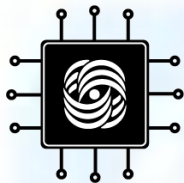


АРХИТЕКТУРА СЕТЕВЫХ УСТРОЙСТВ

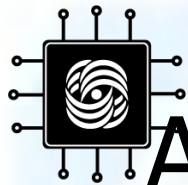
Лекция 01: Сетевые устройства

ВМК МГУ им. М.В. Ломоносова, Кафедра АСВК
Доцент, к.ф.-м.н. Волканов Д.Ю.



План лекции

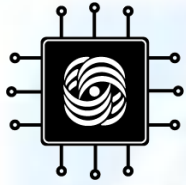
- Развитие компьютерных сетей
- Общая схема коммутатора
- Основные функции сетевого процессора
- Обзор существующих сетевых процессоров



Содержание курса

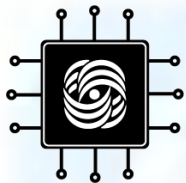
Архитектура сетевых устройств

- Общая схема коммутатора.
- Основные функции сетевого процессора.
- Жизненный цикл пакета в СПУ
- Подходы к разбору заголовка пакета
- Таблицы классификации
- Постановка задачи классификации пакетов
- Ассоциативная память
- Хэш-функции
- Язык P4, NPL
- Память типа TCAM
- Обзор существующих сетевых процессоров



Практические задания

- Написание сценария обработки пакета на ассемблере СПУ
- Исследование структуры данных для СПУ
- Программа обработки заголовка на P4
- *Задание на моделирование буферов устройств*



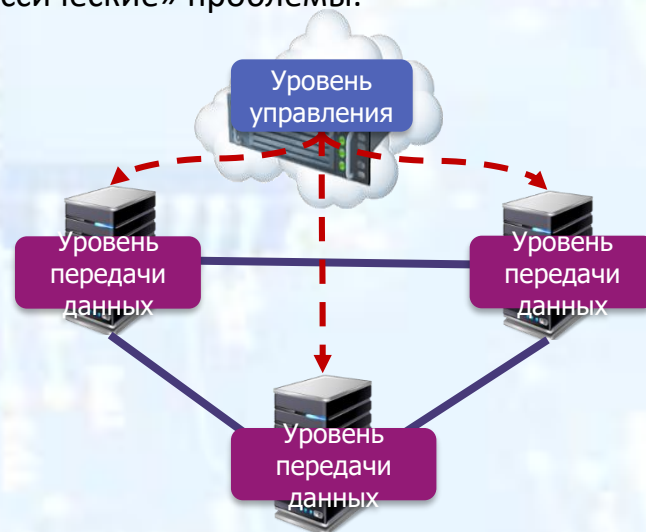
Архитектура сети

Классический подход vs ПКС

Классический подход подразумевает последовательную передачу данных через ряд «узлов», каждый из которых **повторяет идентичные сложные вычисления**, выполняя и управление сетью (Control Plane), и передачу данных (Data Plane).



ПКС – вынос задач управления в единый центр: разделение уровней управления сетью и передачи данных, что позволяет решить «классические» проблемы.

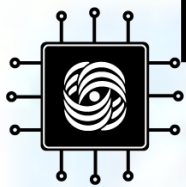


Недостатки:

- Значительное увеличение CAPEX и OPEX.
- Снижение эффективности использования ресурсов сети.
- Высокие требования к количеству и квалификации персонала.
- Зависимость от конкретных производителей.

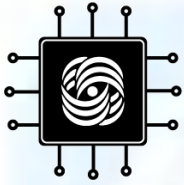
Преимущества ПКС:

- ✓ Сокращение CAPEX/OPEX
- ✓ Централизация управления
- ✓ Ускорение вывода на рынок новых сервисов
- ✓ Использование стандартного оборудования
- ✓ Повышение эффективности использования каналов связи



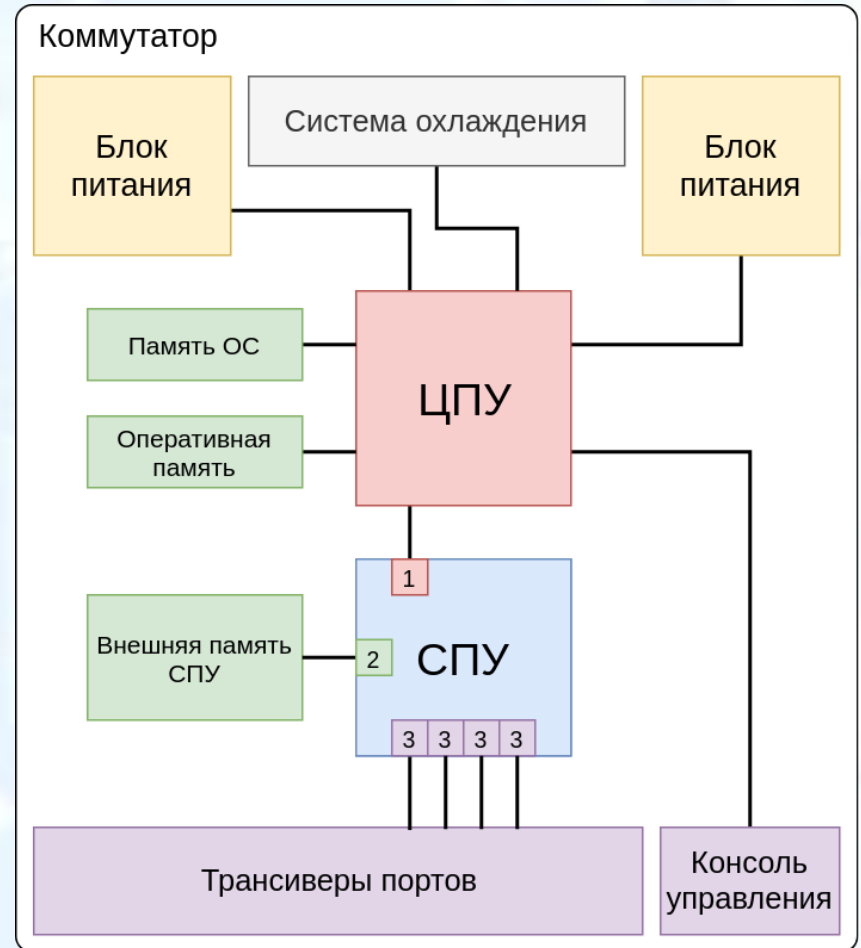
Ключевые особенности технологии ПКС

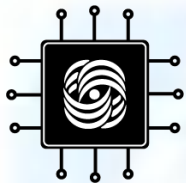
- Изоляция контура управления от контура передачи данных
- Унифицированный интерфейс для приложений управления
- Унифицированный интерфейс для контура передачи данных
- Централизация управления
 - понятие состояния сети
 - резкое сокращение времени сходимости
 - топология на L2 и L3



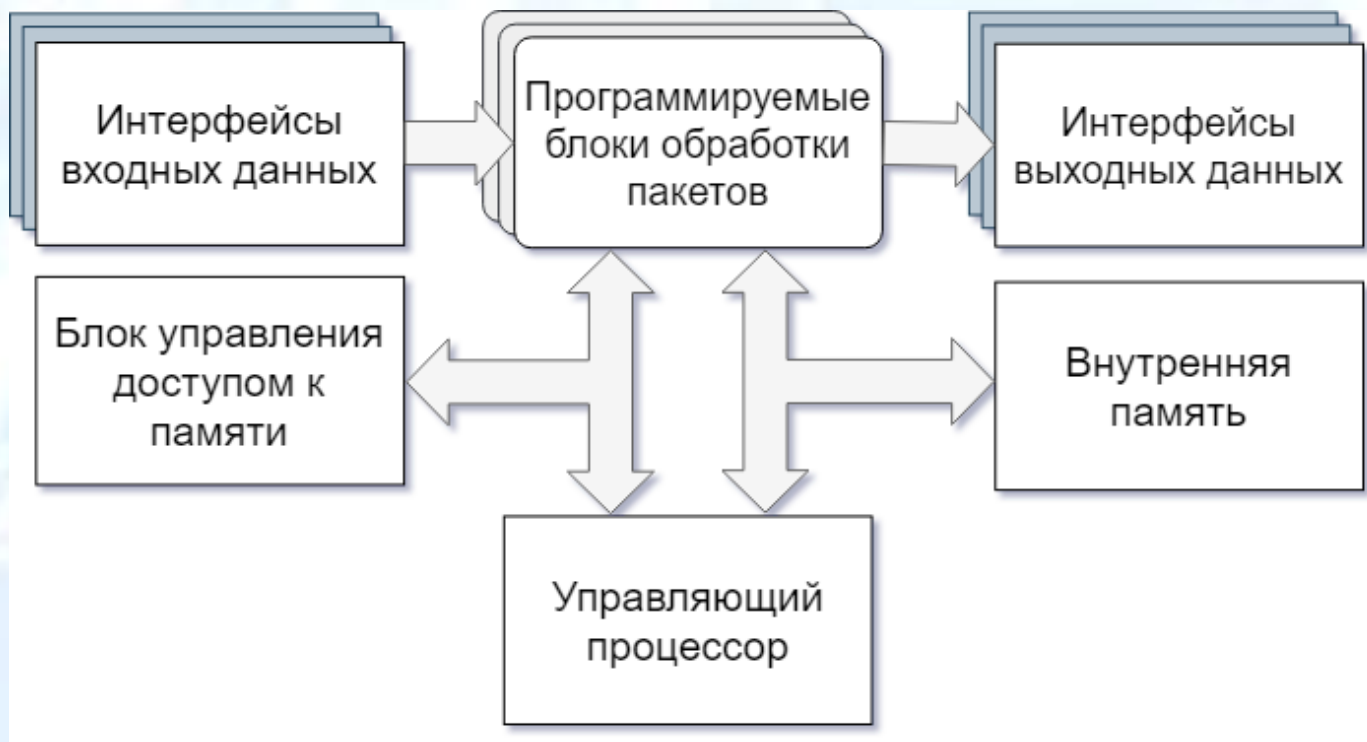
Место СПУ в коммутаторе

- **Сетевое процессорное устройство (СПУ)** – встроенная полупроводниковая система, оптимизированная для выполнения операций передачи данных
- **Функции СПУ:**
 - получение пакета;
 - выделение заголовка из пакета;
 - классификация пакета;
 - модификация заголовка и принятие решения о пути следования пакета;
 - управление очередями;
 - передача пакета.





Обобщенная архитектура СПУ



14 байт

20 байт

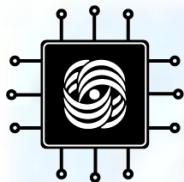
80 байт

Ethernet

IP заголовок

TCP заголовок

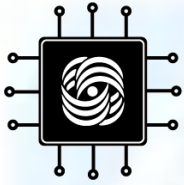
Полезная нагрузка



Рассматриваемые СПУ

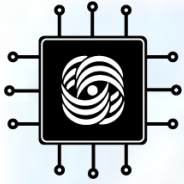
- Barefoot Tofino
- Barefoot Tofino 2
- Mellanox NP-5
- Mellanox SwitchX-2
- Huawei ENP
- Innovium Teralynx 7
- Nokia FP4
- Cisco NPU
- Juniper Q5
- Broadcom Tomahawk 3
- Broadcom Trident 3





Критерии обзора СПУ

- Год выпуска
- **Программируемость СПУ**
- **Тип СПУ**
- **Ключевые особенности архитектуры (6 критериев)**
- Характеристики кристалла (3 критерия)
- Тип интерфейса к ЦПУ
- Управляющий процессор на кристалле (если предусмотрен)
- Производительность
- Допустимые конфигурации сетевых интерфейсов
- Стоимость



Программируемость СПУ

- **Устройства с фиксированной функциональностью**
 - Фиксированный стек протоколов и программа обработки пакетов
- **Конфигурируемые устройства**
 - Загрузка программы обработки пакетов в рамках predetermined протоколов передачи данных
- **Программируемые устройства**
 - Определение новых протоколов передачи данных в загружаемой программе

Broadcom
Tomahawk

Barefoot Tofino,
Broadcom Trident,
Mellanox NP-5,
Cisco NPU...



Подходы к построению коммутаторов

Коммутатор на ядрах общего назначения

Достоинства

- Гибкость настройки и модификации функциональности
- Простота внесения изменений

Недостатки

- Плохое соотношение стоимость / производительность
- При скорости выше 10 Гб/сек потеря пакетов 5-6 %
- Высокое энергопотребление

Схемы специального назначения

Достоинства:

- Наилучшее соотношение стоимость/производительность
- Возможность достижения высокой скорости обработки данных
- Низкое энергопотребление

Недостатки:

- Высокая сложность программирования сервисов
- Необходимость наличия глубокой экспертизы в разработке сетевых устройств
- Необходимость полной переделки при переходе на новый стек протоколов

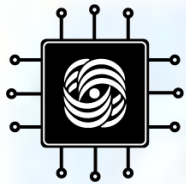
Сетевые процессоры (NPU)

Достоинства:

- Гибкость в программировании новых сервисов
- Промежуточное положение по соотношению стоимость / производительность
- Низкое энергопотребление
- Возможность быстрого развития линейки устройств
- Длительное время нахождения на рынке
- Соответствие имеющемуся опыту разработки в РФ

Недостатки:

- Длительный цикл разработки

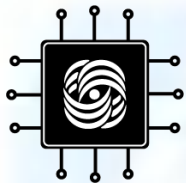


Типы программируемых СПУ

- **Многопроцессорная ИС на базе процессоров общего назначения**
 - Гибкость программирования
 - Невысокая скорость обработки пакетов
- **ASIC**
 - Аппаратная реализация основных функций СПУ (низкая гибкость программирования)
 - Высокая скорость обработки пакетов
- **Сетевой процессор**
 - Специализация к задачам обработки пакетов
 - Компромисс по возможностям программирования

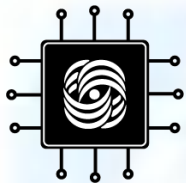
Cisco NPU

Barefoot Tofino,
Innovium
Teralynx,
Broadcom
Trident,
Broadcom
Tomahawk
Mellanox NP-5,
Huawei ENP,
Juniper Q5, ...



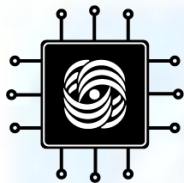
Сравнение СПУ по общим критериям

СПУ	Программируемость	Производительность	Интерфейсы	Стоимость	Стоимость коммутатора
Mellanox NP-5	+, Си	240 Гбит/с	До 100 GbE	1000\$?
Mellanox SwitchX-2	+, Си	До 2 Тбит/с	До 56 GbE	Нет данных	15000\$
Huawei ENP	+	480 Гбит/с	До 100 GbE	Нет в продаже	6000\$
Nokia FP4	+	2,4 Тбит/с	До 400 GbE	Нет данных	Нет данных
Barefoot Tofino	+, P4	6,5 Тбит/с	До 100 GbE	Нет данных	8000\$



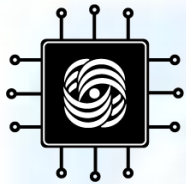
Сравнение характеристик кристалла СПУ

СПУ	Тех. процесс	Интерфейсы с ЦПУ
Mellanox NP-5	28 нм	PCI Express, Ethernet 1×10GbE
Mellanox SwitchX-2	16 нм	PCI Express Gen3
Huawei ENP	16 нм	Нет данных
Nokia FP4	16 нм	Нет данных
Barefoot Tofino	16 нм	4×PCI Express Gen3, 1 или более Ethernet до 100 GbE



Сравнение ключевых особенностей архитектур конвейеров СПУ

СПУ	Состав конвейера	Типы ядер СПУ
Mellanox NP-5	Конвейер из 5 функционально специализированных стадий	Специализированные векторные процессоры
Mellanox SwitchX-2	Нет данных	Нет данных
Huawei ENP	Явной структуры конвейера нет, процессоры объединены в группы, которые могут параллельно выполнять разные задачи	Процессоры со специализированными инструкциями для обработки заголовков Ethernet, IP
Nokia FP4	Нет данных	Нет данных
Barefoot Tofino	Конвейеры входной и выходной обработки, разделенные коммутационной матрицей. Функционально специализированные стадии трех типов	Специализированные процессоры для каждого из типов стадий



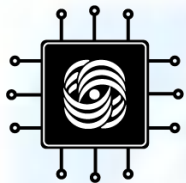
Организация конвейера

Два основных подхода:

- процессорные ядра общего назначения внутри стадий Cisco NPU
- специализация ядер к функциям обработки пакетов Barefoot Tofino,
Mellanox NP-5,
Huawei ENP, ...

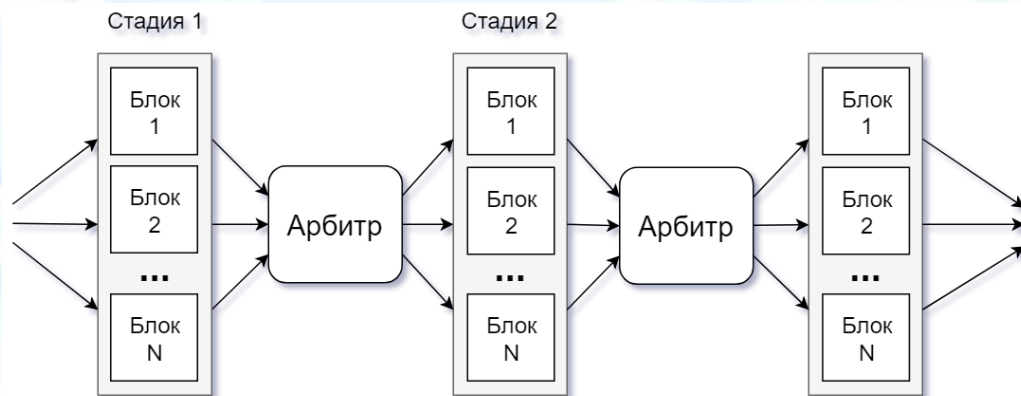
Механизм “разворота” пакетов:

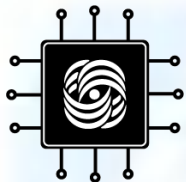
- повторный проход пакета по конвейеру Barefoot Tofino,
Mellanox NP-5,
Juniper Q5
- понижает пропускную способность конвейера



Параллелизм СПУ

- Параллелизм на уровне стадий конвейера
- Параллелизм конвейеров
- Комбинированные подходы



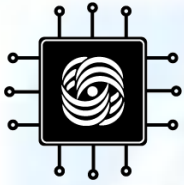


Память СПУ

- Стратегии размещения данных:**
- тела пакетов – внешняя память, Broadcom Trident, Broadcom Tomahawk, Barefoot Tofino, Mellanox NP-5
 - таблицы классификации – внутренняя память
 - все данные во внешней памяти Huawei ENP, Juniper Q5

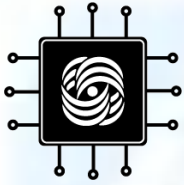
Память тел пакетов: DDR SDRAM, RL DRAM

Память таблиц классификации: SRAM, TCAM



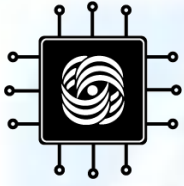
Основные тенденции

- Программируемость разные производители понимают по-разному
- Наибольшая производительность у устройств ASIC
- Принципы построения конвейеров:
 - разделение на 2 части (ingress, egress);
 - коммутационная матрица и репликатор пакетов между частями конвейера;
 - функциональная специализация стадий;
 - масштабируемая архитектура из однотипных конвейеров.
- Размещение тел пакетов во внешней памяти, таблиц классификации – во внутренней

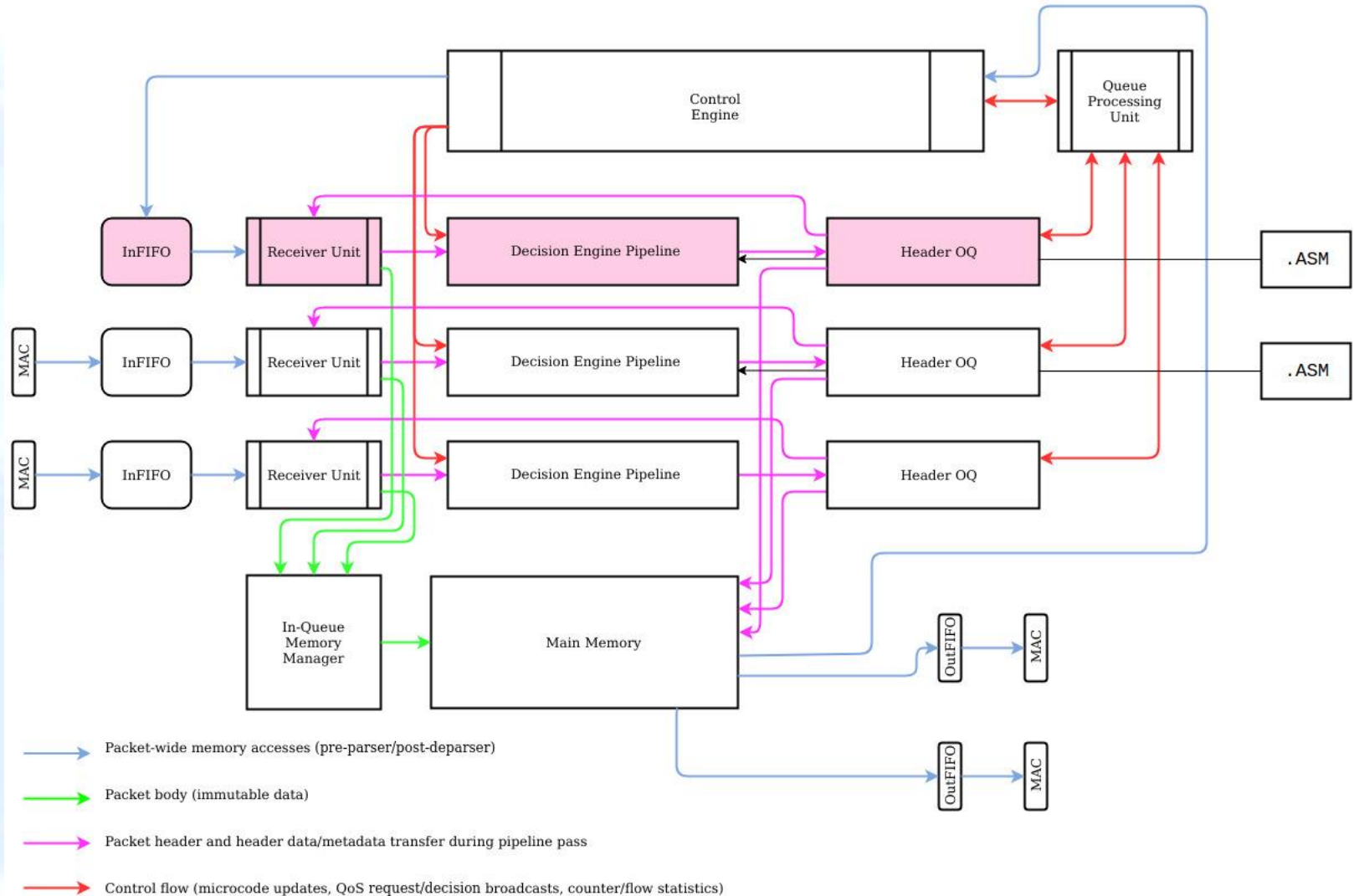


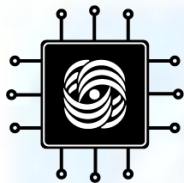
Выводы

- Разработка СПУ типа сетевой процессор
- Масштабируемая архитектура (набор однотипных конвейеров или конвейер стадий с однотипными ядрами)
- Функциональная специализация стадий конвейера
- Использование блоков памяти на кристалле для хранения таблиц классификации



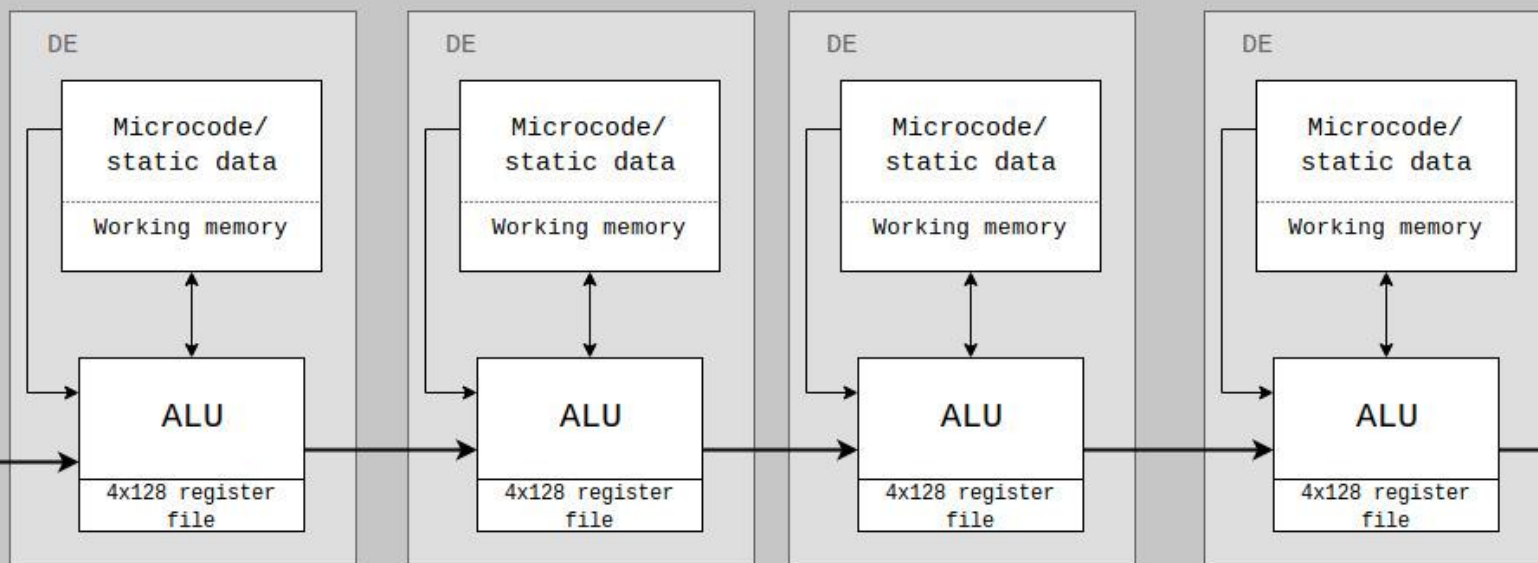
Архитектура со специализированными ядрами

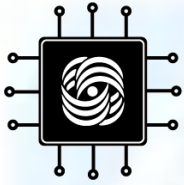




Архитектура со специализированными ядрами

Decision Engine Pipeline

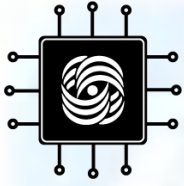




Характеристики предлагаемой архитектуры

- 24 порта по 10 Гбит/с (*+ 4 порта по 100 Гбит/с*)
- Интерфейс PCIe Gen2 x8
- 24*8 специализированных 64-бит потоковых процессоров
- Рабочая частота: 800 МГц
- Общий объем ОЗУ на кристалле: 64 МБайт
- Поддержка таблиц до 32 тыс. элементов.*
- Размер кристалла до 160 мм².*
- Тех. процесс TSMC 28 нм HPC+.*
- Энергопотребление до 45 Вт.*

** - характеристики будут уточнены при проектировании*



Спасибо за внимание!