Multi-core Devices for Safety-critical Systems: A Survey

JON PEREZ CERROLAZA, Ikerlan, Spain
ROMAN OBERMAISSER, Universität Siegen, Germany
JAUME ABELLA and FRANCISCO J. CAZORLA, Barcelona Supercomputing Center (BSC-CNS), Spain
KIM GRÜTTNER, OFFIS – Institute for Information Technology, Germany
IRUNE AGIRRE, Ikerlan, Spain
HAMIDREZA AHMADIAN, Universität Siegen, Germany
IMANOL ALLENDE, Ikerlan, Spain

Multi-core devices are envisioned to support the development of next-generation safety-critical systems, enabling the on-chip integration of functions of different criticality. This integration provides multiple system-level potential benefits such as cost, size, power, and weight reduction. However, safety certification becomes a challenge and several fundamental safety technical requirements must be addressed, such as temporal and spatial independence, reliability, and diagnostic coverage. This survey provides a categorization and overview at different device abstraction levels (nanoscale, component, and device) of selected key research contributions that support the compliance with these fundamental safety requirements.

CCS Concepts: • **Computer systems organization → Multicore architectures**; *Embedded hardware*; *Processors and memory architectures*; *Redundancy*; • **Hardware → Safety critical systems**;

Additional Key Words and Phrases: Fault tolerance, diagnostic coverage, time independence, spatial independence

ACM Reference format:
Jon Perez Cerrolaza, Roman Obermaisser, Jaume Abella, Francisco J. Cazorla, Kim Grüttner, Irune Agirre, Hamidreza Ahmadian, and Imanol Allende. 2020. Multi-core Devices for Safety-critical Systems: A Survey. *ACM Comput. Surv.* 53, 4, Article 79 (August 2020), 38 pages.
https://doi.org/10.1145/3398665

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under Grant No. TIN2015-65316-P, HiPEAC Network of Excellence and the Basque Government under Grant No. KK-2019-00035. The Spanish Ministry of Economy and Competitiveness has also partially supported Jaume Abella under Ramon y Cajal postdoctoral fellowship (No. RYC-2013-14717).

Authors' addresses: J. Perez Cerrolaza, Ikerlan, Arrasate/Mondragon, Pº. J. Mª. Arizmendiarrieta, 2, Spain; email: jmperez@ikerlan.es; R. Obermaisser, Universität Siegen, Siegen, Germany; email: roman.obermaisser@uni-siegen.de; J. Abella and F. J. Cazorla, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain; emails: {jaume.abella, francisco.cazorla}@bsc.es; K. Grüttner, OFFIS-Institute for Information Technology, Oldenburg, Germany; email: kim.gruettner@offis.de; I. Agirre, Ikerlan, Arrasate/Mondragon, Pº. J. Mª. Arizmendiarrieta, 2, Spain; email: iagirre@ikerlan.es; H. Ahmadian, Universität Siegen, Siegen, Germany; email: ahmadian@eti.uni-siegen.de; I. Allende, Ikerlan, Arrasate/Mondragon, Pº. J. Mª. Arizmendiarrieta, 2, Spain; email: iallende@ikerlan.es.

0360-0300/2020/08-ART79 $15.00
https://doi.org/10.1145/3398665

1 INTRODUCTION

Over the past decades, embedded systems have enabled tremendous improvements with respect to functionality, dependability, and performance in many application domains such as transportation and industrial control systems. In these domains, the embedded systems often play a crucial role in guaranteeing the overall safety of the system. These systems are referred to as *safety-critical systems* as their failure can derive into catastrophic consequences such as the loss of lives or severe environmental damage [18] (e.g., automotive cruise control [10], railway signalling [9], wind turbine integrity protection [159], pacemaker [146]). With the aim of reducing the risk of causing such fatalities, safety-critical systems must follow strict certification processes according to domain-specific safety standards. This process usually involves high development efforts and costs. As a general rule, the higher the safety integrity level, the higher the cost of safety certification [12, 158].

In addition, with the increasing digitization trend, where an increasing number of functionalities are implemented by software, the embedded systems often include functions with different safety criticalities that must also co-exist with non-critical software, conforming *mixed-criticality* systems. In the past, the mixed-criticality architecture has generally followed a federated architecture approach in which each major functionality is deployed on a dedicated computing node. The growing demand for additional functionality, leads to an increase in the number of computing nodes, wires and connectors. As a consequence, this leads to an increase in the overall cost, complexity, Size, Weight, and Power (SWaP) that in several cases limits the future scalability of this approach [12, 17, 39, 145, 158, 159]. For example, in the automotive domain, premium cars have more than 20 million lines of code deployed in around 100 computing nodes [35, 39] and electronic components added value ranges between 40% for traditional vehicles up to 75% for electric vehicles [126]. The current automotive evolution targets the development of intelligent Advanced Driver-Assistance Systems (ADAS) and autonomous driving solutions that will further increase the number of functionalities to be integrated [39].

One possible solution is the shift toward an integrated architectural approach [12, 17, 39, 145, 158, 159], where functions of different safety criticality are integrated in a reduced number of centralized computing nodes and processing devices. In this approach, safety certification becomes a challenge, because the integration of mixed-criticality functions requires justifying sufficient independence of implementation and sufficiently low probability of dependent failure between functions [12, 89, 93, 128, 158]. Moreover, this approach requires an increase in the overall computational performance of devices, which could potentially be achieved by means of multi-core devices and mono-core devices with higher frequency. The usage of mono-core devices with higher frequency is considered not competitive in several domains due to increased sensibility to Electromagnetic Interference (EMI) [78], low reliability of thermal dissipation fans [39, 159] and cooling systems volume and weight [170]. However, Commercial Off-The-Shelf (COTS) multi-core devices are becoming dominant in silicon manufacturer roadmaps [17, 28, 41, 80, 91, 126, 198] and provide a cross-domain potential solution, e.g., automotive [10, 96, 128], avionics [17, 99], railway [9, 25], industrial control [3, 158], medical applications [146].

In this scenario, safety-critical system developers of multi-core device-based mixed-criticality systems, need to comply with two sometimes conflicting and contradictory constraints. On the one hand, conservative functional safety standards based on the best safety industrial practices of the last decades, with none or limited consideration of multi-core devices (see advances in ISO 26262-11 [93]). On the other hand, a rapidly evolving and highly innovative semiconductor industry that produces multi-core devices with shrinking technologies and higher integration of cores [94].

This publication is a survey on multi-core devices for safety-critical systems. The bulk of the research in multi-core devices for dependable systems focuses on specific challenges such as device

and component architectures, reliability, time predictability, and safety certification strategies [12, 146]. This survey however, aims to provide a categorization and end-to-end view of safety implications on multi-core devices, starting from nanoscale level and up to device level, by bringing together and summarizing the most relevant research contributions that support the compliance with fundamental safety technical requirements. This survey is aimed at researchers, with the aim of offering a panoramic view (survey of surveys) of the set of fundamental technical challenges that need to be addressed at different abstraction levels. This survey is also aimed at multi-core device and safety system developers. It provides an updated categorized research state-of-the-art, which provides a starting point to go through the (complex) design decision process required in the development of safety-critical systems, as required by diligence and liability requirements when the applicable safety standard has none or limited consideration of multi-core devices [56].

The reminder of the survey is organized as follows. Section 2 summarizes basic concepts and terminology. Section 3 analyzes, categorizes and summarizes the state-of-the-art with selected key research contributions that support the compliance with fundamental safety technical requirements at different device abstraction levels (nanoscale, component, and device). Then, Section 4 summarizes avionics and space domain-specific safety technical considerations, and summarizes relevant example case-studies from several domains. Finally, Section 5 provides links to related research topics, overall conclusion, and future research directions.

2 BASIC CONCEPTS AND TERMINOLOGY

This survey uses the basic concepts, taxonomy and terminology defined by Avižienis et al. [18] for dependable and secure computing. In addition to this, this survey integrates various research fields with specific detailed concepts and terminology described by referenced major survey publications of each specific research field (e.g., nanoscale multi-core device dependability [146]).

The term multi-core device, or device for short, is used within this survey to refer to multi-core processors, System-on-a-Chip (SoC), Multi-Processor System-on-a-Chip (MPSoCs), FPGA with soft-cores and combinations of the previous. And the term computing node is used within this survey to refer to domain-specific terms such as Electronic Control Unit (ECU) for automotive and Line Replaceable Unit (LRU) for avionics.

2.1 Safety Certification and Standards

Safety is defined as the "absence of catastrophic consequences on the user(s) and the environment" [18]. In several cases, safety-critical systems must be approved by an independent certification agency as part of a certification process. IEC 61508 [89] is a generic international safety standard considered a reference for other domain-specific standards such as automotive (ISO 26262 [93]) and railway (EN 5012X [55]). For further details see the survey [176] that describes safety (and security) qualification and certification state-of-the-art and challenges for automotive, railway, space, and avionics domains. Taking into consideration the variability of domain-specific terms and requirements, this survey uses IEC 61508 as the reference safety standard. Section 4 provides a simplified survey of solutions specific for other additional domains such as avionics and space.

In IEC 61508, "Safety Integrity Level (SIL) is a discrete level corresponding to a range of safety integrity values where 4 is the highest and 1 is the lowest" [159]. The most stringent safety systems such as railway signaling systems (SIL4) are implemented with programmable electronics with a probability of dangerous failure in the range of 10^{-9} hours of operation ($\approx$ 114.155 years). This requires stringent technical solutions to reduce to such low level the probability of catastrophe due to systematic faults (e.g., method to reduce the probability of human, process and tool errors) and random faults (e.g., safety mechanism with a given Diagnostic Coverage (DC)).

A safety function is either fail-safe, if a safe state can be reached either by the safety function or by the diagnosis reaction (e.g., wind turbine stop [159], train stop [9]), or fail-operational if the system must guarantee full or degraded operation of the given function even in the presence of a failure (e.g., automotive autonomous driving system, avionics flight control system).

2.2 Fundamental Safety Technical Requirements

The development of safety-critical systems that must comply with safety standards such as IEC 61508, requires an appropriate safety methodology to mitigate systematic errors and compliance with at least two fundamental safety technical requirements (reliability and diagnostic coverage (DC)). The development of safety-critical systems based on multi-core devices extends the previous with spatial and temporal independence requirements [89, 93] (see IEC 61508-3 Annex F). Thus, compliance with at least these four fundamental safety technical requirements is required.

- The *reliability* of a device at time t is the probability that the device will operate correctly during the period of time $[0, t]$. Reliability is measured as the failure rate of the device (λ) and expressed in Failures In Time (FIT), the number of failures expected in 10^9 hours of operation. For a given error, the effect can either be safe (not dangerous) or dangerous with respect to the associated safety function. Thus, as described in Equation (1), device failure rate is made up of the addition of components safe failure rates (λ_s) and dangerous failure rates (λ_d). The reliability of a device can be improved by techniques such as fault-tolerance.
- *Diagnostic Coverage (DC)* denotes the effectiveness of diagnosis techniques to detect dangerous errors, expressed in coverage percentage with respect to all possible dangerous errors. The DC is classified as low (60%– < 90%), medium (90%– < 99%) and high (>= 99%) [89]. As shown in Equation (2), DC is calculated as the ratio between dangerous detected failure rate (λ_{dd}) and dangerous failure rate (λ_d), which includes both dangerous detected (λ_{dd}) and dangerous undetected failure rates (λ_{du}). The diagnostic coverage of digital circuits can be calculated and measured using methods such as Failure Mode and Effects Analysis (FMEA) and fault injection [129]. Dangerous detected errors usually require the definition of an associated reaction (e.g., activate safe state, error correction).
- *Temporal independence* is required to ensure that "one element shall not cause another element to function incorrectly by taking too high a share of the available processor execution time, or by blocking execution of the other element by locking a shared resource of some kind" [89], so "that elements will not adversely interfere with each other's execution behaviour such that a dangerous failure would occur" in the time domain [89]. For this purpose, as recommended by the standard [89, 93], *temporal predictability* (e.g., Worst Case Execution Time (WCET) analysis, scheduling) and temporal diagnosis shall at least be considered.
- *Spatial independence* is required to ensure that "data used by one element shall not be changed by another element" so "that elements will not adversely interfere with each other's execution behaviour such that a dangerous failure would occur" [89]. For example, exclusive access to resources is a common technique used to achieve this purpose.

The probability of dangerous failure per hour (PFH) depends directly on both the DC and the reliability (λ) of the device. PFH is lower with higher diagnostic coverage and lower λ (higher reliability). For a given safety function SIL, the standard [89] defines a system-level average probability of failure range and the device is allocated a portion of this range (e.g., high demand SIL4: 10 FIT $> PFH >=$ 1 FIT). Equation (3) shows the simplest procedure to calculate PFH, for a safety system that puts the system in safe state on detection of any failure [89]. As a general rule, for complex components such as COTS multi-core devices, it is assumed that the probability ratios of

dangerous and safe failures are 50% each ($\lambda_s = \lambda_d = 0.5 \times \lambda$):

$$\lambda = \sum \lambda_s + \sum \lambda_d, \tag{1}$$

$$\text{DC} = \frac{\sum \lambda_{dd}}{\sum (\lambda_{dd} + \lambda_{du})} = \frac{\sum \lambda_{dd}}{\sum \lambda_d}, \tag{2}$$

$$\text{PFH} = \sum \lambda_{du} = \sum \left((1 - DC) \times \sum \lambda_d \right) \approx \sum ((1 - DC) \times 0.5 \times \lambda). \tag{3}$$

2.3 Reliability Degradation Threats

The latest manufacturing processes are already producing 7-nm devices thanks to the advances in CMOS downscaling technology, manufacturing processes and electronic design automation tools [41, 146]. Nevertheless, this CMOS technology downscaling leads to several threats that contribute to reliability degradation such as References [29, 66, 83, 103, 146, 172]:

- *Process, Voltage, and Temperature (PVT) variability*: As the technology continues to shrink, manufacturing process variability introduces higher variability of transistor/gates properties with respect to design-time properties [46, 103, 182]. Runtime variability due to PVT and aging, can also lead to transistor/gates variability but at different locations and operational lifetime points of the device [46, 103]. This variability can lead to timing failure of circuits if design-time margins are exceeded and therefore reduce the device reliability.
- *Transient faults and soft errors*: Soft-errors are caused by environmental conditions (e.g., α-particles, cosmic rays, ionizing radiation, EMI, cross-talk) leading to a transient perturbation that can be manifested as memory bit-flip in memory cells or combinatorial logic result error [29, 101, 103]. If the perturbation affects a single cell, then the event is called Single Event Upset (SEU), and if it affects more than one, then the event is called Multiple Bit Upset (MBU).
- *Permanent faults*: Permanent faults are those whose presence is persistent and permanently affect the correct functional behavior of the system [18, 29]. As the technology continues to shrink and device complexity continues to increase, the probability of physical defects in the hardware (e.g., short-circuit) during lifetime of a system continues to increase, from rare improbable events to probable events, and therefore error tolerance and diagnosis becomes also important to ensure the reliable and safe operation of the device [29].
- *Aging (degradation)*: The temporal degradation of the device can lead to timing degradation and permanent errors. As explained in Reference [83] several factors contribute to this degradation, where *negative bias temperature instability* and temperature are prominent factors.

2.4 Semiconductor Industry Trend

As described by Corradi [41], the semiconductor industry is facing a technology and economic challenge, where fewer chip makers can afford the investments required for 16 nm and below technology. At industry level, the consequence is a reduction in the number of leading semiconductor companies and a general trend toward the mass-production of heterogeneous multi-core SoCs optimized for maximum average performance that target multiple domains with a single SoC solution [41]. Except for the automotive domain, safety-related semiconductor market niche is too small for semiconductor companies to provide economically competitive safety specific solutions

Table 1. Summary of Selected Mass-produced COTS Devices from Major Silicon Manufacturers

Device Type	Core Family	Device Family	Manufacturer
Generic devices	ARM	Zynq 7000	Xilinx
		Jacinto, Sitara, C6000	Texas Instruments
		R-Car V3H, RZ	Renesas
	Power Arch.	iMX8, QorIQ	NXP
	x86	Xeon, Atom C3000	Intel
Safety devices	ARM	Hercules (TMS570)	Texas Instruments
		R-Car H3	Renesas
	Power Arch.	MPC57xx	NXP
		SPC5	STMicroelectronics
	R850	RH850	Renesas
	Tricore	AURIX	Infineon
Hybrid devices	ARM	Zynq UltraScale+	Xilinx

compliant with previously defined certification standards. The automotive domain safety microprocessor solutions market is approximately $5 billion [20, 80], approximately one percent of the global market estimated at $412 billion in 2017 [2].

Therefore, it is expected that multi-core devices used in the development of safety-related systems will have a higher dependency on fewer manufacturers with devices that tend to be either *safety device*, *generic device* or *hybrid device*. Table 1 provides a selection of mass-produced COTS *generic*, *safety* and *hybrid devices* provided by major embedded domain silicon manufacturers.

- *Generic device*: General purpose heterogeneous multi-core device optimized for cross-domain maximum average performance. As shown in Table 1, an average COTS *generic device* is designed for maximum average performance and organized in a shared resources architecture based on components such as [128]: cores, interconnect bus(es) and memory controllers, private and shared caches, Uniform Memory Architecture (UMA)/Non-uniform Memory Architecture (NUMA), and addressable peripherals [28, 41, 47, 107, 126, 170, 197].
- *Safety device*: Multi-core safety device generally designed for the automotive domain that might also support industrial safety standards such as IEC 61508 [80, 126]. As shown in Table 1, an average COTS *safety device* is designed to comply with safety standard(s) and provide maximum average performance in a shared resources architecture based on components such as Reference [128]: cores, shared buses, private caches, UMA/NUMA, and addressable peripherals.
- *Hybrid device*: Multi-core device that combines previous options, e.g., *generic device* with a certifiable "safety island" (e.g., Zynq UltraScale+), *generic device* with integrated FPGA that enables the integration of custom safety designs (e.g., Zynq with ARM and MicroBlaze [12]).

3 MULTI-CORE DEVICE ARCHITECTURES

3.1 Summary

As explained by Furano et al. [63], taking into consideration the limitations of available ASIC technology, the overall computational performance of devices can be increased by means of mono-core devices with higher frequencies, new architectures, Deep Submicron (DSM) and multi-core technologies. The scope of this survey are new architectures, DSM and multi-core technologies.

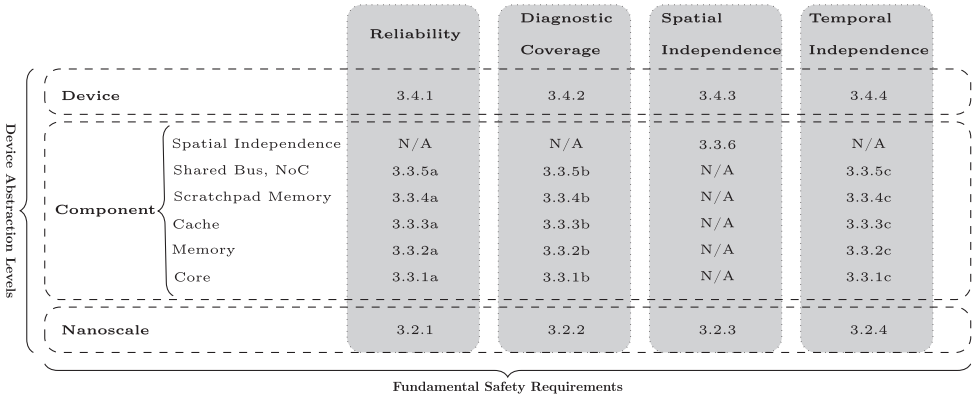

Fig. 1. Structure of the survey.

The developer of a safety-critical system based on multi-core device(s) needs to understand all relevant safety techniques provided by the given device and define required additional complementary measures to comply with applicable safety standards. However, due to the different nature of threats and challenges that must be addressed at different levels of abstraction (e.g., nanoscale manufacturing process variability, cache time predictability, cache coherence) there is a high research fragmentation and a great variability in terminology that makes this analysis difficult. This section contributes with a categorization of a wide and highly fragmented research area in the field of multi-core device architectures for dependable systems, based on three device abstraction levels (nanoscale, component, and device) and a set of fundamental safety requirements that safety-critical system developers must at least address (see Section 2.2). Figure 1 shows the survey structure, mapping survey sections with device abstraction levels and fundamental safety requirements. Table 2 classifies selected research contributions described in this survey with respect to device abstraction levels and fundamental technical safety requirements.

With respect to reliability and DC, several research surveys [101, 117, 146] and contributions [66, 83, 124, 147, 181] provide a comprehensive state-of-the-art of reliability threats and mitigation techniques, DC and fault-tolerance by means of error detection and correction techniques. Even within this specific topic, which focuses on reliability and DC techniques, there is a wide and partially fragmented research area that leads to terminology [146], taxonomy and classification variability. Nonetheless, selected representative techniques have been categorized with the previously described approach and this categorization can be used as a reference for the categorization of the additional techniques described in these contributions. Qualitative and quantitative comparison of techniques are also described in previously referenced contributions using similar (e.g., fault coverage, error coverage) but not always equivalent concepts with respect to IEC 61508 requirements for DC, thus not enabling a direct quantitative comparison of described techniques. Finally, an equivalent approach has been taken with respect to FPGA technology, where several research surveys [24, 40, 142] provide a comprehensive state-of-the-art of mitigation techniques, design standards and fault-tolerant methodologies for FPGAs.

Concerning temporal independence, Maiza et al. [128] provide a survey of timing verification techniques and time-predictability challenges and solutions where selected component and architectural contributions have been categorized at component and device level. At nanoscale level, timing fluctuations are due to physical properties such as temperature and require different techniques. The development of time predictable components provides a solid foundation to achieve device-level temporal independence, simplify WCET analysis and software time composability

Table 2. Selected Research Contributions Classification by Device Abstraction Level and Fundamental Safety Technical Requirement

Level	Requirement	Technical Contribution	Section
Device	Temporal Independence	—Software [12, 42, 75, 151, 158, 159, 186], WCET [6, 23, 38, 60, 107, 123, 128, 134, 140, 150, 155, 165, 173, 195, 203], scheduling [33, 42, 74, 155, 159] —Device architecture [12, 31, 32, 67, 69, 84, 134, 152, 166, 167, 174, 175, 186]	3.4.4
	Spatial Independence	—Software [12, 15, 42, 75, 111, 161, 184] —Device architecture: memory [155, 159], shared bus [12], cache [12, 111, 167], security [115, 160] —SRAM FPGA [73, 79, 114]	3.4.3
	Diagnostic Coverage (DC)	—Software [12, 12, 43, 111, 113, 115, 141, 158, 159] —Device architecture [12, 43, 111, 115, 158, 159, 181]	3.4.2
	Reliability	—Device architecture [81, 82]	3.4.1
Component	Temporal Independence	—Device components: core [49, 121, 122, 152, 175, 204, 205], memory [19, 21, 44, 53, 54, 70, 71, 97, 120, 132, 147, 152, 200, 202], cache [6, 38, 72, 116, 188, 195], NoC [7, 30, 34, 68, 85, 87, 92, 133, 144, 154, 179, 194], SPM [16]	3.3.6 3.3.5 3.3.4 3.3.3
	Spatial Independence	See device-level spatial independence	3.3.2 3.3.1
	Diagnostic Coverage (DC)	—Software [66, 117, 158, 159] —Device components: core [45, 46, 66, 101, 124, 131, 181], memory [117, 147], cache [117, 147], SPM [117, 147], shared bus [9, 12, 112], NoC [85]	
	Reliability	—Software: redundancy (e.g., spatial, temporal) —Device components: core [66, 95, 101, 162, 181]	
Nanoscale	Temporal Independence	—Software, device components and hardware: thermal and power management [59]	3.2.4
	Spatial Independence	—See component and device-level reliability and DC —Mass-produced COTS device design process [89]	3.2.3
	Diagnostic Coverage (DC)	—Software [66, 117, 158, 159] —Device hardware [66, 83, 117, 124]	3.2.2
	Reliability	—Software [5, 59, 83, 117, 149, 158, 159] —Device hardware: [66, 81–83, 117, 124, 147, 181, 182] —SRAM FPGA [24, 40, 142]	3.2.1

[152]. However, with respect to time predictability and WCET, most of the research contributions can be classified as either "how things should be done" (e.g., time predictable components) or "how things can be done" (e.g., how to achieve temporal independence with *generic devices*).

Regarding spatial independence, this is generally supported by specialized architectures (e.g., NUMA) and/or components (e.g., MMU, cache coherency) commonly available in COTS devices that can be used and configured by system designers. However, a major challenge is to ensure the correct configuration (e.g., multiple hardware data paths), reliability and DC (e.g., cache coherency diagnosis) of these components based on which spatial independence justification is based.

Finally, the current research fragmentation leads to a need to address a holistic cross-level and cross safety technical requirement approach as required to develop and certify safety-critical systems. This is observable in the development of domain-specific case-studies described in Section 4.3, where fundamental safety requirements are addressed at different abstraction levels using different application specific and ad hoc combinations of research contributions (e.g., wind turbine [12, 110, 158, 159, 184]). For example, a cross-level approach is required to achieve the DC required

by safety standards, because, as the complexity continues to increase for device architectures, components and interconnections, meeting the required DC at software application level becomes a challenge if sufficient hardware diagnostic support is not provided [12, 83, 158]. But, the addition of these diagnosis techniques in the hardware, can also have an impact on silicon die and lead to cost increase, power overheads and reduced performance. So, the combination of hardware and software techniques combined at different abstraction levels are usually required to detect complex hardware faults such as MBU and transient faults, as supported hardware techniques alone might not be sufficient [83, 101, 124, 149, 172]. However, cross-requirements research contributions that aim at reconciling reliability, DC and time predictability are still scarce (e.g., References [123, 180]).

3.2 Nanoscale Level

At nanoscale level, selected key research contributions focus on the definition of techniques for reliability, diagnostic coverage, circuit spatial independence and circuit time predictability.

3.2.1 Reliability. Reliability laboratory testing data such as References [76, 199] provide FIT rates and wear-out period estimations and measurements. Several surveys [83, 124] provide additional FIT rate data sources and analysis. Intuitively, as CMOS technology downscales and supply voltage decreases, raw reliability should decrease (FIT increase) due to higher probability of random soft errors and transient faults, because less energy is required for a cell upset event. In addition to this, reliability should also decrease due to increasing PVT variability, permanent faults, and aging factors.

However, in multiple cases this trend is mitigated and even reverted as the technology downscales [83, 124]. For example, Xilinx provides reliability laboratory measurements where FIT rates for technologies between 350 and 16 nm vary between 2 and 12 FIT, with better values for 90 nm than for 350 nm [199]. This is due to several reasons, such as to physical properties (e.g., sensitive deployment area decrease leads to memory soft error rate decrease [83, 124]), fabrication process improvements that mitigate PVT variability (e.g., lithographic process), and fault-tolerance techniques deployed in the device such as:

- *Design-time fault-tolerance* that provides hardware-level mitigation support without the run-time execution of additional external circuitry. For example, hardening (e.g., radiation hardening, resilient transistors against faults), design margin-based mitigation techniques (e.g., operating at higher supply voltages, gate sizing, body biasing, circuit guard banding) and diversity [83, 124, 147, 181, 182].
- *Run-time fault-tolerance* that requires the execution of additional circuitry. For example, redundancy (space, time, information), pipeline protection with shadow latches, tunable replica circuits [66, 83, 117, 124, 181].

For aging effect mitigation, thermal management (e.g., Dynamic Voltage and Frequency Scaling (DVFS), specific techniques to reduce temperature spatial and temporal gradients) and previously described circuit-level mitigation techniques (e.g., guard banding) are common methods [83].

With respect to wear-out period, as explained in Reference [5] and reliability laboratory testing data [76], device FIT rates fall sharply at the end of the chip lifetime. This wear-out period is reduced as the technology downscales [147] and imposes a relevant restriction in systems lifetime, e.g., the typical wear-out period for 180-nm technology is higher than 100 years and for 65-nm technology it is expected to be lower than 5 years [5, 76].

Software Techniques. As described in Reference [66], previously described fault-tolerance techniques have an impact on silicon die and can lead to cost increase, power overheads and reduced performance. Therefore, devices targeting industrial markets with high reliability

requirements are prone to include such features as oppose to high volume markets that have lower reliability requirements [182]. Because of this, application specific or generic software fault-tolerance techniques can be used to improve hardware reliability, e.g., multithreading for soft error fault-tolerance [149], *algorithmic-based fault tolerance* [117], software redundant execution (e.g., software application, replicated instructions and check inserted by the compiler) [117], combination of software and hardware techniques with WCET guarantees [5], software-level thermal co-management [59, 83], application-specific techniques (e.g., References [158, 159]).

Hardware Fault Tolerance of One (HFT = 1). IEC 61508-2 (Annex E) [89] defines common minimum requirements for the development of on-chip redundancy with hardware fault-tolerance greater than zero (HFT=1) using integrated circuits with one common substrate [89]. This solution enables the development of devices with on-chip redundancy (e.g., 1oo2) such as HICore1 [81, 82] and Zynq-7000 [77]. For that purpose, at least the following safety technical considerations shall be fulfilled at nanoscale and common substrate level [89]: separate physical blocks on substratum, common cause failure avoidance and mitigation (e.g., temperature, power supply), minimum distance between boundaries of physical blocks (e.g., 50 μm [82]), short circuit, and cross-talk mitigation [89].

FPGAs. SRAM FPGAs use a configuration memory that defines the operations of the electronic circuit implemented by the FPGA. Different hardening techniques can be used to tolerate and mitigate soft and transient errors that could lead to a modification of the intended circuit implementation, e.g., redundancy, scrubbing, partial dynamic reconfiguration, combinations of the previous techniques [24, 40, 142]. However, antifuse-based FPGAs provide higher reliability than SRAM FPGAs, but their higher cost and lack of reprogrammability limits their application outside the aerospace domain.

3.2.2 Diagnostic Coverage (DC). The run-time diagnosis of soft errors and aging errors is a challenge due to the random nature of the error, location, and time distribution [12, 13, 103]. Multiple industrial and research grade techniques can be used to provide required circuit diagnostic coverage, e.g., monitoring circuits, self-tests and diagnostics, transition detector with time-borrowing, and code-based techniques [66, 83, 117, 124]. Example techniques are described at core and memory component level in Sections 3.3.1 and 3.3.2.

3.2.3 Spatial Independence. Non-intended electromagnetic interference between nanoscale-level elements, such as signals, clocks, and power-supplies, due to common cause failures, such as cross-talk, can lead to component and device-level errors. These faults are mitigated by state-of-the-art mass-produced COTS device design processes [89], reliability, and DC techniques described at component and device level (see Section 3.4.3).

3.2.4 Temporal Independence. Device voltage and temperature fluctuations can lead to circuit timing fluctuations. For example, in specific designs 10% supply voltage variation can lead to up to 20% circuit delay and 10°C temperature increase can lead to approximately 5% interconnect delay [66]. While temporal diagnosis techniques can detect and react/recover to circuit timing errors, this is a potential temporal interference source among integrated functions. For example, the computation and resource usage of a non-safety function could potentially generate a temperature increase that leads to minor timing fluctuations in the computation of a safety function. In addition to this, as described in Reference [59], device built-in mechanisms for thermal and power management (e.g., DVFS) that could lead to temporal interference among integrated mixed-criticality functions should be analyzed during design-time and safely managed during run-time.

3.3 Component Level

At component level, key research contributions focus on the definition of techniques for reliability, diagnostic coverage and time predictability for all common device components such as core, memory, cache, scratchpad memory, shared bus, interconnection networks (e.g., Network-on-a-Chip (NoC)) and specialized components for spatial isolation (e.g., Memory Management Unit (MMU)).

3.3.1 Core. A core is a program processing unit that reads and executes program instructions (software application) and reads/writes data in associated memories.

Reliability. The survey [101] and overviews [66, 83, 181] describe and provide quantitative comparisons of processor core fault-tolerant architectures based on techniques such *hardening* and *redundancy*. For the given examples and compared techniques [66, 83, 101], hardware replication can generally provide the highest reliability increase and generally it has a higher hardware overhead (e.g., >= 300% for triplication) than hardening techniques (e.g., < 15% for LEON3FT).

- *Hardening*: LEON3FT and LEON4FT are 32-bit SPARC-V8 RISC architecture soft-cores designed specifically for space domain fault-tolerance, based on an extensive usage of error detection and correction codes such as ECC for memory blocks [162].
- *Redundancy* (triplication): ARM Cortex-R5-based triple core lock-step architecture provides a triple modular on-chip redundancy solution with an error propagation unit that includes majority voter, error detection logic and synchronization logic [95]. With respect to software, software triplication schemes with voting can also be deployed among on-chip redundant cores (e.g., on-chip redundancy with HFT = 1 [159]) to tolerate soft errors but might not be valid to tolerate permanent errors.

Diagnostic Coverage (DC). The survey [101] and overviews [66, 181] describe and provide quantitative comparisons of several online error detection and recovery techniques based on techniques such as "redundancy", "dynamic verification", "anomaly detection" and "temporal diagnosis" for which selected contributions are summarized:

- *Lockstep redundant execution* provides a safety standards compliant (e.g., ISO 26262, IEC 61508) medium to high DC with a high hardware overhead cost (e.g., >= 200%) and minimum performance overhead [66, 181]. This is a cost-competitive and common technique used in COTS *safety devices* listed in Table 1 (e.g., AURIX TC3xx). However, dual-core devices that operate in locked-step, to increase the core DC are considered single-core devices.
- *Dynamic verification* uses dedicated run-time hardware checkers, which are not redundant execution units, to validate specific execution invariants that are assumed to be true in the absence of errors [66]. Analyzed hardware-based techniques have low hardware overhead (e.g., 6%–17%), low detection latency, and low performance overhead [66, 101]. Argus [66, 131] is an example hardware technique that can potentially achieve medium DC with low hardware area overhead (< 17%) [66]. It performs core diagnosis by means of run-time checking of four invariants: the control flow and data-flow are executed as specified in the program binary, correct computations are performed, and memory is not corrupted. It does not perform low-level checking of core components nor provide redundancy of cores for diagnostic purposed. Generic software techniques such as control flow checking [89, 181] can be used to diagnose that the expected control flow is executed (the invariant). Specialized software-based redundant solutions can provide medium potential DC with no hardware overhead cost, low latency but high performance overhead [66]. For example, *software implemented fault tolerance* is a soft error diagnosis solution that integrates control

flow checking with compiler-based transformation of the software that duplicates program instructions and inserts checkpoints to diagnose incorrect computations. Arithmetic codes [181] are also specialized software redundant solutions that add data redundancy to generate codewords that if correctly processed by the associated arithmetic operations will also generate a codeword (the invariant).

- *Anomaly detection* techniques use dedicated or modified circuits to detect errors. For example, periodic Built-in Self Test (BIST) can potentially achieve low to medium DC with low hardware overhead (5%) [66]. Software-based techniques do not require hardware overhead but require a relevant software overhead to perform periodic tests. For example, a periodic *software-based self-test* may require 5%–25% of system execution time [66]. Other example software techniques are assertion and sanity-based fault-tolerance [117] and application specific techniques (e.g., References [158, 159]).
- *Temporal diagnosis* techniques use dedicated or modified circuit to detect timing errors. For example, Razor provides circuit-level error detection and correction of timing errors with low hardware overhead (e.g., $< 3\%$) [45, 46]. It targets the development of power reduction processors where the error rate during operation is monitored to tune the supply voltage.

Temporal Independence. Table 1 provides a summary of selected COTS core families for *generic*, *safety*, and *hybrid devices*. In this summary, it is possible to conclude that a limited set of core families are predominately used by major device manufacturers (e.g., ARM, Power Architecture, x86). These generic purpose complex cores provide high average-case performance but time predictability and WCET estimation becomes a technical challenge as described in Section 3.4.4.

However, as described by Schoeberl et al. [175], a relevant research activity targets the development of time predictable cores that potentially simplifies the WCET estimation and provides lower bounds. The publication [175] provides an overview of time predictable cores, from which most relevant time predictable cores with respect to the scope of the survey are summarized and extended with additional research contributions of interest:

- Patmos [175] is an open-source RISC processor core designed for time predictability and low WCET bounds. The design includes a dual-issue pipeline, specially designed caches for WCET analysis, support for scratchpad memories and NoC interface.
- FlexPRET [204, 205] is a multithreaded RISC-V [102] processor designed for the integration of hard and soft real-time threads, which includes a hardware thread scheduler to schedule hard real-time threads and soft real-time threads. It is an extension of the Precision Timed (PRET) processor [121, 122, 175] that implements ARM ISA with a RISC pipeline, chip-level multithreading, Time-division Multiplexing (TDM) to access shared main memory and scratch memory instead of caches.
- Hard real-time SMT core [152] implements Tricore instruction set [91] with two in-order super-scalar processor pipelines (integer and address) with multithreading support. This core is designed to ensure task-level bounded maximum delay due to inter-task interference, between hard real-time and non-hard real-time tasks.
- SPEAR [49] is a constant-time instruction set processor with minimum interrupt jitter and delay support. It is based on a custom instruction set with fixed and constant execution time.
- PTARM [121] is a soft-core that implements a subset of the ARMv4 ISA with a thread-interleaved pipeline and provides a time predictable DRAM controller.

3.3.2 Memory. The execution of software applications in cores requires program and data memory storage. As explained in References [126, 187], in multi-core devices two basic memory

architectures are used: Non-uniform Memory Architecture (NUMA) and Uniform Memory Architecture (UMA). In NUMA each core has a dedicated physical memory with exclusive access privileges and UMA is a shared memory architecture where the physical memory is uniformly shared among cores and additional peripherals such as Direct Memory Accesses (DMAs).

Reliability. As explained in References [83, 147], memory capacity continuously increases with an associated increase in the amount of occupied silicon area and overall device reliability weight. While memory soft error rate decreases due to downscaling, because the sensitive deployment area decreases, the integrated memory sizes continue to increase at a greater pace [83, 124]. To improve memory reliability several new technologies have emerged such as phase change memories [147].

Diagnostic Coverage (DC). As described by References [117, 147] and supported by COTS devices described in Table 1, error detection in memory is generally performed by information redundancy techniques such as parity bit (low DC) and Error-Correcting Codes (ECCs) that can provide medium DC.

Temporal Independence. Conventional *memory controllers* are not designed to support temporal independence for mixed-criticality software applications and this leads to potential temporal interference between applications that share the same *memory controller*. A conservative solution to bound the temporal interference and WCET, is to consider the worst-case latency for all possible memory transactions. However, this solution focuses on the WCET scenario, ignores application and device specific solutions to improve overall efficiency (e.g., software applications scheduling, bandwidth utilization) and the unpredictability of applications dense memory access patterns. Furthermore, it is more difficult to provide a tight WCET for multi-core devices compared to single-core devices, due to inter-application interference accessing shared resources. As a consequence, the execution time of a given software application depends on other simultaneously executing applications [153]. To support time predictability, specialized memory controllers and management strategies can be used, for example:

- To enhance efficiency, a dynamic command scheduling-based back-end real-time memory controller architecture was proposed [120]. A conservative open-page policy is proposed by Reference [70] that improves average-case performance (e.g., bandwidth, latency) of a firm real-time controller while supporting real-time guarantees.
- Reference [53] offers an interference-free memory for critical applications with a memory controller that supports the integration of applications of different time criticality. In this approach, memory is treated as a set of independent virtual devices that are mapped to mixed critical workloads by a partitioning strategy.
- Reference [97] proposes a memory controller that supports dual-criticality by means of virtual division of memory banks. In this approach, each virtual memory bank type is managed with a specific request scheduler policy.
- MemGuard [44, 201, 202] provides an algorithm to perform memory bandwidth management at core level by using hardware performance counters to monitor the number of last-level cache misses, or accesses via the memory interconnect. Reference [185] introduces a bandwidth regulation module that adapts and extends the MemGuard Linux kernel module algorithm [200] (known as MemGuardXt) for hardware implementation and supports specific operating modes and options, with guarantees for rate-constrained flows and bandwidth rates.

Transactional Memory (TM) controller is a concurrency control mechanism for shared resources. It has been introduced to simplify the concurrency management in multi-core systems, by

supporting the atomic execution of sets of load and store instructions. They suit safety-critical systems as they support time predictability and fault isolation. The *temporal and spatial partitioning* approach is commonly used to establish the required determinism in the implementation of transactional memories. A transactional memory controller can be implemented as software (STM), hardware (HTM), or combination of both. The following selected contributions focus on the WCET pre-estimation for the schedulability of a given software application(s) task set:

- Reference [19] introduces an STM algorithm that prevents starvation by means of update transactions conflicts management. The selection of the conflict detection policy (lazy or eager type), has a significant effect on the temporal schedulability, because it determines when conflicting transactions are aborted. Reference [21] presents a real-time scheduling perspective analysis of STM conflict detection policies. Reference [54] presents an STM concurrency control for multi-core devices with real-time software applications. It also presents transactional conflicts contention manager.
- Reference [152] describes a dynamically partitioned cache interface solution that handles memory transactions with bounded maximum delay guarantees for high criticality software tasks (e.g., hard real-time tasks). This solution, executes on top of the memory controller. In Reference [132], an analysis of conflicts and aborts for different criticality transactions is provided (e.g., hard real-time, best-effort). Moreover, Reference [71] describes a Software Transactional Memory (STM) contention manager that supports priority-based transactions.
- Reference [148] describes a priority-based HTM controller designed specifically for mixed-criticality systems, with support for timing analysis, selection of conflict resolution algorithms and techniques for fault propagation avoidance.

Previously described approaches address the resolution of memory transaction conflicts. However, they do not avoid or address temporal interference in other components such as the interconnect (see Section 3.3.5).

3.3.3 Cache. Cache components [135, 136, 163, 177] reduce memory access time of computing cores, bridging the frequency gap between cores that operate at higher frequencies than memory components. A multi-core device might integrate private caches assigned exclusively to given cores (e.g., L1 cache) and public caches shared by several components such as cores (e.g., L2 cache).

As explained in Section 2.4, an average COTS *generic device* integrates a hierarchy of private and shared caches: one or several private caches per core connected to interconnect(s) and shared caches between the interconnect(s) and shared memories. An average COTS *safety device* integrates a private cache per core connected with the shared bus where memory and addressable peripherals are connected. Few exceptions (e.g., NGMP space multi-core device [106]) have shared caches that can be configured to operate partitioned across cores so that each partition effectively becomes a private cache. However, research devices that target the development of time predictable devices tend to use different approaches (e.g., T-CREST [174, 175]) and NoC-based solutions with private caches (see Section 3.3.5).

Reliability. Most of the area in computing devices is devoted to SRAM storage in the form of cache memories. Hence, their reliability is of prominent importance. Increased integration has led to increased susceptibility to transient faults, which are mostly addressed with parity and ECC [117, 147]. The particular solution to use depends on whether data is stored redundantly, so that parity suffices for error correction, or data is unique and ECC is needed for recovery. The existing options are discussed by Benedicte et al. [22].

Diagnostic Coverage (DC). DC in caches often builds on parity and ECC [117, 147], already deployed for reliability purposes as discussed before, since they allow testing contents correctness on each access at speed. However, cache coherency diagnosis is described at device level (see Section 3.4.2).

Temporal Independence. Although caches improve average performance and WCET, it is hard to prove in general whether cache accesses will be guaranteed to hit, either with abstract interpretation or in a test campaign [6, 195]. Hence, time guarantees are hard to obtain in general, and so caches defeat either WCET estimates tightness or trustworthiness. Moreover, speculation and priority inversion effects in computing devices may make cache hits—whose latency is lower than that of misses—lead to longer execution times than cache misses, thus further challenging time predictability [116].

A thorough survey on cache designs and time predictability can be found in Reference [72]. Details on the pros and cons of cache memories for probabilistic timing analysis is further developed by Cazorla et al. [38]. In this context, Valsan et al. [188] analyze the impact on time predictability of resources surrounding cache memories, such as queues and buffers, and noticing that, if not managed properly, cache (space) partitioning may not suffice to achieve time predictability.

3.3.4 Scratchpad Memory (SPM). SPMs are software-managed memories, thus being interesting alternatives to traditional, hardware-managed, caches [16, 191]. SPMs are usually small and high-speed SRAM. SPMs reduce design complexity, and thus improve power and performance, w.r.t. cache memories, and improve time predictability by granting users and compilers with explicit control of their contents [16]. SPMs disadvantages relate to the burden of explicitly controlling their contents, which increases software development costs and may jeopardize software portability, and thus the ability to reuse legacy software. Multi-core safety devices like the AURIX TriCore-based architecture include private both program and data SPMs for each core in combination with instruction and data caches [91]. Hybrid devices such as the Zynq UltraScale+ also include an SPM, named on-chip memory, shared across the different processor cores in the platform [198].

Reliability. As in the case of caches, SPMs are equally vulnerable to transient faults, and hence, solutions such as coding (in the form of parity or ECC [117, 147]) are highly popular for SPMs.

Diagnostic Coverage (DC). As in the case of caches, parity and ECC already provide means to reach required low to medium DC [117, 147]. Moreover, differently to caches, coherence is not a concern for SPMs, since their contents are explicitly managed by software [16, 191].

Temporal Independence. Time predictability is a key factor that eases WCET estimation and therefore temporal independence, and a key feature of SPMs is that they provide predictable access times. In multi-core architectures with SPMs like the AURIX and Zynq UltraScale+ [91, 198], the SPMs allow achieving higher levels of temporal independence than with cache-based solutions, as they remove the impact of cache hit/misses and the cache coherence problem.

3.3.5 Shared Bus and Interconnection Network. The integration of multiple cores and shared peripherals in a single device requires the integration of inter-core communication mechanisms and shared access to peripherals with ideally low latency and high bandwidth. Different architectural patterns with specific on-chip communication solutions are applicable depending on the device type and number of cores: shared bus, interconnect (or switch) and NoC. Each architectural pattern has different reliability, DC, and time-predictability characteristics.

Reliability. Interconnect and switch components are proprietary solutions with none or limited detailed documentation provided by silicon manufacturers [99]. Thus, component reliability

measurements are provided by the manufacturer and limited information is known about internal fault-tolerance mechanisms. However, error detection and correction for data transmitted is possible by building on end-to-end ECC protection, i.e., codes are sent along with the data. However, such a solution does not provide by itself support to detect deadlocks, livelocks, lost message detection and the like. With respect to NoCs, the survey [86] provides a summary and quantitative comparisons (e.g., bandwidth) of relevant contributions for reliability and fault-tolerance in real-time NoCs.

Diagnostic Coverage (DC). A shared-bus provides a simple solution already considered by current safety standards, with known failure modes and associated diagnosis techniques [89]. Interconnects, on contrast, are complex solutions with none or limited detailed documentation [99] and consideration by safety standards. Due to this, direct diagnosis becomes a challenge and several related research contributions consider it a "black channel" on top of which a Safe Communication Layer (SCL) can be deployed to support the safe communication among software tasks running in different cores [9, 12, 112]. A SCL can be used to provide a high DC of all possible communication errors that can occur in a "black channel," the interconnect or switch. With respect to NoCs, the survey [86] provides a summary of relevant contributions for fault detection in real-time NoCs.

Time Predictability. Generally, a *shared-bus* based on a round-robin scheduling policy has a a moderated impact on WCET variability and analysis complexity [48, 61]. But due to latency and bandwidth limitations, a shared-bus is generally used to interconnect a reduced set of cores. As shown in Table 1, a generic COTS *safety* device integrates 2 or 3 cores with a shared bus.

Interconnects integrated in COTS devices (e.g., P4080 CoreNet [9]), on contrast, have a considerable potential impact on WCET variability and not possible to limit due to the complexity and limited information available for analysis [26, 99]. As shown in Table 1, a *generic* COTS device integrates 2–8 cores with interconnect and switch solutions. The survey [128] describes relevant contributions for temporal predictability and WCET based on interconnect components.

However, as explained in the survey [85], a relevant research activity targets the development of real-time and time predictable *NoC* solutions for which a generic survey and quantitative comparison is provided by References [85, 86]. With respect to NoC communication traffic temporal predictability, two major technical approaches can be defined: full and virtual traffic separation.

Full traffic separation is the most conservative approach to establish time predictability for safety-critical subsystems by completely separating the safety-critical and low-critical traffics. In this approach, the separation of traffics is done in the spatial domain. A subset of resources is reserved for the safety-critical subsystems and no requests from low-critical subsystems will be handled by this subset. However, this approach introduces a low-efficient usage of the network resources, as some resources are blocked for safety-critical subsystems. For example:

- *Fully Disjoint Routes*: Hermes [92] is an example, which introduces a distributed fault-tolerant routing algorithm and utilizes load-balancing routing. This routing algorithm provides pre-configured alternative path selection for fault-tolerance and bypasses faulty network path/region/areas [92]. However, the performance degrades with the number of faulty links.
- *Circuit Switching*: Programmable NoC (PNoC) [87] is a lightweight flexible circuit-switched NoC, which supports a high communication data-path width, dynamic insertion/removal of nodes and guaranteed throughput with low communication jitter. However, although the authors tried to minimize the circuit establishment latency using simple communication protocols, this delay is not deterministic, because it depends on the availability of the resource.

- *Multiple Overlay Networks*: TilePro64 [7] is a complete SoC that offers distributed shared resources (e.g., memories, network controllers) using a packet-switched, wormhole-routed, point-to-point network. The on-chip interconnect network [194] establishes tile-to-memory, tile-to-tile, and tile-to-IO communication by transmitting packets across the network.

In *virtual traffic separation*, temporal partitioning is used to virtually establish the separation between the safety-critical messages and low-critical traffic. The aim of this separation is to eliminate interference between applications and to achieve composable services. This approach can be implemented by defining scheduled injection times (e.g., Time-Triggered solutions) or subdividing the bandwidth between the mentioned traffic types using TDM approach (e.g., Aetherial). In addition, rate controlling (e.g., Nostrum) and prioritized virtual channels (e.g., QNoC) can be used to enhance the interference between two traffic types. The first two approaches (i.e., scheduled injection time and TDM) rely strongly on a notion of global synchronicity. For example:

- *Time-Triggered Network-on-a-Chip (TTNoC)*: In time-triggered solutions, the transmission of messages is performed according to a predefined communication schedule in a periodic manner with a constant period and at a defined instant within the period (the phase) [154]. TTNoC [144, 154] introduces an on-chip time-triggered interconnect that supports the required time-predictability and fault-tolerance for mixed-criticality systems. However, it lacks support for the integration of legacy hardware and transmission of event-triggered messages.
- *Fixed TDM Traffic Slotting*: The Aetherial [68] NoC supports two distinct traffic classes of communication, Best-Effort Services (BES) and Guaranteed Services (GS). The GS offers uncorrupted, loss less, bounded latency, guaranteed throughput and ordered data delivery through resource (wires and buffers) reservation. This reservation is however according to the worst-case. The BES does not reserve any resources and uses unallocated and unused capacity and operates based on packet-switching flow control. In this context, GS establishes a time predictable communication infrastructure for safety-critical subsystems, while BES increases the resource efficiency and suits low-critical subsystems.
- *Priority-Arbitrated Virtual Channels*: The Quality-of-Service NoC (QNoC) [30] offers multiple levels of services (Quality of Service (QoS)) and uses simple architecture by a credit-based back-pressure flow-control. Packets associated to different classes of service (GS, BES) are delivered in an interleaved manner with associated QoS definition priorities such as throughput, relative priority and end-to-end delays. Although packets of high priority incur a low delay, low-priority flows suffer interference from higher priority flows [179].
- *Rate Controlling*: Rate controlling, or bandwidth management, is the process of measuring and regulating the communication (traffic, packets) on a network link to avoid overloading the link, which leads to network congestion or poor performance. This technique obviously cannot eliminate the interference but minimizes it. Nostrum [133] defines a packet-switched 2D mesh topology, which offers GS as well as BES using deflecting routing algorithm. It supports design-time bandwidth guarantees on a path by means of prioritized looped containers.

3.3.6 Specialized Components for Spatial Independence. To achieve multi-core device-level spatial independence in compliance with IEC 61508, several specialized components for spatial independence can be used, e.g., MMU. These components are described in Section 3.4.3.

3.4 Device Level

This section describes selected key research contributions toward device-level reliability, diagnostic coverage, spatial independence, and temporal independence.

3.4.1 Reliability. As explained in Section 2.2 and Equation (1), the device failure rate is composed by the addition of safe and dangerous failure rates of all components, or at least the components that take part in the execution of safety functions. Thus, device-level reliability is defined by the device architecture and building components reliability (e.g., Reliability Block Diagram (RBD) with serial/parallel composition of components).

Specialized device architectures such as HFT=1 (see Section 3.2.1) support on-chip redundancy with hardware fault-tolerance of one (e.g., 1oo2) [77, 81, 82], with a potential reliability improvement with respect to standard multi-core devices (e.g., 1oo1 (single channel)).

3.4.2 Diagnostic Coverage (DC). This section describes several diagnosis techniques specific to multi-core devices: temporal diagnosis, spatial diagnosis and safe start-up/shut-down.

- *Temporal diagnosis* is required to diagnose that safety related time constraints of interest are met and safely react/recover in case of error. For example, in *safety devices* it is required to detect unexpected errors due to systematic and random errors, and in *generic devices* for detecting rare event situations not considered in the design phase. For this purpose, several common application specific techniques can be used (e.g., time constraint monitoring, trapping unexpected interrupts, watchdog [12, 158, 159]) and run-time monitoring diagnosis (e.g., device performance counters [43], software framework [141]).
- *Spatial diagnosis* is required to ensure that data of a given function is not modified by another element or function. In a multi-core device, the parallel execution of software tasks that perform read and write operations in memory require a coherency management, to ensure that a write operation of a given data value is consistently updated for read(s) operation by tasks in other core(s) within a given time frame. Cache coherency management is generally performed by dedicated built-in components (e.g., Snoop Control Unit [111]) that ensure the coherency of distributed and hierarchical caches (e.g., L1/L2). Device built-in cache and memory coherency management are subject to systematic and random faults, thus, diagnosis is required to ensure spatial independence. Cache and memory coherency diagnosis can be performed by built-in hardware diagnosis [181], generic software techniques (e.g., References [12, 111]) or a combination of both. In addition to this, device built-in spatial protection components (e.g., MMU) and error detection and correction techniques (e.g., ECC) require periodic diagnosis to detect random or systematic faults in these components and techniques [113, 115, 159]. Finally, application specific software techniques based on information redundancy for application safety data areas might be used to increase the spatial diagnostic coverage (e.g., CRC, memory duplication, memory duplication with bit inversion [89]).
- *Safe start-up and shut-down* diagnosis is required to ensure that the device is started-up and shut-down in a safe, design-time-defined, and repeatable manner [115, 159], using internal and/or external hardware and software.

3.4.3 Spatial Independence. Key contributions with respect to device-level spatial independence can be classified in two major groups: achieving spatial independence with shared resources and specialized architectures and components that support spatial independence. Spatial diagnosis techniques and associated components are also required to detect errors, must be considered in reliability analysis and temporal independence analysis (see Sections 3.4.2 and 3.4.4).

Spatial Independence with Shared Resources. Ensuring exclusive access to addressable resources is a common basic technique to ensure spatial independence avoiding the threats associated to shared resources. In UMA and NUMA memory architectures where addressable resources are shared resources, the usage of MMU and Memory Protection Unit (MPU) components is a common technique to support the exclusive access to shared resources such as addressable peripherals and memories [155, 159]. Both components are common in state-of-the-art multi-core devices (see Table 1). Moreover, in *generic devices* shared components might be accessible by means of different buses and networks, thus, possible combinations shall also be analyzed and restricted [12]. In an UMA memory architecture with shared memories and local cache per core, coherency threats need to be mitigated managing the device-level coherency of shared data read and write transactions. When a shared data is locally updated, this change needs to be timely updated at device level by means of device specific cache coherency support [12, 111].

In addition to this, modern multi-core devices provide built-in security technology that in some cases can be used to support spatial independence. For example, TrustZone technology provides device-level hardware isolation for trusted software, which can be used to define a trusted execution environment for the safety software with exclusive access to assigned resources [115, 160].

Specialized Architectures and Components for Spatial Independence. In a NUMA memory architecture, where each core has a dedicated physical memory with exclusive access privileges, spatial independence of local memory can be potentially achieved by design [126]. This is a common approach used in NoC architectures such as specialized time independent architectures described in Section 3.4.4. This approach is also used in HFT of one solution (e.g., References [81, 82, 119]).

Device virtualization features enable the development of software-level solutions that support spatial independence among software partitions and tasks. For example:

- A hypervisor is a layer of software that uses device virtualization support to provide independent execution environments to software partitions (e.g., XtratuM, Linux-KVM, PikeOS, Jailhouse [42]). Hypervisor virtualization in combination with multi-core devices can be used for the development of mixed-criticality systems combining safety and non-safety partitions, including safety partitions of different criticality. The development of hypervisor solutions for mixed-criticality applications is an active field of research [12, 42, 75, 184] in combination with the usage of such technology with multi-core devices for the development of mixed-criticality safety systems [9, 10, 98, 157, 158, 159]. Gu [75] provides a survey of hypervisor virtualization solutions for real-time and safety-critical systems. Domain-specific safety certified COTS hypervisors are already commercially available (e.g., PikeOS, XtratuM).
- With respect to *real-time operating systems*, several research initiatives are analyzing container-based mechanisms, which are well-known in the security domain, to provide software isolation mechanisms similar to previously defined software partitions that can also be deployed in multi-core device architectures (e.g., Reference [15]). For example, several initiatives aim to pave the way toward safety certification of Linux (e.g., SIL2LinuxMP [15, 161]).

In addition, and based on this, specialized software strategies can be defined to handle specific spatial independence challenges such as shared addressable registers used to manage input, output, and communication peripherals. If more than one function requires the shared usage of such peripheral registers, then previous strategies to support exclusive access to memory addressable resources might not be valid. For example, digital input/output and communication

servers describe software solution patterns based on partitions that exclusively manage device registers and provide read/write services to other software partitions [12, 111, 167].

With respect to FPGAs, modern SRAM FPGAs provide different logical and spatial isolation mechanisms and supporting tools. FPGAs with multiple dies on a substrate can support the route and placement of redundant channels in different silicon die blocks [79]. In addition to this, isolation fences for spatial independence can be defined to provide logical and spatial isolation between logical blocks using specialized qualified tools such as Isolation Design Flow (IDF) [73, 115].

3.4.4 Temporal Independence. Key contributions with respect to device-level temporal independence can also be classified in two major groups: achieving temporal independence with shared resources and development of specialized architectures that provide or support temporal independence. In both cases, WCET estimation is required for timing verification [128] and temporal diagnosis to detect errors (see Section 3.4.2).

Temporal Independence with Shared Resources. As explained by Maiza et al. [128], Mitra et al. [134], and in Section 2.4, an average COTS device is based on a shared resources architecture. The concurrent access to shared resources and components that optimize average performance at the cost of time predictability, leads to potential temporal interference among executing software tasks that could jeopardize the required temporal independence [17, 47, 107, 116, 128, 134, 156, 159, 170] and WCET estimation with low bounds becomes a challenge [6, 128, 134].

Nonetheless, as explained in References [12, 158, 159], it is feasible to define application and device specific time independent solutions, using COTS *generic devices* with shared resources and associated intrinsic time interference sources [107, 134]. To achieve temporal independence, at least a design systematic approach and diagnostic coverage to detect temporal constraint violations are required. If an unexpected dangerous violation occurs, then the diagnosis error reaction should lead to safe-sate, which does not jeopardy safety but availability. The design systematic approach should at least consider analysis of device and components temporal characteristics such as WCET estimation, partitions scheduling, partition time slot assignment and interrupt sources management. The temporal diagnosis strategy should consider techniques such as monitoring of time constraints and trap unexpected interrupts. As summarized in Table 3, this approach has been extended in several domain and application specific safety concepts, e.g., railway [9], automotive [10].

But, this approach considers a simplistic scenario where a single fail-safe safety function is integrated. As explained by Paulitsch [155], when several fail-safe functions are integrated the requirement for function integrity and availability needs to be analyzed to ensure safe operation. A safety function might not be allowed to sacrifice the operation of another safety function due to the unforeseen consequences of the effect (e.g., cascading effects to other functions that could lead to dangerous pilot cognitive overload) [155]. Thus, these scenarios require also application and device specific solutions to achieve the required temporal independence.

Specialized Architectures for Temporal Independence. Several research contributions define specialized architectures and devices that provide rules, patterns, and building-blocks that can be used in combination with application software to achieve temporal independence:

- The Composable and Predictable Multi-Processor System on Chip (CoMPSoC) architecture [67] supports resources between applications on the same multi-core system-on-a-chip, while avoiding mutual interference. The temporal behavior of an application does not depend on other applications despite the sharing of on-chip resources. CoMPSoC employs the Aetherial NoC to ensure predictability for the communication. This NoC [69] uses time-division multiple-access and static resource allocation for the network interfaces,

Table 3. Summary of Selected Example Case-studies

Domain	Device	Description
Automotive	AURIX (TC27x)	A safety concept example for ISO 26262 ASILD compliant automotive cruise-control safety system based on an automotive *safety device* and virtualization technology (hypervisor) [10]
	AURIX (TC2xx)	Powertrain control example based on a *safety device* using the proposed software development and test environment [125].
	Hercules (TMS570)	Example safety Anti-Lock Braking System (ABS) implementation in a single-core (TMS470) and dual-core *safety device* (TMS570). Both implementations are compared with respect to safety, task scheduling, real-time performance and power consumption [100].
	Generic	Simplified state of the art of automotive COTS multi-core devices, operating systems and timing analysis tools. It proposes a legacy code migration pattern for automotive multi-core devices taking into consideration associated safety standards [118].
Avionics	QorIQ (T4240)	Design and time analysis of a mixed-criticality system that integrates safety and non-safety software applications based on a *generic device* and virtualization technology (hypervisor) [12].
	QorIQ (P4080), ARM A15, Cyclone V, C6000	Selection criteria and assessment of multiple COTS devices against avionics requirements, together with suggestions for their appropriate use and modifications to EASA guidelines [99].
	QorIQ (P4080)	Qualitative and quantitative assessments of the use of this COTS multi-core device for avionic systems, with emphasis on timing analysis [143] and compliance against CAST-32A guidance [8].
	MPPA-256 Bostan	Assessment of software (PikeOS) and MPPA-256 device against avionics requirements in the context of multi-core devices [170].
	LEON3+	Part of the flight control system is evaluated on a 4-core LEON3 design implemented on a FPGA, together with ARINC 653 compliant PikeOS in the context of probabilistic timing analysis [37].
Railway	QorIQ (P4080)	Safety concept for EN 5012x SIL4 on-board railway signalling system based on a *generic device* and hypervisor technology [9].
Industrial Control	Core 2 Duo Processor	Industrial robot controller that integrates mixed-criticality functions such as real-time control and IEC 61508 SIL2 safety soft-PLC in a dual core processor [3].
	Intel Atom (x86) and LEON3FT	Safety certification strategy and safety concept examples for IEC 61508 SIL3 compliant wind-turbine safety system based on *hybrid safety devices* and hypervisor technology [110, 158, 159, 184].
	Zynq-7000 (ARM, MicroBlaze)	Safety concept for IEC 61508 SIL3 compliant wind-turbine safety system based on *generic device* and hypervisor technology [158, 159, 184], extended with certification strategy that supports variability, product line and modular safety cases [12].
Medical Systems	Zynq-7000	Real-time assessment of dependable health-care case study that integrates Linux-KVM with Memguard, on-chip communication and security services [12].
	Several	Overview of processor fault-tolerant techniques and solutions that target medical applications [146]

where time-slots remain empty in case no data needs to be transmitted by a network interface.

- The DREAMS architecture [12] provides structuring rules according to several integration levels. At the chip-level, a multi-core chip is decomposed into tiles that are interconnected by a NoC, where each tile is composed by several cores with associated caches, memories and I/O resources. The architecture defines generic platform services as a baseline for the development of applications including a integrated resource management and fault-tolerant global time base, based on which secure communication and execution of services are deployed. The architecture avoids interference of shared memory resources using MemGuard with a dynamic reclaim mechanism [185]. Each tile is allocated a certain bandwidth for each period and unused bandwidth is assigned to a global repository, which can be exploited by other cores.
- T-CREST is a time predictable multi-core architecture based on Patmos time predictable processor core, specially designed cache for WCET, scratchpad memory and time predictable TDM-based NoC [32, 166, 174, 175].
- The Multi-core Execution of Parallelised Hard Real-time Applications Supporting Analysability (parMERASA) architecture [31] provides an execution environment for hard real-time applications on a scalable multi-core processor. The parMERASA multi-core processor contains scalable timing analyzable NoCs. Processor cores are organized into clusters that are connected with a dedicated NoC [186]. It includes an extended on-demand coherent cache controller [167] that enables tight static WCET analysis [150]. This cache prevents unintended interference between cores by distinguishing between private and shared cache lines.
- The LEON-based Probabilistically Analyzable pRocessor Design (LEOPARD) architecture [84], implemented in a commercial LEON3-based multi-core for the space domain—although applicable to any domain—provides specific timing properties needed for probabilistic timing analysis. In particular, LEOPARD randomizes shared bus arbitration and cache placement and replacement policies. It also upper bounds the latency of some input dependent floating-point operations such as division and square root. This facilitates probabilistic timing analysis [38] and relieves developers from having to control design details such as memory placement and how tasks in different cores share hardware resources.

Worst Case Execution Time (WCET). The WCET estimation challenge can be viewed from different angles, each one with different paradigms to address the challenge:

- *Static vs. measurement-based timing analysis.* Static timing analysis builds upon a timing model of the device where real-time applications should be executed, and performs abstract interpretation of the execution of the software on the device. This way, static timing analysis determines or, quite often, upper bounds the latency of the instructions, basic blocks, and functions of the application under analysis on such a device. Static timing analysis reliability and tightness strongly depend on the accuracy and reliability of the timing model of the device, as well as on the amount of—detailed—information obtained from the application under analysis, mostly related to memory placement and timing of the events. Conversely, measurement-based timing analysis relies on actual execution time measurements of the application running on the actual device, thus avoiding any reliability and tightness issue due to models and detailed information for abstract interpretation. However, measurement-based timing analysis reliability strongly depends on the representativeness of the tests executed w.r.t. the WCET, as well as on the control exercised on the initial conditions of the device.

- *Deterministic vs. probabilistic timing analysis.* Deterministic timing analysis aims at providing an absolute WCET bound that cannot be exceeded under any circumstance. However, the residual risk of exceeding that bound is unknown and not null by construction. Conversely, probabilistic timing analysis aims at delivering a probabilistic distribution intended to upper bound the real execution time distribution so that for any given exceedance threshold (e.g., 10^{-12} per run) a probabilistic WCET bound is obtained. In general, the latter, probabilistic timing analysis, offers a more convenient and flexible trade-off to obtain WCET estimates, but relies on some timing characteristics of the device and/or the application that may not hold unless appropriate hardware or software support is provided.

A plethora of works have analyzed and reviewed the different WCET estimation paradigms. The seminal work by Wilhelm et al. [195] compares static and measurement-based approaches, and hybrids thereof. An updated taxonomy of the state-of-the-art with focus on multi-core aspects for timing analysis can be found in Reference [60]. A comparative perspective of all paradigms above with specific discussion for multi-core devices has been given by Abella et al. [6], whereas a deeper analysis for multi-core aspects can be found in Reference [128]. Puschner [165], instead, considers the multi-core problem with particular focus on the management of mixed criticalities. Finally, tooling aspects for the different paradigms have also been specifically addressed [195, 203].

Measurement-based (deterministic) timing analysis has been object of significant attention with the development of multiple methods applied to specific safety regulations and domains such as automotive [173], avionics [123], and industrial control [110] domains. It is also a common industrial practice. The hardware and software support needed to facilitate deterministic timing analysis [23, 140] as well as probabilistic timing analysis [38] has also been object of deep analysis.

Note that quantitative comparisons across devices are not provided, since WCET estimates are only partially dependent on the device used, its components and the nanoscale implementation aspects. Instead, WCET estimate trustworthiness and tightness strongly depend on how the device is configured for use (e.g., number of cores active, cache replacement policies used) and how the software is deployed—which limits potential interference, and, more importantly, on the particular timing analysis method used, where the cost to apply it, and the amount and type of input needed to apply it, vary dramatically across methods [6]. Overall, to the best of our knowledge, relevant quantitative comparisons across devices on the same ground (i.e., comparable configuration settings, same application, same timing analysis method) with realistic assumptions do not exist, and any conclusion for such a comparison could only be extremely dependent on the settings, applications, and timing analysis method used, thus precluding from obtaining a general conclusion for devices compared. Instead, only deep—but qualitative—comparisons exist for few devices (e.g., Reference [99]).

Software Scheduling. Generally, safety standards demand simple, static, and design-time-defined scheduling solutions. For example, industrial safety standards such as IEC 61508-3 recommend the usage of design-time deterministic scheduling methods (e.g., cycling scheduling with pre-assigned time slots) and strict priority scheduling by means of priority inversion avoidance [89, 159]. Hypervisors, such as XtratuM, support design-time cyclic scheduling policies for partitions [42].

To improve the efficiency of resource sharing, and therefore the overall system average performance, several scheduling solutions have been proposed for mixed-criticality systems as summarized in the survey [33]. This is an active research topic in the scheduling research community that combines functions of different criticality to be scheduled with multi-core devices. However, as explained by Paulitsch et al. [155], with respect to industrial safety systems the contribution of this research is closer to the concepts of "survivability" and "graceful degradation" [74] rather than criticality and safety compliance (with respect to safety standards such as IEC 61508).

4 DOMAIN-SPECIFIC APPROACHES

This section extends the survey with other domains that do not use IEC 61508 as reference safety standard, avionics and aerospace domain. It also provides a summary of selected domain-specific relevant example case-studies.

4.1 Avionics Domain

Avionic systems are subject to very strict certification processes that need to be approved by specific certification authorities. Thus, certification of safety-related systems based on multi-core devices in avionics is often more challenging than for other domains and, moreover, the avionic industry is even more reluctant than other industries to redesign already certified software.

The Integrated Modular Avionics (IMA) architecture, used by most avionic systems in the last decade, has enabled the integration of more functions in fewer computers, thus a reduction of computer systems SWaP. However, increasing performance demands in avionics (e.g., Airbus A-380 has 80 Mbytes code size [5]) cannot longer be satisfied with single core processors [171], and multi-core devices become mandatory to increase performance without increasing the number of on-board computers [1, 12, 17, 170, 171].

However, the multi-core integration (multi-IMA) of legacy single-core IMA systems is a certification challenge, because legacy partitions temporal and spatial isolation must be guaranteed without incurring in huge re-certification costs [1, 51, 52, 62, 88, 99, 104, 143]. The updated ARINC 653 supports the usage of multi-core devices but restricting the execution of safety partitions to a single core at a time with the remaining cores disabled, which is against SWaP reduction.

CAST-32A position paper [36], developed by avionics certification authorities, provides some guidance on how to use multi-core devices for safety-critical systems, allowing multiple cores to be used while running at least one safety partition, but there is still a significant gap between those guidelines and what COTS multi-core devices offer in terms of partition and control of interference channels. This gap relates to the abstract objectives provided in the CAST-32A position paper, which are hard to match—if at all possible—with the actual characteristics of COTS multi-core devices offering the performance level needed by avionic systems. Agirre et al. [8] analyze qualitatively and quantitatively such gap for the particular case of the NXP P4080 multi-core device, showing that COTS multi-core device design in general, and for the NXP P4080 device in particular, is at odds with the objectives imposed by CAST-32A position paper.

Efforts toward easing the adoption of multi-core devices in avionics safety-related systems have been abundant. Paulitsch et al. [17, 155] provide a research state-of-the-art description of mixed-criticality systems, with insights into real avionics systems. Some authors aim at characterizing the timing characteristics of shared hardware resources in COTS multi-core devices, either from a hardware perspective [143] or from an application perspective [27]. The required hypervisor support needed to limit interference due to contention in shared resources has also been evaluated and multiple alternatives provided [65, 98] as well as IMA compliant solutions for timing analysis of legacy code [138]. Agrawal et al. [11] provide a specific solution to manage memory bandwidth dynamically on COTS multi-core devices while still preserving a sufficient degree of time predictability. The impact of fault-tolerance mechanisms on time predictability for multi-core devices has also received some attention [123], since this aspect is too often overlooked. Finally, some authors consider the use of probabilistic timing analysis as a way of relieving end users from the cumbersome task of mastering execution time variability in complex multi-core devices [38, 189, 190].

4.2 Space Domain

Similarly to the avionics domain, the space domain has started building its systems on the IMA for Space (IMA-SP) architecture to facilitate the consolidation of multiple functions onto a single

computer for SWaP reduction. However, as for the avionics domain, IMA-SP does not provide explicit solutions for the integration of multiple safety partitions on multi-core devices.

Space systems are subject to a number of specific requirements far different from those of terrestrial ones [4]. For example, legacy code is relatively scarce, since each mission is basically unique and they require higher levels of fault tolerance support, because they are generally exposed to much higher levels of radiation than terrestrial systems. However, as in many other safety-related domains, performance demands grow sustainedly, with code size scaling by a factor of 10× every 10 years, with missions in 2010 already reaching few millions of lines of code [4].

These characteristics impose the use of high-performance radiation-hardened hardware devices for spacecraft control and process data retrieved through instruments, as well as the consideration of approaches for system verification and validation such as timing aspects of multi-core devices. Such radiation-hardened devices, either in the form of ASIC processors or FPGA implementations, have been developed for several space missions (e.g., dual-core LEON3-FT [96]). In this context, multi-core devices are becoming dominant, with space industry considering increasingly parallel hardware devices such as the Next-generation Microprocessor (NGMP) [106]. The landscape of future computing devices for space for on-board processing is analyzed in Reference [63]. Apart from the aforementioned NGMP device, authors also identify the RAD5500 processor family, a radiation-hardened processor based on the e5500 core of the QorIQ Power Architecture processor. For instance, the RAD5545 processor employs four RAD5500 cores and uses 45-nm technology [63].

A detailed survey of the available (in 2012) space processors for space missions is given in Reference [64], including radiation and non-radiation hardened processors, as well as those with no redundancy, dual-modular redundancy and triple-modular redundancy for fault tolerance. From a different perspective, other works aim at reconciling safety and security concerns for Integrated Modular Avionics for Space (IMA-SP) systems, but still without considering multi-core devices [196].

4.3 Example Case Studies

Table 3 provides a summary of selected domain-specific relevant example case-studies based on a variety of multi-core devices and covering several fundamental safety technical requirements.

5 CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This section describes links to other related research topics, overall conclusion and future research directions.

5.1 Links to Other Research Topics

When several software applications of different criticality are integrated in a multi-core device, multiple challenges need to be addressed and managed, to support the cost-effective safety certification of the given integrated solution. This section explores the links between the surveyed research scope and some additional research challenges to be considered.

5.1.1 Parallel Application Software. The development of software applications for multi-core devices needs to consider the parallel/concurrent execution of multiple software applications on a single device with shared resources.

- Software development: Previously described hypervisor virtualization and operating system extensions that support partitioning provide a foundation for the integration of software partitions with parallel/concurrent execution [159].

- Software re-usability: The re-usability of sequential legacy code (mono-core) becomes a challenge, because it needs to consider both the parallel execution of software applications and the usage of shared resources [107, 126, 183]. For example, Macher [126] describes generic steps that can be used to migrate legacy software to multi-core devices.
- Software integration: The integration of multiple mixed-criticality parallel application software partitions on multi-core devices is another field of research with the development of technical solutions (e.g., hypervisor and partitions) and strategies at system (e.g., Reference [158]) and on-chip level (e.g., performance and run-time monitors [43]).

5.1.2 Parallel Programming Models. The definition of heterogeneous and parallel programming languages [50] with support for mixed-criticality safety systems is a challenge to be addressed [169], which could help exploit the performance delivered by multi-core devices enabling the execution of parallel software applications, rather than only multiple sequential applications in parallel. In this context, OpenMP has been regarded as a potentially appropriate solution, enabling functional verification, and being already supported by Keystone II and MPPA multi-core devices [169]. Regarding timing verification for real-time systems, two different OpenMP models have been considered so far in the literature, namely, fork-join [109] and tasking [178] models, being the latter the one gaining more popularity due to its ability to materialize parallelism even at fine granularity, and not being restricted to only structured parallelism. Irregular task graphs can also be parallelized with the tasking model, and research initiatives aim at leveraging support for heterogeneous and asynchronous parallelism to deliver time predictability [168].

5.1.3 Heterogeneous Computing Platforms. In addition to multi-core devices, heterogeneous computing platforms including GPUs and application specific architectures (e.g., tensor cores) can potentially be used for the development of mixed-criticality systems. For example, GPUs provide high performance computing platforms for application domains such autonomous driving solutions [14, 137] (e.g., NVIDIA Xavier ASILC)). Mittal et. al [137] provide a survey of GPU-based heterogeneous computing techniques and Alcaide et al. [14] an analysis and techniques for the usage of GPUs in high-integrity autonomous driving solutions with respect to ISO 26262 standard.

Besides, the development of *artificial intelligence*-based autonomous systems is leading to open research problems such as the distribution of learning and inference activities over heterogeneous computing platforms. Upcoming SoC-FPGAs platforms (e.g., Xilinx Versal) combine these heterogeneous resources, but challenges remain with respect to hardware support for safety-critical systems such as predictable interconnects, avoidance of temporal interference in memory and safety monitors. For example, while the portability to different GPU architectures and programming interfaces was addressed in prior work [130], portability to other resource types and the simultaneous usage of heterogeneous computing resources is also considered a challenge, with few works currently addressing this challenge [164].

5.1.4 Product Line and Modular Certification Strategies. During the product lifetime different aspects need to be considered to support complexity management [105] and support the incremental certification of systems [12], e.g., product line and variability management [12], periodic updates of hardware that require re-certification (e.g., 10–15 years in automotive [39]), modular safety cases, and modular certification approaches [12, 113, 115].

5.1.5 Security/Cybersecurity. Safety and security requirements are common in multiple domains ("no safety without security") such as avionics [17, 155, 176], control systems [108], automotive [39], and railway [25, 176]. In addition to this, the industrial trend toward Industry 4.0 and digitalization, and the automotive trend toward autonomous vehicles, further exacerbates the need to comply with safety and security requirements. The definition of security standards

compatible with safety standards development processes (e.g., IEC 62443 [90] and IEC 61508) and the definition of cross-domain verification and validation methods [176], paves the way toward safety and security systems certification. From a technical perspective this is an active field of research [108]. For example, the Multiple Independent Levels of Security/Safety (MILS) architectural approach combines safety and security requirements in different domains [155] and it is supported by several hypervisor solutions (e.g., XtratuM). In addition to this, several COTS *generic* and *safety* devices already provide support for the development of safe and secure applications.

Other important aspect is related to software updates. From a security point of view, it is essential to keep the system up-to-date with the latest security patches, what requires regular software updates. On the contrary, the service life of safety-critical systems is rather static, where operation time modifications involve a well established procedure that can not be trivially applied to frequent updates. Accordingly, a safety and security co-engineering is essential for the development and maintenance of such systems with *over-the-air software updates* [139].

5.1.6 Energy, Power, and Thermal Management. As CMOS technology downscales, potentially more transistors can fit on the same silicon die area. However, power, thermal, and energy-efficiency constraints limit this potential transistor density increase [46]. This design paradox is referred to as "dark silicon" [58]. To overcome these constraints, different approaches are considered such as thermally aware chiplets [57] and balance between fault-tolerance and power consumption [127]. Energy and power consumption is a critical factor for diverse safety-critical systems such as portable medical devices [146] and railway signalling autonomous remote object controllers [59]. In this context, Fakih et al. [59] describe techniques and approaches for energy, power and thermal management in the development of multi-core device-based mixed-criticality systems, where described low power techniques must not jeopardize integrated safety functions.

5.1.7 Integrated Development Environments (IDEs) and Tools. The development of multi-core device-based mixed-criticality systems require the development or extension of IDEs and tools to support the complexity management derived from the integration of software functions of different criticality allocated to one or several cores (e.g., partition allocation) taking into consideration several technical constraints such as heterogeneous computing platforms, hypervisor configuration and partition scheduling, real-time guarantees, safety constraints, diagnosis, WCET, parallel programming models, product line variability and design space exploration [12]. For example, several research projects (e.g., ARAMiS II, EMC2, DREAMS) have addressed this challenge. A tool platform with common technical interfaces and seamless work-flows was defined in the ARAMiS II project, with a focus on development environments supporting partitioning, allocation, binding, scheduling and design space exploration for software components on multi-core devices [192]. Application models and design tools for multi-core device-based mixed-criticality systems were defined in the EMC2 project [193]. For example, the developed tools support the optimized allocation of a system's functionality to a target hardware architecture, while ensuring bounded interference between different components in mixed-criticality settings. Development process and tools for mixed-criticality systems with networked multi-core devices were defined in the DREAMS project [12]. The tools support model-transformations, scheduling and design space exploration for mixed-criticality systems based on multi-core devices and virtualization technology.

5.2 Conclusions and Future Research Directions

The integration of functions of different safety criticality (e.g., safety and non-safety) in a multi-core device leads to safety certification challenges. This survey has categorized and summarized at different device abstraction levels (nanoscale, component, and device) the state-of-the-art of selected key research contributions that support the compliance with fundamental safety technical

requirements (reliability, diagnostic coverage, spatial, and temporal independence). Most of the surveyed research state-of-the-art can be classified as either "how things should be done" (e.g., temporal independence with T-CREST time predictable device) or "how things can be done" (e.g., how to achieve temporal independence with current COTS *generic devices*). Taking into consideration the business oriented and innovative nature of the semiconductor industry and the required conservative nature of safety standards, in the future this divergent trend is expected to even increase.

With respect to safety certification, as described by the research state-of-the-art and example case-studies summarized in this survey, it is technically feasible to develop multi-core device-based mixed-criticality safety systems. However, the techniques and strategies described in the case-studies have a high dependency with the integrated multi-core device and software applications (ad hoc solutions). Therefore, there is a need to consider cross-level and cross fundamental technical safety requirements research contributions. Although re-certification costs are also high, periodic system updates are commonly required due to hardware obsolescence, software updates and security patches. This leads to a need to manage either device stocks for the product life period [116] and/or techniques to decouple the dependency among the multi-core device and software applications, by means of techniques such as software partitioning and virtualization, safety and security co-engineering, safety product lines, modular safety cases, and overall complexity management.

In addition to this, compliance with time independence requirement and supporting time predictability in multi-core devices is a technical and research challenge. The availability and development of novel architectures and components that provide time predictability (e.g., T-CREST) can simplify the WCET estimation and overall safety argumentation with respect to temporal independence. However, shared-resources-based COTS multi-core devices are not generally designed with a focus on hard real-time applications and time predictability. Thus, there is a need for the definition of novel WCET analysis techniques, temporal diagnosis techniques, and temporal independence strategies. In addition to this, parallel programming languages and parallel software application techniques that consider the parallel/concurrent execution of software are required for new software development, software re-usability, and software integration.

Finally, as described by the International Technology Roadmap for Semiconductors (ITRS) [94], multi-core devices computational performance will continue increasing with further innovation and research in several fields such as DSM technologies, thermal management and device architectures. This will further require the development of novel hardware and software fault-tolerance, diagnosis techniques and strategies, which will also need to consider device energy-power-thermal management requirements and temporal-predictability restrictions.

And this innovation is required for the development of next-generation industrial and transportation systems that integrate autonomous systems, intelligent ADAS, machine learning, and Industry 4.0 solutions, which will further accelerate the ever-increasing demand for additional functions to be integrated in a reduced number of computing nodes and processing devices with higher computational and on-chip communication performance. And this integration trend will further exacerbate current system integration engineering challenges, such as off-chip communication capabilities, distribution of applications and services, distributed resource management, system-level global-time and scheduling, IDEs and tools, system-level safety, and security engineering [12].

REFERENCES

[1] Xavier Jean, Marc Gatti, Guy Berthon, and Marc Fumey. 2012. *MULCORS-Use of Multicore Processors in Airborne Systems (EASA Project. 2011/6)*. Technical Report. EASA. Retrieved from https://www.easa.europa.eu/sites/default/files/dfu/CCC_12_006898-REV07%20-%20MULCORS%20Final%20Report.pdf.

[2] WSTS. 2018. World semiconductor trade statistics (WSTS). Retrieved from https://www.wsts.org/.
[3] Intel. 2020. Case Study, Intel Core 2 Duo Processor on KUKA Robot Controller. Retrieved from https://www.intel.com/content/dam/doc/case-study/industrial-core-kuka-study.pdf.
[4] J. Abella and F. J. Cazorla. 2017. Chapter 9-Harsh computing in the space domain. In *Rugged Embedded Syst.*, Augusto Vega, Pradip Bose, and Alper Buyuktosunoglu (Eds.). Morgan Kaufmann, Boston, 267–293. DOI: https://doi.org/10.1016/B978-0-12-802459-1.00009-9
[5] J. Abella, F. J. Cazorla, E. Quiñones, A. Grasset, S. Yehia, P. Bonnot, D. Gizopoulos, R. Mariani, and G. Bernat. 2011. Towards improved survivability in safety-critical systems. In *Proceedings of the IEEE 17th International On-Line Testing Symposium (IOLTS'11)*. 240–245.
[6] J. Abella, C. Hernandez, E. Quiñones, F. J. Cazorla, P. R. Commy, M. Azkarate-Askasua, J. Perez, E. Mezzetti, and T. Vardanega. 2015. WCET analysis methods: Pitfalls and challenges on their trustworthiness. In *Proceedings of the 10th IEEE International Symposium on Industrial Embedded Systems (SIES'15)*. DOI: https://doi.org/10.1109/SIES.2015.7185039
[7] A. Agarwal, L. Bao, J. Brownet al. 2007. The tile processor (TM) architecture: Embedded multicore for networking and digital multimedia. In *Proceedings of the 19th IEEE Hot Chips Symposium (HCS'07)*. 1–12. DOI: https://doi.org/10.1109/HOTCHIPS.2007.7482495
[8] I. Agirre, J. Abella, M. Azkarate-Askasua, and F. J. Cazorla. 2017. On the tailoring of CAST-32A certification guidance to real COTS multicore architectures. In *Proceedings of the 12th IEEE International Symposium on Industrial Embedded Systems (SIES'17)*. 1–8.
[9] I. Agirre, M. Azkarate-Askasua, A. Larrucea, J. Perez, T. Vardanega, and F. J. Cazorla. 2015. A safety concept for a railway mixed-criticality embedded system based on multicore partitioning. In *Proceedings of the IEEE International Conference on Computers and Information Technology; Ubiquitous Computer and Communication; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICom'15)*. 1780–1787. DOI: https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.268
[10] I. Agirre, M. Azkarate-askasua, A. Larrucea, J. Perez, T. Vardanega, and F. J. Cazorla. 2016. Automotive safety concept definition for mixed-criticality integration on a COTS multicore. In *Computer Safety, Reliability, and Security*, A. Skavhaug, J. Guiochet, E. Schoitsch, and F. Bitsch (Eds.). Springer, 273–285.
[11] A. Agrawal, G. Fohler, J. Freitag, J. Nowotsch, S. Uhrig, and M. Paulitsch. 2017. Contention-aware dynamic memory bandwidth isolation with predictability in COTS multicores: An avionics case study. In *Proceedings of the 29th Euromicro Conference on Real-Time Systems (ECRTS'17)*.
[12] H. Ahmadian, R. Obermaisser, and J. Perez. 2018. *Distributed Real-Time Architecture for Mixed-Criticality Systems*. CRC Press, Taylor & Francis Incorporated.
[13] M. A. Alam, K. Roy, and C. Augustine. 2011. Reliability and process-variation aware design of integrated circuits—A broader perspective. In *Proceedings of the International Reliability Physics Symposium* 4A.1.1–4A.1.11. DOI: https://doi.org/10.1109/IRPS.2011.5784500
[14] S. Alcaide, L. Kosmidis, C. Hernandez, and J. Abella. 2019. High-integrity GPU designs for critical real-time automotive systems. In *Proceedings of the Conference on Design, Automation & Test (DATE'19)*. 824–829. DOI: https://doi.org/10.23919/DATE.2019.8715177
[15] I. Allende, N. Mc Guire, J. Perez, L. G. Monsalve, N. Uriarte, and Obermaisser R.2019. Towards Linux for the development of mixed-criticality embedded systems based on multi-core devices. In *Proceedings of the 15th European Dependable Computing Conference (EDCC'19)*. 47–54. DOI: https://doi.org/10.1109/EDCC.2019.00020
[16] B. Anuradha and C. Vivekanandan. 2012. Usage of scratchpad memory in embedded systems—State of art. In *Proceedings of the 3rd International Conference on Computing, Communication, and Networking Technologies (ICCCNT'12)*. 1–5. DOI: https://doi.org/10.1109/ICCCNT.2012.6396100
[17] J. Athavale, R. Mariani, and M. Paulitsch. 2019. Flight safety certification implications for complex multi-core processor based avionics systems. In *Proceedings of the IEEE 25th International Symposium on On-Line Testing and Robust Syst. Design (IOLTS'19)*. 38–39. DOI: https://doi.org/10.1109/IOLTS.2019.8854415
[18] A. Avižienis, J. C. Laprie, B. Randell, and C. Landwehr. 2004. Basic concepts and taxonomy of dependable and secure computing. In *IEEE Transactions on Dependable and Secure Computing*, Vol. 1. 11–33.
[19] A. Barros and L. M. Pinho. 2011. Software transactional memory as a building block for parallel embedded real-time systems. In *Proceedings of the 37th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. 251–255. DOI: https://doi.org/10.1109/SEAA.2011.46
[20] H. Bauer, M. Patel, N. Santhanam, and B. Wiseman. 2015. *McKinsey on Semiconductors*. Technical Report.
[21] C. Belwal and A. M. K. Cheng. 2011. Lazy versus eager conflict detection in software transactional memory: A real-time schedulability perspective. *IEEE Embed. Syst. Lett.* 3, 1 (2011), 37–41. DOI: https://doi.org/10.1109/LES.2010.2099104
[22] Pedro Benedicte, Carles Hernandez, Jaume Abella, and Francisco J. Cazorla. 2018. HWP: Hardware support to reconcile cache energy, complexity, performance and WCET estimates in multicore real-time systems. In *Proceedings of*

the 30th Euromicro Conference on Real-Time Systems (ECRTS'18). 3:1–3:22. DOI: https://doi.org/10.4230/LIPIcs.ECRTS.2018.3

[23] S. Bensalem, K. Goossens, C. M. Kirsch, R. Obermaisser, E. A. Lee, and J. Sifakis. 2011. Time-predictable and composable architectures for dependable embedded systems. In *Proceedings of the 9th ACM International Conference on Embedded Software (EMSOFT'11)*. 351–352. DOI: https://doi.org/10.1145/2038642.2038697

[24] C. Bernardeschi, L. Cassano, and A. Domenici. 2015. SRAM-based FPGA systems for safety-critical applications: A survey on design standards and proposed methodologies. *J. Comput. Sci. Technol.* 30, 2 (2015), 373–390. DOI: https://doi.org/10.1007/s11390-015-1530-5

[25] A. Bilbao, I. Yarza, J. L. Montero, M. Azkarate-askasua, and N. Gonzalez. 2017. A railway safety and security concept for low-power mixed-criticality systems. In *Proceedings of the IEEE 15th International Conference on Industrial Informatics (INDIN)*. 59–64. DOI: https://doi.org/10.1109/INDIN.2017.8104747

[26] J. Bin. 2014. *Controlling Execution Time Variability Using COTS for Safety Critical Systems*. Thesis, Université Paris-Sud.

[27] J. Bin, S. Girbal, D. Gracia Pérez, A. Grasset, and A. Merigot. 2014. Studying co-running avionic real-time applications on multi-core COTS architectures. In *Proceedings of the Embedded Real Time Software and Systems Conference*

[28] G. Blake, R. G. Dreslinski, and T. Mudge. 2009. A survey of multicore processors. *IEEE Signal Process. Mag.* 26, 6 (2009), 26–37. DOI: https://doi.org/10.1109/MSP.2009.934110

[29] C. Bolchini, M. K. Michael, A. Miele, and S. Neophytou. 2018. *Dependability Threats*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9

[30] E. Bolotin, I. Cidon, R. Ginosar, and A. Kolodny. 2004. QNoC: QoS architecture and design process for network on chip. *J. Syst. Archit.* 50, 2-3 (2004), 105–128. DOI: https://doi.org/10.1016/j.sysarc.2003.07.004

[31] C. Bradatsch, F. Kluge, and T. Ungerer. 2013. A cross-domain system architecture for embedded hard real-time many-core systems. In *Proceedings of the IEEE 10th International Conference on High Performance Computing and Communications (HPCC'13)*. 2034–2041. DOI: https://doi.org/10.1109/HPCC.and.EUC.2013.293

[32] D. Bui, E. Lee, I. Liu, H. Patel, and J. Reineke. 2011. Temporal isolation on multiprocessing architectures. In *Proceedings of the 48th ACM/EDAC/IEEE Design Automation Conference (DAC)*. 274–279.

[33] A. Burns and R. I. David. 2018. A survey of research into mixed criticality systems. *ACM Comput. Surv.* 50, 6 (2018). DOI: https://doi.org/10.1145/3131347

[34] A. Burns, J. Harbin, and L.S. Indrusiak. 2014. A wormhole NoC protocol for mixed criticality systems. In *Proceedings of the IEEE Real-Time Syst. Symposium (RTSS'14)*. 184–195. DOI: https://doi.org/10.1109/RTSS.2014.13

[35] D. Buttle. 2012. *Real-Time in the Prime-Time—ECRTS Keynote Talk*. Technical Report. ETAS Gmbh.

[36] CAST. 2016. *Multi-core Processors—Position Paper CAST-32A*. Technical Report.

[37] F. J. Cazorla, J. Abella, J. Andersson, T. Vardanega, F. Vatrinet, I. Bate, I. Broster, M. Azkarate-Askasua, F. Wartel, L. Cucu, F. Cros, G. Farrall, A. Gogonel, A. Gianarro, B. Triquet, C. Hernandez, C. Lo, C. Maxim, D. Morales, E. Quinones, E. Mezzetti, L. Kosmidis, I. Aguirre, M. Fernandez, M. Slijepcevic, P. Conmy, and W. Talaboulma. 2016. PROXIMA: Improving measurement-based timing analysis through randomisation and probabilistic analysis. In *Proceedings of the Euromicro Conference on Digital Systems Design (DSD'16)*. 276–285. DOI: https://doi.org/10.1109/DSD.2016.22

[38] F. J. Cazorla, L. Kosmidis, E. Mezzetti, C. Hernandez, J. Abella, and T. Vardanega. 2019. Probabilistic worst-case timing analysis: Taxonomy and comprehensive survey. *ACM Comput. Surv.* 52, 1 (2019). DOI: https://doi.org/10.1145/3301283

[39] S. Chakraborty and S. Ramesh. 2015. Guest editorial special section on automotive embedded systems and software. *IEEE Trans. Comput.-Aided Design Integr. Circ. Syst.* 34, 11 (2015), 1701–1703. DOI: https://doi.org/10.1109/TCAD.2015.2488378

[40] Jason A. Cheatham, John M. Emmert, and Stan Baumgart. 2006. A survey of fault tolerant methodologies for FPGAs. *ACM Trans. Des. Autom. Electron. Syst.* 11, 2 (2006), 501–533. DOI: https://doi.org/10.1145/1142155.1142167

[41] G. Corradi. 2017. Tools, architectures and trends on industrial all programmable heterogeneous MPSoC (KeyNote). In *Proceedings of the 29th Euromicro Conference on Real-Time Systems (ECRTS'17)*.

[42] A. Crespo, P. Balbastre, K. Chappuis, J. Coronel, J. Fanguède, P. Lucas, and J. Perez. 2018. *Execution Environment*. CRC Press.

[43] A. Crespo, P. Balbastre, J. Sim, J. Coronel, D. Gracia Perez, and P. Bonnot. 2018. Hypervisor-based multicore feedback control of mixed-criticality systems. *IEEE Access* 6 (2018), 50627–50640.

[44] N. Dagieu, A. Spyridakis, and D. Raho. 2016. Memguard: A memory bandwith management in mixed criticality virtualized systems memguard KVM scheduling. In *Proceedings of the 10th International Conference on Mobile Ubiquitous Computing, Systems, Services, and Technologies (UBICOMM'16)*. Retrieved from https://www.thinkmind.org/index.php?view=article&articleid=ubicomm_2016_1_40_10072.

[45] Shidhartha Das. 2009. *Razor: A Variability-tolerant Design Methodology for Low-power and Robust Computing*. Thesis, The University of Michigan.

[46] S. Das. 2018. *Variation-Mitigation for Reliable, Dependable and Energy-Efficient Future System Design*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9

[47] D. Dasari, B. Akesson, V. Nélis, M. A. Awan, and S. M. Petters. 2013. Identifying the sources of unpredictability in COTS-based multicore systems. In *Proceedings of the 8th IEEE International Symposium on Industrial Embedded Systems (SIES)*. 39–48. DOI: https://doi.org/10.1109/SIES.2013.6601469

[48] D. Dasari and V. Nelis. 2012. An analysis of the impact of bus contention on the WCET in multicores. In *Proceedings of the IEEE 14th International Conference on High Performance Computing and Communications (HPCC'12)*. 1450–1457. DOI: https://doi.org/10.1109/HPCC.2012.212

[49] M. Delvai, W. Huber, P. Puschner, and A. Steininger. 2003. Processor support for temporal predictability—The SPEAR design example. In *Proceedings of the 15th Euromicro Conference on Real-Time Systems (ECRTS'03)*.

[50] J. Diaz, C. Muñoz Caro, and A. Niño. 2012. A survey of parallel programming models and tools in the multi and many-core era. *IEEE Trans. Parallel Distrib. Syst.* 23, 8 (2012), 1369–1386. DOI: https://doi.org/10.1109/TPDS.2011.308

[51] EASA. 2011. Development Assurance of Airborne Electronic Hardware. EASA CM - SWCEH - 001, European Union Aviation Safety Agency (EASA). Retrieved from https://www.easa.europa.eu/sites/default/files/dfu/certification-docs-certification-memorandum-EASA-CM-SWCEH-001-Development-Assurance-of-Airborne-Electronic-Hardware.pdf.

[52] EASA. 2013. *Certification Memorandum—Software Aspects of Certification—EASA*. Technical Report.

[53] L. Ecco, S. Tobuschat, S. Saidi, and R. Ernst. 2014. A mixed critical memory controller using bank privatization and fixed priority scheduling. In *Proceedings of the IEEE 20th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA'14)*. 1–10. DOI: https://doi.org/10.1109/RTCSA.2014.6910550

[54] M. El-Shambakey and B. Ravindran. 2012. STM concurrency control for embedded real-time software with tighter time bounds. In *Proceedings of the 49th Design Automation Conference (DAC'12)*. ACM, 437–446. DOI: https://doi.org/10.1145/2228360.2228437

[55] EN 2011. EN50128—Railway Applications: Communication, signalling and processing systems—Software for railway control and protection systems.

[56] Meinhard Erben, Wolf Günther, Tobias Sedlmeier, Dieter Lederer, and Klaus-Jürgen Amsler. 2006. Legal aspects of safety designed software development, especially under european law. In *Proceedings of the 3rd European Embedded Real Time Software (ERTS'06)*. 6.

[57] F. Eris, A. Joshi, A. B. Kahng, Y. Ma, S. Mojumder, and T. Zhang. 2018. Leveraging thermally-aware chiplet organization in 2.5D systems to reclaim dark silicon. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE'18)*. 1441–1446. DOI: https://doi.org/10.23919/DATE.2018.8342238

[58] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger. 2012. Dark silicon and the end of multicore scaling. *IEEE Micro* 32, 3 (2012), 122–134. DOI: https://doi.org/10.1109/MM.2012.17

[59] M. Fakih, A. Lenz, M. Azkarate-Askasua, J. Coronel, A. Crespo, S. Davidmann, J. C. Diaz Garcia, N. Romero Gonzalez, K. Grüttner, S. Schreiner, R. Seyyedi, R. Obermaisser, A. Maleki, J. Öberg, M. T. Mohammadat, J. Perez-Cerrolaza, I. Sander, and I. Söderquist. 2017. SAFEPOWER project: Architecture for safe and power-efficient mixed-criticality systems. *Microprocess. Microsyst.* 52 (2017), 89–105. DOI: https://doi.org/10.1016/j.micpro.2017.05.016

[60] G. Fernandez, J. Abella, E. Quiñones, C. Rochange, T. Vardanega, and F. J. Cazorla. 2014. Contention in multicore hardware shared resources: Understanding of the state of the art. In *Proceedings of the 14th International Workshop on Worst-Case Execution Time Analysis*, H. Falk (Ed.). 31–42.

[61] Gabriel Fernandez, Javier Jalle, Jaume Abella, Eduardo Quiñones, Tullio Vardanega, and Francisco J. Cazorla. 2015. Increasing confidence on measurement-based contention bounds for real-time round-robin buses. In *Proceedings of the 52nd Design Automation Conference (DAC'15)*. ACM. DOI: https://doi.org/10.1145/2744769.2744858

[62] S. Fisher. 2013. *Certifying Applications in a Multi-Core Environment: A New Approach Gains Success*. Technical Report. SYSGO AG.

[63] G. Furano and A. Menicucci. 2018. *Roadmap for On-Board Processing and Data Handling Systems in Space*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9

[64] R. Ginosar. 2012. Survey of processors for space. In *Proceedings of the Conference on Data Systems In Aerospace (DASIA'12)*.

[65] S. Girbal, X. Jean, J. Le Rhun, D. Gracia Pérez, and M. Gatti. 2013. Deterministic platform software for hard real-time systems using multi-core COTS. In *Proceedings of the IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC'13)*. IEEE.

[66] D. Gizopoulos, M. Psarakis, S. V. Adve, P. Ramachandran, S. K. S. Hari, D. Sorin, A. Meixner, A. Biswas, and X. Vera. 2011. Architectures for online error detection and recovery in multicore processors. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE'11)*. 1–6. DOI: https://doi.org/10.1109/DATE.2011.5763096

[67] K. Goossens, A. Azevedo, K. Chandrasekar, M. D. Gomony, S. Goossens, M. Koedam, Y. Li, D. Mirzoyan, A. Molnos, A. B. Nejad, A. Nelson, and S. Sinha. 2013. Virtual execution platforms for mixed-time-criticality systems: The CompSOC architecture and design flow. *SIGBED Rev.* 10, 3 (2013), 23–34. DOI: https://doi.org/10.1145/2544350.2544353

[68] K. Goossens, J. Dielissen, and A. Radulescu. 2005. Aethereal network on chip: Concepts, architectures, and implementations. *IEEE Design Test Comput.* 22, 5 (2005), 414–421.

[69] K. Goossens and A. Hansson. 2010. The aethereal network on chip after ten years: Goals, evolution, lessons, and future. In *Proceedings of the 47th Design Automation Conference (DAC'10)*. ACM, 306–311. DOI: https://doi.org/10.1145/1837274.1837353

[70] S. Goossens, B. Akesson, and K. Goossens. 2013. Conservative open-page policy for mixed time-criticality memory controllers. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE'13)*. 525–530. DOI: https://doi.org/10.7873/DATE.2013.118

[71] J. E. Gottschlich and D. A. Connors. 2008. Extending contention managers for user-defined priority-based transactions. In *Proceedings of the Workshop on Exploiting Parallelism with Transactional Memory and other Hardware Assisted Methods (EPHAM'08)*.

[72] Giovani Gracioli, Ahmed Alhammad, Renato Mancuso, Antônio Augusto Fröhlich, and Rodolfo Pellizzoni. 2015. A survey on cache management mechanisms for real-time embedded systems. *ACM Comput. Surv.* 48, 2 (2015). DOI: https://doi.org/10.1145/2830555

[73] E. Grade, A. Hayek, and J. Börcsök. 2016. Implementation of a fault-tolerant system using safety-related Xilinx tools conforming to the standard IEC 61508. In *Proceedings of the International Conference on System Reliability and Science (ICSRS'16)*. 78–83. DOI: https://doi.org/10.1109/ICSRS.2016.7815842

[74] P. Graydon and I. Bate. 2013. Safety assurance driven problem formulation for mixed-criticality scheduling. In *Proceedings of the 1st International Workshop on Mixed Criticality Systems (WMC'13)*. 19–24.

[75] Z. Gu and Q. Zhao. 2012. A state-of-the-art survey on real-time issues in embedded systems virtualization. *J. Softw. Eng. Appl.* 5, 1 (2012), 277–290. DOI: https://doi.org/10.4236/jsea.2012.54033

[76] S. Guertin and M. White. 2010. CMOS reliability challenges - The future of commercial digital electronics and NASA. In *Proceedings of the NEPP Electronics Technology Workshop*.

[77] E. Hallet, G. Corradi, and S. McNeil. 2015. *WP461—Xilinx Reduces Risk and Increases Efficiency for IEC61508 and ISO26262 Certified Safety Applications*. Technical Report. Xilinx.

[78] J. Han, M. Deubzer, J. Seo Park, J. Harnisch, and P. Leteinturier. 2014. Efficient multi-core software design space exploration for hybrid control unit integration. In *SAE Tech. Paper*. DOI: https://doi.org/10.4271/2014-01-0260

[79] A. Hayek and J. Börcsök. 2012. SRAM-based FPGA design techniques for safety related systems conforming to IEC 61508 a survey and analysis. In *Proceedings of the 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*. 319–324. DOI: https://doi.org/10.1109/ICTEA.2012.6462892

[80] A. Hayek and J. Börcsök. 2014. Safety chips in light of the standard IEC 61508: Survey and analysis. In *Proceedings of the International Symposium on Fundamentals of Electrical Engineering (ISFEE'14)*. 1–6. DOI: https://doi.org/10.1109/ISFEE.2014.7050579

[81] A. Hayek, B. Machmur, M. Schreiber, J. Börcsök, S. Gölz, and M. Epp. 2014. HICore1: Safety on a chip; turnkey solution for industrial control. In *Proceedings of the IEEE 25th International Conference on Application-Specific Systems, Architectures, and Processors (ASAP'14)*. 74–75. DOI: https://doi.org/10.1109/ASAP.2014.6868636

[82] A. Hayek, M. Schreiber, B. Machmur, and J. Börcsök. 2013. Design and implementation of on-chip safety controller in terms of the standard IEC 61508. In *Proceedings of the Conference on Recent Advances in Circuits; Systems and Automatic Control*.

[83] J. Henkel, L. Bauer, N. Dutt, P. Gupta, S. Nassif, M. Shafique, M. Tahoori, and N. Wehn. 2013. Reliable on-chip systems in the nano-era: Lessons learnt and future trends. In *Proceedings of the 50th ACM/EDAC/IEEE Design Automation Conference (DAC'13)*. 1–10. DOI: https://doi.org/10.1145/2463209.2488857

[84] Carles Hernández, Jaume Abella, Francisco J. Cazorla, Alen Bardizbanyan, Jan Andersson, Fabrice Cros, and Franck Wartel. 2017. Design and implementation of a time predictable processor: Evaluation with a space case study. In *Proceedings of the 29th Euromicro Conference on Real-Time Systems (ECRTS'17)*, Vol. 76. 16:1–16:23. DOI: https://doi.org/10.4230/LIPIcs.ECRTS.2017.16

[85] S. Hesham, J. Rettkowski, D. Goehringer, and M. A. Abd El Ghany. 2017. Survey on real-time networks-on-chip. *IEEE Trans. Parallel Distrib. Syst.* 28, 5 (2017), 1500–1517. DOI: https://doi.org/10.1109/TPDS.2016.2623619

[86] Salma Hesham, Jens Rettkowski, Diana Göhringer, and Mohamed A. Abd El Ghany. 2015. Survey on real-time network-on-chip architectures. In *Applied Reconfigurable Computing*, Kentaro Sano, Dimitrios Soudris, Michael Hübner, and Pedro C. Diniz (Eds.). Springer, Cham, 191–202.

[87] C. Hilton and B. Nelson. 2006. PNoC: A flexible circuit-switched NoC for FPGA-based systems. *IEEE Proc. Comput. Dig. Techn.* 153, 3 (2006), 181–188. DOI: https://doi.org/10.1049/ip-cdt:20050175

[88] P. Huyck. 2012. ARINC 653 and multi-core microprocessors—Considerations and potential impacts. In *Proceedings of the IEEE/AIAA 31st Digital Avionics Systems Conference (DASC'12)*. 6B41–6B47.

[89] IEC 2010. IEC 61508(-1/7): Functional safety of electrical/electronic/programmable electronic safety-related systems. Technical Report.
[90] IEC 2018. IEC 62443-4-1: Security for industrial automation and control systems—Part 4-1: Secure product development lifecycle requirements. Technical Report.
[91] Infineon. 2018. *AURIX 32-bit Microcontrollers for Automotive and Industrial Applications*. Technical Report. Infineon.
[92] C. Iordanou, V. Soteriou, and K. Aisopos. 2014. Hermes: Architecting a top-performing fault-tolerant routing algorithm for networks-on-chips. In *Proceedings of the IEEE 32nd International Conference on Computer Design (ICCD'14)*. 424–431. DOI: https://doi.org/10.1109/ICCD.2014.6974715
[93] ISO 2018. ISO 26262(-1/11) Road vehicles—Functional safety. Technical Report.
[94] ITRS. 2018. *International Roadmap for Devices and Systems—Executive Summary*. Technical Report. IEEE.
[95] X. Iturbe, B. Venu, J. Jagst, E. Ozer, P. Harrod, C. Turner, and J. Penton. 2018. Addressing functional safety challenges in autonomous vehicles with the Arm TCLS architecture. *IEEE Design Test* 35, 3 (2018), 7–14. DOI: https://doi.org/10.1109/MDAT.2018.2799799
[96] V. Izosimov, A. Paschalis, P. Reviriego, and H. Manhaeve. 2018. *Application-Specific Solutions*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9
[97] J. Jalle, E. Quinones, J. Abella, L. Fossati, M. Zulianello, and F. J. Cazorla. 2014. A dual-criticality memory controller (DCmc): Proposal and evaluation of a space case study. In *Proceedings of the IEEE Real-Time Systems Symposium (RTSS'14)*. 207–217. DOI: https://doi.org/10.1109/RTSS.2014.23
[98] X. Jean. 2015. *Hypervisor Control of COTS Multi-Cores Processors in Order to Enforce Determinism for Future Avionics Equipment*. Ph.D. Dissertation. Telecom ParisTech.
[99] X. Jean, M. Gatti, G. Berthon, and M. Fumey. 2011. *The Use of Multicore Processors in Airborne Systems (EASA 2011.C31)*. Technical Report. EASA, Thales Avionics.
[100] S. K. Jena and M. B. Srinivas. 2012. On the suitability of multi-core processing for embedded automotive systems. In *Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC'12)*. 315–322. DOI: https://doi.org/10.1109/CyberC.2012.60
[101] R. Kalayappan and S. R. Sarangi. 2013. A survey of checker architectures. *ACM Comput. Surv.* 45, 4 (2013), 1–34. DOI: https://doi.org/10.1145/2501654.2501662
[102] B. Keller, M. Cochet, B. Zimmer, J. Kwak, A. Puggelli, Y. Lee, M. Blagojević, S. Bailey, P. F. Chiu, P. Dabbelt, C. Schmidt, E. Alon, K. Asanović, and B. Nikolić. 2017. A RISC-V processor SoC with integrated power management at submicrosecond timescales in 28 nm FD-SOI. *IEEE J. Solid-State Circ.* 52, 7 (2017), 1863–1875. DOI: https://doi.org/10.1109/JSSC.2017.2690859
[103] S. Kiamehr, M. B. Tahoori, and L. Anghel. 2018. *Manufacturing Threats*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9
[104] L. M. Kinnan. 2009. Use of multicore processors in avionics and its potential impact on implementation and certification. *SAE Tech. Papers* (2009).
[105] Hermann Kopetz. 2019. *Simplicity is Complex: Foundations of Cyber-Physical System Design*. Springer. DOI: https://doi.org/10.1007/978-3-030-20411-2
[106] Leonidas Kosmidis, Jerome Lachaize, Jaume Abella, Olivier Notebaert, Francisco J. Cazorla, and David Steenari. 2019. GPU4S: Embedded GPUs in space. In *Proceedings of the 22nd Euromicro Conference on Digital Systems Design (DSD'19)*.
[107] O. Kotaba, J. Nowotsch, M. Paulitsch, S. M. Petters, and H. Theilingx. 2013. Multicore in real-time systems - Temporal isolation challenges due to shared resources. In *Proceedings of the Workshop on Industry-Driven Approaches for Cost-effective Certification of Safety-Critical, Mixed-Criticality Systems (WICERT'13)*.
[108] S. Kriaa, L. Pietre-Cambacedes, M. Bouissou, and Y. Halgand. 2015. A survey of approaches combining safety and security for industrial control systems. *Rel. Eng. Syst. Safety* 139 (2015), 156–178.
[109] K. Lakshmanan, S. Kato, and R. Rajkumar. 2010. Scheduling parallel real-time tasks on multi-core processors. In *Proceedings of the 31st IEEE Real-Time Systems Symposium* 259–268. DOI: https://doi.org/10.1109/RTSS.2010.42
[110] A. Larrucea, I. Agirre, C. F. Nicolas, J. Perez, M. Azkarate-Askasua, and T. Trapman. 2015. Temporal independence validation of an IEC-61508 compliant mixed-criticality system based on multicore partitioning. In *Proceedings of the Forum on Specification and Design Languages (FDL'15)*. 1–8.
[111] A. Larrucea, I. Martinez, H. Ahmadian, R. Obermaisser, V. Brocal, S. Peiró, and J. Perez. 2016. DREAMS: Cross-domain mixed-criticality patterns. In *Proceedings of the Mixed-Criticality Workshop on Real Time System Symposium (RTSS'16)*.
[112] A. Larrucea, I. Martinez, R. Obermaisser, J. Perez, and C. F. Nicolas. 2017. Modular development of dependable mixed-criticality embedded systems. In *Proceedings of the 20th Euromicro Conference on Digital Systems Design (DSD'17)*. 419–426. DOI: https://doi.org/10.1109/DSD.2017.93

[113] A. Larrucea, J. Perez, I. Agirre, V. Brocal, and R. Obermaisser. 2015. A modular safety case for an IEC 61508 compliant generic hypervisor. In *Proceedings of the 18th Euromicro Conference on Digital Systems Design (DSD'15)*. 571–574. DOI: https://doi.org/10.1109/DSD.2015.27

[114] A. Larrucea, J. Perez, and R. Obermaisser. 2015. A modular safety case for an IEC 61508 compliant COTS multi-core device. In *Proceedings of the 13th International Conference on Dependable, Autonomatic and Secure Computing (DASC'15)*. 8. DOI: https://doi.org/10.1109/DSD.2016.66

[115] A. Larrucea, J. Perez, and R. Obermaisser. 2015. A modular safety case for an IEC 61508 compliant generic COTS processor. In *Proceedings of the IEEE International Conference on Computing and Information Technology; Ubiquitous Computing and Communication; Dependable, Autonomic and Secure Computing; Pervasive Intellegence and Computing (CIT/IUCC/DASC/PICom'15)*. 1788–1795. DOI: https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.269

[116] E. A. Lee. 2008. Cyber physical systems: Design challenges. In *Proceedings of the 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing (ISORC'08)*. 363–369.

[117] Ikhwan Lee, Michael Sullivan, Evgeni Krimer, Dong Wan Kim, Mehmet Basoglu, Doe Hyun Yoon, Larry Kaplan, and Mattan Erez. 2012. *Survey of Error and Fault Detection Mechanisms*. Technical Report. The University of Texas. Retrieved from http://lph.ece.utexas.edu/merez/uploads/MattanErez/detection_mechanisms_TR_LPH_2011_002.pdf.

[118] P. Leteinturier, S. Brewerton, and K. Scheibert. 2008. MultiCore benefits & challenges for automotive applications. In *SAE Tech. Paper*. DOI: https://doi.org/10.4271/2008-01-0989

[119] Paul S. Levy. 2017. *WP495—Using Zynq-7000 SoC IEC 61508 Artifacts to Achieve ISO 13849 Compliance*. Technical Report. Xilinx.

[120] Y. Li, B. Akesson, and K. Goossens. 2014. Dynamic command scheduling for real-time memory controllers. In *Proceedings of the 26th Euromicro Conference on Real-Time Systems (ECRTS'14)*. 3–14. DOI: https://doi.org/10.1109/ECRTS.2014.18

[121] I. Liu, J. Reineke, D. Broman, M. Zimmer, and E. A. Lee. 2012. A PRET microarchitecture implementation with repeatable timing and competitive performance. In *Proceedings of the IEEE 30th International Conference on Computer Design (ICCD'12)*. 87–93. DOI: https://doi.org/10.1109/ICCD.2012.6378622

[122] I. Liu, J. Reineke, and E. A. Lee. 2010. A PRET architecture supporting concurrent programs with composable timing properties. In *Proceedings of the 44th Asilomar Conference on Signals, Systems, and Computing (ASILOMAR'10)*. 2111–2115. DOI: https://doi.org/10.1109/ACSSC.2010.5757922

[123] A. Löfwenmark and S. Nadjm-Tehrani. 2018. Fault and timing analysis in critical multi-core systems: A survey with an avionics perspective. *J. Syst. Architect.* 87 (2018), 1–11. DOI: https://doi.org/10.1016/j.sysarc.2018.04.001

[124] H. Lu. 2013. *Low-Cost Highly-Efficient Fault Tolerant Processor Design for Mitigating the Reliability Issues in Nanometric Technologies*. Thesis, Université de Grenoble.

[125] G. Macher, M. Bachinger, and M. Stolz. 2017. Embedded multi-core system for design of next-generation powertrain control units. In *Proceedings of the 13th European Dependable Computing Conference (EDCC'17)*. 66–72. DOI: https://doi.org/10.1109/EDCC.2017.32

[126] G. Macher, A. Höller, E. Armengaud, and C. Kreiner. 2015. Automotive embedded software: Migration challenges to multi-core computing platforms. In *Proceedings of the IEEE 13th International Conference on Industrial Informatics (INDIN'15)*. 1386–1393. DOI: https://doi.org/10.1109/INDIN.2015.7281937

[127] A. Maheshwari, W. Burleson, and R. Tessier. 2004. Trading off transient fault tolerance and power consumption in deep submicron (DSM) VLSI circuits. *IEEE Trans. Very Large Scale Integr. Syst.* 12, 3 (2004), 299–311. DOI: https://doi.org/10.1109/TVLSI.2004.824302

[128] Claire Maiza, Hamza Rihani, Juan M. Rivas, Joël Goossens, Sebastian Altmeyer, and Robert I. Davis. 2019. A survey of timing verification techniques for multi-core real-time systems. *ACM Comput. Surv.* 52, 3 (2019). DOI: https://doi.org/10.1145/3323212

[129] R. Mariani, G. Boschi, and F. Colucci. 2007. Using an innovative SoC-level FMEA methodology to design in compliance with IEC 61508. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE'07)*. 1–6. DOI: https://doi.org/10.1109/DATE.2007.364641

[130] Arya Mazaheri, Johannes Schulte, Matthew Moskewicz, Felix Wolf, and Ali Jannesari. 2019. Enhancing the programmability and performance portability of GPU tensor operations. In *Proceedings of the 25th Euro-Par Conference (Lecture Notes in Computer Science)*, Vol. 11725. Springer, 213–226. DOI: https://doi.org/10.1007/978-3-030-29400-7_16

[131] A. Meixner, M. E. Bauer, and D. Sorin. 2007. Argus: Low-cost, comprehensive error detection in simple cores. In *Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'07)*. 210–222. DOI: https://doi.org/10.1109/MICRO.2007.18

[132] S. Metzlaff, S. Weis, and T. Ungerer. 2013. Leveraging transactional memory for a predictable execution of applications composed of hard real-time and best-effort tasks. In *Proceedings of the 21st International Conference on Real-Time Networks and Systems (RTNS'13)*. ACM, 45–54. DOI: https://doi.org/10.1145/2516821.2516832

[133] M. Millberg, E. Nilsson, R. Thid, and A. Jantsch. 2004. Guaranteed bandwidth using looped containers in temporally disjoint networks within the nostrum network on chip. In *Proceeding of the Conference on Design, Automation and Test in Europe (DATE'04)*, Vol. 2. 890–895.

[134] T. Mitra, J. Teich, and L. Thiele. 2018. Time-critical systems design: A survey. *IEEE Design Test* 35, 2 (2018), 8–26. DOI: https://doi.org/10.1109/MDAT.2018.2794204

[135] S. Mittal. 2016. A survey of recent prefetching techniques for processor caches. *ACM Comput. Surv.* 49, 2 (2016). DOI: https://doi.org/10.1145/2907071

[136] S. Mittal. 2017. A survey of techniques for cache partitioning in multicore processors. *ACM Comput. Surv.* 50, 2 (2017). DOI: https://doi.org/10.1145/3062394

[137] Sparsh Mittal and Jeffrey S. Vetter. 2015. A survey of CPU-GPU heterogeneous computing techniques. *ACM Comput. Surv.* 47, 4 (2015). DOI: https://doi.org/10.1145/2788396

[138] S. R. Msirdi. 2017. *Modular Avionics Software Integration on Multi-Core COTS: Certification-Compliant Methodology and Timing Analysis Metrics for Legacy Software Reuse in Modern Aerospace Systems*. Thesis, Université de Toulouse.

[139] Imanol Mugarza, Jorge Parra, and Eduardo Jacob. 2017. Software updates in safety and security co-engineering. In *Computer Safety, Reliability, and Security*, Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch (Eds.). Springer, Cham, 199–210.

[140] H. Mushtaq, Z. Al-Ars, and K. Bertels. 2015. Calculation of worst-case execution time for multicore processors using deterministic execution. In *Proceedings of the 25th International Workshop on Power and Timing Modeling, Optimization and Simulation (PATMOS'15)*. 33–39. DOI: https://doi.org/10.1109/PATMOS.2015.7347584

[141] G. Nelissen, D. Pereira, and L. M. Pinho. 2015. A novel run-time monitoring architecture for safe and efficient inline monitoring. In *Reliable Software Technologies—Ada-Europe*, Juan Antonio de la Puente and Tullio Vardanega (Eds.). Springer, 66–82.

[142] T. S. Nidhin, Anindya Bhattacharyya, R. P. Behera, and T. Jayanthi. 2018. A review on SEU mitigation techniques for FPGA configuration memory. *IETE Tech. Rev.* 35, 2 (2018), 157–168. DOI: https://doi.org/10.1080/02564602.2016.1265905

[143] J. Nowotsch and M. Paulitsch. 2012. Leveraging multi-core computing architectures in avionics. In *Proceedings of the 9th European Dependable Computing Conference (EDCC'12)*. 132–143.

[144] R. Obermaisser, C. El Salloum, B. Huber, and H. Kopetz. 2008. The time-triggered system-on-a-chip architecture. In *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE'08)*. 1941–1947. DOI: https://doi.org/10.1109/ISIE.2008.4677135

[145] R. Obermaisser, C. El Salloum, B. Huber, and H. Kopetz. 2009. From a federated to an integrated automotive architecture. *IEEE Trans. Comput.-Aided Design Integr. Circ. Syst.* 28, 7 (2009), 956–965. DOI: https://doi.org/10.1109/TCAD.2009.2014005

[146] M. Ottavi, D. Gizopoulos, and S. Pontarelli. 2018. *Dependable Multicore Architectures at Nanoscale*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9

[147] M. Ottavi, S. Pontarelli, D. Gizopoulos, C. Bolchini, M. K. Michael, L. Anghel, M. Tahoori, A. Paschalis, P. Reviriego, O. Bringmann, V. Izosimov, H. Manhaeve, C. Strydis, and S. Hamdioui. 2015. Dependable multicore architectures at nanoscale: The view from europe. *IEEE Design Test* 32, 2 (2015), 17–28. DOI: https://doi.org/10.1109/MDAT.2014.2359572

[148] Z. Owda and R. Obermaisser. 2017. Mixed-criticality transactional memory controller for embedded systems. In *IEEE International Conference on Industrial Informatics (INDIN'17)*. 104–110. DOI: https://doi.org/10.1109/INDIN.2016.7819142

[149] I. Oz and S. Arslan. 2019. A survey on multithreading alternatives for soft error fault tolerance. *ACM Comput. Surv.* 52, 2 (2019), 1–38. DOI: https://doi.org/10.1145/3302255

[150] Haluk Ozaktas, Christine Rochange, and Pascal Sainrat. 2013. Automatic WCET analysis of real-time parallel applications. In *Proceedings of the 13th International Workshop on Worst-Case Execution Time Analysis (WCET'13)*.

[151] Miloš Panić, Sebastian Kehr, Eduardo Quiñones, Bert Boddecker, Jaume Abella, and Francisco J. Cazorla. 2014. RunPar: An allocation algorithm for automotive applications exploiting runnable parallelism in multicores. In *Proceedings of the International Conference on Hardware/Software Codesign and System Synthesis (CODES'14)*. ACM, Article 29, 10 pages. DOI: https://doi.org/10.1145/2656075.2656096

[152] M. Paolieri, J. Mische, S. Metzlaff, M. Gerdes, E. Quiñones, S. Uhrig, T. Ungerer, and F. J. Cazorla. 2013. A hard real-time capable multi-core SMT processor. *ACM Trans. Embed. Comput. Syst.* 12, 3 (2013). DOI: https://doi.org/10.1145/2442116.2442129

[153] M. Paolieri, E. Quinones, F. J. Cazorla, and M. Valero. 2009. An analyzable memory controller for hard real-time CMPs. *IEEE Embed. Syst. Lett.* 1, 4 (2009), 86–90. DOI: https://doi.org/10.1109/LES.2010.2041634

[154] C. Paukovits. 2008. *The Time-triggered System-on-chip Architecture*. Thesis. Institut für Technische Informatik.

[155] M. Paulitsch, O. M. Duarte, H. Karray, K. Mueller, D. Muench, and J. Nowotsch. 2015. Mixed-criticality embedded systems—A balance ensuring partitioning and performance. In *Proceedings of the Euromicro Conference on Digital System Design (DSD'15)*. 453–461. DOI: https://doi.org/10.1109/DSD.2015.100

[156] R. Pellizzoni, A. Schranzhofer, Chen Jian-Jia, M. Caccamo, and L. Thiele. 2010. Worst case delay analysis for memory interference in multicore systems. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE'10)*. 741–746. DOI: https://doi.org/10.1109/DATE.2010.5456952

[157] J. Perez, M. Coppola, M. Faugère, D. Gracia Perez, M. Grammatikakis, A. Larrucea Ortube, A. Mouzakitis, A. Papagrigoriou, P. Petrakis, V. Piperaki, I. Sarasola, and G. Tsamis. 2018. *Evaluation*. CRC Press.

[158] J. Perez, D. Gonzalez, C. F. Nicolas, T. Trapman, and J. M. Garate. 2014. A safety certification strategy for IEC-61508 compliant industrial mixed-criticality systems based on multicore partitioning. In *Proceedings of the 17th Euromicro Conference on Digital Syst. Design (DSD'14)*. 394–400. DOI: https://doi.org/10.1109/DSD.2014.38

[159] J. Perez, D. Gonzalez, S. Trujillo, and A. Trapman. 2015. *A Safety Concept for an IEC 61508 Compliant Fail-safe Wind Power Mixed-criticality Embedded System Based on Multi-core Partitioning*. Lecture Notes in Computer Science, Vol. 9111. Springer. 3–17. DOI: https://doi.org/10.1007/978-3-319-19584-1_1

[160] S. Pinto, A. Oliveira, J. Pereira, J. Cabral, J. Monteiro, and A. Tavares. 2017. Lightweight multicore virtualization architecture exploiting ARM TrustZone. In *Proceedings of the 43rd IEEE Industrial Electronics Society (IECON'17)*. 3562–3567. DOI: https://doi.org/10.1109/IECON.2017.8216603

[161] A. Platschek, N. Mc Guire, and L. Bulwahn. 2018. Certifying Linux: Lessons learned in three years of SIL2LinuxMP. In *Proceedings of the Embedded World Conference*.

[162] S. Pontarelli, Juan A. Maestro, and P. Reviriego. 2018. *Dependability Solutions*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9

[163] B. Krishna Priya, Amit D. Joshi, and N. Ramasubramanian. 2016. A survey on performance of on-chip cache for multi-core architectures. In *Proceedings of the International Conference on Informatics and Analytics (ICIA'16)*. ACM. DOI: https://doi.org/10.1145/2980258.2980336

[164] Roger Pujol, Hamid Tabani, Leonidas Kosmidis, Enrico Mezzetti, Jaume Abella, and Francisco J. Cazorla. 2019. Generating and exploiting deep learning variants to increase heterogeneous resource utilization in the NVIDIA Xavier. In *Proceedings of the 31st Euromicro Conference on Real-Time Systems (ECRTS'19) (Leibniz International Proceedings in Informatics (LIPIcs))*, Sophie Quinton (Ed.), Vol. 133. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 23:1–23:23. DOI: https://doi.org/10.4230/LIPIcs.ECRTS.2019.23

[165] P. Puschner. 2013. Embedded systems for safety-critical and mixed-criticality applications. In *Proceedings of the 2nd Mediterranean Conference on Embedded Computers (MECO'13)*. 1–2.

[166] P. Puschner, B. Cilku, and D. Prokesch. 2016. Constructing time-predictable MPSoCs: Avoid conflicts in temporal control. In *Proceedings of the IEEE 10th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSOC'16)*. 321–328. DOI: https://doi.org/10.1109/MCSoC.2016.55

[167] Arthur Pyka, Mathias Rohde, and Sascha Uhrig. 2014. A real-time capable coherent data cache for multicores. *Concurr. Comput.: Pract. Exper.* 26, 6 (2014), 1342–1354.

[168] Eduardo Quiñones, Marko Bertogna, Erez Hadad, Ana Juan Ferrer, Luca Chiantore, and Alfredo Reboa. 2018. Big data analytics for smart cities: The H2020 CLASS project. In *Proceedings of the 11th ACM International Systems and Storage Conference (SYSTOR'18)*. ACM, 130. DOI: https://doi.org/10.1145/3211890.3211914

[169] S. Royuela, A. Duran, M. A. Serrano, E. Quiñones, and X. Martorell. 2017. *A Functional Safety OpenMP* for Critical Real-Time Embedded Systems*. Springer, Book section 16. DOI: https://doi.org/10.1007/978-3-319-65578-9_16

[170] S. Saidi, R. Ernst, S. Uhrig, H. Theiling, and B. D. de Dinechin. 2015. The shift to multicores in real-time and safety-critical systems. In *Proceedings of the 10th International Conference on Hardware/Software Codesign and System Synthesis*. IEEE Press, 220–229. DOI: https://doi.org/10.1109/CODESISSS.2015.7331385

[171] O. Sander, F. Bapp, L. Dieudonne, T. Sandmann, and J. Becker. 2017. The promised future of multi-core processors in avionics systems. *CEAS Aeronaut. J.* 8, 1 (2017), 143–155. DOI: https://doi.org/10.1007/s13272-016-0228-x

[172] I. Schagaev and T. Kaegi-Trachsel. 2016. *Software Design for Resilient Computer Systems*. Springer. DOI: https://doi.org/10.1007/978-3-319-29465-0

[173] K. Schmidt, J. Harnisch, D. Marx, A. Mayer, A. Kohn, and R. Deml. 2014. Timing analysis and tracing concepts for ECU development. *SAE World Congr. Exhibit.* 1 (2014). DOI: https://doi.org/10.4271/2014-01-0190

[174] M. Schoeberl, S. Abbaspour, B. Akesson, N. Audsley, R. Capasso, J. Garside, K. Goossens, S. Goossens, S. Hansen, R. Heckmann, S. Hepp, B. Huber, A. Jordan, E. Kasapaki, J. Knoop, Y. Li, D. Prokesch, W. Puffitsch, P. Puschner, A. Rocha, C. Silva, J. Sparsø, and A. Tocchi. 2015. T-CREST: Time-predictable multi-core architecture for embedded systems. *J. Syst. Architect.* 61, 9 (2015), 449–471. DOI: https://doi.org/10.1016/j.sysarc.2015.04.002

[175] M. Schoeberl, W. Puffitsch, S. Hepp, B. Huber, and D. Prokesch. 2018. Patmos: A time-predictable microprocessor. *Real-Time Syst.* 54, 2 (2018), 389–423. DOI: https://doi.org/10.1007/s11241-018-9300-4

[176] Erwin Schoitsch. 2013. *D6.1—State of the Art for System Qualification and Certification, V&V Survey (EMC2)*. Technical Report.

[177] A. Scolari, F. Sironi, D. Sciuto, and M. D. Santambrogio. 2014. A survey on recent hardware and software-level cache management techniques. In *Proceedings of the IEEE International Symposium on Parallel and Distributed Processing with Applications*.242–247. DOI: https://doi.org/10.1109/ISPA.2014.41

[178] Maria A. Serrano, Sara Royuela, and Eduardo Quinones. 2018. Towards an OpenMP specification for critical real-time systems. In *Proceedings of the 14th International Workshop on OpenMP (IWOMP'18)*. 143–159. DOI: https://doi.org/10.1007/978-3-319-98521-3_10

[179] Z. Shi and A. Burns. 2008. Real-time communication analysis for on-chip networks with wormhole switching. In *Proceedings of the 2nd ACM/IEEE International Symposium on Networks-on-Chip (NOCS'08)*. 161–170. DOI: https://doi.org/10.1109/NOCS.2008.4492735

[180] M. Slijepcevic, L. Kosmidis, J. Abella, E. Quinones, and F. J. Cazorla. 2014. Timing verification of fault-tolerant chips for safety-critical applications in harsh environments. *IEEE Micro* 34, 6 (2014), 8–19. DOI: https://doi.org/10.1109/MM.2014.59

[181] Daniel J. Sorin. 2009. *Fault Tolerant Computer Architecture*. Morgan & Claypool Publishers. DOI: https://doi.org/10.2200/S00192ED1V01Y200904CAC005

[182] V. Sridharan and S. Gurumurthi. 2018. *Resilience Proportionality—A Paradigm for Efficient and Reliable System Design*. Springer. DOI: https://doi.org/10.1007/978-3-319-54422-9

[183] H. Sutter and J. Larus. 2005. Software and the concurrency revolution. *Queue* 3, 7 (2005), 54–62. DOI: https://doi.org/10.1145/1095408.1095421

[184] S. Trujillo, A. Crespo, A. Alonso, and J. Perez. 2014. MultiPARTES: Multi-core partitioning and virtualization for easing the certification of mixed-criticality systems. *Microprocess. Microsyst.* 38, 8, Part B (2014), 921–932. DOI: https://doi.org/10.1016/j.micpro.2014.09.004

[185] G. Tsamis, S. Kavvadias, A. Papagrigoriou, M. D. Grammatikakis, and K. Papadimitriou. 2016. Efficient bandwidth regulation at memory controller for mixed criticality applications. In *Proceedings of the 11th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC'16)*. 1–8. DOI: https://doi.org/10.1109/ReCoSoC.2016.7533902

[186] Theo Ungerer, Christian Bradatsch, Martin Frieb, Florian Kluge, Jörg Mische, Alexander Stegmeier, Ralf Jahr, Mike Gerdes, Pavel Zaykov, Lucie Matusova, Zai Jian Jia Li, Zlatko Petrov, Bert Böddeker, Sebastian Kehr, Hans Regler, Andreas Hugl, Christine Rochange, Haluk Ozaktas, Hugues Cassé, Armelle Bonenfant, Pascal Sainrat, Nick Lay, David George, Ian Broster, Eduardo Quiñones, Milos Panic, Jaume Abella, Carles Hernandez, Francisco Cazorla, Sascha Uhrig, Mathias Rohde, and Arthur Pyka. 2016. Parallelizing industrial hard real-time applications for the parMERASA multicore. *ACM Trans. Embed. Comput. Syst.* 15, 3 (2016). DOI: https://doi.org/10.1145/2910589

[187] A. Vajda. 2011. *Multi-core and Many-core Processor Architectures*. Springer US, Boston, MA, 9–43. DOI: https://doi.org/10.1007/978-1-4419-9739-5_2

[188] P. K. Valsan, H. Yun, and F. Farshchi. 2016. Taming non-blocking caches to improve isolation in multicore real-time systems. In *Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'16)*. 1–12. DOI: https://doi.org/10.1109/RTAS.2016.7461361

[189] F. Wartel, L. Kosmidis, A. Gogonel, A. Baldovino, Z. Stephenson, B. Triquet, E. Quinones, C. Lo, E. Mezzetta, I. Broster, J. Abella, L. Cucu-Grosjean, T. Vardanega, and F. J. Cazorla. 2015. Timing analysis of an avionics case study on complex hardware/software platforms. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE'15)*. 397–402. DOI: https://doi.org/10.7873/DATE.2015.0189

[190] F. Wartel, L. Kosmidis, C. Lo, B. Triquet, E. Quinones, J. Abella, A. Gogonel, A. Baldovin, E. Mezzetti, L. Cucu, T. Vardanega, and F. J. Cazorla. 2013. Measurement-based probabilistic timing analysis: Lessons from an integrated-modular avionics case study. In *Proceedings of the 8th IEEE International Symposium on Industrial Embedded Systems (SIES'13)*. 241–248. DOI: https://doi.org/10.1109/SIES.2013.6601497

[191] Saud Wasly. 2018. *Architecture Design for Distributed Mixed-criticality Systems Based on Multi-core Chips*. Thesis, Universität Siegen.

[192] Raphael Weber. 2019. ARAMIS II. Tool interoperability and exchange formats. In *Proceedings of the International Workshop on Automated Engineering of Autonomic and Run-time Evolving Systems*. Final Workshop, Stuttgart.

[193] W. Weber, A. Hoess, J. v. Deventer, F. Oppenheimer, R. Ernst, A. Kostrzewa, P. Dore, T. Goubier, H. Isakovic, N. Druml, E. Wuchner, D. Schneider, E. Schoitsch, E. Armengaud, T. Soderqvist, M. Traversone, S. Uhrig, J. C. Perez-Cortes, S. Saez, J. Kuusela, M. v. Helvoort, X. Cai, B. Nordmoen, G. Y. Paulsen, H. P. Dahle, M. Geissel, J. Salecker, and P. Tummeltshammer. 2016. The EMC2 project on embedded microcontrollers: Technical progress after two years. In *Proceedings of the Euromicro Conference on Digital Systems Design (DSD'16)*. 524–531. DOI: https://doi.org/10.1109/DSD.2016.72

[194] D. Wentzlaff, P. Griffin, H. Hoffmann, L. Bao, B. Edwards, C. Ramey, M. Mattina, C. C. Miao, J. F. Brown III, and A. Agarwal. 2007. On-chip interconnection architecture of the tile processor. *IEEE Micro* 27, 5 (2007), 15–31. DOI: https://doi.org/10.1109/MM.2007.4378780

[195] R. Wilhelm, J. Engblom, A. Ermedahl, N. Holsti, S. Thesing, D. Whalley, G. Bernat, C. Ferdinand, R. Heckmann, T. Mitra, F. Mueller, I. Puaut, P. Puschner, J. Staschulat, and P. Stenstr. 2008. The worst-case execution-time problem—Overview of methods and survey of tools. *ACM Trans. Embed. Comput. Syst.* 7, 3 (2008), 1–53. DOI: https://doi.org/10.1145/1347375.1347389

[196] J. Windsor, K. Eckstein, P. Mendham, and T. Pareaud. 2011. Time and space partitioning security components for spacecraft flight software. In *Proceedings of the 30th IEEE/AIAA Digital Avionics Systems Conference (DASC'11)*. 8A5–1–8A5–14. DOI: https://doi.org/10.1109/DASC.2011.6096140

[197] W. Wolf, A. A. Jerraya, and G. Martin. 2008. Multiprocessor system-on-chip (MPSoC) technology. *IEEE Trans. Comput.-Aided Design Integr. Circ. Syst.* 27, 10 (2008), 1701–1713. DOI: https://doi.org/10.1109/TCAD.2008.923415

[198] Xilinx. 2018. *UG1085—Zynq UltraScale+ Device—Technical Reference Manual*. Technical Report. Xilinx.

[199] Xilinx. 2018. *UG116—Device Reliability Report—First Half 2018*. Technical Report. Xilinx.

[200] H. Yun. 2013. Improving Real-Time Performance on Multicore Platforms Using MemGuard. Retrieved from http://www.ittc.ku.edu/%7Eheechul/papers/memguard-rtlws13.pdf.

[201] H. Yun, G. Yao, R. Pellizzoni, M. Caccamo, and L. Sha. 2013. MemGuard: Memory bandwidth reservation system for efficient performance isolation in multi-core platforms. In *Proceedings of the IEEE 19th Real-Time and Embedded Technology and Applications Symposium (RTAS'13)*. 55–64. DOI: https://doi.org/10.1109/RTAS.2013.6531079

[202] H. Yun, G. Yao, R. Pellizzoni, M. Caccamo, and L. Sha. 2016. Memory bandwidth management for efficient performance isolation in multi-core platforms. In *IEEE Trans. Comput.* IEEE, 562–576.

[203] R. Zalman, A. Griessing, and P. Emberson. 2011. Timing correctness in safety-related automotive software. In *SAE Tech. Papers*. SAE. DOI: https://doi.org/10.4271/2011-01-0449

[204] M. Zimmer, D. Broman, C. Shaver, and E. A. Lee. 2014. FlexPRET: A processor platform for mixed-criticality systems. In *Proceedings of the IEEE 19th Real-Time and Embedded Technology and Applications Symposium (RTAS'14)*. 101–110. DOI: https://doi.org/10.1109/RTAS.2014.6925994

[205] M. P. Zimmer. 2015. *Predictable Processors for Mixed-Criticality Systems and Precision-Timed I/O*. Thesis, University of California, Berkeley.

Received October 2019; revised March 2020; accepted May 2020