Zephyr® Project
Developer Summit 2022

June 8-9, 2022
Mountain View, CA + Virtual

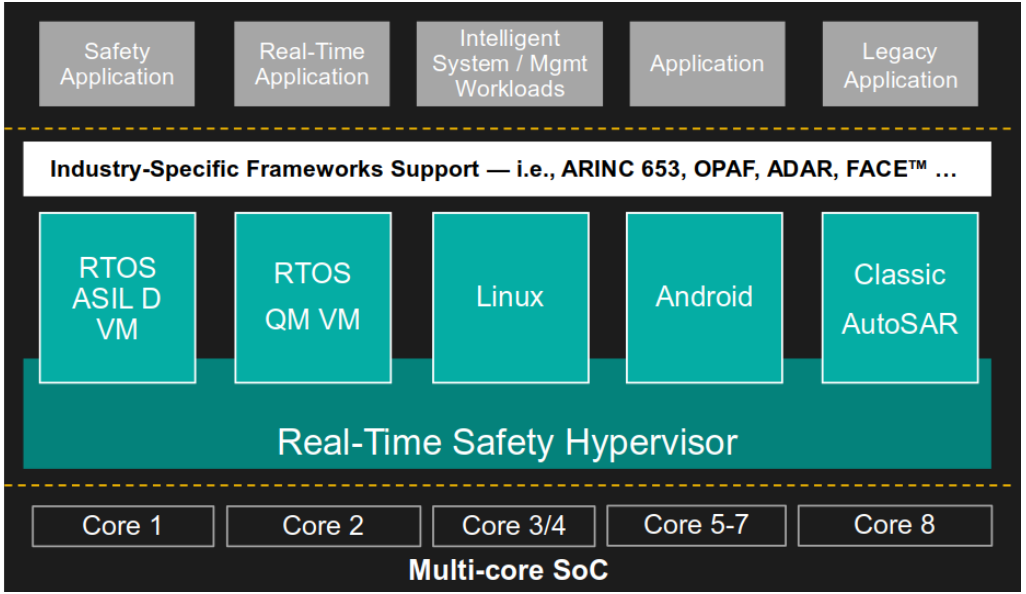
Hypervisor-less virtio: Assembling Multi-OS systems using standards-based protocols for intra-SoC connectivity and device sharing

Dan Milea
Wind River

Intelligent Edge

- Modern SoCs are powerful heterogeneous multiprocessing compute platforms aggregating multiple silicon building blocks like application processors, real-time processors, programmable logic, sensors, accelerators (FPGAs, GPUs, DSPs, TPUs, NPUs)
- Increasing consolidation of embedded systems onto shared HW
  - safety certified operating systems and bare metal applications for safety critical functions (ADAS, engine control, connectivity)
  - Real time OS for early functions (remote cameras, audio, etc.)
  - Rich operating systems (Android, Linux) for dashboard, IVI, navigation
- Increased software complexity, diversity and volume of software
- Individual runtime elements have separate life-cycles and need to be updated and reused across product releases and product variants

Partitioned systems



- Complex SoCs as flexible integration platforms
- Integration of many levels of safety, real-time, and manageability on one platform
- Decoupled multi-vendor software development
- Improved lifecycle management
- Increased testability
- Separation and collaboration

- OpenAMP Project (openampproject.org): standardizing software for heterogeneous environments
- OpenAMP: open source standard and API for asymmetric multiprocessing (AMP) systems
- OpenAMP library:
  - Lifecycle management: remoteproc
  - Messaging: RPMsg (Remote Processor Messaging) framework for inter-processor communication (IPC)
- System resource partitioning (System Device Tree)
- High level services (OpenAMP App Services WG)
  - Covers the application space that sits on top of OpenAMP and describes higher level services for: file sharing, debug proxying, high level IPC APIs, etc.

OpenAMP App Services: Isled

ISLED (v1)

- Daemon which provides access to non-real time services and drivers for Safety / Real-time OSs
- Custom RPC framework built on top of OpenAMP / RPMsg
- Features:
  - Client console
  - Host file system access
  - Port forwarding for TCF connectivity
- Proprietary code base previously en route to being open-sourced
- Closed source RTOS implementation

ISLED (v2)

- Daemon which provides access to non-real time services and drivers for Safety / Real-time OSs
- Standard communication infrastructure (virtio)
- Features:
  - Virtio console
  - 9p file system
  - Virtual sockets
- KVM Tool (lkvm) virtio back-end provider – open source

VIRTIO

Why use VIRTIO for intra-SoC workload integration?

- Open-source standard
- Enables decoupling the software from the hardware and supports modular system integration
- Solves the fragmentation problems caused by hypervisor specific interfaces and device drivers implementations
- Allows the creation of software defined architectures
- Increases efficiency (e.g., not using page cache for buffer reads and writes to files for block devices which are memory-like)
- Enables code reuse

Hypervisor-less VIRTIO

Define and prototype a framework for using VIRTIO as a communication infrastructure, while removing the constraints usually associated with the presence of a hypervisor.

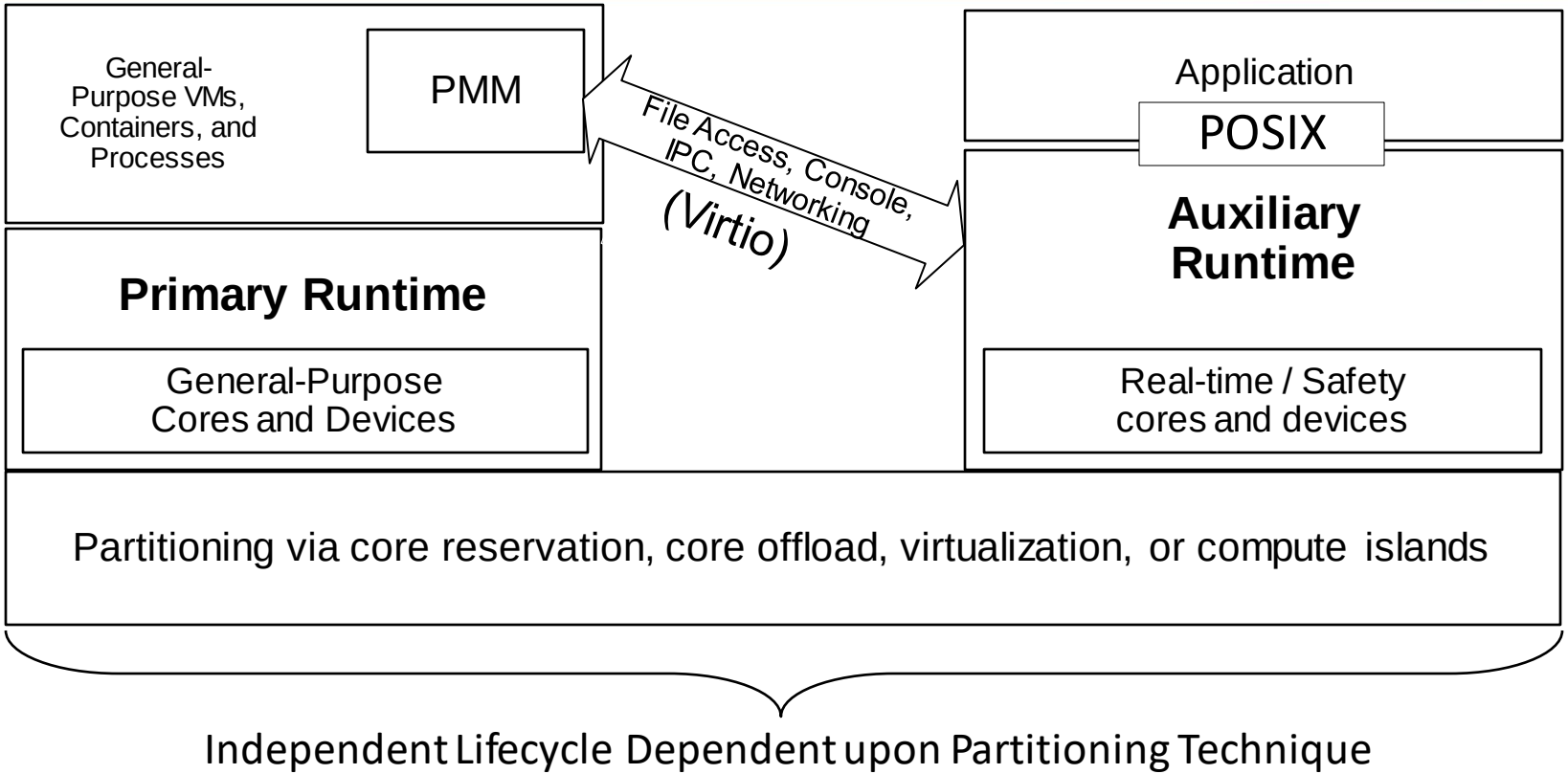
Hypervisor-less VIRTIO:

- 64-bit Intel x86_64 and ARM support
- MMIO transport over shared memory
- Unsupervised AMP support
- Static configuration (features, queues)
- Hardware notifications
- Linux KVM tool used as a Physical Machine Monitor (no longer a Virtual Machine Monitor)
- VIRTIO device support: console, 9P virtual file system, vsock, virtio-net, etc.

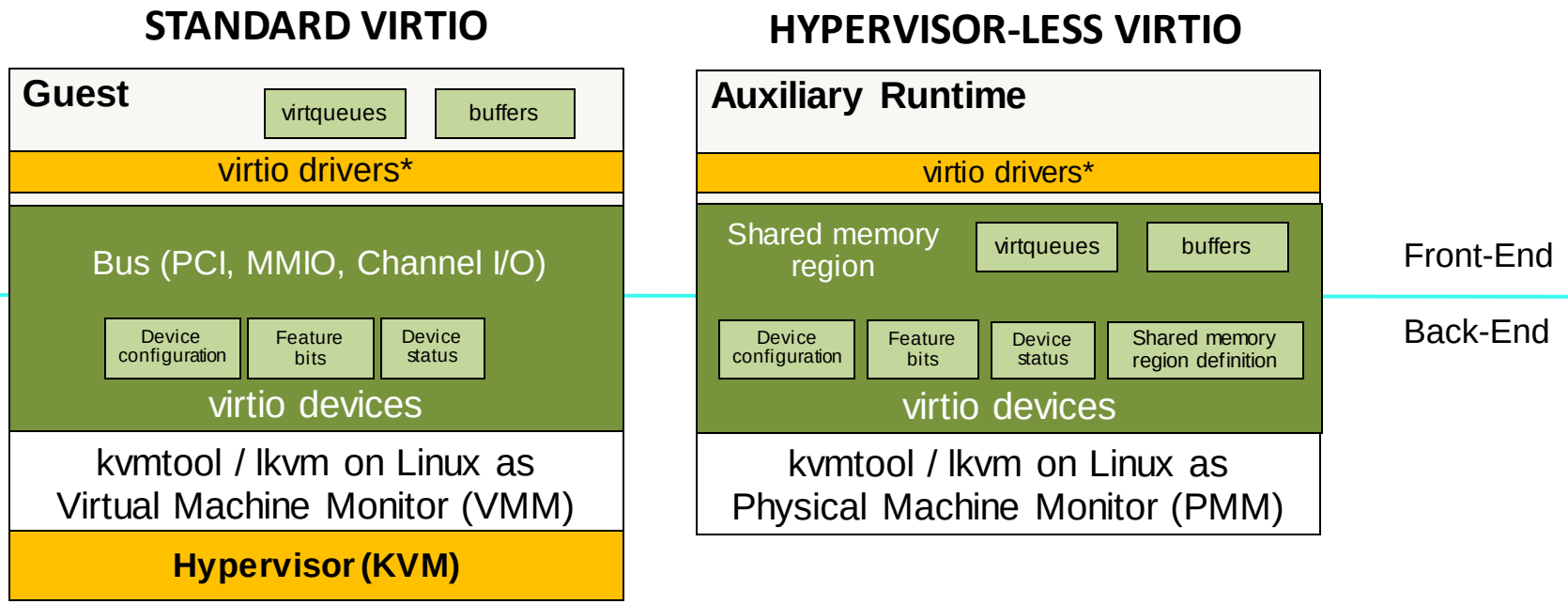
Hypervisor-less VIRTIO

- Commoditizes the creation and validation of real-time / security / safety islands by providing a common VIRTIO-based hardware abstraction layer
- Enables the use of multiple, dedicated, application-specific VIRTIO backends
- Introduces high-level APIs for OpenAMP applications

Hypervisor-less VIRTIO

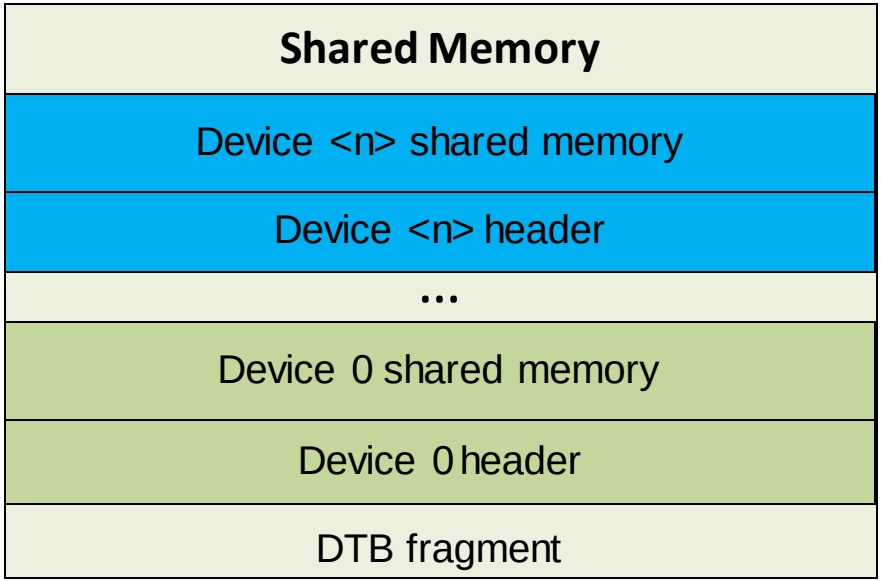


Similarities / differences



* File system (9P), Console (serial), Network (virtual ethernet), IPC (vsock)

Shared Memory Layout



Per-device shared memory	
VIRTIO console	24 KB
vsock	64 KB
9p	24 KB
VIRTIO net	64 KB

OpenAMP App Services

- Application space on top of OpenAMP
- Services:
  - High-level IPC APIs
  - File-sharing
  - Proxy and / or forwarding of IP ports
  - Debug proxy
  - Bare metal APIs using RPC
- Enable OS level APIs for high level applications (e.g., sockets, file systems)
- Use middleware and OS services of one runtime from another

Why Zephyr?

- Zephyr is great open-source development platform
- Vibrant and engaged community
- Small code size and footprint
- Full OpenAMP integration
- Zephyr originated at Wind River and my group was heavily involved in its creation

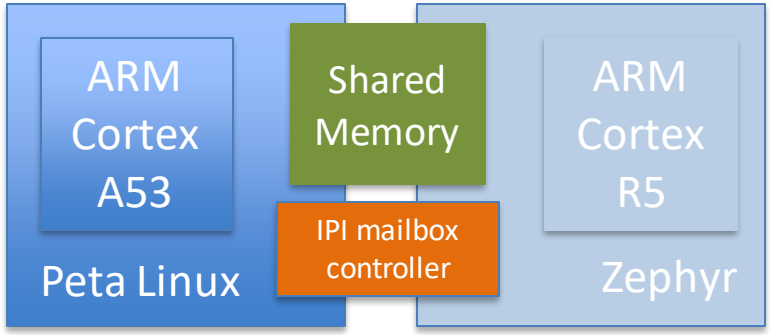
Progress

- Implemented Zephyr VIRTIO framework with support for network, console and entropy devices
  - qemu_cortex_a53
  - MMIO only
  - Zephyr devices are enumerated at compile time, so you can't discover and create devices on the fly
- Enabled hypervisor-less mode for VIRTIO network and entropy devices
- Sources available on GitHub, upstream efforts under discussion
- Build tools for reference hypervisor-less virtio deployment for Xilinx ZCU 102 (QEMU) available on GitHub

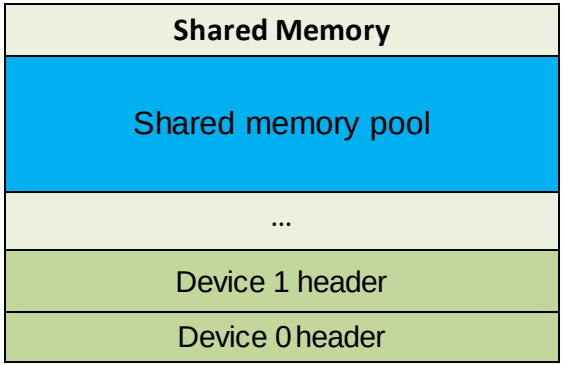
Hypervisor-less virtio

- Physical Machine Monitor (PMM): derived from kvmtool
  - Linux user-space application
  - Virtio network, virtio console, virtual sockets, entropy device, 9P file system
  - Small code base, easy to enhance
  - Notification proxying: vhost-net, vhost-vsock
- Vrings and data buffers need to be in the pre-shared memory region
  - Bounce buffers
  - Request buffers from pre-shared memory region

Implementation notes



Xilinx UltraScale+ MPSoC ZCU102



Pre-shared memory layout

Implementation notes

- Xilinx UltraScale+ MPSoC ZCU102, PetaLinux – Cortex A53, Zephyr – Cortex R5
- Pre-shared memory reserved in PetaLinux device tree, presented to Linux as UIO device
- Vrings in their own sections, placed by custom link script in pre-shared memory region
- The remainder of the shared memory is marked by variables placed by the linker script, used for bounce buffer allocation in hypervisor-less VIRTIO mode
- Xilinx IPI (Inter Processor Interrupt) mailbox controller
- Physical Machine Manager (PMM) on PetaLinux mmaps shared memory region, includes pseudo-state machine for handling VIRTIO requests

Call to action

Looking for developers interested in contributing code and users to provide critical feedback

References

- www.openampproject.org
- Zephyr VIRTIO framework:
  - https://github.com/danmilea/zephyr - branch main, with support for virtio network, console and entropy devices
- Open-source hypervisor-less virtio environment:
  - Physical Machine Monitor (PMM): https://github.com/OpenAMP/kvmtool
  - Hypervisor-less VIRTIO auxiliary runtime: https://github.com/danmilea/zephyr (branch hvl-virtio-0)
  - Build environment: https://github.com/danmilea/hypervisorless_virtio_zcu102/

Zephyr® Project

Developer Summit 2022

June 8-9, 2022

Mountain View, CA + Virtual