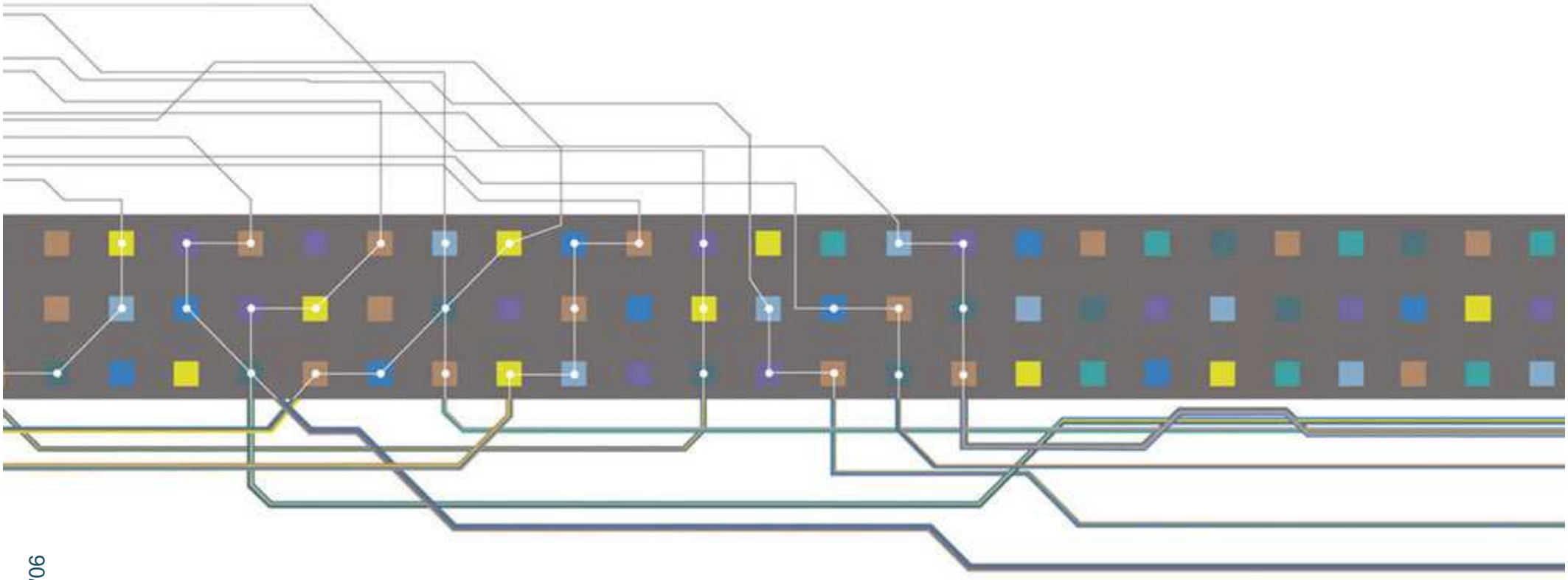
THALES



RTOS and HdS for MPSOC
A Software Infrastructure

August 14-19, 2006 - MPSoC'06

Philippe KAJFASZ

Land & Joint Systems

OUTLINE

- RTOS on MPSoC:
  - How to achieve Real-Time on SMP systems ?
  - MPSoC with shared memory: What can we do ?
  - MPSoC RTOS requirements & cartography of constraints
  - Our experience with SMP RTOS
- Software Architecture:
  - GPOS capabilities
  - Extending capabilities for Real Time
- RTOS & HdS Power Management issues
  - Energy in Battery Powered Systems
  - Energy Saving Techniques
  - Power Optimisation Strategies
  - Framework Infrastructure
- Conclusion

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

RTOS on MPSoC: how to achieve ?

Two approaches, conditioned by MPSoC architecture:

- Distributed memory: not much choice
  - One RTOS for each CPU, communicating with IPC
- Shared memory: some choice
  - One RTOS for each CPU, communicating with IPC
  - One RTOS for all CPU, with SMP capabilities
  - Or a combination of the above
    - Requires memory partitioning capabilities

Comments

- Mix of HPC-based and SMP-based systems on a SoC
  - This SoC is indeed a special kind of blade server (functionality)
- How is RT ensured
  - HPC: beyond one CPU, apply the “do it yourself” methodology
  - SMP: rely on memory access controller performance and cache policies, then on the rare SMP RTOS available.

  None is fully satisfactory.

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Real-Time on SMP systems

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Motivations

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Motivations

- Ability to develop and port **complex** applications
  - Portability: Standard POSIX API
  - Completeness: network stacks, Intellectual Property drivers
  - Versatility: tailored to application needs
  - Simplicity: easy to learn programming model, high level languages and tools

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Motivations

- Ability to develop and port **complex** applications
  - Portability: Standard POSIX API
  - Completeness: network stacks, Intellectual Property drivers
  - Versatility: tailored to application needs
  - Simplicity: easy to learn programming model, high level languages and tools
- While providing support for **determinism**
  - A common constraint in embedded systems
  - Mandatory for strongly reactive systems (Synthetic environments, Software Radio systems, …)
  - Needed for timing measurements, thread and process synchronization, signaling infrastructure, I/O operations.

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

MPSoC with shared memory: What can we do ?

Find a SMP RTOS for MPSoC

- Not so many SMP RTOS on the market (market is quite small presently).

RTOS first, then SMP

- Being able to extend an RTOS.
  - Benefits: all application elements are RT, small footprint
  - Drawbacks: proprietary API, over-specification for non-RT parts.

SMP first, then RTOS

- Being able to extend a SMP GPOS
  - Benefits: complexity handling, standard API, no over-specification
  - Drawbacks: no hard RT features, large footprint

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

MPSoC RTOS requirements at a glance

Development made easy …

- As many capabilities as a host system

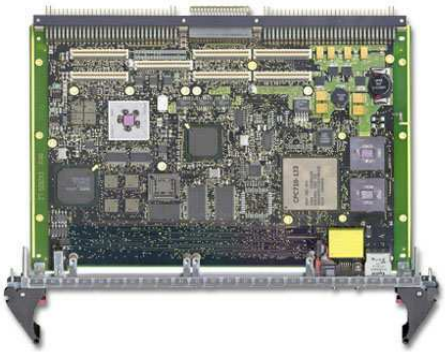
… while performance level is controlled

- Perform efficient, application driven computations
  - added flexibility in co-existing SMP and HPC models
- Optimize software infrastructure footprint
- Manage power consumption
- … and many others …

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Cartography of constraints

Supercomputers



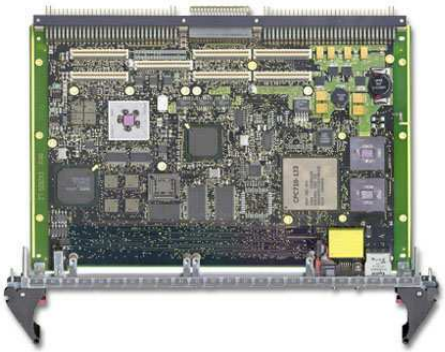
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Cartography of constraints

Supercomputers

HPC constraints:

- Multiple users
- High level programming capabilities
- System partitioning
- Price/performance ratio
- High I/O throughput
- High availability
- no RT considerations
  except for some embedded systems



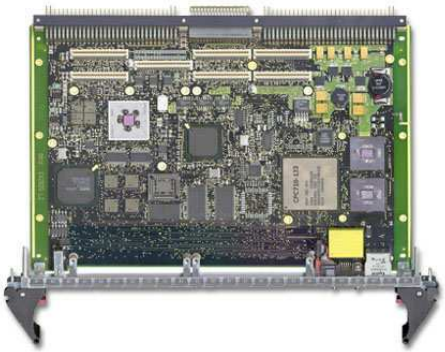
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Cartography of constraints

HPC constraints:

- Multiple users
- High level programming capabilities
- System partitioning
- Price/performance ratio
- High I/O throughput
- High availability
- no RT considerations
  except for some embedded systems

Supercomputers



Embedded constraints:

- Remote operation
- Failure detection
- Configuration reporting
- Resource usage
- strong RT considerations

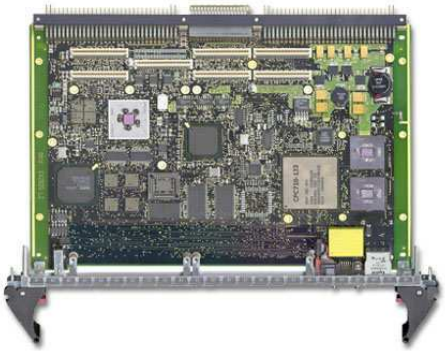
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Cartography of constraints

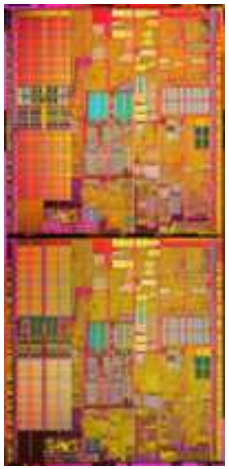
HPC constraints:

- Multiple users
- High level programming capabilities
- System partitioning
- Price/performance ratio
- High I/O throughput
- High availability
- no RT considerations
  except for some embedded systems

Supercomputers



MPSoC



Embedded constraints:

- Remote operation
- Failure detection
- Configuration reporting
- Resource usage
- strong RT considerations

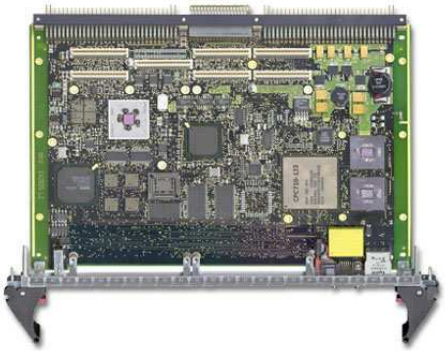
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Cartography of constraints

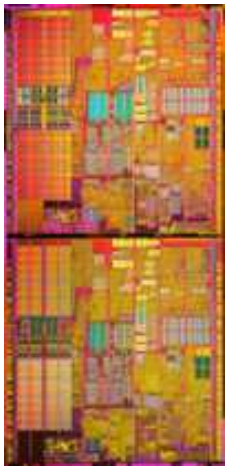
HPC constraints:

- Multiple users
- High level programming capabilities
- System partitioning
- Price/performance ratio
- High I/O throughput
- High availability
- no RT considerations
  except for some embedded systems

Supercomputers



MPSoC



Embedded constraints:

- Remote operation
- Failure detection
- Configuration reporting
- Resource usage
- strong RT considerations

On Chip constraints:

- Little access to system components
- Flexible resource sharing

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Our experience with SMP RTOS

SMP GPOS with RTOS features

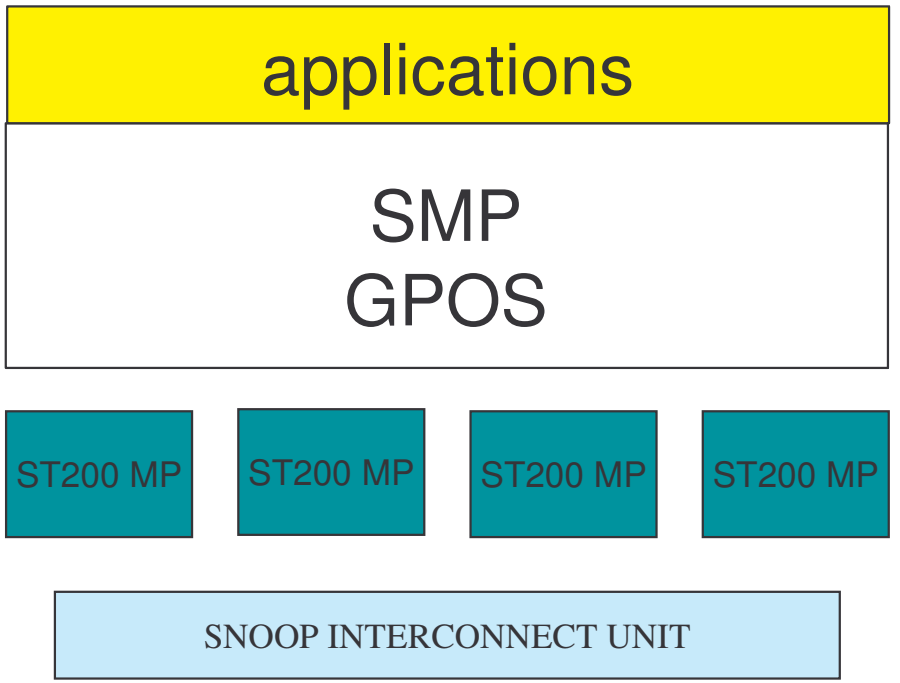
- On supercomputer architectures: up and running
  - 4-way Intel Itanium 2
  - 2-way Intel Xeon
- On embedded systems: under development
  - Power PC-based SBC (82xx, 83xx)

Software components

- “pre-emptible” Linux kernel
- Adeos interrupt pipeline layer
- Real-time domain: DIC (Deterministic Intensive Computing)

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: GPOS capabilities

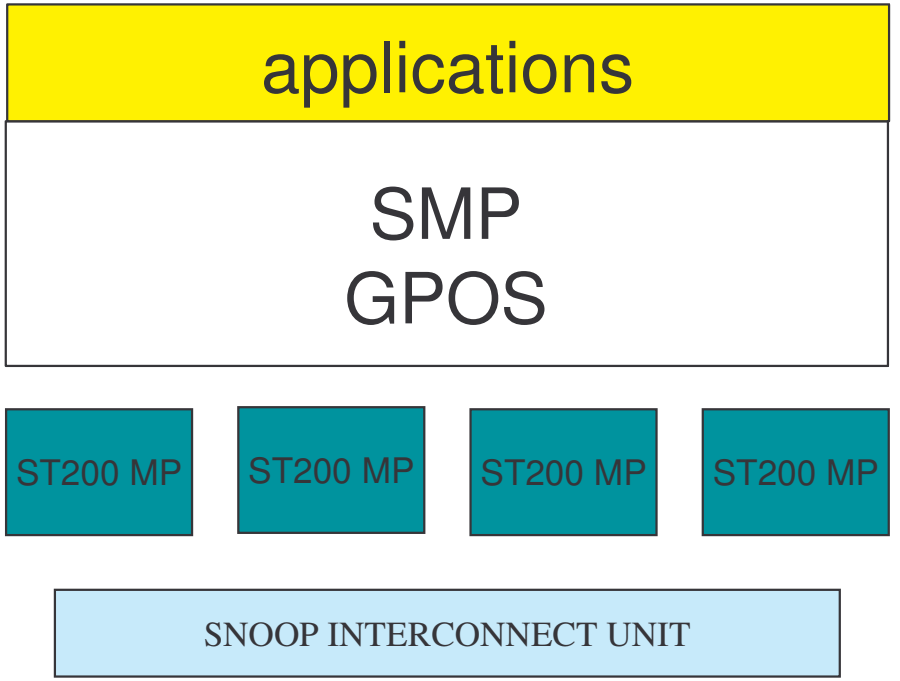


This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: GPOS capabilities

Single System Image

- SMP-capable GPOS



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

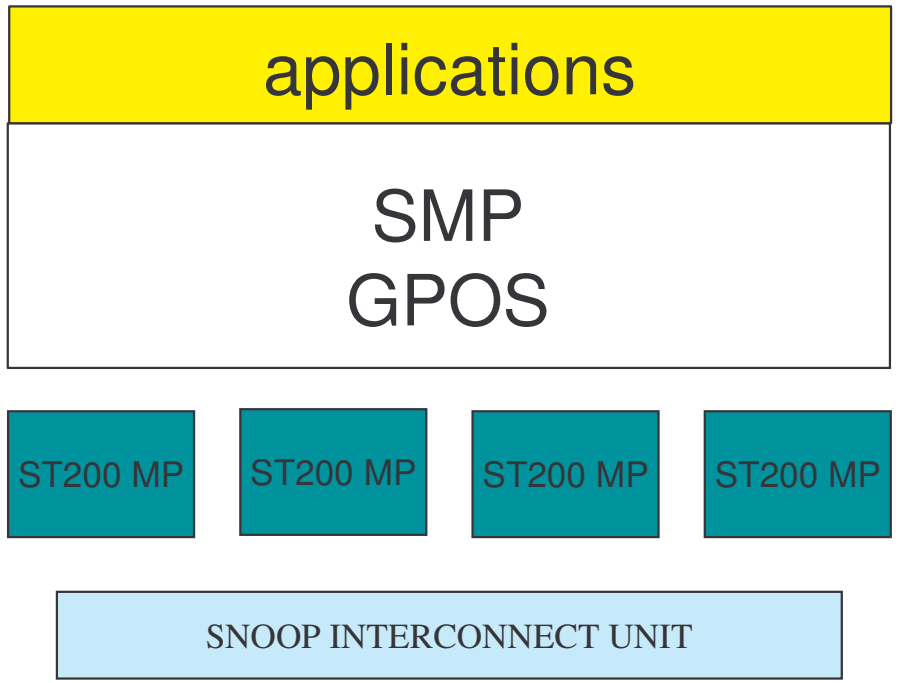
Software Architecture: GPOS capabilities

Single System Image

- SMP-capable GPOS

POSIX compliance

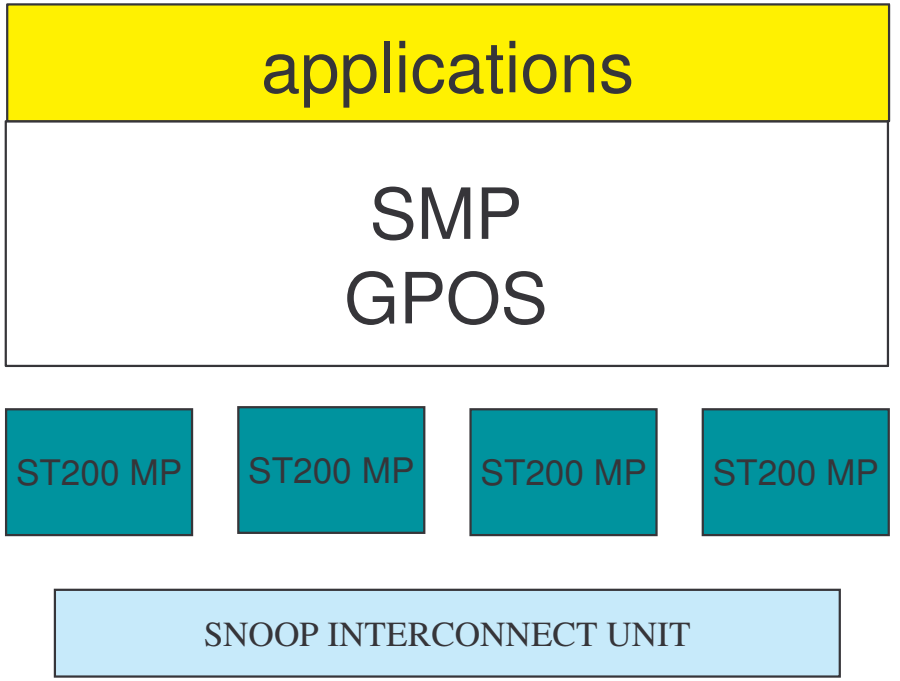
- multithreading
- multi processing
- shared data and semaphores



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: GPOS capabilities

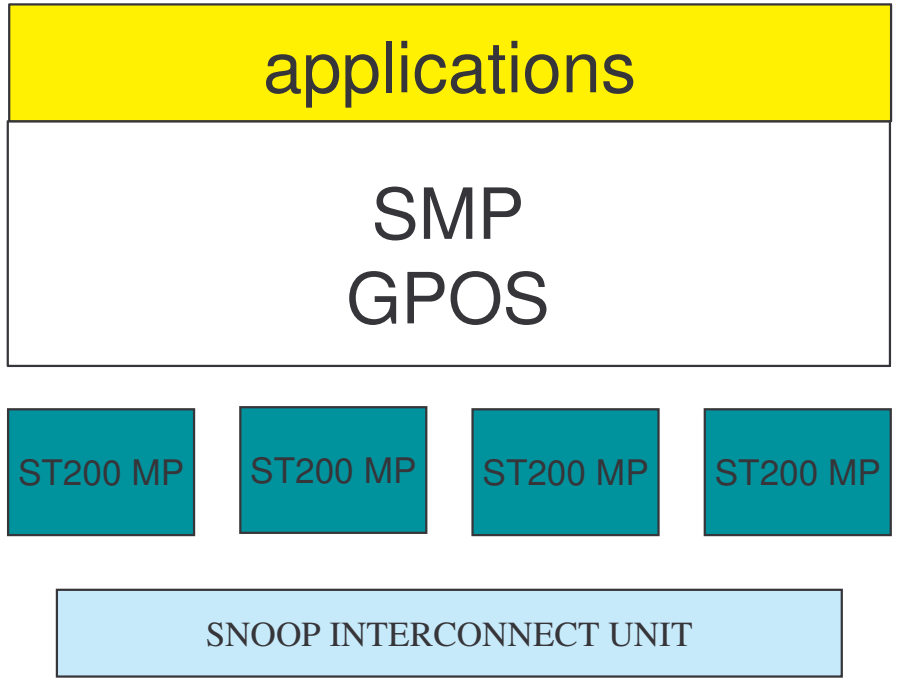
- Single System Image
  - SMP-capable GPOS
- POSIX compliance
  - multithreading
  - multi processing
  - shared data and semaphores
- Load-balancing capabilities
  - customizable



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: GPOS capabilities

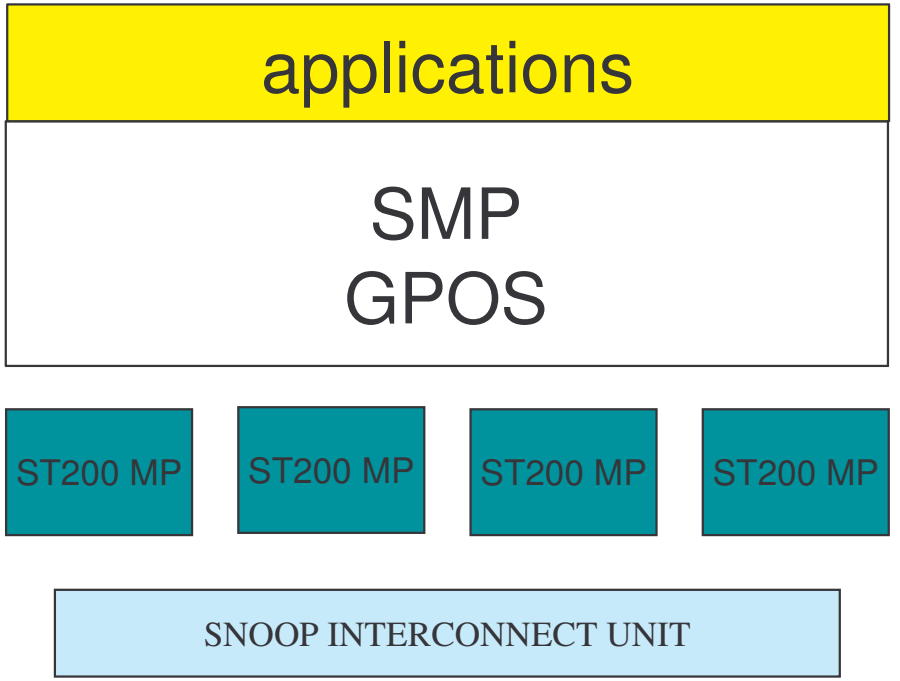
- Single System Image
  - SMP-capable GPOS
- POSIX compliance
  - multithreading
  - multi processing
  - shared data and semaphores
- Load-balancing capabilities
  - customizable
- Processor affinity
  - process groups support



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: GPOS capabilities

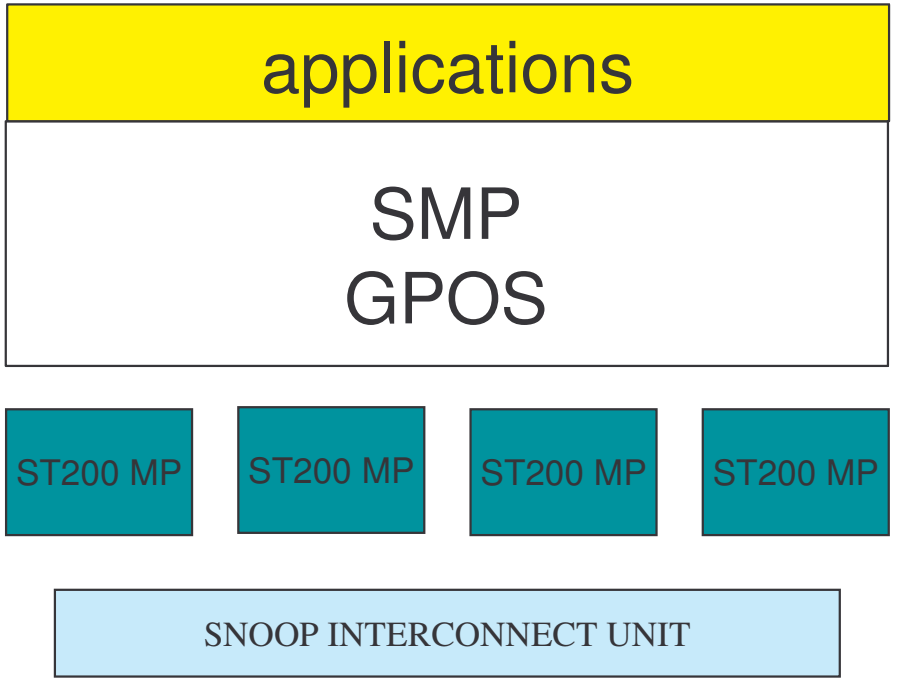
- Single System Image
  - SMP-capable GPOS
- POSIX compliance
  - multithreading
  - multi processing
  - shared data and semaphores
- Load-balancing capabilities
  - customizable
- Processor affinity
  - process groups support
- Variable-priority scheduling



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: GPOS capabilities

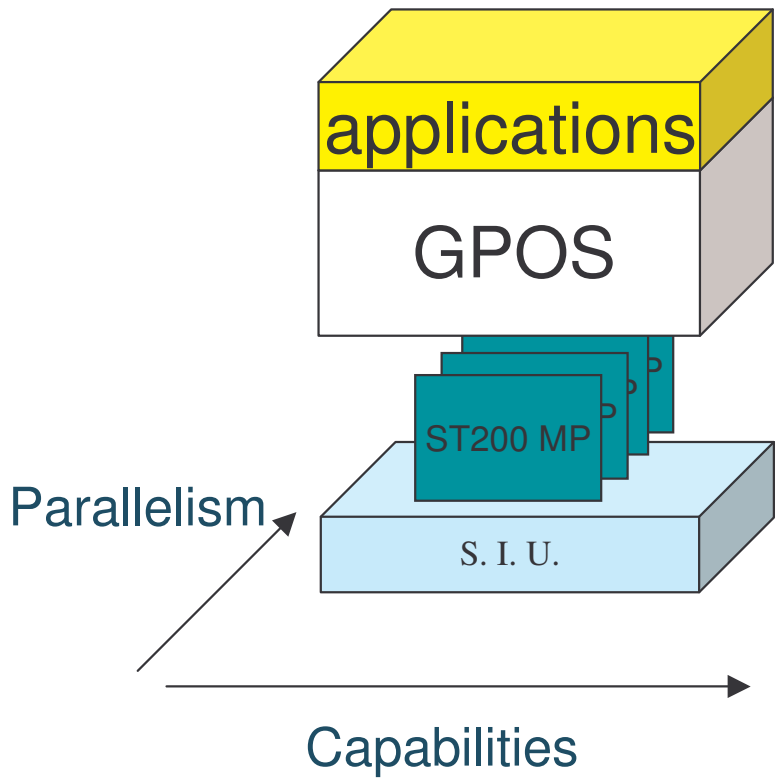
- Single System Image
  - SMP-capable GPOS
- POSIX compliance
  - multithreading
  - multi processing
  - shared data and semaphores
- Load-balancing capabilities
  - customizable
- Processor affinity
  - process groups support
- Variable-priority scheduling



Optimizes global system throughput and CPU usage efficiency

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: extending capabilities

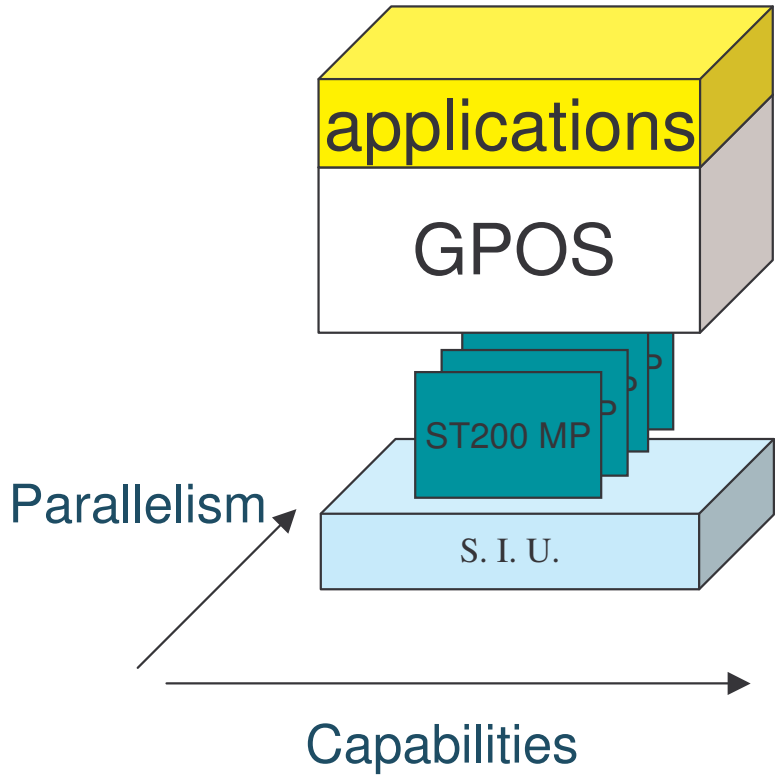


This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: extending capabilities

Another view of the same system

- 3D vs. 2D



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

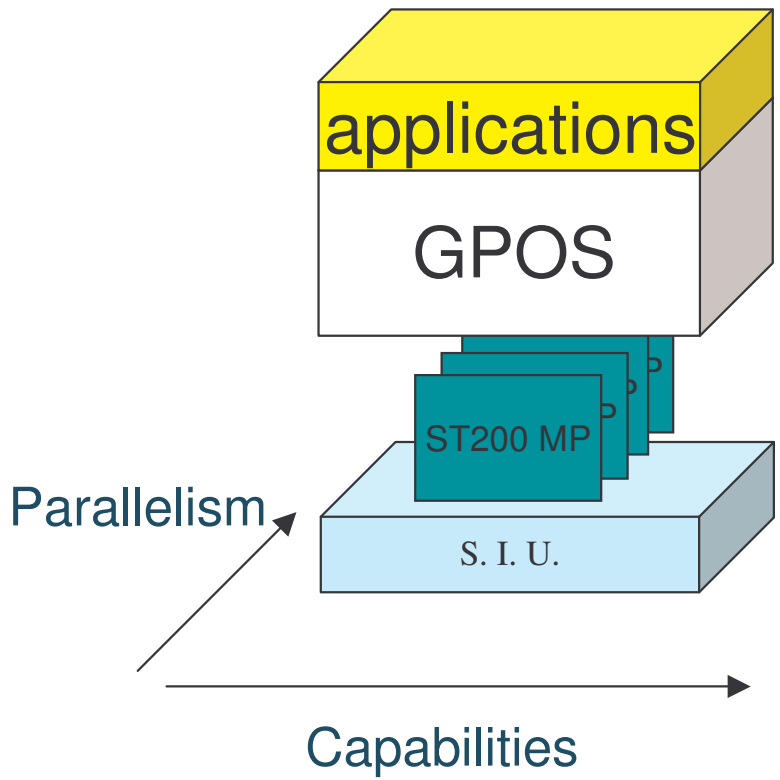
Software Architecture: extending capabilities

Another view of the same system

- 3D vs. 2D

GPOS on multi-processor:

- SMP: automated parallelism of execution …
- … but limited control by default



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Software Architecture: extending capabilities

Another view of the same system

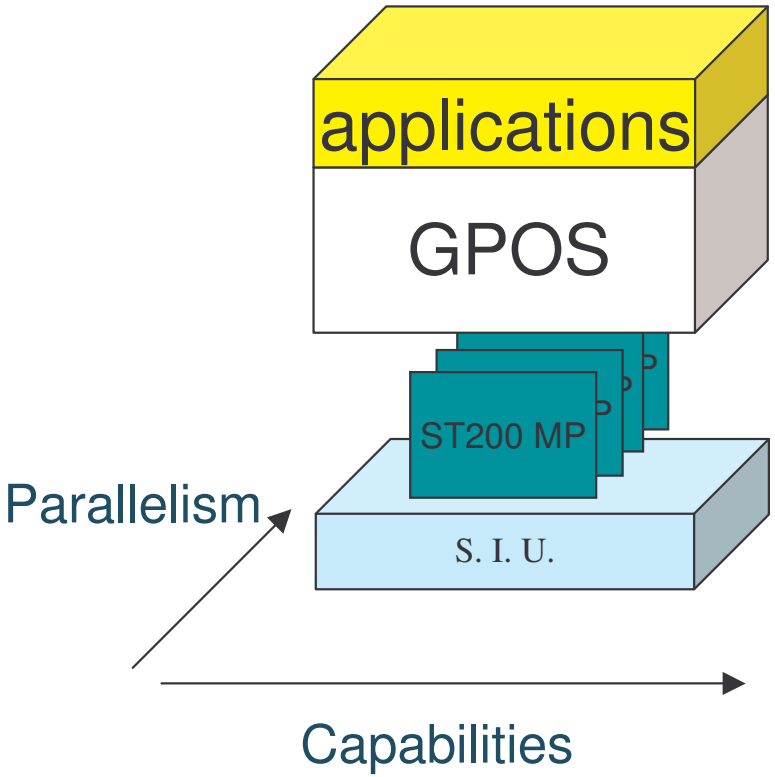
- 3D vs. 2D

GPOS on multi-processor:

- SMP: automated parallelism of execution …
- … but limited control by default

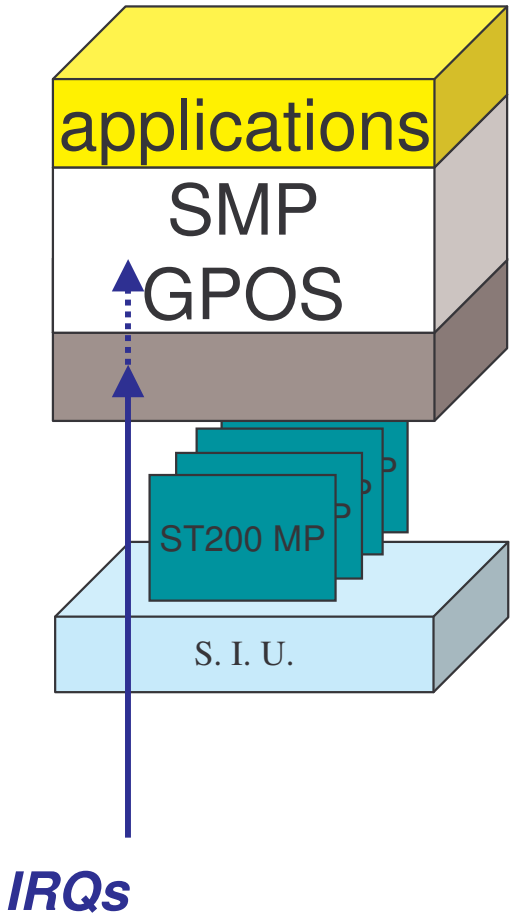
RTOS capabilities: adding determinism

- control process migrations
  - limit cache coherency operations
- control interrupt dispatching
- control resource sharing
  - RTOS tasks first
  - GPOS tasks next



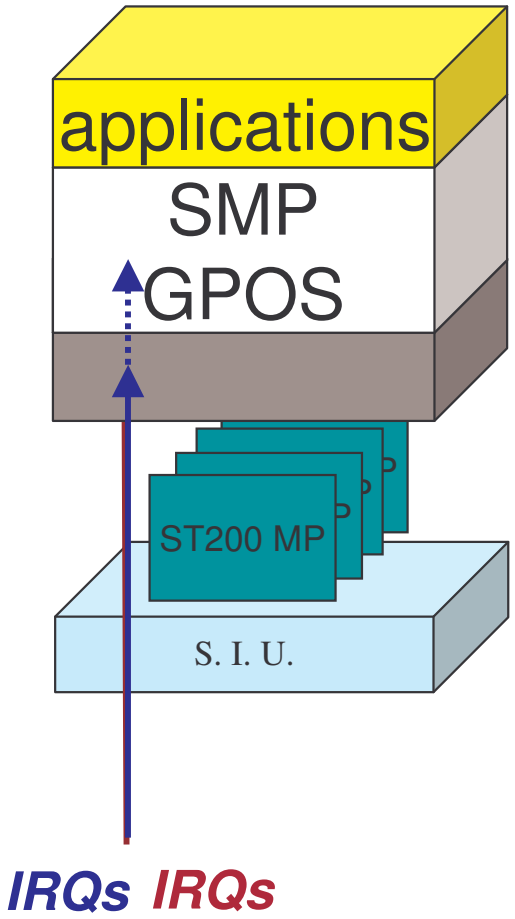
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adding RTOS features: prerequisites



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

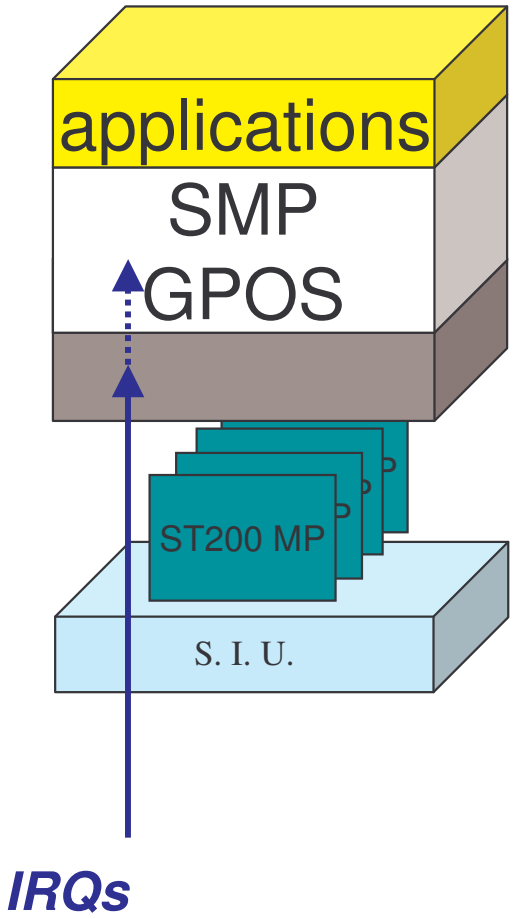
Adding RTOS features: prerequisites



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adding RTOS features: prerequisites

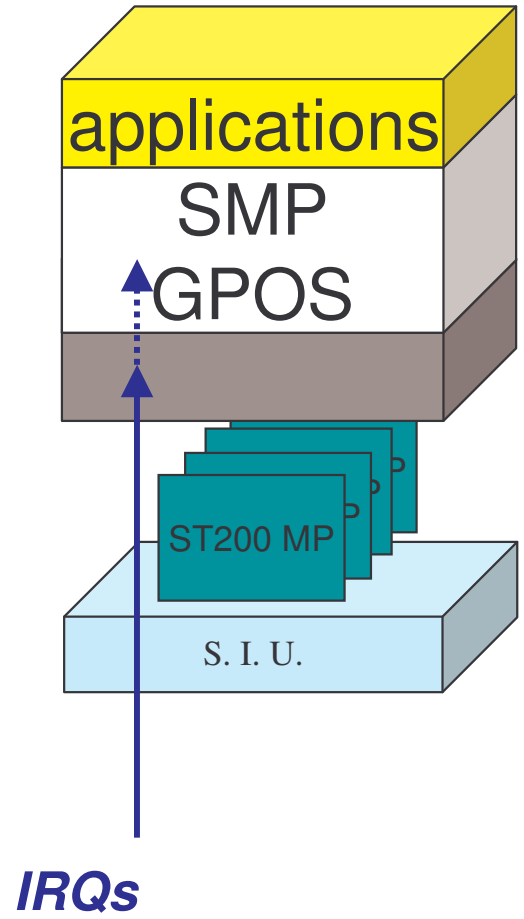
Interrupt virtualization



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adding RTOS features: prerequisites

Interrupt virtualization

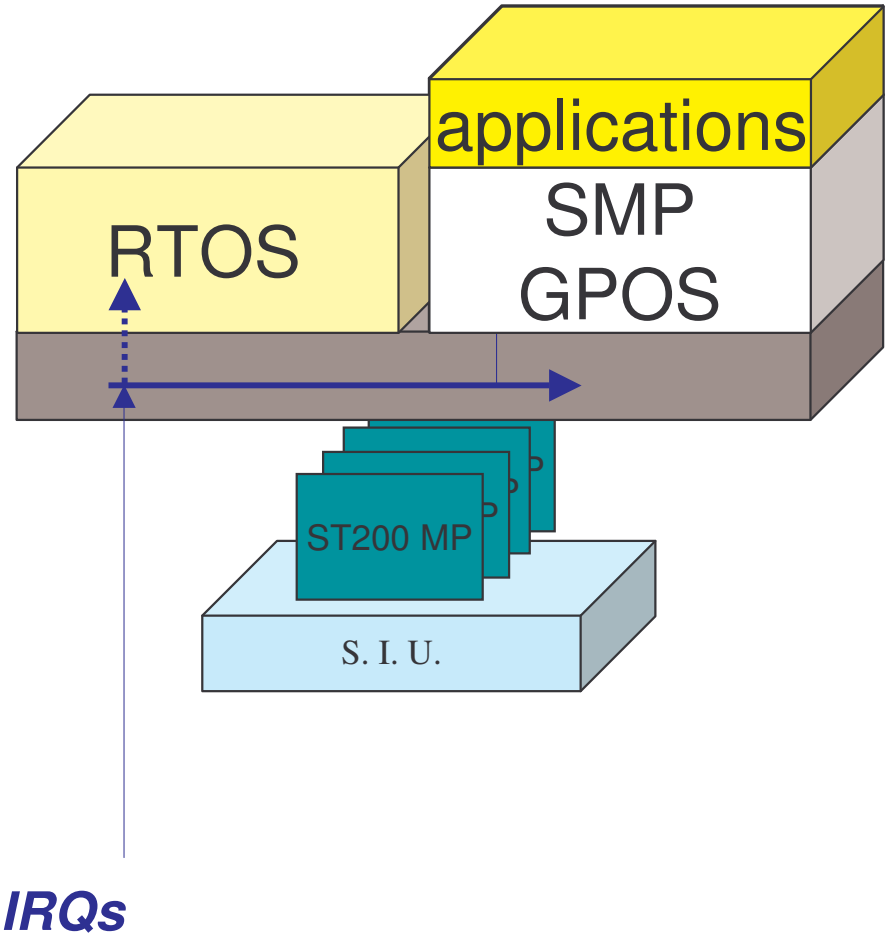


Provision for managing interrupt priorities – no RTOS features yet

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adding RTOS domain

Interrupt virtualization

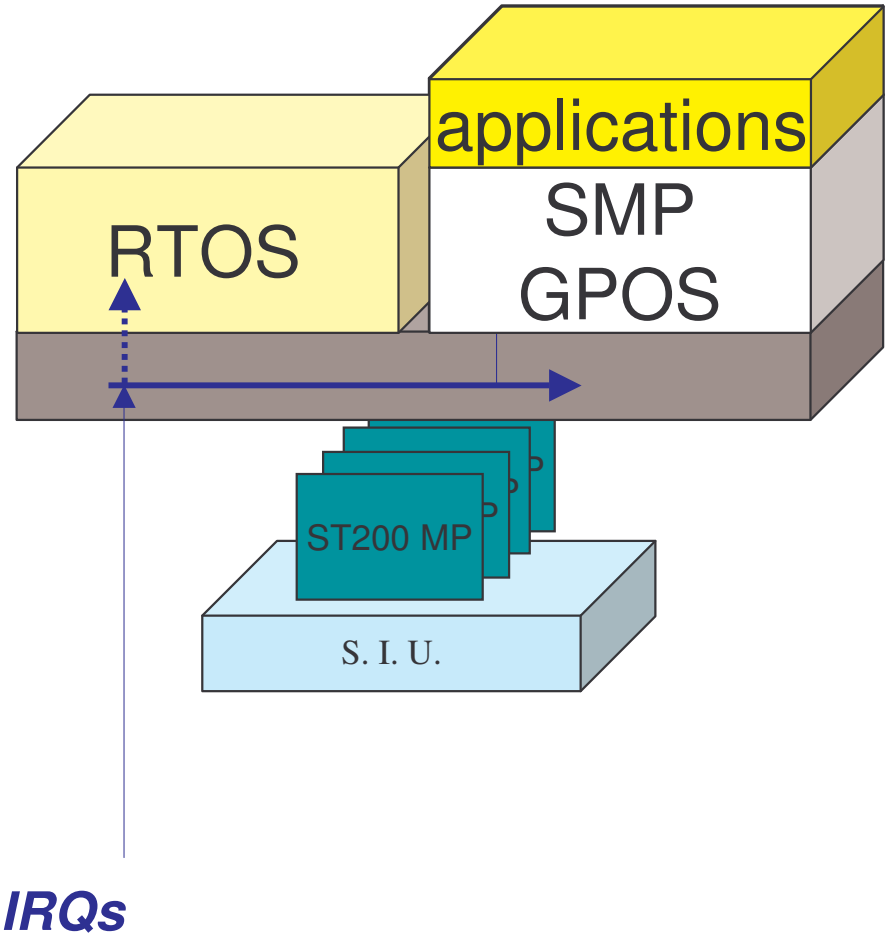


This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adding RTOS domain

Interrupt virtualization

Interrupt pipelining



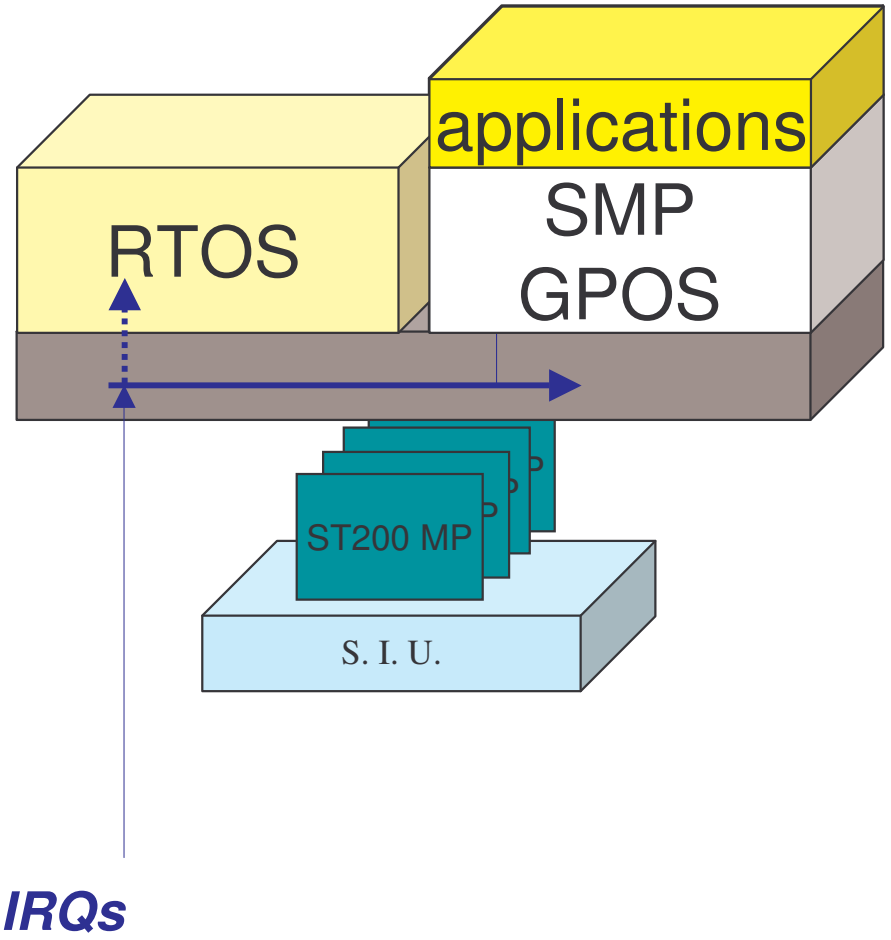
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adding RTOS domain

Interrupt virtualization

Interrupt pipelining

RTOS execution domain



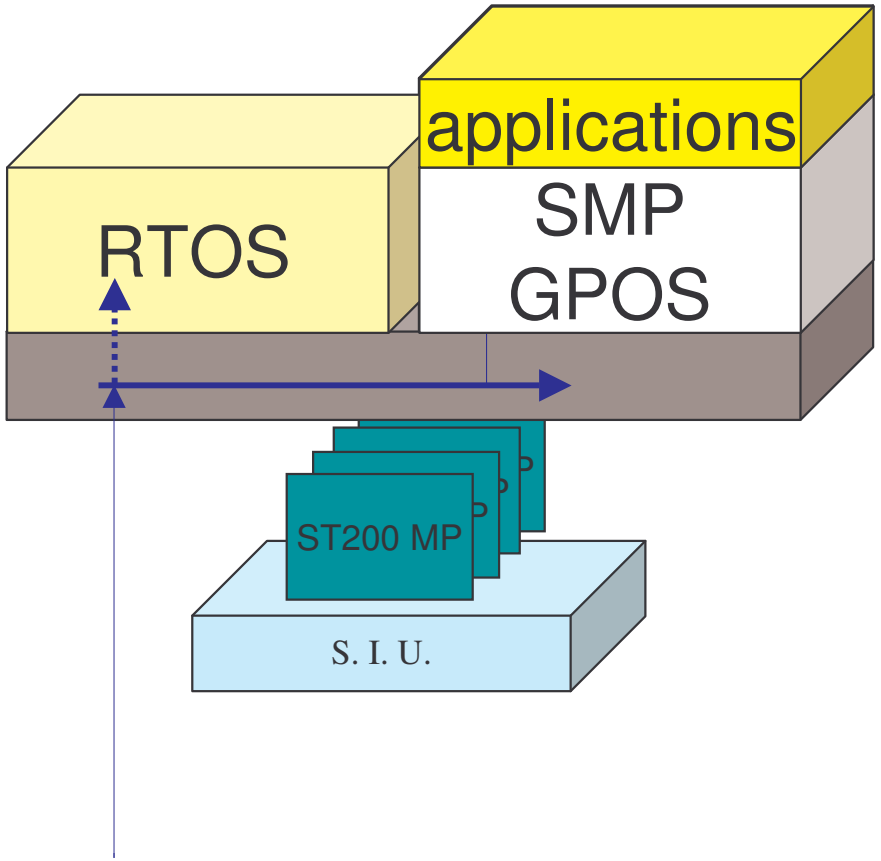
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adding RTOS domain

Interrupt virtualization

Interrupt pipelining

RTOS execution domain



RTOS features added – BUT:

processes are exclusively RTOS or GPOS, no migration
applications processes do not have direct access to RTOS

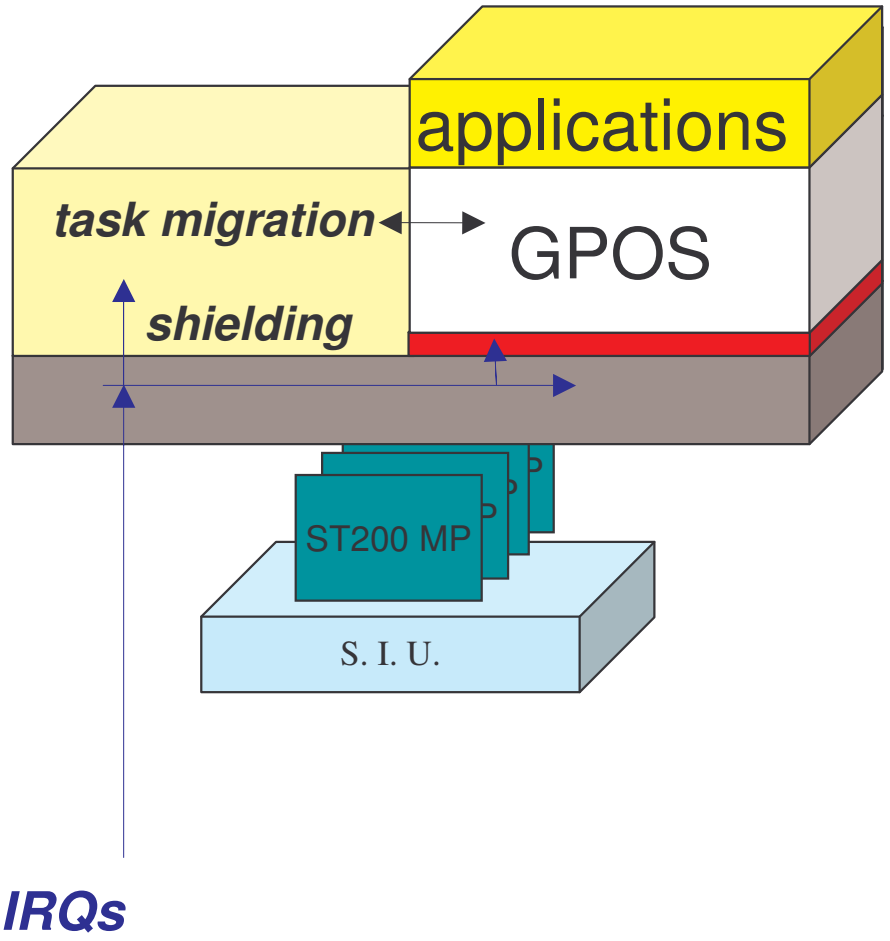
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

RTOS task migration to GPOS

Interrupt virtualization

Interrupt pipelining

RTOS execution domain



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

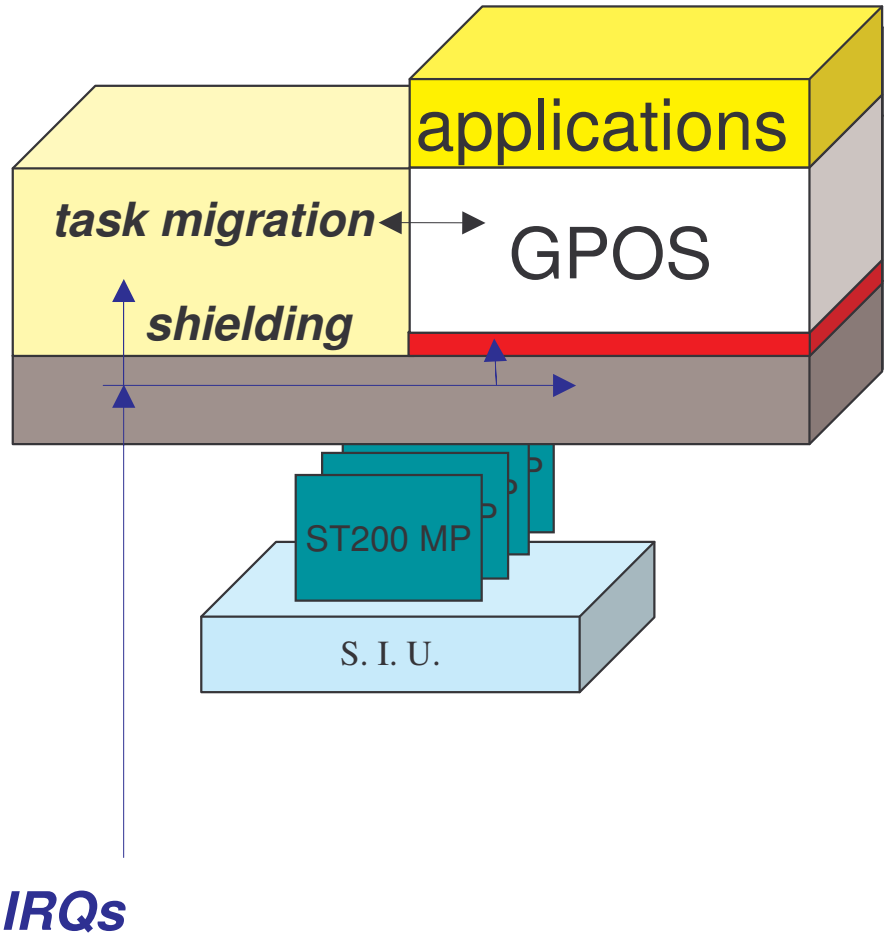
RTOS task migration to GPOS

Interrupt virtualization

Interrupt pipelining

RTOS execution domain

Interrupt shielding



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

RTOS task migration to GPOS

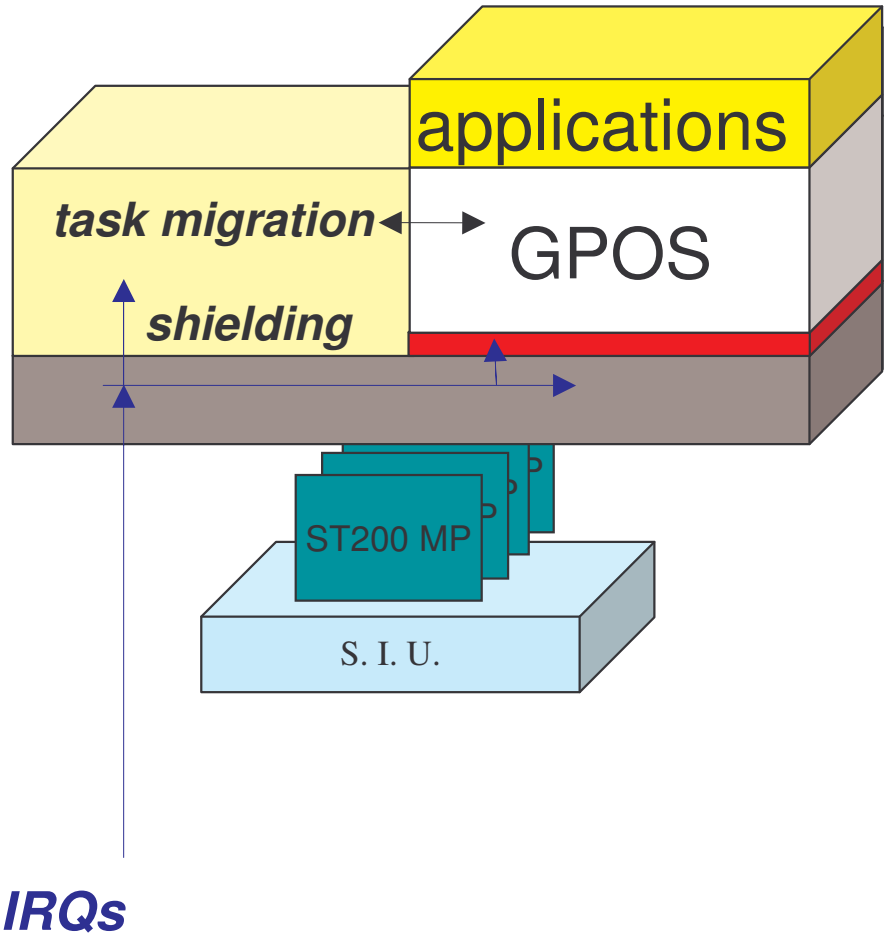
Interrupt virtualization

Interrupt pipelining

RTOS execution domain

Interrupt shielding

Task migration in kernel



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

RTOS task migration to GPOS

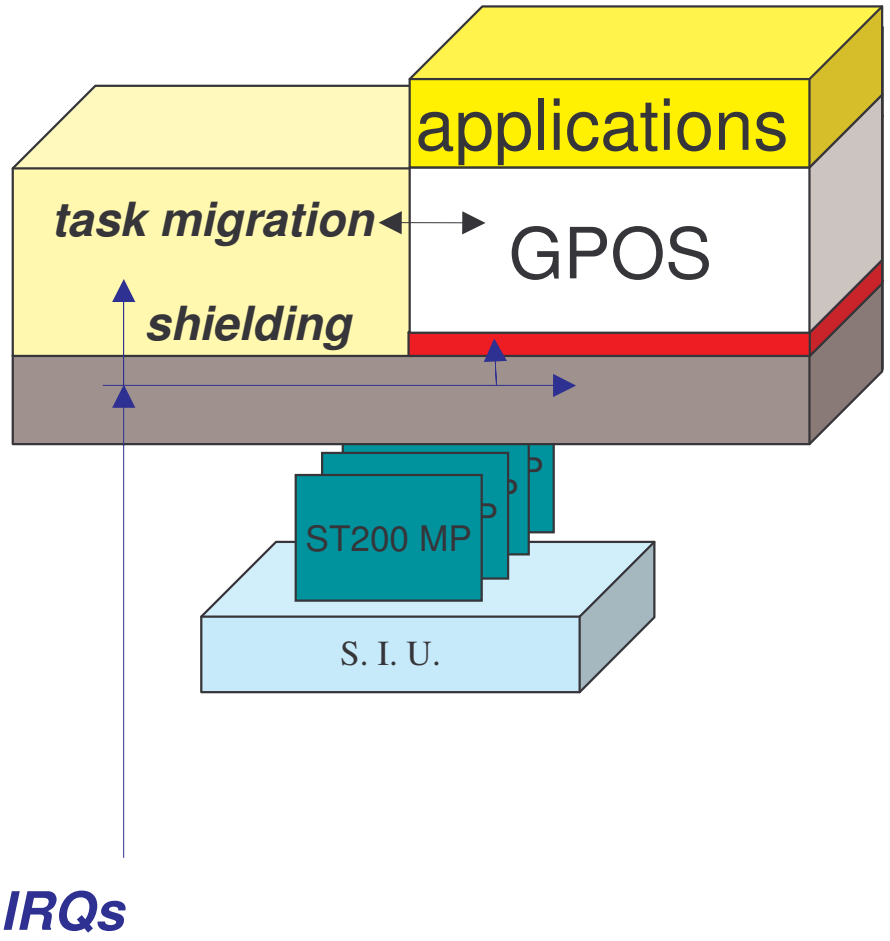
Interrupt virtualization

Interrupt pipelining

RTOS execution domain

Interrupt shielding

Task migration in kernel



benefits :

RTOS tasks first (shielding)

RTOS tasks can use GPOS services (triggers migration)

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Using RTOS features from applications

Interrupt virtualization

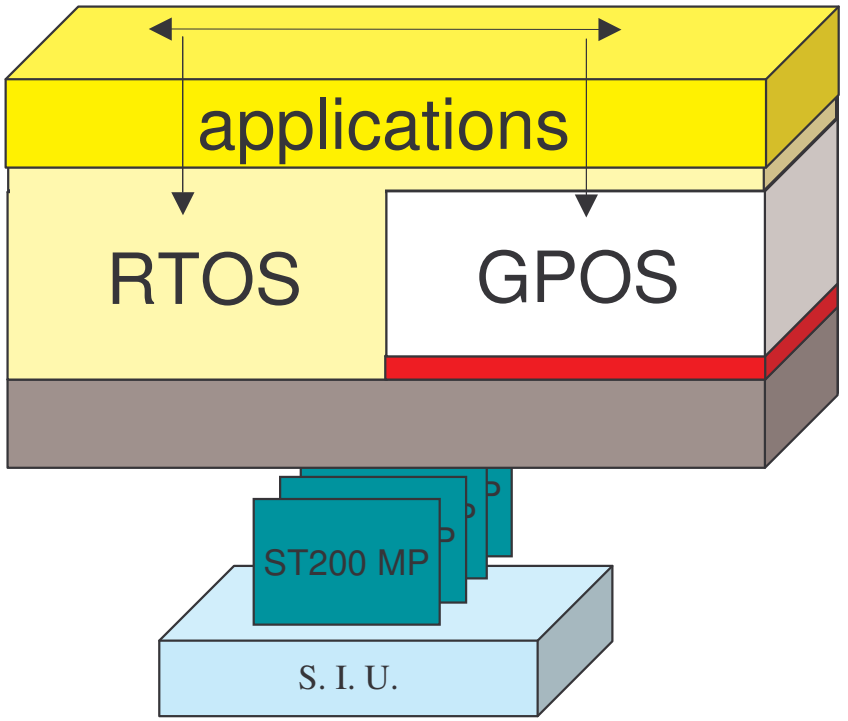
Interrupt pipelining

RTOS execution domain

Interrupt shielding

Task migration in kernel

task migration



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Using RTOS features from applications

Interrupt virtualization

Interrupt pipelining

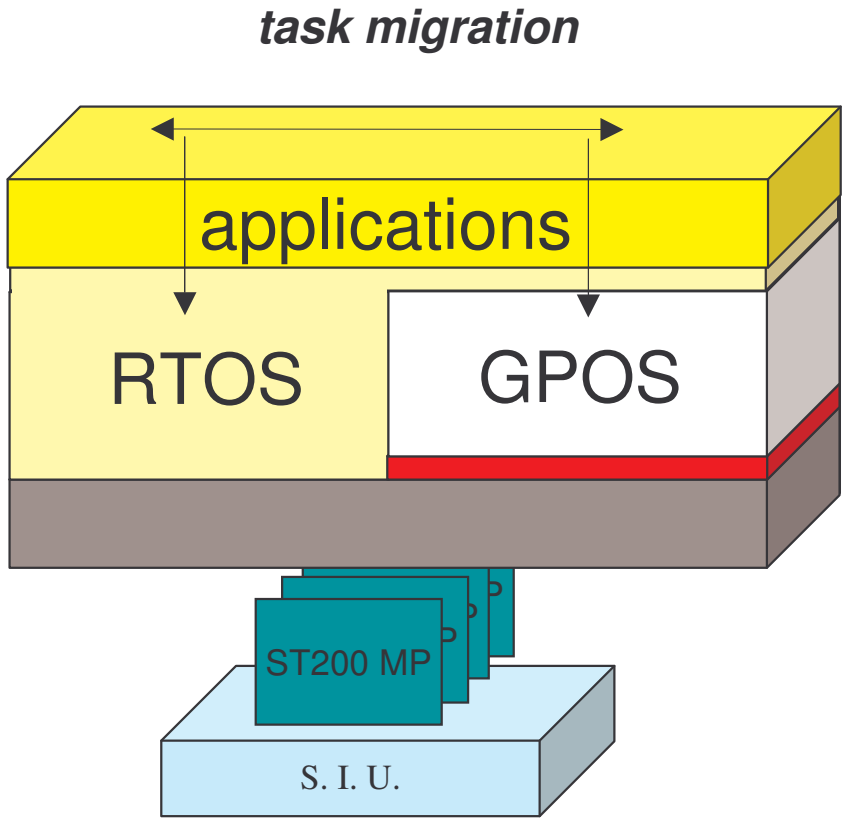
RTOS execution domain

Interrupt shielding

Task migration in kernel

RT API addition

- Application tasks can migrate seamlessly from RT domain to GP domain



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Using RTOS features from applications

Interrupt virtualization

Interrupt pipelining

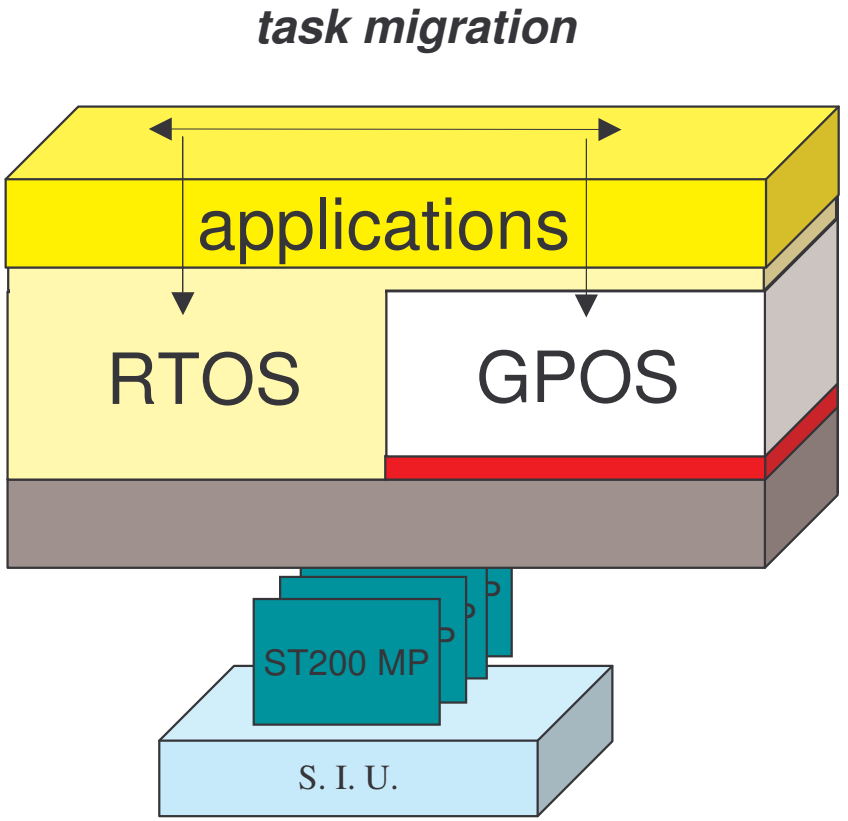
RTOS execution domain

Interrupt shielding

Task migration in kernel

RT API addition

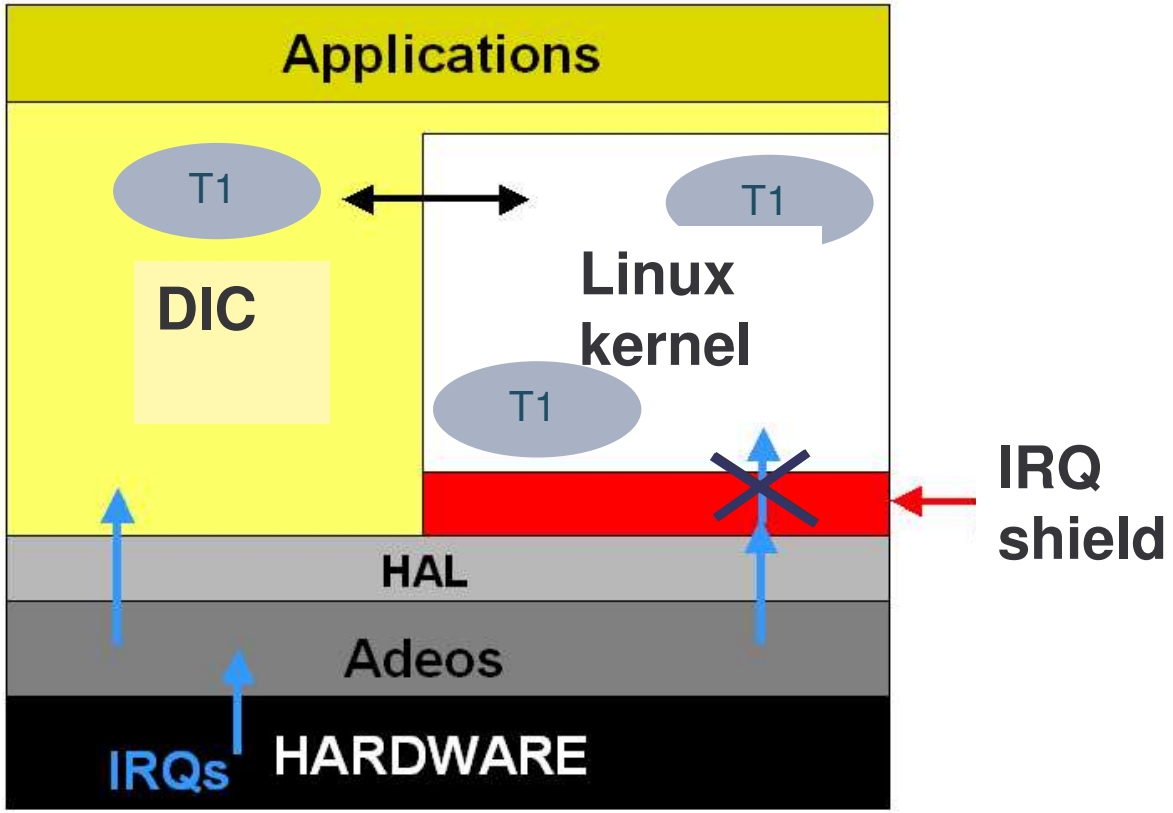
- Application tasks can migrate seamlessly from RT domain to GP domain



applications have direct access to RTOS (e.g. POSIX RT API) and unmodified access to GPOS

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Adeos/DIC thread and IRQ management

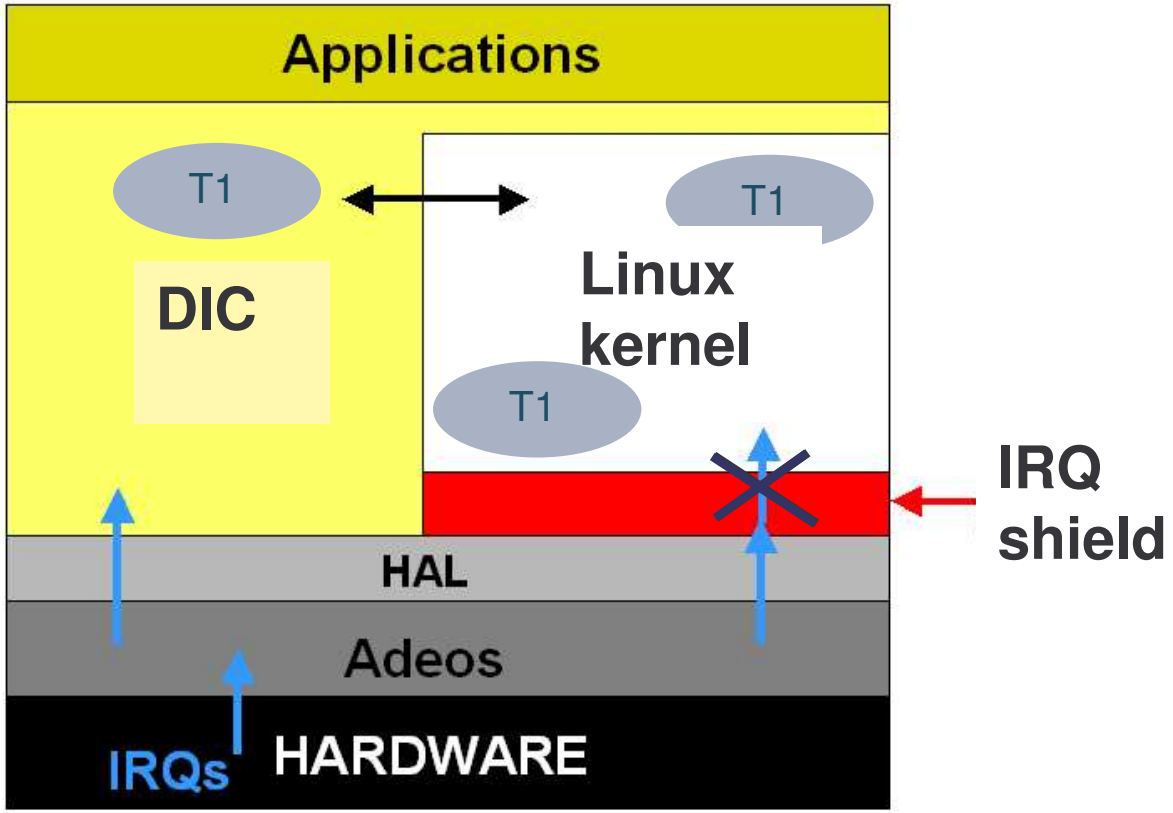


© Thales

This doc

Adeos/DIC thread and IRQ management

- Linux thread T1 is running



: Thales

This doc

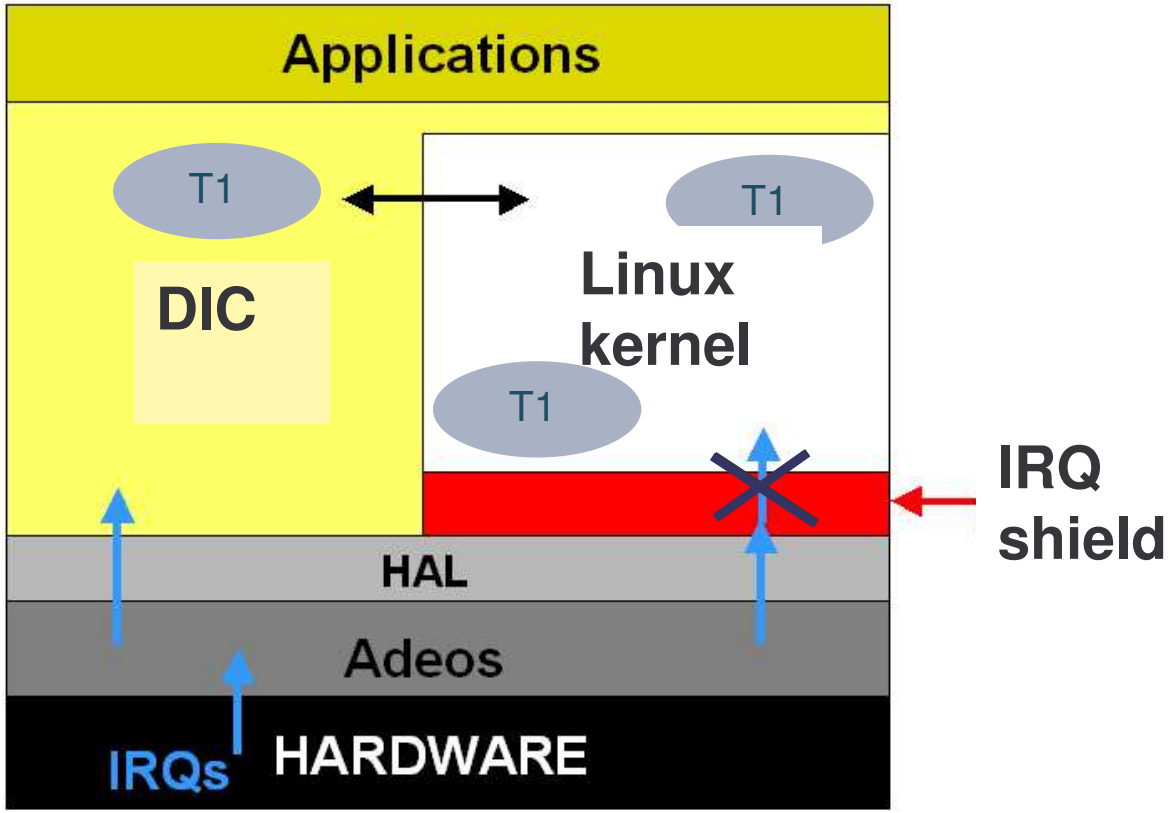
Adeos/DIC thread and IRQ management

- Linux thread T1 is running
- Calls pthread_init_rt()

⇒ Passes under DIC control

⇒ Activates the interrupt shield

PRIMARY MODE



Thales

This doc

Adeos/DIC thread and IRQ management

- Linux thread T1 is running
- Calls pthread_init_rt()

⇒ Passes under DIC control

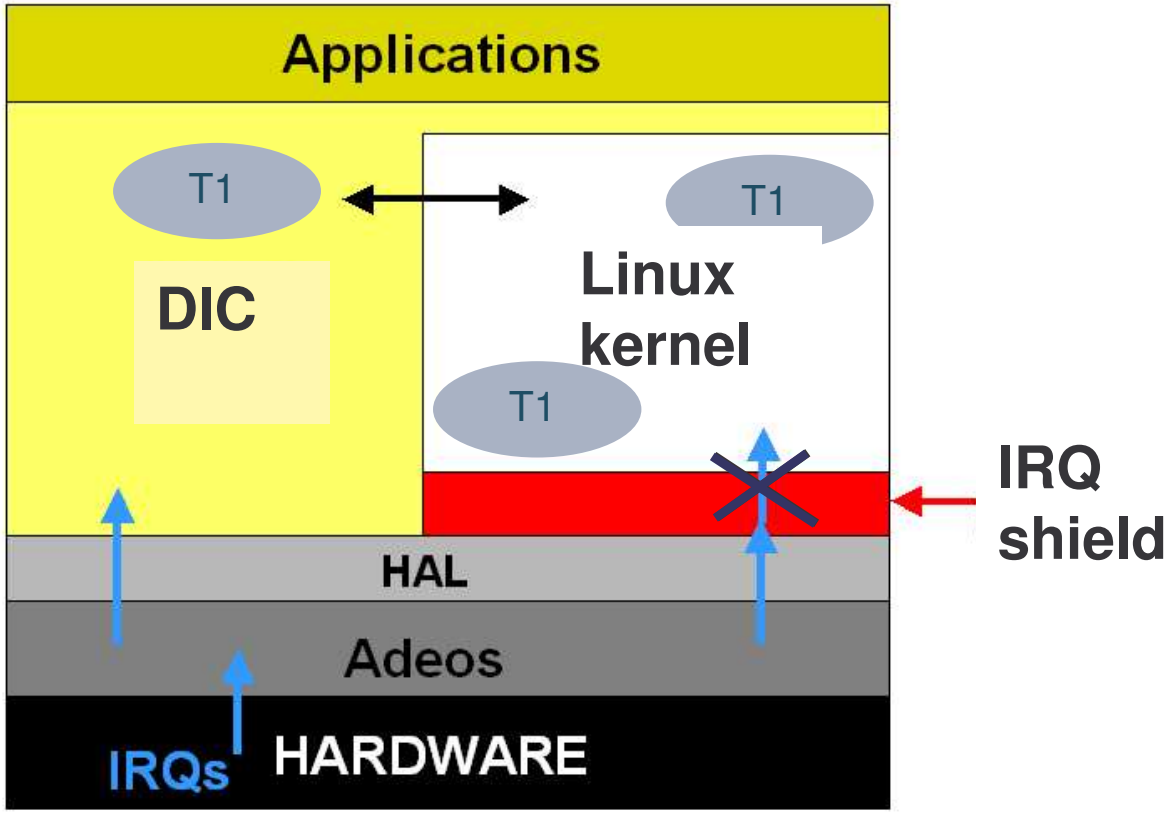
⇒ Activates the interrupt shield

PRIMARY MODE

- Issues Linux system call

⇒ Interrupt shield remains on

SECONDARY MODE



Thales

This doc

Adeos/DIC thread and IRQ management

- Linux thread T1 is running
- Calls pthread_init_rt()

⇒ Passes under DIC control

⇒ Activates the interrupt shield

PRIMARY MODE

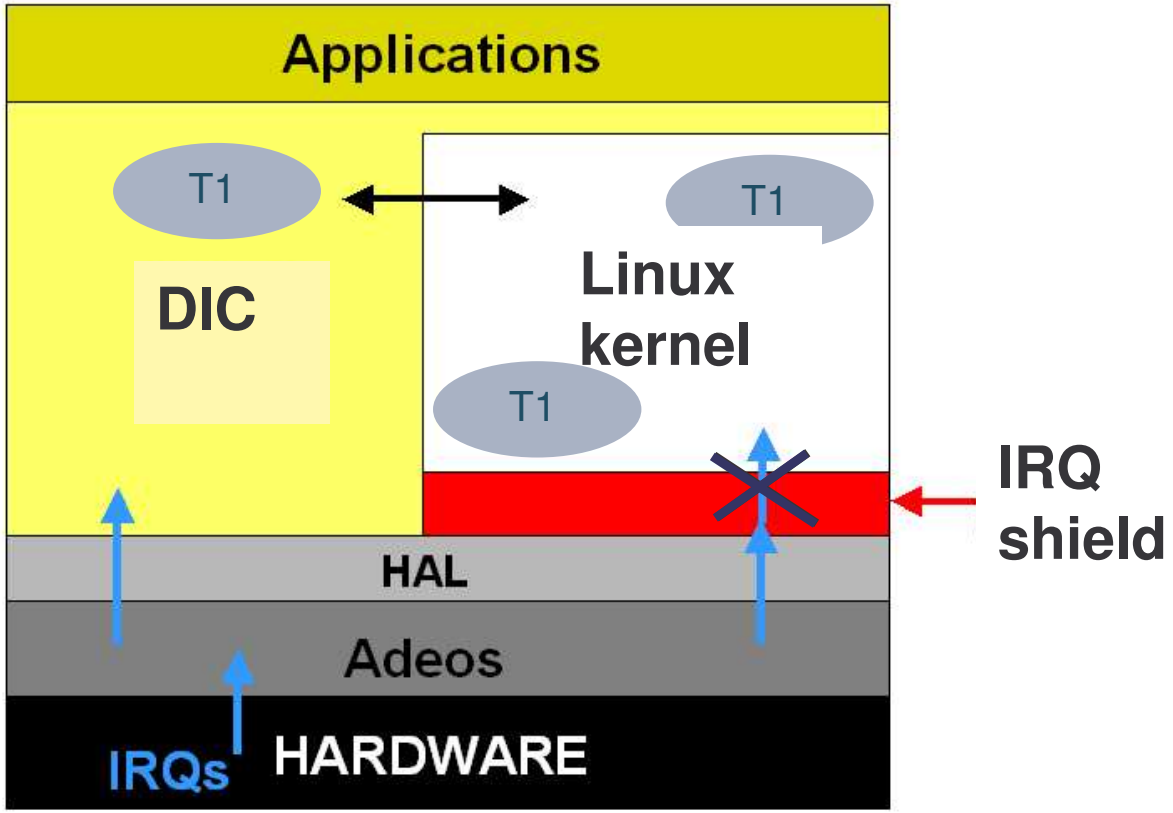
- Issues Linux system call

⇒ Interrupt shield remains on

SECONDARY MODE

- Suspends execution

⇒ Interrupt shield off



© Thales

This doc

Power/Energy Clarifications

Definitions

- Power: used in the context of high performance processors or in layout design, and is mostly related to thermal dissipation issues
- Energy: used in the context of embedded and battery powered systems as it considers the repartition of the power demands over time
  - e.g.: Dynamic Voltage Scaling (DVS) scales down power but not energy. Though, it is beneficial to battery due to Loss Capacity Effects.
  - In this presentation, we will only use the word “Energy” to clarify our interest in optimal management of the supply source. Though the power management remains significant in the present initiative.

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Energy in Battery Powered Systems

Energy is a bottleneck

- Battery powered system imposes a limited Power/Energy budget
- Increased applications and hardware complexity
  - More computation
  - More memory
  - More resource to allocation
  - Higher bandwidth required
  - Higher operating frequencies
  - Important Dynamicity…

Energy features needed:

- Hardware/Software monitors *(for the sake of dynamicity)*
- Efficient decisions on using hardware (e.g. Allocations, Scalings)
- "Performance" should include both notions of time and power

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Energy Saving Techniques

What is effective for power optimisation ?

- Hardware architecture is at the basis of the optimisation
- Operating System should manage properly time and resources
- New Software Developments techniques have to be used

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

What is effective for power optimisation ?

- Hardware architecture is at the basis of the optimisation
  - Technology is not the only factor
  - Appropriate hardware capabilities will greaten “software” saving
    - Power and clock domains, dedicate hardware modules (IDCT,…)
    - Monitors and/or performance counters
- Operating System should manage properly time and resources
  - Scheduling policies
  - Allocation policies (processors, memory)
  - Efficiency depends on:
    - Sensitivity to systems variations (e.g. tasks, external operating conditions)
    - Appropriate balance of inline/offline information
    - Reliability of Hardware indicators

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

What is effective for power optimisation ?

- Software Developers *(PA-API)*
  - Balance performance and power cost (e.g. multimedia applications)
  - Efficient software architecture (e.g. parallelism, task granularity, …)
  - Avoid bad practice (e.g. wait loops, random accesses to memory)
  - Encourage good practice (e.g. Producer-Consumer architecture)

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Balance power & performance

- Evaluation of HW & SW Monitors
  - Battery awareness
  - Drivers (antenna, Display, Serial ports)
  - Performance profiling (Memory, Bus Activity, Tasks)
- Management of Dynamicity (QoS)
  - Scheduling policies
  - Memory Management policies
  - Drivers (eg. power modes)
  - Tasks QoS (PA-API)
- Manage Concurrency (Cache Coherency, Multiple Power Domains)

Build a coherent system globally optimised *(multi-criteria)*

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Dynamic information on system activity

- Hardware activity monitors
  - Performance counters (memory,…)
  - Embedded sensors (Temp.), state of the supply source (e.g. battery, wall-supply)
  - Interrupts reveal system's activity
- System Monitoring
  - Task related information (pre-emption, context switch, successivity)
  - Workload of each processors and buses
  - Software events, Inter-Process Communication
  - Profiling to favour determinism (e.g. periodical events, communications,…)
  - Power Aware drivers (e.g. embed power models, report activity)

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

APM

- Implemented in the BIOS, which has the control
  - Notifications through hardware events
  - Monitoring of Battery State
- APM has 5 power modes:
  - Run APM
  - APM Standby
  - APM Suspend
  - Power Off
- Limitations
  - APM cannot observe devices activity
  - APM cannot determine users activity (e.g. screen-saver interruptions)
  - No individual management of devices

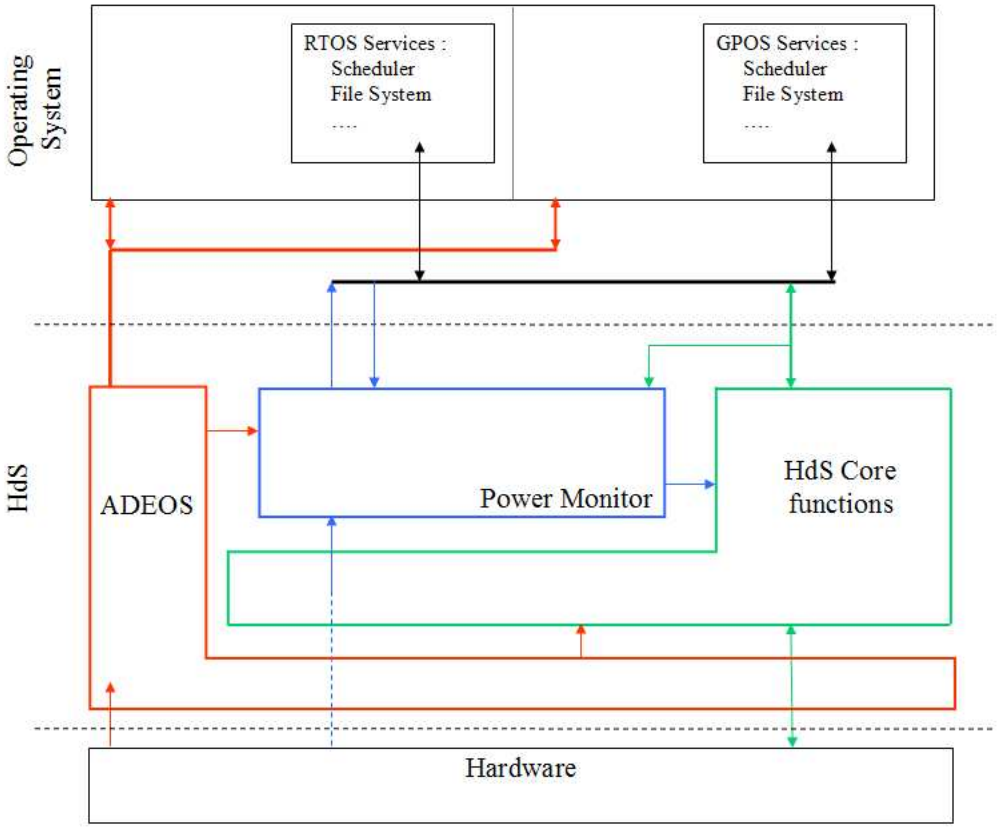
This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

- **ACPI**
  - Extended capabilities of APM
  - All components in the system can be power managed: ACPI provides a well defined way to discover and
    - Control all the components in the system
    - Power management decisions are made by the Operating System.
    - No executable Power Management code is required in the BIOS
    - Standardized description of power management capabilities according to devices

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

At the drivers level

- Power monitor
  - Abstracts and tracks the power/energy consumption
  - Track CPU workload, bus utilisation, interruptions profiling, …
  - Could compensate the lack of sensors through estimation (e.g. battery)



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

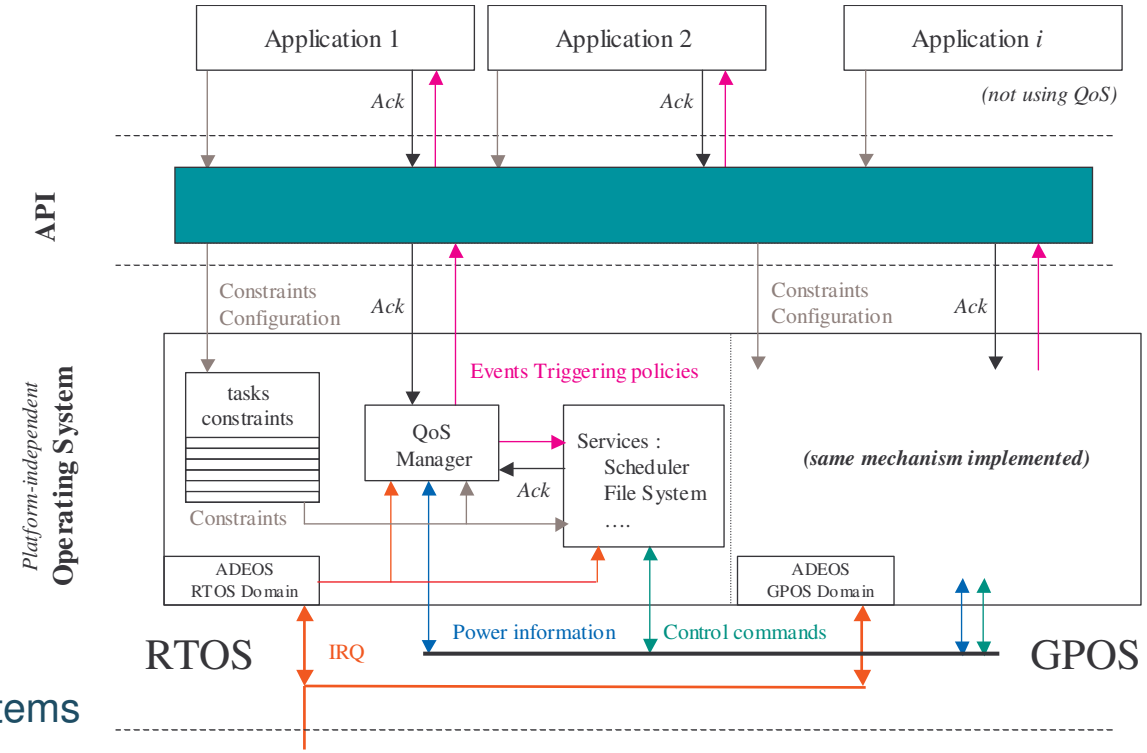
At System Level

- Global QoS manager
  - Ensures coherency of choices between services
  - Tracks system's activity (using triggers and policies)
    - Warns services when the global behaviour of the system has drifted
    - Analyses performance bottlenecks
    - Checks task constraints (e.g. deadline violation)
  - Example of tracked indicators
    - Most consuming task and resources
    - Least consuming task and resources
    - Timing profiling : ACET, Actual WCET, Actual BCET,…
    - Number of task migration per task
    - Number of cache miss per tasks
    - Amount of communication per task
    - Affinity between tasks

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

At Application Level

- API
  - Register a sensitivity list to system predefined events
  - Associates policies with this event
    - Modify their own behaviour
    - Deliver new requirements to the OS (e.g. status, constraints)



This document is the property of Thales Group and may not be copied or communicated without written consent of Thales

Conclusion

Main benefits for MPSoCs:

- Real-time capabilities for MPSoC in addition to GPOS functionalities
- Pluggable scheduling policies for handling both real-time and power consumption
- Notion of performance is generalized to include power efficiency (i.e. through policies)
- New Software QoS capabilities
- Transparent at application level
  - Use of specific API remains optional
    - Real-Time activity
    - Power awareness and definition of management strategies
- Ready for HdS Code Generation (cf Ahmed Jerraya's presentation)

Present drawback:

- *Adherence to Linux. We do not support other OS yet.*

This document is the property of Thales Group and may not be copied or communicated without written consent of Thales