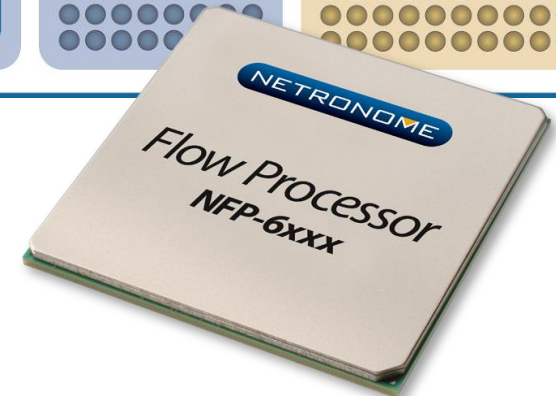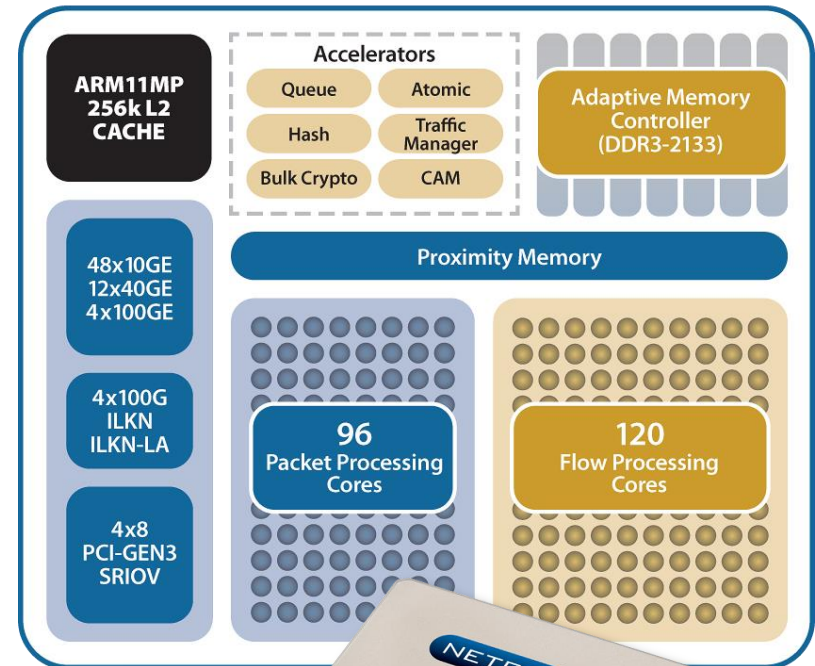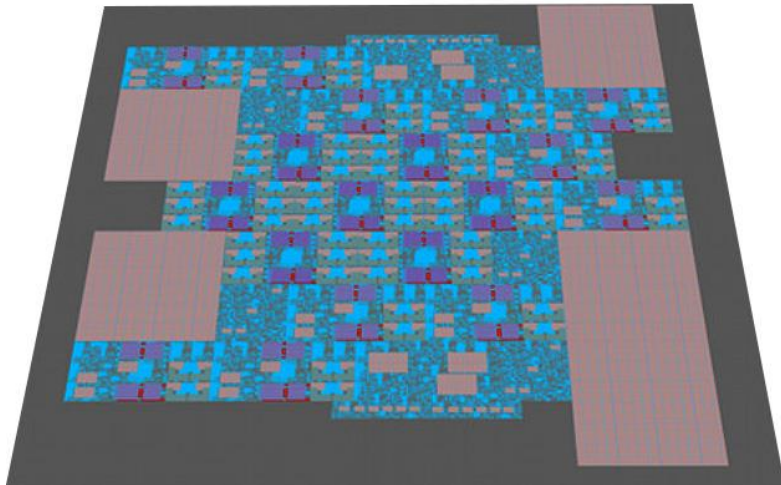# NFP-6xxx – A 22nm High-Performance Network Flow Processor for 200Gb/s Software Defined Networking

Gavin Stark (CTO), Sakir Sezer (Fellow)

Agenda

- Software Defined Networking (SDN)

- Design challenges: Flexibility vs. Performance

- NFP-6xxx Architecture

- Data plane acceleration

ARM11MP 256k L2 CACHE

Accelerators
- Queue
- Atomic
- Hash
- Traffic Manager
- Bulk Crypto
- CAM

Adaptive Memory Controller (DDR3-2133)

48x10GE 12x40GE 4x100GE

4x100G ILKN ILKN-LA

4x8 PCI-GEN3 SRIOV

Proximity Memory

96 Packet Processing Cores

120 Flow Processing Cores

NETRONOME
Flow Processor
NFP-6xxx

NETRONOME

Software Defined Networking

- A new term for defining the evolution of IP networks

- A paradigm for converging networking and cloud computing

- Virtualization of networks and the ability to provide "X"-as-a-Service (XaaS).

- Open Network Foundation (ONF) Definition:
  "In the SDN architecture, the control and data planes are decoupled, network intelligence and state are logically centralized, and the underlying network infrastructure is abstracted from the applications."

SDN Architecture
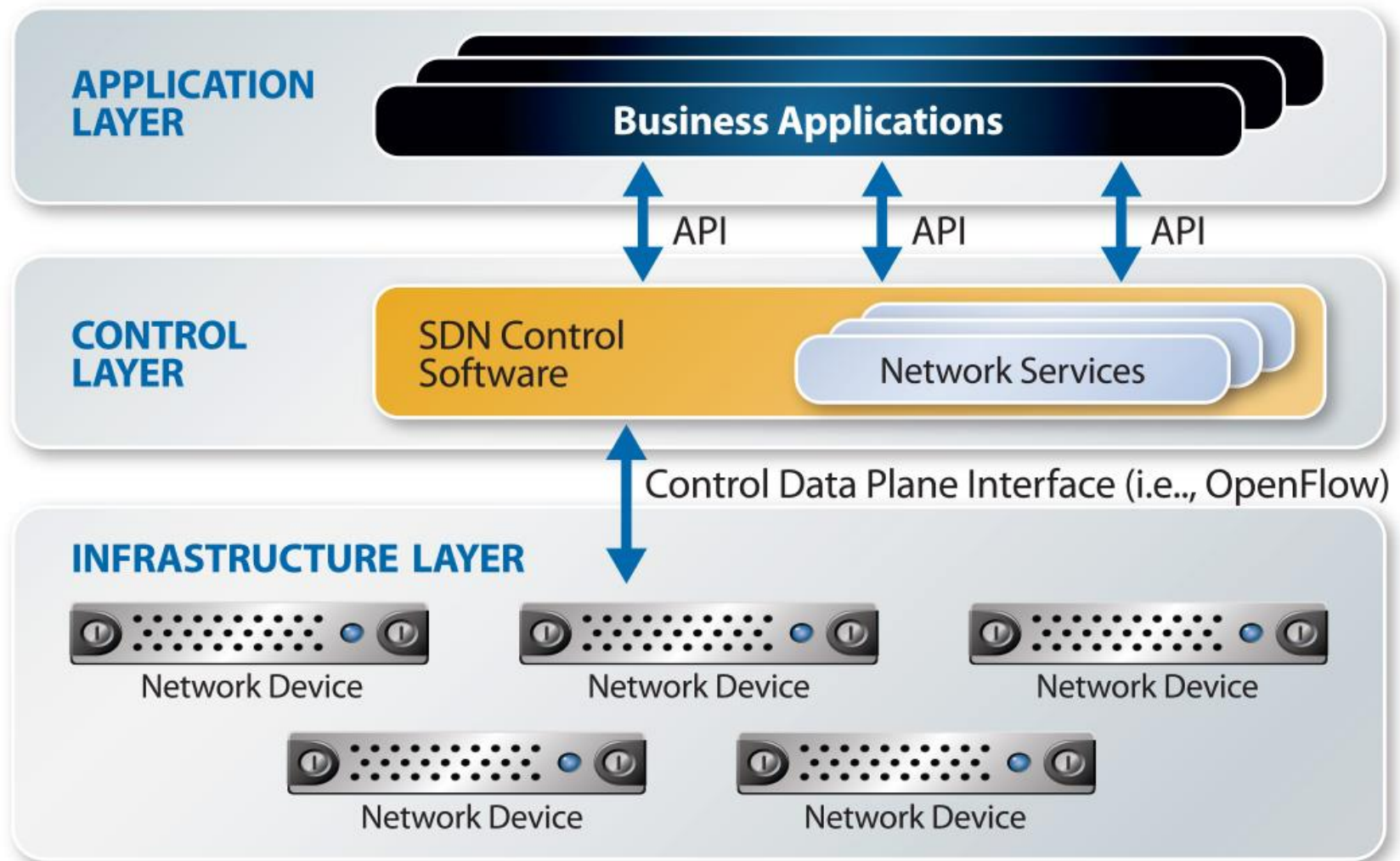
Source: Open Network Foundation

Key SDN Challenges

- ## Performance vs Flexibility
  - Network nodes *"without control plane"* must provide effective configuration of the data-plane without compromising performance and latency (more than a bare-metal switch)
  - Requirements vary based on network architecture and device location

- ## Scalability
  - SDN controllers must meet the scalability challenges of vast heterogeneous networks in a multi-tenant environment, while providing a global network view
  - Infrastructure needs to scale from 10 to 100Gbps

- ## Security
  - Security must be an integral part of the SDN architecture rather than an add-on
  - Networks will be exposed to new types of network intrusion, hijacking and DDoS attacks

- ## Interoperability
  - The transition to SDN requires the coexistence and interoperability of both networks for a very long time, with challenges in management, control and migration towards SDN

NETRONOME

Performance and Technological Challenges

- SDN data-plane must facilitate complex packet and flow processing capability, programmable via OpenFlow.
    - ⇒ Complex flow lookup, highly threaded packet processing, application specific flow processing, and effective resources management.

- Service and application oriented nature of SDN requires intelligence at the network edge and advanced network interface.
    - ⇒ Integration of SDN edge within the data-center fabric with the capability to extend application into the network.
    - ⇒ Network and switch virtualization and intelligent L2-L7 data-path decisions
    - ⇒ SDN security, NGFW and high-throughput low-latency tunneling of flows

SDN requires uniquely architected silicon solutions

- ⇒ **Highly-threaded heterogeneous processor architectures**
- ⇒ **Memory and I/O centric SoC optimized for flow and session processing**
- ⇒ **Deterministic on-chip interconnect eliminating coherency and data duplication**

Flow Processor Roadmap

First silicon Q4 2013

Intel IXP 28xx + FPGA	Netronome NFP-32xx	Netronome NFP-6xxx

- Intel 130nm
- 10 Gbps Flow Processor
- 16 microengine cores
- 120/30 Gbps memory hierarchy bandwidth
- 1.4GHz

- TSMC 65nm
- 40 Gbps Flow Processor
- 40 microengine cores
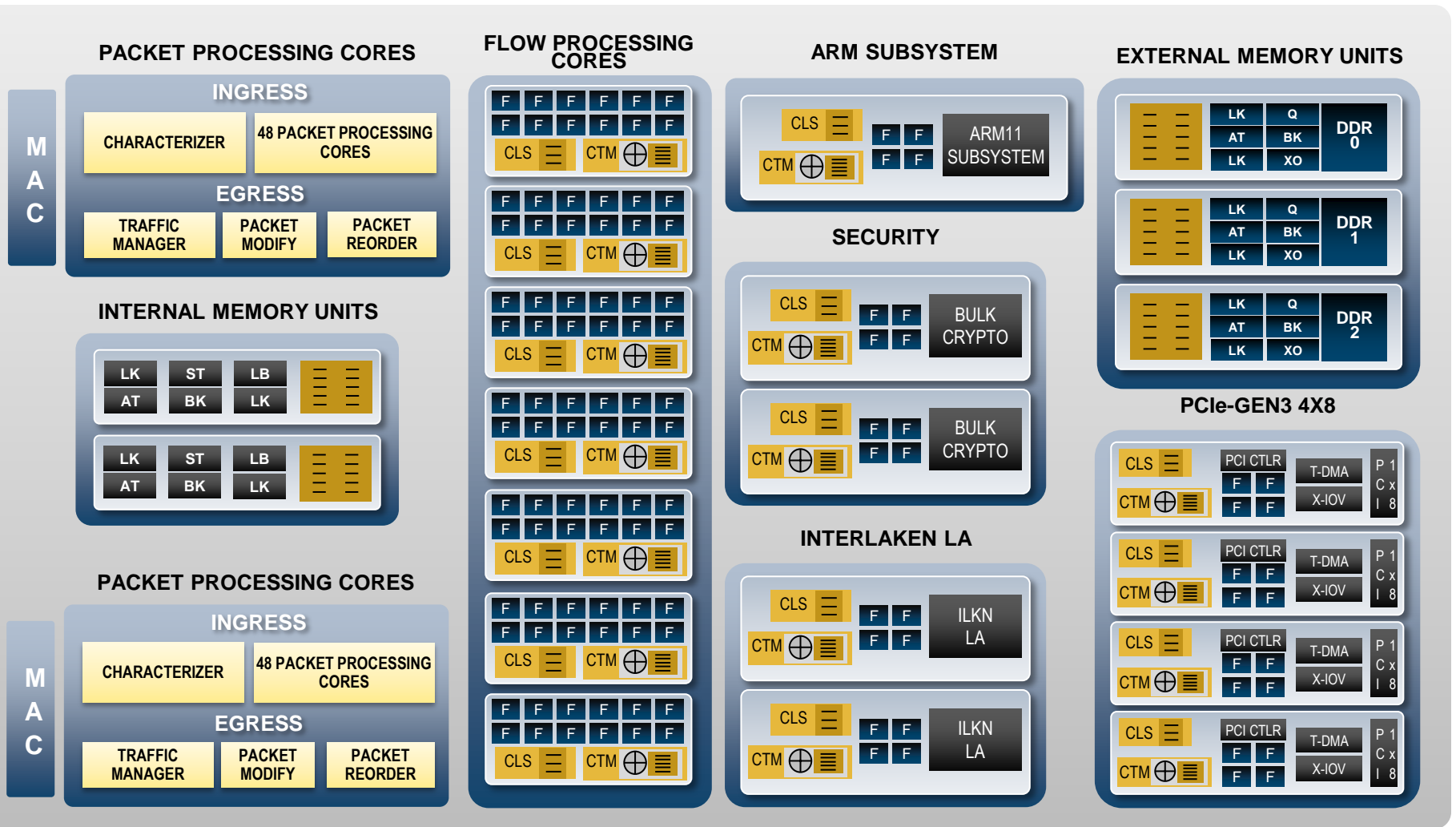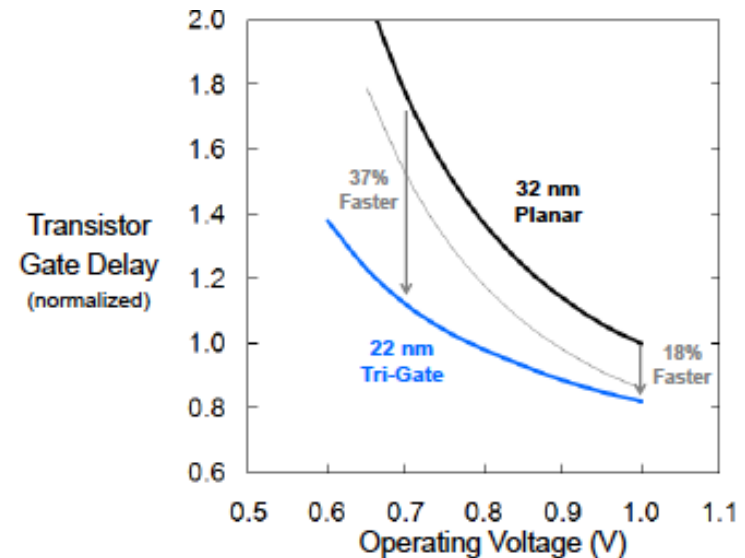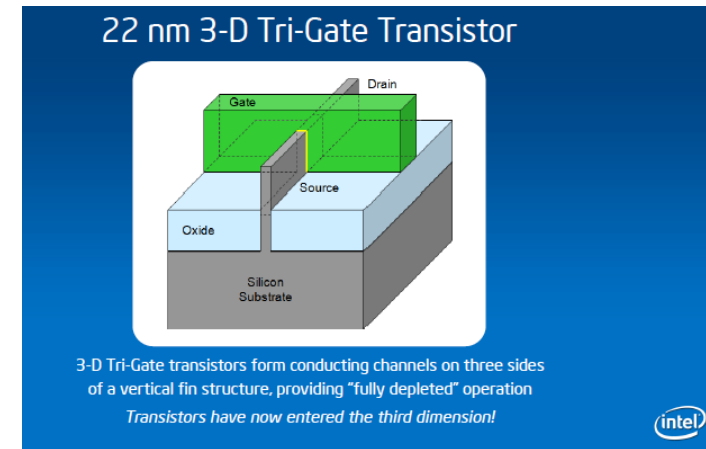- 180/220 Gbps memory hierarchy bandwidth
- 1.4GHz
- 325M Transistors

- Intel 22nm
- 200 Gbps Flow Processor
- 120 flow processing cores (FPC)
- 96 packet processor cores (PPC)
- 400/7400 Gbps memory hierarchy bandwidth
- 1.2GHz
- >3B transistors

NETRONOME

NFP-6xxx Detailed Block Diagram

PACKET PROCESSING CORES

INGRESS
CHARACTERIZER | 48 PACKET PROCESSING CORES
EGRESS
TRAFFIC MANAGER | PACKET MODIFY | PACKET REORDER

INTERNAL MEMORY UNITS

LK ST LB
AT BK LK

LK ST LB
AT BK LK

PACKET PROCESSING CORES

INGRESS
CHARACTERIZER | 48 PACKET PROCESSING CORES
EGRESS
TRAFFIC MANAGER | PACKET MODIFY | PACKET REORDER

M A C

FLOW PROCESSING CORES

CLS CTM

ARM SUBSYSTEM

CLS CTM F F F F ARM11 SUBSYSTEM

SECURITY

CLS CTM F F F F BULK CRYPTO

CLS CTM F F F F BULK CRYPTO

INTERLAKEN LA

CLS CTM F F ILKN LA

CLS CTM F F ILKN LA

EXTERNAL MEMORY UNITS

LK Q DDR 0
AT BK
LK XO

LK Q DDR 1
AT BK
LK XO

LK Q DDR 2
AT BK
LK XO

PCIe-GEN3 4X8

CLS CTM PCI CTLR F F T-DMA X-IOV P C I 1 x 8

CLS CTM PCI CTLR F F T-DMA X-IOV P C I 1 x 8

CLS CTM PCI CTLR F F T-DMA X-IOV P C I 1 x 8

CLS CTM PCI CTLR F F T-DMA X-IOV P C I 1 x 8

NETRONOME

Technology - Intel 22nm SOC

- Intel's 3-D Tri-Gate transistor manufactured at 22nm
  - 37% performance increase at low voltage (0.7V)
  - 50% power reduction at constant performance (versus 32nm)
  - Low leakage (SOC process)

- >3B transistors, 1.2GHz

- 35W-70W, 200mW/Gbps, 300nW/pkt

- Standard methodology optimized for multi-vendor tool-chain

22 nm 3-D Tri-Gate Transistor

3-D Tri-Gate transistors form conducting channels on three sides of a vertical fin structure, providing "fully depleted" operation
Transistors have now entered the third dimension!

Transistor Gate Delay (normalized) vs Operating Voltage (V)

NETRONOME

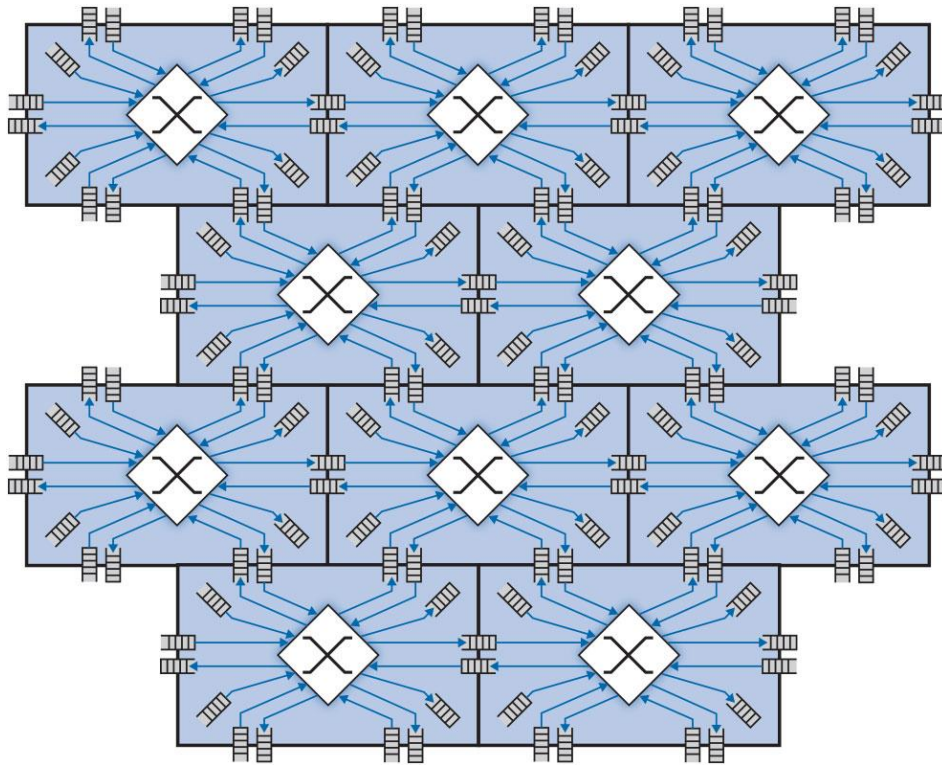# Innovative Heterogeneous Island Architecture

- Identical overlay mesh for all islands spreads traffic
- Regular bus structures, JTAG, scan, power, clocks...
- Many varieties of island contents, reused as GDSII
- Requires latency-tolerant processing architecture
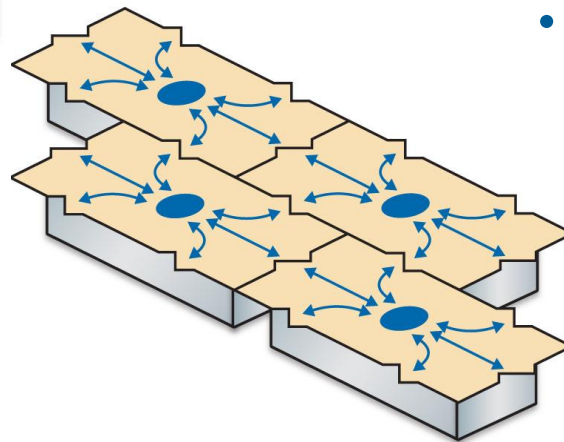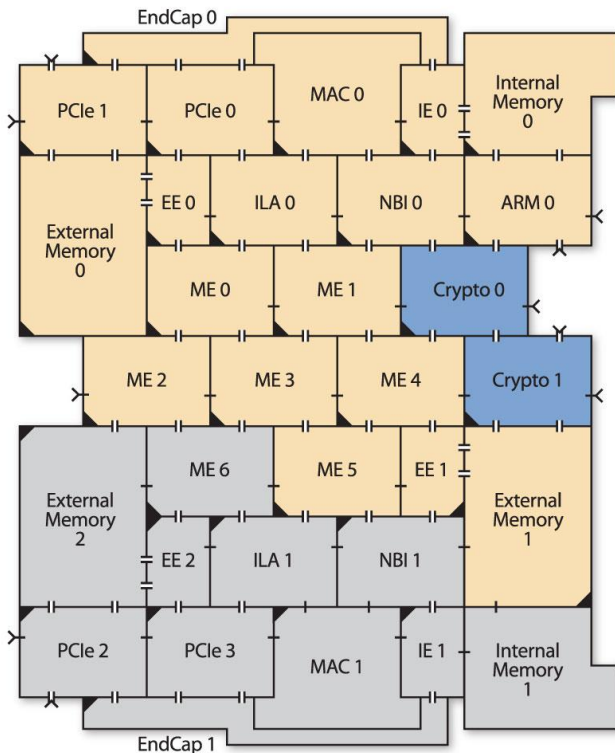
- Single storage location for any one piece of data
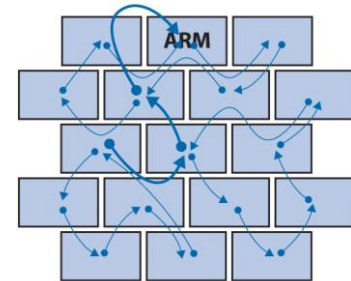 - No coherence traffic
  - No data replication

Works optimally with memory-centric distributed processing

Perfect for heterogeneous flow processing

Distributed Bus Structures

- Event chains and rings
  - System alerts / events / interrupts
  - Localized events from accelerators to FPCs

- Global Control
  - System-wide control/status/configuration/debug
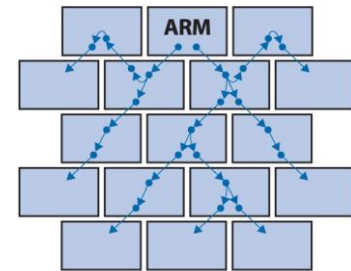 - Lower bandwidth, independent of global data bus

- Global Data - Distributed Switch Fabric
  - Hexagonal crossbar routing
  - Island-centered switch fabric elements
  - 768Gbps bandwidth across each island
 - Configurable routing permits application-driven traffic balancing

- Testability, JTAG, clock distribution, observability
  - System-wide control/status/configuration/debug
  - Lower bandwidth, independent of global data bus

NETRONOME

Latency Tolerant Processing

- Flow processing cores (FPC) optimized for memory-processing latency
  - Multiple cooperating contexts, large register files partitioned by context
  - Multiple posted memory/accelerator commands
  - Copes with 50-500 cycles of memory latency
  - Perfect fit for memory-centric processing

Example single FPC with eight contexts
Posted memory transactions for just one context shown
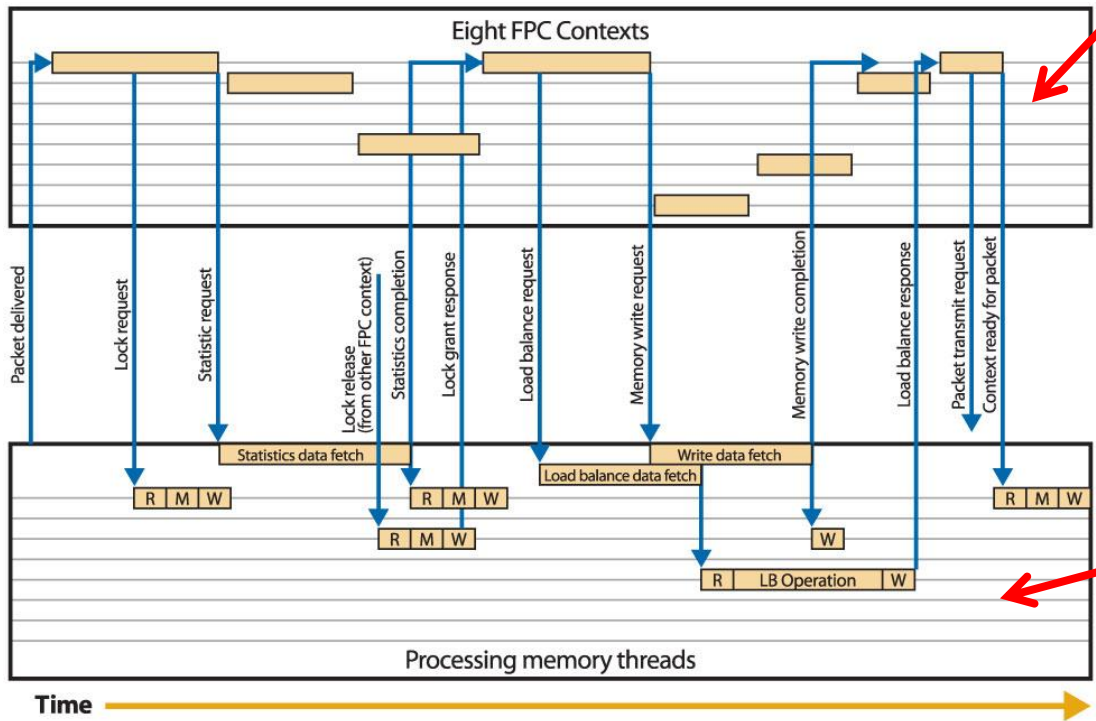
Multithreaded Memory Processing

- Multithreaded data fetch
- Multiple processing pipelines per unit
- Multiple distributed units across a device
- Perfect for multibanked SRAMs

Example single processing memory with 8 processing threads
Exclusive access to memory handled inside processing memory for atomic memory transactions, locks, statistics

NETRONOME

PPC and FPC Pipelines

PPCs (96 per device)

- Single threaded, lightweight
- In-order, round-robin operation
- Fed with hardware characterization and packet data
- Software-programmed classification result
- No inter-packet state
- Access to hierarchy of lookup memories

FPCs (120 per device)

- Derived from and source code compatible with Intel IXP Microengines
- Eight cooperative contexts (2 cycle switch)
- Pool-of-threads, run-to-completion software model
- Integrated with memory-processing hierarchy
- Organized in clusters (16) each with 4kB local scratch / 256kB local memory
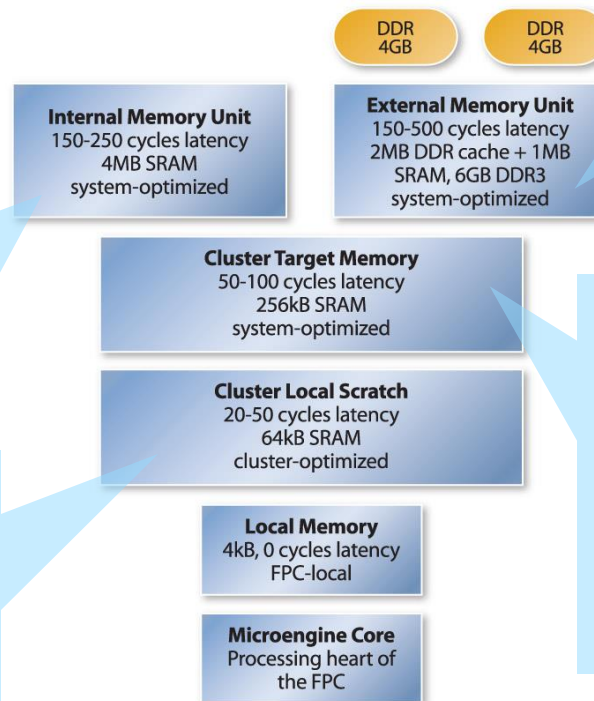
Hierarchical Processing Memories

Philosophy: Processing in the optimal location

Process data where the data resides

Three external DDR memory units (MU) per device

- Locks, hash tables, microqueues
- Linked lists, rings
- Recursive lookups
- >300 different processing operations
- >200 threads per unit

Two internal memory units (MU) per device

- Locks, hash tables, microqueues
- Recursive lookups
- Statistics, load balancing
- >300 different processing operations
- >200 threads per unit

Cluster Local Scratch (17 per device)

- Locks, hash tables, microqueues
- Rings, stacks
- Regular expression NFA
- >100 different processing operations

- 17 per device
- Locks, hash tables, microqueues
- Packet buffering, delivery, transmit offload
- Rings
- >250 different processing operations
- >100 threads per unit

DDR 4GB

DDR 4GB

Internal Memory Unit
150-250 cycles latency
4MB SRAM
system-optimized

External Memory Unit
150-500 cycles latency
2MB DDR cache + 1MB
SRAM, 6GB DDR3
system-optimized

Cluster Target Memory
50-100 cycles latency
256kB SRAM
system-optimized

Cluster Local Scratch
20-50 cycles latency
64kB SRAM
cluster-optimized

Local Memory
4kB, 0 cycles latency
FPC-local

Microengine Core
Processing heart of
the FPC

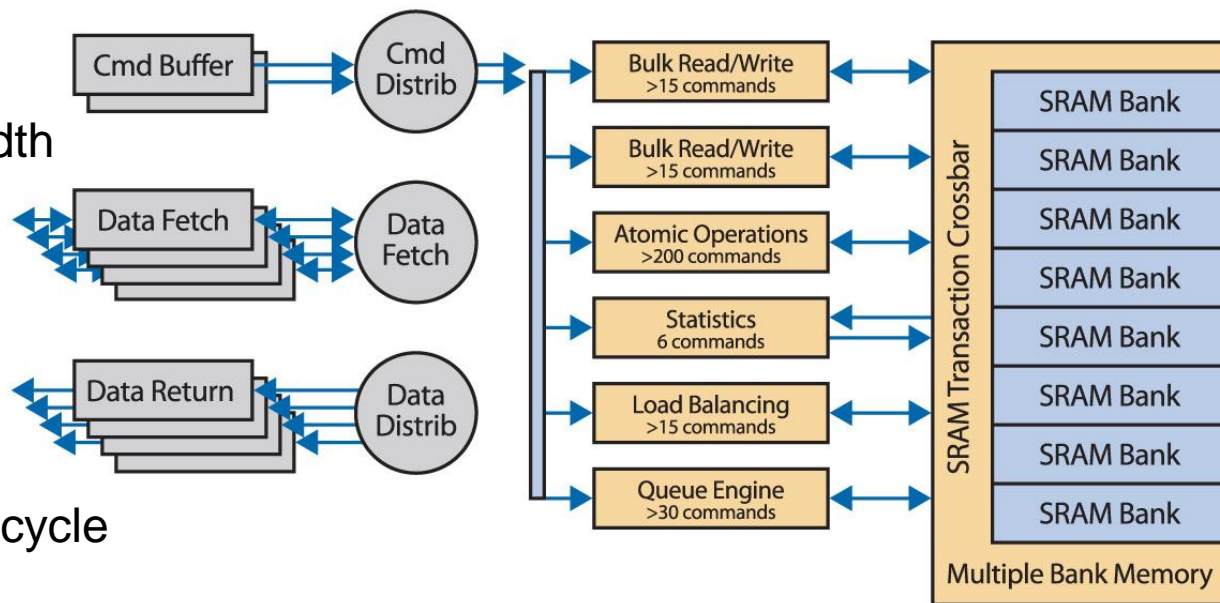Includes 32MB of memory

NETRONOME

Multi-Pipelined Processing Memory

Switch Fabric Interface

- 2 billion commands per second
- 500Gbps data bandwidth

Multi-Bank SRAM

- Eight crossbar inputs
- 1Tbps bandwidth
- Eight transactions per cycle

Multiple Processing Engines

- No locking between engines
- Different engines in different processing memories in the device
- Different engines support different processing operations
- Highly threaded to maintain 100% throughput when required

Highly Threaded Memory Engines

- Example memory processing engine types
  - Atomic transactions eg semantic locks, micro-queues
  - Multi statistics update with single command
  - Recursive lookup - Algorithmic TCAM and packet ACL analysis lookups
    - 7 different engines in the device, >80 instances
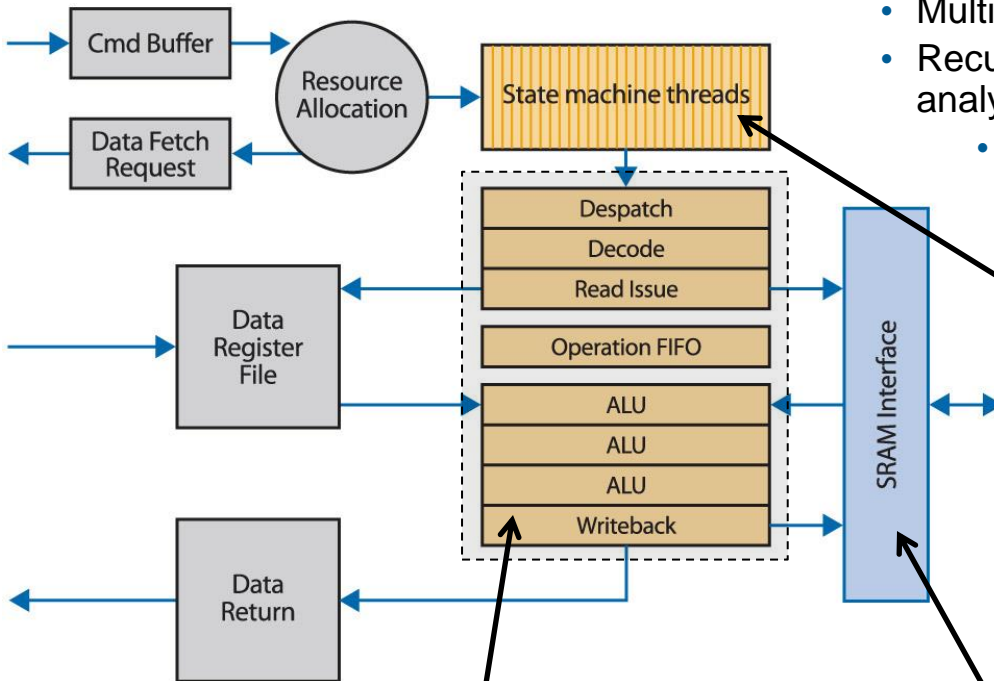
- State machine threads
  - 16-64 per engine
  - In-order operation chaining for overlapping addresses
  - Dynamically allocated data register file resources
  - For externally-backed memory, include DDR cache tags

- Operation pipeline
  - Engine-function specific
  - No stalling
  - Performance limited by SRAM bandwidth (up to 1GOps per engine)

- SRAM interface
  - Forwards data from writes to reads
  - Posted reads; latency non-deterministic
  - Requires operation FIFO to accommodate

Diagram labels: Cmd Buffer, Resource Allocation, State machine threads, Data Fetch Request, Despatch, Decode, Read Issue, Operation FIFO, ALU, ALU, ALU, Writeback, Data Register File, SRAM Interface, Data Return

NETRONOME

Flow Processing Acceleration

Hardware Accelerators are a Good Thing

Hash Engines
Consume limited set of packet contents and produce result

Bulk Cryptography
Consume packet contents and keys, produce new packet contents
Ideal in unit with separate DMA and multithreaded control path

Data Distributors
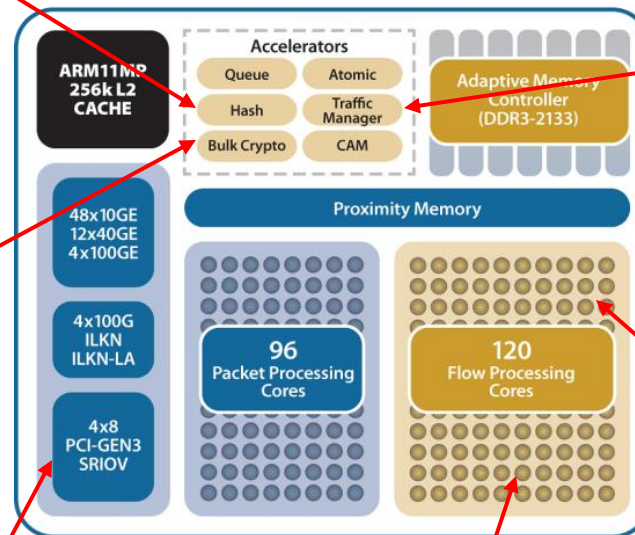Interconnect to multisocket multicore x86 for session processing

Traffic Management & Packet Modification
Order / schedule packets, modify and transmit

Lock Handling, Statistics, Load Balancer
Consume packet/byte count data, maintain statistics/balancing databases

String Matching
Consume limited set of packet contents and produce result

ARM11MP 256k L2 CACHE

Accelerators
Queue | Atomic
Hash | Traffic Manager
Bulk Crypto | CAM

Adaptive Memory Controller (DDR3-2133)

48x10GE 12x40GE 4x100GE

4x100G ILKN ILKN-LA

4x8 PCI-GEN3 SRIOV

Proximity Memory

96 Packet Processing Cores
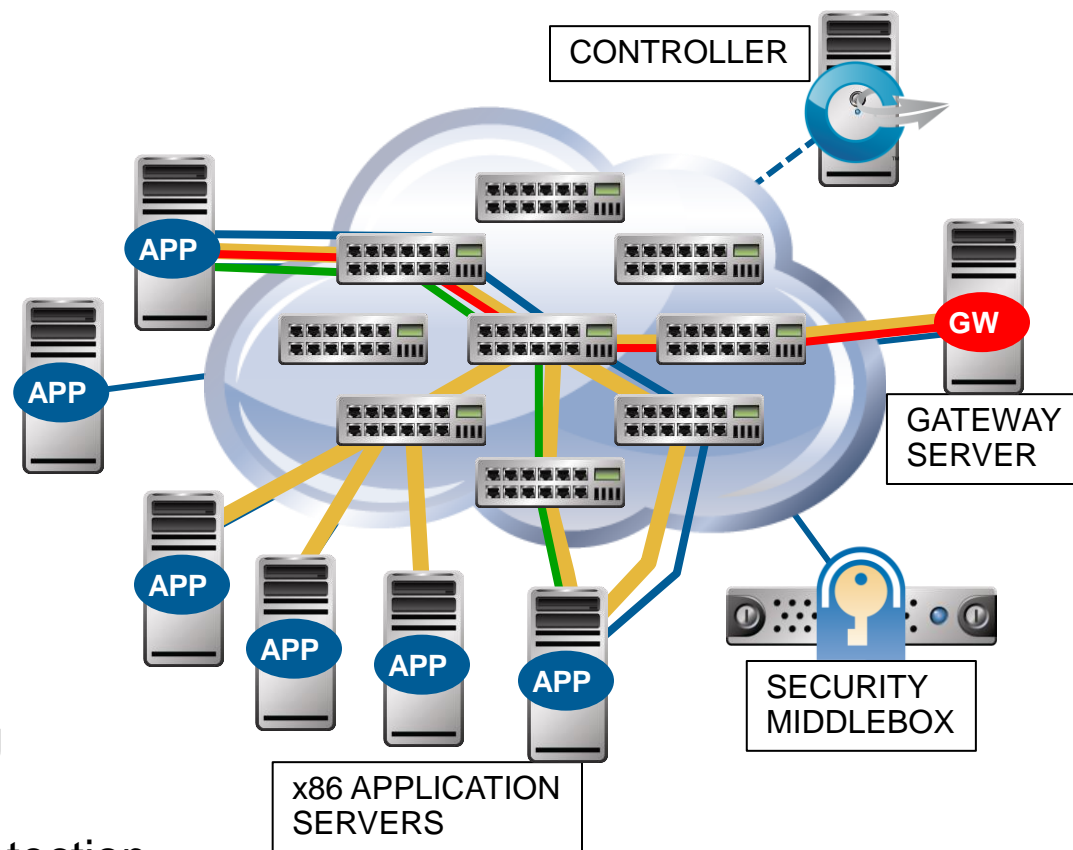
120 Flow Processing Cores

NETRONOME

SDN Data Plane Acceleration
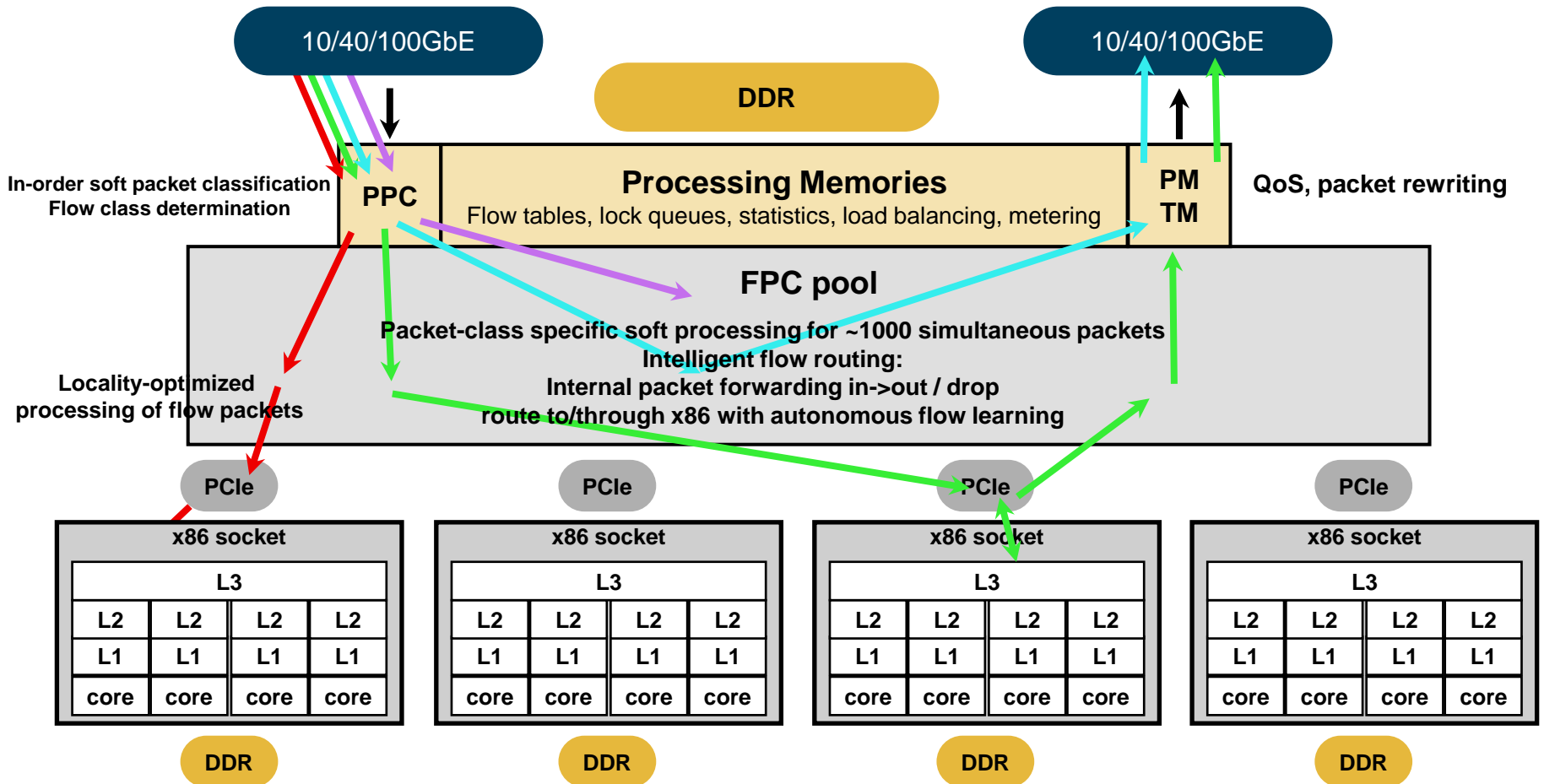
Acceleration at the Edge

- Port-to-port packet transfer
- Stateful flow classification and state tracking
- Virtualization
- Tunnelling/detunnelling
- Quality of Service (QoS)
- Complex control handling
- Security processing

Intelligent Network Interface

- Load balancing
- Flow routing, packet forwarding
- TCP offload
- Protocol parsing, application detection
- TCP offload / stream reassembly
- IP defragmentation / packet reorder

CONTROLLER

APP

APP

GW

GATEWAY SERVER

APP

APP

APP

APP

SECURITY MIDDLEBOX

x86 APPLICATION SERVERS

NETRONOME
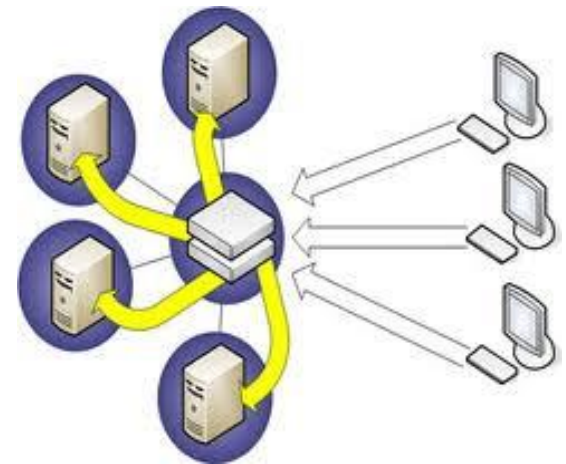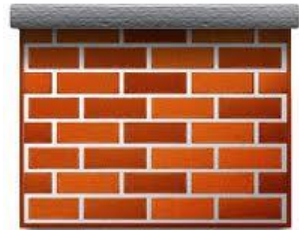
Heterogeneous Processing for SDN

Cyber Security

- Stateful Firewalls
- Advanced auditing and policing of user/application security policies
- Deep Packet/Flow inspection (DPI/DFI)
- Intrusion Detection and Prevention Systems (IDS/IPS)
- Botnet / Malware Detection
- Application security, Anti-virus
- Web Security, DDoS filter
- Web/email content filter, Spam filter
- Lawful Intercept

The notion of a flow and network state are critical for many of today's network and cloud security applications