1 Introduction

1.1 Motivation

Across a broad range of industries and markets, an increasing share of a product's value is provided by embedded computer systems. This trend is supported by the continued innovation in industry and academia that envisions new electronical functions on top of what can be achieved by electrical or mechanical measures alone, and that provide more cost-efficient implementations of formerly physical designs. The new functionality is eagerly awaited on the markets by a steady consumer demand for lower cost, more reliable, and more feature-rich products.

The added electronical features inevitably require more computational resources that allow performing more complex algorithms and generally run more sophisticated applications in a real-time environment. In the automotive industry this approach has led to the introduction of numerous embedded control units (ECUs) that each contribute a specific functionality to the car. As long as the number of such functions was small, and the functions were mainly tied to the mechanical counterparts which they control, this procedure was feasible. But by the mid-2000s it has led to more than 70 distinct controllers in a single high-end vehicle that must communicate over a set of field-buses and dedicated interconnects [Bro07]. The integration of some of these functions on a reduced set of controllers promises a tremendous cost saving potential.

The requirements for future computing platforms are thus to enable the integration of multiple functions onto a single ECU, to increase the computing performance provided to each application, and to deliver higher system reliability. These goals can not be achieved with the traditional, single-processor ECU design, because it does not scale without entailing new challenges: For example, increasing the processor's frequency leads to a quadratic increase in the component's power consumption [Cha92] which generates even more concerns about the required energy and heat dissipation, and raises issues with regard to electro-magnetic compatibility.

Consequently, the industries are turning to multicore solutions to deliver the necessary performance. Multicore designs have been used successfully in data-dominated systems, such as multimedia, or around applications that are by nature parallelizable, such as in server-based data centers. According to [Gul07], the shipping volume of multicore processors was expected to quadruple in the time from 2007 to 2009 and continue to do so through 2011. Although this development was offset by the global

recession, recent studies indicate that it is now back on track [Sch09a]. The application of multicore processors in control-dominated systems has focused on heterogeneous setups [Fre00, Inf08], in which a dedicated architecture can suitably be used to address a mix of low-latency and high-bandwidth service requirements. Recently, also homogeneous multicores have been proposed [Fre09a, inf] for use as powerful computing platforms.

In order to exploit the technological advantages of such new technologies for real-time systems, it is mandatory to have a thorough understanding of the resulting timing. Any failure to accurately predict the behavior of the final product will lead misdimensioning. In the best case this causes annoying cost increases, but in the worst case it endangers the correct operation and introduces safety risks.

The integration of multicore components into distributed cyber-physical systems leads to highly heterogeneous setups that consist of several processing nodes that are connected via a communication network. To provide sufficient computing performance, the nodes themselves are complex systems, sometimes consisting of multiple processors with a local bus and shared resources. The resulting hierarchical setup entails a multitude of challenges with respect to the timing predictably and indirectly the established design process. In this work, we provide answers to the question on how the relevant performance guarantees can be established in such setups.

This thesis introduces a formal performance analysis for today's and future networked embedded systems with real-time requirements. The analysis can accompany the development process during system design, component dimensioning, performance verification, and product certification. The performance predictions are not only safe, but also accurate by addressing the versatile behavior of typical embedded systems.

In the remainder of this chapter, we will take a closer look at the predominant trends in the embedded systems industry in order to project the hardware and software architectures of today's and upcoming designs (Section 1.2). We then highlight the classical steps in the development process in Section 1.3, which allows us to see how timing problems are classically considered. The challenges to this process that are introduced by the new architectures are then identified in Section 1.4.1. This leads us to a discussion of the previously proposed countermeasures in Section 1.5 and the identification of the benefits of our methodology in Section 1.6.

1.2 Embedded System Trends

In the past years, the value generated in embedded systems has grown in many, if not all economically important domains [ZVE09] such as industrial automation, consumer electronics, or medical technology.

In the following section, we take a closer look at the situation in the automotive industry. This industry is subject to representative trends that can also be observed in other domains: The strict constraints of embedded systems are combined with the

large markets of consumer electronics and a general openness to evolving technologies in light of a global competition.

1.2.1 Market Trends

A growing share of the value generated in a modern car is created in its electric, electronic, and programmable electronic (E/E/PE) components. The value of electrical systems and electronics in the average automobile will in 2015 approach EUR 4,150 (as opposed to EUR 2,220 in 2004) [Dan04]. This trend is reflected in the continued growth of the automotive electronics market that continuously beats the growth of the overall vehicle markets as shown in Figure 1.1. The growth could be sustained even during the recession of 2008/2009, and is expected to continue to average around 7% annually [Sch09a].
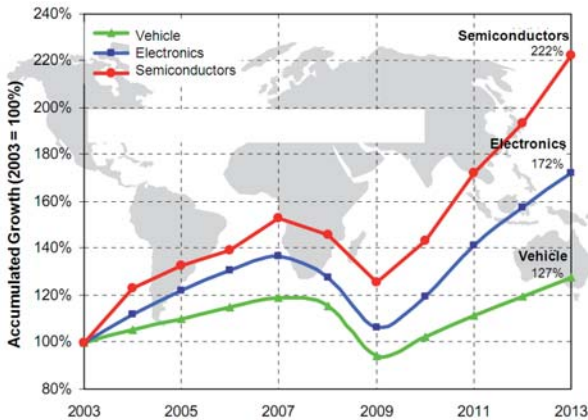

Figure 1.1: Relative Growth of Semiconductors, Embedded Modules and Vehicles Markets in Automotive (Source [Sch09a])

Several major trends can be identified that indicate that the amount of electronic and software-based solutions will significantly grow in the future. To accommodate the additional software, powerful controllers are required that deliver the computing power, without sacrificing non-functional requirements such as power, reliability, or electro-magnetic radiation.

Reducing the Environmental Impact Consumers and lawmakers worldwide are becoming increasingly aware of the scarcity of common natural resources and the environmental impact of its uncontrolled exploitation. The transportation sector is

a significant contributor to the release of recognized climate gases[1]. In Germany 2007, one third of the country's energy was consumed by transportation, with more than 80% of it attributed to road traffic [Sta09b]. This has led to stringent emission regulation in this area [Rod10].

A key concept to tackle this challenge is the optimization of the combustion engine's control to allow a more fuel-efficient injection, and further innovations in the powertrain platform (start-stop technology [Wei07], regenerative breaking, and the move towards hybrid or even fully electrical engines). These developments imply more complex control applications to regulate the physical components. Another effective method to cut the fuel consumption is to reduce the vehicles weight [Bun07]. This can be achieved by replacing mechanical and hydraulic control schemes by electronic X-by-wire solutions [Lee02] — which again increases the control complexity. Weight is also saved by decreasing the number of ECUs and in the car, which to a certain extend allows streamlining the network topology [Obe09]. All of these means directly or indirectly demand for the availability of more computing power per control unit.

Increasing Passenger and Pedestrian Safety Mechanical measures to increase the passengers safety have long been explored and are mostly exploited [Lee02]. Additional improvements can only be achieved with active measures that better assist the driver in avoiding collisions in the first place. The complexity of such measures reaches from simple break control [Lei80] to camera-based lane, pedestrian, and object recognition systems. The recent and upcoming applications consist of several sensors and actors, and significant data processing in between [Cur00]. The introduction of high data volumes into the domain of reliable control applications makes this field particularly challenging. Not least does it lead to an increased stress on the underlying computing hardware and design methodology.

Increasing the Car's Reliability The electrical and electronical components are responsible for a rising share of car breakdowns. The German automobile association (ADAC) identified the electronics as the source for car breakdowns in 40% of the cases in 2007[2] (up from 35% in 2003[3]). This highlights that the correct vehicle operation increasingly relies on the correct functioning of the applied electronics hardware and software.

One has to be aware that the required microelectronics progress with respect to performance, low power, and cost relies on the increase on the amount of available transistors and a reduction of the feature size. These improvements are lately bought at the cost of a growing susceptibility to transient faults that may directly lead to erroneous internal states [Mat97, Shi02, Bor05]. To compensate for this tendency, the

---

[1] `http://www.ipcc.ch/` (retrieved 2010-05-23)

[2] `http://www.sueddeutsche.de/auto/adac-pannenstatistik-autos-die-aerger-machen-1.411175-2` (retrieved 2010-05-23)

[3] `http://www.sueddeutsche.de/auto/adac-test-pannenursache-nr-die-elektrik-1.569566` (retrieved 2010-05-23)

computation and communications have to be secured with additional error-detection and correction layers. This will again lead to an increased communication and computation requirement [Man07, Seb09]. To accommodate the additional computation, multicore solutions have been proposed [Smo06, Wel09]. These solutions have the added benefit that they introduce a new dimension of spacial redundancy by separating the operations not only over time, but also over physical cores, which reduces the susceptibility to permanent faults. Besides the said positive economic, environmental and reliability benefit, the envisioned reduction of ECUs also leads to a reduction of other error sources – for example, in automotive environments more than 30% of electrical failures are ascribed to connector problems [Pet06] (citing [Swi00]).

Keeping the Cars Affordable With all these upcoming challenges in mind, car's have to remain affordable. If anything, the global recession of 2009 has increased the attention given to the the cost-efficiency of the automobile products.[4] But changes are also happening on a global scale [May08]. A new segment, the "ultra low-cost vehicles" with price points of less than USD5000 per vehicle is emerging.

As a large share of the value is created in electronics (see above), this segment also has to bear a significant share of the cost pressure. Fortunately, the current E/E/PE topologies provide a vast amount of optimization opportunities. A modern car can easily carry around more than 60 ECUs that are interconnected via a set of 6 field buses [Gri03]. Significant cost saving can be expected if the car's (software) functionality can be implemented on a reduced number of ECUs. In [Obe09] it was demonstrated how quickly an increased component cost is offset by the reduced number of components and the decrease in the wiring effort. Software standardization initiatives such as AUTOSAR [aut06] make this consolidation feasible.

Delivering Improved Customer Experience Finally, customers are to a growing extent perceiving the value of sophisticated electronic and software functions. Then-DaimlerChrysler estimated in 2003 that 80% of all future automotive innovations will be driven by electronics, 90% of which attributed to software [Gri03]. This trend is also indicated by the strong efforts by the OEMs to use built-in navigation and infotainment systems for brand differentiation [Sch09a].

But the standards and applications in these areas are subject to much shorter innovation cycles than given by the average car's lifecycle. Updates are now commonly applied during a product cycle, and also in the field. This strongly suggests software-based solutions on versatile platforms even in domains where dedicated processors and architectures have been predominant (as in [Moo05]).

The cited trends will cause a significant increase of the software and communication load in the automotive electronics that can not be handled by the past approach, in

---

[4]This has in particular hit the high-class segment, which classically acts as a door-opener to new technologies [Sch09a]

which the number of control units in a vehicle grows proportionally with the number of functions. The more complex applications and higher functional integration demands for sophisticated platforms that deliver the necessary computing power — without sacrificing the non-functional constraints of power, reliability, and weight.

1.2.2 Technology Trends

Finding an optimal architecture is a challenging task due to the stringent constraints of embedded, often mobile, systems that precludes many design options that are available to general purpose computing. A typical embedded design specification demands for functional correctness and correct timing, sufficient reliability, adherence to power constraints, and robustness to future design changes. This diversity of requirements often leads to heterogeneous solutions that are tailored to the application.

Dedicated Hardware Solutions The challenge of a large computing workload can for example be tackled with dedicated hardware, in particular when the required operations are highly regular and sufficiently parallel. The spectrum of implementations reaches from FPGA (field-programmable gate array) based solutions to the manufacturing of dedicated ASICs (application specific integrated circuits), which are not reconfigurable. FPGA's have been proposed for several niche aspects in the automotive electronics infrastructure, e.g. in gateways [San07] and image processing [Ang08, Won10]. The online reconfigurability of such systems is an increasingly interesting aspect, because many applications traverse through a number of dedicated scenarios during a typical drive (e.g. the image processing may be different in the modes "parking", "cruising on highway", or "cruising in rain").

Compared to FPGAs, dedicated ASICs promise a lower power consumption and faster processing speed at a lower cost per unit. However, the setup cost is relatively high, so that a large amount of identical units must be expected. Consequently, this approach is chosen typically only for low-level operations (such as bus controllers, sensor interfaces, ...). In particular, the applications projected in Section 1.2.1 commonly exhibit a complex and irregular control flow, making the dedicated hardware solutions inappropriate.

Programmable Platforms But there are also other design aspects to consider apart from the provisioning of pure processing power. Many high-level automotive functions have been implemented in software (e.g. the engine control, gateway functions, ...), although in theory other approaches would have been possible. This leaves the industry with a large base of existing solutions and significant know-how in the domain to which future solutions must be compatible.

However, continuously increasing the processor's clock speed is not feasible, because this usually implies also adopting the supply voltage to higher levels, which leads to an unreasonable surge in the overall power consumption [Cha92], which quickly introduces thermal problems [Fen03]. Moreover, the higher clock frequency yields

more electromagnetic radiation and susceptibility [Not10]. Thus on the one hand, power and clock constraints are increasingly strict [Sch97], but on the other hand the number of transistors continues to grow [Moo65, ITR]. This suggests measures to "translate transistors into performance" [Fly05]. The driving logic behind this trend can be observed in just about every computing domain: the desktop environment [Gee05], in server and super-computer environments, in embedded consumer devices, and in mobile devices.

One approach is to invest transistors to make the processor pipeline faster (i.e. longer) and better utilized by implementing measures such as speculative execution and prefetching that are often beneficial for the average-case throughput. Besides the debatable benefit of average-case improvements for real-time systems, this path has already been explored to an extent that makes any incremental improvements tremendously complex to design, program, or verify. A more predictable approach is to augment the processor pipeline with a configurable hardware unit. This approach is followed in the application specific instruction-set processors (ASIPs) [Raz94], which attempt to efficiently combine software execution with (reconfigurable) hardware. The approach allows to speed up and reduce the power requirement of the most common functions and instructions. Alternatively, the additional chip area can also be invested into fast memories (e.g. caches) in order to tackle the problem of the increasing speed gap between processing performance and memory bandwidth [Wul95].

Finally, parallel processing has been identified as the key option to tackle the growing performance requirements [ITR]. Through this, an equivalent amount of operations per second can be achieved at far less time than in a purely sequential processing. The parallelization of the workload can be achieved on different levels of granularity (instruction-level parallelism, data-level parallelism, task-level parallelism), and the achievable gain largely depends on the application itself. For example, a digital filter application typically exhibits significant instruction-level parallelism, while a heterogeneous application such as a user interface offers only parallel tasks. Commonly, different levels of parallelism support can be provided by an architecture.

Multicore Processors Multicore controllers (also called multicore processors), provide an efficient means to supply additional performance by combining several benefits. Firstly, they are fully programmable, enabling a relatively easy functional port of existing single-core applications. The architecture provides task-level parallelism that is very easy to exploit, especially if formerly independent functions are to be integrated. Secondly, the comparably low clock frequency makes multicores power-efficient, avoids heat problems, and provides a resistance to electromagnetic interference (and limits the emitted radiation). Thirdly, the independent cores offer the option for independent scheduling, which opens the way to physical redundancy as well as an evolutionary path to the high-performance applications through scheduling partitions and operating system virtualization. These benefits make multicore processors an attractive design target in across virtually all computing domains.