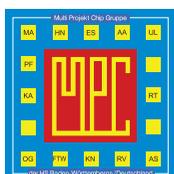


MPC

MULTI PROJEKT CHIP GRUPPE
BADEN - WÜRTTEMBERG

Herausgeber: Hochschule Ulm Ausgabe: 52 ISSN 1868-9221 Workshop: Künzelsau Juli 2014

- 1 Low Power Communication Electronics**
K. v. Arnim, Intel Mobile Communications GmbH, München
- 5 Electronic Displays: Technologies – Systems – Optimizations**
K.-H. Blankenbach, HS Pforzheim
- 11 A Millimeter-Wave Power Amplifier Concept in SiGe BiCMOS Technology for Investigating HBT Physical Limitations**
J. Wursthorn, H. Knapp, B. Wicht, HS Reutlingen
- 15 Substratkoppeln in schnell schaltenden integrierten Leistungsendstufen**
C. Rindfleisch, J. Wittmann, B. Wicht,
Robert Bosch Zentrum für Leistungselektronik, Reutlingen
- 21 Optimierte, wiederverwendbare OTA-Schaltungen für moderne Power BiCMOS-Technologien**
A. Gerlach, M. Junge, J. Scheible, T. Rosahl,
Robert Bosch Zentrum für Leistungselektronik, Reutlingen
- 27 Eine effiziente Dekodierarchitektur für verallgemeinert verkettete Codes**
J. Spinner, J. Freudenberger, HS Konstanz
- 33 Entwicklung einer innovativen Softwarearchitektur für Cloud-basierte Cyber-physische Systeme**
M. Engelsberger, T. Greiner, HS Pforzheim



Cooperating Organisation
Solid-State Circuit Society Chapter
IEEE German Section



Inhaltsverzeichnis

Low Power Communication Electronics	1
K. v. Arnim, Intel Mobile Communications GmbH, München	
Electronic Displays: Technologies – Systems – Optimizations	5
K.-H. Blankenbach, HS Pforzheim	
A Millimeter-Wave Power Amplifier Concept in SiGe BiCMOS	11
Technology for Investigating HBT Physical Limitations	
J. Wursthorn, H. Knapp, B. Wicht, HS Reutlingen	
Substratkoppeln in schnell schaltenden integrierten Leistungsendstufen	15
C. Rindfleisch, J. Wittmann, B. Wicht, Robert Bosch Zentrum für Leistungselektronik, Reutlingen	
Optimierte, wiederverwendbare OTA-Schaltungen	21
für moderne Power BiCMOS-Technologien	
A. Gerlach, M. Junge, J. Scheible, T. Rosahl, Robert Bosch Zentrum für Leistungselektronik, Reutlingen	
Eine effiziente Dekodierarchitektur für verallgemeinert verkettete Codes	27
J. Spinner, J. Freudenberger, HS Konstanz	
Entwicklung einer innovativen Softwarearchitektur	33
für Cloud-basierte Cyber-physische Systeme	
M. Engelsberger, T. Greiner, HS Pforzheim	

Tagungsband zum Workshop der Multiprojekt-Chip-Gruppe Baden-Württemberg
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie.

Die Inhalte der einzelnen Beiträge dieses Tagungsbandes liegen in der Verantwortung der jeweiligen Autoren.

Herausgeber:

Gerhard Forster, Hochschule Ulm, Prittwitzstraße 10, D-89075 Ulm

Mitherausgeber (Peer Reviewer):

Jürgen Giehl, Hochschule Mannheim, Paul-Wittsack-Straße 10, D-68163 Mannheim

Frank Kesel, Hochschule Pforzheim, Tiefenbronner Straße 65, D-75175 Pforzheim

Axel Sikora, Hochschule Offenburg, Badstraße 24, D-77652 Offenburg

Alle Rechte vorbehalten

Diesen Workshopband und alle bisherigen Bände finden Sie im Internet unter:

<http://www.mpc.belwue.de>



Low Power Communication Electronics

Klaus v. Arnim

Abstract—This paper gives an overview of challenges on state-of-the-art SoC design for mobile applications. Technology trends of scaling beyond 28 nm as well as low-power design techniques that are necessary in these technology nodes are described.

Index Terms—Low-power design, SoC.

I. INTRODUCTION

Until 5 years ago, low power design for mobile applications was fundamentally different from the low-power requirements for high-performance. Typical applications for mobile phones ran at 250 MHz, with a system power of 0.1 W and standby time was expected to be up to a month. At the same time, high-performance cores for desktops and notebooks achieved frequencies of 4 GHz, but at system power values of up to 400 W and battery runtimes of 3 hours for active use. Today's applications, such as tablets and smartphones, require the best of the two worlds. The user of a tablet should experience the same performance of the formerly bulky and expensive system in a power-efficient, fan-less, compact, and light format.

Neither the shrink of process technologies nor new low-power design techniques alone can enable such systems. It is the challenging task of the system engineering, circuit design and technology definition to build an efficient system-on-chip together.

State-of-the-art digital System-on-Chips (SoC) include a high number of system components, such as cores (microprocessors, DSPs and accelerators), lots of “random” logic, up to 1000 memory instances (SRAM, ROM, eFuses), a large variety of I/O interfaces including a full-chip ESD concept (General Purpose I/Os, LP-DDR2, USB2/3, etc.) as well as PLLs and voltage regulators.

This long and continuously growing list of components on a single chip result in a tremendous challenge on all design levels, such as handling multiple power and clock domains, manage timing and power re-

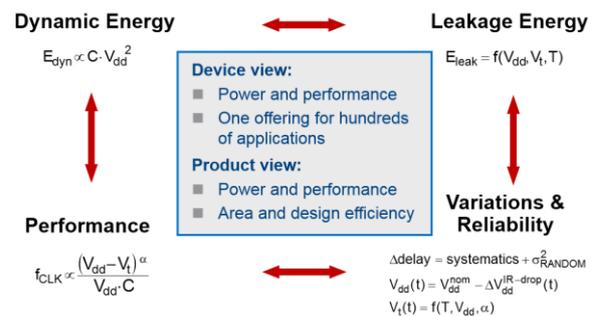


Figure 1: The magic square of power/performance optimization.

quirements across the chip, or to physically connect the blocks with more and more complicated design rules.

These requirements result in simple design guidelines for successful SoC integration:

- Keep it simple: hand-optimize only where it really pays off
- Design automation: avoid manual interaction in order to reduce errors
- Consistent design style: follow same design methodology in all components

The remainder of this paper is organized as follows: In the first half, technology trends such as lithography challenges and recent device development trends are presented. In the second half, low-power design techniques are discussed.

II. TECHNOLOGY TRENDS

Today, technology scaling faces some very significant limitations in transistor design and lithography.

A. Near-Term Device Options

Device scaling with classical planar-bulk structure and Poly-SiON gate stacks come to an end at the 28 nm CMOS technology node. The requirement to reduce area and dynamic power enforce the reduction of the supply voltage V_{dd} (Figure 1). This results either in a lower performance (if the threshold voltage V_t is kept constant) or in an increased leakage (for reduced V_t). In addition, the reduced dimensions in smaller technologies result in higher variations, e.g. by increased short-channel effects.

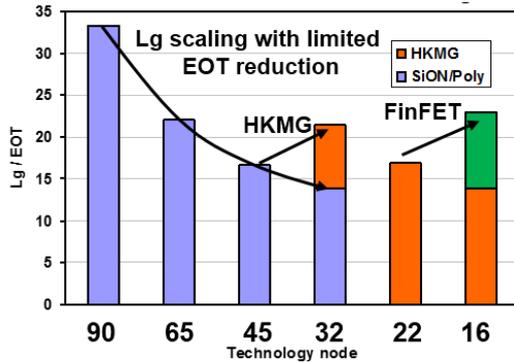


Figure 2: The transistor's ability for short-channel control (SSC) can be measured by the ratio of gate length (L_g) and Equivalent Oxide Thickness (EOT). In the 28/32 nm technology, high-k/metal gates (HKMG) boost the SSC [1]. In 22 nm or the 16/14 nm technology, the introduction of FinFETs again improve the device behavior [2], [3].

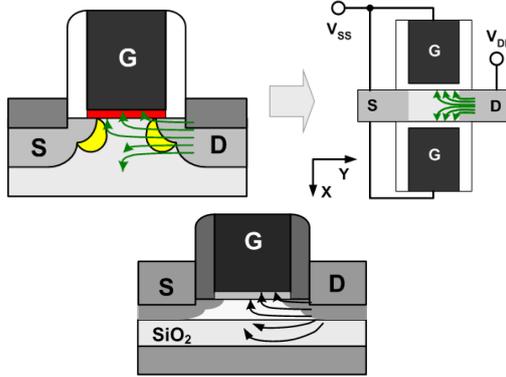


Figure 3: While in the classical planar-bulk CMOS transistor (top-left) the field below the gate has to be shielded by the bulk dopants, the channel of FinFETs (top-right) and Fully-Depleted SOI transistors (bottom) can be controlled without channel or bulk doping.

As shown in Figure 2, the short-channel control (SSC) decreased from 90 nm to 45 nm CMOS technologies. Only by introducing of device innovations like high-k/metal gates and FinFETs the SSC can be brought back to an acceptable level.

For technologies 22 nm and beyond, not only FinFETs, but also planar fully-depleted SOI transistors become an option. In contrast to the thinner Equivalent Oxide Thickness (EOT) with high-k/metal gates, the improved SSC in these devices is achieved by geometry, not just by material innovations (Figure 3).

FinFETs can be implemented in bulk or on SOI wafers (Figure 4). While the SOI version provides better leakage control, simpler process and lower power, the bulk implementation with lower process cost and better heat dissipation has become the preferred option. Due to the geometry-based gate control, the new device concepts do not require dopants in the channel region, resulting in a better mobility and thus in higher on-currents (I_{on}). In addition, the improved SSC results in higher currents in all relevant regions of the

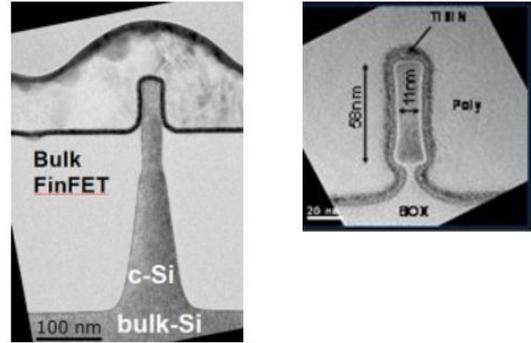


Figure 4: Bulk FinFET and SOI FinFET from technology pre-development implementations.

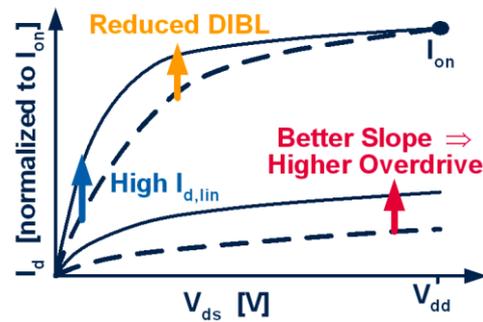


Figure 5: Even for the same on-current I_{on} , FinFETs provided higher drive currents in other regions of the output characteristic that are important for digital and analog circuit performance [4].

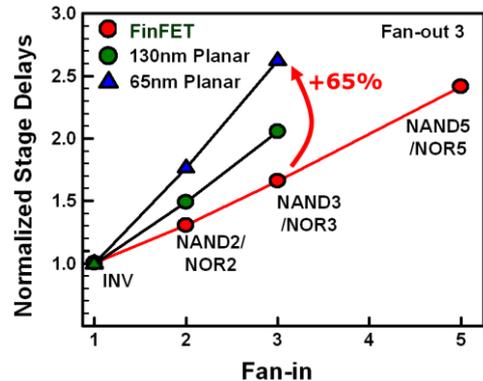


Figure 6: The improved circuit performance of FinFETs becomes obvious when plotting the NAND/NOR performance relative to the inverter performance [5].

output characteristic (Figure 5). The superior transistor behavior becomes visible especially in performance of stacked transistor configurations, e.g. in NAND and NOR gates (Figure 6).

B. Lithography Trends

Since the introduction of immersion lithography in the 45 nm CMOS technology there have been basically no further improvements of the wavelength $\lambda = 193$ nm or the numerical aperture $NA = 1.35$. The

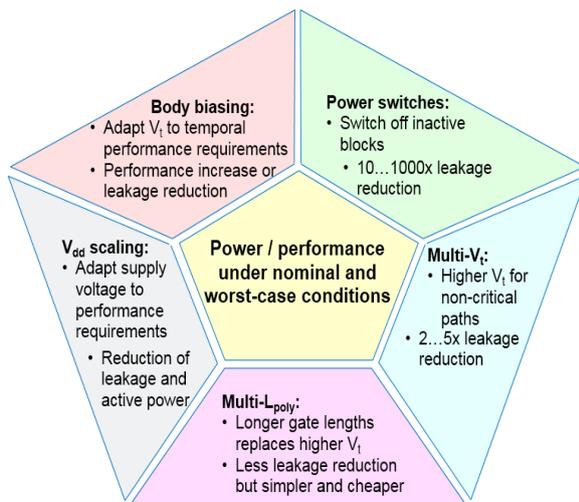


Figure 7: The five low-power techniques need to be applied in an optimized way in order to build power-efficient high-performance systems.

key figure λ/NA will also stay constant for many years. While the target pitches in the 28 nm CMOS technology can still be printed with this, new patterning techniques are required in 20 nm and beyond.

On the one hand, the introduction of new patterning techniques such as double patterning, restricted design rules, advanced illumination techniques, or spaced-based definition techniques pave the way to technology scaling to 20, 14 and 10 nm. On the other hand, these new techniques result in highly challenging design requirements.

III. LOW-POWER DESIGN TECHNIQUES

While the pool of low-power design techniques is well-known since many years (Figure 7), the requirements of new technologies as well as new application result in the highly challenging task of combining these techniques into an efficient system [6].

Recently, V_{dd} scaling and the use of multiple V_{dd} 's on the same chip have become the most important low-power techniques for mobile applications. This includes dedicated supplies for memories, optimized supplies for low-power and high-performance circuit blocks as well as adaptive voltages in dependence of performance requirements, process conditions or temperature. As shown in Figure 8, the voltage scaling is limited by increasing design efforts towards low voltages and by reliability issues for high voltages.

While combining different gate lengths (Multi- L_{poly}) becomes difficult in advanced technology nodes due to restricted design rules, the wide-spread application of multiple V_t 's is still an important design measure.

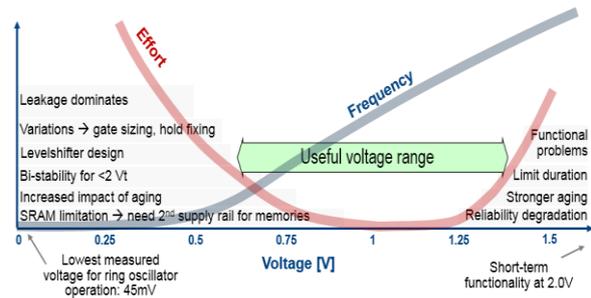


Figure 8: When adapting the voltage to performance requirements, the effort for extremely low and extremely high voltages become much higher.

IV. SUMMARY

Mobile applications such as smartphones and tablets have become the major drivers for CMOS technology scaling. After reaching the 28 nm node, industry faces significant obstacles. In this paper, challenges and solutions are presented with respect to device technology and lithography. On the device level, low gate length scaling can be overcome by introducing high-k/metal gates and FinFETs.

On the design side, additional measures are required to lever the potentials of the new technologies. Leakage and active power would even increase if optimized low-power design techniques were not applied.

REFERENCES

- [1] H. S. Yang et al., "Scaling of 32 nm low power SRAM with high-k metal gate", *IEDM* 2008.
- [2] C. Auth et al., "A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," *Symposium on VLSI Technology*, pp. 131-132, June 2012.
- [3] Wu, Shien-Yang et al., "A 16 nm FinFET CMOS technology for mobile SoC and computing applications." *Proc Int. Electron Devices Meeting (IEDM)*, 2013
- [4] K. von Arnim et al., "An effective switching current methodology to predict the performance of complex digital circuits", *IEDM* 2007.
- [5] K. von Arnim et al., "A low-power Multi-Gate FET CMOS technology with 13.9 ps inverter delay, large-scale integrated high performance digital circuits and SRAM", *Symposium on VLSI Technology* 2007.
- [6] C. Pacha et al., "Efficiency of low-power design techniques in multi-gate FET CMOS circuits", *ESSCIRC* 2007.



Klaus von Arnim received the degree of Dipl.Ing. and Dr-Ing. in Electrical Engineering from the Christian Albrecht University in Kiel in 2002 and 2006, respectively. From 2002 until 2011 he was with Infineon Technologies in Munich and Leuven. In 2011 he joined Intel as a Design System Architect.

Electronic Displays Technologies – Systems – Optimizations

Karlheinz Blankenbach

Abstract—Electronic displays are one of the key components of the information age. Ubiquitous computing needs mobile devices used outdoors. This sets the main requirements for displays: multimedia, low power and outdoor readability. Neither LCD, OLED nor e-paper meets all those needs. Optimizations of displays dedicated for applications are therefore necessary. Selected examples for GUI, LCD power saving and image enhancement for outdoor readability are presented.

Index Terms—Displays, LCD, OLED, e-paper, interface, display controller, adaptive dimming, image enhancement, GUI, HMI.

I. INTRODUCTION

The rise of microelectronics led to smartphones, PCs and large sized TV sets – all equipped with high resolution displays, which are manufactured in Asia. Beyond those consumer applications (CE), many professional systems with microcontrollers or processors and various sizes of displays are on the market. They follow the trends of CE displays like increasing resolution and touch. Examples for professional displays are

- Automotive infotainment
- Household appliances
- Industrial and medical monitors
- Public displays

Their displays are manufactured as well in Asia, but added value and integration effort (e.g. microcontroller, interfacing, cover lens) is done in western countries by many companies. The design of a system with a display requires several “looks” to this system as hardware, software, display, and the user.

As this paper is published in a microelectronics journal, some selected common fundamentals of microelectronics and displays are presented here as well: Figure 1 visualizes typical substrates (left) and structures (center) for both disciplines. While today’s wafers have a diameter of 300 mm, latest LCD production mother glass has dimensions of about 300 cm,

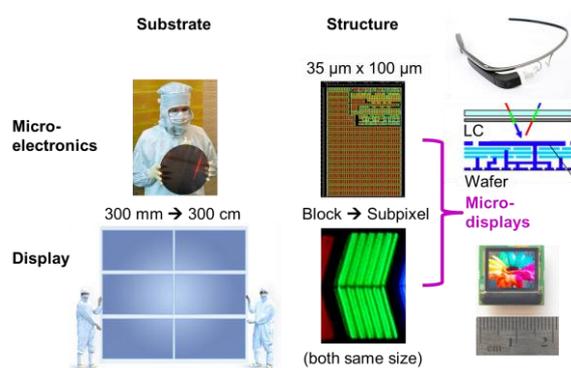


Figure 1: Comparison of substrates (left) and structures (center) of microelectronics and electronic displays. Microdisplays (right) are displays stacked on wafer [1].

which is a factor of ten larger. However, typical production technologies are the same for both like photolithography. Design rules for microelectronics are in the 30 nm range compared to $\sim 1 \mu\text{m}$ for TFTs (thin film transistors) of Active Matrix displays produced in thin film technology.

Building blocks of microcontrollers are e.g. CPU, UARTs, ports and ADCs (of very different die size) while a display is built by pixels (red, green, blue subpixel for color displays, green chevron structure in Figure 1). Since the presentation of Google Glass, head worn displays (HWD) have gained public attention, however, this technology was used for decades in e.g. medical and military applications. HWDs are built with so called microdisplays (Figure 1 right) – the unique combination of microelectronics and displays with pixel sizes in the range of $10 \mu\text{m}$ for a 0.5” display. Liquid crystal (LC) or OLED material and dedicated layers are put on top of the semiconductor (wafer), which contains the driving electronics (transistors for individual pixel drive and matrix drivers, details see Figure 5).

This paper provides an overview of display technologies, display driving and optimizations. When designing systems with a display, the point to start with is the data (content) which is intended to be displayed. This set in combination with the observer distance the resolution and the size of the display. From the resolution (also relevant for software development), the requirements for the display interface and computing power (μC , display controller) can be derived.

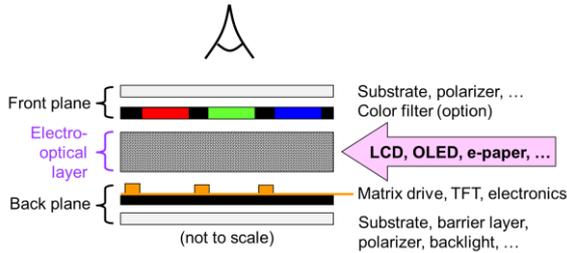


Figure 2: Cross section of a typical (color) display.



Figure 3: Display technologies: Reflective (top) with low power consumption and sunlight readability but mostly monochrome visualization. Emissive displays (bottom) with multimedia have high power consumption and are suitable for indoor use.

II. FUNDAMENTALS OF DISPLAYS

All major display technologies are somewhat similar in their basic structure (Figure 2). Detailed information on materials, layers, layouts, electrical and optical characteristics can be found in the corresponding chapters of [1] - [4]. The fundamental building blocks are:

- Substrate on front- and backside, mostly glass.
- TFTs and driving electronics (see Figure 5).
- Electro-optic layer which is determined by the display technology used like LCD, OLED and e-paper. This layer sets the requirements and properties of e.g. TFTs, panel electronics and front-(e-paper) or backlight (LCD).
- Color is mostly generated from white light (LEDs for LCDs, white OLED material) by color filters in front of the RGB subpixels.
- A polarizer towards the observer (top) is used due to function principle (LCD) or for reflection reduction (OLED).

Basically there are two types of displays which are distinguished by their optical behaviour (Figure 3):

- Reflective displays like monochrome LCDs with low resolution and e-paper displays (e.g. e-book readers). Their features are low power (as no light is generated) and the reflection of ambient light

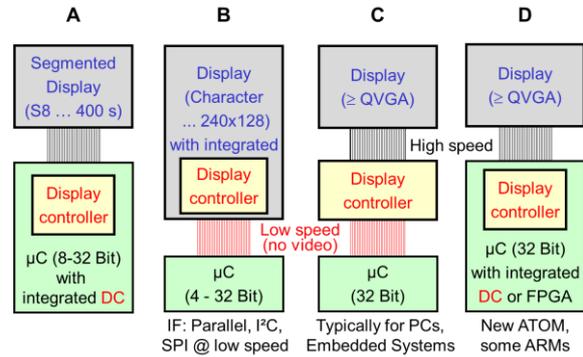


Figure 4: Approaches to connect microcontrollers (µC) and displays via display controller.

which ensures sunlight readability. However these technologies are not capable of reproducing colors suitable for images and modern HMIs.

- Emissive displays convert electrical power to light via the electro-optic layer (OLED, Plasma) or backlight (LCD). These displays are suitable for multimedia content with high switching speed incl. low lag touch operation. As the amount of the emitted light is “small” compared to sunlight, those displays are hardly readable in bright light.

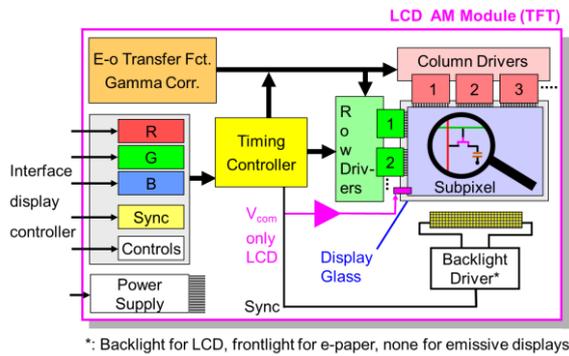
In consequence, there is no display technology yet (2014) which fits for all lighting conditions. Therefore, display system development deals often with the optimization of ambient light readability. If color and daylight readability is for example required (like in cars), high display luminance and reflection reduction is the way to choose e.g. AM LCDs and OLEDs.

III. DISPLAY ELECTRONICS

To show data on the display, a microcontroller (or microprocessor) has to deliver the content pixelwise. As displays require a steady stream of display data, a display controller is inserted between the microcontroller and the display. This display controller frees the real time load of the microcontroller, so the latter can perform tasks like A/D acquisition and calculations for the content shown on the display. It is worth to mention that modern high resolution display controllers (graphics adapters, graphics processing units) outperform the CPU(s) of the system regarding computing power and area on the die.

There are basically four possibilities (see Figure 4) for µC, display controller and display. The best solution depends on the resolution and system design:

- A: Low resolution displays like “Segment 8” (Figure 3 top left) are usually driven by microcontroller derivatives with built-in display controller (often named in data sheets as “LCD output”). Such systems are usually of low power and visualize values like time or temperature. For systems which are connected to a network, this might be



*: Backlight for LCD, frontlight for e-paper, none for emissive displays

Figure 5: Color LCD module with parallel interface as example for “panel electronics”, details see e.g. [4].

also a reasonable approach, as basic status data can be shown.

- B: Displays with built-in display controller are often named as “display modules” like character displays (e.g. 2 lines with 20 characters each) or low resolution black/white graphics (resolution e.g. 128 x 64). The microcontroller has to send only data via the interface when the display content has to be updated. Therefore this interface can be of low speed as typically no animations are shown and the display resolution is relatively low (up to QVGA). There are some QVGA color displays with built-in display controller which are able to show video-like animations at a reasonable update rate.
- C: High-resolution multimedia systems like PCs and high end embedded systems consist of three separate components (microcontroller, display controller, display). The benefit of this approach is that all three components can be chosen to their requirements as microcontrollers with built-in display controller (see D) are not widespread. Challenges are the two different high-speed interfaces (microcontroller to display controller and display controller to display) with different pinning and connectors.
- D: Microcontrollers with built-in display controllers have one interface less than “C” which saves cost and PCB space. However, there are only a few types available. The CPU and/or bus occupation load can reach more than 50% for VGA and larger resolutions. This approach is also highly suitable for FPGAs as IP blocks enable tailored performance.

Therefore the design of a system with a display is mainly determined by the display resolution and the data to be displayed (incl. video). Another important topic is the software effort to generate the data to be visualized. An operating system supports basic display features like text and graphics; more advanced GUI software is available from many suppliers.

In order to provide more examples of microelectronics and displays, “panel electronics” (Figure 5) is a highly suitable topic. This figure shows as an example a high resolution Active Matrix LCD (often named as TFT). Interfaces between display controller and higher resolution displays have either a parallel (TTL, only for low resolutions) or a serial (e.g. LVDS, HDMI, DISPLAYPORT) interface. It is easily noticeable that many ICs are needed to bring the interface signal to individual subpixel: The interface data are processed by the timing controller which modifies them according to the electro-optic layer (see Figure 2) characteristics and serializes them for the column drivers. The driving principle of the “display glass” is that rows (vertical) are subsequently selected and all grey level data are transferred to the pixel by the column drivers (horizontal). If the last row is reached, the next frame starts with the first row. This matrix principle keeps the number of contacts from the display glass to the driver ICs low as only the sum of rows (vertical resolution) plus the number of subpixels (horizontal resolution) applies. For direct driven displays like low-resolution segmented ones (see Figure 3 top left, Figure 4 “A”) each segment has to be connected to the display controller. Thus the number of segments is limited typically to about 100; higher numbers can be reached via multiplexing (“simple” matrix drive).

In consequence, systems with a display have many mutual relations which make optimizations challenging. The next paragraph provides three examples.

IV. SELECTED EXAMPLES FOR OPTIMIZATION

The following three examples are realized by Electronics Engineering students of Pforzheim University at different levels. All three demonstrate optimizations in terms of the whole system: The first one deals with microcontroller interfacing, the second with power reduction for AMLCDs and the third with improving the readability of emissive displays outdoors by software.

A. Microcontroller, Interface and GUI Software

As pointed out above, high resolution displays are not easy to glue to a microcontroller. Furthermore, the widespread 8-Bit microcontrollers are not able to support displays with higher resolution like monochrome QVGA (320 x 240) because their addressable memory is too low. One solution is to integrate an “intelligent” processor (as display controller) into the display module so that only simple commands like “print_text” or “draw_box” are needed. Such an intelligent display module causes basically a higher BOM (bill of material) than e.g. using a 32-Bit μ C with built-in display controller (Figure 4 “D”). However, for existing designs which need an upgrade of the

Example Code for Slider

```

uint16_t val = 50;
cmd_track(x, y, l, h, 34);

while(1) {
    Cmd(TAG(34));
    cmd_slider(x, y, l, h, 0, val, max);
    uint32_t tracker = rd32(REG_TRACKER);

    if ((tracker & 0xff) = 34)
        val = tracker >> 16;
}
    
```

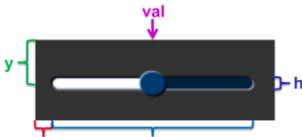


Figure 6: Code example for advanced display interfacing by an “intelligent” display controller with software support for quick and easy programming of GUIs.

display or low volume production, such a solution is highly suitable.

This approach has been evaluated during a study project. FTDI’s FT800 (EVE) was used as an “intelligent” display controller with I²C or SPI interface to the μ C and parallel interface to the display. This IC is capable of handling resolutions up to VGA and resistive touch screens (positions are sent to the μ C for processing of the inputs). An Arduino based μ C-system was selected as microcontroller because of some sample code. Figure 6 shows an example of the software to control RGB LEDs with ambient light and mood control by sliders. It is obvious that the workload for the programmer is reduced at large extent due to simple commands as shown for the slider control: A few lines of code are enough to set the position of the slider and its value (set by touch screen) can be read.

B. LCD power reduction

Modern color LCDs are equipped with a backlight (see Figure 5). The transmission of a color LCD glass is very low (~5%), so 95% of the power for the LED backlight is lost. Therefore it is highly attractive to reduce the power consumption. Most of the possible improvements can only be made by the display manufacturer like increasing the active pixel area. One possibility, which can be done as “added value” by system integrators, is to modify the backlight LED driving towards individual output. This method is called adaptive (or local) dimming and adjusts the output of each LED (via PWM) to the content of the display. More details can be found in [1] and [5]. The power reduction can reach 50% for typical video content; the value for HMIs depends strongly on the design (black background is favourable).

In a Master’s thesis, all necessary tasks have been implemented in a low cost FPGA. Figure 7 shows the fundamental approach for adaptive dimming:

- The light spread functions of each LED must be measured (top right) and stored in a compressed way (using functions and segments).
- With the RGB grey levels of the image to be displayed known, the impact of every LED to the

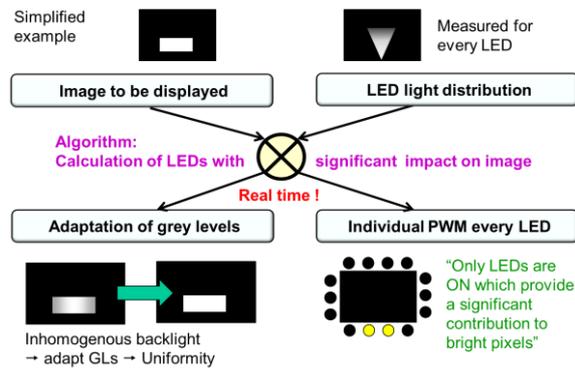


Figure 7: LCD power reduction by content based dimming of LED backlight. All data processing must be executed in real time.

luminance of this content has to be calculated (circle, center). This could be only performed in real time with reasonable effort by using e.g. thresholds and sector segmentation.

- For effective power reduction, not all LEDs are fully switched on (bottom right). Therefore some areas of the display are not backlit homogeneously (bottom left). This effect needs to be compensated via grey level modifications.

All tasks (including processing of the LED emission characteristics) must be performed in real time for all subpixels. As this method requires high computing power, it is only implemented nowadays in high end consumer electronic devices. However, professional displays would profit from this method as well like in central information displays in cars: Reducing display power (and in consequence the cars fuel consumption) is only one benefit, the larger ones might be less heat generation and longer lifetime of the LEDs due to lower temperature.

C. Image Enhancement for Outdoor Applications

Many displays are used outside sheltered rooms like smartphones, tablets, automotive displays and e-signage. Bright light reflections reduce the readability of emissive multimedia displays significantly. Mobile displays can be orientated by the user to avoid disturbing reflections, e.g. by turning the display and/or moving into shadow. This is however not possible in cars and for public displays. High luminance and reflection reduction are improvement strategies but they raise power consumption (causes heat which reduces lifetime) and cost. A more efficient way is to modify grey levels of the image by algorithms which deal with display characteristics and human vision. A typical example is a night scene with a few street lights - all dark content is overridden by ambient reflections. The solution is therefore to raise the lower grey levels without noticeable distortion of the grey scale impression and colors.

Figure 8 provides examples of a PhD thesis performed at the Display Lab of Pforzheim University

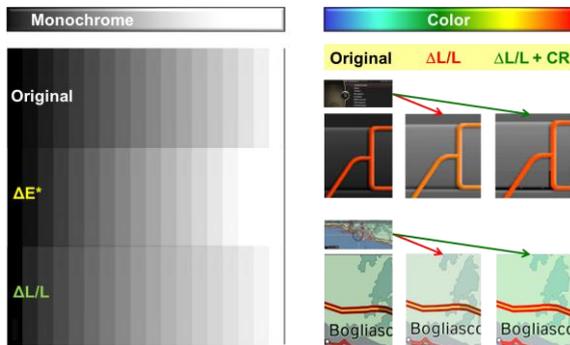


Figure 8: Results of image enhancement algorithms (calculated for 10 klx) for monochrome grey scale (left) and automotive HMI examples (right) with enlarged details demonstrate the effect of color rendering (“+ CR”). These effects can be experienced by showing this figure on a color display at bright light conditions. The actual reproduction depends strongly on the characteristics of the output device (printer, monitor).

(current results see [6]): (current results see [6]): Two algorithms basing on display measurements and human vision (called “ ΔE^* ” and “ $\Delta L/L$ ”) were developed. The left part shows the effect on grey scale reproduction. It is clearly visible, that “ $\Delta L/L$ ” results in distinguishable dark grey levels and discrimination of bright ones (clipping for “ ΔE^* ”).

For color images (right), the original color should be maintained as far as possible. However, this is not achievable in any case due to luminance limits for every color (only white has the maximum luminance, blue is e.g. 10% of that). Using only grey level enhancements like “ $\Delta L/L$ ” (center) results in noticeable color shifts. This might be disturbing for modern automotive HMIs and is not intended by designers. By implementing color rendering algorithms (CR, right), the original color is close to the original one (left).

Our methods were evaluated successfully including the influence of driver glare (light from sky, road, and landscape “bothering” the driver’s eye) and provide better results for automotive non-photorealistic HMIs like other algorithms which are optimized for images and video (photorealistic rendering). The method is applicable to all display technologies (except b/w), even for reflective e-paper there was a noticeable improvement.

V. SUMMARY

Modern systems equipped with a display set various challenges towards optimum implementation. From the point of display technology one has to choose between lowest power, monochrome and outdoor use (all features of reflective displays) and multimedia performance with high resolution but also high power and limited outdoor readability (emissive displays).

Another important topic in display integration is the connection of the system’s microcontroller and the display. This is done via a display controller, which

can be integrated either in the μC or the display module or implemented as separate IC. The optimum solution strongly depends on the display resolution and series volume.

Improvements and optimizations of display systems cover all subassemblies of display hard- and software. Three examples for easy interfacing, power reduction for AMLCDs and enhancement of outdoor readability by software were provided.

ACKNOWLEDGEMENT

The author thanks all organizations (like BMBF and BMWi) and companies for the support of various activities in the field of electronic displays for the Display Lab of Pforzheim University.

REFERENCES

- [1] J. Chen, W. Cranton, M. Fihn (eds), “*Handbook of Visual Display Technology*,” Springer, Heidelberg 2012.
- [2] J-H Lee, D. N. Liu, S-T Wu, “*Introduction to Flat Panel Displays*,” Wiley, New York 2008.
- [3] S.T. Wu, D.-K. Yang, “*Fundamentals of Liquid Crystal Devices*,” Wiley, New York 2006.
- [4] W. den Boer, “*Active Matrix LCDs*,” Newnes, Atlanta 2011.
- [5] S. Kobayashi, “*LCD Backlights*,” Wiley, New York 2009.
- [6] K. Blankenbach, A. Sycev, S. Kurbatfinski, M. Zobl, „Optimization and Evaluation of Automotive Displays under Bright Ambient Light using Novel Image Enhancement Algorithms,” *SID Symposium Digest* 45, 652 – 655 (2014).



Karlheinz Blankenbach holds a M.Sc. (Diplom) in Physics and a Ph.D. degree, both from the University of Ulm, Germany. His industrial experience was at AEG (a subsidiary of DAIMLER), Ulm where he developed display electronics and LCDs for public information systems. In 1995 Karlheinz was appointed as full professor at Pforzheim University. He is chairman of “Electronic Displays” conference (edC) and of the German Flat Panel Forum (DFP).

A Millimeter-Wave Power Amplifier Concept in SiGe BiCMOS Technology for Investigating HBT Physical Limitations

Jonas Wursthorn, Herbert Knapp, Bernhard Wicht

Abstract—A millimeter-wave power amplifier concept in an advanced silicon germanium (SiGe) BiCMOS technology is presented. The goal of the concept is to investigate the impact of physical limitations of the used heterojunction bipolar transistors (HBT) on the performance of a 77 GHz power amplifier. High current behavior, collector-base breakdown and transistor saturation can be forced with the presented design. The power amplifier is manufactured in an advanced SiGe BiCMOS technology at Infineon Technologies AG with a maximum transit frequency f_T of around 250 GHz for npn HBT's [1]. The simulation results of the power amplifier show a saturated output power of 16 dBm at a power added efficiency of 13%. The test chip is designed for a supply voltage of 3.3 V and requires a chip size of 1.448 x 0.930 mm².

Index Terms—Millimeter-wave, Power Amplifier, SiGe, BiCMOS.

I. INTRODUCTION

The demand for vehicle safety is steadily being redefined and tightened by the European New Car Assessment Programme (EURONCAP). Since 2014 it is practically impossible to get a 5-star EURONCAP rating for a new vehicle without an autonomous emergency breaking system (AEB). Automotive manufacturers use frequency modulated continuous wave (FMCW) based radar systems to realize such systems. FMCW radar systems can detect the relative velocity and the distance to the vehicle ahead. These attributes make them also suitable for advanced driver assistance systems like lane change assistants or blind spot detectors. Video and imaging systems might be used additionally to improve the object recognition on the road.

The block diagram of a bi-static (separate transmitter/receiver antenna) FMCW based radar system is shown in Figure 1. A voltage controlled oscillator (VCO) is used as a radio-frequency source that feeds the power amplifier as well as the mixer in the receiver

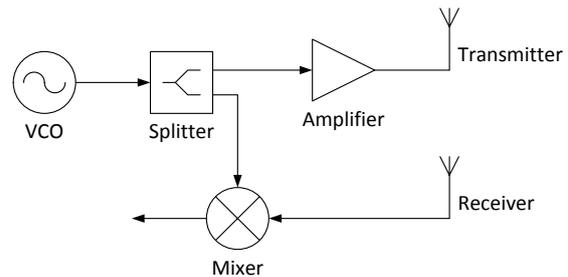


Figure 1: Simplified block diagram of a bi-static FMCW radar system.

channel. The power amplifier offers a high signal level to the transmitter antenna. The receiver antenna delivers the signal to the mixer input, where it is mixed with the actual local oscillator signal of the VCO. From the time and frequency shift between the two signals the distance and the relative velocity are calculated.

In order to detect the reflected signal at the radar module properly, a minimum signal to noise ratio is required. This ratio depends on several parameters but is in the end limited by the power level delivered from the amplifier to the transmitter antenna. The design of the power amplifier is challenging, because physical limitation effects like high current operation or breakdown are often not represented accurately in all transistor models. Measurement results on the power amplifier described in this work are expected to show how an operation close to the physical limits affects its performance, leading to more confident decisions regarding power amplifier design.

The design considerations are described in detail starting with the amplifier topology in Section II. Section III explains how high current effects can be forced with the design. To run the transistors in saturation or breakdown, the bias voltage at the common-base stage can be varied. The bias voltage generation is explained in IV. The implemented test circuit layout is described in Section V. Overall simulation results of the power amplifier can be found in VI. A conclusion is given in Section VII.

II. POWER AMPLIFIER TOPOLOGY

As further automotive circuit designs will be based on the investigations of the designed power amplifier,

Jonas Wursthorn (Jonas.Wursthorn@infineon.com) and Herbert Knapp are with Infineon Technologies AG, Am Campeon 1-12, 85579 Neubiberg, Germany.

Bernhard Wicht is with the Robert Bosch Center for Power Electronics, Reutlingen University, Alteburgstraße 150, 72762 Reutlingen, Germany.

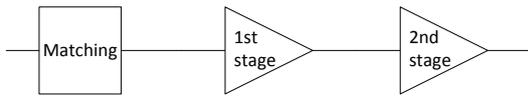


Figure 2: Block diagram of the power amplifier topology.

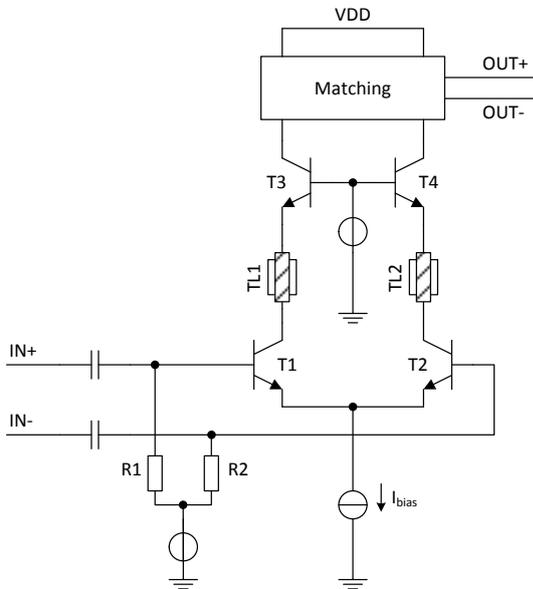


Figure 3: Simplified schematic of the second power amplifier stage.

it has to cover ambient temperatures from -40°C to 125°C . To ensure a preferably constant output power over the temperature range a multi-stage amplifier (2 stages) is chosen wherein the single stages are run in compression mode. This operation mode acts like a buffer if the output power of the previous stage drops due to higher temperature. The gain reduction at high temperatures is mainly caused by the lower current gain and transit frequency in the transistor [2].

In order to have a defined $50\ \Omega$ input impedance, a matching network is designed. It consists of a transmission line in series and a capacitance to ground. For the interstage matching between the first and second stage the maximum output power is the main criteria – not a defined impedance level. Therefore the length of the transmission lines between the stages is adjusted. Figure 2 shows a block diagram of the signal path resulting from the mentioned considerations. The matching is realized with transmission lines. The topology of the first and second amplifier stage is similar – the only differences are the device dimensions.

A simplified schematic of the second amplifier stage is shown in Figure 3. The RF signal (IN+/IN-) is fed to a differential common-emitter stage which is DC biased by R1 and R2. A current source I_{bias} at the emitter node is preferred over a resistor for mainly two reasons. On the one hand, the base-emitter diodes of T1 and T2 have a rectifying effect on the applied RF signal, which results in a different DC voltage at the

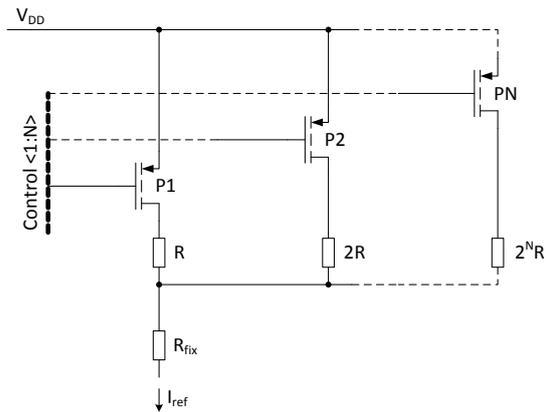


Figure 4: Generation of different reference currents for the current mirror network.

emitters leading to different currents depending on the signal peak when using a resistor. On the other hand the current is limited in case of breakdown when using a current source.

Transmission lines TL1 and TL2 represent the parasitics of the connection between the common-emitter stage (T1/T2) and the common-base stage (T3/T4). The common-base stage offers a low impedance ($1/g_{m3/4}$) to the common-emitter stage. This is essential because otherwise the Miller capacitance between base and collector of T1/T2 would have a huge impact on the amplifiers gain (low pass behavior). The Cascode topology reduces this impact to a minimum as it keeps the collector of T1/T2 practically grounded in terms of RF. For the common-base stage, C_{BC} does not appear between the input and output and is therefore less critical because it does not act as a Miller capacitance.

III. HIGH CURRENT EFFECTS

If a certain limit for the current density in a bipolar transistor is exceeded, the transit frequency is reduced due to the Kirk effect. For a SiGe HBT this effect is shifted to higher frequencies but then f_T drops even faster than for common bipolar transistors [3].

The current density suggested for product design in the used technology is $13\ \text{mA}/\mu\text{m}^2$. As there is a margin for products, the current density of the considered power amplifier should be increasable to around twice this value. This is realized with a digitally adjustable current mirror. CMOS transistors are switched on/off to increase/decrease the reference current of the current mirror like shown in Figure 4. The design uses $N=5$ bits to vary the current density from 4 to $25\ \text{mA}/\mu\text{m}^2$.

High current effects are not represented in the used transistor models so far. Accordingly, there are no simulations showing this effect. For measurements it is expected that the output power will increase with higher currents exceeding the nominal current density. For values far above the suggested current density the

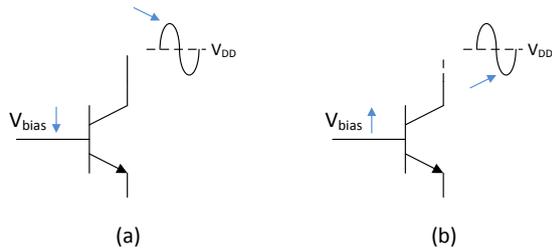


Figure 5: Bias voltage conditions for (a) breakdown and (b) saturation of the used HBT.

output power is expected to drop rapidly due to high current effects in the transistor [2] [3].

IV. BIAS VOLTAGE GENERATION

The region for a useful biasing voltage of the common-base stage is mainly defined by the voltage swing at the collector of the transistor and is nominally between 2 and 2.2 V. If the voltage swing reaches its maximum, the collector-emitter breakdown voltage may be exceeded (see Figure 5(a)) which leads to avalanche multiplication. This effect defines the lower limit of the biasing voltage.

If the output swing is at its minimum, the transistor might enter the saturation region (see Figure 5(b)). This means the base-collector diode will be forward biased, resulting in a malfunction of the power amplifier.

The base bias voltage is generated in a circuit according to Figure 6. The `pmosCM_enable` bit allows to switch between a PMOS-based and an NMOS-based current mirror. The PMOS-based circuit is used to generate low bias voltages, the NMOS-based part covers the upper voltage range close to V_{DD} . This results in an overlapping bias voltage range from 0.4 to 2.9 V. For measurements it is expected that there is a range within these two voltage limits where the output power is nearly constant.

V. IMPLEMENTED TEST CIRCUIT

The various digital input pins for adjusting the current density or setting the common-base bias voltage are controlled by a serial control interface. The available measurement equipment is only capable of single ended RF signal generation. As the core part of the power amplifier expects a differential signal a balun (balanced-unbalanced) network is required. This network consists of transmission lines and capacitors and is based on the principle described in [4].

A layout of the complete test circuit is shown in Figure 7. Biasing and serial control interface are placed in the left part of the die. The right-hand side contains the RF part with the input pads on the bottom and the differential output pads on the top. Except for

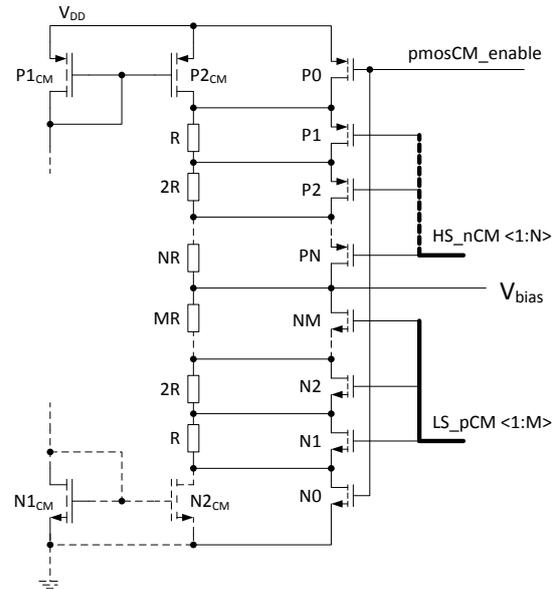


Figure 6: Simplified schematic of the common-base biasing network.

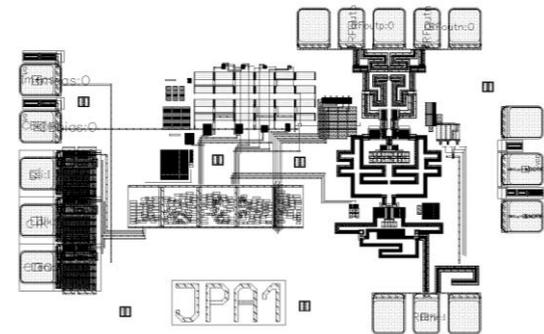


Figure 7: Layout of the implemented test circuit. The RF part is on the right-hand side, biasing and digital part on the left-hand side. The chip size is 1.448 mm x 0.930 mm.

the balun and matching network, the RF layout is highly symmetrical.

VI. SIMULATION RESULTS

The simulation results of the complete chip for the output power P_{out} and the power added efficiency

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}}$$

are shown in Figure 8 and Figure 9, respectively. A linear gain of ≥ 25 dB is achieved. The saturated output power is 16 dBm for room temperature. Due to the multi-stage approach the difference in output power (in compression) is less than 1 dBm over the complete temperature range. The PAE simulation also includes the power consumption of the biasing network and the serial control interface and can therefore be further increased for a stand-alone power amplifier. In this configuration the simulation shows a PAE of 13%.

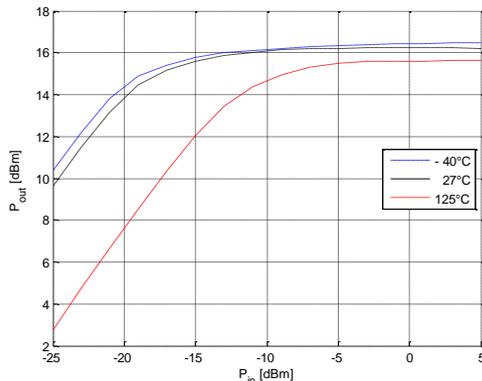


Figure 8: Simulated output power P_{out} vs. input power P_{in} of the complete chip for the automotive temperature range.

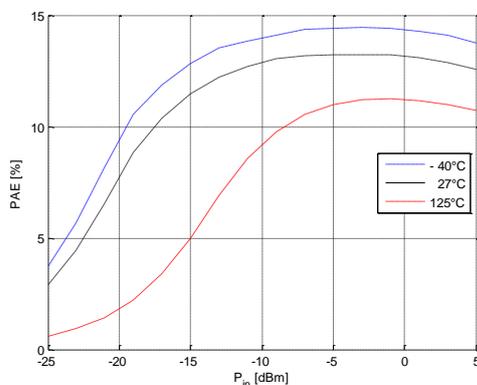


Figure 9: Simulated power added efficiency PAE vs. input power P_{in} of the complete chip for the automotive temperature range.

VII. CONCLUSION

A 77 GHz SiGe BiCMOS power amplifier concept for investigating physical limitations has been presented. The power amplifier simulation shows an output power of 16 dBm with a 13% PAE at 27°C. A serial control interface allows setting the current density in the power amplifier stages and the bias voltage for the common-base stage. The bias voltage can be varied from 0.4 to 2.9 V. The current density can be increased up to 25 mA/μm² to force high current effects.

VIII. ACKNOWLEDGEMENT

The authors wish to acknowledge the DOTSEVEN project (316755) supported by the European Commission through the Seventh Framework Programme (FP7) for Research and Technology Development. Furthermore, I would like to thank Thomas Kurth for giving me the opportunity to write my Master Thesis at Infineon Technologies AG which was the basis for this work.

REFERENCES

- [1] P. Chevalier, T. F. Meister, B. Heinemann, S. van Huylbroeck, W. Liebl, A. Fox, A. Sibaja-Hernandez and A. Chantre. „Towards THz SiGe HBTs”, *IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 57-65, 2011.
- [2] M. Reisch, *High-Frequency Bipolar Transistors*, Springer Verlag, 2003.
- [3] L.E. Larson, „Silicon Bipolar Transistor Design and Modeling for Microwave Integrate Circuit Applications”, *Proceedings of the 1996 Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 142-148, 1996.
- [4] W. Bakalski, W. Simbürger, H. Knapp, H.-D. Wohlmuth and A. L. Scholz, „Lumped and Distributed Lattice-type LC-Baluns”, *IEEE MTT-S International Microwave Symposium Digest*, volume 1, pp. 209-212, 2002.



Jonas Wursthorn received the Bachelor degree in Electrical Engineering from DHBW Stuttgart in 2010. After working one year with Infineon Technologies AG as a technical assistant for radio frequency circuit design and test he started a Master degree program for Power and Microelectronics at Reutlingen University which he finished 2013. Afterwards he joined Infineon as a Ph.D. student.



Herbert Knapp received the Diploma and Ph.D. degrees in Electrical Engineering from the Technical University Vienna, Austria, in 1997 and 2000, respectively. In 1993 he joined Siemens, Corporate Technology, in Munich, Germany, where he worked on circuits for wireless communications and high-speed data transmission. He is now with Infineon Technologies, Munich, Germany, and is engaged in the design of circuits for automotive radar applications.



Bernhard Wicht received the Diploma degree from Technical University Dresden in 1996 and the Ph.D. degree from the Technical University Munich in 2002. 2003-2010, he was with the Mixed Signal Automotive business unit of Texas Instruments in Freising, Germany, responsible for the development of automotive smart power ICs. Since September 2010 he is professor for integrated circuits at Reutlingen University, Robert Bosch Center for Power Electronics.

Substratkoppeln in schnell schaltenden integrierten Leistungsendstufen

Christoph Rindfleisch, Jürgen Wittmann, Bernhard Wicht

Zusammenfassung—Durch schnell schaltende Leistungsendstufen werden durch kapazitive Umladeströme Störungen ins Substrat und in empfindliche Schaltungselemente eingekoppelt, die dort zur Störung der Funktion führen können. In dieser Arbeit werden Substratstrukturen zur gezielten Ableitung dieser Störungen vorgestellt und ihre Wirksamkeit mit Hilfe von Device Simulation evaluiert. Ohne Ableitstrukturen kann eine Potentialanhebung des Substrats bis zu 20 V entstehen. Die Untersuchungen belegen, dass die Potentialanhebung durch p-Typ Guard-Ringe um 75 %, durch leitende Trenches um 88 % sowie durch Rückseitenmetallisierung um nahezu 100 % reduziert werden kann.

Schlüsselwörter—Substratkoppeln, Störeinkopplung, Leistungsendstufe, schnell schaltend, High-Side Transistor.

I. EINLEITUNG

Leistungsendstufen dienen der aktiven Steuerung des Energieflusses in vielfältigen Anwendungen wie der Motorsteuerung oder Spannungswandlung. In Schaltwandlern wird durch ein pulsweitenmoduliertes (PWM) Ein- und Ausschalten eines Leistungstransistors und durch anschließendes Filtern beispielsweise eine niedrigere Ausgangsspannung erzeugt. Die Erhöhung der Schaltfrequenz erlaubt eine Verkleinerung des Filters und reduziert somit die Kosten eines Schaltwandlers. Abbildung 1 zeigt die Implementierung eines Abwärtsschaltwandlers, der Eingangsspannungen bis 40 V auf Ausgangsspannungen unter 5 V regelt und mit Schaltfrequenzen über 10 MHz arbeitet. Ein NMOS Leistungstransistor ist vor allem bei höheren Schaltfrequenzen gegenüber einem PMOS bevorzugt, da der benötigte Einschaltwiderstand mit weniger parasitären Kapazitäten realisiert werden kann und somit der Wirkungsgrad des Wandlers steigt.

Da sich das Referenzpotential HSGND (Source-Anschluss des NMOS Transistors) mit dem Schaltzustand

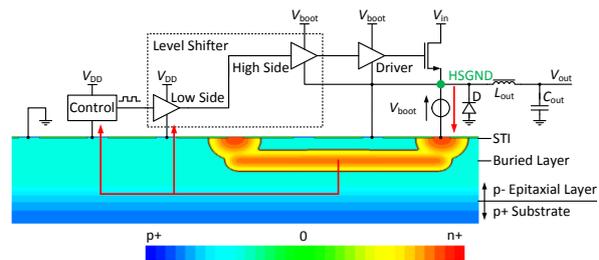


Abbildung 1: Schaltplan sowie Substratstruktur des untersuchten Abwärtsschaltwandlers.

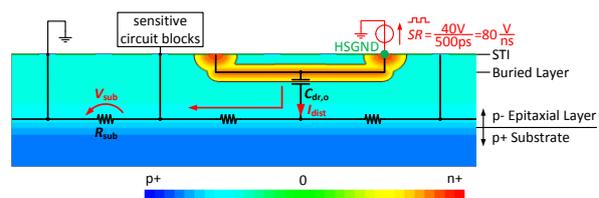


Abbildung 2: Substratstruktur mit parasitären Bauelementen und Koppelpfad.

des Transistors ändert, muss dieser durch einen Gate-treiber auf der High-Side zwischen HSGND bzw. $HSGND + 5\text{ V}$ (V_{boot}) angesteuert werden. Das PWM Signal, das durch eine Regelschaltung auf der Low-Side mit einer Versorgungsspannung von 5 V erzeugt wird, wird durch einen Levelshifter auf das Potential der High-Side übertragen.

Der in Abbildung 1 gezeigte Wandler wurde in einer 180nm BiCMOS-Technologie mit hoch dotierter p-Substratschicht und niedrigdotierter p-Epitaxieschicht implementiert. Um sowohl die High-Side, als auch die Low-Side Schaltungselemente, auf dem gleichen Substrat integrieren zu können, ist aufgrund deren unterschiedlicher Spannungsdomänen eine Isolation erforderlich. Die Raumladungszone einer n-dotierten Schicht wird als Isolationswanne eingesetzt. Diese besteht aus einer hoch dotierten vergrabenen Schicht (Buried Layer) unterhalb der High-Side-Schaltungselemente zur vertikalen Isolation und einem hoch dotierten tiefen n-Gebiet (Sink) zur lateralen Isolation.

Damit die Isolation in allen Betriebszuständen aufrechterhalten wird, muss sich das n-dotierte Gebiet der Wanne immer auf einem höheren Potential befinden als das p-dotierte Gebiet im Inneren und Äußeren der Wanne. Im Schaltungsdesign wird die Wanne daher

mit der Versorgungsspannung der High-Side-Schaltelemente (V_{boot}) verbunden. Da V_{boot} durch eine Bootstrap Schaltung immer auf 5 V über HSGND geladen wird, hat diese Verbindung zur Folge, dass sich das Potential V_{boot} der Wanne mit dem Schaltungszustand des High-Side-Transistors zwischen ~ 5 V und $V_{in} + 5$ V ändert, was eine Umladung der Raumladungszonenkapazität zwischen der Isolationswanne und dem Substrat ($C_{dr,o}$) zur Folge hat. Der daraus resultierende Störmechanismus ist in Abbildung 2 gezeigt. Die ins Substrat gelangten Umladeströme verbreiten sich als Substratströme (I_{dist}) über das Substrat und führen zu einem Spannungsabfall (V_{sub}) am Substratwiderstand (R_{sub}) und somit zu einem sich ändernden Substratpotential. Da das Substratpotential für die meisten Schaltelemente dem Referenzpotential der Low-Side entspricht, führt der Spannungsabfall zu einer Einkopplung in die Low-Side-Schaltelemente.

Das Ableiten des Störstroms erfolgt über die Massekontakte des Substrats und über Schaltelemente, die intern eine Masseverbindung aufweisen. Innerhalb der Schaltelemente kann es zu einer Überlagerung des eingekoppelten Stroms mit Signalströmen und so ebenfalls zu einer Funktionsstörung kommen.

Die Höhe des Umladestroms hängt einerseits von der Raumladungszonenkapazität $C_{dr,o}$ und dem Substratwiderstand R_{sub} ab, welche fest durch die gewählte Technologie und die realisierte Schaltung vorgegeben sind. Andererseits hat die Höhe des Spannungssprungs und dessen Anstiegsgeschwindigkeit an der High-Side-Isolationswanne, d.h. an HSGND und V_{boot} , maßgeblichen Einfluss. Je höher die Anstiegsgeschwindigkeit, umso höher sind die enthaltenen hochfrequenten Anteile und damit die Amplitude des Umladestroms I_{dist} .

Kleine Tastverhältnisse des PWM Signals bei hohen Umsetzverhältnissen von $V_{IN} > 40$ V auf $V_{OUT} < 5$ V erfordern bei Schaltfrequenzen über 10 MHz Anstiegs- und Abfallzeiten im Bereich von 500 ps am Schaltknoten (HSGND). Eine höhere Flankensteilheit wirkt sich außerdem vorteilhaft auf die Reduzierung der Schaltverluste aus. Somit kann die Amplitude der Substratströme kaum beeinflusst werden. Daher muss entweder die Immunität der gestörten Schaltelemente verbessert oder die Einkopplung der Störungen in diese Schaltelemente verringert werden. Nachfolgend werden Substratstrukturen zur gezielten Ableitung der Störungen aus dem Substrat vorgestellt und ihre Effektivität mittels Prozess- und Device-Simulationen evaluiert.

II. REDUKTION DES SUBSTRATKOPPELNS

Für die Evaluation der Effektivität verschiedener Ableitstrukturen mittels einer Prozess- und Device-Simulation wurde zunächst eine Isolationswanne („Well1“) sowie eine Störsecke („Sink1“) nach Abbildung 3a implementiert. Um die Simulationszeit klein zu halten, wurde aus Symmetriegründen nur die Hälfte einer Iso-

lationswanne mit einer Größe von $70 \mu\text{m} \times 70 \mu\text{m}$ simuliert, was etwa einem Fünftel der Größe einer Isolationswanne entspricht, die im untersuchten Abwärts-wandler für die Isolation aller High-Side-Schaltelemente eingesetzt wird. Da die Raumladungszonenkapazität und damit auch der Störstrom mit der Größe der Wanne skalieren, wird erwartet, dass auch die eingekoppelten Störungen dementsprechend skalieren.

Die Störsecke, bestehend aus einer großflächig hochdotierten p-Struktur zwischen zwei niederdotierten Substratkontakten, ist mit einem Abstand von $217 \mu\text{m}$ zum Mittelpunkt der Isolationswanne platziert und mit Masse verbunden. Während der durchgeführten Transientensimulation werden die Kontakte der Isolationswanne mit einem Rechteckpuls mit einer Amplitude von 40 V und einer Anstiegs- und Abfallzeit von 500 ps beaufschlagt.

A. Simulation ohne Ableit- und Isolationsstrukturen

Die Simulation der Dotierstruktur in Abbildung 3a ergab die in Abbildung 3b gezeigte Stromverteilung. Aufgrund des erheblich geringeren Widerstandes der Substratschicht gegenüber der Epitaxieschicht fließt der größte Teil des Substratstroms direkt in die Substratschicht, wo er sich über das gesamte Substrat verbreitet und somit jede auf dem Substrat befindliche Schaltung, unabhängig von ihrer Entfernung zur Isolationswanne, stören kann.

Aufgrund der Masseanbindung des Substrats an der Oberfläche der Epitaxieschicht müssen die eingekoppelten Störungen über die hochohmige Epitaxieschicht gegen Masse abfließen. Die Aufteilung der Ableitströme auf die einzelnen eingebrachten Masseanbindungen erfolgt gemäß dem Widerstand zwischen der Masseanbindung und der Substratschicht. Je geringer der Widerstand, umso mehr Strom fließt über die jeweilige Struktur. Bestätigt wird diese Überlegung durch die hohe Stromdichte unter der großflächig hochdotierten Struktur gegenüber den kleinflächigen nieder dotierten Substratkontakten. Folglich lässt sich der Stromfluss über eine Struktur über ihre Fläche und ihren flächenspezifischen Widerstand beeinflussen. Um die Stromeinkopplung in Schaltelemente zu verringern ist es daher erforderlich, Ableitstrukturen mit einem geringen Widerstand zwischen Masse und der Substratschicht ins Substrat einzubringen. Nur so kann der relative Anteil des über Schaltelemente abfließenden Störstroms effektiv verringert werden. Damit die Kosten durch die zusätzlich benötigte Chipfläche klein gehalten werden, sollten darüber hinaus Strukturen mit einem niedrigen flächenspezifischen Widerstand verwendet werden.

Abbildung 3c zeigt, dass für die simulierte Wannengröße ein Spannungshub des Substratpotentials von bis zu 4 V zu erwarten ist. Hochskaliert auf die tatsächliche Wannengröße entspricht dieser einem Spannungshub von bis zu 20 V bei Spitzenströmen im Substrat von bis zu 130 mA. Der Betrag der Spannung wird dabei vom

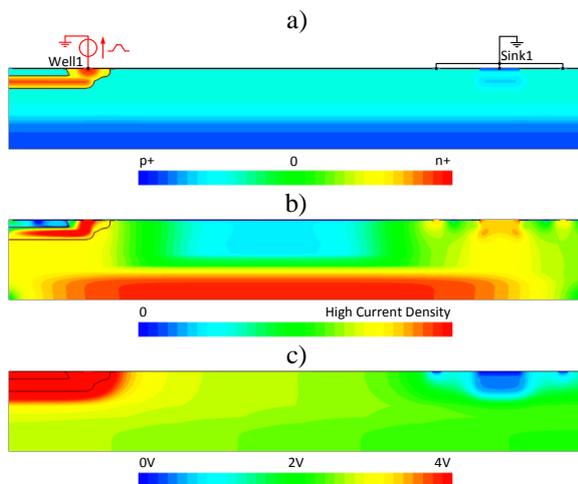


Abbildung 3: a) Dotierstruktur ohne Ableit- und Isolationsstrukturen; b) Simulierte Stromverteilung zum Zeitpunkt des größten Stromflusses während der steigenden Flanke; c) Resultierende Potentialverteilung im Substrat.

Widerstand zwischen der Substratschicht und der Masse bestimmt. Je geringer der Widerstand der eingebrachten Ableitstruktur, umso geringer auch der resultierende Spannungshub.

B. P-Typ Guard-Ringe und n-Typ Guard-Ringe

P-Typ Guard-Ringe [1] zur Ableitung der Störströme und n-Typ Guard-Ringe [1] zur Unterdrückung des lateralen Stromflusses wurden als eine der Optionen zur Reduktion der Störeinkopplung untersucht. Abbildung 4a zeigt die simulierte Dotierstruktur. Neben der Wanne und der Störsecke befinden sich zwei 17,5 μm breite p-Typ Guard-Ringe mit einer Entfernung von 56 μm („pGR1“) und 98 μm („pGR2“) zum Mittelpunkt der Isolationswanne. Weiterhin wurden zwei n-Typ Guard-Ringe mit einer Entfernung von 77 μm („nGR1“) und 119 μm („nGR2“) vom Mittelpunkt der Isolationswanne eingebracht.

P-Typ Guard-Ringe und n-Typ Guard-Ringe sind oft als Standardbauelemente in vielen Technologie verfügbar oder können durch die Verwendung aller p bzw. n-Strukturen einer Technologie erzeugt werden. Nachteilig ist, dass sie aufgrund von Ausdiffusionseffekten einen erhöhten zusätzlichen Flächenbedarf aufweisen.

Abbildung 4b zeigt die simulierte Stromverteilung im Substrat bei Verwendung von p-Typ und n-Typ Guard-Ringen. Verglichen mit der Stromverteilung der Simulation ohne Ableitstrukturen kann eine erhebliche Reduktion der Stromdichte unterhalb der Störsecke beobachtet werden. Simulationen ergaben eine Reduktion des in die Störsecke eingekoppelten Stroms um 64 %. Die Effektivität von p-Typ Guard-Ringen wird maßgeblich durch die zur Verfügung stehenden Dotierschichten bestimmt und ist somit abhängig von der genutzten Technologie. Da ein p-Typ Guard-Ring eine höhere Dotierung zwischen dem Massekontakt und der

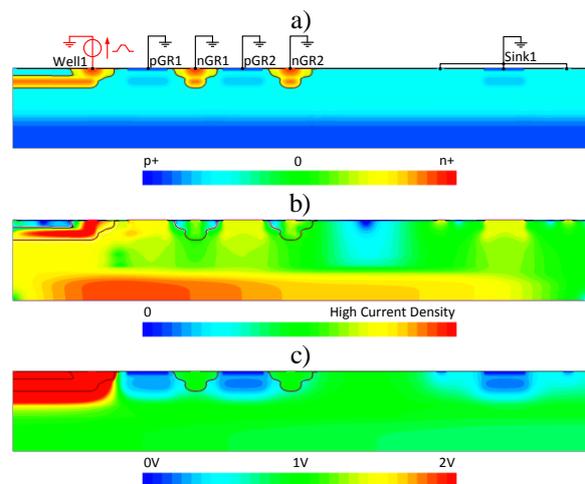


Abbildung 4: a) Zugrunde gelegte Dotierstruktur mit p-Typ Guard-Ringen und n-Typ Guard-Ringen; b) Simulierte Stromverteilung der Dotierstruktur aus a) zum Zeitpunkt des größten Stromflusses während der steigenden Flanke; c) Aus der simulierten Stromverteilung in b) resultierende Potentialverteilung.

Substratschicht aufweist, verringert sich der flächenspezifische Widerstand. Mehr Substratstrom kann abgeleitet werden.

In Abbildung 4b ist weiterhin zu erkennen, dass die in die Epitaxieschicht eingebrachten n-Typ Guard-Ringe aufgrund des nahezu nicht vorhandenen lateralen Stromflusses in der Epitaxieschicht keine Isolationswirkung aufweisen. Da der laterale Stromfluss hauptsächlich in der niederohmigen Substratschicht beobachtet werden kann, müsste der n-Typ Guard-Ring für eine hohe Isolationswirkung über das gesamte Substrat reichen. Tiefere n-Typ Guard-Ringe stehen in den meisten Technologien jedoch nicht zur Verfügung. Abbildung 4c zeigt den resultierenden Spannungsabfall über dem Substrat. Gegenüber der Simulation ohne Ableit- und Isolationsstrukturen konnte dieser um 75 % reduziert werden.

C. Leitende Trenches

Einen noch geringeren flächenspezifischen Widerstand, und somit eine höhere Effektivität pro Fläche und einen geringeren Flächenbedarf, bieten leitende Trenches [2]. Ein bis zum Substrat geätzter Graben wird mit einem hoch leitfähigem Material (z.B. Polysilizium) aufgefüllt und erzeugt somit eine niederohmige Verbindung des Masseknotens zum Substrat. Abbildung 5a zeigt die simulierte Dotierstruktur, bestehend aus der Isolationswanne, der Störsecke, sowie einem mit 56 μm Abstand zum Mittelpunkt der Isolationswanne platzierten leitenden Trench („Trench1“). Die Breite des leitenden Gebietes des Trenches entspricht in etwa der Breite einer der zuvor simulierten p-Typ Guard-Ringe.

Die simulierte Stromverteilung in Abbildung 5b zeigt, dass der Stromfluss über die Störsecke gegenüber der Simulation ohne Ableit- und Isolationsstrukturen

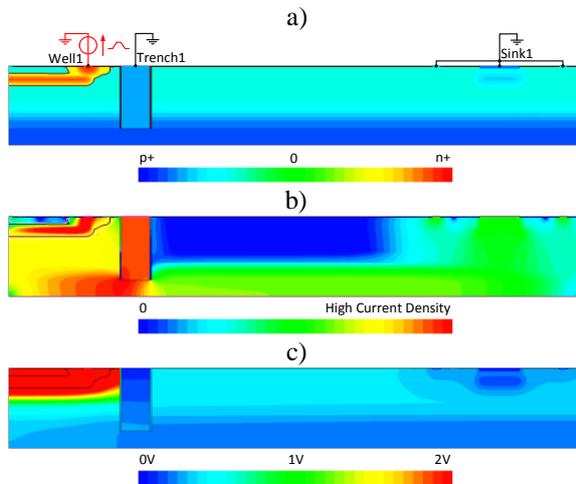


Abbildung 5: a) Dotierstruktur mit leitendem Trench; b) Simulierte Stromverteilung zum Zeitpunkt des größten Stromflusses während der steigenden Flanke; c) Resultierender Potentialverteilung im Substrat.

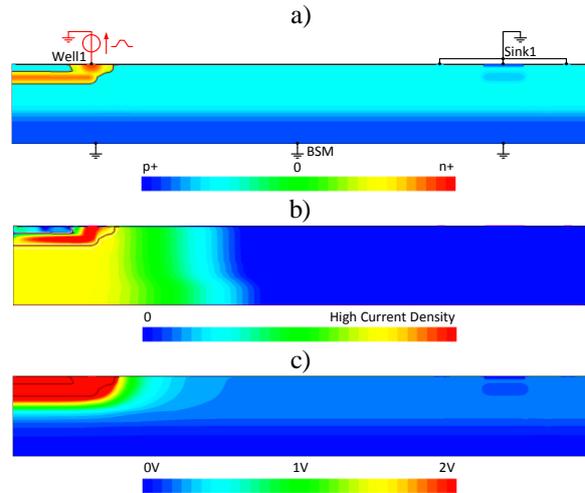


Abbildung 6: a) Dotierstruktur mit Rückseitenmetallisierung; b) Simulierte Stromverteilung zum Zeitpunkt des größten Stromflusses während der steigenden Flanke; c) Resultierender Potentialverteilung im Substrat.

(siehe Abbildung 3a) um 92 % reduziert wird. Aber auch für leitende Trenches gilt, dass die Tiefe des Trenches dessen Effektivität bestimmt. Die höchste Ableitfähigkeit kann nur mit Trenches erreicht werden, die mindestens bis zur niederohmigen Substratschicht reichen. Wie der simulierte Spannungsabfall in Abbildung 5c zeigt, konnte aufgrund des geringen Widerstandes des Trenches der Spannungsabfall im Substrat gegenüber der Simulation ohne Ableit- und Isolationsstrukturen um 88 % reduziert werden.

D. Rückseitenmetallisierung

Bei Verwendung einer Rückseitenmetallisierung (Backside Metallization „BSM“) [1, 3] wird eine metallisch leitende Schicht direkt auf die Substratschicht auf der Rückseite des Wafers aufgetragen und direkt mit Masse verbunden. Die Substratströme müssen so nicht über die hochohmige Epitaxieschicht nach Masse abgeleitet werden. Eine Rückseitenmetallisierung kann unabhängig von der verwendeten Technologie durch zusätzliche Fertigungsschritte realisiert werden, erhöht aber die Fertigungs- und Verpackungskosten eines ICs.

Abbildung 6a zeigt die Dotierstruktur bei aufgetragener Rückseitenmetallisierung und Abbildung 6b die simulierte Stromverteilung. Es ist erkennbar, dass nahezu der gesamte ins Substrat eingekoppelte Strom über den Rückseitenkontakt abgeleitet wird. Lediglich in unmittelbarer Nähe der Wanne kann ein lateraler Stromfluss ausgemacht werden, der mit einer zusätzlichen Ableitstruktur (beispielsweise einem p-Typ Guard-Ring) nahe der Isolationswanne verringert werden könnte. Mittels Rückseitenmetallisierung wird der in die Störsenke einkoppelnde Strom um nahezu 100 % reduziert.

Das Simulationsergebnis in Abbildung 6c zeigt, dass aufgrund der niederohmigen Anbindung des Substrats

auch ein Spannungshub des Substratpotentials während des Umladens der High-Side Isolierung nahezu vollständig verhindert werden kann.

III. ZUSAMMENFASSUNG

Schnell schaltende High-Side NMOS-Leistungstransistoren verursachen beim Umschalten ein signifikantes Koppeln ins Substrat. Für einen Abwärtsschaltwandler für hohe Umsetzverhältnisse von > 40 V zu < 5 V mit einer Schaltfrequenz von > 10 MHz wurden während des Umladens der High-Side-Isolationswanne am Schaltknoten Störströme bis zu 130 mA und eine Potentialanhebung des Substrats von bis zu 20 V durch Device Simulation ermittelt. Mittels gezielter Beeinflussung der Substratstruktur durch eingebrachte Ableitstrukturen können die Störungen effektiv reduziert werden. Eine kostengünstige Ableitstruktur kann über p-Typ Guard-Ringe realisiert werden, wodurch der in empfindliche Schaltungselemente einkoppelnde Strom um 64 % und der Spannungshub des Substratpotentials um 75 % reduziert werden konnte. Durch die Verwendung von leitenden Trenches wurde eine Reduktion des einkoppelnden Stromes in Höhe von 92 % und eine Reduktion des Spannungsabfalls am Substratwiderstand in Höhe von 88 % erzielt bei halbiertem Flächenbedarf gegenüber p-Typ Guard-Ringen. Als effektivste Lösung stellte sich die Rückseitenmetallisierung heraus, welche sowohl den eingekoppelten Strom, als auch den Spannungshub des Substratpotentials um nahezu 100 % reduziert. Die Simulationen ergaben, dass n-Typ Guard-Ringe oder nichtleitende Trenches, deren Tiefe nicht über das gesamte Substrat reicht, nicht als Isolationsstruktur gegen einen lateralen Stromfluss im Substrat verwendet werden können.

LITERATURVERZEICHNIS

- [1] A. Afzali-Kusha, M. Nagata, N. Verghese, D. Allstot, „Substrate Noise Coupling in SoC Design: Modeling, Avoidance, and Validation“, *Proceedings of the IEEE*, Band 94, Nummer 12, Dez. 2006.
- [2] P. Schroter, S.Jahn, F. Klotz, „Improving the Immunity of Automotive ICs by Controlling RF Substrate Coupling“, 8. *Workshop der Electromagnetic Compatibility of Integrated Circuits* (EMC Compo), Nov. 2011.
- [3] M. Pfost, P. Brenner, T. Huttner, A. Romanyuk, „A Comprehensive Experimental Study on Technology Options for Reduced Substrate Coupling in RF and High-Speed Bipolar Circuits“, *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting*, Sept. 2003.



Christoph Rindfleisch erhielt den akademischen Grad des Bachelor of Engineering in Elektrotechnik mit dem Profil Automatisierungstechnik im Jahr 2011 von der Dualen Hochschule Baden Württemberg in Mosbach. Das duale Studium fand in Kooperation mit der Bosch Rexroth AG in Lohr am Main statt. Seit 2011 studiert er im Masterstudiengang Leistungs- und Mikroelektronik an der Hochschule Reutlingen.



Jürgen Wittmann erhielt den akademischen Grad des Dipl.-Ing. im Jahr 2006 von der Technischen Universität München. Zwischen 2006 und 2011 arbeitete er in der Mixed-Signal Automotive Abteilung von Texas Instruments in Freising als Analog Designer. Seit Februar 2011 ist er als akademischer Mitarbeiter am Robert Bosch Zentrum für Leistungselektronik der Hochschule in Reutlingen tätig. Er arbeitet zurzeit an seiner Doktorarbeit im Bereich Leistungs- und Mikroelektronik.



Bernhard Wicht erhielt den akademischen Grad Dipl.-Ing. im Jahr 1996 von der Technischen Universität Dresden und den Grad Dr.-Ing. im Jahr 2002 von der Technischen Universität München. 2003-2010 war er im Geschäftsbereich Mixed Signal Automotive bei Texas Instruments, Freising, verantwortlich für die Entwicklung von Smart Power ICs. Seit September 2010 ist er Professor für integrierte Schaltungen am Robert Bosch Zentrum für Leistungselektronik an der Hochschule Reutlingen.

Optimierte, wiederverwendbare OTA-Schaltungen für moderne Power BiCMOS-Technologien

Andreas Gerlach, Moritz Junge, Jürgen Scheible, Thoralf Rosahl

Zusammenfassung—Es wird das Ziel verfolgt, eine Möglichkeit für die sichere Wiederverwendbarkeit von Schaltungen aus der OTA-Schaltungsklasse bereitzustellen. Hierfür werden ausgewählte OTA-Schaltungstopologien für die „Copy-and-Paste“-Methode vorgestellt. Es wurde im industriellen Umfeld gezeigt, dass sie sich unter der Voraussetzung einer repräsentativen Topologieauswahl – vordimensioniert für den typischen Anwendungsbereich – schon in dieser Form für die Wiederverwendung eignen.

Schlüsselwörter—Chipentwicklung, Analog-Reuse, OTA, Copy-and-Paste, Reuse-Bibliothek, Schaltungsklasse, Analog-IP, Topologieauswahl, Dimensionierung.

I. EINLEITUNG

In der Entwicklung analoger integrierter Schaltungen werden im Vergleich zur Entwicklung digitaler Chips nur wenige Schaltungen wiederverwendet. Eine erfolgreiche, breite Wiederverwendung von Analog-IP würde zu einer Kostenersparnis in der Entwicklung führen und wird daher generell in der Industrie angestrebt [1]. Bei einer Wiederverwendung über den gesamten Design-Flow würde sowohl der Aufwand bei der Schaltungsentwicklung, wie auch bei der Layoutentwicklung sinken.

Grundsätzlich lässt sich zwischen der einfachen „Copy-and-Paste“-Methode (fest dimensionierte Schaltungen) und weiterentwickelten Konzepten zur Wiederverwendung analoger Schaltungen mit gezielt zugelassenen Freiheitsgraden unterscheiden. Im Weiteren werden Erfahrungen mit der „Copy-and-Paste“-Methode dargestellt. Die „Copy-and-Paste“-Wiederverwendungsmethode wurde am Beispiel ausgewählter Schaltungstopologien der OTA-Schaltungsklasse praktisch umgesetzt. Die Vorkommenshäufigkeit und der einfache Aufbau der Schaltungen innerhalb dieser Schaltungsklasse waren die Entscheidungskriterien bei der Auswahl. Eine Schaltungsklasse enthält Schaltungen, die eine bestimmte Schaltungsfunktion erfüllen.

Andreas Gerlach, andreas.gerlach@reutlingen-university.de, Robert Bosch Zentrum für Leistungselektronik, Oferdinger Str. 50, 72768 RT-Rommelsbach; Moritz Junge, mojunge@uni-bremen.de, Universität Bremen; Jürgen Scheible, juergen.scheible@reutlingen-university.de, Robert Bosch Zentrum für Leistungselektronik; Thoralf Rosahl, thoralf.rosahl@de.bosch.com, Robert Bosch GmbH.

Bei OTAs ist das die Umwandlung einer Spannungsdifferenz am Eingang in einen dazu proportionalen Strom am Ausgang (näheres hierzu siehe Kap. 3). Eine Schaltung kann mehreren Schaltungsklassen angehören.

II. EINE REUSE-BIBLIOTHEK

Eine Reuse-Bibliothek ist eine Bibliothek, die in der EDA-Entwurfsumgebung eingebunden ist und Schaltungen für einen breiten Wiederverwendungseinsatz beinhaltet. Man kann nicht eine Reuse-Bibliothek, die alle möglichen Schaltungstopologien bereitstellt, einrichten. Dazu ist die Anzahl der möglichen Schaltungstopologien zu groß. Wenn man Schaltungen für Reuse bereitstellt, muss die Topologieauswahl auf geeignete Beispiele mit zu erwartendem hohem Wiederverwendungsgrad beschränkt werden.

Bei der Auswahl sollte darauf geachtet werden, möglichst einfache Topologien auszuwählen, die den typischen Aufgaben genügen. Eine einfache Topologie, die eine Aufgabe genauso gut lösen kann wie eine komplexere, sollte immer bevorzugt werden, da sie vielerlei Vorteile bietet: geringerer Entwurfsaufwand, geringere Fehleranfälligkeit, geringere Chipfläche und daraus resultierende geringere Kosten, um nur die wichtigsten zu benennen. Diese Regel gilt allgemein für die Entwicklung analoger Schaltungen und auch speziell für die Entwicklung von Schaltungen für den Reuse-Einsatz. Bei den Reuse-Schaltungen ist mit dem Begriff Entwurfsaufwand speziell der Aufwand gemeint, der benötigt wird, um eine Schaltung so auszulegen, dass sie als Element einer Reuse-Bibliothek genutzt werden kann.

Um erfolgreich die „Copy-and-Paste“-Methode umzusetzen, wurden die für eine Reuse-Bibliothek einzuhaltenden Eigenschaften definiert. Diese Eigenschaften werden nachfolgend aufgelistet und jeweils kurz erläutert:

- Jede für Reuse geeignete Schaltungstopologie wird in Form eines Schaltplans samt dazugehörigem Layout und Symbol in der Entwurfsumgebung bereitgestellt.
- Zu jeder Schaltungstopologie wird ein sie beschreibendes Datenblatt erstellt, denn eine genaue Spezifikation der Schaltungen ist Voraussetzung für ihre Wiederverwendung.

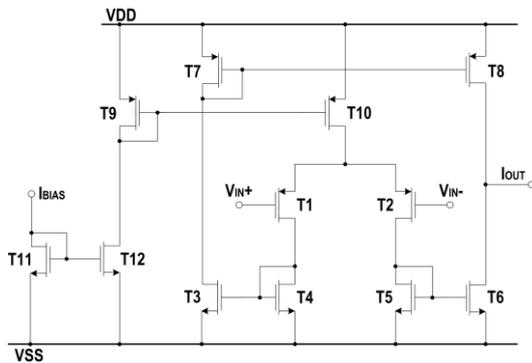


Abbildung 1: Schaltplan des symmetrischen „P-Eingang“-OTA.

- Damit die Schaltungen einfach in die Design-Umgebung eingebunden werden können, wird die Biasstromrichtung für alle Topologien eindeutig definiert. So werden die hier vorgestellten Schaltungen mit einem „Pull-up“-Bias-Strom gespeist.
- Alle Schaltungen sind fest dimensioniert. Dabei erfüllt die Dimensionierung typische Anforderungen aus der Praxis.

Die Reuse-Bibliothek kann bei Bedarf um weitere, sinnvolle Topologien erweitert werden.

III. DIE OTA-SCHLUNGSKLASSE

OTA-Schaltungen (Operational Transconductance Amplifier) sind Operationsverstärker, die eine Differenzspannung an den beiden Eingängen in einen dazu proportionalen Ausgangsstrom umwandeln [2]. Ihre Hauptanwendungsbereiche sind Regelung, Kleinsignalverstärkung und Schwellwernererkennung.

OTAs zählen zu den am häufigsten eingesetzten Schaltungen im analogen und im Mixed-Signal-Chipentwurf [3]. Zudem sind sie in ihrer Komplexität noch recht überschaubar (Umfang hier betrachteter OTA-Beispiele: 10 bis 25 Transistoren). Aus diesen Gründen wurden sie als erstes Anwendungsbeispiel für die Entwicklung von Reuse-Methoden ausgewählt.

IV. AUSGEWÄHLTE OTA-SCHLUNGSTOPOLOGIEEN FÜR DIE REUSE-BIBLIOTHEK

In diesem Kapitel werden drei ausgewählte OTA-Schaltungstopologien für Wiederverwendungszwecke präsentiert. Mit diesen OTA-Varianten hat man die Wahl zwischen OTA-Schaltungen für die Auswertung der am häufigsten in der Praxis vorkommenden Arten von Signalen. Die Reihenfolge der Vorstellung der einzelnen Schaltungen entspricht der zu erwartenden Häufigkeit der Einsetzbarkeit.

Alle hier vorgestellten Schaltungen haben einen Single-Ended-Ausgang, denn OTAs mit Single-Ended-Ausgängen werden im Vergleich zu denen mit symmetrischen Ausgängen häufiger eingesetzt und

Tabelle 1: Symmetrischer-„P-Eingang“-OTA: Dimensionierung.

T1, T2:	$W = 30 \mu\text{m}, L = 5 \mu\text{m}, M = 2$
T4, T5:	$W = 8 \mu\text{m}, L = 22 \mu\text{m}$
T3, T6:	$W = 8 \mu\text{m}, L = 22 \mu\text{m}, M = 2$
T7, T8:	$W = 10 \mu\text{m}, L = 15 \mu\text{m}$
T9, T10:	$W = 2 \mu\text{m}, L = 2 \mu\text{m}$
T11, T12:	$W = 2 \mu\text{m}, L = 2 \mu\text{m}$

Tabelle 2: Symmetrischer-„P-Eingang“-OTA: Typische Werte der eingestellten Ausgangsparameter. Sie wurden simulatorisch bei Raumtemperatur und unter Einbezug der Bauelementtoleranzen bei einer Versorgungsspannung von 5 V ermittelt.

Steilheit	102 μS
Offsetspannung (3σ -Wert)	2,24 mV
Abweichung der Steilheit	5,15 %
Eingangsspannungsbereich	-0,3 V – 2,8 V

weisen einen einfacheren Aufbau auf. Beim „P-Eingang“- und „N-Eingang“-OTA handelt es sich um einstufige Verstärker mit einstellbarem Übersetzungsverhältnis. Der hier präsentierte „Rail-to-Rail“-OTA ist ebenfalls ein einstufiger Verstärker, jedoch ohne Übersetzungsverhältnis.

Die Schaltungen wurden dimensioniert, um typische Anforderungen aus der Praxis zu erfüllen. Das Augenmerk wurde hierbei auf die Steilheit, den Steilheitsverlauf, die Offsetspannung und das Rauschen gelegt. Zukünftig wären zusätzliche Dimensionierungsvarianten denkbar, z.B. eine auf Geschwindigkeit optimierte.

A. Symmetrischer „P-Eingang“-OTA

Bei diesen Schaltungen lassen sich, bedingt durch P-Kanal- oder (optional) pnp-Eingangstransistoren, Eingangsspannungen im unteren Versorgungsspannungsbereich analysieren, d.h. vom unteren Rail¹ (auch knapp darunter) bis etwas über die halbe Versorgungsspannung². Sie erreichen das obere Rail jedoch nicht.

¹ Mit unterem und oberem Rail bezeichnet man die unteren und oberen Level der Versorgungsspannung.

² Diese qualitative Aussage gilt für die im Kfz-Bereich typischen Versorgungsspannungen von 3 V – 5 V und ist deutlich vereinfacht. Quantitativ wird der Wert der maximalen Eingangsspannung durch den Ausdruck

$$U_{in,max} = V_{DD} - |U_{DS10,min}| - |U_{GS1,min}|$$

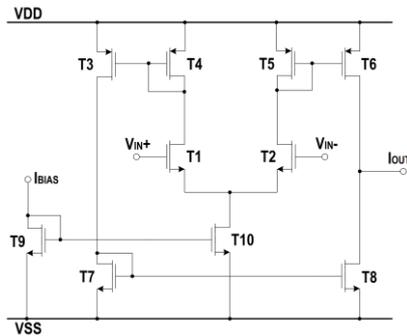


Abbildung 2: Schaltplan des symmetrischen „N-Eingang“-OTA.

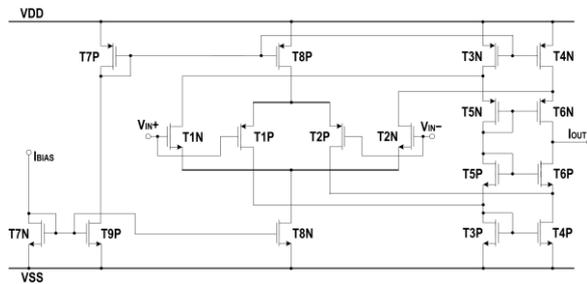


Abbildung 3: Schaltplan des „Rail-to-Rail-Eingang“-OTA.

Da die meisten in der Praxis auszuwertenden Signale im unteren Versorgungsspannungsbereich liegen, wird der „P-Eingang“-OTA innerhalb der OTA-Schaltungs-kategorie am häufigsten eingesetzt. Die Topologie der symmetrischen „P-Eingang“-OTA-Schaltung, die hier für die Wiederverwendung ausgewählt worden ist, wird in Abbildung 1 gezeigt. Die gewählten Parameter, sowie die damit erzielten Ergebnisse sind in Tabelle 1 und 2 aufgelistet.

B. Symmetrischer „N-Eingang“-OTA

Das Eingangspaar des Differenzverstärkers dieser Schaltung besteht im Gegensatz zum symmetrischen „P-Eingang“-OTA aus N-Kanal- oder (optional) npn-Transistoren. Hierdurch ist die Schaltung für die Auswertung von Signalen im oberen Eingangsspannungsbereich ausgelegt. Der Arbeitsbereich reicht von knapp unterhalb der halben Betriebsspannung bis über das obere Rail. Schaltungstopologisch gesehen ähnelt der Aufbau sehr dem des „P-Eingang“-OTAs (vergl. Abbildungen 1 und 2); die Schaltpläne sind fast komplementär identisch.

beschrieben; wobei V_{DD} die Versorgungsspannung, $|U_{DS10,min}|$ der Betrag der minimalen Drain-Source-Spannung vom Transistor T10 und $|U_{GS1,min}|$ der Betrag der minimalen Gate-Source-Spannung vom Transistor T1 ist.

Tabelle 3: Symmetrischer-„N-Eingang“-OTA: Dimensionierung

T1, T2:	$W = 30 \mu\text{m}, L = 15 \mu\text{m}$
T4, T5:	$W = 15 \mu\text{m}, L = 15 \mu\text{m}$
T3, T6:	$W = 15 \mu\text{m}, L = 10 \mu\text{m}, M = 2$
T7, T8:	$W = 15 \mu\text{m}, L = 10 \mu\text{m}, M = 2$
T9, T10:	$W = 2 \mu\text{m}, L = 2 \mu\text{m}, M = 2$

Tabelle 4: Symmetrischer-„N-Eingang“-OTA: Typische Werte der eingestellten Ausgangsparameter. Sie wurden simulatorisch bei Raumtemperatur und unter Einbezug der Bauelementtoleranzen bei einer Versorgungsspannung von 5 V ermittelt.

Steilheit	83 μS
Offsetspannung (3σ -Wert)	2,49 mV
Abweichung der Steilheit	6,41 %
Eingangsspannungsbereich	1,2 V – 5,3 V

Um die das Biasing betreffende Regel einzuhalten (Kapitel II), werden für die „N-Eingang“-OTA-Schaltung im Vergleich zum „P-Eingang“-OTA zwei Transistoren weniger benötigt. Dies ist darauf zurückzuführen, dass bei dem hier festgelegten, einzuspeisenden „Pull-up“-Bias-Referenzstrom in dieser Topologie lediglich ein Stromspiegel für die Spiegelung des Stroms zum Differenzverstärker benötigt wird. Beim „P-Eingang“-OTA war an dieser Stelle eine doppelte Stromspiegelung vorhanden. Die gewählten Parameter, sowie die damit erzielten Ergebnisse sind in Tabelle 3 und 4 aufgezeigt.

C. „Rail-to-Rail-Eingang“-OTA

In der hier gewählten „Rail-to-Rail-Eingang“-OTA-Topologie (Abbildung 3), einem typischen Repräsentanten des „Rail-to-Rail-Eingang“-OTA-Typs, sind ein „P-Eingang“- und ein „N-Eingang“-OTA so verschaltet, dass die Auswertung von Signalen innerhalb des gesamten Versorgungsspannungsbereichs möglich ist. Die Steilheitsfunktion des „Rail-to-Rail-Eingang“-OTAs ergibt sich aus einer Superposition der Steilheitsverläufe der beiden Einzelschaltungen (Abb. 4).

Durch die Kombination der Einzelschaltungen werden die Teilströme dieser Schaltungen am Ausgang überlagert, um den Ausgangsstrom I_{out} zu erzeugen. Die Transistorzahl wird gering gehalten, indem zwei gefaltete Kaskoden eingesetzt werden. Die Vorteile dieser Lösung sind nebst der geringen Transistorzahl ein vergrößerter Ausgangswiderstand der Schaltung, eine Erhöhung der auswertbaren Signalfrequenz (erzielt durch Minimierung des Miller-Effekts), sowie im Vergleich zu einer Topologie mit normalen Kaskoden ein größerer Eingangsspannungsbereich.

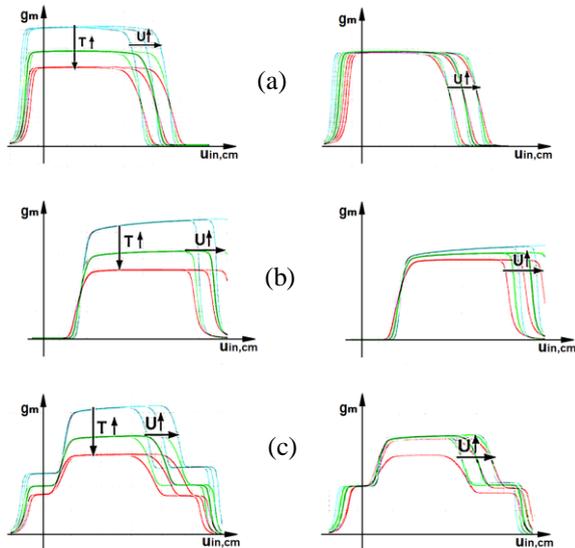


Abbildung 4: Temperatureinflüsse beim Biasing mit einer Konstantstromquelle (linke Spalte). Korrektur von Temperatureinflüssen durch Biasing mit einer PTAT-Referenzstromquelle (rechte Spalte). Steilheitsverlauf dargestellt gegenüber der Common-Mode-Eingangsspannung für unterschiedliche Temperaturen des „P-Eingang“ (a), „N-Eingang“ (b) und des „Rail-to-Rail-Eingang“ (c). Die kleinen Pfeile neben den Buchstaben U und T bedeuten jeweils den Anstieg der Versorgungsspannung und der Temperatur. Die etwas größeren Pfeile zeigen die Richtungen der dadurch verursachten Änderungen der Steilheitsverläufe.

Die hier vorgestellte „Rail-to-Rail“-OTA-Schaltung hat eine Mindeststeilheit von $50 \mu\text{S}$. Das Maximum des Steilheitsverlaufs, aufgetragen über dem Eingangsspannungsverlauf, beträgt jedoch $102 \mu\text{S}$, wobei sich ein inhomogener Steilheitsverlauf ergibt, mit einer charakteristischen Stufe in der Mitte (Abb. 4c). Es gibt schaltungstechnische Maßnahmen, die zu einem homogenen Steilheitsverlauf führen [4]. Es hat sich jedoch in der Praxis bestätigt, dass viele Aufgaben auch ohne diese zusätzlichen Aufwendungen gelöst werden können. Nach der Devise, immer die einfachste Topologie zur Lösung eines schaltungstechnischen Problems einzusetzen (vgl. Kapitel II), fiel die Entscheidung auf die hier vorgestellte Variante. Die gewählten Parameter, sowie die damit erzielten Schaltungsparameter sind in den Tabellen 5 und 6 aufgelistet.

Es wäre denkbar, in Zukunft weitere „Rail-to-Rail“-Topologien der OTA-Schaltungsbibliothek hinzuzufügen, auch solche, die keinen „Treppenverlauf“ der Steilheit aufweisen, sondern einen homogenen Verlauf. Dies würde zu einer möglichen breiteren Einsetzbarkeit führen.

V. REDUKTION DER TEMPERATURABHÄNGIGKEITEN

Bei der Entwicklung von Elektronik für den Kfz-Einsatz wird das Verhalten von Schaltungen bei unterschiedlichen Temperaturen gründlich analysiert. Im Auto können große Temperaturschwankungen auftreten.

Tabelle 5: „Rail-to-Rail-Eingang“-OTA: Dimensionierung

T1, T2:	$W = 30 \mu\text{m}, L = 5 \mu\text{m}, M = 2$
T4, T5:	$W = 8 \mu\text{m}, L = 22 \mu\text{m}$
T3, T6:	$W = 8 \mu\text{m}, L = 22 \mu\text{m}, M = 2$
T7, T8:	$W = 10 \mu\text{m}, L = 15 \mu\text{m}$
T9, T10:	$W = 2 \mu\text{m}, L = 2 \mu\text{m}$
T11, T12:	$W = 2 \mu\text{m}, L = 2 \mu\text{m}$

Tabelle 6: „Rail-to-Rail-Eingang“-OTA: Typische Werte der eingestellten Ausgangsparameter. Sie wurden simulatorisch bei Raumtemperatur und unter Einbezug der Bauelementtoleranzen bei einer Versorgungsspannung von 5 V ermittelt.

Steilheit	$102 \mu\text{S}$
Offsetspannung (3σ -Wert)	$2,36 \text{ mV}$
Abweichung der Steilheit	50%
Eingangsspannungsbereich	$-0,3 \text{ V} - 5,3 \text{ V}$

ten. Sie werden durch Wittereinflüsse bewirkt, sowie durch überwiegend chemische und mechanische Wärmequellen im Auto (insbesondere Verbrennungsmotor, Abgasleitungsbereich und Bremsanlage). Analoge Schaltungen, die in unmittelbarer Nähe dieser Wärmequellen verbaut werden, müssen im gesamten möglichen Temperaturbereich zuverlässig arbeiten. Um dies für die hier für die Wiederverwendung vorgestellten OTA-Schaltungen sicherzustellen, wurde das Biasing mit einer PTAT-Stromquelle als Maßnahme zur Unterdrückung von Temperatureinflüssen überprüft.

Bei der heutigen Entwicklung analoger und Mixed-Signal-Schaltungen werden – neben der die eigentlichen, spezifizierten Nutzfunktionen erfüllenden Schaltungsblöcke – Referenzstromquellen auf dem Chip mitintegriert. Man kann davon ausgehen, dass eine Konstantreferenzstromquelle für das Biasing immer vorhanden ist. In vielen Fällen steht zusätzlich eine PTAT-Referenzstromquelle (Proportional To Absolute Temperature) bereit. Die hier gezeigten Schaltungen können sowohl durch eine Konstantstromquelle, wie auch durch eine PTAT-Stromquelle gespeist werden. Durch das Verwenden einer PTAT-Stromquelle lässt sich jedoch der Temperatureinfluss auf die Schaltung deutlich verringern. Dies wird am Beispiel des Steilheitsverlaufs grafisch gezeigt (Abb. 4, rechte Spalte).

Die Verwendung einer PTAT-Stromquelle für das Biasing der hier vorgestellten Schaltungen hat sich als Maßnahme zur Temperatureinflussunterdrückung als wirkungsvoll erwiesen und lässt sich in der Praxis gut umsetzen.

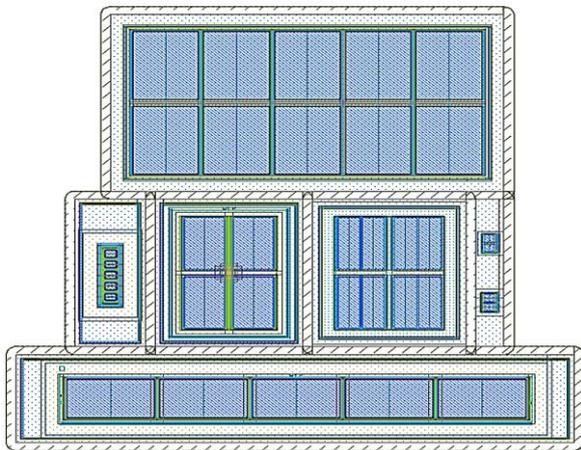


Abbildung 5: Layout der „Rail-to-Rail-Eingang“-OTA-Schaltung.

VI. GESAMTFLOWBETRACHTUNG

Alle OTA-Schaltungen für den „Copy-and-Paste“-Einsatz wurden im Rahmen einer Gesamtflowbetrachtung entwickelt. Durch die Berücksichtigung von Randbedingungen und Möglichkeiten des an die Schaltungsentwicklung anschließenden physikalischen Entwurfs können die entstehenden Schaltungen so ausgelegt werden, dass in Zukunft nicht nur fest vorgegebene Layouts in der Reuse-Bibliothek ihren Platz finden, sondern auch eine Automatisierung des Layoutentwurfs bei sich ändernden Transistorparametern möglich wird. Hierzu wurden Constraint-Vorgaben zum Floorplanning und zur Umsetzung der notwendigen Symmetrie – insbesondere der Aufteilung und Verschachtelung der Bauelemente – erarbeitet. Abbildung 5 zeigt ein aktuelles Layout-Beispiel für fest eingestellte Transistorparameter.

VII. RESÜMEE

Am Beispiel der OTA-Schaltungsklasse konnte gezeigt werden, dass die „Copy-and-Paste“-Wiederverwendungsmethode in der Praxis gut umsetzbar und damit prinzipiell wirkungsvoll ist, wenn sie auf einer Reuse-Bibliothek mit den in Kap. II beschriebenen Eigenschaften aufsetzt. Da die entsprechende methodische Unterstützung in der Praxis noch weitgehend fehlt, wird die Methode in der hier vorgestellten Form bislang leider selten angewendet.

Damit Schaltungen wiederverwendet werden können, bedarf es einer umfangreichen Vorarbeit. Es muss untersucht werden, welche Schaltungen oft genutzt werden und sich damit für eine Reuse-Bibliothek überhaupt eignen. Für den breiten Einsatz dieser Schaltungen sind Standardwerte der relevanten Größen zu spezifizieren. Die Schaltungen müssen entsprechend dimensioniert und beschrieben den Entwicklern in einer Bibliothek zur Verfügung gestellt werden. Nur unter diesen Bedingungen ist damit zu rechnen, dass An-

wender diese Schaltungen in Projekten auch nutzen werden. Diese Arbeiten sind mit einem hohen Zeitaufwand verbunden. Zusätzlicher Aufwand entsteht mit der Pflege der Bibliothek und Migrationen in andere Technologien, wenn man sich nicht nur auf eine bestimmte Technologie beschränken möchte. Eine Hilfe bei der Dimensionierung und Technologiemigration dieser Schaltungen bringt eine universelle und weitgehend technologieunabhängige Schaltungsklassentestbench [3].

Es wäre auch erstrebenswert, für Reuse geeignetes Analog-IP, welches während laufender Projekte entsteht, in der Reuse-Bibliothek festzuhalten. In der Praxis bleibt aber oft nicht genügend Zeit, um während laufender Projekte einzelne Teilschaltungen zu dokumentieren und in eine Bibliothek zu integrieren. Würde das gelingen, dann wäre mit erfolgreichen Reuse mit der „Copy-and-Paste“-Methode zu rechnen. Eine Sammlung von Analog-IP dieser Art wäre ein wertvolles Kapital für die Firma.

Es ist damit zu rechnen, dass die hier präsentierten, für den typischen Anwendungsfall optimierten und in Form von Datenblättern beschriebenen Schaltungen bald in Projekten eingesetzt werden. Die ersten Anwendungen sind bereits geplant.

VIII. AUSBLICK

Es könnte zukünftig eine Erweiterung der „Copy-and-Paste“-Bibliothek erfolgen, indem mehrere Dimensionierungsvarianten für die jeweiligen angebotenen Schaltungstopologien bereitgestellt werden, um ein breiteres Anwendungsspektrum abzudecken. Ein weiterer Schritt zu höherem Wiederverwendungsgrad, der schon über den Rahmen einer „Copy-and-Paste“-Methode hinausgeht, wäre definierte Freiheitsgrade in den Schaltungstopologien zuzulassen. Es ist möglich, generalisierte Aussagen zum Dimensionierungsvorgehen innerhalb einer Schaltungsklasse zu treffen. Diese werden als Grundlage zur Ableitung von Dimensionierungsalgorithmen für diese Schaltungsklasse dienen.

Ein noch höherer Wiederverwendungsgrad lässt sich erreichen, indem die Reuse-Bibliothek auf weitere häufig verwendete Topologievarianten ergänzt wird. Bei einer Vielzahl unterschiedlicher Varianten wird ein Topologieauswahlalgorithmus die Grundlage zur intelligenten Topologieauswahl bilden. Die Kombination der automatisierten Parametereinstellungen und des Topologieauswahlmechanismus [5], basierend auf einer repräsentativen Reuse-Bibliothek, bildet die Zukunft der Wiederverwendung analoger Schaltungen und ist Ziel weiterer Arbeiten.

LITERATURVERZEICHNIS

- [1] M. Barros, J. Guilherme, N. Horta: *Analog Circuits and Systems Optimization Based on Evolutionary Computation Techniques*, Springer-Verlag Berlin Heidelberg, 2010, ISBN 978-3-642-12345-0.

- [2] U. Tietze, Ch. Schenk: *Halbleiter-Schaltungstechnik*, 12. Aufl., Springer-Verlag Berlin Heidelberg New York, 2002, ISBN 3-540-42849-6.
- [3] A. Gerlach, M. Junge, J. Scheible: *Universelle OTA-Testbench*, ASIM-Workshop STS/GMMS 2014, 20.-21.02.2014, ARGESIM / ASIM Wien – Hochschule Reutlingen, ISBN 978-3-901608-42-1, S. 83 – 87.
- [4] W. Sansen: *Analog Design Essentials*, Springer-Verlag, Netherlands, 2006, ISBN-10 0-387-25746-2.
- [5] W. Kruiskamp, D. Leenaerts: DARWIN: “CMOS opamp synthesis by means of a genetic algorithm”. In: *Proc. ACM/IEEE Design Automation Conference*, ISBN 0-89791-725-1, pp. 550–553 (1995).



Andreas Gerlach erhielt den akademischen Grad des Dipl.-Ing. in Elektro- und Informationstechnik im Jahr 2010 von der Ruhr-Universität Bochum und ist Doktorand am Robert Bosch Zentrum für Leistungselektronik, wo er an Methoden zur verbesserten Wiederverwendung analoger Schaltungen arbeitet.



Moritz Junge studierte Elektrotechnik an der Universität Bremen und war zum Zeitpunkt des Entstehens dieses Artikels Diplomand unter der Betreuung von Andreas Gerlach. Das Thema seiner Diplomarbeit lautet: „Analyse vorhandener Transkonduktanzverstärker und Optimierung hinsichtlich Wiederverwendbarkeit von Schaltung und Layout“.



Jürgen Scheible studierte Elektrotechnik an der TH Karlsruhe, wo er im Jahre 1991 auf dem Gebiet der Electronic Design Automation zum Dr.-Ing. promovierte. Anschließend war er bei der Robert Bosch GmbH im Geschäftsbereich Automotive Electronics in verschiedenen Positionen tätig, zuletzt als Leiter der Entwicklungsabteilung ASIC Layout Design. Seit 2010 ist er Professor für EDA am Robert Bosch Zentrum für Leistungselektronik der Hochschule Reutlingen. Sein Forschungsinteresse gilt der Automatisierung des Entwurfs analoger integrierter Schaltkreise.



Thoralf Rosahl schloss sein Studium der Elektrotechnik an der TU Dresden im Jahre 1990 als Dipl.-Ing. ab. Seit 1991 ist er bei der Robert Bosch GmbH im Geschäftsbereich Automotive Electronics im Bereich „Analoges Schaltungsdesign“ tätig. Der Schwerpunkt seiner Arbeiten liegt auf dem Gebiet von Power-Anwendungen, speziell lineare und getaktete Spannungsversorgungen. Als „Chief Expert Analog Design“ betreut er auch Themen zur Designmethodik.

Eine effiziente Dekodierarchitektur für verallgemeinert verkettete Codes

Jens Spinner, Jürgen Freudenberger

Zusammenfassung—Zukünftige Flashspeichertechnologien benötigen eine leistungsfähige Fehlerkorrektur, um eine zuverlässige Datenspeicherung zu gewährleisten. Die Fehlerkorrektur muss dabei einen hohen Datendurchsatz erzielen, der nur durch eine ASIC-Implementierung der Fehlerkorrektureinheit erreicht werden kann. Eine Fehlerkorrektur auf Basis der Generalized Concatenated (GC) Codes ermöglicht eine geringere Komplexität und damit geringere Chip-Fläche als mit den heute üblichen BCH-Codes. In diesem Beitrag wird eine Architektur zur Decodierung von GC-Codes vorgestellt und eine Optimierung des Flächenverbrauchs beschrieben.

Schlüsselwörter—Fehlerkorrektur, Error Checking & Correction (ECC), SRAM, Flash, GCC.

I. EINLEITUNG

Flashspeicher haben in eine Vielzahl von eingebetteten Systemen Einzug genommen. Je nach Anwendung benötigen diese Systeme eine Garantie für die Zuverlässigkeit der Persistenz. Bei der Datenspeicherung in Flashspeichern treten jedoch, wie in anderen Speichermedien auch, mit einer gewissen Wahrscheinlichkeit Fehler auf. Die Fehlerwahrscheinlichkeit steigt aufgrund der fortschreitenden Reduktion der Strukturgrößen mit jeder neuen Generation an. Zur Reduktion der Fehler wurden bislang Kanalcodes wie z.B.: Bose-Chaudhuri-Hoquenghem (BCH) und Reed-Solomon (RS) Codes eingesetzt [1]. Zukünftig werden aufgrund der hohen Fehlerwahrscheinlichkeiten auch Low-Density-Parity-Check (LDPC) Codes und Code-Verkettungen eingesetzt.

Ein Speichersystem besteht meist aus dem eigentlichen Flashspeicher und einem Controller (siehe Abbildung 1) [2]. Der Controller führt unter anderem die Fehlerkorrektur durch und leitet die Daten über ein Bussystem an eine Verarbeitungseinheit weiter. Für industrielle Anwendungen kann ein Nachweis der Fehlerkorrektureigenschaften des Kanalcodes erforderlich sein. Diese Nachweisbarkeit besitzen Kanal-

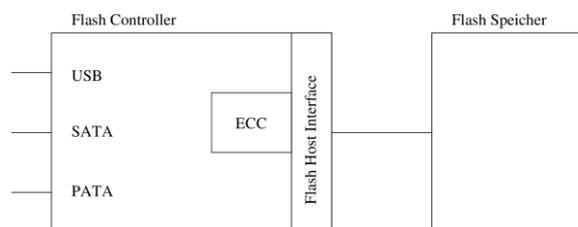


Abbildung 1: Flashspeicher mit Controller für die ECC.

codes, die mithilfe der sogenannten Bounded-Minimum-Distance (BMD) Decodierung decodiert werden. Diese Eigenschaften haben die algebraischen Codes, zu denen BCH- und RS-Codes gehören. Die Decodier-Komplexität dieser beiden Codes steigt allerdings mindestens quadratisch mit der Fehlerkorrekturfähigkeit, was sich in der Durchsatzrate und der Chipfläche bemerkbar macht. Um diese Performance-Einbußen zu reduzieren, wurde die Verwendung von anderen Codes untersucht.

In diesem Beitrag wird ein Decoder für die sogenannten Generalized Concatenated (GC) Codes vorgestellt. Ein verketteter Code besteht aus mehreren inneren BCH-Codes und mehreren äußeren RS-Codes. Durch die Konstruktion des Codes auf Basis mehrerer kürzerer Codes wird eine Verringerung der Komplexität erreicht [3].

Ein erster Pipeline-Decoder für GC-Codes wurde in [4] vorgestellt. In diesem Beitrag wird der Flächenbedarf der Decoder-Einheit reduziert, ohne die Leistungsfähigkeit des Verfahrens zu reduzieren. Bei der Decodierung muss das gesamte Codewort in einem Puffer gespeichert werden. Der bisherige Pipeline-Ansatz erforderte eine Speicherung der Codebits in Schieberegistern. Der Flächenverbrauch für die Speicherstruktur entspricht dabei etwa dem Bedarf für die komplette Decodier-Logik. Ziel dieser Arbeit ist es daher, die Schieberegisterstruktur durch SRAM-Zellen zu ersetzen und dadurch eine Flächeneinsparung zu erzielen.

In Abschnitt II wird zunächst auf die Codekonstruktion eingegangen. Anschließend wird in Abschnitt III die Architektur des Pipeline-Decoders vorgestellt. Im Anschluss werden verschiedene Optionen zur Flächenoptimierung diskutiert und eine Lösung mit mehreren SRAM-Bausteinen vorgeschlagen. Die Ergebnisse dieser Verbesserung werden in Kapitel V präsentiert.

Jens Spinner, jens.spinner@htwg-konstanz.de und Jürgen Freudenberger, juergen.freudenberger@htwg-konstanz.de, sind Mitglieder der HTWG-Konstanz, Institut für Systemdynamik, Brauneggerstrasse 55, 78462 Konstanz.

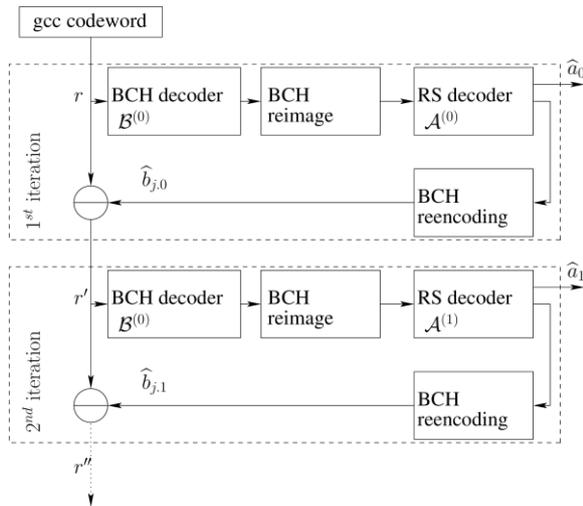


Abbildung 2: Mehrstufige Dekodierung.

II. CODE-KONSTRUKTION UND DECODIERUNG

Wie bereits einleitend erwähnt, besteht der GC-Code aus inneren BCH-Codes und äußeren RS-Codes. Das Codewort kann durch eine Matrix dargestellt werden, in der zunächst je m Zeilen einen RS-Code über das Galois-Feld-Alphabet $GF(2^m)$ bilden. Jede einzelne Spalte ist wiederum mit einem BCH-Code geschützt.

Die Komponentencodes werden mithilfe der Syndrom-Decodierung decodiert. Hierbei werden zunächst die Syndrome ermittelt. Anhand dieser Syndrome errechnet der Berlekamp-Massey-Algorithmus (BMA) das Fehlerstellenpolynom. Mit diesem Fehlerstellenpolynom werden im nächsten Schritt die Fehlerpositionen durch Suche der Nullstellen des Fehlerstellenpolynoms ermittelt. Dies gilt für den BCH-Code sowie für den RS-Code, wobei im Falle des RS-Codes zusätzlich die Fehlerwerte mittels des sogenannten Forney-Algorithmus berechnet werden müssen.

In Abbildung 2 ist der Zyklus dieser mehrstufigen Decodierung der GC-Codes dargestellt. Dieser Prozess beginnt mit der Decodierung des Codes B_1 der ersten Ebene des inneren BCH-Codes. Da es sich um einen verschachtelten BCH-Code handelt, muss das geschätzte Codewort durch erneute Codierung (*bch reimage*) ermittelt werden. Ist dies für alle Zeilen erfolgt, kann der Zeilencode A_1 decodiert werden. Im letzten Iterationsschritt wird das Ergebnis der Decodierung angewandt, d.h. die Symbole des Codes A_1 werden wieder auf Codeworte des BCH-Codes abgebildet (*bch reencoding*). Diese Codeworte werden vom GC-Codewort abgezogen. Dieser Zyklus wiederholt sich nun mit der nächsten Stufe, bis alle Stufen durchlaufen wurden.

Da es sich um eine BMD-Decodierung handelt, kann man exakte Aussagen über die Fehlerwahrscheinlich-

keiten in jeder Stufe der Decodierung treffen. Die Parameter der RS-Codes werden dann so gewählt, dass die Anforderungen an die Restfehlerwahrscheinlichkeit bei einer gegebenen Kanalfehlerwahrscheinlichkeit eingehalten werden können.

III. BESCHREIBUNG DER DECODER-PIPELINE

Die Architektur des Decoders ist in Abbildung 3 dargestellt und besteht aus zwei Funktionseinheiten. Die erste Funktionseinheit enthält die Codewort-Puffer für das GC-Codewort (*data matrix buffer*) und das RS-Codewort (*RS buffer*). Die zweite Einheit enthält die Decodier-Logik. Die kurze Codewortlänge der Spaltencodes und die geringe Fehlerkorrekturfähigkeit der inneren Codes ermöglicht eine Pipelining-Struktur mit einer geringen Latenz. Dieses Pipelining erlaubt es, je ein Spaltencodewort pro Zyklus zu decodieren und das daraus resultierende Symbol für die Syndrom-Berechnung des äußeren Codes zu verarbeiten. Die Decodier-Logik der Pipeline besteht aus folgenden Komponenten:

- BCH-Syndrom-Berechnung
- BCH-BMA
- BCH-Chien-Search
- RS-Syndrom-Berechnung

Haben alle Spalten die Pipeline durchlaufen, so sind alle Syndrome des äußeren Codes bereit, um durch den RS-BMA verarbeitet zu werden. Solange die Berechnung des Fehlerstellenpolynoms des äußeren Codes durch den RS-BMA aktiv ist, steht die Pipeline still. Ist das Fehlerstellenpolynom des äußeren Codes ermittelt, so wird die Pipeline in einem zweiten Teil fortgesetzt. Dieser besteht aus:

- RS-Chien-Search
- RS-Forney-Algorithmus
- Reencodierung und Subtraktion des Ergebnisses vom GC-Codewort

Für die nächste Iteration geht das Ende der Pipeline-Struktur nahtlos in den Anfang über. Synchron zur Decodier-Logik müssen die Spaltencodes gespeichert, ausgelesen und korrigiert werden. Die Zugriffe auf diese Spalten haben für jeden GC-Code einen konstanten Spaltenabstand. Diese Eigenschaft und die spaltenweise Decodierung machte die Implementierung des Codewortpuffers mit Hilfe eines Schieberegisters zu einer einfachen Lösung. Die Flächenanalysen der Module nach einer Synthese des Decoders zeigten jedoch, dass die Implementierung des Puffers für das GC-Codewort als Schieberegister die Hälfte der gesamten Fläche ausmachte.

IV. BESCHREIBUNG DER SRAM-LÖSUNG

Die Registerzellen benötigen ab einer bestimmten Speichergröße eine größere Fläche als Speicherzellen in einem SRAM-Verbund. Die Anforderungen an den Zugriff auf den GC-Puffer sind:

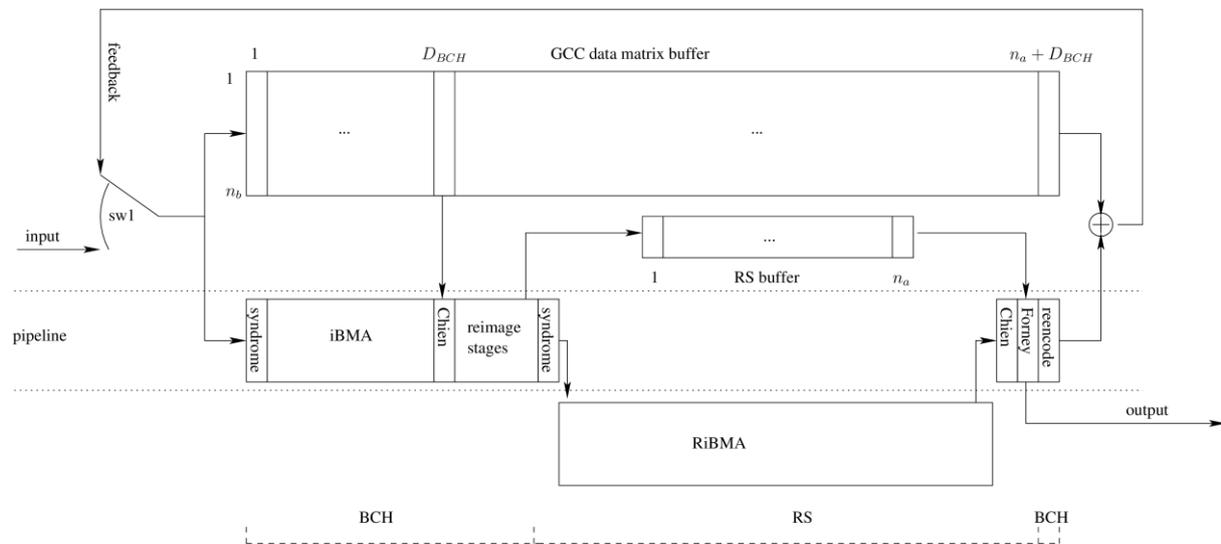


Abbildung 3: GC-Pipeline Struktur.

- spaltenweiser Zugriff
- inkrementelle Adressierung
- zwei Lese- sowie ein Schreibzugriff mit konstanten Spaltenabständen

A. Einzelnes SP-SRAM

Ein einzelnes Single-Port SRAM benötigt für den GC-Puffer eine Datenbreite von n_b Bits. Seine Länge wäre n_a Zeilen. Die Spalten des GC-Codewortes wären in ihrer Position statisch an die Adresse im SRAM gebunden. Diese einfache und flächenmäßig kleinste Lösung würde dazu führen, dass die drei Zugriffe je einen Zyklus benötigen und ein Pipeline-schritt somit mit einem Drittel des Taktes des Speichers getaktet werden müsste. Die Decodier-Logik ließe allerdings dieselbe Taktrate zu, die ein SRAM hätte. Daher würde mit dieser Lösung die Geschwindigkeit des Decoders nicht ausgenutzt.

B. Einzelnes DP-SRAM

Ein einzelnes Dual-Port SRAM wäre etwas größer als ein SP-SRAM aber noch deutlich kleiner als ein Schieberegister. Auch hier wären die Spalten des GC-Codeworts in ihrer Position statisch. Die drei Zugriffe könnten in je einen Doppelzugriff im ersten Zyklus und einen Einzelzugriff im zweiten Zyklus verteilt werden. Die Anzahl der Zugriffszyklen auf das DP-SRAM können nun von drei auf zwei reduziert werden.

C. Zwei DP-SRAM

Ein weiterer Lösungsweg ist die Zerlegung des Schieberegisters in zwei DP-SRAM. Das Schieberegister wird an der Stelle D_{BCH} aufgeteilt und durch je ein DP-SRAM ersetzt. In diesem Szenario wird die Funktionalität eines Schieberegisters durch einen

Ringpuffer nachgebaut. Die Spalten des GC-Codeworts würden zunächst in den ersten Ringpuffer eingelesen. Das Ende des Ringpuffers wird ausgelesen und in den zweiten Ringpuffer eingefügt sowie zur Korrektur nach der BCH-Chien-Search und anschließenden Berechnung des RS-Syndroms verwendet. Ist anschließend der zweite Teil der Pipeline aktiv, werden die Spalten des GC-Codeworts nach dem Abziehen der decodierten Ebene vom Spaltencode wieder in den ersten Ringpuffer eingefügt und gleichzeitig der ersten Pipeline zur Verarbeitung zur Verfügung gestellt.

Diese Lösung könnte alle nötigen Speicherzugriffe je Zyklus bedienen. Sie benötigt allerdings mehr Fläche als eine Lösung mit SP-SRAM. Dieser Mehraufwand kommt einerseits vom höheren Speicheraufwand des ersten Ringpuffers, zum anderen benötigen DP-SRAM pro Bit-Zelle mehr Fläche als SP-SRAM.

D. Drei SP-SRAM

Die in dieser Arbeit genauer untersuchte Lösung basiert auf drei SP-SRAM. Diese drei SRAM werden gleichzeitig im Wechsel an den drei Zugriffspunkten der Pipeline auf den Codewort-Puffer geschrieben bzw. ausgelesen. Die Spalten des GC-Codeworts sind im Wechsel auf die SRAM verteilt. Die SRAM-Zuordnung lässt sich durch die Modulo Funktion bestimmen.

$$\text{Spalte \% 3} = \text{SRAM}_{id}$$

Die Adresse innerhalb des SRAM ist der abgerundete Quotient aus:

$$\frac{\text{Aktuelle}_{\text{Spalte}}}{3} = \text{SRAM}_{\text{Adresse}}$$

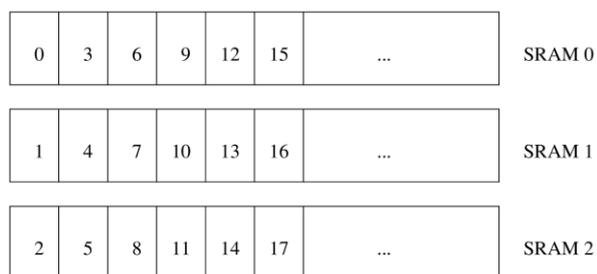


Abbildung 4: Verteilung der Spalten auf drei SRAMs.

In Abbildung 4 wird die Verteilung der GC-Spalten schematisch dargestellt. Mithilfe dieser Konstruktion ist gleichzeitiger Zugriff auf den Speicherverbund möglich, wenn die Zugriffe einen Abstand von Modulo der Anzahl der SRAM plus einen spezifischen Wert für den Puffer-Zugriff einhalten.

Für den Fall, dass die Abstände der Zugriffe aus der Decodier-Logik diese Regel nicht erfüllen, können zusätzliche Puffer verwendet werden. Man würde im schlechtesten Fall vier zusätzliche Register mit der Größe einer Spalte benötigen. Um das bereits vorhandene Schieberegister zu ersetzen, setzt sich, wie in Abbildung 5 zu sehen, das Modul SRAM-Puffer aus den Modulen *SRAM-Iterator* und dem SRAM-Speicherverbund (*multiple sram*) zusammen.

Das Modul *SRAM-Iterator* steuert die Adressierung und den Zugriff auf die drei SRAM. Es setzt und inkrementiert diese Adressen, die sich aus einem Minor- und mehreren Major Feldern, wie in Abbildung 4 dargestellt, zusammensetzt. Die Minor-Nummer wird mit jedem Zyklus inkrementiert und läuft von 0 bis 2. Je nach Minor-Adresse wird die dazugehörige Major-Adresse inkrementiert. Dieser SRAM-Iterator inkrementiert für jeden Zugriffspunkt der Pipeline eine Adresse. Damit beinhaltet dieses Modul drei Major-Adressregister. Diese Major-Adressen müssen vom Iterator so gestartet werden, dass die Zugriffe auf die logischen Spalten denen der Schieberegisterimplementierung entsprechen.

Das Modul SRAM-Speicherverbund beinhaltet drei SRAM-Instanzen. Es wertet die Minor-Adresse aus und leitet dementsprechend die Major-Adressleitungen sowie die Daten- und Steuerleitungen (Schreib- oder Lesezugriff) an das jeweilige SRAM weiter, welches hiermit durch die Minor-Adresse eindeutig bestimmt wird.

Diese Lösung hat den Vorteil, dass der Speicher-Overhead aus dem Schieberegister und den aufgeteilten DP-SRAM nicht benötigt wird. Dieser Overhead verschwindet, da die Positionen der GC-Codewortspalten statisch im Speicher sind. Darüber hinaus wird der Logik- und Leitungs-Overhead eines DP-SRAM durch die Verwendung von SP-SRAM vermieden. Die

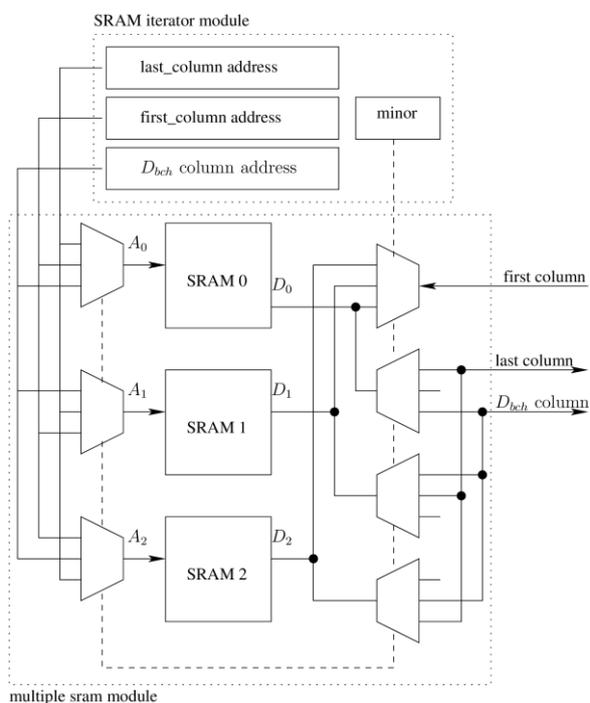


Abbildung 5: Steuerung des SRAM Zugriffs.

Tabelle 1: Flächenbedarf.

Decoder	Fläche μm^2	Einsparung
BCH über einen Sektor	350258	
GC-Code mit Flip-Flop	251434	28%
GC-Code mit SRAM	131242	62%

Speicherzugriffe für alle drei Zugriffspunkte der Pipeline sind in jedem Takt gleichzeitig und kollisionsfrei durchführbar.

Der vernachlässigbare Nachteil ist zum einen, dass die Speichersteuerlogik, wie z.B. der Adress-Decodierer, für jeden der drei Speicher vorhanden sein muss. Darüber hinaus benötigt diese Version zusätzliche Logik, wie das Multiplexen der Daten und Adressleitungen auf die drei Speicher, sowie die Ansteuerung der drei Speicher über die beschriebene Adressierungstechnik.

V. ERGEBNISSE

Für einen Vergleich wurden die Verilog-Implementierungen des BCH-Codes über einen gesamten Sektor, die GC-Variante mit einem Schieberegister und die GC-Variante des mehrfachen SP-SRAM mit dem Design Compiler von Synopsys

synthetisiert. Die Flächenergebnisse für eine Strukturgröße von 40 nm sind in Tabelle 1 gegenübergestellt.

Die Decoder-Logik kann mit einer Frequenz von 182-263 MHz getaktet werden. Ein SP-SRAM in dieser Technologie würde mit deutlich weniger als 600 MHz gerade noch einen sicheren Datenzugriff erlauben. Daher müsste die Taktrate für die Variante aus Abschnitt IV-A reduziert werden und würde somit weniger Durchsatz erzielen.

VI. ZUSAMMENFASSUNG

In diesem Beitrag wurde gezeigt, wie die Implementierung eines GC-Codes optimiert werden kann. Die Optimierung konzentriert sich dabei auf den Codewort-Puffer. Hierbei wurde ausgenutzt, dass ein Schieberegister durch ein deutlich kleineres SP-SRAM ersetzt werden kann. Die auftretenden Kollisionen der Zugriffspunkte der beschriebenen Pipelinestruktur werden durch Aufteilen des Speichers in drei Segmente und eine geschickte Ansteuerung gelöst. Mit dieser Verbesserung wurde eine beachtliche Flächenreduktion erzielt. Im Vergleich zu einem Decoder für einen vergleichbaren BCH-Code [5] wird eine Verdopplung des Datendurchsatzes erzielt. Der GC-Decoder benötigt dabei nur rund ein Drittel der Chip-Fläche des BCH-Decoders.

DANKSAGUNG

Wir danken der Firma Hyperstone für die Unterstützung dieser Arbeit.

LITERATURVERZEICHNIS

- [1] R. Michelsoni, A. Marelli, and R. Ravasio, *Error Correction Codes for Non-Volatile Memories*. Springer, 2008.
- [2] J. Spinner, J. Freudenberger, C. Baumhof, A. Mehnert, R. Willems, "A BCH Decoding Architecture with Mixed Parallelization Degrees for Flash Controller Applications", *26th IEEE International SoC Conference (SOCC)*, Erlangen, Germany, Sep. 2013.
- [3] J. Freudenberger, U. Kaiser, J. Spinner: "Concatenated code constructions for error correction in non-volatile memories", *International Symposium on Signals, Systems and Electronics (ISSSE)*, Potsdam, Germany, Oct. 2012.
- [4] J. Spinner, J. Freudenberger, „Design and Implementation of a Pipelined Decoder for Generalized Concatenated Codes“, *27th Symposium on Integrated Circuits and Systems Design (SBCCI)*, Aracaju, Brazil, 2014.
- [5] J. Freudenberger, J. Spinner: "A configurable Bose-Chaudhuri-Hocquenghem codec architecture for flash controller applications", *Journal of Circuits, Systems, and Computers*, 15 pages, 2013.



Jens Spinner erhielt den akademischen Grad des BSc in Technischer Informatik im Jahr 2009 sowie den MSc in Informatik im Jahr 2011 von der HTWG-Konstanz. Er ist wissenschaftlicher Mitarbeiter der Hochschule Konstanz.



Dr. Jürgen Freudenberger ist seit 2006 Professor an der Hochschule Konstanz. Dort leitet er das Institut für Systemdynamik. Seine Forschungsarbeit beschäftigt sich vorrangig mit der Entwicklung von Algorithmen im Bereich der Signalverarbeitung und der Codierung für zuverlässige Datenübertragung sowie mit der effizienten Umsetzung der Verfahren in Hard- und Software.

Entwicklung einer innovativen Softwarearchitektur für Cloud-basierte Cyber-physische Systeme

Maximilian Engelsberger, Thomas Greiner

Zusammenfassung—Die Integration von Cloud Computing in Cyber-physische Systeme ermöglicht die flexible Nutzung von IT-Diensten und stellt nahezu unbegrenzte Mengen an Speicherplatz und Rechenkapazität für eingebettete Systeme bereit [1]-[3]. Durch leistungsfähige Cloud-basierte Algorithmen werden verbesserte und situationsabhängige Entscheidungen anhand des aktuellen Ist-Zustands eines physischen Systems möglich, ohne dabei die Rechenleistung des eingebetteten Systems vor Ort zu erhöhen [4]-[6]. Einige der Herausforderungen bei Cloud-basierten Steuerungen liegen in der nahtlosen Integration, dem Umgang mit Verbindungsausfällen, der Abmilderung des Vendor-Lockins [7] sowie in der Verkürzung der Entwicklungszeiten. Im Beitrag wird ein neues Softwareframework für die Entwicklung Cloud-basierter Algorithmen nach dem SaaS-Modell (Software-as-a-Service) vorgestellt. Das Framework unterstützt durch eine Codewiederverwendung, sowie durch die Anwendung und Anpassung von etablierten Entwurfsmustern, den Aufbau Cloud-basierter Steuerungssysteme und hilft die Entwicklungszeiten zu verkürzen. Durch eine angepasste Multi-Tier- und Multi-Layer-Architektur [8] wird gezeigt, wie eine nahtlose Integration von Sensoren und Aktuatoren in Cloud-basierte Steuerungen möglich ist und die Migration zwischen verschiedenen Anbietern erleichtert werden kann. Eine Fehlerbaumanalyse und daraus abgeleitete Fehlerbehandlungsstrategien liefern Lösungsansätze für den Umgang mit Verbindungsausfällen. Die entwickelten Lösungen werden anhand eines Anwendungsbeispiels aus dem Bereich der Regenerativen Energiesysteme erprobt und bewertet.

Schlüsselwörter—Cyber-physische Systeme, eingebettete Systeme, Cloud Computing, Software-as-a-Service (SaaS), Multi-Layer-Architektur, Multi-Tier-Architektur, Middleware, Solar Tracking.

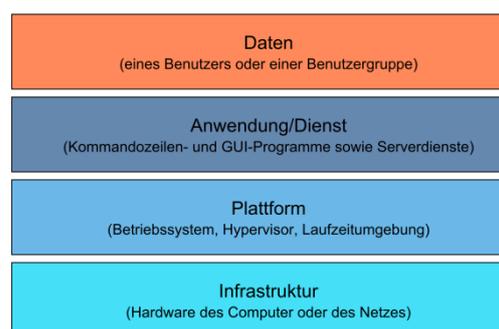


Abbildung 1: Ressourcen-Typen eines Computersystems, die als XaaS-Modell im Cloud Computing über Netze bereitgestellt werden.

I. EINLEITUNG

Cyber-physische Systeme (CPS) bilden das Rückgrat der gerade anlaufenden vierten industriellen Revolution. Es handelt sich hierbei um eingebettete Systeme, die eng mit ihrer physischen Umgebung verbunden sind und über öffentliche Netze Dienste in Anspruch nehmen oder selbst anbieten [1], [2]. Herkömmliche Steuerungssysteme können, aufgrund ihrer beschränkten Rechenkapazität und ihrer abgeschlossenen Systemumgebung, nur vergleichsweise starre Abläufe realisieren [9]. Cloud Computing beschreibt ein Paradigma, mit welchem IT-Ressourcen wie z.B. Infrastrukturen, Plattformen, Software und Daten flexibel über Kommunikationsnetze bereitgestellt und genutzt werden können [3]. Bestehende Ansätze versuchen die Vorteile des Cloud Computings, wie z.B. die Skalierbarkeit der bereit gestellten Ressourcen, mit Maschinen- und Anlagensteuerungen zu verbinden [5] [6]. Durch die Einbindung Cloud-basierter IT-Dienste werden dem CPS dabei nahezu unbegrenzte Mengen an Speicherplatz und Rechenkapazität bereitgestellt, die innerhalb des Systems genutzt werden können [3].

Aufbauend auf den Paradigmen des Cloud Computings und der CPS wird das im Beitrag vorgestellte Smart Systems Framework (SSF) entwickelt. Dieses soll eine nahtlose Integration von Sensoren und Aktuatoren in eine Cloud-basierte Steuerung ermöglichen und sinnvoll auf Verbindungsausfälle reagieren. Auch das sog. Vendor-Lockin, das entsteht, wenn zwischen verschiedenen Anbietern von Cloud Computing oder zwischen verschiedenen Hardwareanbietern gewech-

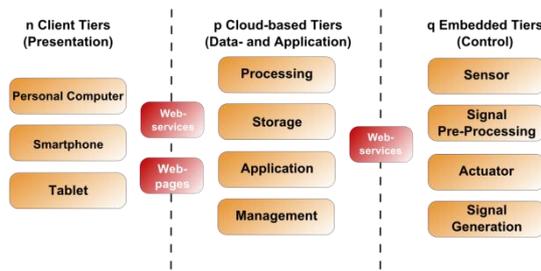


Abbildung 2: Die Multi-Tier-Architektur beinhaltet Embedded Tiers mit netzwerkfähigen Sensoren und Aktuatoren, Cloud Tiers für die Datenverarbeitung und Client Tiers für den Benutzerzugriff.

selt werden soll, wird durch die neue Architektur abgemildert. Ziel des SSF ist es darüber hinaus, dem Entwickler die notwendigen Werkzeuge bereit zu stellen, die ihn beim Aufbau von Cloud-basierten Steuerungen unterstützen. Im Idealfall muss sich der Entwickler dann lediglich um die Implementierung der eigenen Anwendungslogik kümmern, nicht aber um Details der Kommunikation mit den Cloud-basierten Ressourcen oder bspw. das Verbindungsmanagement. Dies hilft dabei, die zunehmende Komplexität solcher Systeme zu beherrschen und gleichzeitig die Entwicklungszeiten zu verkürzen. Diese verlängern sich tendenziell bisher im Vergleich zu klassischen Steuerungen durch den Einsatz unterschiedlicher Plattformen und Programmiersprachen. Frameworks, welche diesen Problemen entgegenreten, stehen Cloudanbieter- und Systemhersteller-unabhängig bisher nicht zur Verfügung. Mit dem SSF wird die für diese Aufgabe geeignete Softwarearchitektur umgesetzt. Die Architektur basiert auf dem, im Cloud Computing üblichen, Servicemodell SaaS (Software-as-a-Service) [3].

II. NEUER ARCHITEKTUR-ANSATZ FÜR CLOUD-BASIERTE STEUERUNGEN

A. Multi-Tier-Architektur

Das angewendete Software-Modell ist angelehnt an das in der Softwareentwicklung übliche Multi-Tier-Entwurfsmuster (siehe Abb. 2) [8]. Hierbei werden verschiedene Teile eines Software-Systems auf unterschiedliche physische Systeme (Tiers) verteilt. Das Entwurfsmuster wurde für die Verwendung im SSF auf drei spezialisierte Tier-Typen angepasst. Die Aufgaben und Eigenschaften dieser neuen Tier-Typen seien im Folgenden erläutert. Das Gesamtsystem, bestehend aus allen Tiers, wird dabei als Cluster bezeichnet.

1) Embedded Tiers

Embedded Tiers verfügen über keine erweiterten Analyse- und Verarbeitungsressourcen, sondern sind möglichst leichtgewichtig, energieeffizient und kostengünstig ausgelegt. Ihre Infrastrukturschicht (siehe

Abbildung 1) bilden eingebettete Systeme, also Hardware, die über ein sehr anwendungsspezifisches Design verfügt [10]. Sie dienen zum Ansteuern der Sensoren und Aktuatoren und zum Datenaustausch mit den Cloud-basierten Tiers (siehe Abschnitt 2). Wird ein eingebettetes System als Teilkomponente eines CPS nach Broy verwendet, so verfügt das Gerät über eine Netzchnittstelle, z.B. in Form einer drahtgebundenen oder drahtlosen Schnittstelle [1, 2].

2) Cloud-basierte Tiers

Die Cloud-basierten Tiers stellen die Ressourcen für die Ausführung umfangreicher Algorithmen und wichtige Teile der Steuerlogik bereit. Sie nehmen Sensordaten von den Embedded Tiers entgegen und generieren neue Steueranweisungen, welche sie an die Aktuatoren senden. Außerdem stellen sie die Anwendungslogik für die Client Tiers (siehe Abschnitt 3) bereit und dienen als Speicherort für alle globalen Daten des Clusters. Cloud-basierte Tiers können weitere Datenquellen mit in die Verarbeitung einbinden, z.B. durch über Webservices erreichbare Drittquellen.

Bei Cloud-basierten Tiers kann es sich prinzipiell um alle Typen von Cloud-Ressourcen handeln: Infrastrukturen, Plattformen, Anwendungen und Dienste sowie Daten (siehe Abb. 1). Voraussetzung ist, dass die Ressourcen über Webservices erreichbar sind, also nach dem XaaS-Modell (Everything as a Service) implementiert sind. Im vorgestellten Beispiel werden Cloud-basierte Algorithmen nach dem SaaS-Modell (Software-as-a-Service) auf einem Cloud-basierten Tier implementiert, welche ein PaaS-Modell (Platform-as-a-Service) erweitern.

Durch die Skalierungsfähigkeiten einer Cloud-Ressource kann die genutzte Speicher- und Rechenkapazität jederzeit an die Erfordernisse des Clusters angepasst werden. Erhöht sich die Anzahl der Sensoren und Aktuatoren im Feld, so können die bereitgestellten Steuer-, Verarbeitungs- und Speicher-Ressourcen in der Cloud flexibel erweitert werden. Das System skaliert somit optimal. Des Weiteren werden die in Anspruch genommenen Cloud-Ressourcen nutzungsabhängig abgerechnet, was den Auf- und Ausbau von Cloud-basierten Cyberphysischen Systemen nach diesem Prinzip erleichtert [11].

3) Client Tiers

Die Client Tiers dienen als primäre Human-Machine Interfaces (HMI) bzw. grafische Benutzeroberflächen (Graphical User Interface, GUI) gemäß dem hier an angewendeten Entwurfsmuster [8]. Dabei kann es sich z.B. um herkömmliche Endanwendergeräte wie x86- oder ARM-basierte PCs, Tablet-Computer oder Smartphones handeln. Auf ihnen muss nicht zwingend eine Instanz des SSF laufen, sondern lediglich ein herkömmlicher Webbrowser. Über Cloud-basierte Tiers stellt das Cluster Webanwendungen bereit, wel-

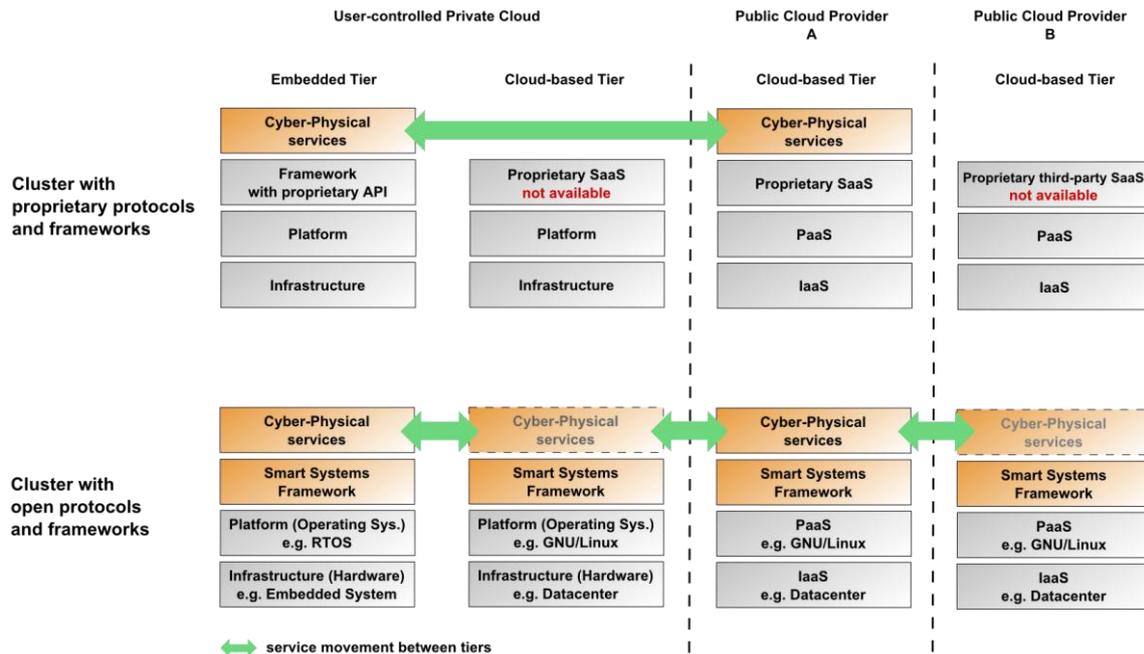


Abbildung 3: Bei Clustern, welche proprietäre Protokolle und Frameworks verwenden, besteht die Gefahr eines Vendor-Lockins. Durch Cluster mit offenen Protokollen und Frameworks wird die Migration eigener SaaS-Anwendungen zwischen verschiedenen Cloud-Dienstanbietern und Private Clouds vereinfacht.

che von den Client Tiers genutzt werden. Über ebenfalls Cloud-basierte Datenbankinstanzen erfolgt der Datenaustausch mit den anderen Tiers desselben Clusters.

4) Clustermanagement-Datenbank und Webfrontend

Über die auf einem Cloud-basierten Tier untergebrachte Clustermanagement-Datenbank werden die Konfigurationsparameter des Clusters abgelegt. Dies umfasst bspw. die Kommunikationsraten zwischen den Embedded Tiers und den Cloud-basierten Tiers. Über eine weitere Datenbank-Entität wird außerdem die Schnittstelle zum Webfrontend bereitgestellt. Das Webfrontend ist hier eine PHP-basierte Webanwendung, über welche die Client Tiers Statusinformationen über das Cluster abrufen oder Steuerkommandos an das Cluster senden können. Die Clustermanagement-Datenbank kann anwendungsabhängig erweitert werden.

5) Plattforminteroperabilität

Um ein nützliches Werkzeug für die Entwicklung Cyber-physischer Systeme mit unterschiedlichsten Infrastrukturschichten zu realisieren, ist es wichtig, dass das zu Grunde liegende Framework auf möglichst vielen unterschiedlichen Systemen ausführbar ist. Aus Gründen der einfachen Portierbarkeit wurde daher auf C99 und Java gesetzt [12]. Für die Kommunikation zwischen den Tiers kommen ausschließlich freie Protokolle zum Einsatz (siehe Abb. 3). Dies sind in den

unteren Schichten die üblichen Internet-Protokolle wie UDP, TCP/IP und HTTP. Darauf aufbauend werden Webservices nach dem REST-Architekturstil betrieben [13]. Als API hierfür wird die freie REST-Implementierung JAX-RS/Jersey von Oracle eingesetzt [14].

B. Multi-Layer-Architektur

Das für das Cluster angewendete Software-Modell ist angelehnt an das in der Softwareentwicklung übliche Multi-Layer-Entwurfsmuster [8]. Es dient klassischer Weise zur funktionalen Abgrenzung einzelner Programmdomänen und wird durch das SSF als logische Abstraktionsschicht zwischen der zugrundeliegenden Plattform und der eigentlichen Steuerungslogik verwendet. Die Abstraktionsschicht und die Unterteilung der Layer in Module dienen hierbei zur schrittweisen Abstraktion der Peripherie-Funktionen und Cloud-Dienste. Über die hierfür entwickelten Layer und Module des SSF sei im Folgenden ein Überblick gegeben.

1) Schichten und Module

Das SSF wird, aufbauend auf der Plattformschicht, ausgeführt oder kann Teil davon sein, z.B. als Kernelmodul. Insbesondere bei eingebetteten Systemen ist es durchaus üblich, Anwendungsmodule direkt in das Betriebssystem zu integrieren [10]. Dies ist auch beim hier eingesetzten Echtzeitbetriebssystem Nut/OS auf einem ARM9 SoC der Fall [15]. Das SSF besteht auf oberster Ebene aus der Cyber-physischen

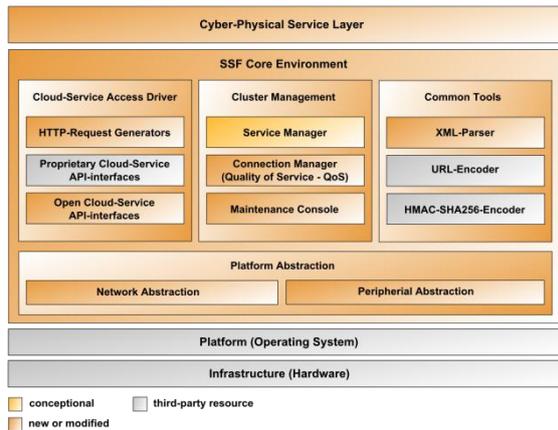


Abbildung 4: Die Multi-Layer-Architektur des Smart Systems Framework dient als Abstraktionsschicht zwischen der Plattform und der Anwendungslogik im Cyber-physical Service Layer.

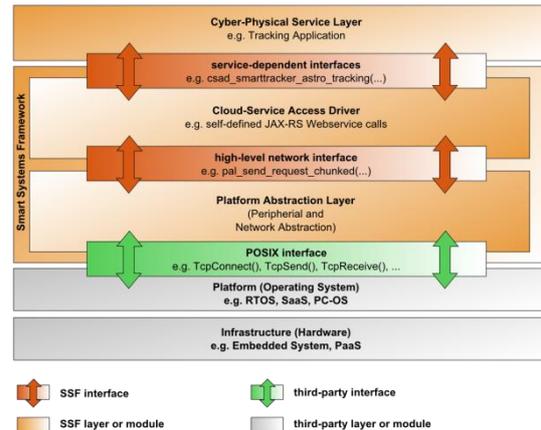


Abbildung 5: Das Smart Systems Framework setzt eine schrittweise Abstraktion der genutzten IT-Ressourcen um und ermöglicht so ein anwendungsfokussiertes Entwickeln.

Diensteschicht (Cyber-physical Service Layer, CPSL), in welcher die Anwendungslogik eines Tiers implementiert ist. Das SSF ist modular aufgebaut und beinhaltet alle anwendungsübergreifenden Module, auf die im Folgenden genauer eingegangen wird.

2) Kernmodule

Die Kernumgebung gliedert sich in derzeit vier Hauptmodule (siehe Abb. 4). Die beiden wichtigsten sind die Plattformabstraktionsschicht (Platform Abstraction Layer, PAL) sowie der Clouddienste-Zugriffstreiber (Cloud Service Access Driver, CSAD). Dazu kommen die Kernmodule für das Clustermanagement (CM), sowie ein Modul für oft benötigte Hilfsfunktionen, die sog. Common Tools (CT). Der CSAD bietet dem CPSL die Schnittstellen zum Zugriff auf die Dienste eines Cloudanbieters. Hierbei kann es sich um hinzugelinkte proprietäre application programming interfaces (API) handeln, welche direkt vom Drittanbieter stammen. Dieses Verfahren hat sich auf x86-basierten Tiers bewährt, da hier in den meisten Fällen eine passende API bereits existiert. Für eingebettete Systeme ist u.U. keine proprietäre API in der für die Infrastruktur geeigneten Programmiersprache - meistens C99 - verfügbar.

In diesem Fall müssen die API-Methoden für die RPC-Aufrufe direkt im CSAD implementiert werden, wie das im hier gezeigten Anwendungsbeispiel umgesetzt wurde. Innerhalb des CSAD befindet sich das für die REST-basierte Kommunikation benötigte Set an HTTP-Request-Generatoren für alle unterstützten Anbieter von Cloud Computing Diensten. Für die im Beitrag vorgestellte Beispielanwendung sind das momentan Amazon Webservices (AWS) [16].

Das Clustermanagement (CM) fasst alle zum Betrieb eines Cloud-basierten Cyber-physischen Systems notwendigen Funktionalitäten zusammen. Dazu zählen z.B. die Methoden zum Zugriff auf die Clustermanagement-Datenbank (siehe Abschnitt II) und die

Wartungskonsole zur lokalen Konfiguration eines Tiers über ein Standard-I/O-Terminal, als auch z.B. der Verbindungs-Manager. Innerhalb des Verbindungs-Managers stehen Routinen bereit, um fortlaufend eine Bewertung der Verbindungsqualität (Quality-of-Service, QoS) zwischen den kommunizierenden Tiers vorzunehmen (siehe Abschnitt A).

In den Common Tools (CT) sind weitere, Tier-weit benötigte Komponenten untergebracht. Hierzu zählen bspw. ein URL- und ein HMAC-256-Encoder zur Signaturerzeugung sowie ein einfacher XML-Parser [16]. Der PAL bildet das Kernmodul für die Low-Level-Peripheriekomponenten wie bspw. angebundene Sensoren und Aktuatoren auf Embedded Tiers. Die beiden wesentlichen Komponenten innerhalb des PAL sind die Netzabstraktion sowie die Peripherieabstraktion. Die Netzabstraktion bietet einen komfortablen und abstrahierten Zugriff auf den TCP/IP-Stack der zugrundeliegenden Plattform und abstrahiert die POSIX-Schnittstelle für das Framework [17]. In der Peripherieabstraktion sind die für das System benötigten Treibermodule organisiert.

C. Mehrstufige Abstraktion

Während herkömmliche Treiberbibliotheken lediglich dafür sorgen, dass eine gewisse Infrastruktur-Ressource aus dem Anwendungsprogramm heraus nutzbar wird, arbeitet das SSF konsequent mit schrittweise aufeinander aufbauender Abstraktion mit Hilfe von Nested-Function-Calls. So ist im Beispiel das Socket-Interface, als POSIX-kompatible Schnittstelle zum TCP/IP-Stack der Plattform, für eine Benutzung vollkommen ausreichend. Allerdings müssen Verbindungsaufbau, Fehlerbehandlung und das Senden und Empfangen von Daten stets von Neuem implementiert werden. Das kostet Zeit und ist generell fehleranfällig. Durch die Netzabstraktion innerhalb des PAL wird das Abstraktionsniveau eine erste Stufe angehoben (siehe Abb. 5). Diese kann nun von höheren Schichten be-

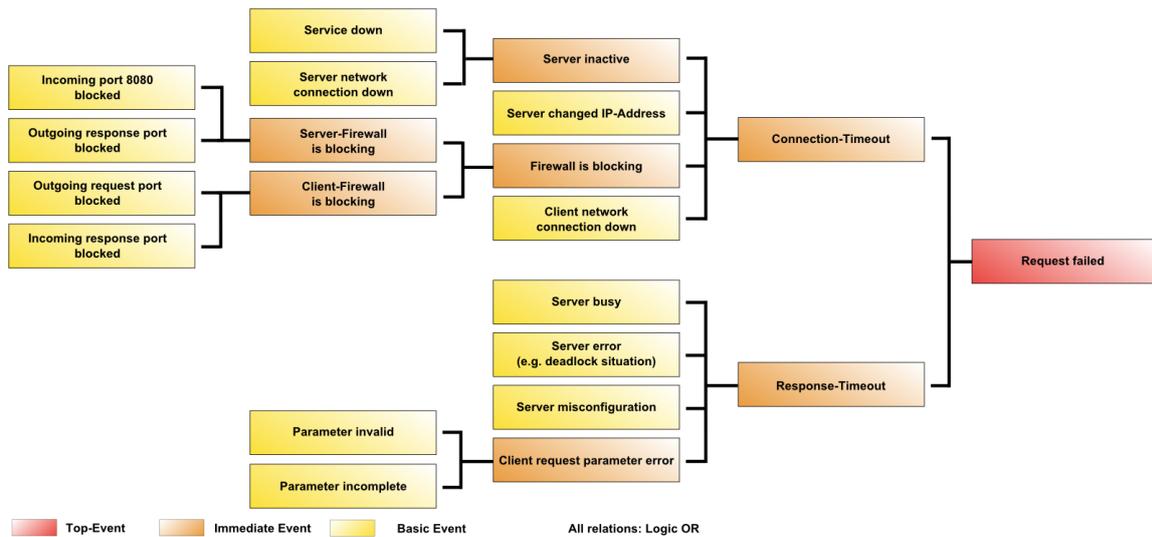


Abbildung 6: Eine qualitative Fehlerbaumanalyse (Fault-Tree-Analysis, FTA) ermöglicht die Verbesserung der Fehlererkennung und Fehlerbehandlung bei RPC-Aufrufen (Remote Procedure Calls) zwischen Embedded Tiers und Cloud-basierten Tiers.

nutzt werden, um weitere, Applikations-spezifische Abstraktionen vorzunehmen. Im Beispiel greift der CSAD auf die Netzabstraktion des PAL zu und nutzt diesen für die Request-Abwicklung. Der CSAD kapselt wiederum diese Funktionalität auf logisch-funktionalem Level und stellt diese nach oben hin dem CPSL bereit.

Von dort kann der Anwendungsentwickler innerhalb des CPSL diesen Dienst, z.B. auf einem eingebetteten System für die Implementierung seiner Anwendungslogik nutzen, ohne konkretes Wissen über Cloud Computing, RESTful Webservices oder die konkreten API-Funktion zu benötigen.

III. VERBINDUNGSMANAGEMENT

A. Fehlerbaumanalyse

Bei einer Cloud-basierten Datenverarbeitung ist die Kommunikation über die Internet-Infrastruktur ein kritischer Vorgang. Schlägt diese fehl, so versagt u.U. die gesamte Steuerung. Aus diesem Grund wurde das Scheitern eines Kommunikationsvorgangs zwischen zwei Tiers als Top-Fehlerereignis für die qualitativ durchgeführte deduktive Fehlerbaumanalyse (Fault-Tree-Analysis, FTA) identifiziert (siehe Abb. 6) [18]. Da für die Internet-Kommunikation, selbst bei Verwendung einer Wegwahltechnik wie IP-basiertes Source Routing, keine Aussage über die konkreten Kommunikationssysteme wie z.B. Router und Switches gemacht werden kann, muss dieser Teil der FTA ebenfalls als Blackbox-Ereignis zusammengefasst werden. In Abbildung 6 ist schematisch die durchgeführte FTA mit den wichtigsten Fehlerereignissen zu sehen. Sie dient in diesem Fall als Ausgangspunkt zur Ermittlung der Ausfallmöglichkeiten, also zur Identifikation der Basisereignisse. Eine konkrete Verknüp-

fung von Ausfallwahrscheinlichkeiten zur Berechnung einer Gesamtausfallwahrscheinlichkeit ist aus den oben genannten Gründen nicht sinnvoll möglich.

B. Fehlererkennung

Die Fehlererkennung ermöglicht es, zur Laufzeit auftretende Fehlersituationen automatisch zu detektieren, anhand der durch die FTA identifizierten Fehlerereignisse einzuordnen und schließlich der Fehlerbehandlung zuzuführen. Eine der im SSF implementierten Fehlererkennungsstrategien ist das Austauschen eines Heartbeat-Signals [19] zwischen den einzelnen Tiers eines Clusters. Bei einem dauerhaften Ausbleiben des Signals kann dann wechselseitig auf einen Ausfall der Gegenstelle geschlossen werden. Da die Kommunikation über drahtlose Schnittstellen wie bspw. Bluetooth, WLAN oder GSM vergleichsweise viel Energie kostet, sollte die Heartbeat-Kommunikation auf das Allernötigste minimiert werden. Auch das verursachte Übertragungsvolumen über Mobilfunknetze sowie von und zum Clouddiensteanbieter sollte minimiert werden, da ein größeres Übertragungsvolumen in der Regel auch höhere Kosten verursacht. Bei der Nutzung von eigenen drahtlosen Sensornetzen ist hingegen die Auslastung der Übertragungskanäle zu minimieren, insbesondere bei einer großen Anzahl von Teilnehmern pro Kanal. Aus diesen Gründen ist ein ständiges Austauschen eines zusätzlichen Heartbeat-Signals, speziell bei Systemen mit einer großen Anzahl an Teilnehmern, am besten zu vermeiden. Stattdessen wird in der hier vorgestellten Lösung ein sog. virtuelles Heartbeat-Signal verwendet (siehe Abb. 7). Dabei sendet z.B. der Sensor periodisch Nutzdaten an die Cloud-basierten Algorithmen, das SSF wertet die übertragenen Nutzdaten zusätzlich durch den Verbindungsmanager (CM) aus und ermittelt daraus fortlaufend die auftretende La-

tenzzeit für einen vollständigen Übertragungsvorgang. Der CM stellt dabei die notwendigen Schnittstellen zum CPSL bereit und ermöglicht der Anwendungslogik, entsprechend auf einen Ausfall oder zu hohe Latenzzeiten zu reagieren. Der CPSL kann sich dann z.B. an die veränderte Verbindungsqualität (Quality of Service, QoS) anpassen, indem er die Übertragungsrate reduziert oder - im Falle eines längerfristigen Verbindungsausfalls - vordefinierte Notfallroutinen aufruft.

C. Fehlerbehandlung

Beim Erreichen von zur Entwurfszeit festgelegten Schwellwerten werden vordefinierte, anwendungsbezogene Notfallroutinen ausgeführt. Ein Beispiel hierfür wird in Abschnitt IV gegeben. Entlang der durch die FTA ermittelten Fehlerfälle können außerdem automatische Versuche unternommen werden, die bestehende Verbindung zu reparieren. Ein Beispiel hierfür ist, dass in der FTA ein Basisereignis identifiziert wurde, wonach eine fehlerhafte Netzkonfiguration vorliegt. Wird dieser Fehlerzustand durch die Fehlererkennung bestätigt, kann der betroffene Tier versuchen, seine Konfiguration mittels DHCP zu erneuern und den Request zu wiederholen.

IV. ANWENDUNG

Die Architektur des SSF wird anhand eines Anwendungsbeispiels aus dem Bereich der Regenerativen Energiesysteme umgesetzt und erprobt. Hierbei wird ein System zur Umwandlung von Solarenergie in Elektrizität mittels Photovoltaik und Solarthermie betrachtet. Im Vordergrund steht dabei die Verbesserung informationstechnischer und steuerungstechnischer Aspekte von zweiachsig nachführbaren Solaranlagen, sogenannter Solartracker [20, 21].

A. Solartracking

Die durch die Erdrotation verursachte scheinbare Bewegung der Sonne am Taghimmel führt dazu, dass ein starr montiertes Photovoltaikmodul, wie es z.B. auf Hausdächern üblich ist, nur für eine relativ kurze Zeit des Tages optimal von der Sonne beschienen wird. Die Leistungsausbeute von Photovoltaikmodulen ist allerdings nur dann maximal, wenn möglichst viel Lichtenergie auf die Fläche der Module trifft. Dies ist lediglich bei einer orthogonalen Ausrichtung der Ebene der Module zum Einfallswinkel der Sonnenstrahlen der Fall.

Solartracker sind elektromechanische Systeme, die es ermöglichen, eine Nutzlast, wie z.B. ein Photovoltaikmodul, dem scheinbaren Lauf der Sonne am Taghimmel nachzuführen. Dabei wird das azimutale Koordinatensystem bzw. Horizont-Koordinatensystem verwendet. Die Nachführung erfolgt klassischer Weise mit Hilfe einer orts- und zeitabhängigen, vorprogrammierten Route (Track) der Sonne. Die Nachfüh-

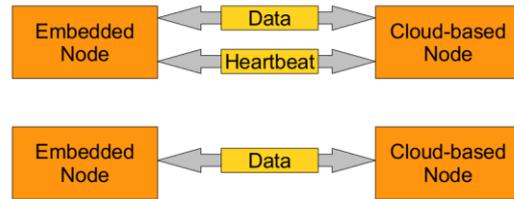


Abbildung 7: Der Verbindungsmanager ermittelt die Verbindungsqualität (Quality-of-Service, QoS) anhand des periodischen Nutzdatenstroms. Das spart Energie und entlastet den Übertragungskanal.

zung kann dabei einachsig (z.B. über die Azimut-Ordinate) oder zweiachsig (über Azimut und Altitude erfolgen). Diese sog. astronomische Nachführung hat allerdings einige Nachteile: So entspricht bspw. bei sich häufig ändernden Bewölkungsverhältnissen die aktuelle Position der Sonne nicht immer der Richtung, aus der die meiste Lichtenergie kommt. Dies ist z.B. dann der Fall, wenn die Sonne durch Bewölkung verdeckt wird. In diesem Fall kann es Sinn machen, einen Solartracker sensorisch nachzuführen (Sensorisches Tracking). Hierbei wird mit Hilfe von Lichtsensoren der Beleuchtungsgradient am Himmel ermittelt und die Anlage auf den derzeit tatsächlich hellsten, also energiereichsten Bereich ausgerichtet.

Das sensorische Tracking ist bei vielen zweiachsigen Solartracker-Anlagen in Verwendung. Es weist aber ebenfalls einige prinzipbedingte Schwachstellen auf. So können z.B. defekte oder verschmutzte Sensoren zu einem fehlerhaften Tracking der Anlage führen. Ebenso können kurzzeitige Störungen, etwa durch Reflexionen an Fensterflächen oder Autoscheiben, eine Umpositionierung der Anlage auslösen, die aber keinen Mehrertrag an Energie bringt. Besonders kritisch sind häufig veränderliche Beleuchtungsverhältnisse, welche über den Lichtsensor zu häufigen Bewegungen der Anlage führen. Hierbei besteht die Gefahr, dass die Energie, die zum Bewegen der Anlage benötigt wird, den durch die Bewegung angestrebten Mehrertrag an Energie überschreitet. Die Energiebilanz fällt dann negativ aus. Cloud-basierte, smarte Tracking-Algorithmen sollen diese Ausgangssituation verbessern.

B. Zielsetzung des Smart Trackings

Die Zielsetzung eines Smart Tracking-Ansatzes ist es, die Anzahl der energetisch nicht sinnvollen Bewegungen eines Solartrackers durch eine intelligente Sensordatenauswertung zu minimieren. Dazu ist die Implementierung sog. smarterer Algorithmen notwendig. Diese unterscheiden sich von einer herkömmlichen Nachführsteuerung dadurch, dass sie intelligent auf sich dynamisch verändernde Umgebungsbedingungen reagieren können. Dies wird einerseits durch die Miteinbeziehung externer Datenquellen möglich und zum anderen durch eine aufwändigere Analyse der Ist-Situation des Systems, als sie durch eine herkömmliche Nachführsteuerung durchgeführt wird. Für



Abbildung 8: Das Kompositionsstrukturdiagramm nach UML 2.0 zeigt den Systemaufbau der Smart Tracker-Modellanlage. Links befinden sich die Schnittstellen zu den Client Tiers. Die beiden Module der Tracking Global Unit (TGU) bilden die Cloud-basierten Tiers und die Tracking Control Unit (TCU) stellt den Embedded Tier dar.

die Implementierung solcher smarten Algorithmen wird eine erhöhte Menge an Rechenkapazität und Speicherplatz benötigt. Diese Ressourcen können auf einer üblichen Steuereinheit für Solartracker nicht vorausgesetzt werden. Des Weiteren liegt die Stärke des Konzepts gerade in der Flexibilität zum sukzessiven Ausweiten der Analysefähigkeiten eines Steueralgorithmus zur Laufzeit – und der flexiblen Skalierbarkeit des damit verbundenen erhöhten Ressourcenbedarfs.

C. Modellanlage

Zur Demonstration von Smart Tracking-Algorithmen wird eine Modellanlage eines kleinen Feldes von Solartrackern konstruiert (siehe Abb. 8). Diese sind durch eine lokale Steuerbaugruppe (Tracking Control Unit, TCU) individuell ansteuerbar. Die TCU ist verbunden mit der Cloud-basierten Steuereinheit (Tracking Global Unit, TGU). Hierbei ist die TCU innerhalb des Software-Architekturmodells als eingebettetes System und die TGU als Cloud-basiertes System ausgeführt. Auf TCU und TGU laufen jeweils Instanzen des Smart Systems Framework. Die TGU besteht aus zwei Hauptkomponenten: Die erste ist die Cluster-Kommunikation, welche die REST-basierten Webservices für die Kommunikation mit den angebotenen TCUs abwickelt und in welcher auch die Smart Tracking Algorithmen implementiert sind. Die zweite Komponente ist die Schnittstelle zur Clustermanagement-Datenbank, über welche Konfigurationsparameter und anwendungsbezogene Trackingdaten in die jeweils zuständigen Tabellen geschrieben werden. Die TCU bildet die lokale Steuerbaugruppe des Trackerfeldes. Die Sensoren und Aktuatoren der Tracker-systeme sind über ein gemeinsames Switchboard und verschiedene Adapterboards an die Steuerbaugruppe angebunden.

D. Regelkreise

Da die Modellanlage aus drei Trackern mit jeweils zwei Freiheitsgraden (Azimut und Altitude) besteht, wurden insgesamt sechs quasi-parallel arbeitende Regelkreise realisiert. Jeder dieser Regelkreise besteht aus zwei gegensätzlich angeordneten lichtempfindlichen Phototransistoren. Sie ermöglichen die Auswertung der Richtung der Lichtintensität für jeweils eine Achse. Die Sensordaten werden durch die TCU an die

TGU übermittelt und dort prozessiert. Das in Abhängigkeit von den aktuellen Eingangswerten berechnete Stellensignal wird von der TGU über die TCU zurück an den jeweils zuständigen Servomotor übertragen. Dieser verfügt intern über einen weiteren, untergeordneten Regelkreis, der mittels eines Spannungsteilers das Eingangssignal als Sollwert mit dem Ist-Signal der Achsenposition des Positions-Encoders vergleicht.

E. Beispiel-Algorithmus: Smarte Sensordatenvalidierung

Ziel der hier gezeigten Softwarearchitektur ist die nahtlose Integration von eingebetteten Systemen und Cloud-basierten Steuerungen. Anhand einer intelligenten Sensordatenvalidierung wird ein Beispiel für einen Smart Tracking-Algorithmus gegeben. Wie bereits erläutert, können fehlerhafte Sensoren oder kurzzeitige Reflexionen zu einer schlechten Energiebilanz beim Tracking führen und so unter Umständen den Wirkungsgrad eines kompletten Solartracker-Feldes negativ beeinflussen. Der hier vorgestellte sensorische Trackingalgorithmus schafft hiergegen Abhilfe. Die Cloud-seitige Signalvalidierung kann, aufgrund der vorhandenen Rechenkapazität, viel umfangreicher ausgelegt und flexibler erweitert werden, als das auf einer herkömmlichen Steuerung möglich ist. Im hier umgesetzten Beispiel werden die beiden gängigen Tracking-Methoden intelligent kombiniert. Hierbei wird durch den Algorithmus standardmäßig ein einfaches, astronomisches Tracking durchgeführt und zusätzlich werden die Sensordaten ausgewertet. Kommt es nun über einen längeren Zeitraum zu umweltbedingten Abschattungen, wechselt der Algorithmus in den sensorischen Tracking-Modus und sucht sich die energieintensivste Stelle am Himmel. Stimmen die Output-Werte des astronomischen und sensorischen Trackings über einen längeren Zeitraum wieder überein, wird zum astronomischen Tracking-Modus zurück gewechselt. Dieses kleine Beispiel zeigt, wie durch intelligentere Algorithmen kurzfristige Reflexionen und Sensorstörungen nicht zu einem Umpositionieren der kompletten Tracker-Anlage führen und somit die Energieeffizienz gegenüber üblichen Steuerungen durch Cloud-basierte Algorithmen verbessert werden kann.

Im Falle eines Verbindungsausfalls zwischen TCU und TGU findet eine Fehlererkennung und Fehlerbehandlung auf der TCU-Seite gemäß der in Abschnitt

III durchgeführten Fehlerbaumanalyse statt. Kommt innerhalb des für die jeweilige Verbindung als sinnvoll ermittelten Timeouts keine Antwort von der TGU, so wird der Request wiederholt, bis die Verbindung wieder hergestellt werden konnte oder ein voreingestelltes Timeout-Quantity-Limit erreicht ist. In diesem Fall geht die TCU von einer dauerhaft ausgefallenen Verbindung zur TGU aus und führt eine vordefinierte Notfallroutine aus. In diesem Beispiel ist das die Ausrichtung der Tracker nach Süden. Im Hintergrund wird periodisch versucht, die Verbindung wiederherzustellen. Sobald dies gelungen ist, wird das unterbrochene Tracking fortgesetzt.

V. BEWERTUNG UND FAZIT

Es konnte gezeigt werden, dass elektromechanische Steuerungssysteme die keinem harten Zeitlimit unterliegen (Periodendauer z.B. ≥ 1000 ms) über Cloud-basierte Steuerungen mit Hilfe der hier vorgestellten Softwarearchitektur einfach zu realisieren sind. Das Ziel von nahtlos integrierten Sensoren und Aktuatoren in eine durchgängig Cloud-basierte Datenverarbeitung wurde erreicht. Durch das Smart Systems Framework kann die Architektur mit geringem Aufwand auf weitere Aufgabenstellungen aus dem Bereich der Cloud-basierten Cyber-physischen Systeme angewendet werden.

Die Vorteile wurden beispielhaft durch eine Automatisierungsanwendung aus dem Bereich der Regenerativen Energiesysteme demonstriert. Über leistungsfähige, Cloud-basierte Algorithmen konnten differenziertere Analysen der Ist-Situation des Systems vorgenommen werden, als dies auf eingebetteten Steuerungen möglich wäre. Mit dem SSF steht somit ein Werkzeug zur einfachen Implementierung von Cloud-basierten Steuerungsalgorithmen bereit. Ein weiterer Vorteil ist die Möglichkeit der sukzessiven Verbesserung der Cloud-seitigen Algorithmen in der Hochsprache Java ohne die Anpassung der Soft- oder Hardware auf den eingebetteten Systemen.

Das Framework ermöglicht die Entwicklung von extrem leichtgewichtigen und energiesparenden Sensoren und Aktuatoren im Anwendungsfeld, da durch den modularen Aufbau des SSF das Speicherfootprint für eingebettete Systeme minimiert werden kann.

Die Architektur skaliert, dank des angepassten SaaS-Modells, optimal mit der Anzahl der Sensoren und Aktuatoren. Ein wichtiges Alleinstellungsmerkmal zu bestehenden Ansätzen ist das hierdurch abgeschwächte Vendor-Lockin beim Wechsel zwischen verschiedenen Cloudanbietern oder Hardwareherstellern. Mit der gezeigten Lösung wird somit eine Anbieter-unabhängige Migration von Automatisierungsprojekten möglich. Der Nachteil der mangelhaften Interoperabilität wird abgeschwächt. Es ist nun außerdem möglich, die selbst erstellten Dienste auch Inhouse, auf herkömmlichen Servern oder z.B. auf Private Clouds zu migrieren.

Durch das integrierte Fehlererkennungs- und Fehlerbehandlungs-Konzept werden die Folgen von Verbindungsausfällen abgeschwächt und es wird dem System ermöglicht, so lange in einen sicheren Zustand zu wechseln, bis die Verbindung wieder hergestellt werden kann. Die Umsetzung des Konzeptes des virtuellen Heartbeat-Signals spart außerdem Energie auf den Embedded Tiers, trägt zur Reduzierung des Übertragungsvolumens bei und kann Übertragungskanäle einsparen.

Durch die über das SSF bereit gestellten Basisfunktionen wird darüber hinaus eine umfangreiche Codewiederverwendung erreicht, die es dem Entwickler ermöglicht, sich auf die Implementierung der Anwendungslogik zu konzentrieren. Die schrittweise Abstraktion innerhalb des Frameworks erleichtert die Beherrschung von zunehmender Komplexität, speziell bei wachsendem Projektumfang.

Durch die Verwendung des Frameworks entsteht ein signifikanter Zugewinn an Entwicklungsgeschwindigkeit gegenüber der Entwicklung ohne Framework-Unterstützung. Dies ermöglicht die Senkung von Entwicklungskosten und eine schnellere Überführung von Automatisierungsprojekten aus der Planungsphase in den Produktivbetrieb.

VI. AUSBLICK

Die Anwendung des Smart Systems Framework im Bereich der Regenerativen Energiesysteme eröffnet eine große Vielfalt von neuen Optimierungsansätzen. Speziell im Hinblick auf dezentrale Energienetze und nicht gleichmäßig zur Verfügung stehende Energiequellen ist ein intelligentes Management extrem wichtig. Durch die Integration von Umgebungsmodellen in die Tracking-Algorithmen werden eine Vielzahl von Optimierungsansätzen möglich. So können z.B. topographische Modelle des Geländes, Vegetationsmodelle sowie Modelle der Bebauungssituation integriert und deren komplexe Schattenläufe automatisch mit in die Tracking-Optimierung einbezogen werden. Eine weitere Möglichkeit ist auch die Realisierung einer stufenlosen Leistungsanpassung durch vernetzte Solartracker. Solch ein virtuelles Kraftwerk kann die Energienetze vor Überlastsituationen schützen und somit zur Netz- und Versorgungssicherheit beitragen.

DANKSAGUNG

Diese Arbeit wurde mit Mitteln des Ministeriums für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg gefördert.

LITERATURVERZEICHNIS

- [1] M. Broy, *Cyber-Physical Systems* (acatec DISKUTIERTE), Springer, 2010, ISBN 9783642149016.
- [2] M. Broy, *Cyber-Physical Systems – Innovationsmotor für Mobilität, Gesundheit, Energie und Produktion*, Springer, 2011, ISBN 9783642275661.

- [3] R. Buyya, *Cloud Computing: Principles and Paradigms*, John Wiley & Sons, 2010, ISBN 1118002202, 97811180-02209.
- [4] B. Kölmel, „Cloud Computing und Eingebettete Systeme: System-Entwurf, Realisierung und Bewertung“, *VDI/VDE Automation* 2013, 14. Brachentreffen der Mess- und Automatisierungstechnik, 2013, ISBN 978-3-18-092209-6.
- [5] J. Schlechtendahl, “Communication Mechanisms for Cloud based Machine Controls,” *Procedia CIRP* Volume 17, Variety Management in Manufacturing — Proceedings of the 47th CIRP Conference on Manufacturing Systems, 2014.
- [6] J. Schlechtendahl, “Study of network capability for cloud based control systems,” 24th *International Conference on FAIM*, San Antonio, 20.-23.05.2014, 2014.
- [7] K. C. Laudon, *Wirtschaftsinformatik: eine Einführung*, Pearson Deutschland GmbH, 2010, ISBN 3827373484, 9783827373489.
- [8] M. Fowler, *Patterns of Enterprise Application Architecture*, Pearson Education, 2012, ISBN 9780133065213, ab Seite 19.
- [9] W. Weller, *Automatisierungstechnik im Überblick: Was ist, was kann Automatisierungstechnik?*, Beuth Verlag, 2008, ISBN 3410167609, 9783 410167600.
- [10] E. Lee, *Introduction into Embedded Systems*, UC Berkeley, 2011, ISBN 9780557708574.
- [11] J. Ellenberg, *Preispolitik im Software-as-a-Service Markt*, Diplomica Verlag, 2014, ISBN 3842893973, 9783842893979, ab Seite 2.
- [12] C. Ullenboom, *Java 7 – Mehr als eine Insel*, Addison-Wesley, 2012, ISBN 9783836218023, Abschnitt 13.2.
- [13] J. Webber, *REST in Practice: Hypermedia and Systems Architecture*, O’Reilly Media Inc., 2010, ISBN 0596805829, 9780596805821.
- [14] S. Sumathi, *Fundamentals of Relational Database Management Systems*, Springer Science & Business Media, 2007, ISSN 1860-949X, ab Seite 24.
- [15] Echtzeitbetriebssystem Nut/OS Real-Time Operating System, <http://ethernut.de/software.html>.
- [16] Amazon Webservices (AWS Development Toolkit), <http://aws.amazon.de>.
- [17] D. Lewine, *POSIX Programmers Guide*, O’Reilly Media Inc., 1991, ISBN 9780937175736.
- [18] V. E. Veseley, *Fault Tree Handbook (NUR-EG-0492)*, U.S. Nuclear Regulatory Commission, 1991.
- [19] G. Laußer, *Nagios - Das Praxisbuch: Open Source-Monitoring im Unternehmen*, Pearson GmbH, 2010, ISBN 3827328004, 9783 827328007.
- [20] F. Konrad, *Planung von Photovoltaik-Anlagen: Grundlagen und Projektierung*, Vieweg Verlag, 2008, ISBN 9783834895424.
- [21] K. Heuck, *Elektrische Energieversorgung: Erzeugung, Übertragung und Verteilung elektrischer Energie*, Springer, 2013, ISBN 978383421744.
- [22] JAX-RS-Implementierung Jersey Development Toolkit, <https://jersey.java.net>: Oracle.



Maximilian Engelsberger erhielt den akademischen Grad des Bachelor of Engineering in Technische Informatik im Jahr 2012 von der Hochschule Pforzheim. Derzeit studiert er im Master-studiengang Embedded Systems an der Hochschule Pforzheim und ist wissenschaftlicher Mitarbeiter am Institut für Smart Systems und Services.



Prof. Dr. Thomas Greiner ist wissenschaftlicher Direktor des Instituts für Smart Systems und Services und einer der Sprecher des in Zusammenarbeit mit der Universität Tübingen angebotenen Promotionskollegs „Entwurf und Architektur Eingebetteter Systeme“, Hauptarbeitsgebiete: System- und Softwaremodellierung, Entwurf und Architektur signalverarbeitender eingebetteter Systeme.

MULTI PROJEKT CHIP GRUPPE

Hochschule Aalen

Prof. Dr. Bürkle, (07361) 576-2103
heinz-peter.buerkle@htw-aalen.de

Hochschule Albstadt-Sigmaringen

Prof. Dr. Rieger, (07431) 579-124
rieger@hs-albsig.de

Hochschule Esslingen

Prof. Dr. Lindermeir, (0711) 397-4221
walter.lindermeir@hs-esslingen.de

Hochschule Furtwangen

Prof. Dr. Rülling, (07723) 920-2503
rue@hs-furtwangen.de

Hochschule Heilbronn

Prof. Dr. Gessler, (07940) 1306-184
gessler@hs-heilbronn.de

Hochschule Karlsruhe

Prof. Dr. Koblitz, (0721) 925-2238
rudolf.koblitz@hs-karlsruhe.de

Hochschule Konstanz

Prof. Dr. Schick, (07531) 206-657
cschick@htwg-konstanz.de

Hochschule Mannheim

Prof. Dr. Giehl, (0621) 292-6860
j.giehl@hs-mannheim.de

Hochschule Offenburg

Prof. Dr. Sikora, (0781) 205-416
axel.sikora@hs-offenburg.de

Hochschule Pforzheim

Prof. Dr. Kesel, (07231) 28-6567
frank.kesel@hs-pforzheim.de

Hochschule Ravensburg-Weingarten

Prof. Dr. Siggelkow, (0751) 501-9633
siggelkow@hs-weingarten.de

Hochschule Reutlingen

Prof. Dr. Wicht, (7121) 271-7090
bernhard.wicht@reutlingen-university.de

Hochschule Ulm

Prof. Dipl.-Phys. Forster, (0731) 50-28338
forster@hs-ulm.de

www.mpc.belwue.de