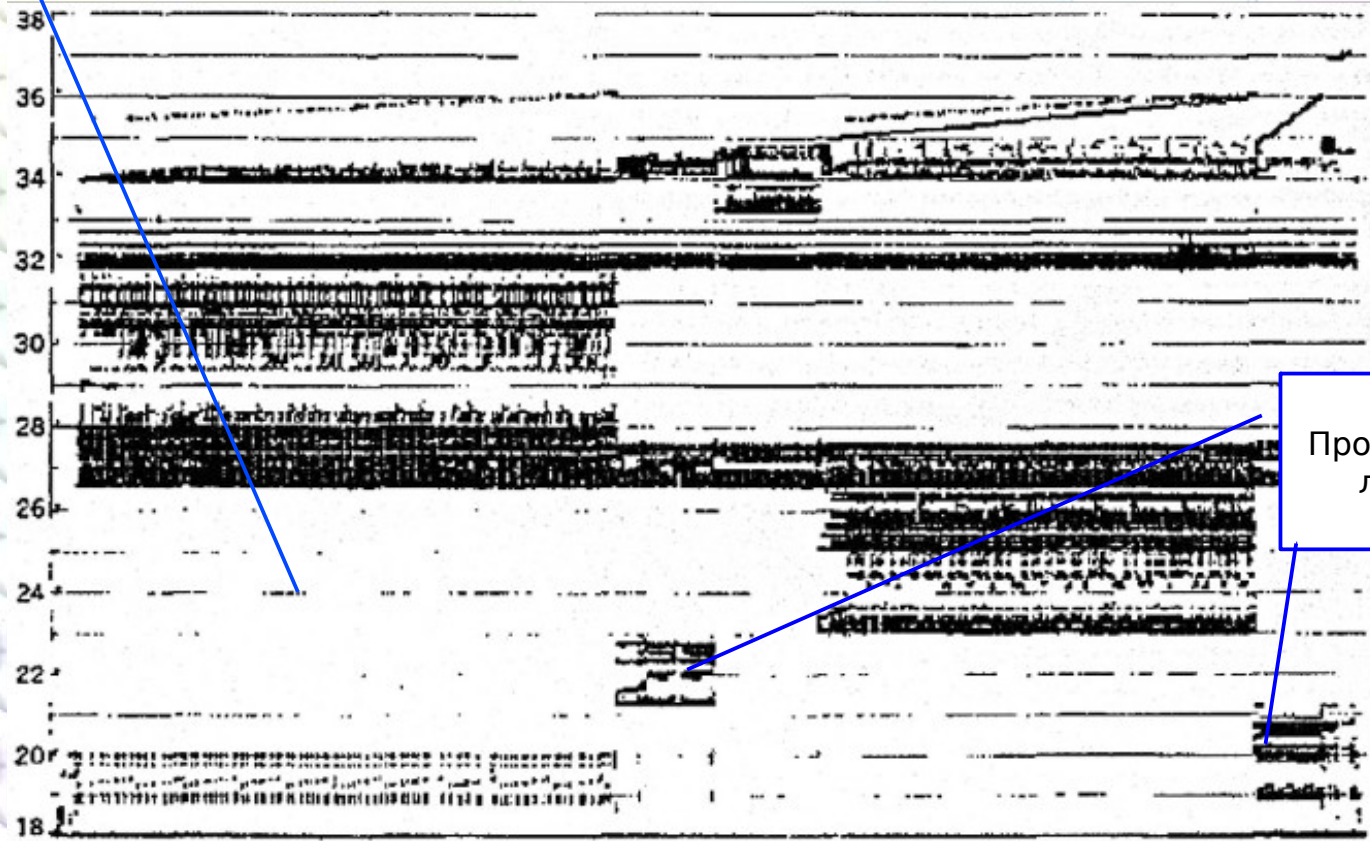
# Ускорение доступов в память

- Быстрое, небольшое хранилище
- Прозрачно для ISA
- Хранит недавно использованные данные, а также прилегающие к ним блоки

# Локальность (locality) данных

Временная  
локальность

Адрес

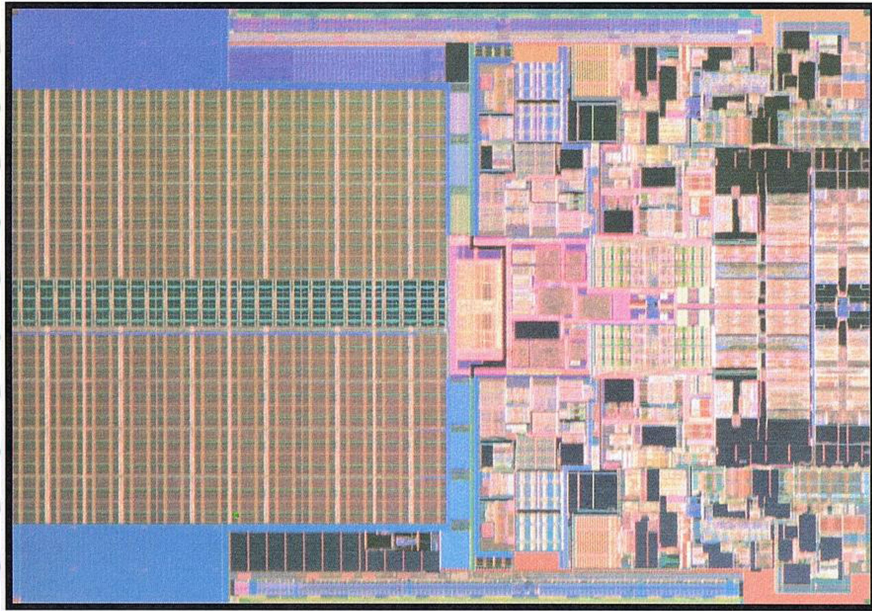


Пространственная  
локальность

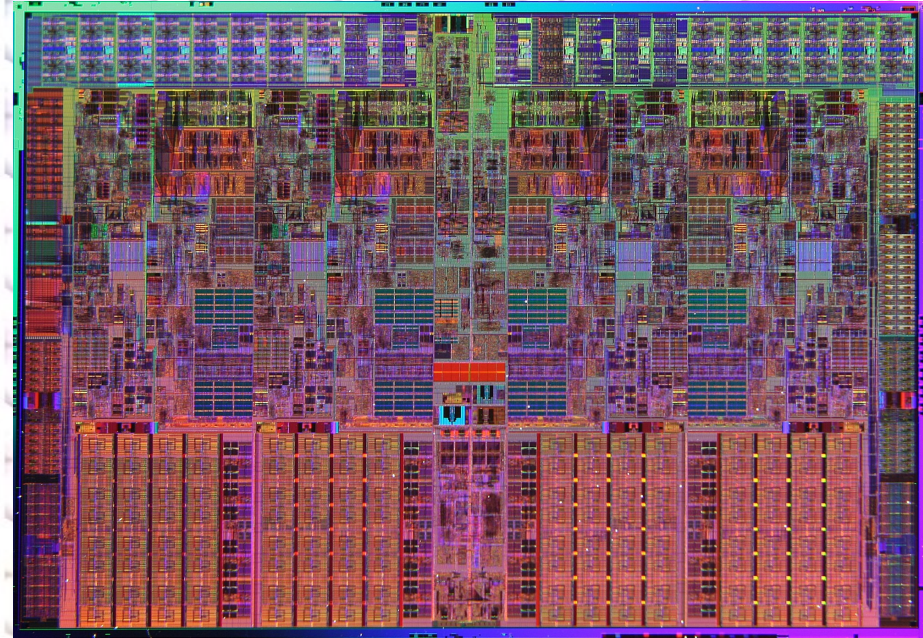
Время

# Реальные системы

Penryn



Nehalem EX



# КЭШ

$N_{ways}$

$N_{sets}$

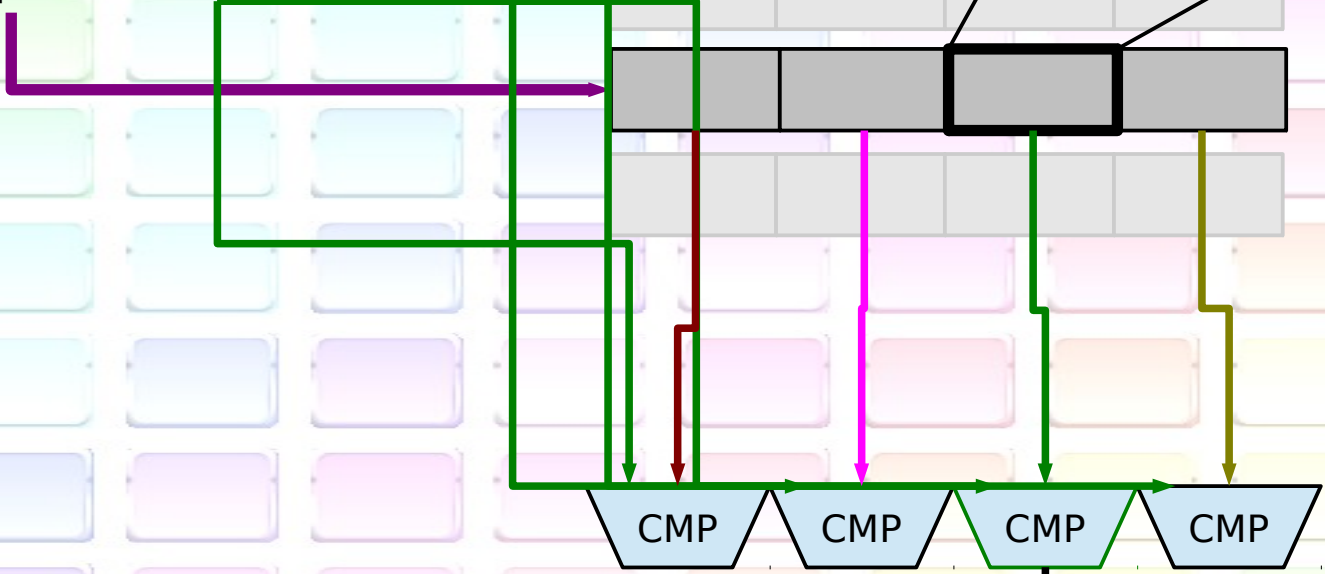
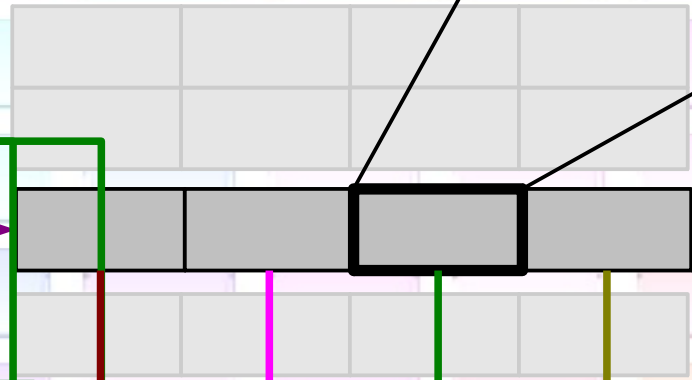


# Поиск в кэше

0x **aa** **bbb** 37

39      23      7      0

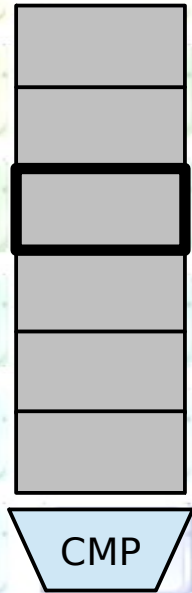
n\_set      tag      offset



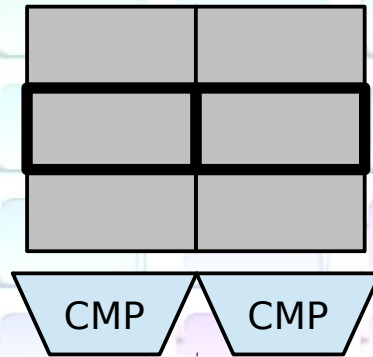


# Ассоциативность

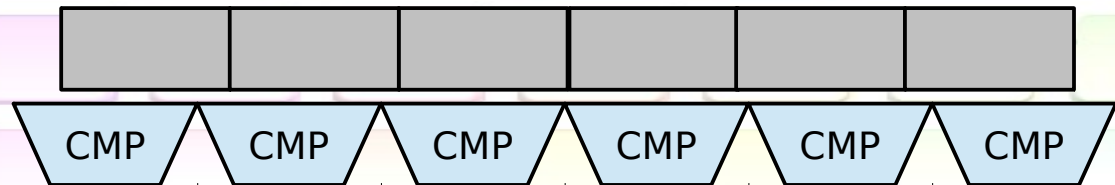
Прямого  
отображения



Двухпутевой



Полностью ассоциативный



# События при поиске в кэше

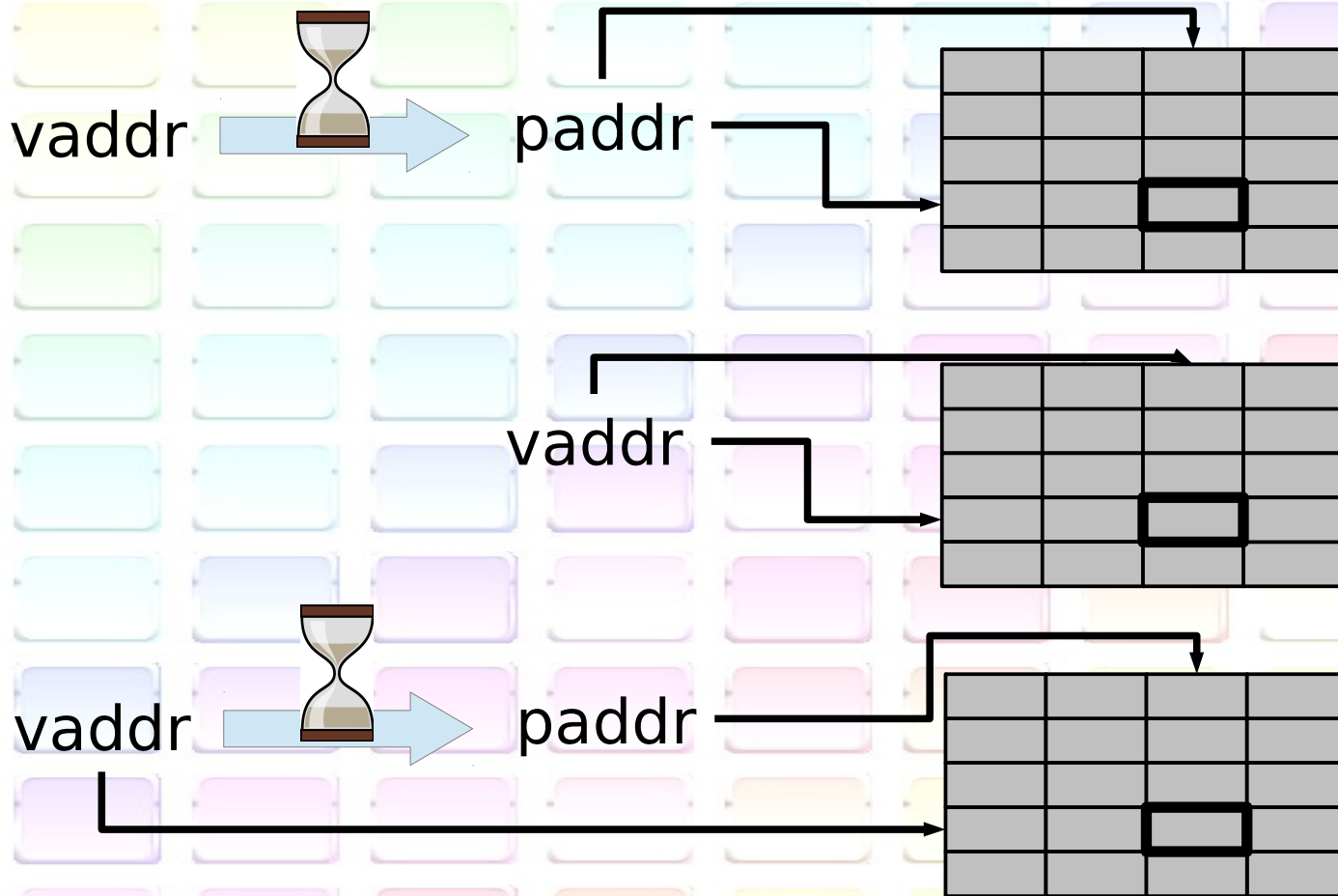
## Внешние

- Чтение – попадание
- Чтение – промах
- Запись – попадание
- Запись – промах

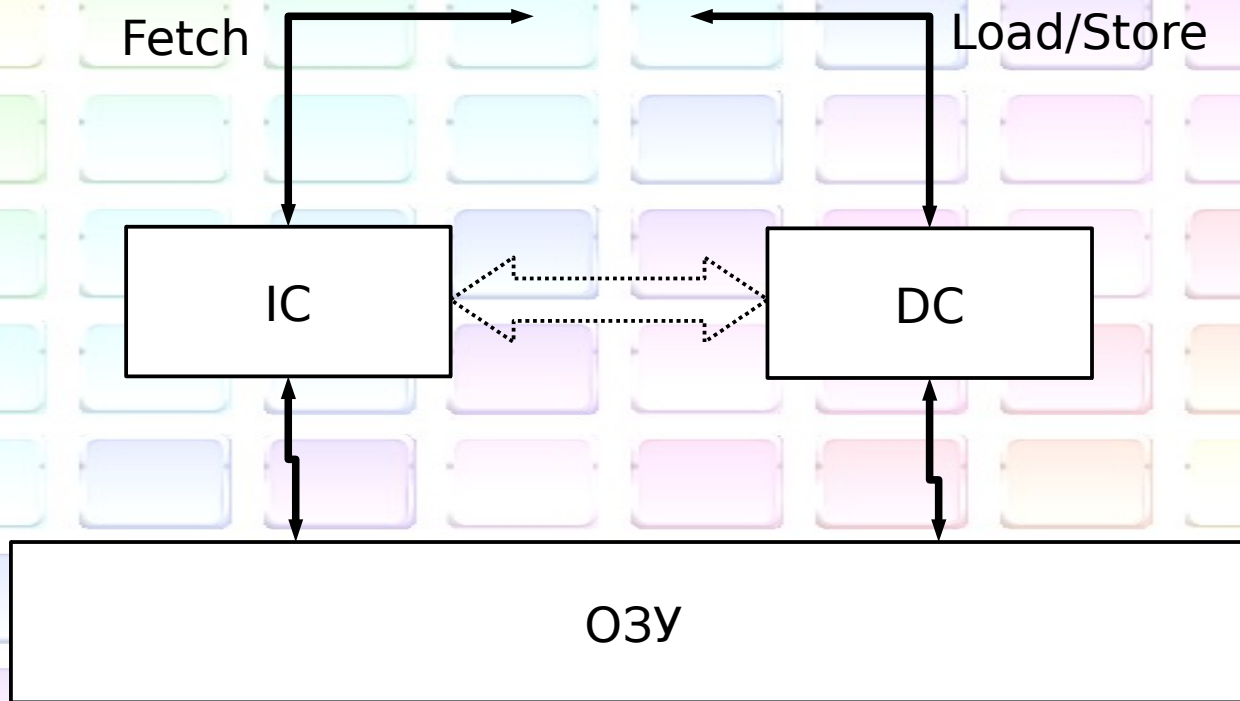
## Внутренние

- Поиск – тэг не найден
- Поиск – флаг V снят
- Добавление – нет свободной ячейки

# Кэши и физические/виртуальные адреса



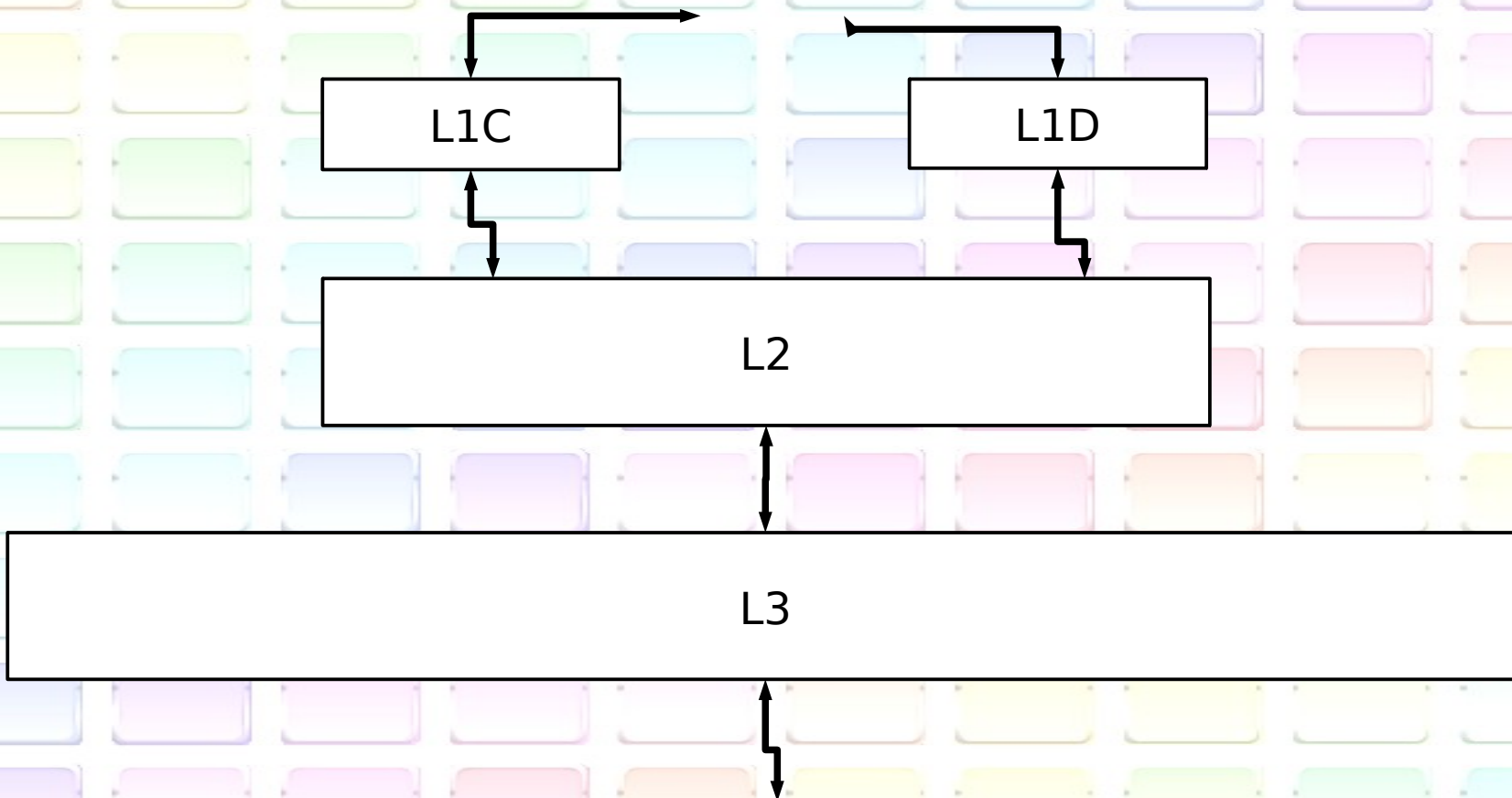
# Кэши инструкций и данных



# Многоуровневые кэши

- Один уровень кэша имеет ограничения по ёмкости/скорости
- Решение – иерархическая система
- Кэши высоких уровней имеют большую ёмкость и большее время задержки

# Многоуровневые кэши



# Политики запись в память

- WT (write through)
- WB (write back)
- WA (write allocate)
- WC (write combined)
- UC (uncacheable)

# Кэши в многопроцессорных системах

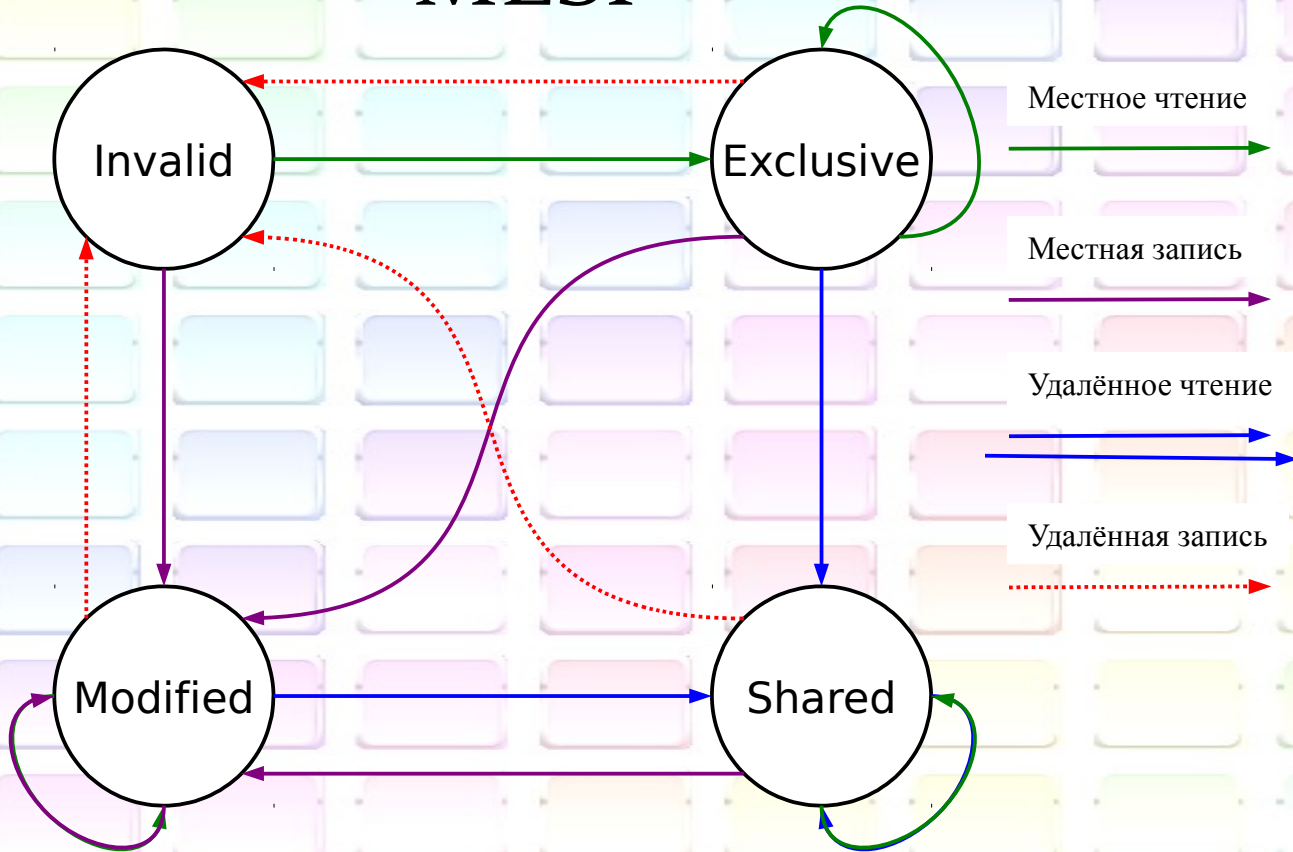
- В качестве кэширующих агентов выступают ЦПУ/другие устройства, имеющие DMA
- Исполняющимся программам необходимо иметь “консистентный” вид на память
- Он обеспечивается “подглядыванием” в чужие кэши/явными передачами сообщений и определяется протоколом когерентности системы



# Когерентность

CPU	Адрес								
	0хаabbccd0	0x15	0x12	###					
	0хаabbcсе0	0x11	0x11	0x11					

# MESI



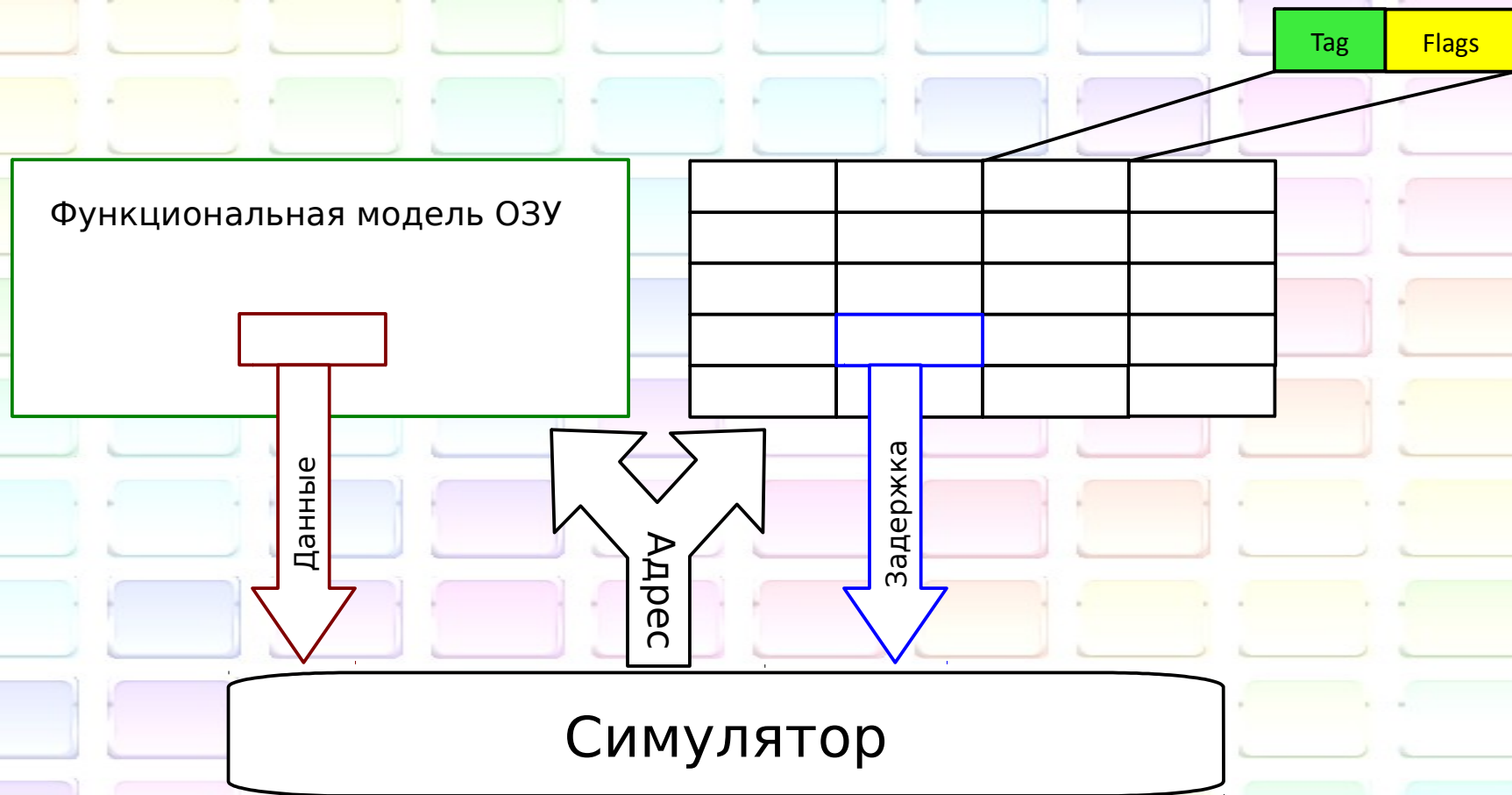
# Симуляция

- Для функциональных моделей наличие кэша необязательно
- Моделирование кэшей замедляет симуляцию на порядки
-

# «Честное» моделирование

- Прямая реализация модели по спецификации
- Параметризуемые модели:
  - Количество линий, их ёмкость
  - Ассоциативность, политики вытеснения, WB
  - Задержки операций
  - Соединения с другими уровнями
  - Протокол когерентности

# Модель задержки (timing model)

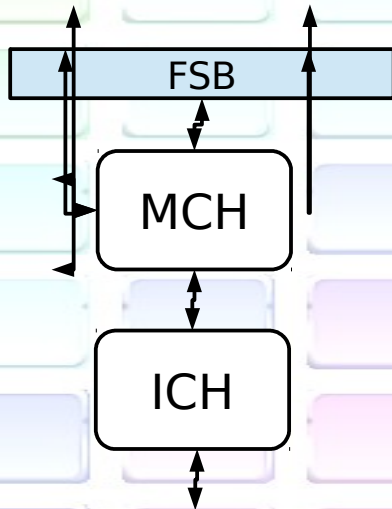


# Разогрев Кэша

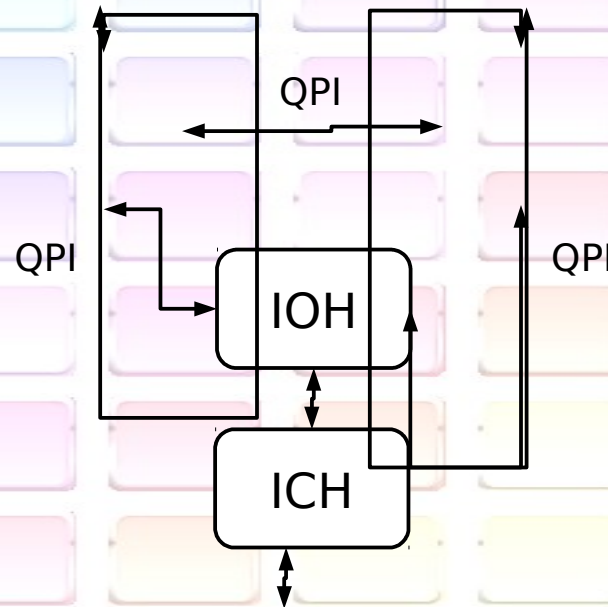
- TODO

# Кэши и масштабируемость

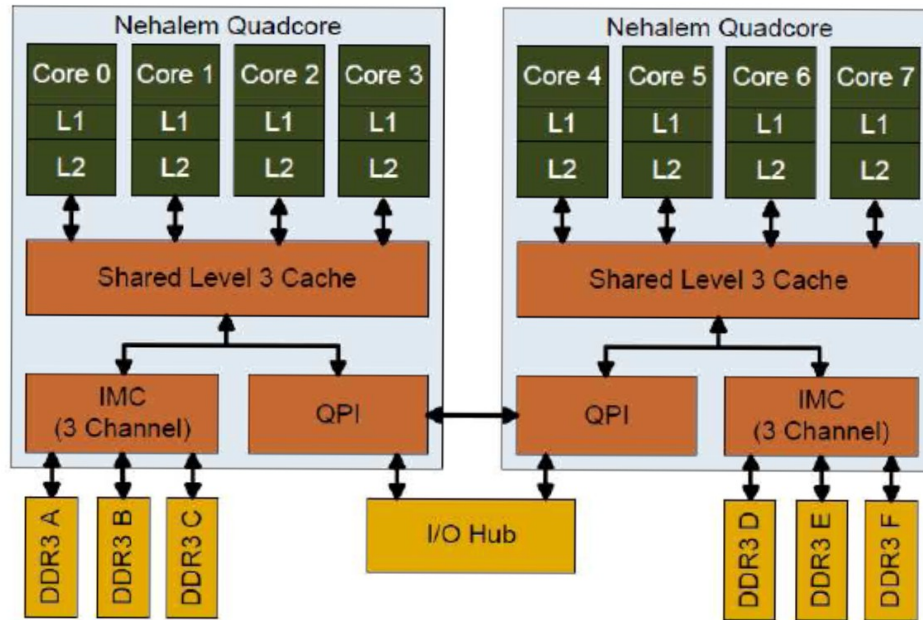
## FSB



## QPI



# Nehalem 4 cores x 2 sockets





# Рекомендуемая литература

- Хэннеси, Паттерсон.
- См. литературу к главе 9
- GEMS:<http://research.cs.wisc.edu/gems/>
- Gem5 [http://gem5.org/Main\\_Page](http://gem5.org/Main_Page)
- Ulrich Drepper “What Every Programmer Should Know About Memory”

# На следующей лекции:

- Не центральным процессором единым
  - Полноплатформенная симуляция
  - Исполняющие и неисполняющие устройства
- Моделирование многопроцессорных систем
  - Квант (квота) времени
  - Гиперсимуляция

# Спасибо за внимание!

Все материалы курса выкладываются на сайте лаборатории:

[http://iscalare.mipt.ru/material/course\\_materials/](http://iscalare.mipt.ru/material/course_materials/)

Замечание: все торговые марки и логотипы, использованные в данном материале, являются собственностью их владельцев.  
Представленная здесь точка зрения отражает личное мнение автора, не выступающего от лица какой-либо организации.