FTF 2016
TECHNOLOGY FORUM

USER-SPACE NETWORKING WITH ODP AND DPDK

FTF-NET-N1840

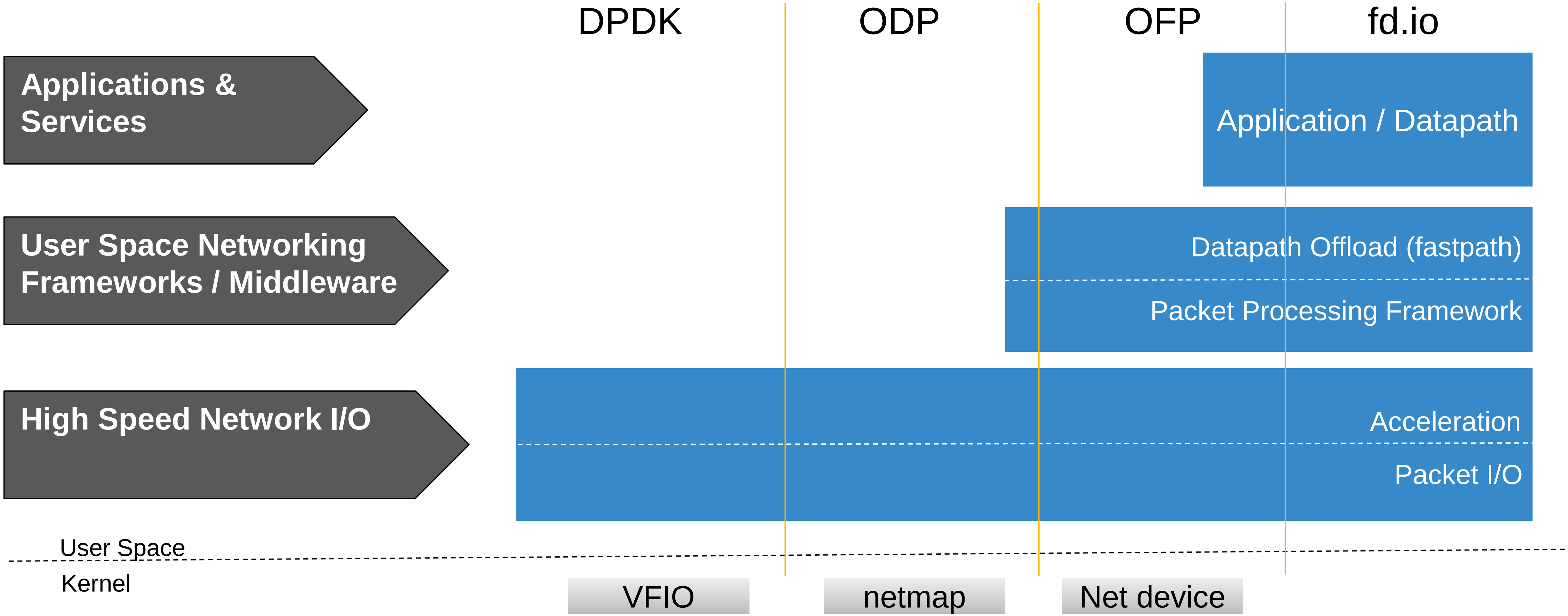
RAVI MALHOTRA
PRODUCT MARKETING
FTF-NET-N1840
MAY 17, 2016

AGENDA

- User-space Networking – Trends
- DPDK and ODP
- FD.IO and Open Fast Path
- NXP Solutions for User-space Networking

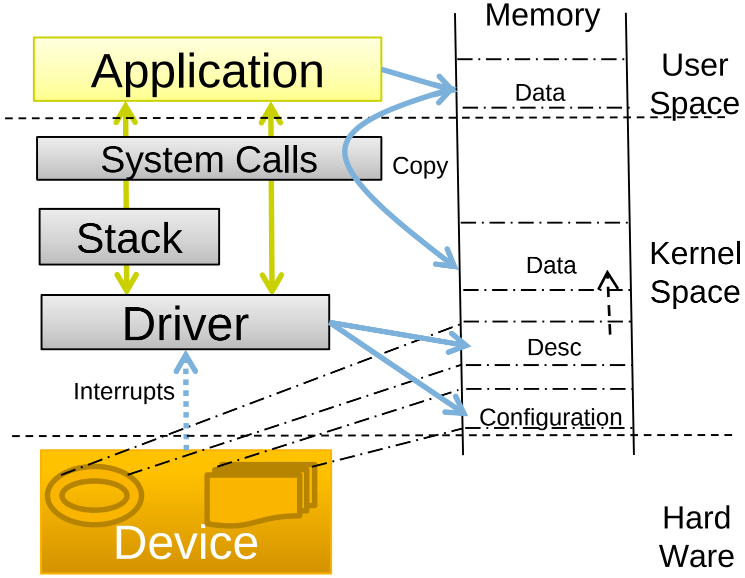
USER-SPACE NETWORKING – TECHNOLOGY AND ADOPTION TRENDS

Key Open Initiatives for User Space Networking

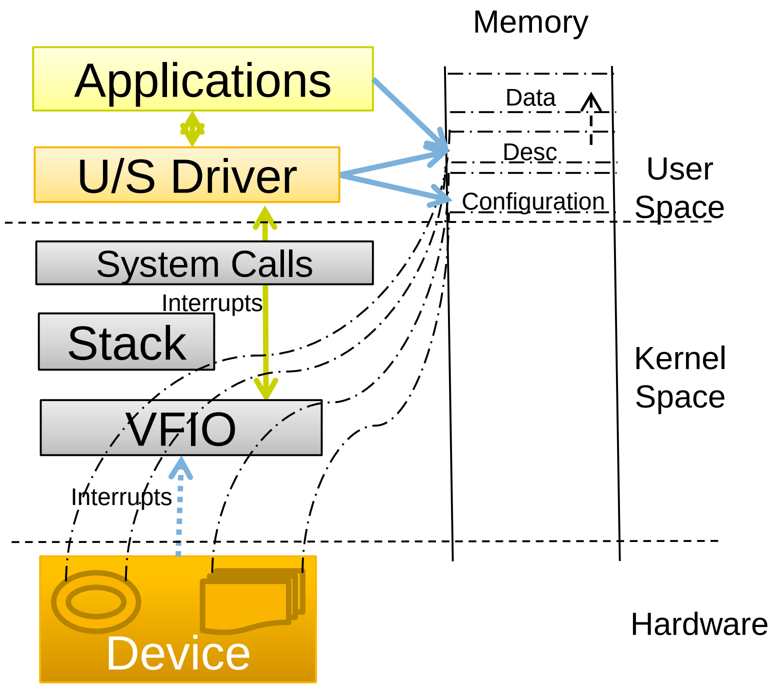


User space network allows network I/O and packet processing frameworks to co-reside with Application, resulting in improved performance, flexibility and agility

Kernel Vs. User Space Applications



Kernel drivers e.g. eTSEC



User-space drivers e.g. USDPAA

- Benefits of user-space applications
  - Flexible threading/process model
  - Isolation of memory
  - Easy to re-start
  - Simpler management of resources
  - Standardized System call interface & libraries
  - Freedom of licensing – not necessarily GPL

- User-space drivers remove overhead of data copying & configuration.
- Mapping entire device memory in application space provides isolation

Traditional User-space Offerings – Vendor Proprietary

Vendor	Offering	Platform	Year introduced
Broadcom	Hyper-Exec, NetOS	XLR, XLP, XLS	2004
Cavium	Simple-exec, US App layer	Octeon	2005, 2009
Freescale	Lightweight-exec, USDPAA	QorIQ DPAA	2008, 2009
LSI	Run-time environment	Axxia	2010
Intel	DPDK	x86	2011

- **Traditionally, user-space offerings evolved from bare-metal counterparts**
  - Very low-level API
  - Designed for highest performance, and not for ease-of-use or portability

- **Use-cases were targeted – e.g.**
  - Routing/Gateway fast-path
  - Base-band transport and L2/L3 processing

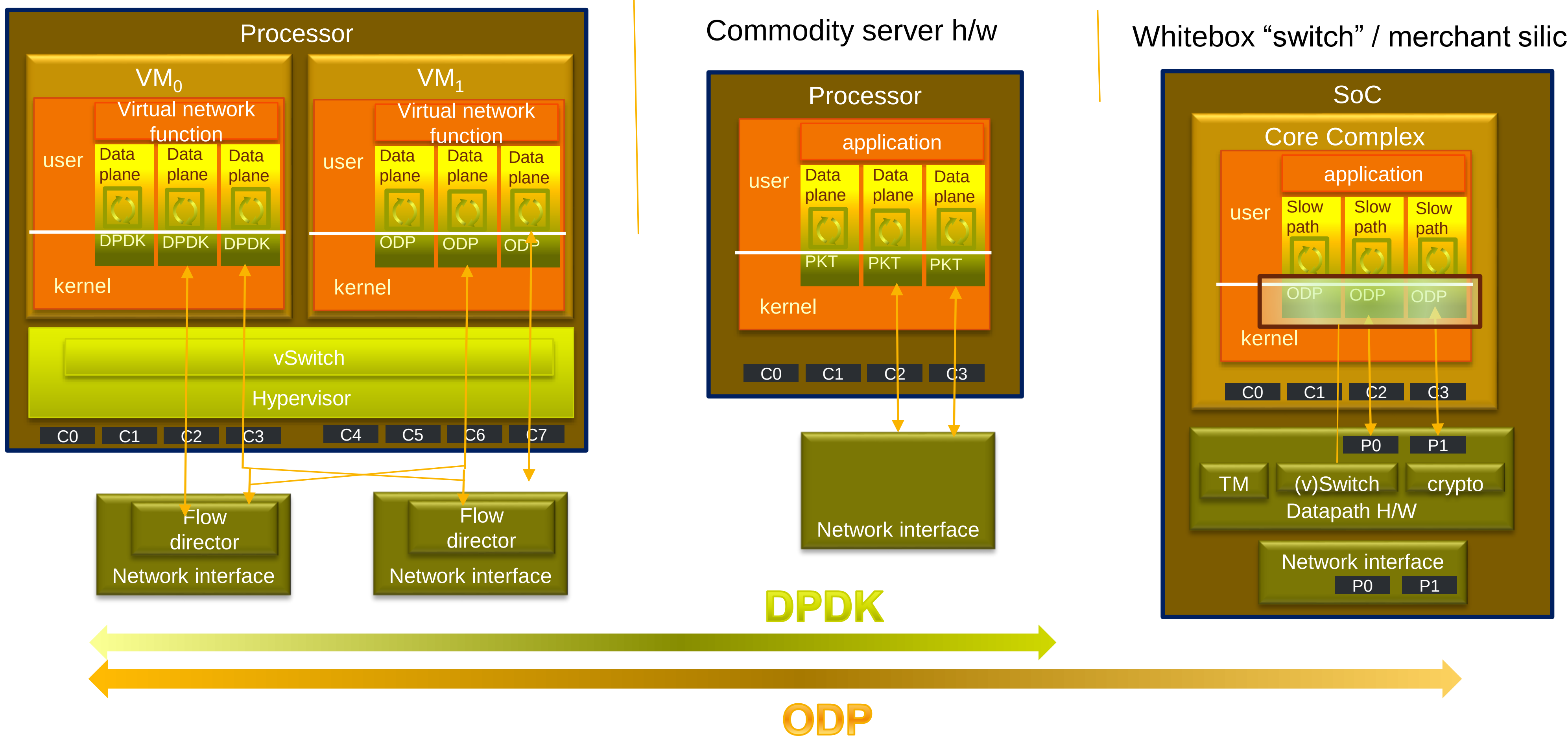
New Market Drivers – NFV & SDN

NFV

Network appliances virtualized to run on commodity hardware throughout the network

SDN

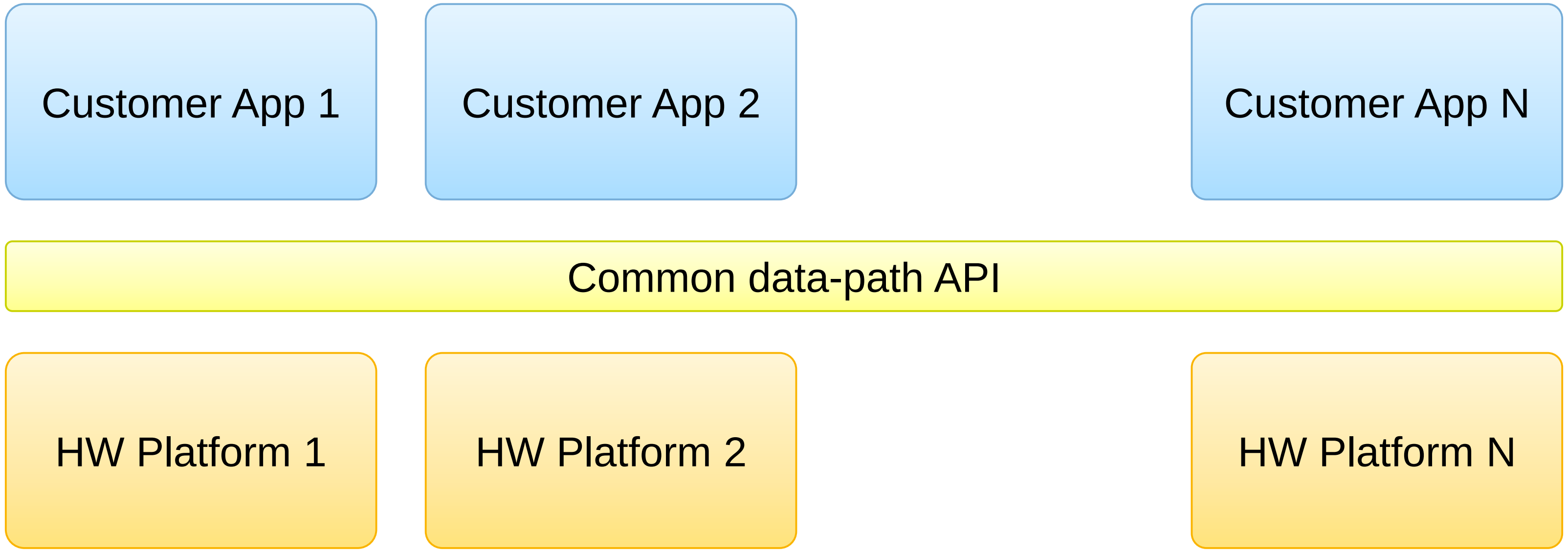
Network infrastructure defined through software, running on commodity server hardware or whitebox / merchant silicon switches.



Commodity server h/w

Whitebox “switch” / merchant silicon

Need For a Common Data-path API



Customer App 1

Customer App 2

Customer App N

Common data-path API

HW Platform 1

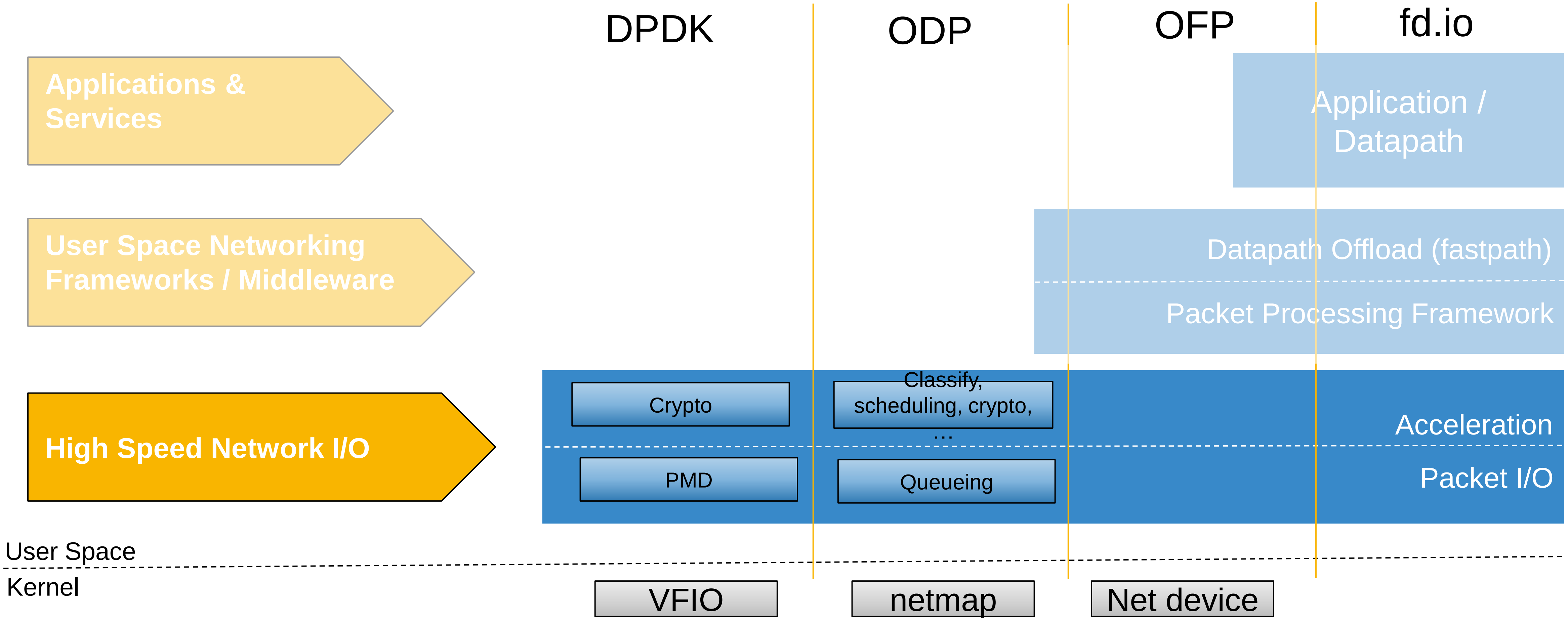
HW Platform 2

HW Platform N

- **A common API**
  - Increases portability of applications across several HW platforms
  - Increases the number of applications that can run on a HW platform.

- **Is it possible, even probable?**
  - Basic I/O, acceleration and run-time services – Yes.
  - HW vendors will continue to add differentiation, value-added services for advanced functionality.
  - Provisioning and management also needs standardization – especially for NFV deployment.

Key Open Initiatives for User Space Networking



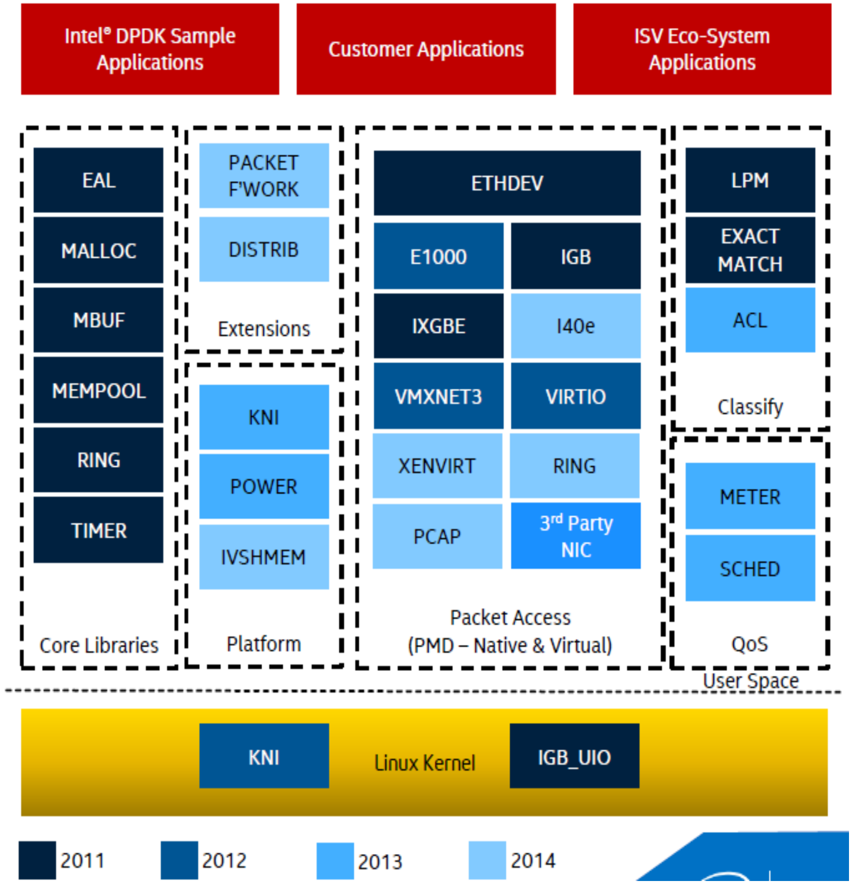
Focus of Network I/O APIs is to present traffic to user space applications with greatest performance and allow access to packet acceleration functions

Data Path Development Kit – DPDK

- **Now Part of Linux Foundation**
  - Targeted towards SDN and NFV
  - Large developer/user community

- **Open to Other Platforms**
  - IBM-PPC (Power8), Tensilica as host
  - ARMv8, SoC platforms

- **Key Features**
  - Core run-time libraries
    - x86 optimized run-time services
    - EAL abstracts processing model
  - Packet-access (I/O)
    - Ethernet device framework (PMD)
    - Intel NICs, virtual IO & 3rd Party NICs
  - Classification & QoS
    - Leverages HW support+ SW libraries.
  - Platform services
    - Kernel NW Interface, Power-mgmt
    - Shared-mem (inter-VM, app)
  - Crypto API



Intel® DPDK Sample Applications

Customer Applications

ISV Eco-System Applications

EAL

MALLOC

MBUF

MEMPOOL

RING

TIMER

Core Libraries

PACKET F'WORK

DISTRIB

Extensions

KNI

POWER

IVSHMEM

Platform

ETHDEV

E1000

IGB

IXGBE

I40e

VMXNET3

VIRTIO

XENVIRT

RING

PCAP

3rd Party NIC

Packet Access (PMD – Native & Virtual)

LPM

EXACT MATCH

ACL

Classify

METER

SCHED

QoS

User Space

KNI

Linux Kernel

IGB_UIO

2011

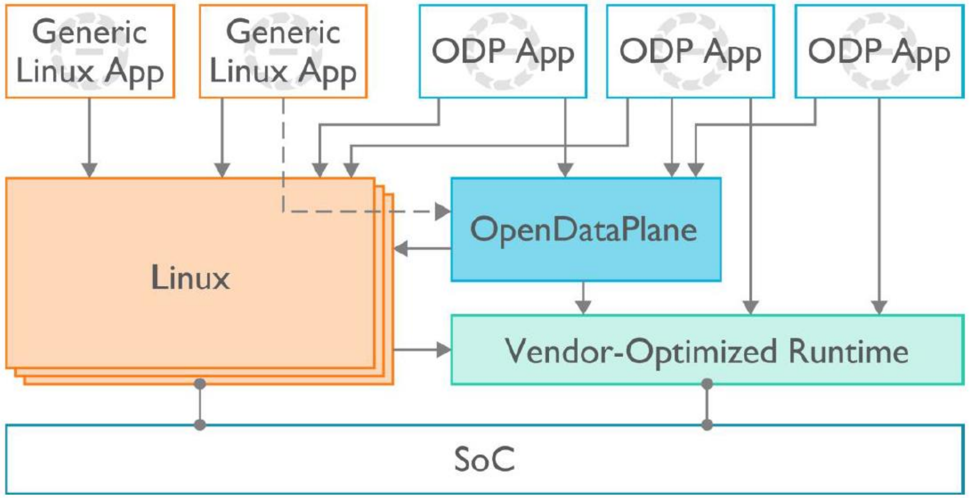
2012

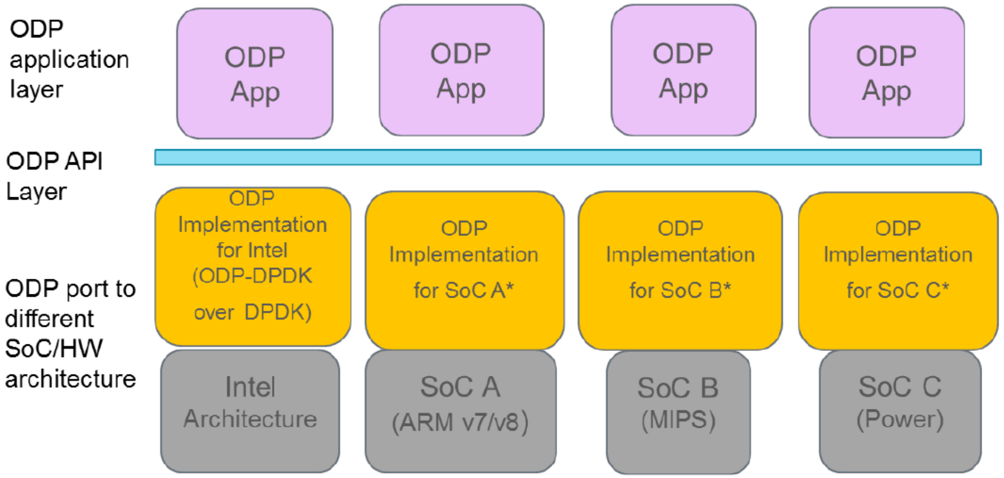
2013

2014

Linaro Open Data Plane – ODP

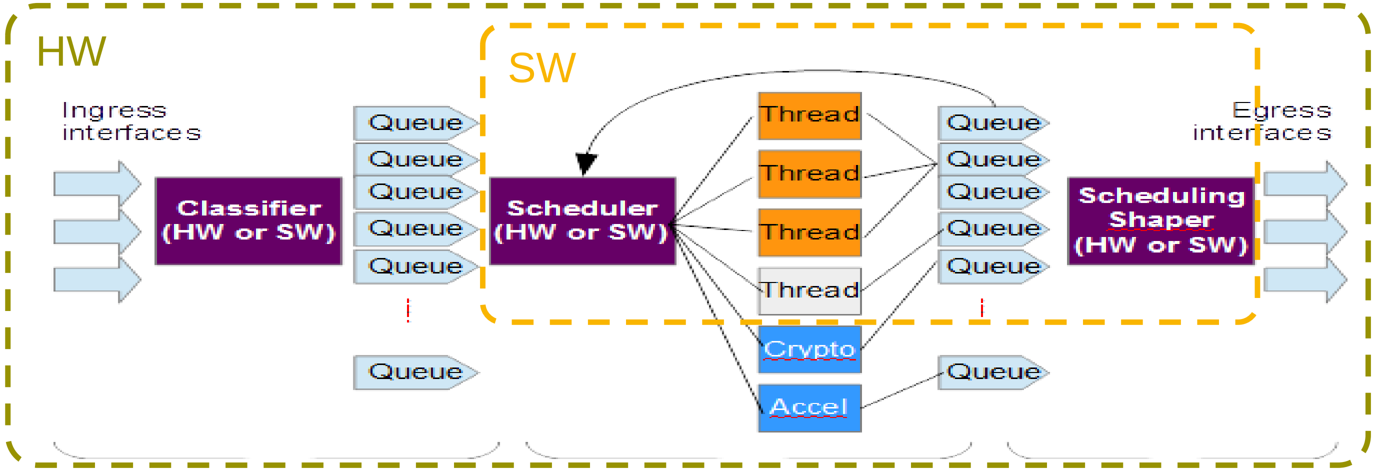
- **Community Effort**
  - Driven by Linaro NW group
  - Broadcom, Cavium, TI, Freescale – vendors
  - Cisco, NSN – key customers
- **Main Focus**
  - Define a common ODP API
  - Sample implementation for
    - Linux-generic
    - ODP API mapping to DPDK.
  - Implementations provided by HW vendors (not linaro.org)
  - Allows applications to use vendor-specific extensions.
- **Key Features**
  - Run-time – timers, sync, memory, buffers.
  - Multiple Packet I/O modes
  - Flexible queuing and scheduling.
  - Crypto offload – IPSec
  - Classification and QoS





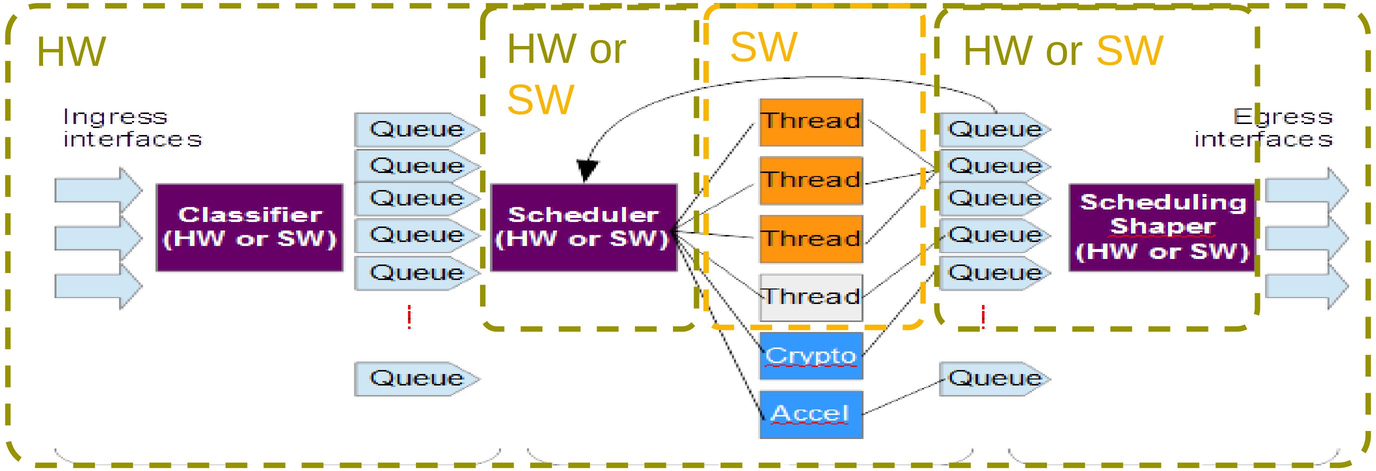
ODP application layer
ODP API Layer
ODP port to different SoC/HW architecture
ODP App
ODP App
ODP App
ODP App
ODP Implementation for Intel (ODP-DPDK over DPDK)
ODP Implementation for SoC A*
ODP Implementation for SoC B*
ODP Implementation for SoC C*
Intel Architecture
SoC A (ARM v7/v8)
SoC B (MIPS)
SoC C (Power)

DPDK vs. ODP – HW Acceleration



DPDK Approach:

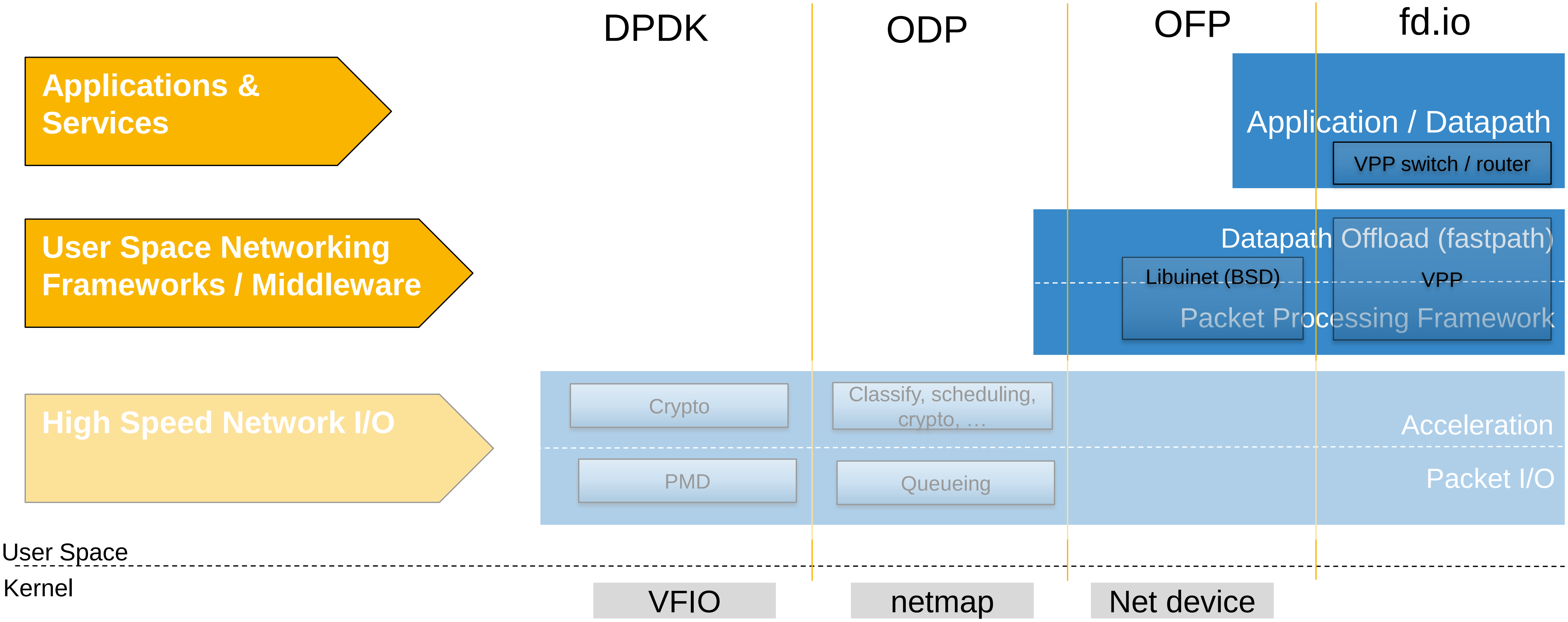
- Designed for Simple NICs
- Works well for large number of balanced flows
- SW implementation comes at a cost.



ODP Approach:

- Flexible design – blocks can be in HW or SW
- Works for balanced & unbalanced traffic flows
- Works well with Accelerators, multiple I/O sources

Key Open Initiatives for User Space Networking



DPDK

ODP

OFP

fd.io

Applications & Services

User Space Networking Frameworks / Middleware

High Speed Network I/O

Application / Datapath

VPP switch / router

Datapath Offload (fastpath)

Libuinet (BSD)

VPP

Packet Processing Framework

Crypto

Classify, scheduling, crypto, …

Acceleration

PMD

Queueing

Packet I/O

User Space

Kernel

VFIO

netmap

Net device

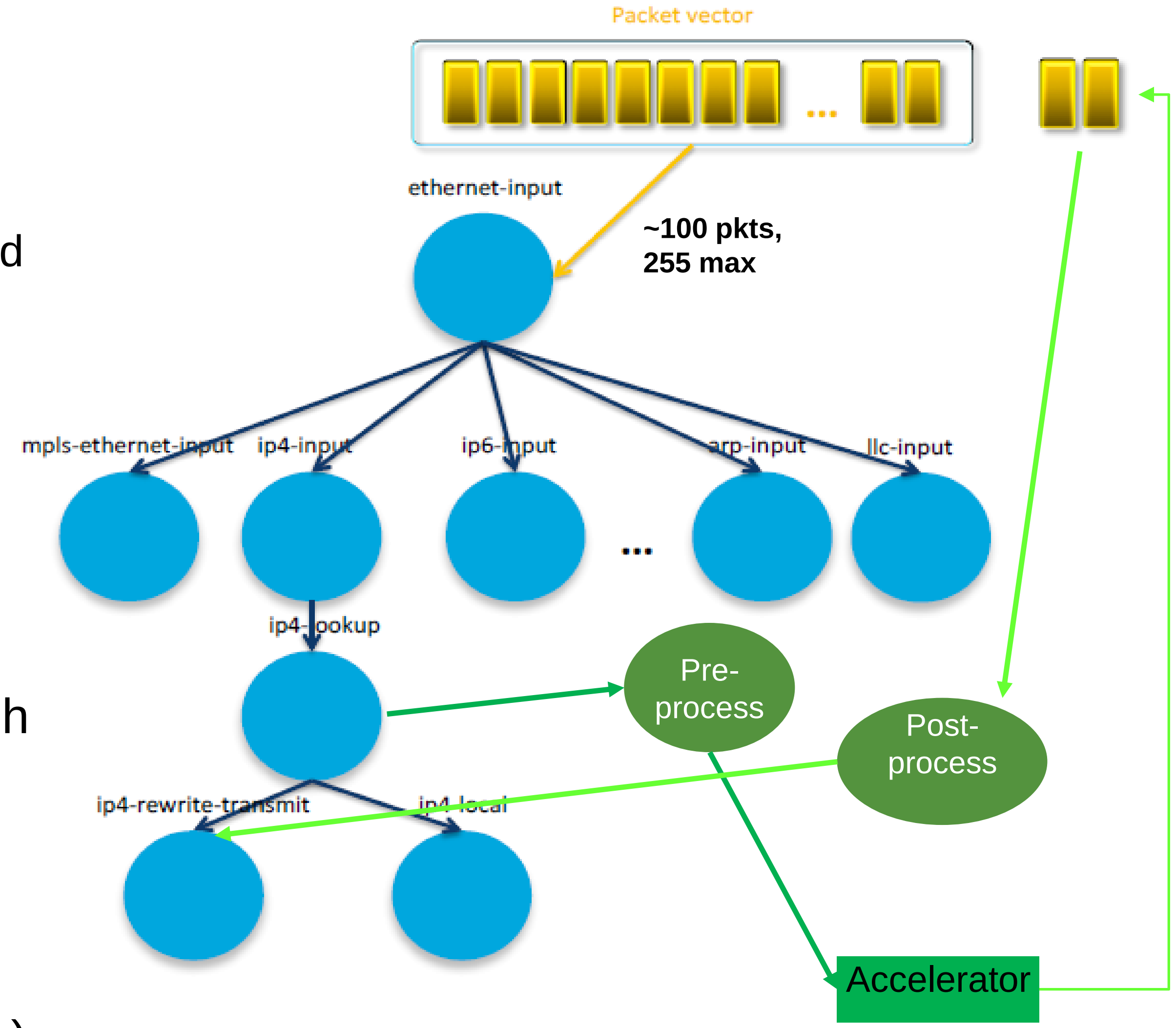
Higher order APIs provide means to compose user space networking stacks and applications in optimal ways that take advantage of high speed network I/O and provide reference applications

Fast Data Project (FD.IO)

- Linux Foundation project “relentlessly focused on data IO speed and efficiency for more flexible and scalable networks and storage”, based on initial contribution from Cisco
  - Network I/O – based on DPDK drivers, performance-optimized for VPP
  - Vector packet processing (VPP) – Core project
    - Modular plug-in architecture for composing data paths as graphs of reusable nodes
    - Suite of function-specific graph node modules for common data path operations (e.g. classify, packet re-write, etc.)
    - Highly optimized for gen purpose CPUs (memory hierarchy, superscalar, pipelining, etc.)
    - Acceleration can be easily substituted for software-implemented graph nodes
  - Application integration – SDN / NFV
    - Production quality switch/router reference application composed from VPP framework
    - Netconf/Yang southbound controller dataplane management agent for OpenDaylight Integration

Pluggable Packet Processing Graphs

- **Batching (vector) model**
  - Process batch of packets through pipeline for optimal performance (e.g. i-cache utilization)
    - Drain ingress “queue” until up to 255 packets vectorized
  - Vector is traversed through nodes of graph (data path)
  - Process 2-3 packets within a graph node in non-blocking fashion (avoid pipeline stalls)
- **Pluggable graph node model**
  - Binary plugins may substitute any node in the graph
  - Hardware acceleration integrated using plugins
  - Accelerated graph node enqueues packets to hardware accelerator
  - Accelerated nodes exit graph when invoking synchronous acceleration operations (non-blocking)

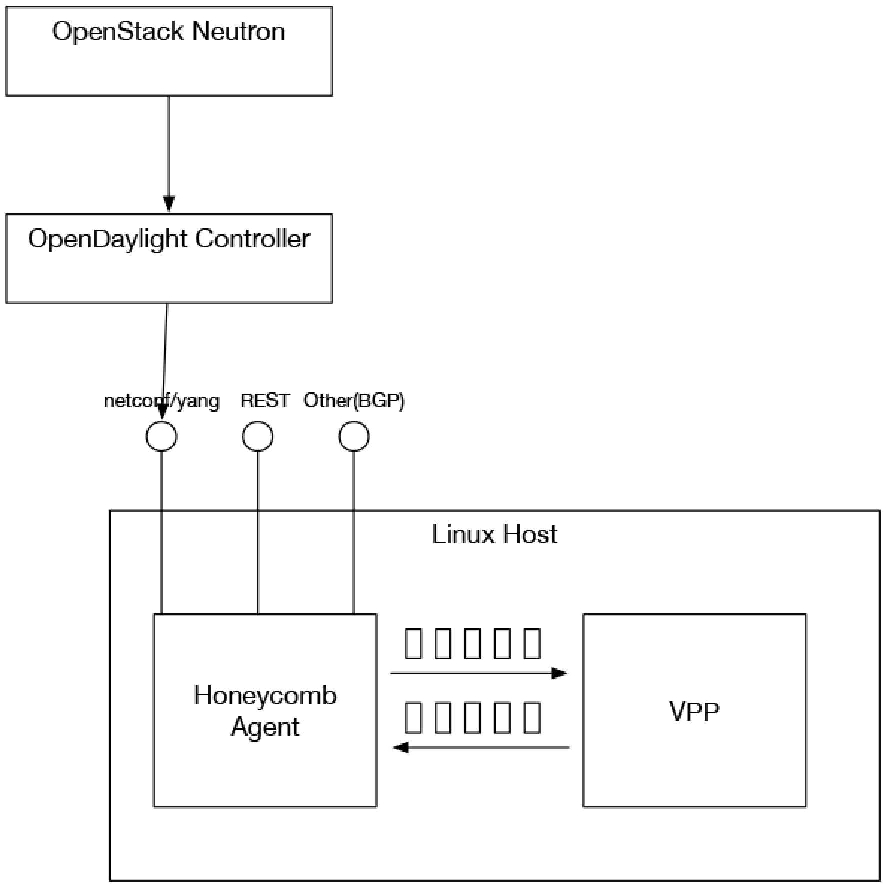


High Performance Switch/Router

- **VPP switch / router to function as plugin replacement for virtual forwarding engine (OpenStack Neutron)**
  - Dataplane management agent (Honeycomb) handles southbound controller interface allowing OpenDaylight to configure dataplane
  - Based on Yang/netconf
  - Provides BGP support allowing FIB synchronization (e.g. FPM)*

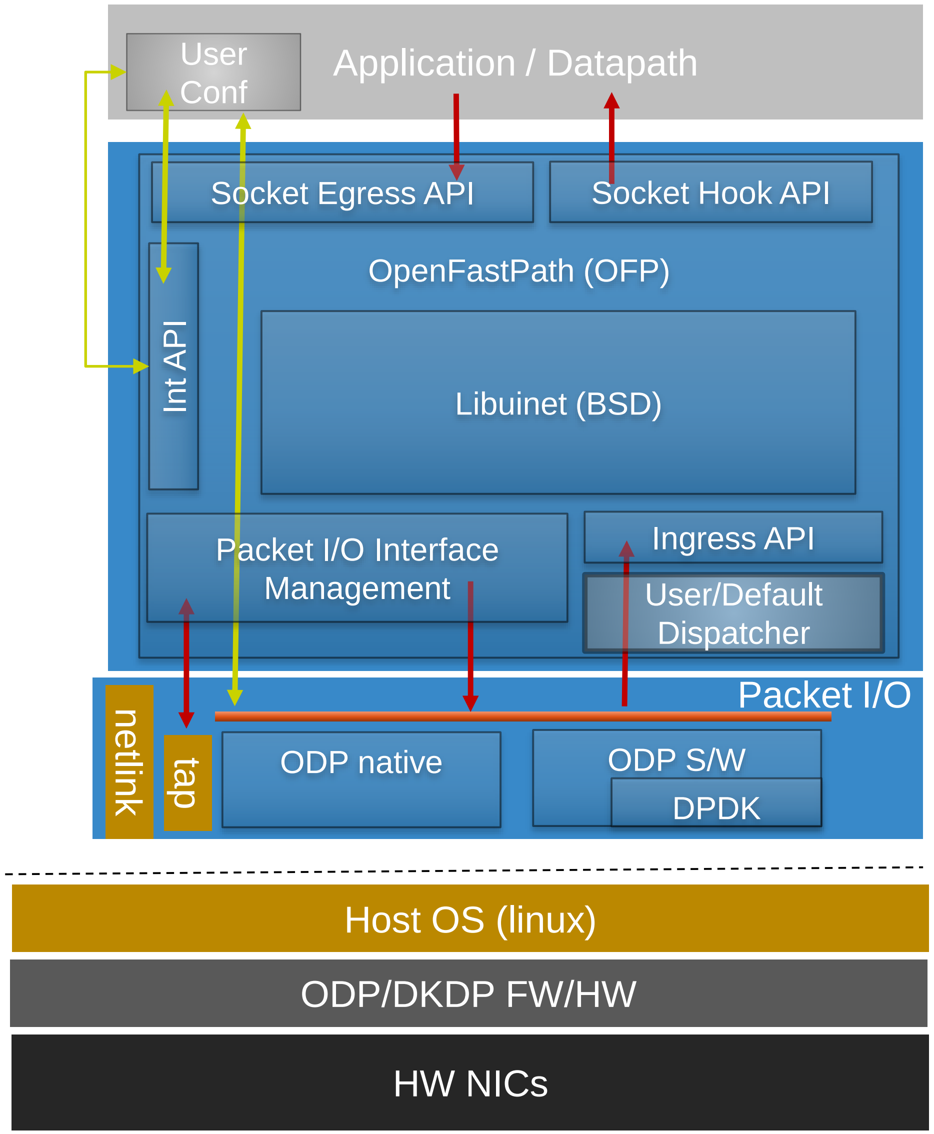
- **Performs better on NIC-to-VM and VM-VM performance:**
  - Compared with OVS-DPDK (openflow model)
  - Exhibits significantly better scaling with FIB size

- **Comparison against OpenFlow**
  - Uses traditional (sequential code) model of packet processing, rather than flow-based
  - Doesn’t rely on primitives like flows, tables, match/action directives as in OpenFlow



Open Fast Path (ofp)

- User space port of the BSD TCP/IP networking stack
- Accelerated IPv4/IPv6 routing/forwarding
- Improved performance for termination / tunneling applications
- Derived from libuinet project
- Synchronized CP (e.g. FIBs) processing via netlink
- Optimized for use with Open Dataplane:
  - Takes advantage of scheduling, classification, acceleration
  - Interoperable with DPDK (using ODP-DPDK)

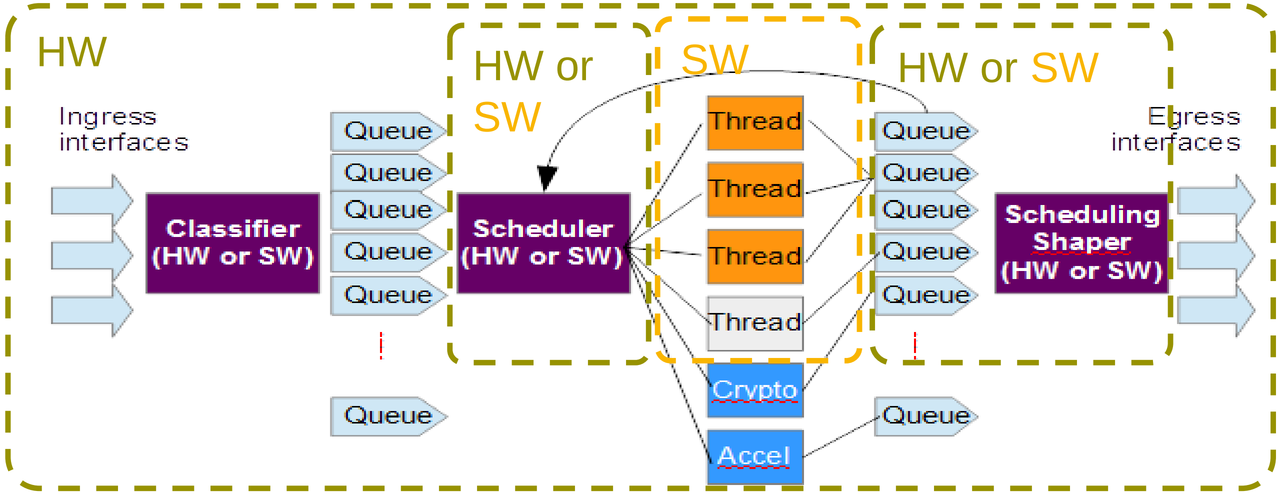


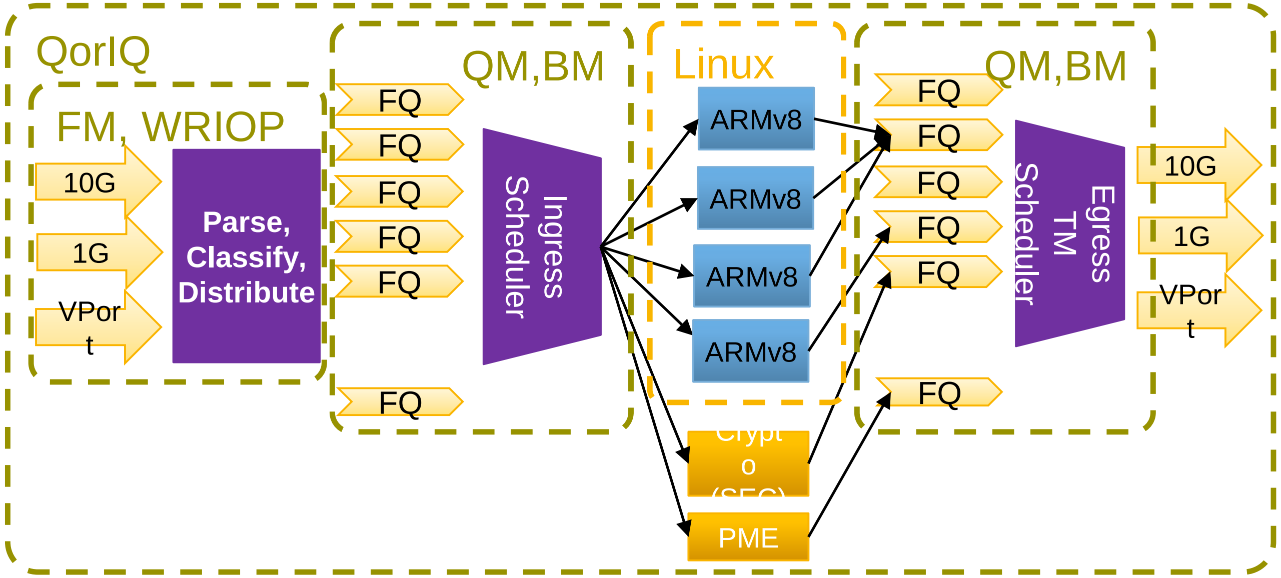
User Space

Kernel

QorIQ USER-SPACE INFRASTRUCTURE

DPAA – Compatible With DPDK & ODP Since 2008





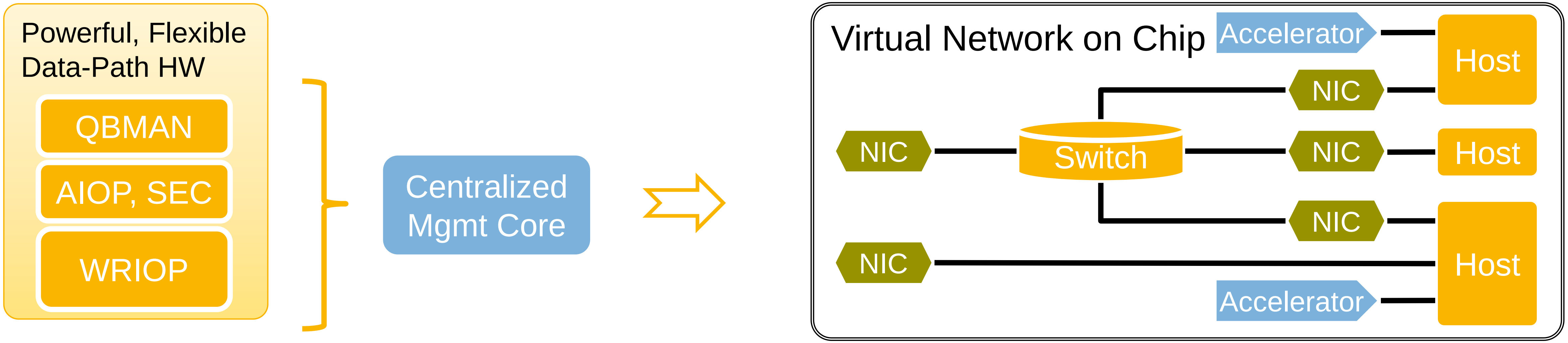
DPDK/ODP Approach:

- Leverage hardware for ingress and egress processing
- ODP adds on HW scheduling offload
- Accelerators – Crypto offload
- Complete user-space processing model

NXP Approach:

- FMan offloads parsing, classification, distribution.
- QMan, BMan offload scheduling, buffering
- Virtualized accelerators – SEC, PME, DCE
- User-space driver, threading model
- Doing all this since 2008 – now into 3rd generation

Data-Path Infrastructure for the Evolving Network



Powerful, Flexible Data-Path HW

QBMAN

AIOP, SEC

WRIOP

Centralized Mgmt Core

Virtual Network on Chip

Accelerator

Host

NIC

NIC

Switch

NIC

Host

NIC

NIC

Host

Accelerator

Ease of use

- NW Object model
- Flexible resource management

Virtualized

- Logical Network resources
- Fully isolated
- Runtime configuration

Standard

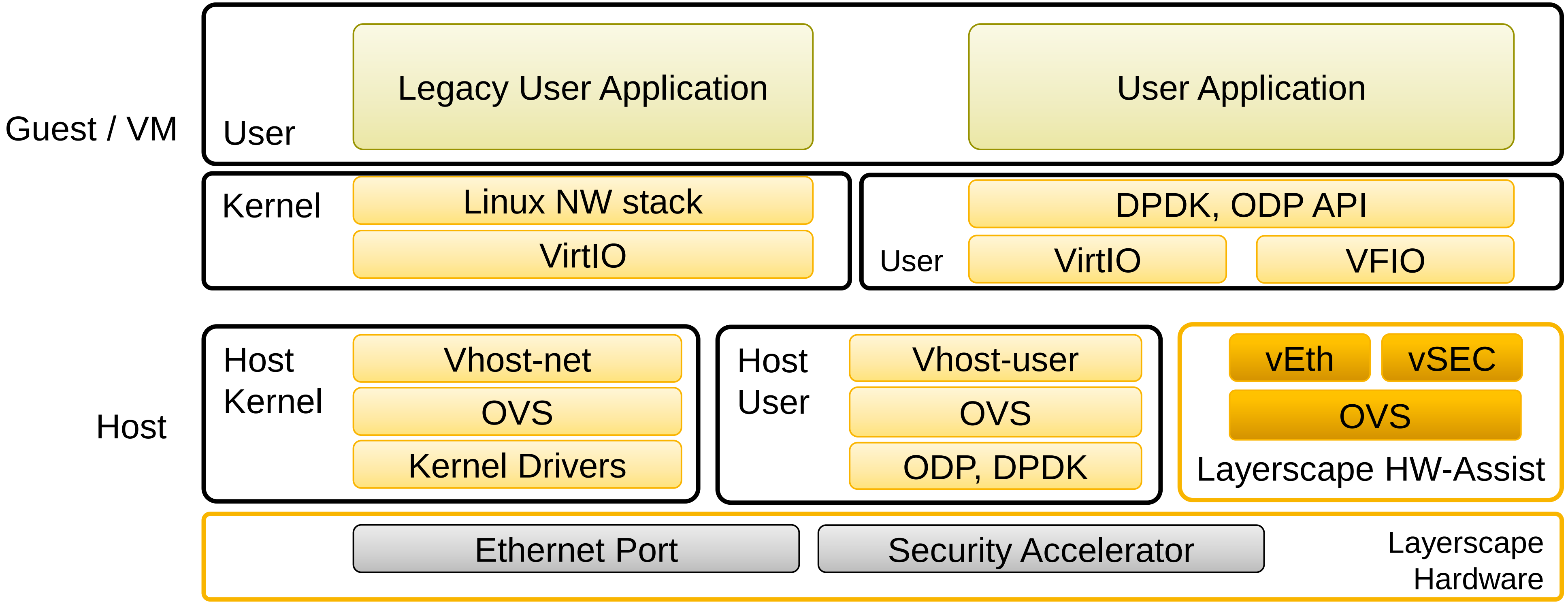
- Standard Linux drivers and DPDK/ODP API
- Virtual NW topology same as physical topology

Programmable

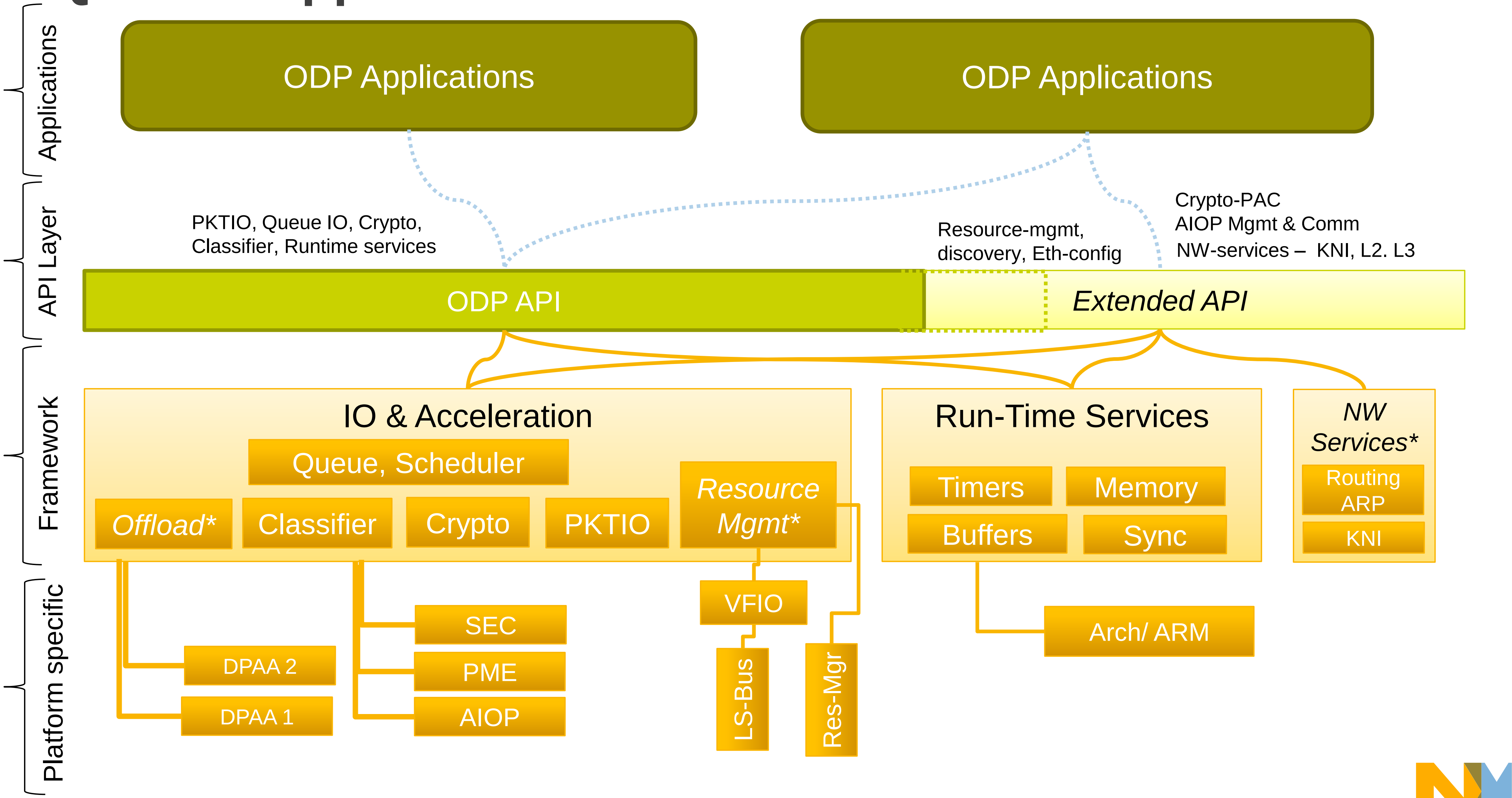
- ANSI C-programmable AIOP packet engine, tools
- Powerful Security, DPI, Compression Engines

User-space Networking for Virtualized Environments

Front/Back-end	Kernel/Kernel	Kernel/User	User/User	User/HW
Portability	Highest	High	High	Medium
Performance	Low	Medium	Medium	Highest
Differentiation	Low	Medium	Medium	High



QorIQ ODP Support



QorIQ ODP Support

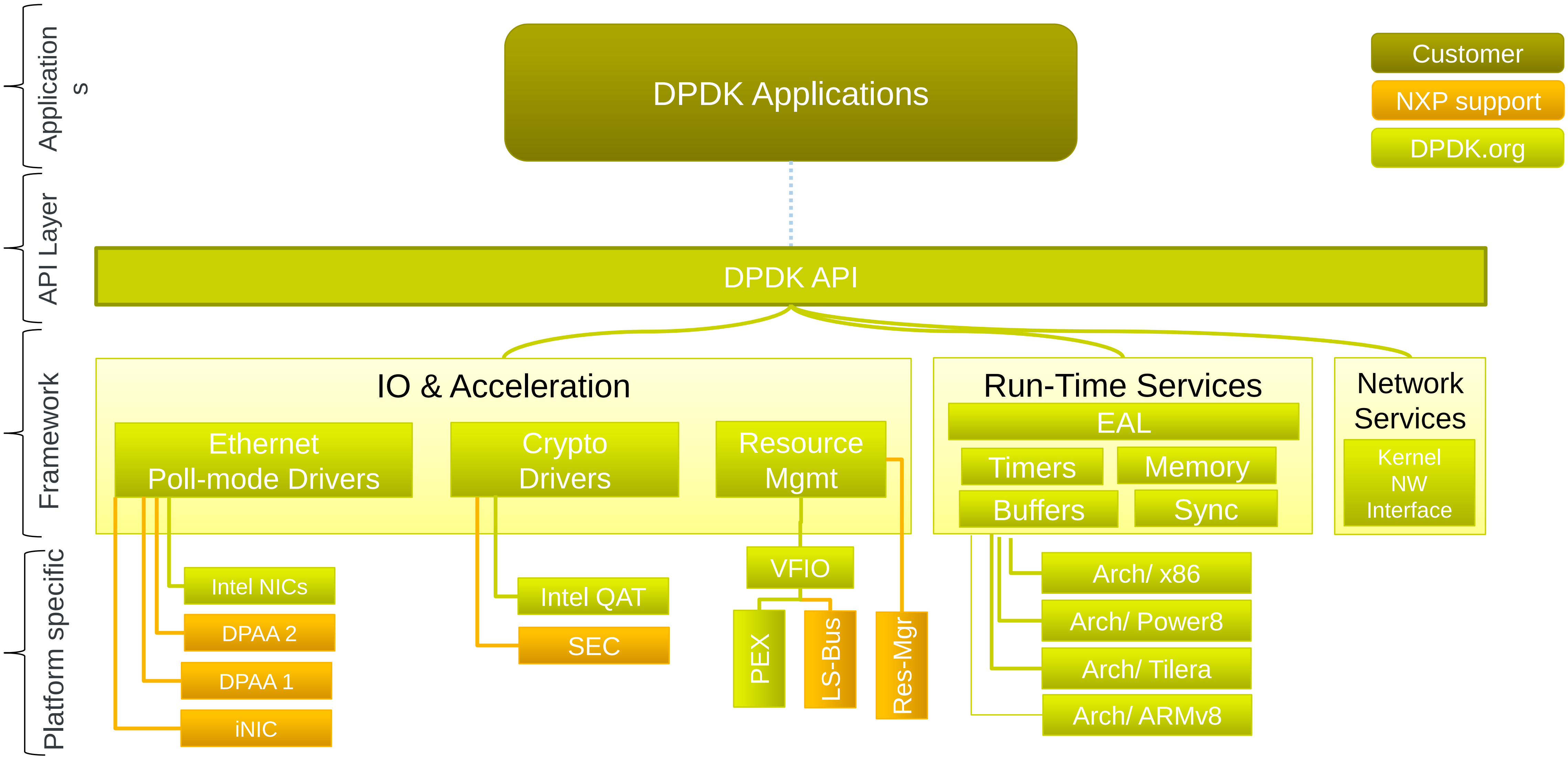
- **QorIQ HW is inherently aligned to ODP**
  - Classification and scheduling
  - HW queue and buffer mgmt
  - Crypto & other HW offloads
  - ARM 64-bit cores

- **Complete ODP-API coverage**
  - Queue and Scheduler API
  - PKTIO and Classifier API
  - Crypto API – algorithmic and protocol
  - Runtime services incl. pkt-buffers
  - Support for both DPAA1 & DPAA2 platforms
    - LS1043, LS1046
    - LS2088, LS1088

- **QorIQ HW have additional capabilities**
  - Switching, demuxing
  - Application level offloads
  - Virtual networking and resource mgmt
  - Provided as extensions to ODP-API
  - *Efforts underway to make them part of ODP*

- **Value-added ODP extensions**
  - Complete Ethernet capabilities
    - MAC/Phy, IPR/IPF, *GRO/GSO, Smart-NIC*
    - Physical and Virtual Ethernet ports
  - NW services
    - Provide Linux network stack services, visibility
    - Network-devices (KNI), Routing, ARP
  - Resource management
    - VFIO and VirtIO based assignment of resources.
    - Dynamic re-configuration and discovery
    - Multiple application support, flexible process model

QorIQ DPDK Support



QorIQ DPDK Support

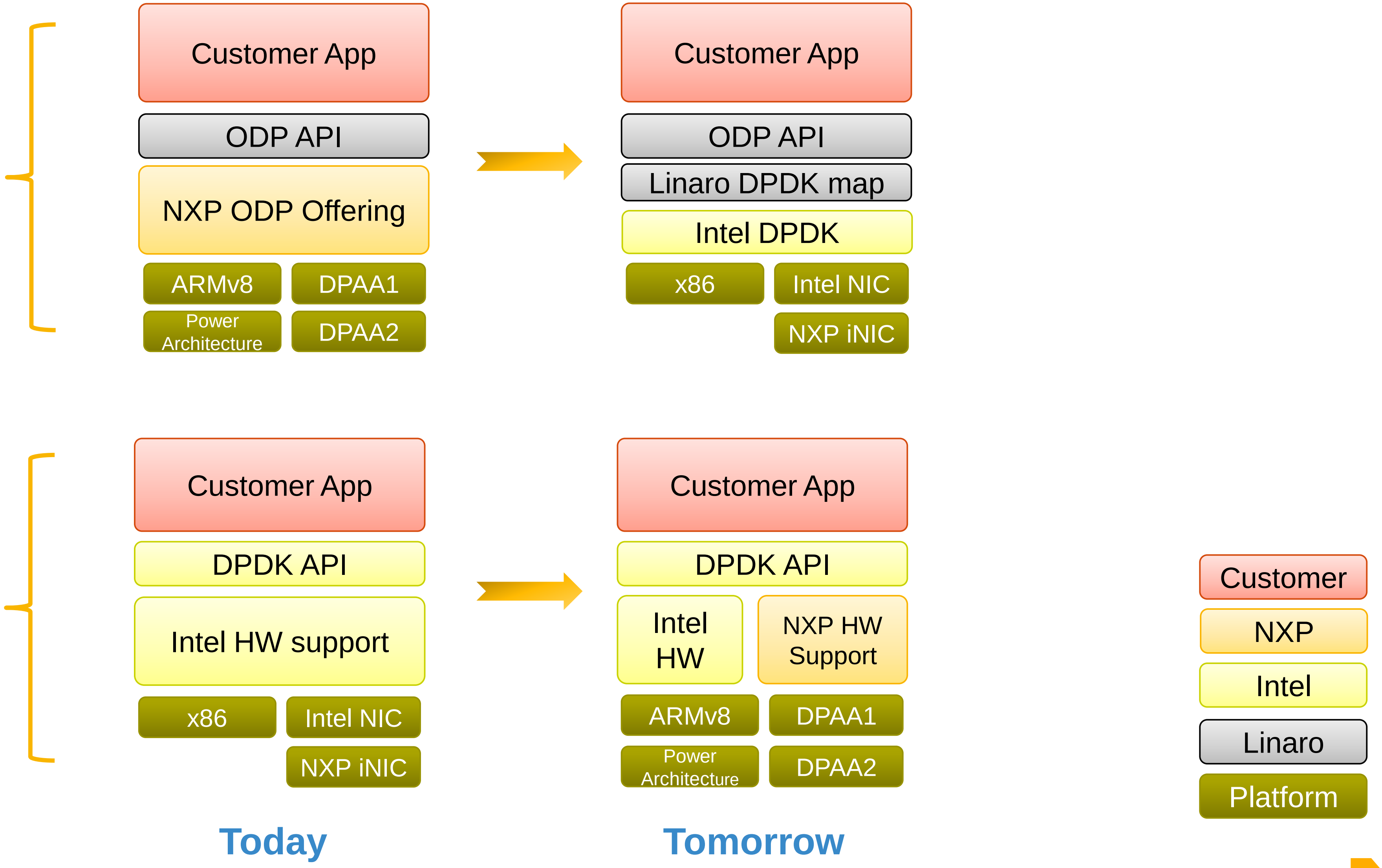
- **Basic Platform support**
  - DPAA1 Poll-mode driver
  - DPAA2 Poll-mode driver
  - iNIC Poll-mode driver
  - Crypto offload to local SEC
  - LS1043, LS1046
  - LS2080, LS2088, LS1088

- **Architectural enhancements**
  - Support for SoC/platform drivers
  - Hardware buffer management
  - Optimizations for ARMv8 run-time

- **Virtualization support**
  - OVS over DPDK in host user-space
  - Vhost-user
  - DPDK in guest/VM
  - Virtio poll-mode driver
  - DPAA2 VFIO poll-mode driver

- **Future work**
  - Ingress scheduling, load balancing.
  - Protocol aware crypto
  - Egress scheduling, QoS offload.

The Quest for Migration



ODP Path
First dev on ARMv8. Power Architecture

Customer App
ODP API
NXP ODP Offering
ARMv8 | DPAA1
Power Architecture | DPAA2

→

Customer App
ODP API
Linaro DPDK map
Intel DPDK
x86 | Intel NIC
NXP iNIC

DPDK Path
First dev on x86

Customer App
DPDK API
Intel HW support
x86 | Intel NIC
NXP iNIC

→

Customer App
DPDK API
Intel HW | NXP HW Support
ARMv8 | DPAA1
Power Architecture | DPAA2

Today

Tomorrow

Customer
NXP
Intel
Linaro
Platform

Summary

- **Need for a common user-space API**
  - Mainly driven by NFV and SDN
  - Best of portability, re-use and acceleration

- **Open Data Plane and Data Path Development Kit**
  - Different origins, communities - but lot of convergence
  - Both will continue to be adopted
  - FD.IO and Open Fast Path provide user-space frameworks and applications on top of DPDK/ODP

- **NXP provides optimized solutions for both ODP and DPDK**
  - Our Data-Path architecture has been compatible since 2008
  - Working with the community to add more acceleration, features
  - Actively engaged in and tracking FD.IO and Open Fastpath communities

NXP

SECURE CONNECTIONS
FOR A SMARTER WORLD

ATTRIBUTION STATEMENT

NXP, the NXP logo, NXP SECURE CONNECTIONS FOR A SMARTER WORLD, CoolFlux, EMBRACE, GREENCHIP, HITAG, I2C BUS, ICODE, JCOP, LIFE VIBES, MIFARE, MIFARE Classic, MIFARE DESFire, MIFARE Plus, MIFARE FleX, MANTIS, MIFARE ULTRALIGHT, MIFARE4MOBILE, MIGLO, NTAG, ROADLINK, SMARTLX, SMARTMX, STARPLUG, TOPFET, TrenchMOS, UCODE, Freescale, the Freescale logo, AltiVec, C 5, CodeTEST, CodeWarrior, ColdFire, ColdFire+, C Ware, the Energy Efficient Solutions logo, Kinetis, Layerscape, MagniV, mobileGT, PEG, PowerQUICC, Processor Expert, QorIQ, QorIQ Qonverge, Ready Play, SafeAssure, the SafeAssure logo, StarCore, Symphony, VortiQa, Vybrid, Airfast, BeeKit, BeeStack, CoreNet, Flexis, MXC, Platform in a Package, QUICC Engine, SMARTMOS, Tower, TurboLink, and UMEMS are trademarks of NXP B.V. All other product or service names are the property of their respective owners. ARM, AMBA, ARM Powered, Artisan, Cortex, Jazelle, Keil, SecurCore, Thumb, TrustZone, and μVision are registered trademarks of ARM Limited (or its subsidiaries) in the EU and/or elsewhere. ARM7, ARM9, ARM11, big.LITTLE, CoreLink, CoreSight, DesignStart, Mali, mbed, NEON, POP, Sensinode, Socrates, ULINK and Versatile are trademarks of ARM Limited (or its subsidiaries) in the EU and/or elsewhere. All rights reserved. Oracle and Java are registered trademarks of Oracle and/or its affiliates. The Power Architecture and Power.org word marks and the Power and Power.org logos and related marks are trademarks and service marks licensed by Power.org. © 2015–2016 NXP B.V.