

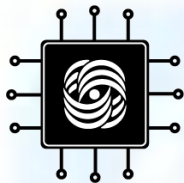
# **АРХИТЕКТУРА СОВРЕМЕННЫХ ЭВМ**

## **Лекция 09: Нейрокомпьютеры и современные процессоры**

ВМиК МГУ им. М.В. Ломоносова, Кафедра АСВК

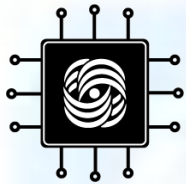
Чл.-корр., профессор, д.ф.-м.н. Королёв Л.Н.,

Ассистент Волканов Д.Ю.



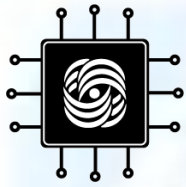
# План лекции

- Биологический нейрон
- Модель нейровычислений
- Проблемы организации
- Примеры
- Современные микропроцессоры
  - Машины фон-Неймана
  - Поточковые машины (dataflow)
- Intel Core i7 (Nehalem)
- E2

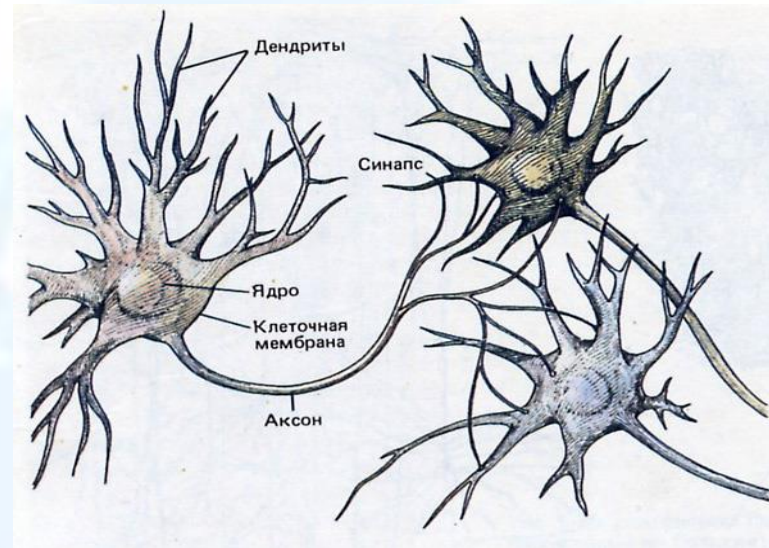


# Что такое нейрокомпьютер?

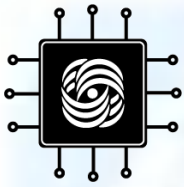
- Вычислительная система с MSIMD архитектурой
- Процессорный элемент упрощён до уровня нейрона
- Резко усложнены связи между элементами
- Программирование перенесено на изменение весовых коэффициентов связей между элементами



# Биологический Нейрон



- Мозг содержит 10 миллиардов нейронов
- Тысячи типов нейронов соединены между собой
- 100 триллионов связей
- Нейрон может находиться, либо в спокойном, либо возбуждённом состоянии
- Порог перехода



# Математическая модель нейрона

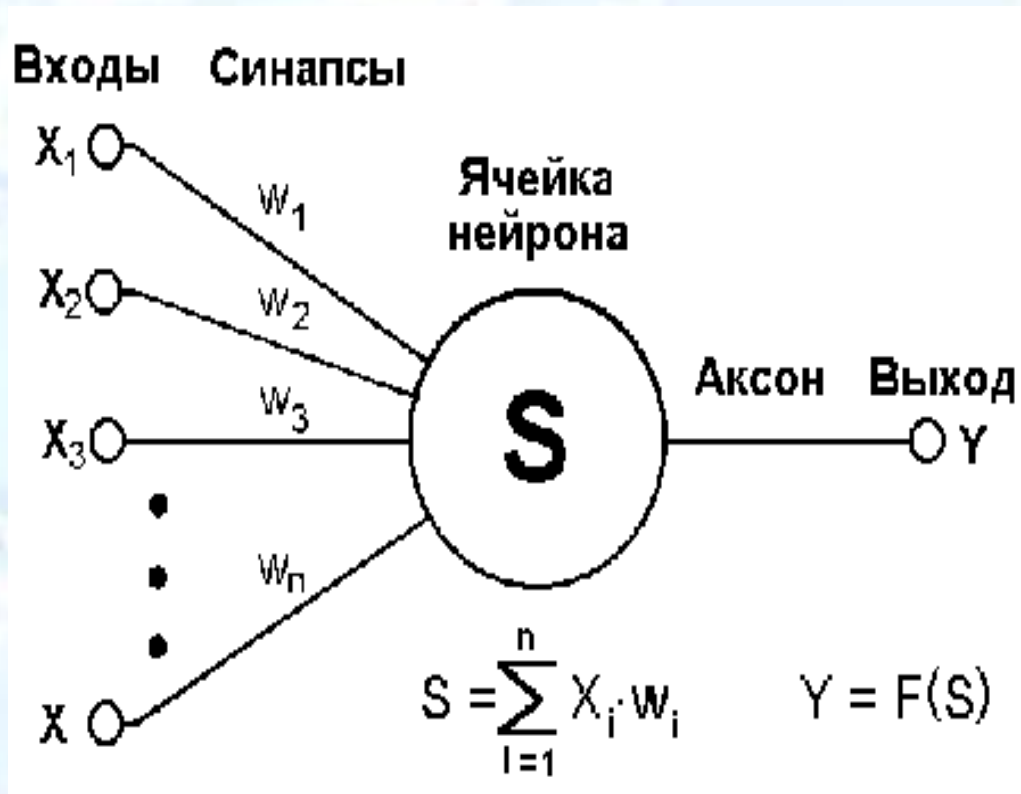
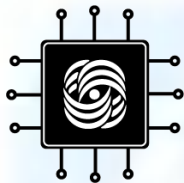
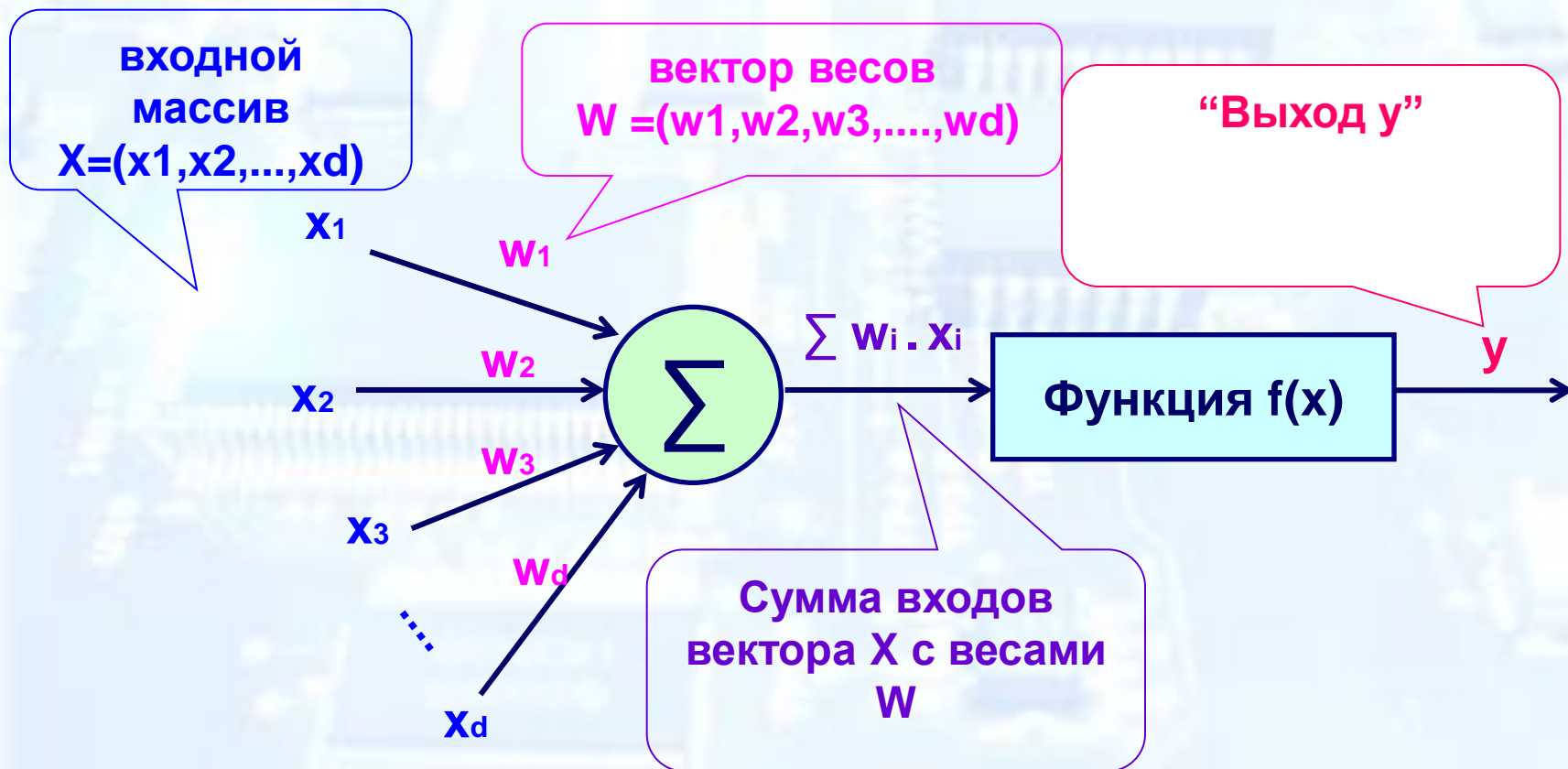
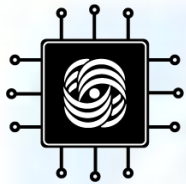


Рисунок 1. Математическая модель нейрона

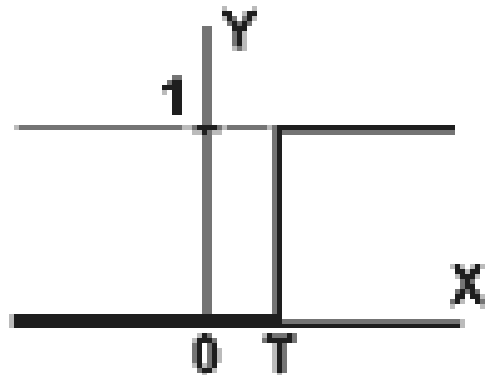


# Имитатор Нейрона

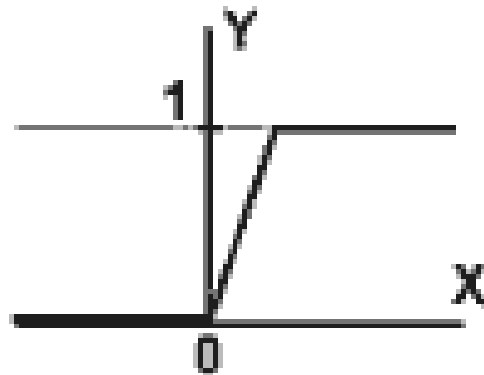




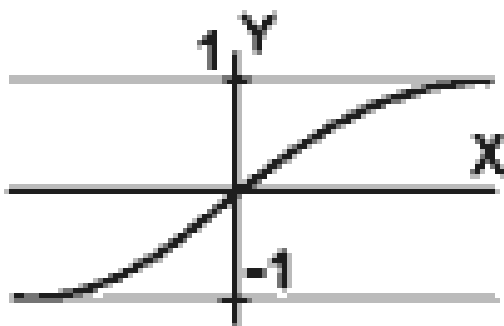
# Примеры функций



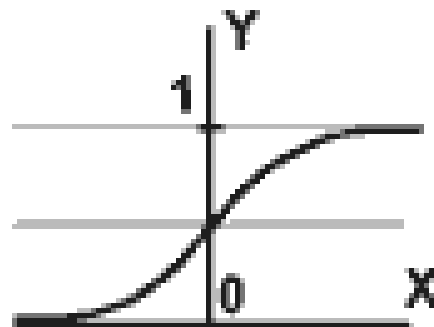
а)



б)

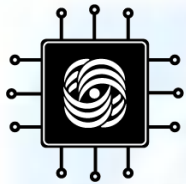


в)

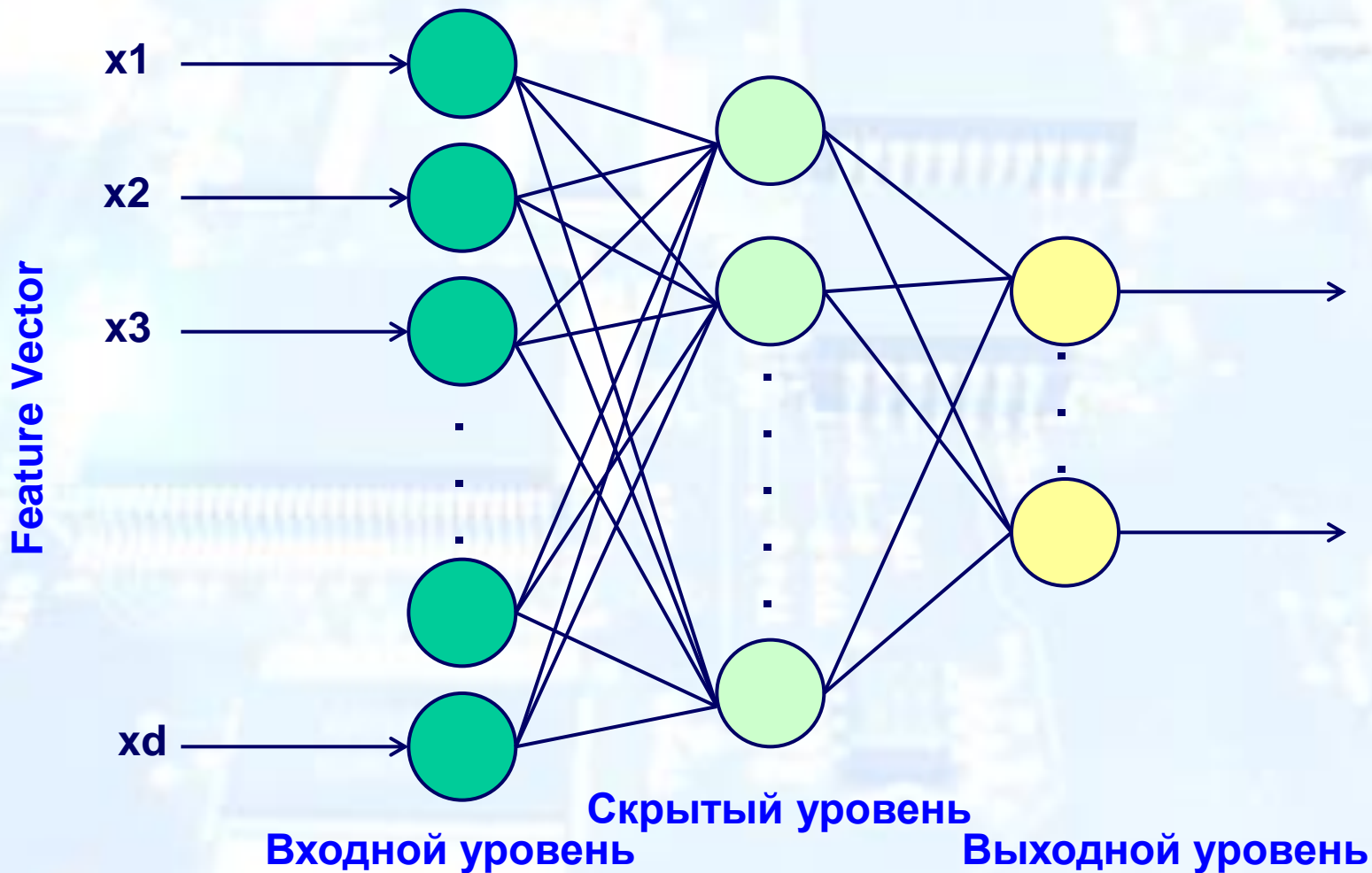


г)

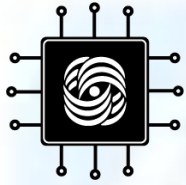




# Многоуровневая нейронная сеть

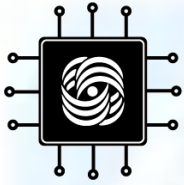






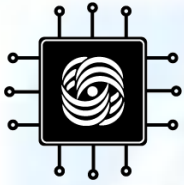
# Проблема обучения

- Обучение с готовыми ответами
- Обучение без ответов



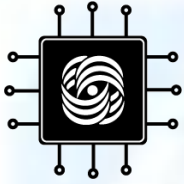
# Применение нейронных сетей

- Управление в реальном времени
- Распознавание образов
- Прогнозирование в реальном времени
- Оптимизация
- Обработка сигналов при наличии больших шумов
- Протезирование
- Психодиагностика



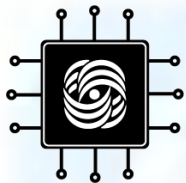
# Варианты построения нейрокомпьютеров

- На основе обычных процессоров
- На основе DSP процессоров
- На специализированной аппаратной базе



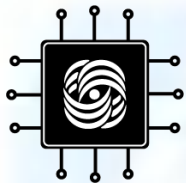
# Характеристики

- CUPS (connections update per second) - число измененных значений весов в секунду (оценивает скорость обучения).
- CPS (connections per second) - число соединений (умножений с накоплением) в секунду (оценивает производительность).
- $CPS_{PW} = CPS/N_w$ , где  $N_w$  - число синапсов в нейроне.
- CPPS - число соединений примитивов в секунду,  $CPPS = CPS * B_w * B_s$ , где  $B_w$ ,  $B_s$  - разрядность весов и синапсов.
- ММАС - миллионов умножений с накоплением в секунду.



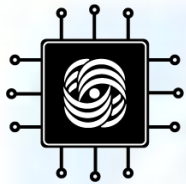
# Особенности DSP процессоров

- Аппаратная поддержка программных циклов, кольцевых буферов
- Один или несколько операндов извлекаются из памяти в цикле исполнения команды
- Нет команд  $R,R \rightarrow R$
- Реализация однотактного умножения и команд, использующих в качестве операндов содержимое ячеек памяти



# Особенности DSP процессоров (2)

- Сложение и умножение требуют:
  - произвести выборку двух операндов
  - выполнить сложение или умножение (обычно и то и другое)
  - сохранить результат или удерживать его до повторения
- Множественный доступ к памяти за один и тот же командный цикл.

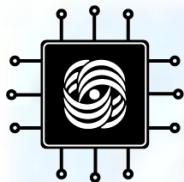


# Процессор NM6403



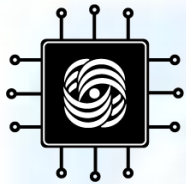
- 50 Mhz
- RISC ядро
  - 32-битные данные
  - 32-битные операции
  - 8 + 8 регистров
- Векторное устройство
  - Переменная разрядность
  - До 2048 параллельных умножений





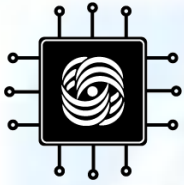
# RISC-ядро

- 5-ти ступенчатый 32-разрядный конвейер;
- 32- и 64-разрядные команды (обычно выполняется две операции в одной команде);
- Два адресных генератора, адресное пространство - 16 GB;
- Два 64-разрядных программируемых интерфейса с SRAM/DRAM-разделяемой памятью;
- Формат данных - 32-разрядные целые;
- Регистры:
  - 8 32-разрядных регистров общего назначения;
  - 8 32-разрядных адресных регистров;
  - Специальные регистры управления и состояния;
- Два высокоскоростных коммуникационных порта ввода/вывода,
- Аппаратно совместимых с портами TMS320C4х.



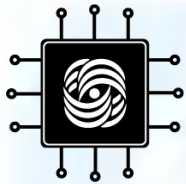
# VECTOR-сопроцессор

- Переменная 1-64-разрядная длина векторных операндов и результатов;
- Формат данных - целые числа, упакованные в 64-разрядные блоки, в форме слов переменной длины от 1 до 64 разрядов каждое;
- Поддержка векторно-матричных и матрично-матричных операций;
- Два типа функций насыщения на кристалле;
- Три внутренних 32x64-разрядных RAM-блока



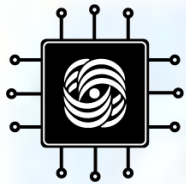
# Производительность

- Скалярные операции:
  - 50 MIPS;
  - 200 MOPS для 32-разрядных данных;
- Векторные операции:
  - от 50 до 50.000+ ММАС (миллионов умножений с накоплением в секунду);
- I/O и интерфейсы с памятью:
  - пропускная способность двух 64-разрядных интерфейсов с памятью - до 800 Мбайт/сек;
- I/O коммуникационные порты - до 20 Мбайт/сек кажд



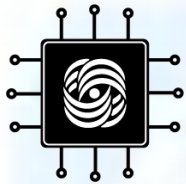
# Особенности NM64003 (1)

- Возможность работы с входными сигналами (синапсами) и весами переменной разрядности (от 1 до 64 бит), задаваемой программно, что обеспечивает уникальную способность нейропроцессора увеличивать производительность с уменьшением разрядности операндов;
- Быстрая подкачка новых весов на фоне вычислений;
- (24 операции умножения с накоплением за один такт при длине операндов 8 бит);
- V аппаратная поддержка эмуляции нейросетей большой размерности;
- Реализация функции активации в виде пороговой функции или функции ограничения;



# Особенности NM64003 (2)

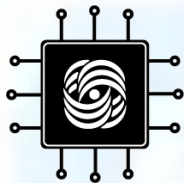
- Наличие двух широких шин (по 64 разряда) для работы с внешней памятью любого типа: до 4Мб SRAM и до 16 Гб DRAM;
- Наличие двух байтовых коммуникационных портов ввода/вывода, аппаратно совместимых с коммуникационными портами TMS320C4x для реализации параллельных распределенных вычислительных систем большой производительности.
- Возможность работать с данными переменной разрядности по различным алгоритмам, реализуемым с помощью хранящихся во внешнем ОЗУ программ



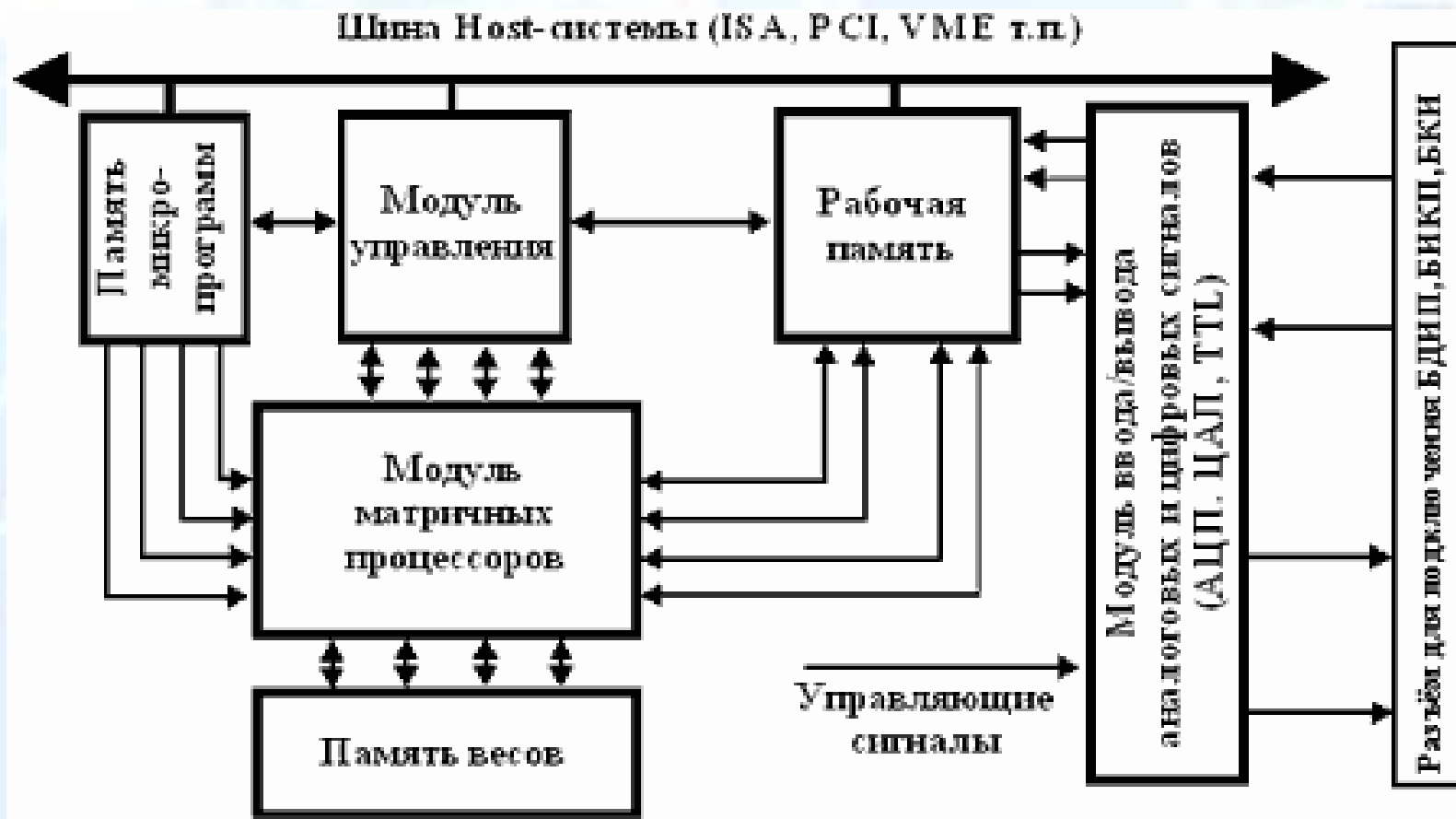
# Системы на NM 6403

- MC431 – однопроцессорная плата
- NM4 – четырехпроцессорная плата
- 6МСВО – 4 платы по 6 процессоров и платы для обработки входных сигналов

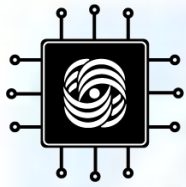




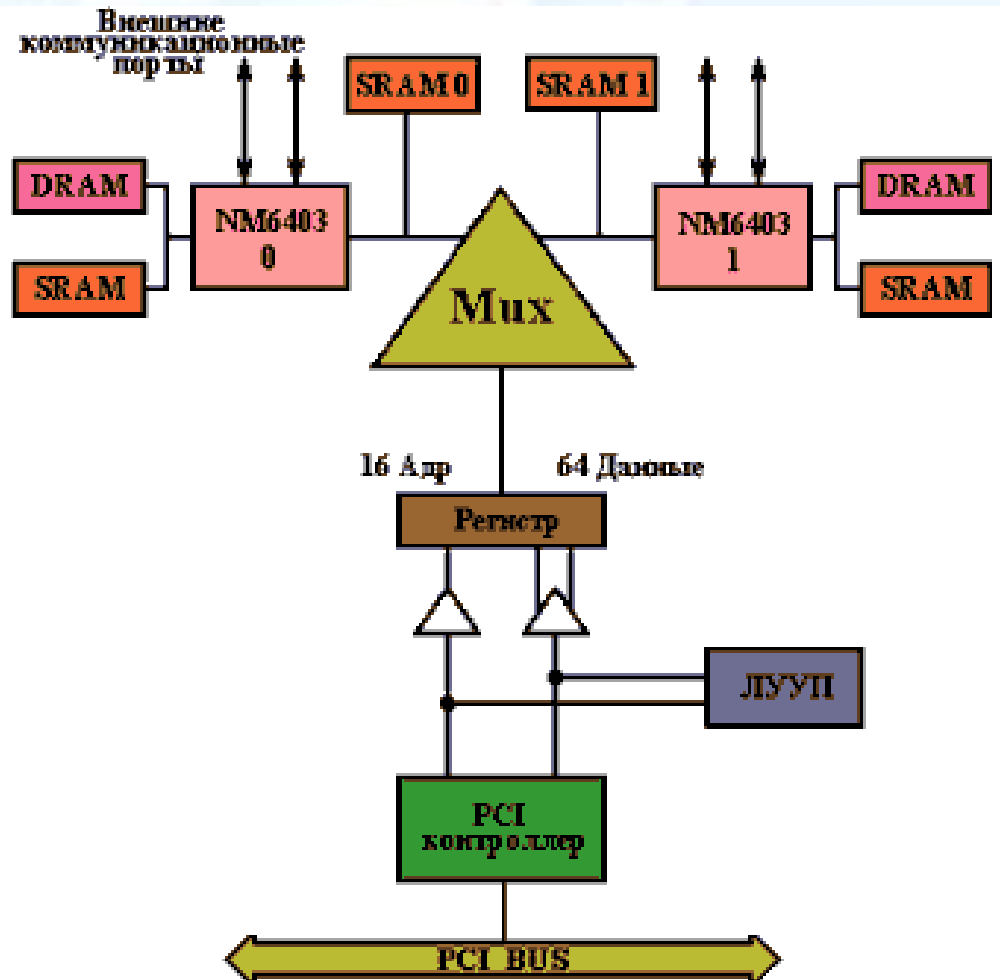
# Схема нейровычислителя

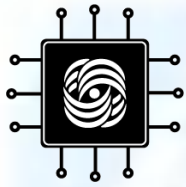






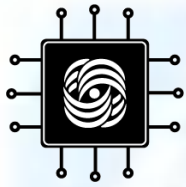
# Нейроускоритель МЦ 4.01





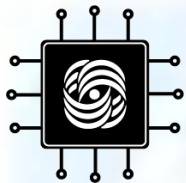
# Преимущества нейрокомпьютеров

- Все алгоритмы нейроинформатики высокопараллельны
- Устойчивость нейросистем к помехам и разрушениям
- Устойчивые и надежные нейросистемы могут создаваться и из ненадежных элементов, имеющих значительный разброс параметров



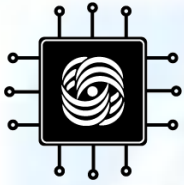
# Особенности нейрокомпьютеров

- Массовый параллелизм
- Распределённое представление информации и вычисления
- Способность к обучению и обобщению
- Адаптивность
- Свойство контекстуальной обработки информации
- Толерантность к ошибкам
- Низкое энергопотребление



# План лекции

- Современные микропроцессоры
  - Машины фон-Неймана
  - Поточковые машины (dataflow)
- Intel Core i7 (Nehalem)
- E2



# Фон-Нейман vs Dataflow

$$\mathbf{x} = \mathbf{y} * \mathbf{a};$$

$$\mathbf{z} = \mathbf{y} + \mathbf{b};$$

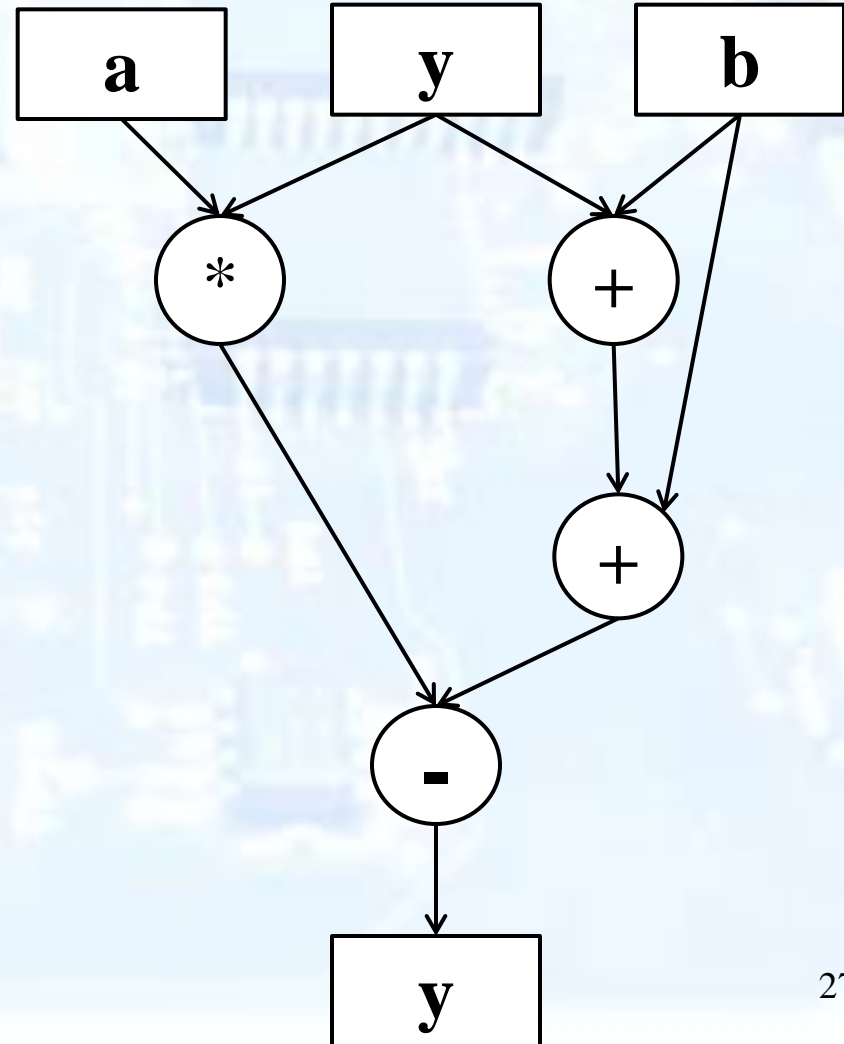
$$\mathbf{z} = \mathbf{z} + \mathbf{b};$$

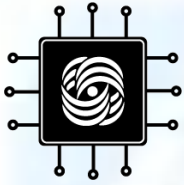
$$\mathbf{y} = \mathbf{x} - \mathbf{z};$$

**Поток сигналов  
управления**

VS

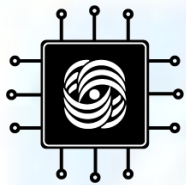
**Поток данных**



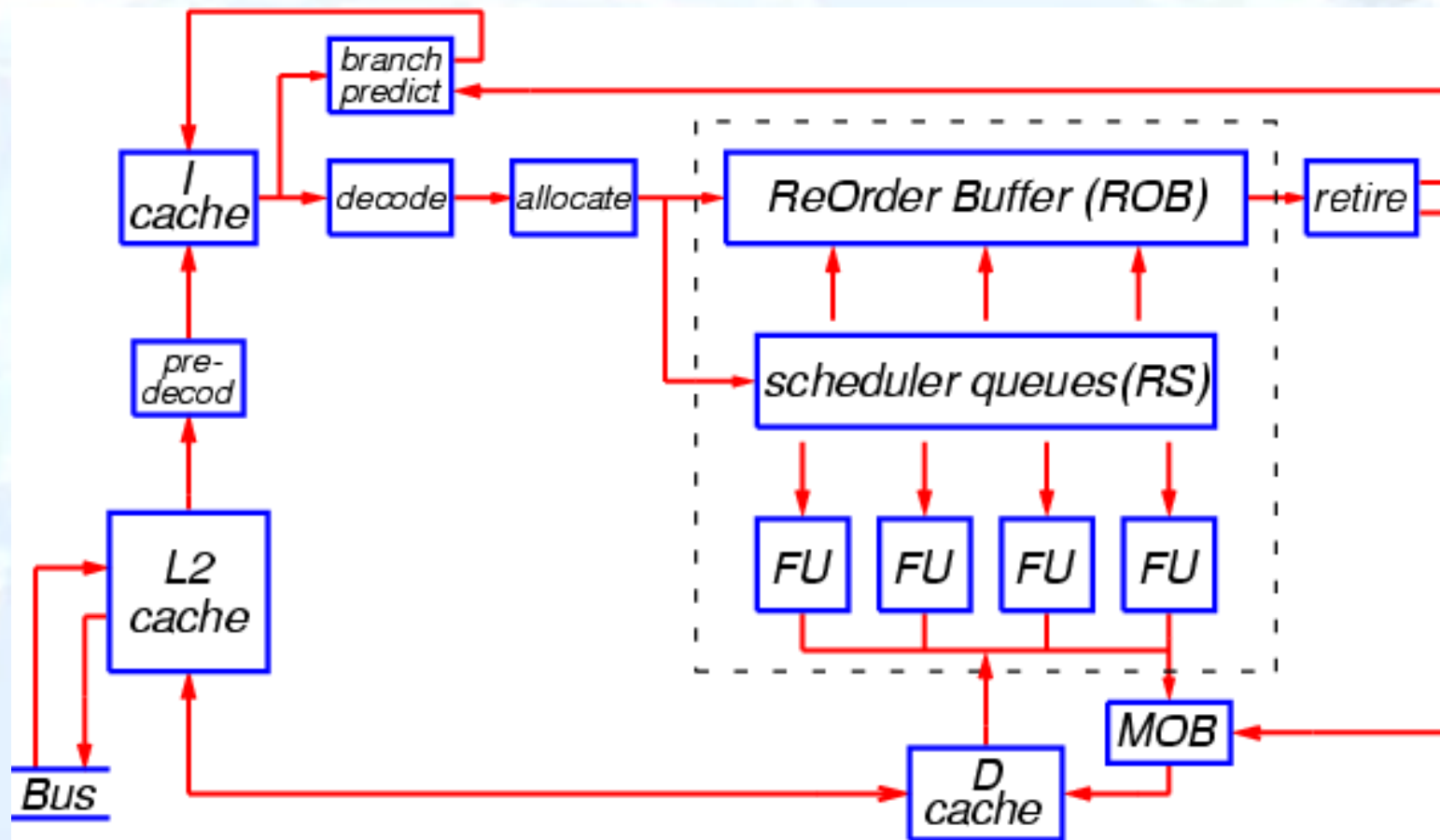


# Современные процессоры

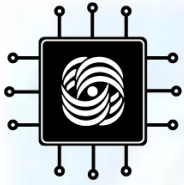
- **Конвейерная организация** – многие сложные действия разбиваются на этапы с небольшим временем выполнения, каждый этап выполняется параллельно;
- **Суперскалярная организация** - на каждом этапе обрабатываются сразу несколько потоков инструкций в параллель;
- **Внеочередное исполнение** – операции не обязаны выполняться в ФУ строго в том порядке, который определен в программном коде;
- **Спекулятивное выполнение** – исполнение инструкций, которые не должны были выполняться (ошибочная ветвь исполнения);
- **Предсказание переходов** – предсказание программного адреса перехода и направления ветвления.



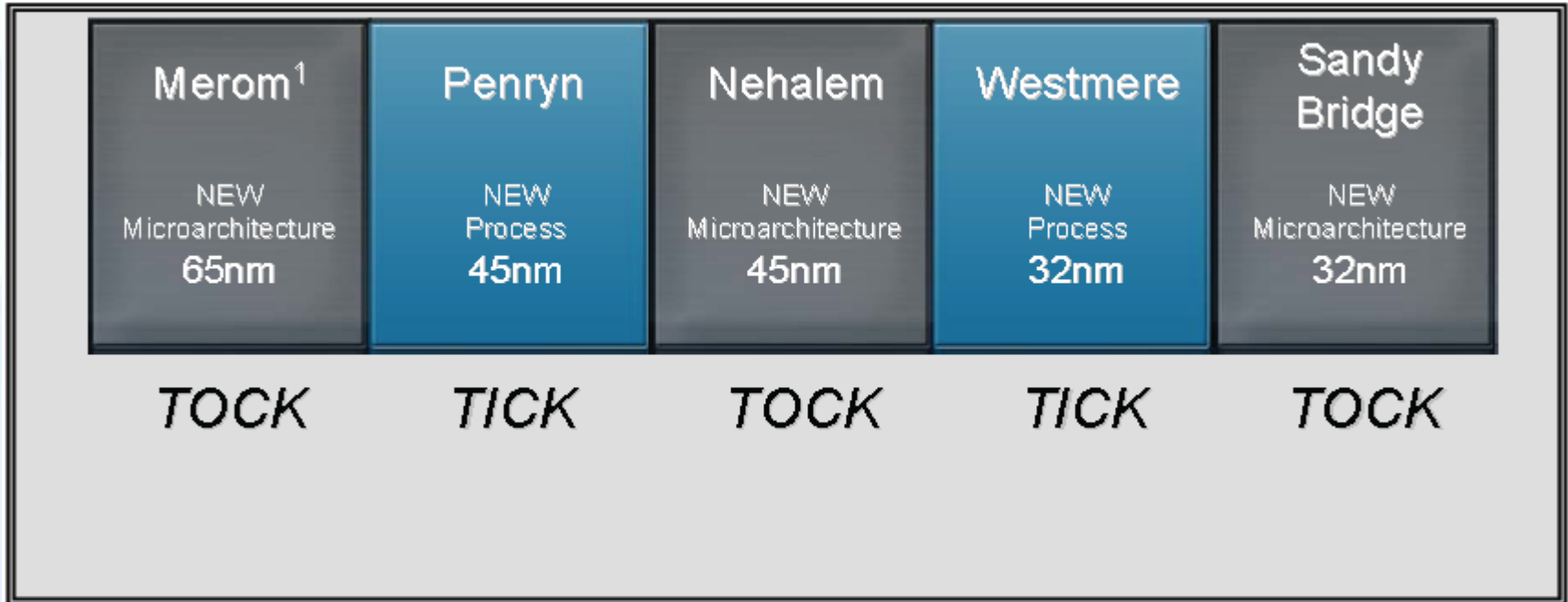
# Схема современного процессора

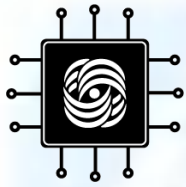






# Intel Core i7 (Nehalem)

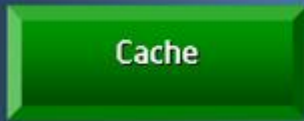




# “Кирпичики” Nehalem

## Nehalem Design Scalable Via Modularity

Nehalem Building Block Library



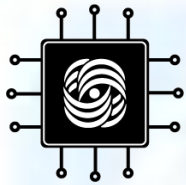
Ex: 4 Core



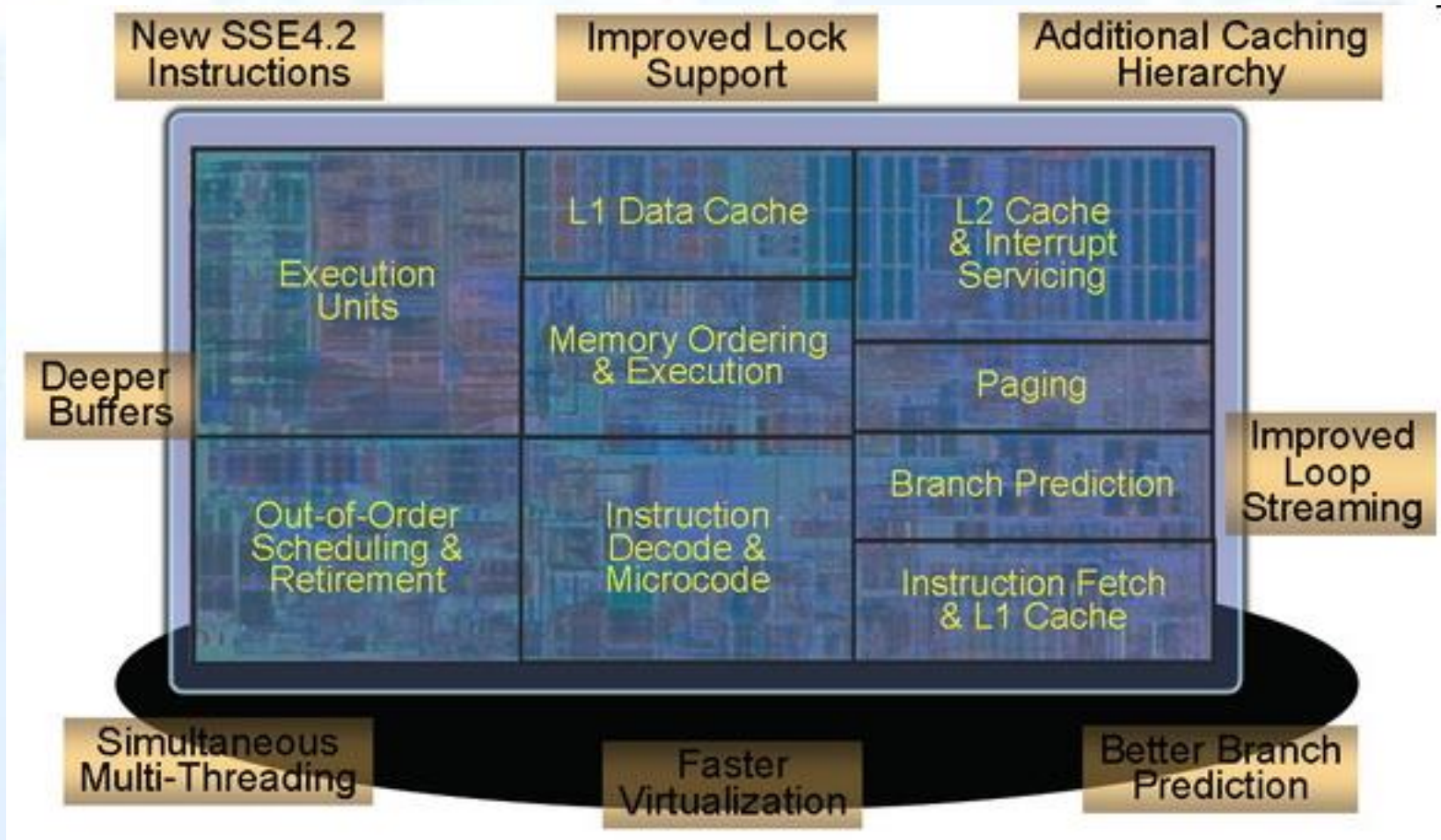
Ex: 8 Core

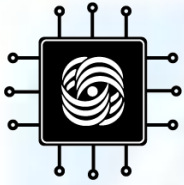


Sample Range of Product Options

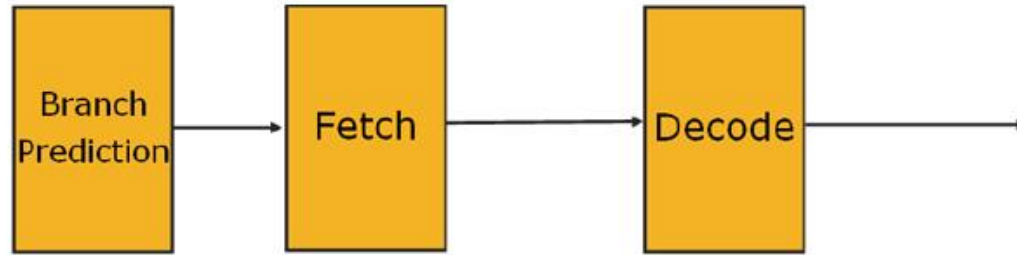


# Архитектура ядра

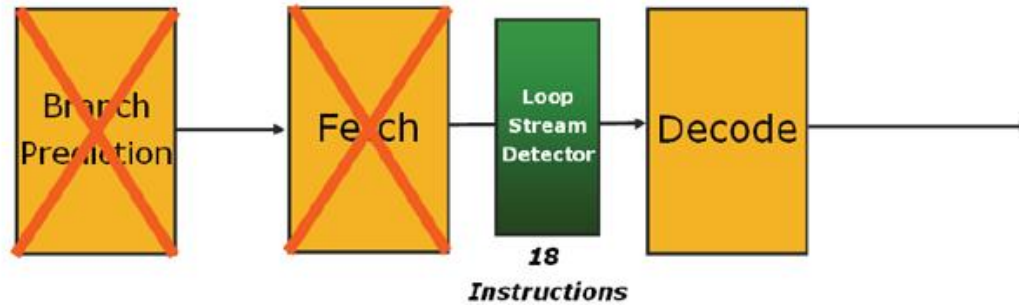




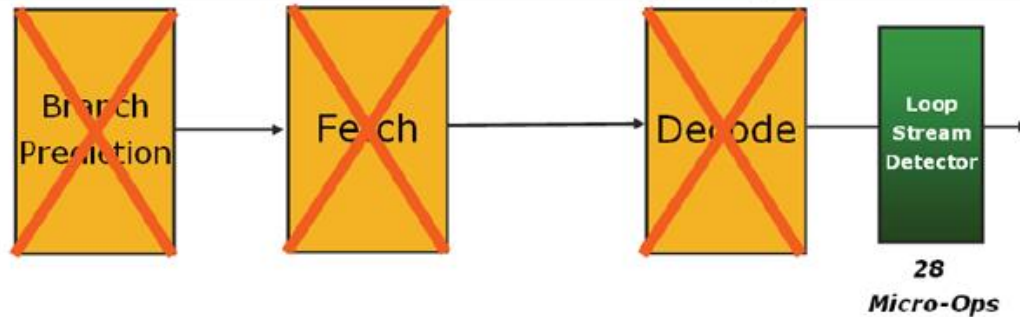
# Loop Stream Detector

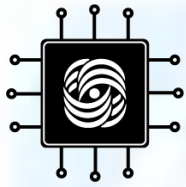


## ***Intel® Core™2 Loop Stream Detector***



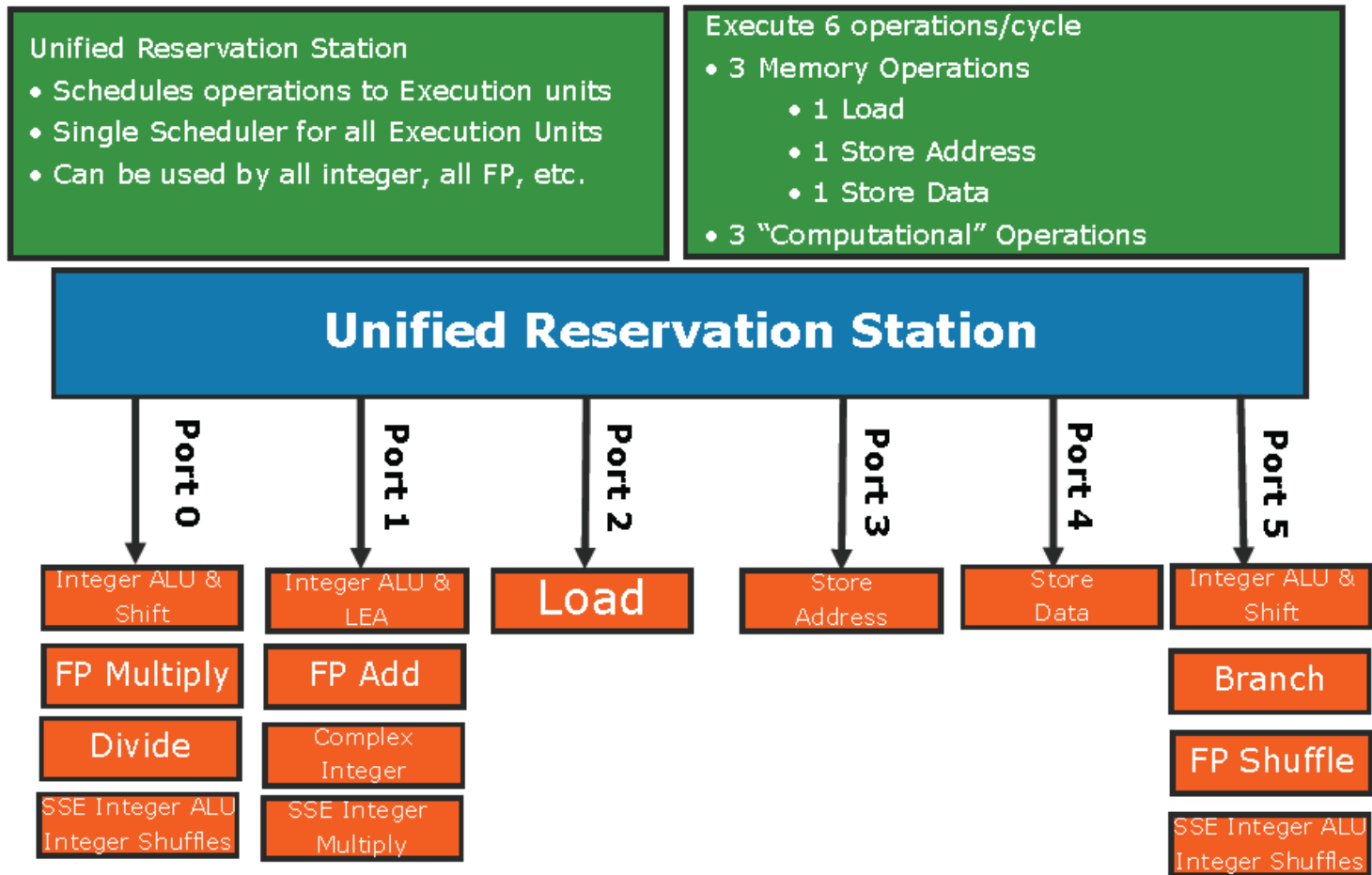
## ***Intel Core Microarchitecture (Nehalem) Loop Stream Detector***

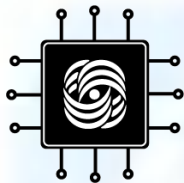




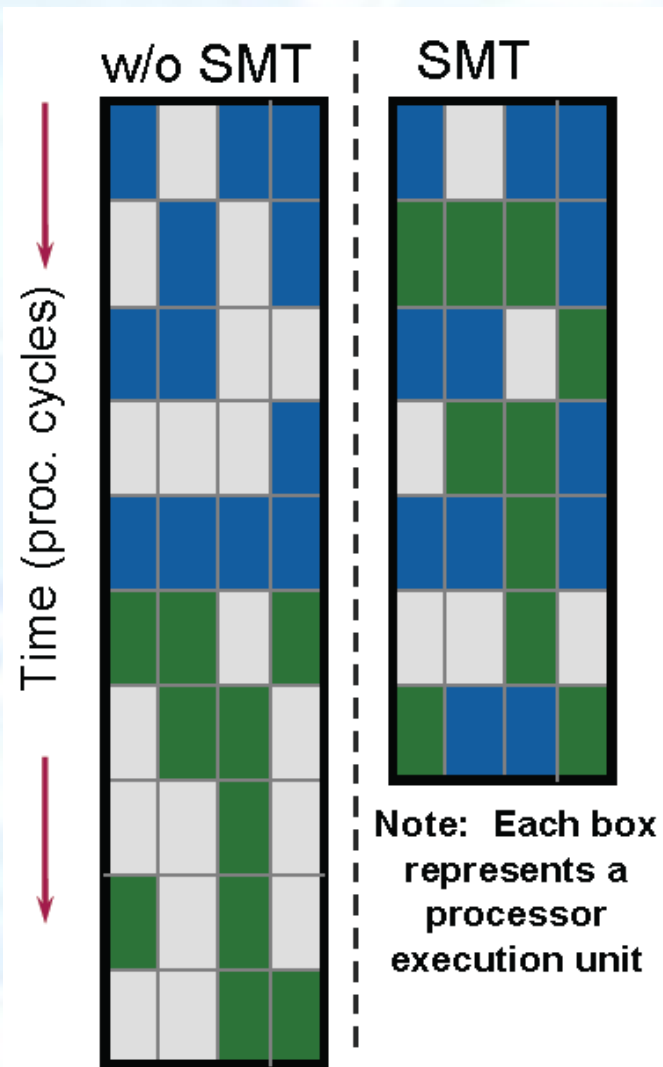
# Функциональные устройства

## Execution Unit Overview

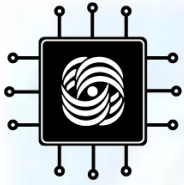




# HyperThreading



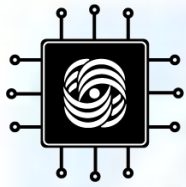




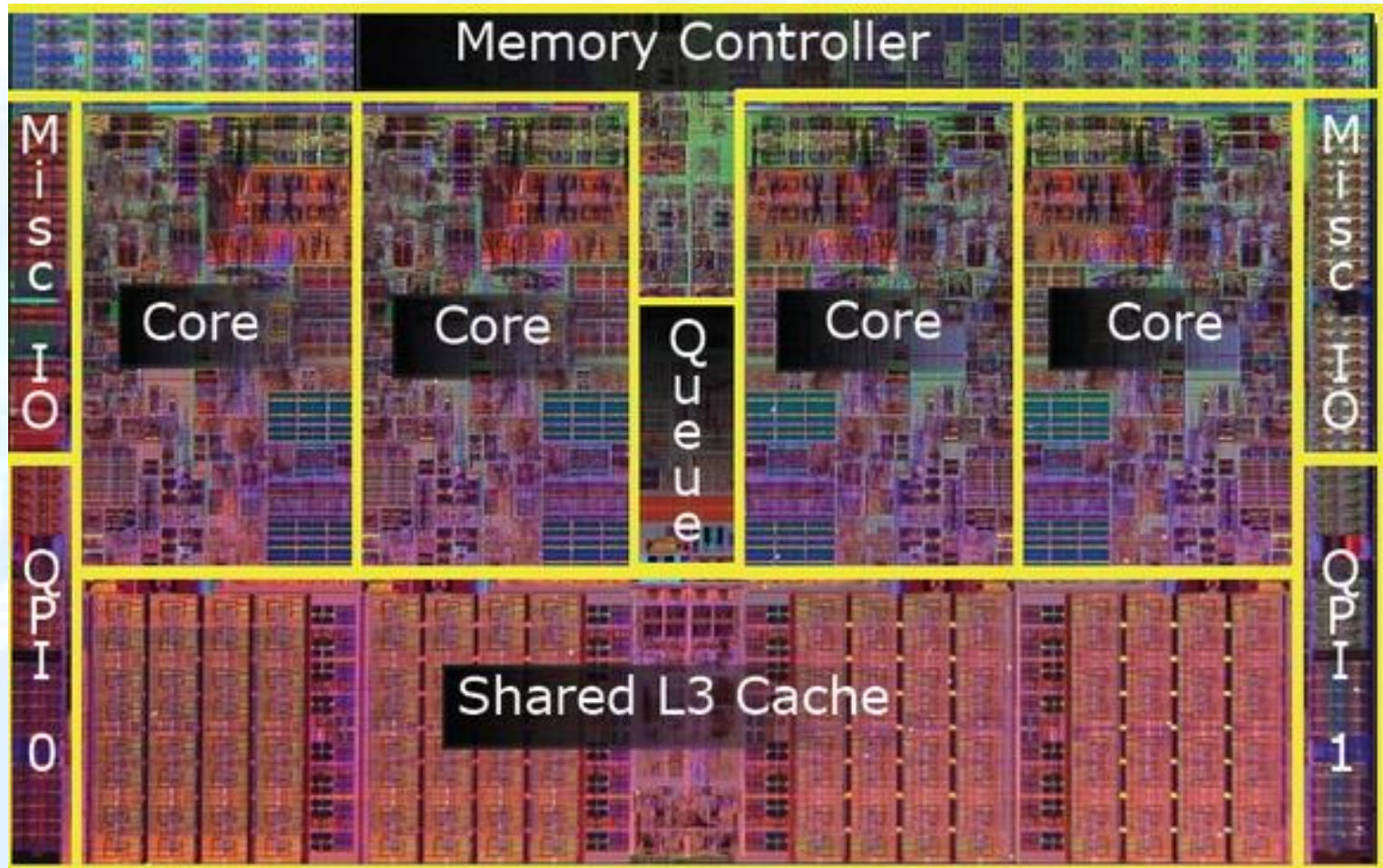
# Иерархия памяти

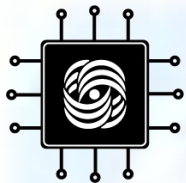






# Микросхема





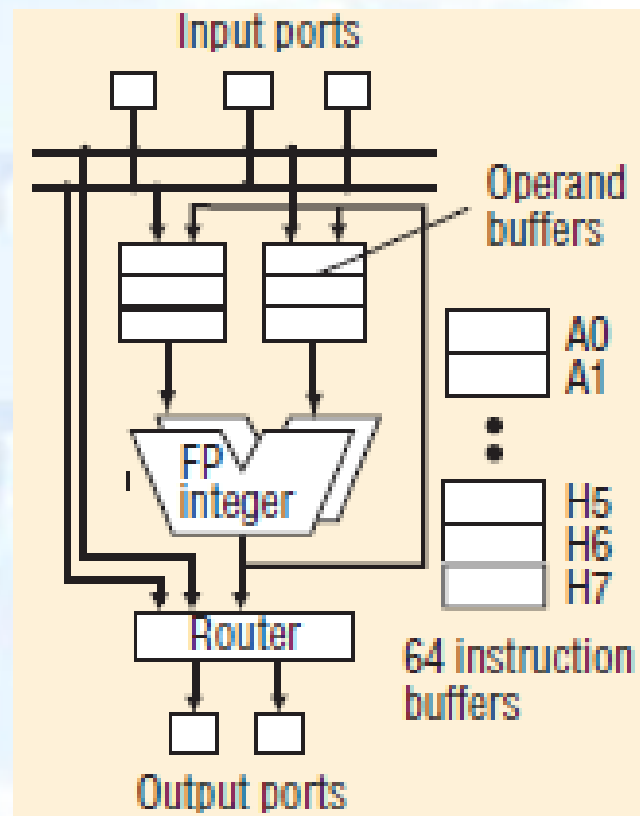
# Потоковые машины

EDGE – Explicit Data Graph Execution

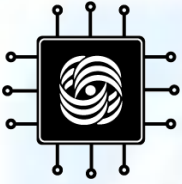
- Основная идея - прямая передача операндов, без записи в регистры;
- Операция выполняется сразу, как доступны все её операнды.

• “Недостатки” современных RISC процессоров

- Много неэффективных и ресурсоемких структур, направленных на построение графа зависимостей по данным;
- Много ресурсоемких операций доступа к регистрам.



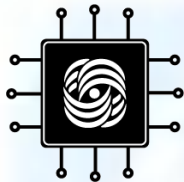
*RISC:*    **ADD** R1, R2, R3  
*EDGE:*    **ADD** T1, T2



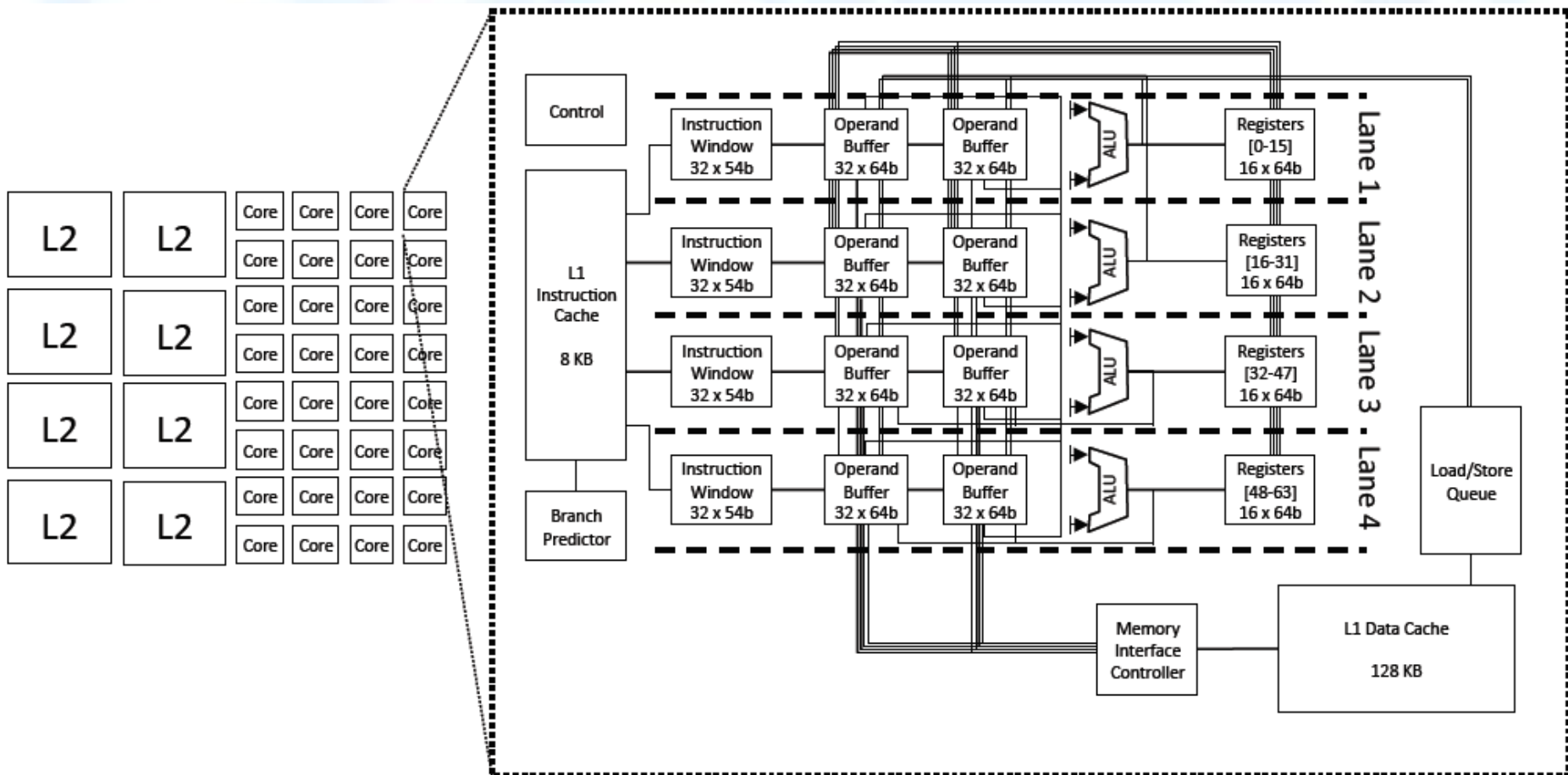
# E2

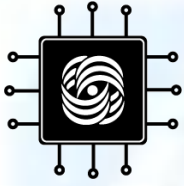
## Особенности:

- Программа разделяется на блоки инструкций (как поток управления), но внутри инструкции исполняются в dataflow режиме;
- Два режима = скалярный + векторный;
- Объединение ядер – физические ядра могут объединяться в логические;
- Спекулятивное исполнение.
  - между блоками;
  - внутри блока (предикатное выполнение).



# Схема E2





**Спасибо за внимание!**