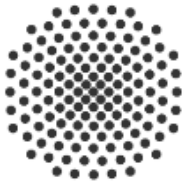
2022



University of Stuttgart
Institute of Robust Power Semiconductor Systems

Seminar “Selected Topics in Power and Microwave Electronics”

PAPER PROCEEDINGS

INSTITUTE OF ROBUST POWER SEMICONDUCTOR SYSTEMS

Seminar Selected Topics on Power and Microwave Electronics

This course, offered by the Institute of Robust Power Semiconductor Systems (ILH) at the University of Stuttgart, deals with current research topics in high data rate communications and wide-bandgap semiconductor technology (e.g. GaN, SiC) for power electronics and millimeter wave applications. In the guided self-study part, students deepen a selected topic of the lecture and prepare a scientific paper in the form of a conference paper and present it in a final presentation.

Students acquire knowledge and experience on writing and presenting a scientific publication through the study and preparation of an individually chosen topic from among a given list of topics in power and microwave electronics.

This document contains the proceedings of the Seminar for the summer term 2022.

On 13th of October 2022, we organized at ILH the festive wrap-up session for the Seminar "Selected Topics in Power and Microwave Engineering" for the summer term 2022. The students Satabdi Bastia and Burak Ozat were distinguished with the best paper awards for their works entitled "Temperature Measurement Methods in WBG Power Semiconductors” and "SiGe BiCMOS - next generation RF node?", respectively! They were selected from 31 participants in the course, based on their submitted paper best originality, value, clarity, and quality ratings. Congratulations to the winners and best of luck to all the seminar participants!

Prof. Dr.-Ing. Ingmar Kallfass and M.Sc. Laura Manoliu with the distinguished students during the best paper award ceremony

Ultra-Wideband Power Amplifier Peaking and Gain Shaping Techniques

Electrical Engineering
Universität Stuttgart
Stuttgart, Germany

***Abstract*— Modern wideband microwave systems often require a flat overall gain response with respect to frequency. This paper illustrates how ultra-wideband power amplifiers help achieve a positive slope gain using gain-equalization techniques to compensate for degradation in the system gain profile due to high-frequency effects. It captures the state-of-the-art analysis and a detailed description of circuit design methodologies used to achieve a positive gain profile. This paper further shows various parameters that are used for gain shape control of the overall system while analyzing various systems designs to achieve a positive slope gain.**

Keywords— peaking amplifier, positive slope shaping, gain shape control

I. INTRODUCTION

With the growth of wireless applications and broadband system applications, the demand for fully integrated, high-power, low-cost power amplifiers is increasing, especially in modern communications, radio links, measuring equipment, and military microwave systems that require a large number of millimeter-wave(mmW)devices. Generally speaking, these systems require low return loss, good gain flatness, and high output power in a wideband range. Distributed amplifiers and feedback amplifiers are frequently utilized to accomplish these objectives. Many factors, including channel losses, the skin effect, temperature changes, and other high-frequency effects, contribute to the degradation of signals in high-frequency applications. These undesirable outcomes cause the gain to drop off with frequency. This negatively affects the gain flatness of the system, thereby impacting the system's dynamic range and performance. To optimize the gain slope, gain slope equalization is preferred. Fig. 1 illustrates the basic principle of gain slope equalization. The equalizer as shown in the figure introduces a positive slope response to a system without an equalizer. This helps achieve a flat system response thus improving system performance.

This brief presents three methods & circuit design techniques used to achieve a desired characteristic gain equalizer implemented on SiGe BiCMOS. [1] presents a system with amplification and filter stages to achieve the maximum slope gain in comparison to other methods in this literature. [2] shows a system with an iNMOS-based tunable resistive gain equalizer and [3] demonstrates a tunable method using transistor-loaded parallel RC circuits.

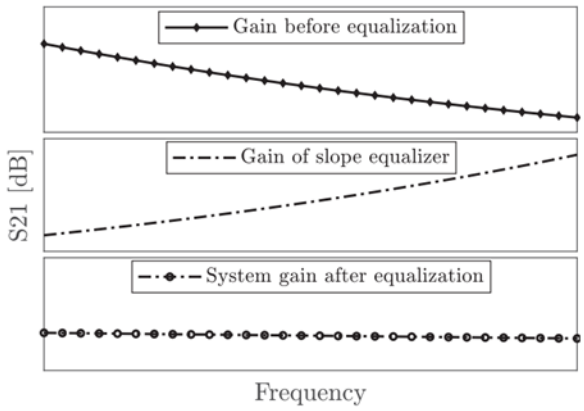


Fig. 1 Gain Slope Equalization [7]

Section II gives a state-of-the-art analysis of the methods presented in [1]. Section III discusses the alternative designs for gain-equalizers as shown in [2] and [3]. Section IV concludes the paper.

II. STATE-OF-THE-ART GAIN SHAPING TECHNIQUE

A wideband active equalizer in 0.25µm SiGe BiCMOS technology is presented in [1], which is packaged using in-house Au-stud bump-based flip-chip bonding. The measured design achieves a positive slope of 2.03 dB/BW over the defined *BW* (5-to-13 GHz).

A. Design Procedure

To accommodate the core-chip blocks' typically negative gain slope of 1 dB/GHz, the active equalizer is divided into two sections. The first section amplifies the signal whose output is connected to a filter unit section, that has a linearly increasing gain. The design procedure starts by selecting an appropriate amplifier topology and a filter architecture to achieve the desired positive gain slope. Fig. 1 presents the schematic of the designed active equalizer.

1) The Cascode Amplification Stage: To mitigate the filter section's loss, the amplifier in this stage should aim for high

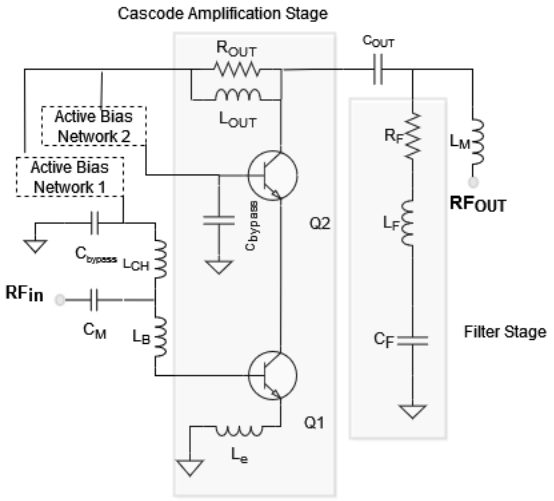


Fig. 2. Topology of state-of-the-art active equalizer circuit [1]

gain and low noise in a wide BW. A cascode topology is preferred over the common-emitter (CE) method for its unconditionally stable high gain. It also lowers reverse transmission and enhances input-output isolation because there isn't a direct coupling from the output to the input[5]. As a result, the Miller effect is removed, leading to a substantially larger bandwidth. It results in increased noise and power dissipation; however, it can suppress the noise of the following stages due to a higher gain level [5]. The optimal device current density for a unit device is picked to achieve maximum gain at the upper end of the desired bandwidth range, i.e., 13 GHz. Generally, it is desirable to search for conditions at which the lowest value of the noise figure can be provided with a maximum associated power gain [6]. This means choosing a base bias for the lowest noise figure (NF) at 13GHz which assists the amplifier in achieving an increasing gain profile over the desired bandwidth. In the second step, the device size is selected to match the circuit to the noise source resistance keeping the current density the same.

Except for the cascode stage's collector, the amplifier biases are supplied via active bias networks. For improved input voltage swing, the base bias of the CE biased stage is supplied by an RF choke inductor, L_{CH}.

2)The Filter Stage: A series RLC network is preferred for the filter topology because it offers greater return loss features and is more compatible with the amplifier transfer characteristic. The desired gain profile and the desired return loss *BW* cannot be achieved when the filter and amplification stages are separated. As a result, they must be designed consistently. Because the design procedure assumes a high output impedance for the cascode amplifier, small device sizes are used. At the output, an LC tank circuit is used, and its quality factor is reduced by $R_L$and the filter resistance R_N, to achieve the desired *BW*. The resonance frequency of the filter determines the component values of the design; explicit analyses are provided in [4]. With less RF feature degradation, the Au-stud bump-based flip-chip bonding can provide increased system integrity.

B. Results

The S-parameters of the equalizer are measured using a 67GHz vector network analyzer. The measured results of the S21 parameter are shown in Fig. 2. It demonstrates the gain performance of the designed equalizer. The design achieves a linearly increasing slope gain with the required *BW* of 5-to-13 GHz. It attains the 10.83 dB gain peak at 12.6 GHz. The

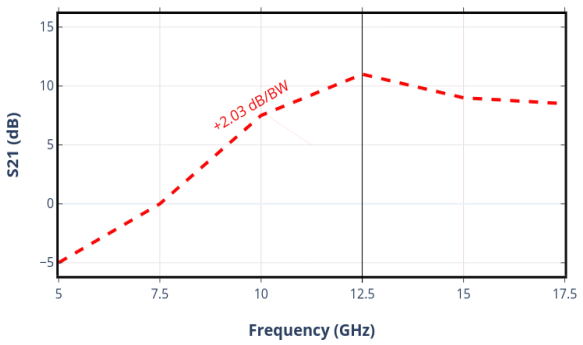


Fig. 3. The measured S21-parameters [1]

measurements are taken up to 18GHz to demonstrate the negligible impact of the applied in-house Au-stud bump bonding on RF performance.

III. Alternative Circuit Design Techniques

A. Tunable SiGe BiCMOS Gain Equalizer based on iNMOS

The gain-equalizer demonstrated in [2] is implemented on a 0.25µm SiGe BiCMOS technology using isolated NMOS (iNMOS) as a variable resistor. To the author's best knowledge, this is the only method to use an iNMOS to obtain a tunable slope across the desired bandwidth.

1) Circuit Description: The conventional gain equalizer topology is modified as shown in Fig. 5(a) to achieve the desired gain profile. The short-circuited quarter-wave transmission line is replaced with a parallel LC-resonator due to their comparable impedance behavior over frequency [2]. The S & Y parameters derived in [2] show that the tunability functionality can be achieved in two ways: by replacing the shunt resistor R_p with a variable resistor or by having L_s and C_s such that they resonate at the same frequency. But, tunable inductors and capacitors consume more area and power. Hence, the first method is preferred and R_p is replaced by an iNMOS-based variable resistor. The on-resistance of an NMOS can be calculated as shown in (1) while $V_{DS} << 2(V_{GS} - V_{TH})$. As observed by the equation, the resistance of the NMOS is inversely proportional to the gate voltage (V_{GS}). This leads to an increase in the gain equalizer slope.

$$R_{on} = \frac{1}{\mu_n C_{OX} \frac{W}{L}(V_{GS} - V_{TH})} \quad (1)$$

Fig. 4(a). shows the cross-sectional of the iNMOS created by a deep n-well & body floating technique. This technique is used to improve the power handling capacity of the variable resistor along with reducing drain-body and source-body capacitances in an NMOS[7]. As shown in Fig. 4(b), the body diodes D_{db} and D_{sb} will turn on at the negative cycle of high power. The current through these diodes will increase smoothly because of

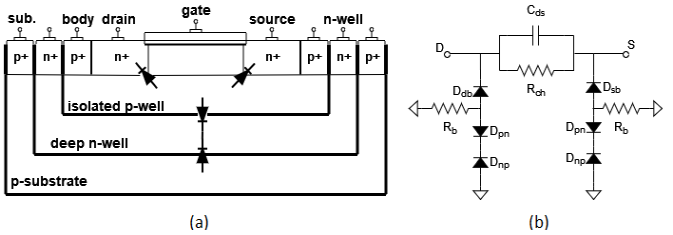


Fig. 4. (a) Cross sectional diagram of an iNMOS [2] (B)Equivalent circuit model of iNMOS with body floating [2]

high R_b and the presence of D_{pn} and D_{np} caused by the deep n-well creation.

2) Results: The designed tunable gain equalizer operates in the 8 to 12.5GHz frequency range as shown by measured results in Fig. 5(b). It has a positive slope of about 1dB/GHz.

B. Tunable SiGe active equalizer using Transistor-Loaded RC Circuits

The proposed equalizer amplifier in [3] realizes both positive and negative gain slopes in the same circuit. This brief will focus on the method used to achieve a positive gain profile within the 0.1 to 0.8GHz range. This design uses parallel RC circuits to achieve this gain profile. Fig. 6 shows the schematic diagram of the active equalizer. To obtain tunable

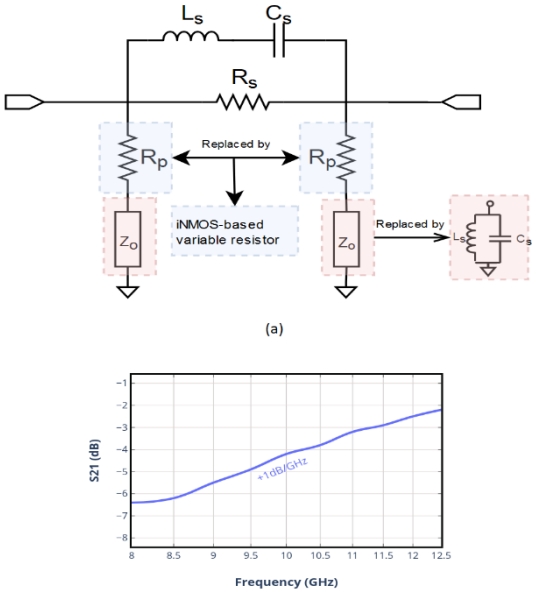


Fig. 5. (a) Conventional and proposed gain-equalizer topology (b) Measured S21 result[2]

positive/negative gain slopes simultaneously, transistor-loaded parallel RC circuits are used in the differential amplifier's load and feedback routes (Z_L and Z_S). The voltage gain G_v is shown in (2)

$$G_V = \frac{R_L}{R_S}\sqrt{\frac{1+(\omega C_S R_S)^2}{1+(\omega C_L R_L)^2}} \quad (2)$$

Z_L and Z_S are assumed to be equal. Therefore, R_L=R_S and C_L=C_S. G_V hence turns into unity. For R_L<R_S and C_L=C_s, the active equalizer outputs a positive slope. The variable resistance is controlled by V_{C1} or V_{C2}. R_1 and L_1 play the role of input impedance matching. The measured gain is plotted in Fig. 7. The measurement was carried out at a supply voltage V_{CC} of 2V with a control voltage V_{C1} of 2-3V, where a positive gain slope is obtained. The active equalizer has achieved a gain slope of +13.2dB/GHz over 0.1 to 0.8GHz.

Table I shows a comparative analysis of the mentioned equalizers in this brief

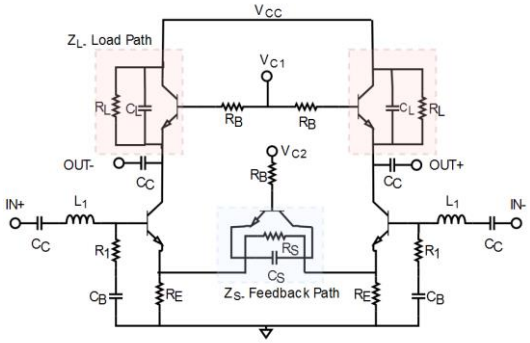


Fig. 6. Schematic Design of Gain Equalizer using Transistor Loaded RC circuits [3]

TABLE I COMPARATIVE ANALYSIS OF MENTIONED GAIN EQUALIZERS

	Technology	***Operating Freq. [GHz]***	***Gain Slope [dB/GHz]***	***Gain [dB]***	***Power diss. [mW]***
[1]	SiGe BiCMOS+ amplification,filter	5-13	2.03	10.83	46
[2]	SiGe BiCMOS + iNMOS based var. resistor	8-12.5	1	-2.2	0
[3]	SiGe BiCMOS + RC Circuits	0.1-0.8	13.2	<+10	n.r.

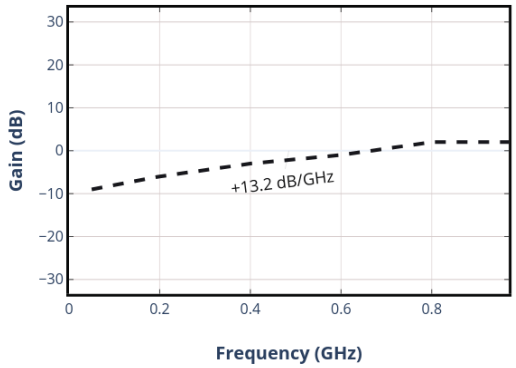


Fig. 7. Measured S21 Parameters [3]

IV. Conclusion

The SiGe BiCMOS active equalizer presented in [1], to the authors' best knowledge is a state-of-art design with wideband and low noise features. It integrates a band-select filter and a cascode amplification stage to achieve a measured positive gain slope of 2.03 dB/GHz with 10.32 as the peak gain. It however has large power dissipation as compared to the other methods. An iNMOS-based variable resistor for a tunable equalizer is proposed in [2]. It shows a positive gain slope of 1dB/GHz within the 8-to-12.5GHz frequency range. The body-floating technique used in this method improves the power handling capability thus minimizing the power dissipation. The equalizer in [3] achieves both negative and positive slope gain of +13.2 dB/GHz over 0.1 to 0.8GHz. This is done using transistor-loaded parallel RC circuits. This design shows equalization at low frequencies making it unsuitable for ultrawideband applications

References

[1] C. Çalışkan, M. H. Mahmud, A. Kasal, M. Yazici and Y. Gurbuz, "Low Noise Wideband Monotonically Raising Gain Active Equalizer in SiGe BiCMOS for Phased Arrays," 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), 2020, pp. 798-801

[2] E. Turkmen, A. Burak, T. Alper Ozkan and Y. Gurbuz, "A Tunable SiGe BiCMOS Gain-Equalizer For X-Band Phased-Array RADAR Applications," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 66, no. 12, pp. 1947-1951, Dec. 2019

[3] Y. Itoh, W. Xiaole and S. Omokawa, "An L-Band SiGe HBT Active Differential Equalizer with Tunable Positive/Negative Gain Slopes Using Transistor-Loaded RC-Circuits," 2018 Asia-Pacific Microwave Conference (APMC), 2018, pp. 708-710

[4] C. Çalışkan, A. Burak, E. Turkmen, M. Yazıcı and Y. Gurbuz, "Active Positive Sloped Equalizer for X-Band SiGe BiCMOS Phased Array Applications," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 66, no. 12, pp. 1952-1956, Dec. 2019

[5] W.. M. L. Kuo, Qingqing Liang, J. D. Cressler and M. A. Mitchell, "An X-band SiGe LNA with 1.36 dB mean noise figure for monolithic phased array transmit/receive radar modules," IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, 2006, 2006, pp. 4 pp.-501.

[6] J. N. Burghartz et al., "RF potential of a 0.18-/spl mu/m CMOS logic technology," International Electron Devices Meeting 1999. Technical Digest (Cat. No.99CH36318), 1999, pp. 853-856.

[7] X. J. Li. And Y.P. Zhang, "Flipping the CMOS Switch", in IEEE Microwave Magazine, vol. 11, no.1, pp. 86-96, Feb. 2010

Analog Pre-Distortion concepts for mmW Amplifiers

Elektrical Engineering
University of Stuttgart
Stuttgart, Germany

Abstract— **In Order to match modern requirements on power Amplifiers for high efficiency and high linearity linearization techniques needs to be applied, as efficiency and linearity of a power amplifier is a contradiction. To achieve higher linearity predistortion can be applied. While digital predistortion is widely used in modern RF Power Amplifiers due to its performance benefits analog predistortion becomes more popular because of its Advantages regarding power consumption which becomes more and more a crucial factor in modern applications because of their handheld or small cell characteristic. This Paper provides an overview and comparison of the most popular analog predistortion concepts and their suitability for millimeter wave amplifiers.**

I. INTRODUCTION

The linearity performance of RF power amplifiers has become one of the most important characteristics of amplifiers as it affects the efficiency, channel density, signal coverage and adjacent channel power ratio (ACPR). Particularly the third order intercept point (TOI) as a figure of merit is used to describe the linearity of power amplifiers. To develop power amplifiers with a high efficiency it is necessary to improve the linear power of an amplifier because with higher power the efficiency also increases. In figure 1.1 the output curve of an amplifier is shown.

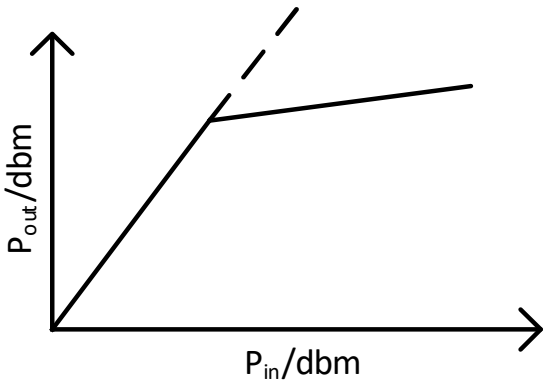


Figure 1.1 Amplifier output curve with gain compression.

The power added efficiency (PAE) can be used as a metric to evaluate the efficiency of an amplifier. In figure 1.2 a typical PAE curve is shown.

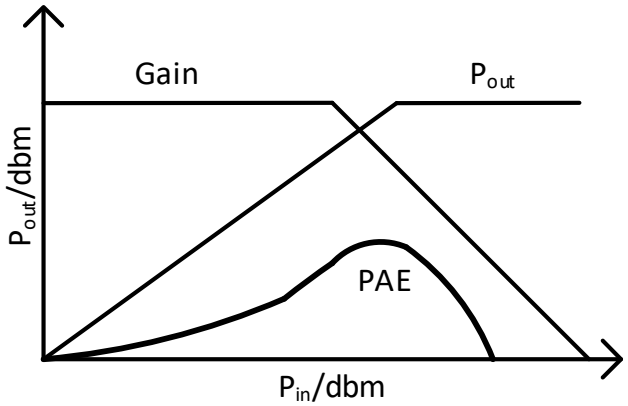


Figure 1.2: A typical PAE curve of an amplifier.

There are different approaches on a system level to address the problem of linearization. One approach is the feedback technique with the disadvantage of a limited correction bandwidth due to delays in the feedback loop. To evade this problem signal predistortion could be applied. There are two categories of predistortion analog and digital. Compared to analog predistortion the digital approach has an outstanding performance advantage but with the never-ending hunger for communication bandwidth the digital approach becomes more complex and less energy efficient. Therefore, the analog approach becomes more and more interesting for modern higher frequency radios.

II. Mathematical point of view

As described in [1] the nonlinear effects of a power amplifier can be modeled with the polynomial expression

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 (1)$$

With the signal x(t) two closely spaced tones are applied to a test amplifier

$$x(t) = A\cos(\omega_1 t) + A\cos(\omega_2 t)$$

Which results in

$$y(t) = a_0 + a_1 A x(t) + a_2 A^2 x^2(t) + a_3 A^3 x^3(t)$$

By illustrating the output signal in frequency domain as shown in Figure 2.1 we can see that the third order intermodulation products (IMD_3) are relatively close to the original tones so they can't be easily eliminated by a band-pass filter. Therefore, it is crucial that the IMD_3 products are minimized.

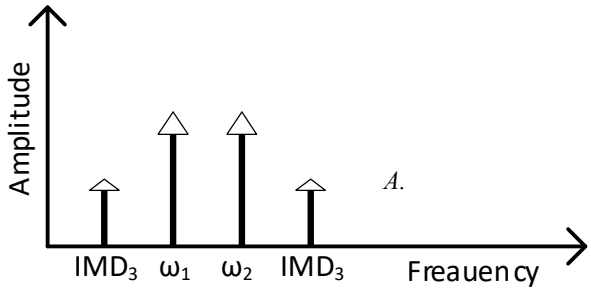


Figure 2.1 Frequency spectrum of the PA output [1].

The aim of every linearization method is to reduce the IMD_3 products by applying a signal in a way that the IMD_3 products are canceled. In figure 2.2 this process is illustrated.

A.

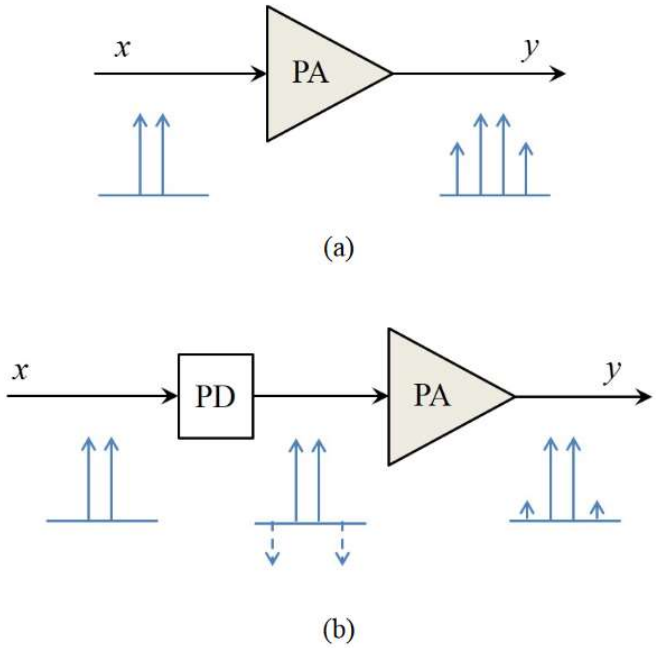


Figure 2.2 (a) PA without PD (b) and with PD [1].

III. Analog predistortion Concepts

Diod based predistortion

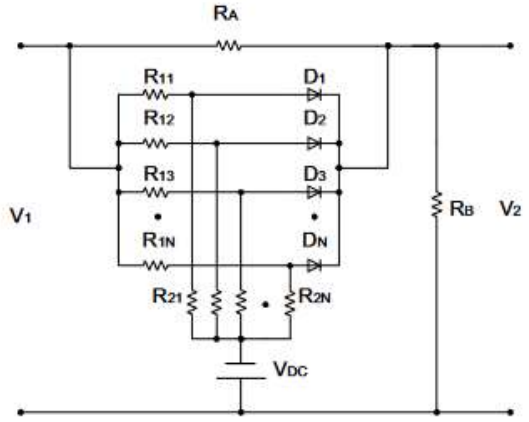


Figure 3.1 Diod based PD circuit [2].

Diode based pre distorter can be used for linearization of a saturating nonlinear amplifier. In [2] such a pre distorter is described for linearization of a GaN power amplifier in digital radio links. The aim of this technique is to curve fit the inverse characteristic of the power amplifier. This is achieved by approximation of the invers characteristic with a certain number of joined straight lines. In Figure 3.1 a example of such a predistortion circuit based on diode is shown. The number of strait lines for the curve fit equals the number of diodes. In [2] the described pre distorter was used to compensate the amplitude-amplitude modulation. The results showed that a

3 db improvement of the overall characteristic cloud be achieved. Improvement in the QPSK an third order intermodulation performance could be achieved for output powers around 40 dBm.

Predistortion by using passive components

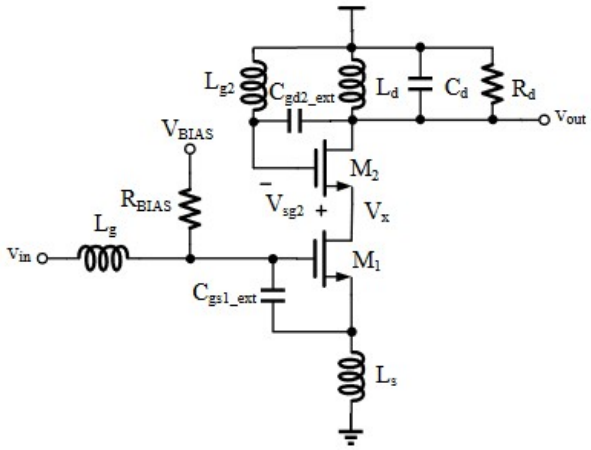


Figure 3.2 A passive predistorted LNA[4].

The passive predistorter linearizes an amplifier by building a inverse funktion of the amplifiers nonlinearity. Therefore the IMD_3 produkts are nullified.
In [4] a 2.4 GHz cascode low noise amplifier LNA with a third order linearity enhancement is described. The enhancement is achieved by including a inductor L_{g2} and a capacitor C_{gd2_ext} to the gate of the cascode transistor of the LNA. An improvement of 4.8 to 11.2 dB in third order linearity is shown.

Variable gain amplifier, phasshifter and logarythmic amplifier

In [1] a system with a Variable Gain Amplifier and a Variable Phase Shifter as a pre distorter for a wireless communication system is presented. In Figure 3.3 the block diagram of this system is illustrated. The log Amps and the phase detector are used to control the VPS and the VGA. This system was designed for LTE applications. A error correction of 5dB at 28 dBm output power could be achieved. This system the VGA is used to compensate the gain error of the PA and the VPS is used to compensate the phase error. The log amplifiers and the phase detector are used to detect the representative error.

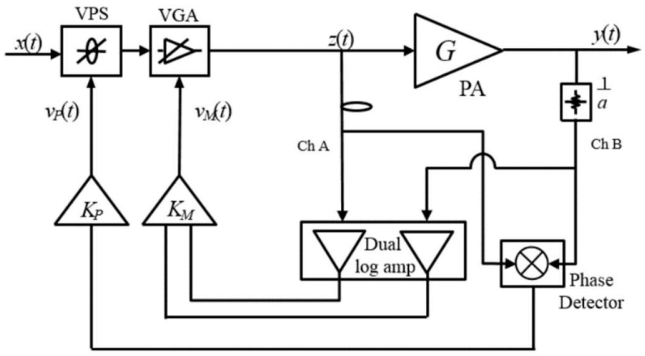
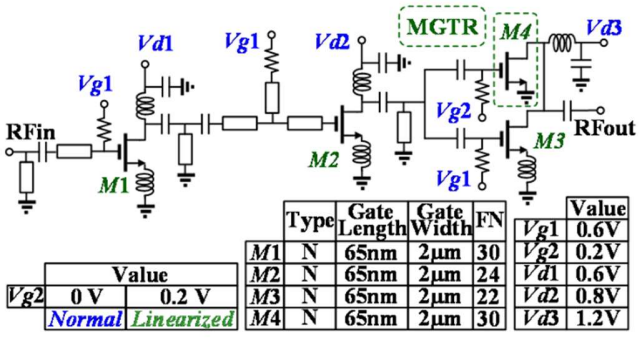


Figure 3.3 Blockdiagram of a VPS and VGA based predistorer [1].

Multi Gate Transistor

In [5] a linearized 35 to 39 GHz CMOS receiver for 5G communication is presented. In figure 3.4 the schematic of the linearized RF amplifier is shown.



Vg2	Value	
	0 V	0.2 V
	Normal	Linearized

	Type	Gate Length	Gate Width	FN
M1	N	65nm	2µm	30
M2	N	65nm	2µm	24
M3	N	65nm	2µm	22
M4	N	65nm	2µm	30

	Value
Vg1	0.6V
Vg2	0.2V
Vd1	0.6V
Vd2	0.8V
Vd3	1.2V

Figure 3.4 Linearized RF amplifier with a MGTR predistorter[5.]

This technique uses a second transistor in parallel to the main transistor with a smaler gate bias voltage to cancel the nonlinearity of the amplifier.
The work shows that the adjustment of the bias of the transistor M4 V_{g2} the IM_3 power decreases between 10 and 16 dB compared to the non linearized operation.

Comparison

The presented techniques improve the linear performance of RF amplifiers but differ in their performance and complexity. Each of the presented techniques are tested with different frequencies and signal bandwidths. Therefor a direct comparison is difficult. The adjacent channel power ration (ACPR) is a figure of merit for communication systems. In [1] a ACPR of -39 dBc is achieved while in [4] a ACPR of -51 dBc could be achieved. It gets obvious that the predistortion technique should be chosen wisely on the application requirements like signal quality, power supply limitations and complexity.

IV. Predistortion in MMW applications

According to [6] modern digital predistortion gets to their limits in wideband applications. Analog predistortion on the other hand has good wideband capabilities causing popularity in modern wireless applications with up to 100 GHz. Especially in mobile or satellite communications where the highly efficient amplifiers are necessary because of power supply limitations. The paper promises an increase in linear power of up to 6dB which is a huge benefit in all applications.

V. conclusion

Analog predistortion offers cost and power saving solution for linearity enhancement in RF amplifier applications. In this paper different analog predistortion concepts for different communication systems and subsystems are presented. Each of the presented concepts has its own advantages and disadvantages regarding performance and complexity. Choosing the right predistortion technique is a key factor for the system improvement.

References

[1] PhdThesis (cho2016analog) Cho, M. Analog predistortion for improvement of RF power amplifier efficiency and linearity Georgia Institute of Technology, Georgia Institute of Technology, 2016 http://hdl.handle.net/1853/556484

[2] M. T. Abuema'atti, A. M. T. Abuelmaatti and T. Yeung, "Predistortion linearization of GaN power amplifier in digital radio links," *2011 Saudi International Electronics, Communications and Photonics Conference (SIECPC)*, 2011, pp. 1-6, doi: 10.1109/SIECPC.2011.5876894.

[3] C. Chang and M. Onabajo, "Linearization of subthreshold low-noise amplifiers," 2013 IEEE International Symposium on Circuits and Systems (ISCAS), Beijing, 2013, pp. 377-380

[4] *Eswaran, U. et al. "Class-E GaAs HBT power amplifier with passive linearization scheme for mobile wireless communications." Turkish Journal of Electrical Engineering and Computer Sciences 22 (2014): 1210-1218.*

[5] C. -N. Chen, Y. Chen, T. -Y. Kuo and H. Wang, "A 35-39 GHz CMOS Linearized Receiver with 2 dBm IIP3 and 16.8 dBm OIP3 for the 5G Systems," *2019 14th European Microwave Integrated Circuits Conference (EuMIC)*, 2019, pp. 92-95, doi: 10.23919/EuMIC.2019.8909517.

[6] A. Katz, M. Chiappetta and R. Dorval, "Predistortion linearization to 100 GHz," *2013 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications*, 2013, pp. 34-36, doi: 10.1109/PAWR.2013.6490180.

Supply Modulation Concepts for Efficiency Enhancement of mmW Amplifiers

dept. Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—This research paper deals with the supply modulation concepts that improve the efficiency of millimeter-wave amplifiers which can be achieved by using load-modulated power amplifier architectures. A digital equalization can be maintained by introducing a time-variable linear distortion of the supply modulator. There are numerous concepts for modulating the supply voltage of millimeter-wave amplifiers. We discuss theoretical analyses of enhancing the efficiency of supply modulation by several PA types. Doherty amplifier, outphasing amplifier, and ET amplifier are three architectures that are most commonly used to increase efficiency. In this paper, the basics of these different architectures are discussed. In each category, a variety of techniques are involved and they are also reviewed in this paper. Furthermore, the advantages and disadvantages of each method is discussed. In addition, the supply modulation concept to improve the efficiency of a back-off power amplifier is discussed.**

***Index Terms*—envelop tracking (ET), Doherty, millimeter wave (mmW), power amplifier (PA), supply modulator, power added efficiency (PAE), peak average power ratio (PAPR), Monolithic Microwave Integrated Circuit (MMIC)**

I. INTRODUCTION

With the development of wireless communication systems, high transmission speed will require a wideband frequency spectrum in the fifth-generation communication systems (5G), especially in millimeter wave (mmW) frequency. The frequency range amounts to the band of the spectrum with a wavelength between 10 millimeters (30 GHz) to 1 millimeter (300 GHz). However, the complexity like 128 QAM or 256 QAM will cause a high peak-to-average power ratio (PAPR) which leads to more challenges in efficiency and linearity. A variety of supply modulation techniques have been developed for mmW amplifiers that allow extending the useable frequency range of supply modulation into the millimeter-wave range, even beyond 100 GHz .Traditional PAs are easier but the supply-modulated PA systems are more complicated. It is not easy to implement supply modulation in a PA, because it involves more challenges. These factors are the main reason to maintain the transmitter PA efficient: (1) efficiency is not constant over output power; (2) highly efficient PAs are nonlinear; and (3) RF load variations can adversely impact efficiency [1]. Signals that result in high-capacity communications, such as currently used LTE, tend to result in high peak-to-average power ratio (PAPR) waveforms the transmitter needs to amplify.

II. ENVELOP TRACKING (ET)

When the supply voltage continuously tracks the inputs envelop the concept is known as envelop tracking. In the past, we evolved bread-broad compounds for power applications ranging from 10-200 W in the RF range below 6 GHz. In recent times, our interest is changed to MMIC solution that targets efficient solutions of satellite communication Ka-band (17-20 GHz) and for 5G-FR2 in the mmW (24-26 GHz).The main advantage of envelop tracking is to enhance the efficiency of a power amplifier which is used in wireless system. To improve power amplifier (PA) efficiency at back-off output power, envelop tracking (ET) architecture is one of the most promising candidates. When ET PA is implemented, the drain voltage is usually generated from the envelope of the RF signal, which has a bandwidth of 4x- 8x greater than modulated bandwidth, which places higher demands on the supply modulator [2]. To reduce the requirements of the supply modulator, several methods of generating bandwidth-limited envelopes have been initiated. One of the methods is to limit the slew rate of the envelope. It is also possible to use the low-pass filters to narrow the envelope‘s bandwidth so that its value is greater than that of the envelope by reducing the amount of residue at the input of the modulator [2]. However, this operation is impractical because it needs several iterations.Due to the envelope bandwidth reduction, there will be additional nonlinearities and memory effects, increasing the requirements for digital predistortion compared to conventional ET PAs. The linearity issue can be resolved, by introducing a dual-input single-output (DISO) behavioral model, so that ET PA with band limited can be achieved without the iteration [2]. This approach minimizes the bandwidth requirement of the supply modulator and sustains the efficiency of ET PA compared with the de-trough envelope [2]. Moreover, the corresponding digital pre-distortion and the bandwidth reduction technique were demonstrated in the test bench, as shown in Fig 1. The modulated supply bandwidth is 100 MHz and the envelope swing range is from 2 V to 6 V [2]. Hence, the digital is also validated to compensate for distortion introduced by band-

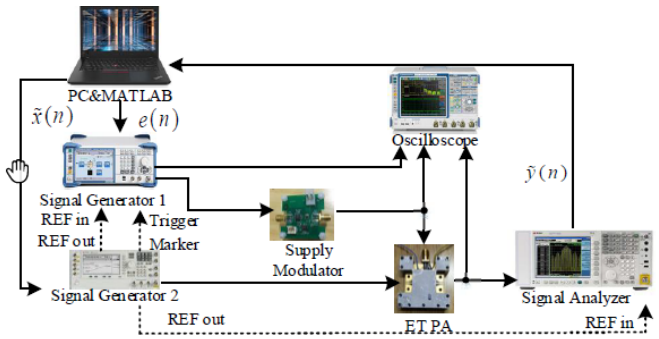


Fig. 1. Test bench set up [9].

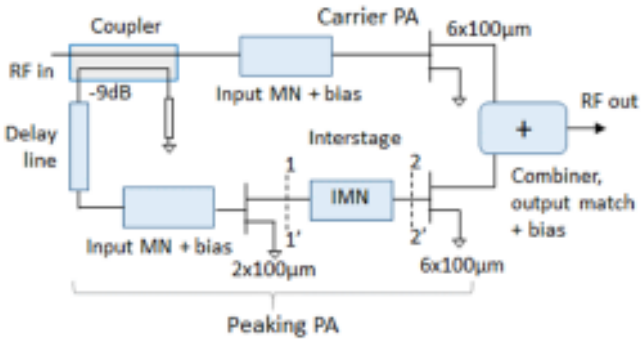


Fig. 2. A block diagram of the Doherty-type load-modulated PA [6].

width reduction, providing a promising solution for wideband mmW ET systems. The bandwidth of the envelope from 2x to 1.2x the modulated bandwidth can be reduced by using this method, but it does not reduce the efficiency of PA [2]. A drawback of the ET is supply output should be very low noise, hence it is difficult maintain. Another disadvantage is higher efficiency needed.

III. DOHERTY AMPLIFIER

Doherty PA are commonly used lower than 6 GHz to improve power-added efficiency (PAE) for communication systems. The black-box design method allows a network to be synthesized from current and voltage behavioral bounds set by the expectations of load-modulated Doherty-type operation. However, the load-modulated Doherty-type MMIC amplifier invented with a black-box method illustrates the first-pass performance [3]. A load-modulated amplifier at 11 GHz, shown in Fig.2 is designed in the 0.25μm Cree Wolfspeed GaN process with 6x100μm slot vias (see Fig.2). It is tested with a continuous wave signal and reaches a maximum output power of 35.5 dBm with 7.6 dB of gain and 38% PAE, and at 7 dB output power back off has 7.4 dB of gain, 32% drain efficiency, and 26% PAE. To overcome potential gain limitations at X-band, the load-modulated Doherty-type PA designed in this work employs an uneven split and additional gain in the peaking path. Numerous Doherty examples are operating in X-band, but recent redistribution and division of commercial spectrum warrant further exploration of frequent scaling in radar and remote sensing applications. In addition of gate-supply modulation has been considered both in the Doherty PAs and also for developing the linearity of ET PAs. The harmonically injected PA is known as another type of PA. The main advantages of this method are able to accomodate signals with higher PAPR. It is also applicable for both lower and higher power amplifier. The disadvantages are this circuit is more complex and it has limited bandwidth.

IV. METHODOLOGY

This paper not only discuss about the three main architectures but also discuss the different methods to enhance the efficiency which can be seen below:

A. Load Modulated Balanced Amplifier (LMBA)

The load modulated balanced amplifier (LMBA) approach was implemented recently to label the bandwidth limitations of Doherty and outphasing techniques, to improve efficiency in back-off. A classically balanced amplifier with wideband input and output quadrature couplers is known as LMBA (see Fig.3) [4]. The terminated isolation port of the output coupler is injected by a control signal, regulating the load extended to the two devices, and generating an active and controllable match [4]. The designed control PA and an input splitter are used to obtain a modulated RF input control signal, which is known as RF-input LMBA. The load modulation has wide bandwidth and efficiency enhancement, hence this architecture is attractive. The modulating supply voltage of the balanced amplifier can be extended by the back-off efficiency of an LMBA. Improved good linearity is obtained in between the range of 9 and 12 percent when discrete supply modulation is cooperating with digital pre-distortion of the wideband RF input LMBA [4].

B. Conventional buck converters

A buck converter is an inductor that acts against the input voltage. The main advantage of the buck-boost converter can produce greater efficiency and lower operating cycles across a wide range of input and output voltages. This type of converter is used in variable applications because it has a better power conversion system that extends the life of the battery and the heating issue can be reduced. Another advantage of these converters is with a small number of external components both step-up and step-down voltages can be achieved. Hence, it provides an adequate solution for smaller external. The disadvantage of the conventional buck converters is the GaN RF power technology is supported in n-type and is not suitable for all types. This isolated driver becomes complex for highspeed drivers, capacitive load is added to the switch node of the supply modulator, hence it reduces the efficiency of conversion. Where the converter operates in the reverse mode of the system, hence this alternative method resolves the previous issue.

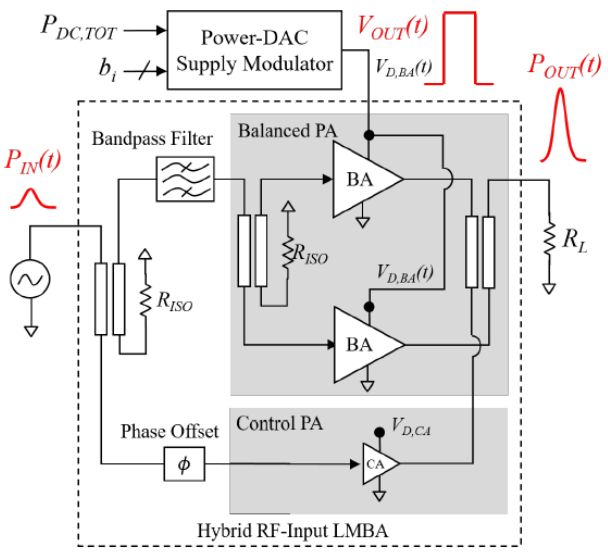


Fig. 3. A block diagram of the load-modulated balanced amplifier (LMBA) [3].

C. The Reverse Buck Converter (RB)

The reverse buck converter (RB) features a lower switching propagation delay, which allows for improving closed-loops control bandwidth. In RF PA, the transistor source is not connected to a system ground but is connected to a floating bias voltage. In such cases, it is hard to design the RF PA using the reverse buck converter. A reverse buck converter topology has been introduced for the foremost in dynamic operation based on a wide-band supply-modulated system. The reverse approach eases the performance of a GaN switch because it operates relative to the ground, hence the isolated switch driver has been eliminated. This causes small parasitic capacitance and smaller delay but needs the use of a floating-ground RF PA. The overall system power-added efficiency exceeds more than 39%, which is a 10.7% points improvement compared to the constant 40 V supply operation [5]. The related bandwidth and power are equivalent to buck converter systems.

V. OUTPHASING AMPLIFIER

Outphasing amplifiers use phase-shift control of multiple saturated or switched-mode branch power amplifiers (PAs) to generate a modulated RF output. By using non-isolating power combining, the branch amplifiers are affected by load modulation, and that results in modulation of the system output. Chrieix is one type of outphasing amplifier, which converts the input signal of the modulated amplitude into differential phase modulation. This modulates and controls the output load and amplitude respectively. The efficiency is achieved at low output powers by using this combiner. The main advantage of discrete supply modulation is the reduced dc power consumption is achieved by the internal PAs. It has two PAs which are the same only in the case of outphasing and it is different in load modulation and the output combiners.

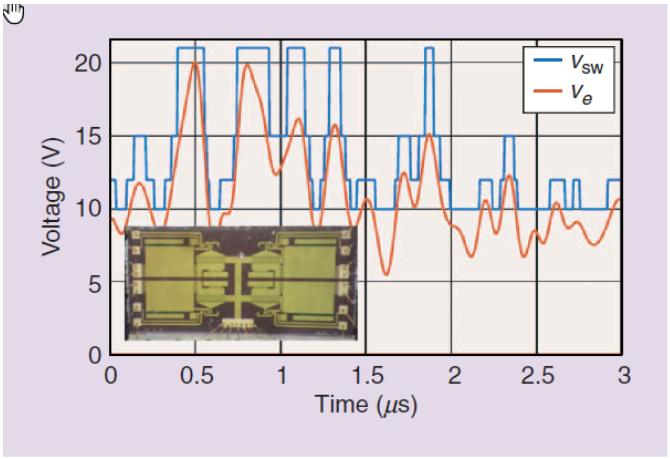


Fig. 4. The output voltage of the four-discrete-level converter is optimized for a 10- MHz LTE envelope signal and (inset) a photo of the GaN die. This converter achieves 97.3% power-stage efficiency [6].

A. Four-level and eight-level supply MMIC

There are two monolithic microwave integrated circuit (MMIC) multilevel discrete tracking supplies that have been illustrated: a four-level supply and an eight-level, three-bit supply. The four-level MMIC reaches higher than 97.3% power-stage efficiency at a 3.5- W average output power level for a 10- MHz LTE envelope signal [6] (see Fig. 4). The eight-level MMIC supply traces an amplitude-modulated radar pulse with a carrier frequency of 10- GHz. The ET continues the high efficiency of both dynamic supply and the PA are tracked by the ET, and modulated amplitude is introduced. The eight-level MMIC integrated with the PA MMIC was also identified. The covered noise signal of a band is limited by 250- MHz and the reduced bandwidth can be traced on four levels, which results in a 20 percent improvement in efficiency.

VI. SUPPLY MODULATION FOR IMPROVING EFFICIENCY

Supply modulation is mainly used for broadband modulated power amplifier systems for modern telecommunications signals with a PAPR. The problem occurred by using supply modulation is the power dissipation in power amplifier with fixed supply voltage (see Fig.5). The efficiency can be improved in Doherty PA by a modulated main amplifier with help of modulated supply. Moreover, this method is used in larger back-off power levels. A Modulation of the supply voltage allows for higher modulation bandwidths. This is known as discrete level supply modulation (class-G) as shown in Fig.6. In this paper, we already discussed both architectures. A drawback of supply modulation is unconstrained to the outputstage drain of a multistage PA. For example, the efficiency can be improved, when the modulation occurred concurrently for both the drivers and final stage PA. In addition, another type of PA being studied is harmonically injected, which operates one PA in class-A mode while the other is operating on the second

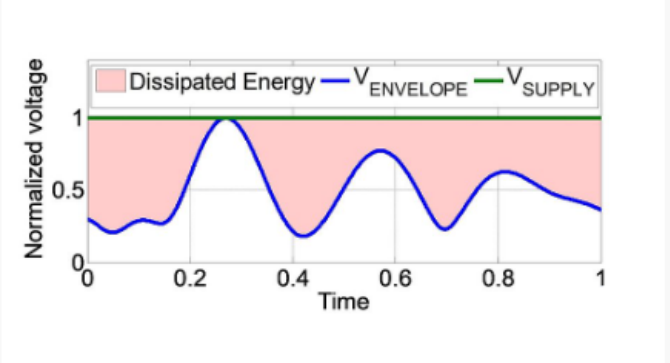


Fig. 5. Power dissipation in power amplifier with fixed supply voltage [7]

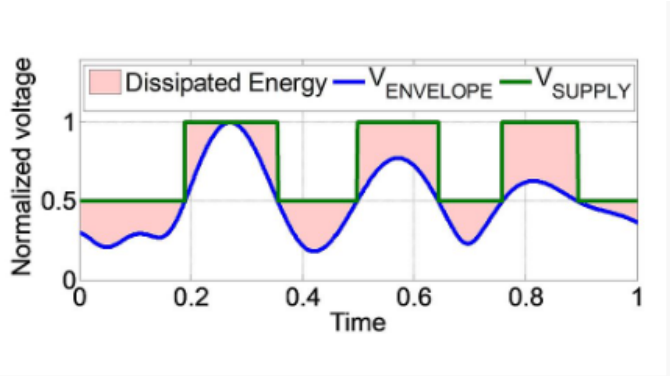


Fig. 6. Power dissipation in power amplifier with two-stage class-G modulation [7]

harmonic. When modulated supply is applied, this type of PA reacts to produce high efficiency.

VII. CONCLUSION

There are a wide variety of methods exist for enhancing the efficiency of mmW amplifiers by using supply modulation concepts. Architectures that use, namely the envelop tracking, the Doherty and outphasing amplifiers. First, the theoretical limits of envelop tracking were discussed. The main superiority is that the performance of ET is greater than other architectures, and also good for a large range of output power. A Doherty can adequately operate with the help of an ET amplifier when the power variation is out of control. Following are the disadvantages of Doherty PA : bandwidth limitation, exact envelop cannot be traced when it has variable ET. An increase in efficiency can be achieved only by reducing the bandwidth. Furthermore, the modulated gate supply has been injected both into the ET PAs to enhance the linearity and also in Doherty PAs. A low-power injected circuit can be used, to cut off the cost. The advantages and disadvantages of all the methods were discussed, with this envelop tracking seems to be more reliable for millimeter-wave frequencies. It is difficult to provide a specific solution for all kinds of approaches with variable frequency ranges, such as this paper is useful to find an appropriate architecture.

REFERENCES

[1] T. Cappello, C. Florian, D. Niessen, R. P. Paganelli, S. Schafer and Z. Popovic, "Efficient X-Band Transmitter With Integrated GaN Power Amplifier and Supply Modulator," in IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 4, pp. 1601-1614, April 2019, doi: 10.1109/TMTT.2019.2898188.

[2] Z. Popovic, "Amping Up the PA for 5G: Efficient GaN Power Amplifiers with Dynamic Supplies," in IEEE Microwave Magazine, vol. 18, no. 3, pp. 137- 149, May 2017, doi: 10.1109/MMM.2017.2664018.

[3] I. Dettmann and M. Berroth, "Supply Modulation of Power Amplifiers with a Class AD Modulator," 2009 German Microwave Conference, 2009, pp. 1-4, doi: 10.1109/GEMIC.2009.4815868.

[4] Z. Popovic, "GaN power amplifiers with supply modulation," 2015 IEEE MTT-S International Microwave Symposium, 2015, pp. 1-4, doi: 10.1109/MWSYM.2015.7167057.

[5] T. Cappello, P. H. Pednekar, C. Florian, Z. Popovic and T. W. Barton, "Supply Modulation of a Broadband Load Modulated Balanced Amplifier," 2018 IEEE/MTT-S International Microwave Symposium - IMS, 2018, pp. 304-307, doi: 10.1109/MWSYM.2018.8439462.

[6] X. Yi et al., "Integrated Bandwidth Reduction and Shaping Technique for Wideband Millimeter Wave Envelope Tracking Power Amplifier with Digital Predistortion," 2020 IEEE MTT-S International Wireless Symposium (IWS), 2020, pp. 1-3, doi: 10.1109/IWS49314.2020.9360052.

[7] S. Paul, N. Wolff, C. Delepaut, V. Valenta, W. Heinrich and O. Bengtsson, "A 14 W wideband supply-modulated system with reverse buck converter and floating-ground rf power amplifier," 2017 IEEE MTT-S International Microwave Symposium (IMS), 2017, pp. 936-939, doi: 10.1109/MWSYM.2017.8058740.

[8] A. Duh, M. Duffy, W. Hallberg, M. Pinto, T. Barton and Z. Po ov ´c, "A 10.8-GHz GaN MMIC Load-Modulated Amplifier," 2019 49th European Microwave Conference (EuMC), 2019, pp. 408-411, doi: 10.23919/EuMC.2019.8910923.

[9] Z. Popovic, "Amping Up the PA for 5G: Efficient GaN Power Amplifiers with Dynamic Supplies," in IEEE Microwave Magazine, vol. 18, no. 3, pp. 137- 149, May 2017, doi: 10.1109/MMM.2017.2664018.

[10] W. Hong, "Thoughts on Millimeter Wave and Terahertz Communications," ZTE Technology Journal, vol. 24, pp. 39-42, Jun 2018. (in Chinese)

GAN Monolithic Bidirectional Switches

Electrical Engineering
University of Stuttgart
Stuttgart

***Abstract*—The structure and construction of monolithic integrated GaN bidirectional switches are discussed. The advantages of using GaN MBDS in the high frequency and high power applications in comparison with existing bidirectional switches made of conventional Si and SiC devices are included. The internal parasitic effects of monolithic integrated devices are studied in detail and an overview on the solutions to eliminate the instabilities in the device are provided. Finally the wide range of applications using the GaN MBDS like matrix converters, AC/AC current source converters, inverters etc, and their benefits are reviewed.**

***Index Terms*—BDS, GAN HEMT, Monolithic Bidirectional Switches**

I. Introduction

The novel multilevel converter topologies like T-type or active neutral point clamped (ANPC) converters and matrix converters consist of higher number of power semiconductor devices and requires bi-directional current control. The usage of conventional devices for bidirectional current control will increase the complexity of the system with increased parasitic effects. The efficient switching operation of such power devices at high frequency with the printed circuit board (PCB) layout optimization is not the best solution for reducing the ringing and overshoot, this would arise the need for monolithic integration of devices. However, application of monolithic integration scheme to bidirectional switches made of conventional semiconductor devices can only minimize the parasitic stray inductance in commutation loops, still the increased conduction losses prevail.

The above challenges in usage of conventional bidirectional switches are overcome by using monolithically integrated bidirectional Gallium Nitride (GaN) switches. The bidirectional GaN high electron mobility transistor (HEMT) provides smaller commutation loops and shows on-state resistance like that of a single unidirectional GaN switch. They can conduct the current in both directions and block the voltage in both polarities, based on the control signals at the gate terminal of the device. The conventional silicon (Si) metal oxide semiconductor field effect transistor (MOSFET) and Si super junction MOSFET (SJMOSFET) are vertical devices, whereas the GaN HEMT have lateral chip architecture. In general, the lateral devices offer more degrees of freedom for monolithic integration, which can be applied for improvising the functionality of GaN HEMT power circuits.

II. Monolithic Integrated Bidirectional GaN HEMT

The regular bidirectional switches comprise of two different unidirectional devices connected in anti-series with common drain-drain or common source-source or common collector-collector configurations [1], [2], [3]. The major drawback of this implementation is that, the series connection of two power devices doubles the on-state resistance, thereby resulting in higher conduction losses compared to that of unidirectional power semiconductor devices. In addition, the increased number of device count causes more commutation loops and thus resulting in increased parasitic stray inductance in the commutation loop. The identified monolithic integration of GaN HEMT overcomes the stated disadvantages. The architecture of GaN MBDS as in Fig. 1 (right) consists of two gates G_a, G_b and additional source S_b which replaces the drain D in unidirectional GaN HEMT as in Fig. 1 (left). The driving scheme and external circuit are made similar to that of two unidirectional GaN with common drain-drain node. As they share a common drift zone, the length between two gates of bidirectional device and length between the gate and drain of unidirectional device are the same. Thus, the on-state resistance of monolithic bidirectional switches(MBDS) GaN are very less compared to the conventional series connected two unidirectional transistors.

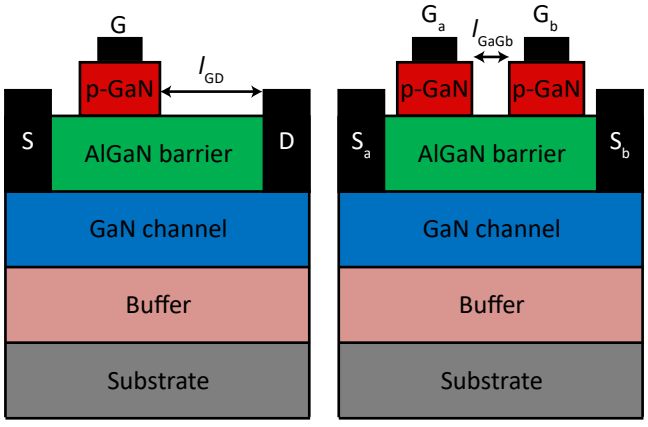


Fig. 1. Cross section of unidirectional GaN HEMT and GaN MBDS (Adapted from [4])

The advantages of GaN HEMT in comparison with discrete Si devices are high breakdown voltage due to wide band gap

and can be used for applications with high switching frequency which results in reduction of size of passive components. With low gate charge and very low area specific on-state resistance, GaN HEMT are capable of high current density with low switching losses. As stated in [6], the GaN HEMT shows ten times higher switching speed in comparison with the silicon devices. Also, GaN devices offer high efficiency and high power density in the power electronic converters. Regarding the application of GaN MBDS in the converters, they have significantly smaller commutation loop and the parasitic inductance as well related as stray inductance is very low. Also, they have on-state resistance same as that of unidirectional GaN transistors.

III. Parasitic Effects in GaN MBDS

Though the monolithic integration of devices eliminates the parasitic inductance in the commutation loop, still there are different parasitic effects at the device level and here we have an overview on two important parasitic effects namely, dynamic on-state resistance (R_{DSON}) and impacts of substrate termination on the device.

A. Dynamic On-state Resistance

In comparison with the series connection of two separate unidirectional GaN switches, the bidirectional GaN switches offer nearly 48% reduced on-state resistance R_{DSON} [4]. However in the dynamic condition, while switching the device at high frequency and high blocking voltage, R_{DSON} increases by multiple factors. There are different factors that causes impact on R_{DSON} which are different off-state times, off-state voltage at different temperatures, drain current at different temperatures and different switching frequencies. All these factors are studied in detail by carrying out various experiments [7], using three different devices namely normally off p-GaN gate GaN HEMT, normally on GaN and super junction MOSFET IPP65R125C7 ($650\,\mathrm{V}, 15\,\mathrm{A}$). The following are the important notable facts for normally off GaN HEMT, on the investigation of applied time of off-state voltage before turning on the device the dynamic R_{DSON} decrease with on-state time, it has been observed that the dynamic R_{DSON} increases with off-state drain bias whereas in both the case the R_{DSON} of super junction MOSFET shows very less deviation [7]. For measurement of R_{DSON} transient for various on-state drain current, with the increase in drain current R_{DSON} increases [7]. While various switching frequencies are applied from $50\,\mathrm{kHz}$ to $400\,\mathrm{kHz}$, it is observed that higher switching frequency and longer pulse duration can cause the on-state resistance to increase with $30-40\%$ higher at $400\,\mathrm{kHz}$ compared to switching at $50\,\mathrm{kHz}$, whereas for SJMOSFET only negligible increase is noted.

With the well-known double pulse test, it is noted that, switching the device from high-bias off-state causes an increase in on-state resistance of GaN switches. This phenomenon can be explained by the trapping effects, however, the impact of conduction loss in the on-state does not influence the significant temperature increase, there is another effect that occurs during the switching event which is the cumulative trapping of hot channel electrons that can also be a reason for the R_{DSON} behaviour. The electrons trapping [10] in different layers of GaN HEMT namely, 2DEG interface, AlGaN barrier layer or GaN buffer layer, can also influence the increase in R_{DSON}. Two different states exists in trapping effects, one is the deep-acceptor state [10] which captures the carriers while switching on the device, this is responsible for the increase in R_{DSON} in the initial phase of R_{DSON} transient and other is the donor state [10] that causes decrease in R_{DSON}. The latter captures the carrier from the source or gate-leakage current during high gate-drain voltage off-state. As the temperature increases, more electrons from gate or drain leakage current can occupy the trap states during off-state bias leading to an immediate R_{DSON} increase after being switched. The dynamic R_{DSON} can be reduced significantly by using double heterostructure epitaxy without any doping or using an iron doped single structure [10].

The hot electrons are the gate leakage electrons in the off state, they are trapped in the buffer layer or the surface, which reduces the channel carrier density resulting in the increase of R_{DSON}. From the experiment performed in [9] using p-GaN gate AlGAN/GaN HEMT, the results show that the major contributing factor for current collapse is high power states during switching with different rise time and fall time with constant off time. Also, from the accurate measurement results of R_{DSON} from [9], it has been observed that the module without field plates have R_{DSON} increased by 78% whereas, with the field plate the R_{DSON} can be increased by only 1.8 times. The reduction of current collapse while using the field plate indicates that the trapping of hot electrons occurs near the surface and not at the buffer. The prevention of hot electrons from moving over to AlGaN barrier and getting trapped at the gate edge is done by adding of field plate, which makes a vertical electric field component to exist between the field plate and channel in opposite direction.

B. Substrate Termination

1) Substrate loop: The monolithic integration of the gate driver circuit and half bridge circuit as in [12], results in the reduction of the parasitic inductance in commutation loop and gate loop drastically. Still, instabilities are observed in the device performance which would be due to the third loop, namely substrate loop, where the substrate must be connected to the source externally in the above monolithic integrated system. The parasitic effect occurring in the substrate loop is comparatively low in the individually packed devices where the source and substrate are internally shorted. This parasitic substrate-loop inductance [11] must be reduced for stable operation of the circuit without any steady-state high voltage oscillations. In order to evaluate effect of substrate loop impedance the experiment has been carried out in [12] and an interesting fact has been identified that there is no intentional damping of substrate-loop on packaging level, instead the resistance of the power transistor or indirectly coupled gate resistance dampens the oscillations in substrate. The chosen external gate resistor not only optimises the gate loop but also the least

value ensures the stable operation with non-zero substrate-loop impedance. Therefore, as a countermeasure, instead of external source to substrate termination, a resistor R_{BS} of low value can be added in the substrate loop. Reference [12] also shows that by increasing R_{BS}, the oscillations in substrate loop can be damped without altering the gate resistance. Despite the merit, there is another drawback in adding the R_{BS}, that is they increase the switching losses to circuit as like RC snubber, which would be major concern for soft switching topologies.

2) Methods of substrate termination: Despite the monolithically integrated bidirectional GaN HEMT consisting of two sub-switches constructed on the same substrate, the performance of bidirectional switches should match with the unidirectional switch with a guarantee on stable switching and limited R_{DSON} degradation due to back-gating effects. Reference [13] studies about both active and passive substrate termination, the latter one is with the substrate being left floating or being fixed to one of the two source nodes and former one comprises of auxiliary switches for controlling the substrate potential with increased complexity and giving additional parasitic impedance. Fig. 2 comprises of GaN MBDS with gates T_a, T_b and auxiliary switches A_a and A_b introduce on-state resistance in the substrate loop, which will be the main contributing factor for R_{BS}, and also these switches increases the chip size without increasing the current capability of transistors. In [13] four different systems are

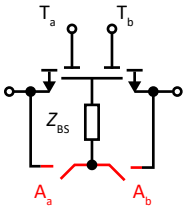


Fig. 2. Bidirectional GaN HEMT with auxillary switches (Adapted from [13])

undergone experiments, which uses bidirectional and unidirectional switches with two systems each, a double pulse test and continuous pulsed operation test with resistive substrate termination are carried out. With the passive termination, it has been observed that, by connecting the substrate to negative DC potential, there is a dynamic increase of R_{DSON} to a limited extent, by leaving the substrate to float, there is prominent degradation of R_{DSON}, and by connecting the substrate to drain potential as in unidirectional transistor, there is tremendous increase in R_{DSON}. However, by connecting the substrate to any of the sources would have controlled R_{DSON} in one direction and severe degradation in other direction, whereas using a floating substrate have symmetrical behaviour which can be used in applications where efficiency is not the main concern.

An external resistor is connected between source and substrate, in order to emulate the on-state resistance of auxiliary switches, the test has been carried out for various values of R_{BS} and observed that the Bidirectional switches have minimum R_{DSON} for R_{BS}= 2 Ω [13], this optimum value independent on blocking voltage and load current. On comparing the results from double pulse test and continuous pulsed operation, the optimum range for R_{BS} for the latter is slightly higher than the former. Using the optimized value of termination resistance at the substrate, around 7% of R_{DSON} is reduced which thereby reduced the conduction losses noticeably. Thereby, the on-state characteristics of both unidirectional and bidirectional HEMT are influenced by the substrate termination.

IV. State of Art Converters

One of the typical applications is the matrix converters, which directly converts the AC to AC with significant reduction of operation loss compared to that of conventional AC/DC/AC converters. These matrix converters achieve high

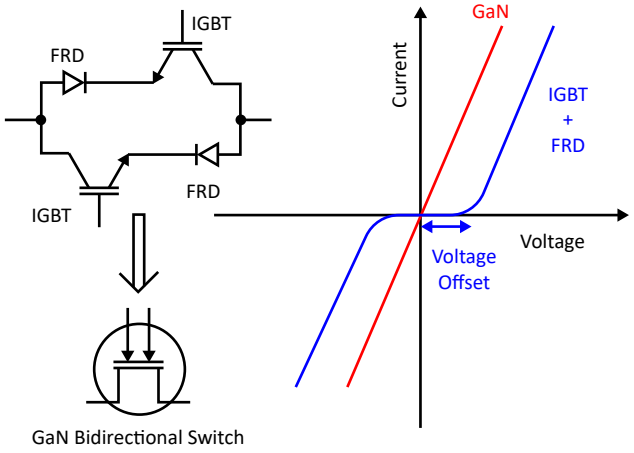


Fig. 3. Characteristics and configuration of GaN MBDS and IGBT-based bidirectional switches (Adapted from [14])

power conversion efficiency without the power factor control circuits. The important component of this matrix converter is the bidirectional switches, consisting of series or anti-parallel connection of two Insulated Gate Bipolar Transistors (IGBTs) and Fast recovery diodes (FRDs) as like in Fig.3, these conventional bidirectional switches are larger and provide more conduction losses. The integrated technology of GaN is capable of fast switching with very less parasitic inductance. In [15] a very compact matrix converter is realized in a size of 25x18 mm^2 using a GaN integrated chip and this size is equivalent to about one-thousandth of the existing three phase matrix converter which is about 30x50 cm^2. As the GaN bidirectional switches doesn't depend on the voltage offsets for conducting the current, the proposed 3x3 matrix converters [14] composed of GaN switches are in compact size providing highly efficient conversion of AC to AC with 98% efficiency at 1 kW output power which is very high in comparison with IGBT based matrix converter. The used GaN switch [14] is made of a recessed gate constructed by p-GaN/i-AlGaN regrowth technique, which reduces the on-state resistance. The results from [14] shows that the operation loss of GaN-based matrix converter is reduced by half in comparison to IGBT based matrix converter, in which the reason for the drastic reduction of conduction loss is the reduced on-state resistance

and elimination of forward voltage offset. The used GaN switches can carry current up to 100 A and the system can achieve high output power in range of 10 kW.

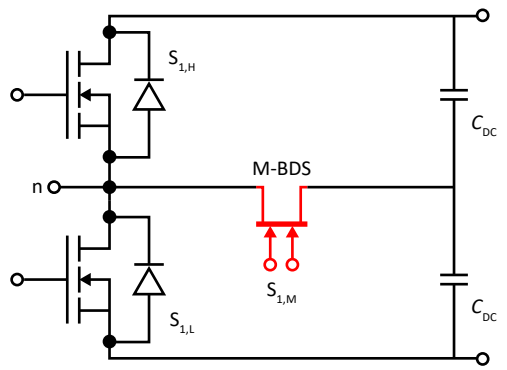


Fig. 4. Construction of TT bridge-leg with 600 V GaN-MBDS and two 1200 V SiC MOSFETs(Adapted from [16])

Another notable application is the T-Type (TT) power factor correction (PFC) rectifier and inverter system. In T-Type PFC rectifier and inverter systems, T-Type bridge legs [16] extends the half bridge to possibly connect switch node to DC link midpoint and this realizes three output voltage levels, which requires to block the voltage in both polarities and conduct the current in both directions. In [16] the system makes use of 600 V/140 mΩ GaN MBDS and the performance has been evaluated by carrying out the calorimetric measurement of TT bridge-leg of a system. This TT bridge-leg comprises of Silicon Carbide (SiC) MOSFETs connected in half bridge configuration and GaN MBDS connected to DC link midpoint Fig. 4. These switches are paired with similar R_{DSON} and calorimetric measurement is carried out for the commutation cell. The proposed system [16] could achieve more than 99% for power density up to 15 W/l.

V. Benefits of MBDS in AC-AC Current Source Converter in Comparison to Voltage Source Inverter

The voltage source inverter is challenging because of the complex current ripple waveform, this inverter is made of GaN based three level ANPC voltage source converters (VSC) [8]. In comparison with SiC MOSFETS, the GaN enable these multilevel inverters to switch faster thereby reducing the size of reactive components. Despite the advantages, the high frequency switching is in the range where the electromagnetic interference (EMI) standard is applied. The studies in [8] make use of 650 V, 60 A GaN-FETs, the major contributor to the switching loss is the output capacitive switching loss. The GaN FET is operated at 200 kHz, where the converter side inductor size is reduced but the filter has been changed to two stages instead of a single stage.

The AC-AC current source converters are the alternative for voltage source converters, since they provide output voltage continuously with very less effort for filters. And this converter removes some issues like high frequency motor losses, the requirement for expensive shielded cables, long motor cables and, common-mode currents which damages the motor bearings. The usage of wide band gap devices decreases the switching frequency and increases the efficiency by using two-third pulse width modulation for current source rectifier (CSR) or current source inverters (CSI). In [5] the proposed system makes use of 600 V, 140 mΩ GaN MBDS, this completely removes the structural drawback of current source converters (CSC). It has been observed that operating a CSR or CSI in 2/3 PWM can cause a reduction of switching losses by 33%. The commutation cell consists of three MBDS where two of the switches in the half bridge undergo active switching whereas the switch connected to DC link midpoint undergoes passive switching. They carry out the calorimetric measurement for CSC commutation cell which includes also the losses from the passive switch. The impact of losses from passive switch is added to the switch that has been hard turned-on and efficiency of 98% is achieved for operating at switching frequency of 96 kHz for driving a 200 V motor at 1.5 kW nominal power, using the GaN MBDS and proposed synergetic control of AC-AC CSC as shown in Fig. 5.

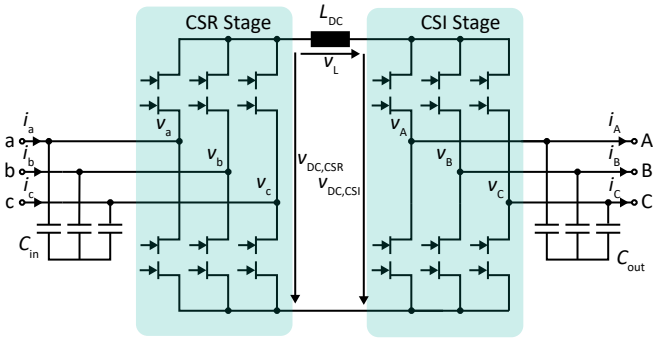


Fig. 5. AC-AC CSC with filter capacitors on AC-side (Adapted from [8])

VI. Conclusion

The bidirectional switches composed of conventional Si semiconductor devices like IGBTS and FRDs would result in increased device count, increased gate loop complexity and increased conduction losses. The GaN MBDS offers low R_{DSON}, higher switching frequency, higher power density and higher efficiency in comparison with conventional bidirectional switches. Despite the advantages of GaN MBDS, the parasitic effects like dynamic R_{DSON} and substrate termination within the monolithically integrated device would be the major challenge to be faced to develop a stable device. The dynamic R_{DSON} being dependent on various operating conditions, R_{DSON} can be reduced by using double heterostructure epitaxy or field plates. Also terminating the substrate loop with suitable R_{BS}, the oscillations in substrate loop can be damped, as well as R_{DSON} is optimised to minimum. If one such stable GaN MBDS is used as a bidirectional switch in any power electronic converters, it would have higher efficiency in their category and also be compact in size.

REFERENCES

[1] E. Gurpinar and A. Castellazzi, "Single-Phase T-Type Inverter Performance Benchmark Using Si IGBTs, SiC MOSFETs, and GaN HEMTs," in IEEE Transactions on Power Electronics, vol. 31, no. 10, pp. 7148-7160, Oct. 2016, doi: 10.1109/TPEL.2015.2506400.

[2] Semikron, Datasheet - SKiiP 28TMLI12F4V1," 2015.

[3] EPC – Efficient Power Conversion Corporation, eGaN® FET Datasheet EPC2110 – Dual Common-Source-Mode," 2017.

[4] C. Kuring, O. Hilt, J. Böcker, M. Wolf, S. Dieckerhoff and J. Würfl, "Novel monolithically integrated bidirectional GaN HEMT," 2018 IEEE Energy Conversion Congress and Exposition (ECCE), 2018, pp. 876-883, doi: 10.1109/ECCE.2018.8557741.

[5] A. Kouchaki and M. Nymand, "Filter design for active neutral point clamped voltage source converter using high frequency GaN-FETs," 2017 IEEE 12th International Conference on Power Electronics and Drive Systems (PEDS), 2017, pp. 349-354, doi: 10.1109/PEDS.2017.8289170.

[6] I. Omura, W. Saito, T. Domon and K. Tsuda, "Gallium Nitride power HEMT for high switching frequency power electronics," 2007 International Workshop on Physics of Semiconductor Devices, 2007, pp. 781-786, doi: 10.1109/IWPSD.2007.4472634.

[7] N. Badawi, O. Hilt, E. Behat-Treidel, J. Böcker, J. Würfl and S. Dieckerhoff, "Investigation of the dynamic on-state resistance of 600V normally-off and normally-on GaN HEMTs," 2015 IEEE Energy Conversion Congress and Exposition (ECCE), 2015, pp. 913-919, doi: 10.1109/ECCE.2015.7309785.

[8] N. Nain, D. Zhang, J. Huber, J. W. Kolar, K. Kin Leong and B. Pandya, "Synergetic Control of Three-Phase AC-AC Current-Source Converter Employing Monolithic Bidirectional 600 V GaN Transistors," 2021 IEEE 22nd Workshop on Control and Modelling of Power Electronics (COMPEL), 2021, pp. 1-8, doi: 10.1109/COMPEL52922.2021.9646006.

[9] I. Hwang et al., "Impact of Channel Hot Electrons on Current Collapse in AlGaN/GaN HEMTs," in IEEE Electron Device Letters, vol. 34, no. 12, pp. 1494-1496, Dec. 2013, doi: 10.1109/LED.2013.2286173.

[10] GaN-based power HEMTs: Parasitic, Reliability and high field issues

[11] S. Moench et al., "Effect of substrate termination on switching loss and switching time using 600 V GaN-on-Si HEMTs with integrated gate driver in half-bridges," 2017 IEEE 5th Workshop on Wide Bandgap Power Devices and Applications (WiPDA), 2017, pp. 257-264, doi: 10.1109/WiPDA.2017.8170557.

[12] S. Moench et al., "Instabilities by Parasitic Substrate-Loop of GaN-on-Si HEMTs in Half-Bridges," 2018 IEEE 6th Workshop on Wide Bandgap Power Devices and Applications (WiPDA), 2018, pp. 242-246, doi: 10.1109/WiPDA.2018.8569066.

[13] C. Kuring et al., "Impact of Substrate Termination on Dynamic On-State Characteristics of a Normally-off Monolithically Integrated Bidirectional GaN HEMT," 2019 IEEE Energy Conversion Congress and Exposition (ECCE), 2019, pp. 824-831, doi: 10.1109/ECCE.2019.8912793.

[14] H. Umeda et al., "High power 3-phase to 3-phase matrix converter using dual-gate GaN bidirectional switches," 2018 IEEE Applied Power Electronics Conference and Exposition (APEC), 2018, pp. 894-897, doi: 10.1109/APEC.2018.8341119.

[15] 30.5 A GaN 3x3matrix converter chipset with Drive-by-Microwave technologies

[16] F. Vollmaier, N. Nain, J. Huber, J. W. Kolar, K. K. Leong and B. Pandya, "Performance Evaluation of Future T-Type PFC Rectifier and Inverter Systems with Monolithic Bidirectional 600 V GaN Switches," 2021 IEEE Energy Conversion Congress and Exposition (ECCE), 2021, pp. 5297-5304, doi: 10.1109/ECCE47101.2021.9595422.

Parallelization of GaN-Transistors

Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—Gallium nitride (GaN) based power devices are rapidly being adopted due to their ability to operate at frequencies and switching speeds beyond the capability of silicon (Si) power transistors. Using GaN high electron mobility transistors (HEMTs) in high current applications, such as pulsed power modulators for particle accelerator systems, requires the parallelization of multiple devices. In order to achieve a dynamically balanced current distribution between the parallel devices, synchronized gate voltages are crucial. In this paper, we will discuss different topologies of GaN transistor parallelization in terms of complexity, efficiency and the effective output voltage and a multi-phase converter circuit using GaN. Furthermore, we will discuss a parallelization design, which has the best outcome among all the listed methods.**

***Index Terms*—GaN, parallelization, multi-phase.**

I. Introduction

GaN devices are becoming extensively popular in the modern world due to characteristics like low switching losses, low drain to source on resistance (R_{DSon}), low parasitic capacitances, high critical field strength and absence of reverse recovery effect. This is ideal for hard-switched, high current and low voltage converters for electro-mobility and hybrid transformers. One of the main challenges in using GaN HEMT is the parallelization concept to increase the current in the circuit and to reduce the conduction losses in the process. These are mainly required in low voltage applications for which GaN power modules are not available in market yet. The parasitics in the gate loop causes severe damages to the circuit while paralleling. Current sharing must be symmetrical in all sense during the switching transient. In this paper, we will be going over different concepts and topologies in GaN parallelization. In section II, we go through different possible topologies of GaN parallelization. Section III shows the current state of the art in parallelization. A very detailed design of each aspect of a parallel GaN circuit is explained in section IV. Section V deals with a thermal simulation performed on a four GaN paralleled system. We discuss various limitations and remedies of parallelization of GaN in section VI. Finally, we conclude the paper by analyzing an example of the multi-phase operation of GaN using a buck converter circuit in multi-phase configuration.

II. Topologies of GaN Parallelization

A half-bridge circuit consists of two switches, in this case two GaN transistors, one on the high side (HS) and the other one on the low side (LS) of the load. In general, four different arrangements for paralleling half-bridges are possible[3]. Half-bridge drivers with one channel for the upper and lower switch are considered in this paper to allow dead times in the range of 10 ns. The classification is done based on the gate driver connection and the load inductor connection.

A. Multi-phase topology

Fig. 1 shows two configurations with individual inductor load. In the first configuration, each HS switch and each LS switch in the parallel connection has a separate gate driver circuit. In the second configuration, all HS switches share a common gate driver circuit and all the LS switches share another gate driver circuit. Individual output inductors for each half-bridge eliminate the problem of current sharing during switching transients as long as current sharing is achieved in steady state. The advantage of having a common gate driver circuit is that the half bridges currents and voltages can be controlled synchronously. However, the large number of connectors and inductors increases the mechanical complexity for a high number of parallel half-bridges

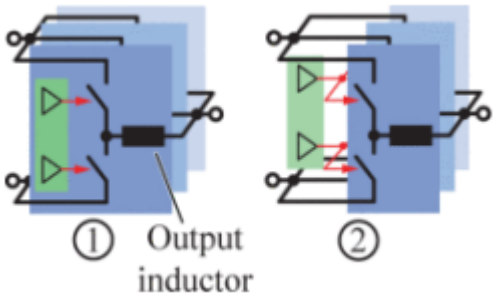


Fig. 1. Paralleling half-bridges with multiple load in multi-phase configuration ① with separate gate drivers ② with common gate drivers[3]

B. Common load topology

All the half bridges are connected to a common output inductor. If the half-bridges share a common output inductor, the gates of all parallel half-bridges must be controlled synchronously to enable an equal current sharing during switching transients. Due to component tolerances and jitters, this can only be ensured with a common gate driver (Fig 2) since gate

driver integrated circuits (ICs) with output synchronization are not available yet. Consequently, Fig. 2.4 is the preferable arrangement for paralleling GaN half-bridges with currently available devices.

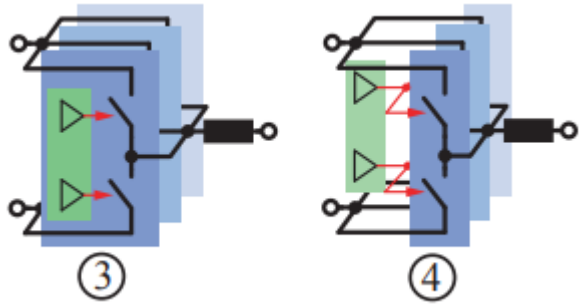


Fig. 2. Paralleling half-bridges with common load topology ③ with separate gate drivers ④ with common gate drivers[3]

III. State of the Art

The fourth topology shown in fig. 2.4 is the most ideal case for parallelization and is now utilized as the state of the art concept. The currently available and efficient technology of GaN parallelization is the GaN PX package which has been developed by GaN Systems. In [3], we discuss the detailed design of the circuit based on the above topology.This package greatly improves the paralleling performance and stability compared to the tradition GaN package due to the following reasons. The PX package has very low source inductance when compared to the traditional package and immensely improves the paralleling performance. The PX package provides dual gate pads for easier layout and reduces the total size of gate drive loop in the design and creates a symmetry in the loop for parallelization[6]. Furthermore, the package has positive R_{DSon} coefficient, thermally stable gate threshold voltage and negative temperature coefficient of trans-conductance.

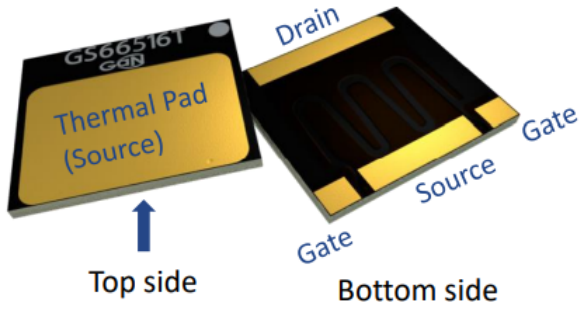


Fig. 3. $GaN_{PX}T$ package GS66516T (650V/25mΩ)[6]

IV. PCB Layout Design for Paralleling GaN

A. Gate loop design

The main aim of this design is to minimize the total parasitic inductance and to divide the existing parasitics equally between each paralleled GaN HEMT. This ensures equal distribution of the gate voltage among the paralleled GaN HEMTs. This comes into relevance especially during fast switching cases. Star connected gate traces can be used for achieving equal length and hence approximately equal parasitics from the driver to gate pads of each GaN HEMT of the parallel design. The dual gate driver pins are internally connected that only one needs to be supplied with the gate input voltage. For example as shown in Fig. 4, only two gate loops are required to drive a 4-GaN paralleled circuit. The Kelvin Source (KS) pins reduce the coupling between power loop and gate loop. Now in-order to minimize the gate loop inductance, we increase the trace width of the current loops with adjacent current conducting layers. Due to this, we will be able to use a small external damping resistance and therefore achieve faster switching speeds.

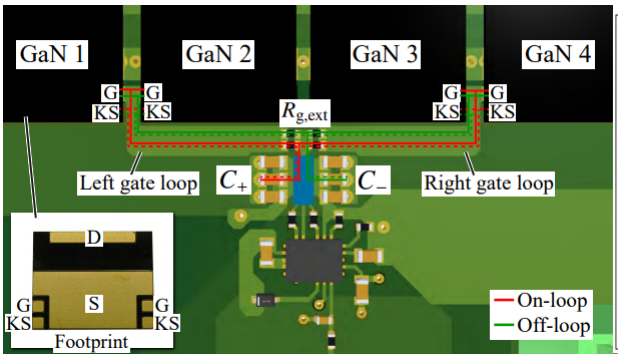


Fig. 4. Symmetric gate driver loop with minimized loop inductance[4]

B. Power loop design

The key factor in the power loop design is to achieve equal parasitics in each GaN HEMT branch. This gets rid of transient current imbalances which further improves the switching speed. The relevant parasitics are depicted in Fig. 5. Certain methods to achieve symmetry and minimize parasitics are, fine tuning of the wire bond lengths of each semiconductor chip and using discrete semiconductor devices[4].

The switching cell of the power loop includes a heat sink for limiting the power loss of the cell. The isolated power supply tends to block the air circulation. This can be prevented by by placing the power supply at the back side of the PCB and attaching the heat sink on the top side. The unique heat sink design is depicted in figure 6[2]. Heat can be dissipated effectively by controlling the airway. Epoxy insulation is used to insulate the screws used to attach the heat sink to the PCB.

V. Thermal Simulation of the Design

A thermal simulation is performed on a four-GaN paralleled system with two half-bridges. The ambient temperature is taken to be 60°C. A similar junction thermal resistance is assumed to determine the loss dependency due to a slightly asymmetrical layout. The power loss is determined at steady state after 0.6ms[8]. The actual experiment was carried out and

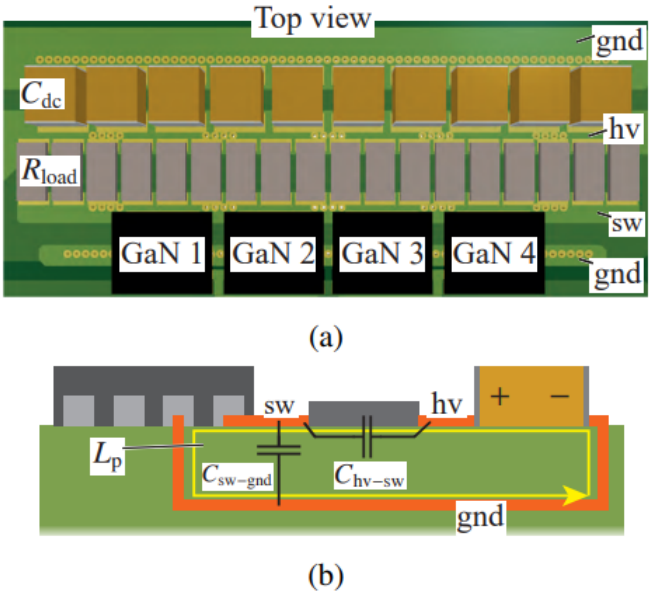


Fig. 5. Power loop design including layout parasitics (a) Top view and (b) cross sectional view[4]

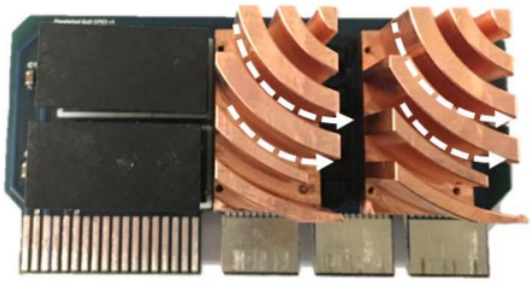


Fig. 6. Unique heat sink design of the switching cell[2]

it was found that the actual temperatures were lower than the simulated temperatures but the temperature differences were comparable. Due to the comparable high parasitic inductance and resistance between the transistors, Low side driver of the second branch therefore suffers higher loss power and is reaching a temperature of 69°C, which was already observed in the comprehensive simulations, where the device temperature was nearly constant. The low-side driver of the first branch again is approximately 12K colder.

VI. Limitations of Parallelization

A. Unsymmetrical PCB Design

Because of the high voltage and current transients of GaN HEMT switching, the parasitic inductance of the PCB traces is critical to the GaN HEMTs safe operation and the electromagnetic induction (EMI) noise control of converter.[1] Several compact layouts for paralleling bare die E-mode GaN devices are demonstrated, with minimal parasitic inductance. High device switching speed and considerable device package size are the limitations for the PCB layout design. A commonly used half bridge gate driver chip is utilized for the

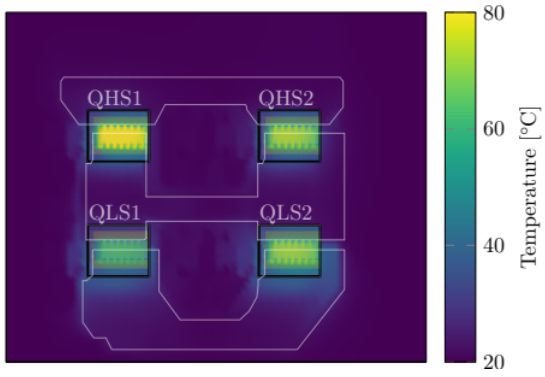


Fig. 7. Thermal steady state simulation after neglecting the influence of thermal interface material and heat-sink[8]

phase leg design. The circuit layouts are shown in Fig. 8a and 8b respectively. By increasing the number of devices in the circuit, the main loop layout will possess certain constraints. In case of the parallel configuration, the loop inductance in each parallel device becomes unsymmetrical due to the presence of decoupling capacitors. This leads to unbalanced values of peak current and voltages on the paralleled GaN HEMTs.

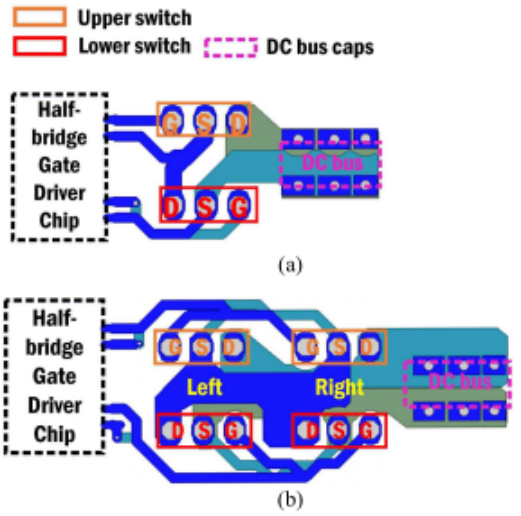


Fig. 8. Phase-leg circuit (a) the single GaN HEMT layout, and (b) the parallel GaN HEMTs layout[1]

B. Gate Drive Signal Mismatch

The mismatch of the gate drive signals leads to dynamic current sharing with the paralleled GaN HEMTs. Even if the paralleled devices share the same gate drive chip, signal mismatch rises as a result of gate drive circuit impedance mismatch. All the components inside the gate drive circuit needs to be carefully selected and placed, including the gate resistor, the ferrite bead, the diode and the parasitic inductance of the PCB trace[1]. When paralleling a large number of devices, unique gate drivers should be supplying each of

them. The single stage common emitter buffers should be u instead of a complete gate driver chip[1].

C. Miller Turn-on

When one switch of the half bridge is turned on by ha switching, the drain-source voltage V_{DS} of the second swi increases, which in turn leads to an increase in the dra gate voltage. A drain to gate current i_{DG} flows through Miller capacitance C_{DG} as illustrated in Fig. 9. The share i_{DG} charging C_{GS} is proportional to the total impedance of gate turn-off path and hence the half-bridge becomes m susceptible to unintended turn-on. This destructive even called Miller turn-on. The total turn-off impedance compri the internal driver resistance R_{DrOff}, the external gate tu off resistance R_{Goff}, the gate loop inductance L_G and internal gate resistance of the GaN device R_G,GaN. Prese of larger turn on resistance $R_{G,on}$ will reduce the occurre of Miller turn on but may lead to increased switching los Another alternative to prevent this destructive event is provide bipolar gate voltages.

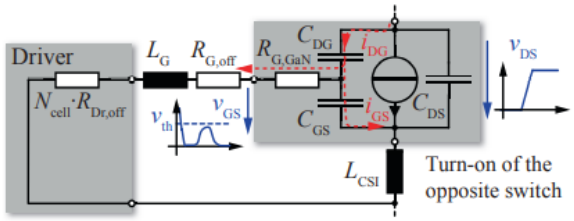


Fig. 9. i_{DG} partially charges C_{GS}. Depending on the impedance in the turn-off path, this can result in an unintended turn-on of the turned-off device[3].

VII. Example of a Multi-phase Configuration of GaN

An example of multi-phase operation of GaN can be explained in a multi-phase buck converter. In a buck converter, there are three main electrical components: a switch, such as a insulated gate bipolar transistor (IGBT) or Metal oxide semiconductor field effect transistor (MOSFET), a diode, and an inductor. There are different configurations of how buck converters look, but the most common setup is shown in figure 10. Through rapid switching, this design can achieve a voltage step-down, from source to load, in an efficient manner[5]. This design can be repeated to higher number of phases to reach higher load current specifications as well. The iterations of these stacked circuits, shown in figure 10, are referred to as "phases".

In a multi-phase buck converter design the "power stage" of each phase is paralleled together to supply a common load. The power stage in a buck converter is composed of the high-side and low-side switching device, and the inductor for smoothing the current. In a synchronous buck converter, the low-side switching device is a transistor which works as similar to a diode. When the power stages are paralleled together, each phase carry a portion of the overall current that is being delivered to the load[5]. The multi-phase buck converter improves the overall power efficiency of the point-of-load (POL) system by evenly spreading out heat dissipated in the design, as opposed to a single-phase design with one centralized source of heat.

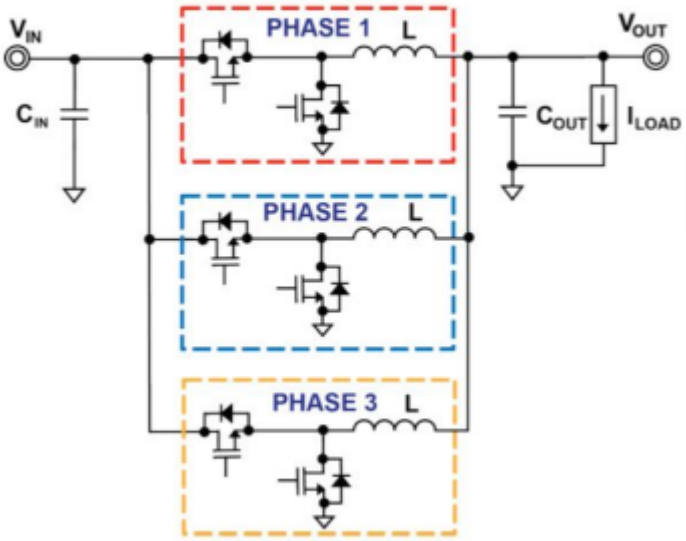


Fig. 10. A Multi-phase Buck converter circuit with GaN switches[5]

VIII. Summary and Conclusion

The paper presents detailed summary of all aspects of Parallelization of GaN transistors. GaN HEMTs are now preferred over traditional Si and SiC MOSFETS because of their operating frequency ranges and high switching speeds. The main motive of parallelization is to utilize GaN transistors in high current applications. A proper design of the paralleled GaN circuits is mandatory to overcome the limitations mentioned in this paper. Furthermore, we discuss different topologies that can be realized and concludes by choosing the most efficient topology based on gate driver and load symmetries. Symmetry is more intensively discussed in this paper whether it is in the gate loop or in the power loop. A small difference in any characteristics in the parallel branches may lead to severe consequences like damage to the device and must be prevented at all cost. The paper describes the state of the art system of parallelization of the device based on the most efficient topology. The PCB designs of the gate loops and power loops are designed by considering all possible parasitics and discusses different ways to minimize them. Finally, we discuss a multi-phase operation of GaN HEMTs which is an alternative concept of parallelization by differentiating each branch into a separate phase. This paper can be taken as a base reference for all future possible developments in GaN parallelization.

REFERENCES

[1] H. Li et al., "Paralleled Operation of High-Voltage Cascode GaN HEMTs," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 4, no. 3, pp. 815-823, Sept. 2016, doi: 10.1109/JESTPE.2016.2557316.

[2] B. Sun, R. Burgos, N. Haryani, S. Bala and J. Xu, "Design, characteristics and application of pluggable low-inductance switching power cell of paralleled GaN HEMTs," IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, 2017, pp. 1077-1082, doi: 10.1109/IECON.2017.8216185.

[3] J. Burkard and J. Biela, "Paralleling GaN switches for low voltage high current half-bridges," 2019 IEEE Energy Conversion Congress and Exposition (ECCE), 2019, pp. 3245-3252, doi: 10.1109/ECCE.2019.8912830.

[4] R. Risch and J. Biela, "Low Voltage GaN-Based Gate Driver to Increase Switching Speed of Paralleled 650 V E-mode GaN HEMTs," 2020 22nd European Conference on Power Electronics and Applications (EPE'20 ECCE Europe), 2020, pp. P.1-P.11, doi: 10.23919/EPE20ECCEEurope43536.2020.9215773.

[5] "Multiphase Buck Design From Start to Finish (Part 1),"ti.com, Apr-2017 [Online] Available: https://www.ti.com/lit/an/slva882b/slva882b.pdf?ts=1656313393629

[6] GaN Systems Inc., "GN004 Design Considerations of Paralleled GaN HEMTs 200910," 2021. [Online]. Available : https://gansystems.com/wp.content/uploads/2020/09/GN004-Design-Considerations-of-Paralleled-GaN-HEMTs_200910.pdf

[7] D. Reusch and J. Strydom, "Effectively paralleling gallium nitride transistors for high current and high frequency applications," 2015 IEEE Applied Power Electronics Conference and Exposition (APEC), 2015, pp. 745-751, doi: 10.1109/APEC.2015.7104433.

[8] D. Koch, J. Weimer, M. Weiser, J. Hueckelheim and I. Kallfass, "Gate Driver Concept for Parallel Operation of Low-Voltage High-Current GaN Power Transistors for Mild-Hybrid Applications," 2021 IEEE Applied Power Electronics Conference and Exposition (APEC), 2021, pp. 1755-1760, doi: 10.1109/APEC42165.2021.9487194.

A review on Variable Gain Amplifiers and Automated Gain Control in MMICs

Electrical engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—This paper is a review on the use of variable gain amplifiers and automated gain control in microwave and millimeter-wave monolithic integrated circuits (MMIC) Research, summarize and compare state-of-the-art circuit design techniques in MMICs to realize VGAs AGC implementations for different semiconductor technologies (Si CMOS, SiGe BiCMOS, III-V compound semiconductors GaAs, InP, GaN)**

***Index Terms*—Monolithic Microwave integrated circuits (MMICs), Variable Gain Amplifier(VGA), Automated Gain Control (AGC), Low Noise Amplifier (LNA), High electron Mobility Transistors(HEMT)**

I. Introduction

Amplifiers in play a fundamental role in wireless communication links. With time-varying link parameters like the free-space path loss, atmospheric attenuation, or modulation formats, To run the link at its best efficiency, such as achieving the highest data rate or the longest transmission distance, it is preferable to change the output signal power or the reception gain.it is advantageous to adjust the transmit power or the receiver gain to operate the link at its optimum performance, e.g. achieving maximum data rate or transmission distance. To achieve this functionality, variable gain amplifiers are used, as the name suggests it's an amplifier with the gain controlling feature. A VGA is an attenuation network, which decreases the maximum gain of the amplifier or adds another block specified gain stage based on a control signal that either increases or decreases the gain. Hence a VGA should have three components 1) Gain stage(s) as its core. 2) A network that adds or subtracts gain to total output gain. 3) A control signal which varies the total output gain of the VGA. A desired function in wireless receivers is AGC because the power received from the wireless channel is varying with time. The AGC circuitry maintains a fixed output signal amplitude with input signals of variable amplitude, such VGAs are embedded into an AGC in conjunction with e.g. power detectors and digital control circuitry. This concept of AGC in receivers is applied to transmitters as well, where the transmitter has to maintain the desired output power, Ex: In mobile phones monitoring transmitted power. AGC-equipped variable gain amplifiers are utilized in a range of communications and remote sensing devices. For signal chains that demand a high dynamic range, variable gain amplifiers provide unparalleled performance. VGAs have been used to give industry-leading performance in applications such as voice processing, ultrasound, radar, LIDAR, and wireless communications.

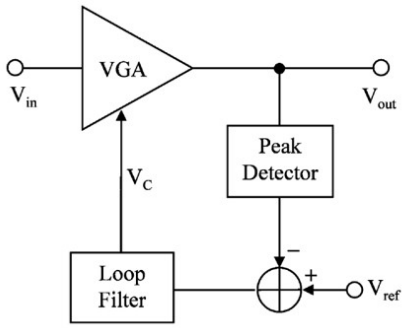


Fig. 1. Conventional AGC loop.

II. VGA implementation techniques

There are many different ways in which a VGA can be implemented: 1) A open or closed loop system. 2) Analog or Digital control voltage. 3) Use of non-linear components (Ex: BJT or MOSFET) or passive components in a feedback loop. VGAs are mostly implemented with linear components such has resistors and capacitors in binary-weighted or digitally switched network, references [2], [5], and [14] demonstrate this technique is depicted in Fig (2). A Wide-band variable gain low noise amplifier using a ladder attenuator with a gain range from 0 to 40 dB is voltage controlled resistor ladder network [5] is an example for an open loop system.

A programmable gain/loss circuit using an 8-bit digital word providing up-to 256 0.l- dB steps in gain or loss fabricated on 3 μm n-well process with a maximum gain of 25.5 dB is designed in [2] using a resistor network is an example for closed loop system. In the above three examples, it can be seen that the control signal is a digital signal switching ON/OFF a certain branch of the network to control the total gain of the VGA. Though conceptually simple and appealing, the preceding method for producing VGAs in CMOS is made possible by the easy accessibility of analog switches and precisely ratioed capacitors. It does have some restrictions, though. The ratio of maximum to minimum, or spread, of feedback element values increases exponentially and may occupy a significant chip surface when a wide decibel range

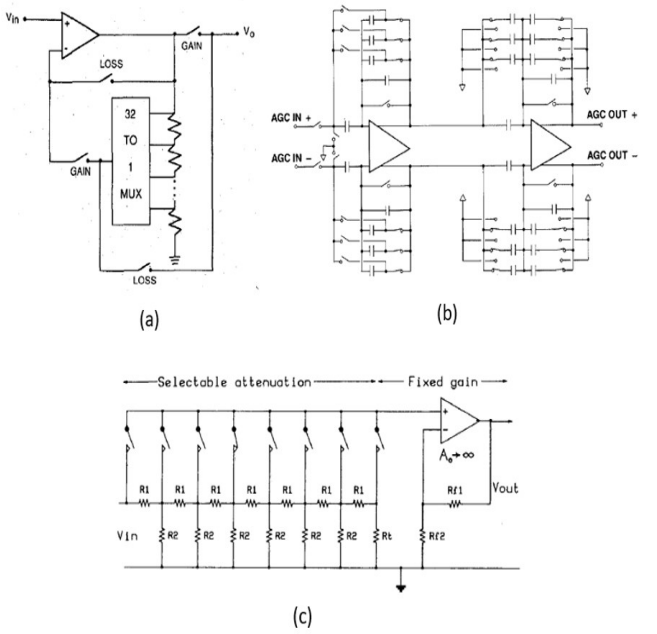


Fig. 2. (a) Closed loop VGA circuit of a 5 bit stage. [2] (b) VGA circuit of 6 bit switched capacitor. [14] (c) Open loop VGA circuit with discrete step scheme. [5]

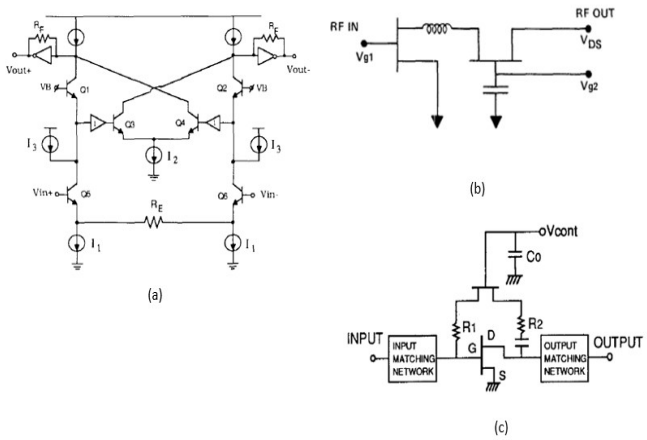


Fig. 3. (a) VGA designed using Gilbert cell [10] (b) Conventional cascode VGA circuit (c) Proposed design in [9] for low distortion VGA.

of adjustable gain is needed. The operational amplifier must be corrected for stable operation in feedback, which further limits the frequency of operation. [10].

The utilization of non-linear components directly in the signal channel has been a recurrent theme in the implementation of voltage-controlled amplifiers. Ex: wide-band VGA often invokes the exponential I_C-V_{BE}/I_D-V_{GS} relationship in a BJT or MOSFET to implement a controlled current fork or a transistor multiplier core [5]. Bipolar ICs have developed a traditional method suitable for broad dynamic range and wide-band VGAs. The Gilbert gain cell, whose gain is precisely controlled by the ratio of bias currents, serves as its foundation. Trans-linear circuits are those that make use of the bipolar transistor's exponential I-V characteristic, or the linear dependence of transconductance on bias current. It has been implemented to use a variable gain amplifier (VGA) with a gain range of 50 dB in a standard 3 μm CMOS process using parasitic lateral and vertical bipolar transistors to form the core of the VGA circuit [10].

The transconductance (superscriptsg_m) dependency of superscriptsV_G or superscripts$V_D D$ is used in conventional variable gain FET amplifiers. The typical drain-voltage controlled VGAs, such as cascode amplifiers and twin gate amplifiers, are gain control amplifiers. In reference [9], a new variable gain amplifier is suggested as a solution to the aforementioned issues. Fig. 3 depicts the amplifier's fundamental setup. Capacitors, fixed resistors, matching networks, a FET for amplification, a FET for a varistor, and a FET make up the amplifier. Variable negative feedback is employed by the amplifier. The linearity of amplifiers can be increased using negative feedback. The FET varistor's control voltage (superscriptsV_{cont}) can be adjusted to modify the gain. Strong negative feedback occurs when the varistor is adjusted to a low resistance value, compressing the amplifier gain ([9]).

In [13], As depicted in Fig. 4(a), a variable gain mixer is suggested that operates from RF frequency range of 18-33 GHzin the 0.13 μm SiGe: BiCMOS technology. A micromixer structure, as opposed to the conventional Gilbert cell, is chosen as the mixer core due to its superior linearity performance. Additionally, it does away with the necessity for any input balun, conserving chip space. NMOS transistors acting as a variable resistor and consuming no additional power regulate the mixer's gain. As a result, the VGA can be removed from the receiver to save energy. The linearity of the output buffer is the only constraint on the mixer gain, which has a range of -10 dB to 15.5 dB and a maximum input 1 dB compression point (IP1 dB) of -10 dBm.

Due to its advantages of having a large bandwidth, 5G mm-wave communication has recently attracted attention from all around the world. By combining the phase and acquiring control capabilities to reduce the sidelobe and take into consideration path loss, the phased array transmitter and receiver design may be created. Variable Gain Low Noise Amplifiers (VGLNAs) and Variable Gain Power Amplifiers (VGPAs) have been developed in response to the aforementioned specifications (see references [18] and [21]). These amplifiers are preferred because they have phase inverting capabilities and minimal phase variation, which simplify the complexity of the system calibration. A Phase-Inverting Variable Gain Power Amplifier in 65 nm CMOS for 5G Communication is shown in Reference [21]. The VGPA achieved the dB-linear gain control features as well as 180°phase inversion by using a Gilbert structure-based VGA stage. With a dB-linear gain tuning range of 5.6/5.2 dB, this VGPA produces a maximum gain of 38.7/38.9 dB. Additionally, its phase inversion error is less than 5.9°, and its RMS phase error from 38 to 43 GHz is less than 3.6/2.2°. In the Gilbert-based cascade gain stage, four

Ref	[19]	[8]	[15]	[1]	[4]	[16]	[3]
Process	65-nm CMOS	120-nm SiGe	65-nm CMOS	55-nm BiCMOS	100-nm GaN HEMT	130-nm SiGe BiCMOS	35-nm HEMT InGaAs/lnAIAs
Topology	2-stage current steering	1-stage current steering	2-stage current steering and 1 stage splitting cascode	π-type DAC attenuator	cascode with voltage divider control	cascode voltage varying the bias point	cascode voltage varying the bias point
Gain control mechanism	Digital (4-bit)	Digital (3-bit)	Analog	Digital (5-bit)	Analog	Analog	Analog
Frequency (GHz)	38-40	32-34	57-66	71-76	75-110	180-200	250 - 300GHz
Gain control range(dB)	16	11	33	12	24.2	24.7	15
RMS Phase error (°)	<2.67 0.89@40GHz	<2.5 0.8@33GHz	N/A	<2.6 1.9@73.5GHz	N/A	N/A	N/A
Phase variation (°)	3.18@40GHz	3.18@40GHz	7	0.5@73.5GHz	12.6@94GHz	N/A	N/A
$\|S11\|$(dB)	-10	-11	-6.5	-10	N/A	-10	-5
$\|S22\|$ (dB)	-4.5	-15	-5	-6	N/A	-6	-4
Max OP1dB (dBm)	2.5@40GHz	-7	-4	-1.5@73.5GHz	15@94GHz	-6.5@190GHz	3@260GHz
DC power (mW)	38	33	10.9	18.4	N/A	37.2	N/A
NF (dB)	7.9	4.3	N/A	N/A	N/A	9.8	4

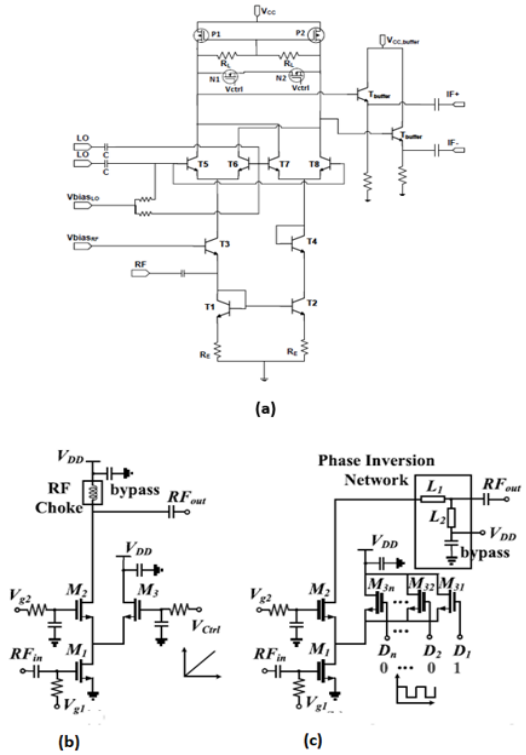


Fig. 4. (a) VGA designed has mixer [13] (b) VGA with analog current steering (c) VGA with digital current steering [17]

common-gate transistor units of various sizes are introduced in tandem. The associated transistors are activated or deactivated using the digital control bits, resulting in the effective total transistor width size (Wa and Wb) having various ON/OFF states.

A digital gain control and low phase variation 40-GHz VGA with 4-bit is designed and manufactured on the 65-nm CMOS technology, according to reference [17]. The VGA's measured gain ranges from 22 to 6 dB at 40 GHz and has 16 distinct linear-in-decibel gain levels. The standard analog form of current guiding VGA is composed of a cascade configuration (M1 and M2) with a supplementary current steering transistor M3, whereas the digital copy of current guiding design is offered as indicated in Fig. 4(b) (c). The device M3 can be broken down into numerous transistor cells (M_{3n}, n = 1 to n) using digital control voltage (M_{3n}, n = 1 to n). By selecting the size of such device cells, one may determine the magnitude of the directing current of M2. By adjusting the size of these device cells, the value of M2's steering current can be controlled, and a linear-in dB gain change characteristic may be accomplished.

III. AGC IMPLEMENTATION TECHNIQUES

AGC circuit's principal function is to maintain a virtually constant output level regardless of the strength of their input signal. Since the invention of radio transmission, these circuits have been in use, and they have frequently been designed using empirical principles. Although their performance and design have improved since then, in most cases their behavior is still enough for the emergence of new and more demanding systems. For applications requiring jitter correction via the linear mode functioning of equalizers, deciding circuits with flexible dynamic range, and sensibility improvement utilizing automatic threshold modification, an AGC is frequently preferred over a simple limiting amplifier. [11]

The forward path also includes two post amplifiers to boost bandwidth and sensitivity, a VGA for gain adjustment, an exponential voltage generator to convert control voltages in to one of differential exponential gain control voltages, a dc offset canceller with such a common-mode voltage adaptation system to cancel the offset voltage and correct the common mode voltages to an adequate point, and a test buffer to drive a 50Ω off-chip load. The suggested PD extracts the output signal's amplitude linearly in the feedback channel. The integrator then adds up the difference between the output of PD V_{PD} and an internal reference voltage to produce the control

voltage. V_{REF} must be generated by a reference level shifter (Ref. Level Shifter) in order to automatically attain the same proportional coefficient as V_{PD} because V_{PD} is proportional to the output signal amplitude. When powered by a 1.2-V power supply, the AGC uses 72 mW.This does not include the output buffer that powers the 50Ω probe. From 200 kHz to 7.5 GHz, the 3-dB bandwidth is available. The gain of the variable-gain amplifiers (VGA) is adjusted by feedback loops in conventional closed-loop analog AGCs to achieve the required output signal intensity. The signal amplitude is commonly calculated in such AGC loops using a peak detector, as shown in Fig. 1. This is reliable for signals with low peak-to-average power ratios (PAPR), such as sinusoids, but is unreliable for signals with high PAPR. The lack of time for amplitude estimation makes the issue worse. [7]

To fix the aforementioned problems A pseudo-RMS detection AGC algorithm is used in [6], which implements an open loop AGC. The efficacy of this method for calculating OFDM amplitude utilizing received 0.8 μs brief training symbols has been established. The AGC adjusts the gain almost within 1 dB of its ideal value in the time for 7 brief training symbols. The AGC operates in three stages: one open-loop phase for precise gain tuning, two switching phases for setting the coarse gain. The gain is set within 5 dB of its ideal value and all gain and filter stages perform linearly thanks to the coarse gain-setting steps. The fine step brings the final gain to its ideal level within +/- 1 dB.

IV. VGAs and AGCs for THz communication

Because of the low oscillation frequency ($fmax$) and lossy substrate of the CMOS process, designing sub-THz CMOS amplifiers is a difficult task. Poor (f_{max}) in the amplifier design results in low intrinsic gains, such as the maximum available/stable (G_{ma}/G_{ms}) and unilateral (U) gains. The high gain amplifier design necessitates numerous stages and results in a significant dc dissipation because of the high loss passive parts resulting from the lossy substrate. Large output power can be provided via large transistor sizes. Although it is challenging to increase the transistor size in the CMOS process, doing so can result in a reduction in (f_{max}) . Low f_t in the CMOS process makes designing sub-THz LNAs difficult [20]. High-speed integrated circuits up to sub-THz frequencies are now practical thanks to the advancement of (sub) THz transistor technology. The performance of optical components, coherent optical fiber communication, and imaging systems can all be enhanced by these high-speed and wide bandwidth ICs. The fact that NMOSFETs in CMOS integrated circuits have f_{max} that are below or close to 300 GHz poses the major obstacle to the development of 300 GHz CMOS transceivers. As a result, implementing an amplifier in the 300 GHz spectrum is challenging. In [3], a multistage balanced variable gain amplifier for the H-band (220-325 GHz) with tandem-X couplers is described. According to the author's knowledge, this is the only paper with a VGA in the 300 GHz frequency band. As is customary for THz frequencies, the fabrication uses 35 nm mHEMT devices with a maximum

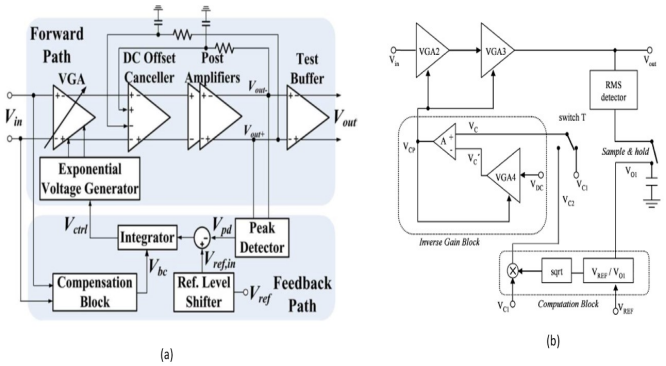


Fig. 5. (a) Conventional AGC loop [2] (b) Open loop AGC with rms algorithm implementation [14].

gain of 20 dB at 261 GHz and a gain control range of 15 dB in the frequency range of 220 to 320 GHz. Sadly, there are no VGA or AGC designs in the few additional articles that discuss simple amplifiers in the 300 GHz frequency ranges.

V. Use of different fabrication technology

For the CMOS VGA, which is generally categorized as either a precisely defined gain-step type with a digitally controlled signal [5] or a continuously variable gain type operated by an analog gain-control signal, it is challenging to create an exponential for a logarithmic function because MOS transistors typically operate with a square-law transfer characteristic. An approximative rational function is utilized rather than an exponential function for the deterministic gain range of up to 30 dB. For CDMA receivers, which need an intermediate frequency (IF) gain of at least 70 dB, the 30 dB gain range is inadequate. Parasitic bipolar devices are employed for precise exponential transformation for a larger gain range. The time to calculate amplitude must be substantially smaller than the time constant of the loop filter for closed-loop AGCs to function properly. The required DC amplitude is matched to the measured output amplitude when the incoming amplitude changes, and the difference error is supplied back to adjust the VGA's gain to maintain a constant output amplitude. The negative feedback loop automatically adjusts for changes in input amplitude, although slow loop response is necessary for stability. As device cutoff frequencies rise, MMIC LNAs are widely being employed as radio front-ends in the 0.1-0.3 THz region. In terms of noise performance, bandwidth, and output power, HEMTs constructed in the GaAs or InP semiconductor system have been the best available technology. Although a small number of HBT results are documented in the literature, HEMTs are typically used to construct W-Band (or high frequency) amplifiers. The greatest power output that can be obtained with the device is determined by the gate width, whereas the cutoff frequency of a HEMT is determined by the gate length of the device. More total output power will result from larger, larger peripheral devices with a wider gate-width. [12]

VI. Conclusion

In this work, numerous VGA design configurations are reported. Variable transistor transconductance, dynamic feedback, current splitting, variable attenuators, and current steering techniques are some of the gain control techniques that are used in VGA. Attenuators and active feedback systems result in high losses. This method, which uses a cascade of transistors in a cross-couple direct bypass design, provides a wide gain control range. The current dividing methods display a constrained gain control range, significant power consumption, and a significant loss at mm-wave frequencies because multiple transistors are utilized to split the current. The current steering approach uses a broad gain control range and minimal power consumption due to its simplicity and compatibility for mm-wave technology. It is discovered that many cascode topologies are used in designs that operate above W-band. Basic AGC topologies, which are utilized in communication systems, are given in this study. AGCs use VGAs to regulate the signal strength in order to supply steady feed to subsequent stages. Peak detection and control voltage supply circuit performance limits the THz communication AGCs. High $fmax$ and ft fabrication processes enable high speed designs.

References

[1] Kimia T. Ansari, Tyler Ross, and Morris Repeta. An e-band variable-gain amplifier using a programmable attenuator. In *2018 48th European Microwave Conference (EuMC)*, pages 1289–1292, 2018.

[2] J.N. Babanezhad and R. Gregorian. A programmable gain/loss circuit. *IEEE Journal of Solid-State Circuits*, 22(6):1082–1090, 1987.

[3] I. Dan, D. Meier, S. Wagner, A. Leuther, A. Tessmann, H. Massler, and I. Kallfass. A 300 ghz multi-stage balanced variable gain amplifier with tandem-x couplers. In *2015 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS)*, pages 1–4, 2015.

[4] S. Diebold, D. Müller, D. Schwantuschke, S. Wagner, R. Quay, T. Zwick, and I. Kallfass. Algan/gan-based variable gain amplifiers for w-band operation. In *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, pages 1–4, 2013.

[5] B. Gilbert. A low-noise wideband variable-gain amplifier using an interpolated ladder attenuator. In *1991 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pages 280–281, 1991.

[6] O. Jeon, R.M. Fox, and B.A. Myers. Analog agc circuitry for a cmos wlan receiver. *IEEE Journal of Solid-State Circuits*, 41(10):2291–2300, 2006.

[7] Chang Liu, Yue-Peng Yan, Wang-Ling Goh, Yong-Zhong Xiong, Li-Jun Zhang, and Mohammad Madihian. A 5-gb/s automatic gain control amplifier with temperature compensation. *IEEE Journal of Solid-State Circuits*, 47(6):1323–1333, 2012.

[8] Byung-Wook Min and Gabriel M. Rebeiz. Ka-band sige hbt low phase imbalance differential 3-bit variable gain lna. *IEEE Microwave and Wireless Components Letters*, 18(4):272–274, 2008.

[9] M. Muraguchi and M. Aikawa. A linear limiter: a 11-ghz monolithic low distortion variable gain amplifier. In *1991 IEEE MTT-S International Microwave Symposium Digest*, pages 525–528 vol.2, 1991.

[10] Pan and Abidi. A 50 db variable gain amplifier using parasitic bipolar transistors in cmos. In *Symposium 1988 on VLSI Circuits*, pages 9–10, 1988.

[11] Sagar Ray and Mona Mostafa Hella. A 10 gb/s inductorless agc amplifier with 40 db linear variable gain control in 0.13 m cmos. *IEEE Journal of Solid-State Circuits*, 51(2):440–456, 2016.

[12] Lorene Samoska. Towards terahertz mmic amplifiers: Present status and trends. In *2006 IEEE MTT-S International Microwave Symposium Digest*, pages 333–336, 2006.

[13] Syed Sharfuddin Ahmed and Hermann Schumacher. A 18-33 ghz variable gain down-conversion mixer in 0.13µm sige:c bicmos technology. In *2021 Austrochip Workshop on Microelectronics (Austrochip)*, pages 9–12, 2021.

[14] Cheng-Chung Shih, Kwai-Kwong Lam, Kuang-Lu Lee, and R. Schalk. A cmos analog front end for 9600bps facsimile modem. In *1987 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, volume XXX, pages 300–301, 1987.

[15] Di-Sheng Siao, Jui-Chih Kao, and Huei Wang. A 60 ghz low phase variation variable gain amplifier in 65 nm cmos. *IEEE Microwave and Wireless Components Letters*, 24(7):457–459, 2014.

[16] Paul Stärke, David Fritsche, Corrado Carta, and Frank Ellinger. A 24.7 db low noise amplifier with variable gain and tunable matching in 130 nm sige at 200 ghz. In *2017 12th European Microwave Integrated Circuits Conference (EuMIC)*, pages 5–8, 2017.

[17] Jeng-Han Tsai and Chia-Lung Lin. A 40-ghz 4-bit digitally controlled vga with low phase variation using 65-nm cmos process. *IEEE Microwave and Wireless Components Letters*, 29(11):729–732, 2019.

[18] Yunshan Wang, Tzu-Yang Chiu, Chun-Chia Chien, Wei-Hsuan Tsai, and Huei Wang. An e-band high-performance variable gain low noise amplifier for wireless communications in 90-nm cmos process. *IEEE Microwave and Wireless Components Letters*, pages 1–4, 2022.

[19] Han-Chih Yeh, Sofiane Aloui, Chau-Ching Chiong, and Huei Wang. A wide gain control range v-band cmos variable-gain amplifier with built-in linearizer. *IEEE Transactions on Microwave Theory and Techniques*, 61(2):902–913, 2013.

[20] Byeonghun Yun, Byeong-Taek Moon, and Sang-Gug Lee. Cmos based sub-thz wireless transceivers: Issues and solutions. In *2022 14th Global Symposium on Millimeter-Waves Terahertz (GSMM)*, pages 135–138, 2022.

[21] Xuexue Zhang, Xiaokang Niu, Qin Chen, Xin Chen, Depeng Cheng, Jing Feng, Jun Feng, and Lianming Li. A 39-ghz phase-inverting variable gain power amplifier in 65-nm cmos for 5g communication. *IEEE Microwave and Wireless Components Letters*, pages 1–4, 2022.

Built-in Self-Test and Calibration in MMICs

Electrical Engineering
Universität Stuttgart
Stuttgart, Germany

***Abstract*—An overview of requirements for built-in self-testing and calibration in the microwave and millimeter-wave monolithic integrated circuits is provided in this paper. The advantages of various circuit design techniques, semiconductor technologies, and architectures for self-testing are discussed along with their drawbacks and challenges. A state-of-the-art Si CMOS-based radar transceiver has also been reviewed.**

***Index Terms*—built-in self-test, CMOS, calibration, mm-wave, transceiver.**

I. Introduction

For over a decade, researchers working in the field of silicon-based semiconductor technology have made developments that enabled achieving a high level of integration of RF circuits into the system on chip (SoC) [1]. These devices are used for various applications which include communication, imaging, and automotive radars. Millimeter-wave integrated circuit (MMIC) testing is a complicated, time-consuming process, and an expensive task. Testing of radio frequency (RF) building blocks and integrated transceiver systems requires a high-performance measurement system, with necessary frequent calibration procedures necessary. In addition, the continuous increase in operational frequencies has led to the fact that testing these radio frequency integrated circuit (RFIC) has escalated to about 40% and become a substantial part of the total manufacturing cost and time [5]. Analog/RF circuits are becoming more susceptible to fluctuations in temperature, noise coupling, and process variations due to advances in manufacturing processes. Consequently, production loss is increased, which must be compensated with higher design margins. One of the solutions for problems that occurred during the test phase is to use special stimulus waveforms as input signals, and the captured responses are processed to predict circuit performance parameters [6]-[11]. The Built-in self-test (BIST) focuses on simplifying the testing procedures at the end of chip fabrication, as well as providing an online diagnosis of critical parameters with on-chip calibration capabilities.

II. Design techniques used to Realize BIST in MMIC

A. Direct conversion type of architecture

1) Architecture: In this architecture, the performance parameters namely gain (G), and noise figure (F) are used to calculate the total performance of the millimeter-wave front-end [2]. Gain is calculated using a low-frequency test signal which is up-converted to produce a test signal at RF frequency using an up-conversion mixer powered by a local oscillator (LO) that supplies a LO power [1]. The coupler feeds the receiver with an RF test signal, and it is measured by a power detector which is proportional to the input. Using a digital signal processor, the output power is the direct conversion of the detector output, and the signal processor output provides the information about the gain of the test device. G is calculated with the equation (1).

$$G = P_{\text{out}}/P_{\text{in}} \tag{1}$$

where P_{out} is output power of the system, P_{in} is input power of the system. The noise figure is also an important performance parameter of a receiver front-end. Direct noise measurement is added by introducing a resistor at the receiver front-end. As a result, the input noise density power is defined. The output noise density power must be measured to compute the F which is calculated from the equation (2).

$$F = P_{\text{out}}/kT_0BG \tag{2}$$

where P_{out} is the output power of the system, k is Boltzmann's constant in joules per kelvin, and T_0 is the receiver system noise temperature in kelvins, B is bandwidth. By using the above-mentioned type of measurement procedures, we can calculate F and G. The noise figure is only determinable if the G and the B is known. Therefore, G must be measured in advance. Due to the use of a 90° phase shifter, the two side-band frequencies are generated with the up-conversion mixer as shown in Fig. 1. A single-side-band modulator is shown in Fig. 2. As a consequence, only one of the two side-band frequencies is generated while the other is canceled out by destructive interference. With the use of the Hilbert modulator as shown in Fig. 3, the total signal power is included with one side-band and therefore built-in self-test measurement is not harmed by interference effects.

2) Analysis: The accuracy of the proposed measurement method is limited by the error that is made by the power measurement at input and output power of the test device [2]. The relative error in gain is calculated using the equation (3).

$$\delta G = \sqrt{\delta^2{}_{\text{Pin}} + \delta^2{}_{\text{Pout}}} \tag{3}$$

where δG is relative error in gain, $\delta^2{}_{\text{Pin}}$ is relative error of input power $\delta^2{}_{\text{Pout}}$ is the relative error of output power. The relative error in built-in test noise figure is defined as in equation (4),

$$\delta F = \sqrt{\delta^2{}_{\text{Pin}} + 2\delta^2{}_{\text{Pout}}} \tag{4}$$

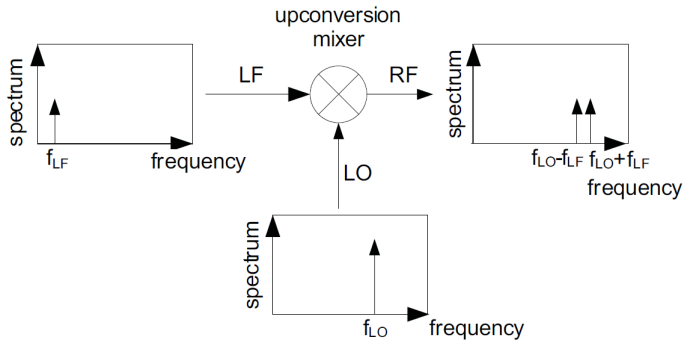


Fig. 1. Generation of two sideband frequencies during up-conversion of a sine wave test signal by a single mixer stage (Adapted from [1])

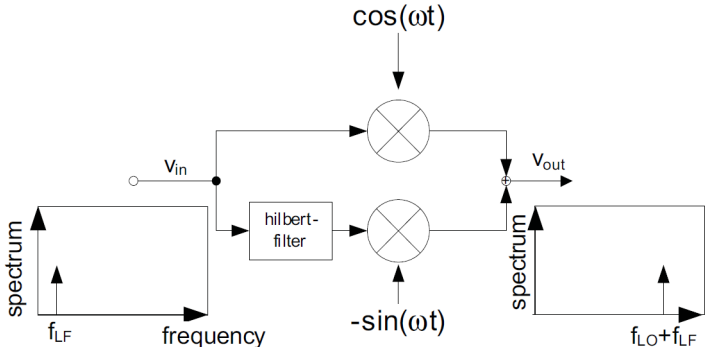


Fig. 2. Generation of a single side-band millimeter wave test signal from a low frequency input utilizing a Hilbert Modulator (Adapted from [1])

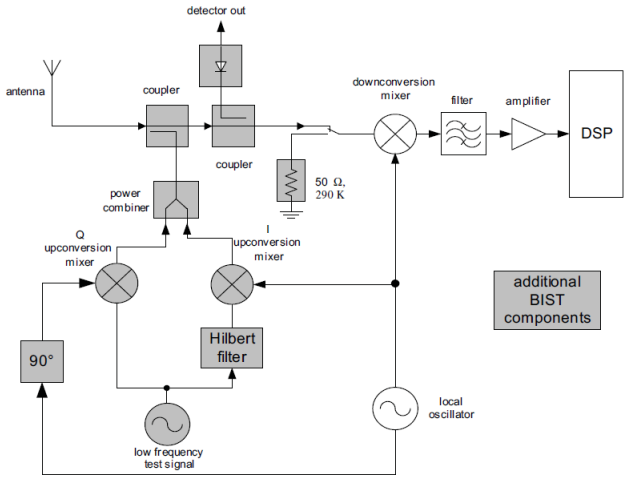


Fig. 3. Test architecture for built-in self test of gain and noise figure for RF receiver frontend using a single sideband modulator (Adapted from [1])

where δF is relative error in noise figure, $\delta^2{}_{\text{Pin}}$ is relative error of input power $\delta^2{}_{\text{Pout}}$ is the relative error of output power.

3) Drawbacks: The major drawback in this system of the relative error in F which is dependent on the accuracy of the of output and input power measurement [1].

B. SiGe based down-conversion mixer architecture

1) Architecture: In this section, a BIST which offers testing procedures as well as online diagnosis through simple low-frequency off-chip interfaces is described. The BIST is developed at the expense of an increased chip area and the power consumption for the additional testing circuitry [12]. An innovative down-conversion mixer architecture that can simultaneously up-convert a low-frequency test signal [2]. A double-sideband signal is up-converted and coupled back into the mixer input path using this approach. This enables a built-in functionality test of the receive mixer. The proposed method minimizes additional chip area consumption and does not generate any more power for test signal creation.

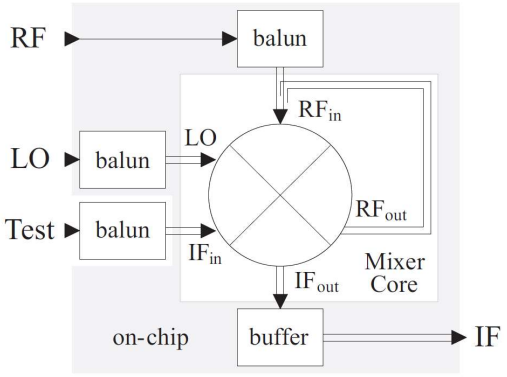


Fig. 4. General scheme of the mixer used simultaneously for up- and down conversion for built-in functionality test (Adapted from [2])

As shown in the above Fig. 4, a direct down-conversion mixer built-in into the system. By using on-chip LC baluns, both single-ended LO and RF input signals are converted to differential signals. The mixer core is driven by an external LO signal and is capable of simultaneous up-conversion of the RF_{in} in signal as well as down-conversion of the IF_{in} signal (test signal). As the input test signal appears at the IF output, the up-converted double-sideband signal is differentially fed back into the RF input path. The local oscillator and mixer will be integrated in the future so that the external high-frequency signals would not be needed for the circuit's functionality test. With the help of a low-frequency balun and an external signal source, the differential low-frequency test signal is generated externally. This direct down-conversion mixer used a Silicon Germanium (SiGe) bipolar technology with a transistor operating at a maximum f_{t} of 200 GHz and f_{max} of 250 GHz. The chip which is fabricated occupies an area of $1028\,\text{x}\,1128\,\mu\text{m}^2$ and draws 22 mA from a single

3.3 V supply [2]. It is possible to convert different inputs simultaneously up and down through the proposed mixer architecture without consuming additional power. The output of the mixer is coupled back into the receiver input path to enable built-in functionality testing of the receiver mixer.

2) Drawbacks: A minimum NF of 21 dB is achieved at an IF power of 3 dB without test signal applied to the IF input in the first case. It is noted that the gain of the receiver decreases by 0.5 dB under heavy test signal input power saturation conditions when an additional 1 MHz 30 mV_{peak} test signal is applied, along with an increase in noise figure. When the frequency of the test signal is increased to 1 MHz [2], there is an increase in the test signal output power along with increasing LO power which until saturation.

C. Phased array type architecture in RFIC

1) Architecture: One of the key bottlenecks of RFIC-based phased arrays is the S-parameter testing of so many different channels on a single chip [13]. The BIST should be able to measure each channel individually and with high accuracy it should not reduce the RF performance of the chip or occupy a large area on the RFIC. The antenna ports are coupled using a 20 dB coupler to receive the test signal. Each channel is then individually turned ON and an on-chip I/Q receiver is used to measure the amplitude and phase response of the injected signal [3]. This determines the vector response of the channel.

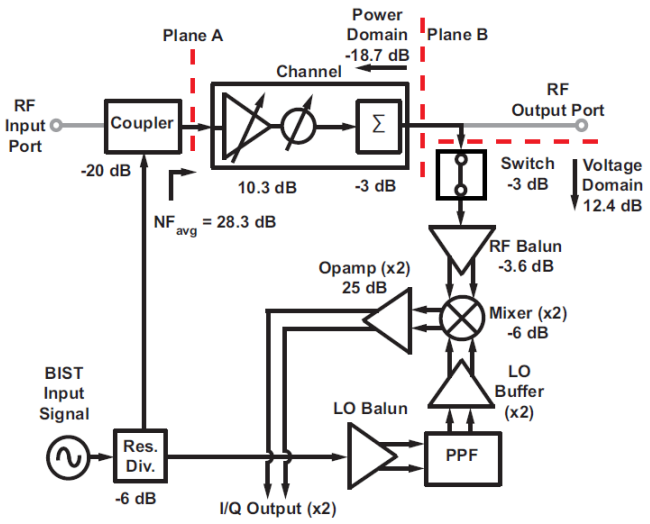


Fig. 5. The BIST phased array chip block diagram showing power or voltage gain of each block (Adapted from [3])

The above Fig. 5. shows that the BIST circuits need to be integrated into a chip, it is implemented without any inductors so that it occupies a small amount of area on the chip. A wideband 3-port resistive splitter with a 6 dB loss is first used to separate the BIST signal between the BIST coupler line and the BIST LO line for the on-chip I/Q receiver. Differential passive mixers made with 0.13 m complementary metal oxide semiconductor (CMOS) serve as the foundation of the I/Q receiver [3]. Operational amplifiers are direct current (DC) connected to the mixer outputs and have a modelled closed-loop voltage gain of 25 dB and a 3 dB bandwidth of 3.15 MHz. The amplitude and phase are calculated with respect to the response obtained from I/Q and they are calculated to obtain the normalized response using the below equation (5) and (6).

$$A = \sqrt{I^2 + Q^2} \tag{5}$$

where A is the amplitde, I is the in-phase element, Q is the quadrature element.

$$\phi = \tan^{-1}(Q/I) \tag{6}$$

where ϕ is the phase, I is the in-phase element, Q is the quadrature element. The BIST system's time domain response has a 98 percent settling time of 340 ns, and as a result, the BIST may function without incident at a 1 MHz rate. As a result, a wide variety of amplitude and phase settings may be tested on an 8-element array in less than 1 ms [3].

2) Drawbacks:

- Antenna impedance is created when connected to a phased array system. This can be reduced by using CMOS switches.
- When using a larger array system it is necessary to carry out proper calibration of amplitude and phase difference between the channels.
- When using a larger array system, additional loss due to power combining network causes an additional 3 dB loss, and this therefore results in a 9 dB more loss for an 8-element array.

III. State-of-art BIST functionalities in MMICs

A. Test Structure

The 143−152 GHz radar transceiver described has a number of self-test and calibration features that make it easier to test the device in production and to repair any deficiencies in the analog front-end [4]. The transceiver was created using available 130 nm SiGe BiCMOS technology, and it has a two channel heterodyne architecture that enables the adoption of a low-IF frequency scheme. In this work, the first fundamental frequency, fully integrated, two channel radar transceiver operating at 143 − 152 GHz is presented. Several BIST features have been included in order to facilitate simple low-frequency, low-cost testing and online monitoring of the circuit performance. First, divide-by-64 chains have been introduced for both the Tx and Rx voltage controlled oscillator (VCO). Apart from allowing the VCO to be locked by external phase locked loop (PLL), the dividers provide a low frequency signal that can be used to independently verify the tuning range and phase noise of the VCO.

The schematic of the bidirectional power detector is illustrated in Fig. 6. This device is based on a coupled-line directional coupler that terminates each output on a bipolar detector with a 10 dB coupling voltage. Due to its larger size and wider bandwidth, coupled-line topologies were preferred

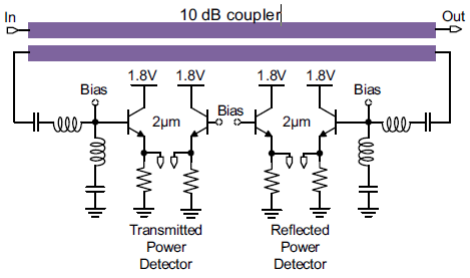


Fig. 6. General scheme of the mixer used simultaneously for up and down conversion for built-in functionality test (Adapted from [4])

instead of ring hybrids. A coupling ratio of $10-\mathrm{dB}$ was selected in order to minimize the impact of the detector on the signal path [4]. By selecting a $10-\mathrm{dB}$ coupling ratio, the detector's impact on the signal path was minimized. The schematic of the proposed, digitally controlled impedance tuner is shown in Fig. 7. A shunt inductor tunes out MOS transistor's parasitic capacitance when they act as two-state resistors. Due to the 45° transmission line, ΓC is rotated by 90° and is rendered purely imaginary, unlike ΓT, which ideally assumes only real values. Therefore, when ΓT and ΓC are varied between 0 and 1, the reflection coefficient at the coupler input equals to the below equation (7).

$$\Gamma_{\text{in}} = \Gamma T + j\Gamma C \tag{7}$$

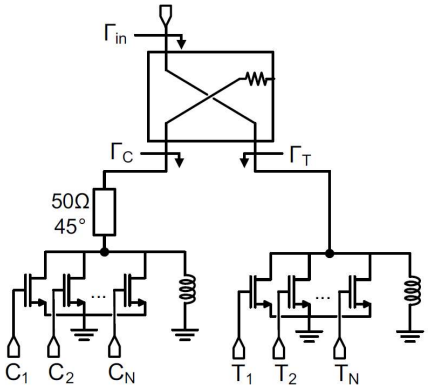


Fig. 7. General scheme of the mixer used simultaneously for up and down conversion for built-in functionality test (Adapted from [4])

In order to measure both forward and reflected waves, bidirectional power detectors have been inserted between the Rx LO distribution tree and the mixers, as well as the Tx LO distribution and the Transmission Amplifier (TA). Their role is to monitor the signal power provided by the LO distribution trees and to isolate the problems raised in the mixer, transmit amplifiers, or the VCO [4]. Both the primary channel and reference channel outputs have similar detectors installed. In addition to measuring the transmit power, the reflection power detector reading can be used to calculate the reflection coefficient at the TX output. A digital impedance tuner has been added at the output of the reference channel to calibrate it as a one-port VNA.

B. Drawbacks

- When used for monitoring internal system nodes, the power detector introduces a very low loss burden. With only a fraction of the power reaching the bipolar detector, which typically exhibits a linear P_{in} - V_{DC}.
- The measured Γ_{in} states of the impedance tuner breakout at $146\,\mathrm{GHz}$. This is due to the nonidealities of the hybrid coupler installed in the tuner as well as the imperfect tuning of the MOS resistor's output capacitance.
- Based on the $6\,\mathrm{dB}$ coupler and the associated interconnect, $7-8\,\mathrm{dB}$ loss between the transmit amplifier output and antenna pad is expected between the antenna port and the power sensor.

IV. Comparison of different architecture

A. Operating frequency and application

- Down-conversion mixer architecture [2] offers an operating frequency from $200-250\,\mathrm{GHz}$. This is the highest compared to other BIST approaches. This architecture is used in automotive radar applications.
- Phased array type RFIC architecture [3] operates in $90-100\,\mathrm{GHz}$ and are used in wireless communication applications.
- Direct conversion type architecture operates at $77\,\mathrm{GHz}$ and are used in radar applications.

B. Power dissipation and chip size

Gain and noise figures are the performance parameter required to be measured. Direct conversion type of architecture [1] requires the generation of the test signal for measurement, resulting in additional components occupying a large area on the chip, dissipating more power. Whereas in the down-conversion mixer type of architecture [2], test signal is generated internally using the input signal, eliminating the use of extra components which keeps the circuit area ($2.6\,\mathrm{mm}\,\mathrm{x}\,2.3\,\mathrm{mm}$) and power dissipation ($3.3\,\mathrm{V}$) at minimum. In paper [4], the architecture uses power detector to measure gain and measures transmitted with a total power consumption of $800\,\mathrm{mW}$.

V. Conclusion

In this work different circuit design techniques, their individual drawback and challenges were studied. A detailed review of state-of-the-art BIST embedded SiGe CMOS radar transceiver was carried out. In addition to power detectors and impedance tuners, the transceiver incorporates a novel level of self-testing, as well as a new calibration approach.

REFERENCES

[1] D. Kissinger, R. Agethen, and R. Weigel, "Analysis of a built-in test architecture for direct-conversion SiGe millimeter-wave receiver frontends," in Proc. IEEE Instrum. Meas. Techn. Conf., Austin, TX, May 2010, pp. 944–948.

[2] D. Kissinger, H. Knapp, L. Maurer, R. Weigel "A 77-GHz down-conversion mixer architecture with built-in test capability in SiGe technology" Published 13 December 2010 Computer Science 2010 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)

[3] O. Inac, Donghyup Shin, Gabriel M. Rebeiz "A Phased Array RFIC With Built-In Self-Test Capabilities" 2012, IEEE Transactions on Microwave Theory and Techniques

[4] I. Sarkas, M. Girma, J. Hasch, T. Zwick, S. Voinigescu "A Fundamental Frequency 143-152 GHz Radar Transceiver with Built-In Calibration and Self-Test" 2012 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)

[5] D. Kissinger, R. Agethen, R. Weigel "A Versatile Built-In Test Architecture for Integrated Millimeter-Wave Radar Receiver Front-Ends"2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings

[6] D. Kissinger, B. Laemmle, L. Maurer, and R. Weigel, "Integrated test for silicon front ends," IEEE Microw. Mag., vol. 11, no. 3, pp. 87–94, May 2010.

[7] D. Kissinger, H. Knapp, L. Maurer, and R. Weigel, "A 77-GHz down-conversion mixer architecture with built-in test capability in SiGe technology," in Proc. Bipolar/BiCMOS Circuits Technol. Meeting, Austin, TX, Oct. 2010, pp.200–203.

[8] K. H. K. Yau, E. Dacquay, I. Sarkas, and S. P. Voinigescu, "Device and IC characterization above 100 GHz," IEEE Microw. Mag., vol. 13, no. 1, pp. 30–54, Jan. 2012.

[9] O. Inac, D. Shin, and G. Rebeiz, "A phased array RFIC with built-in self-test capabilities," IEEE Trans. Microw. Theory Tech., vol. 60, no. 1, pp. 139–148, Jan. 2012.

[10] B. Laemmle, K. Schmalz, C. Scheytt, D. Kissinger, and R. Weigel, "A 122 GHz multiprobe reflectometer for dielectric sensor readout in SiGe BiCMOS technology," in IEEE Compound Semicond. Integr. Circuits Symp. Tech. Dig., Waikoloa, HI, Oct. 2011, pp. 1–4.

[11] ——, "A 62 GHz reflectometer for biomedical sensor readout in SiGe BiCMOS technology," in IEEE Topical Meeting on Silicon Monolithic Integr. Circuits in RF Syst. Dig., Santa Clara, CA, Jan. 2012, pp. 45–48.

[12] D. Kissinger, B. Laemmle, L. Maurer, and R. Weigel, "Integrated test for silicon front ends," IEEE Microw. Mag., vol. 11, no. 3, pp. 87–94, May 2010.

[13] G. M. Rebeiz and F. E. van Vliet, Lunch Panel Session, "SiGe/CMOS RFIC phased arrays: will they be used in defense and commercial systems?" in IEEE MTT-S Int. Microw. Symp., Boston, MA, Jun. 2009.

Vertical GaN-Based Power Transistors

The Department of Electrical Engineering
Institute for Robust Power Semiconductor Systems
University of Stuttgart, Germany

***Abstract*—Due to characteristics like wide band gap (WBG), high breakdown field strength, high electron velocity, and high thermal conductivity, Gallium Nitride (GaN) forms a superior choice for high-frequency and high-power devices. Based on GaN material, vertical transistors have recently emerged as a promising technology in power electronics. This paper reviews the most significant vertical GaN transistor technologies demonstrated throughout the literature. Further, a performance comparison of GaN-based lateral and vertical transistors is presented to emphasize the respective advantages and limitations. Finally, critical challenges in the advancement of vertical GaN transistors are discussed.**

***Index Terms*—Vertical GaN transistor, GaN-on-Si, GaN-on-GaN, GaN-on-Sapphire.**

I. Introduction

Since the beginning of semiconductor technology, Silicon (Si) has been the preferred material for power electronics. However, in recent times, Si-based devices have reached fundamental material limits [1]. Thus, materials such as Gallium Nitride (GaN) with superior properties have emerged as a leading choice. WBG materials like GaN allow devices to be optimized with thinner drift regions due to the higher breakdown field which reduces the lower specific ON-resistance (R_{ON}) in power devices. The R_{ON} is further reduced by the high mobility of GaN. Thus, a given current capability is achievable with a smaller die size. This further lowers the input and output capacitances. Also, GaN enables faster switching transients due to its higher saturation velocity and lower capacitances [2]. Due to these distinct advantages, GaN-based materials have led to the development of a variety of next generation high-performance power devices. The focus of this paper is on a specific category called, GaN-based vertical power transistors.

Section II discusses the rationale for using Vertical GaN transistors over conventional ones. The major vertical GaN transistor technologies are presented in Section III. Section IV compares performances of vertical and lateral GaN transistors based on key parameters. Finally, drawbacks associated with vertical GaN devices and various challenges in their further development are addressed in section V.

II. Motivation for Using Vertical GaN Transistors

In [3], high-performance lateral GaN power devices with low ON-resistance and large breakdown voltage (V_{BR}) were demonstrated. The devices, though, had a major drawback. As the gate-to-drain spacing increased, the V_{BR} scaled proportionally. This in turn demanded a larger device area for high voltage operation. Since, the thickness and quality of the buffer layers determined the V_{BR}, the overall complexity and cost of epitaxial growth increased for high voltage devices. Another important drawback of this category was highlighted in [4] [5]. The trap states and high electric fields present at the surface greatly affected lateral devices, leading to current collapse and other reliability problems. The GaN vertical devices, however, do not have these issues. For them, the electric field peaks far away from the surface. Also, the V_{BR} is depended on the drift layer thickness, immaterial of the device area [6]. Vertical transistors on bulk GaN substrates have been shown to exhibit ideal ON- and OFF-state performances. This is made possible by two main factors: low defect density and the ability to homo-epitaxially grow thick layers on substrates [7]. Due to the vertical nature of the current flow, fully vertical designs are not affected by the current crowding. Therefore, they possess a much larger current capability and a significantly smaller R_{ONsp}.

III. Vertical GaN transistor technologies

In the recent development of vertical GaN transistor technology, devices based on substrates such as Si, GaN, and Sapphire have been proposed. This section discusses the fabrication processes and the performance characteristics of those devices in detail.

A. GaN-on-Si

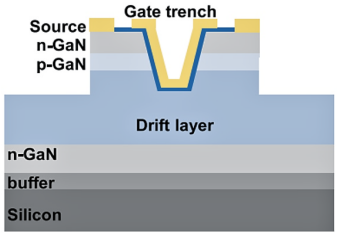


Fig. 1. Device structure of fully-vertical GaN-on-Si MOSFET [3].

Bulk GaN substrates are still very expensive and available only in small diameters. However, the recent advancement in the hetero-epitaxial growth of thick GaN layers on large area Si substrates [3] offers a cost-effective platform. The work in [3] compared two designs, namely, quasi-vertical

and fully-vertical. Since the high voltage GaN-on-Si quasi-vertical transistors were critically limited by current crowding in the bottom n-GaN layer, the compositions required a larger device area. The fully-vertical design, shown in Fig. 1, was not affected by current crowding, therefore offering a much larger current capability and significantly smaller R_{ONsp}. The device showed a high current density (ID) greater than 1.6 kA/cm^2, low R_{ONsp} of 5 mΩcm^2 , high (V_{BR}) of 520 V and excellent transconductance of 300 S/cm^2 despite the absence of field plates. In [8], the design of the device and fabrication have been explained. Using plasma-based dry etching and epitaxial regrowth techniques, the p-GaN CA was constructed.The electrical measurements were performed on both vertical and lateral modes of operation In the vertical configuration device, R_{ON} is relatively higher despite the lower drain current.

B. GaN-on-GaN

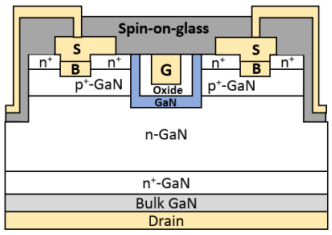


Fig. 2. Cross-sectional OG-FETs schematic for GaN-on-GaN [1].

Excellent switching characteristics like a high blocking voltage, or a low specific R_{ON} or both were shown by vertical GaN transistors on bulk GaN substrates, as reported by [9]. The devices turned out to be potential candidates for high power applications that require high-frequency switching with low switching loss. Further, the cell pitch was miniaturized by stacking the source electrode on the p-body electrode. A field-plate edge termination technique was employed to reduce the electric field at the edge of the p-n junction around the isolation mesa. The result of this was an increase in the blocking voltage and a substantial decrease in the leakage current. The lower doping concentration of the drift layer and the wider cell pitch mentioned in [10], [11], yielded the lower ID and the larger R_{ON} than those of the previously-reported hexagonal trench MOSFET.

In [1], the impact of trench dimensions on V_{BR} and R_{ON} of trench MOSFETs fabricated on bulk GaN substrates was examined. The device structure is presented in Fig. 2. The specifications of the substrate were: threading dislocation density (TDD) = 3X10^6 cm^{-2}, trench depth $\approx$600 nm, and drift region mobility of 850 cm^2/V-s.

C. GaN-on-Sapphire

One of the prominent works in this technology was presented in [12]. The fabrication process was optimized to enhance the ON- and OFF-state performance in the vertical GaN trench MOSFETs. The piranha cleaning process was used prior to the gate dielectric deposition. This reduced

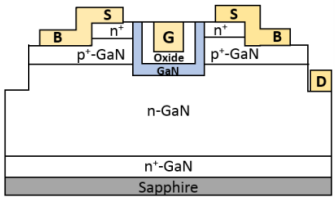


Fig. 3. Cross-sectional OG-FETs schematic for GaN-on-Sapphire [1].

MOS channel interface charges and improved the ON-state device performance. Further, with the help of the thick bottom dielectric process, the V_{BR} of the device was largely enhanced through suppression of the electric field in the gate dielectric close to the gate trench's bottom. This achieved improvement in the OFF-state performance. Also, the suppression of the electric field at the gate trench corner was vital to improving the V_{BR}. An effective solution was introduced that consisted of a thick SiO_2 layer in the gate trench as the field oxide. The use of innovative techniques developed a high performance quasi-vertical GaN-on-sapphire trench MOSFET with a substrate drift layer thickness of 4 µm. A general device architecture of GaN-on-Sapphire is shown in Fig. 3.

Further, the development trends in different device types of vertical GaN transistors are presented through Tables I, II and III. The representative architectures of the CAVET, trench MOSFET, and FINFET device types are shown in Fig. 7.

TABLE I
DEVELOPMENT TRENDS IN CAVET

In 2004 [13] [14], CAVET Similar to a double-diffused MOS	An AlGaN/GaN heterostructure in the channel region, with R_{ON} as low as 2.6 mΩcm^2, V_{th} was negative due to the presence of 2DEG under the gate.
In 2014 [15], First ampere-class demonstration of vertical GaN transistor with max current 2.3 A	V_{BR} was greater than 1 kV. Since, V_{th} was not high enough, the requirement was attained by thinning the gate layer between AlGaN and p-GaN.
In 2016 [16], An advanced structure using a trench CAVET	The use of a p-GaN and semi-polar face gate structure to boost V_{th} to 2.5 V with R_{ON} of 1mΩcm^2 and a carbon doped insulating GaN underneath the channel to improve V_{BR}.
In 2022 [17], Novel Trench-Gated vertical GaN transistor with Dual-Current-Aperture (DCAVET)	A DCAVET creation by inserting p-GaN between two current blocking layer (CBL), allowing E-field distribution in the buried p-n-junction and high V_{BR} of 1504 V and low R_{ON} of 0.77 mΩcm^2, BFOM = 2.94 GW.cm^{-2}.

IV. PERFORMANCE COMPARISON OF LATERAL AND VERTICAL GAN TRANSISTORS

For low and medium power (up to 10 kW) devices, heteroepitaxial growth is suited and the lateral topologies are preferred. However, for applications demanding higher power

TABLE II
DEVELOPMENT TRENDS IN TRENCH MOSFET

In 2008 [18] Vertical GaN trench MOSFET normally-off operation	V_{th} of over 3 V and V_{BR} over 1 KV obtained using CMOS gate structure.
In 2015 and 2016 [11] Vertical GaN trench MOSFET	Low R_{ON} of <2 mΩcm^2 over 20 A operation
In 2022 [12], Quasi-vertical GaN trench MOSFET	Fabrication process optimization to improve ON and OFF state performance, as detailed in section III-C. R_{ON} is 0.95 mΩcm^2, I_d is 3.4 kA/cm^2, V_{th} is 6.1 V, V_{BR} is 485 V.

TABLE III
DEVELOPMENT TRENDS IN FINFET

In 2016 and 2017 [7], [19], Vertical GaN transistors using Fin structures	The current control via a fin-shaped n-GaN channel including all gate stacks allowing normally OFF operation without p-GaN layer and epitaxial regrowth. V_{BR} is 1.2 KV, V_{th} of 0.8 V and low R_{ON} of 1 mΩcm^2.
In 2017 [7], With sub-micron fin shaped channels	R_{ON} is 0.2 mΩcm^2, exceptionally high ON ID, V_{BR} above 1200 V, normally-off operation with BFOM = 7.2GW·cm^{-2}, stable V_{th} up to 150°C.
In 2018 and 2019 [20] Large-Area GaN Vertical FinFET	R_{ON} of 2.1 mΩcm^2, V_{th} of 1.3 V, a 1.2 kV, 5 A FinFET was demonstrated in a chip area of 0.45 mm^2.
In 2021 [21] New interfin structure to reduce parasitic capacitance and increase drift region discharging speed	Three new interfin designs: 1. Oxide full-filling[FF], 2. Reduced fin to fin spacing[RFS]. 3. Split-gate[SG]. The design structures improve parasitic capacitance while causing switching loss.

(>10 kW), higher V_{BR} (>1.2 kV) is required. In this case, the lateral topology becomes costly and has lower manufacturability. This is due to the very large chip area required by the V_{BR} at the given current level (typically over 20 A) [22]. In contrast to lateral topology, vertical topology is economical and feasible for this range of power applications. Additionally, the difference between the source, drain, and gate positions in the lateral versus vertical device structure is shown in Fig. 4.

The following four different possible combinations of the

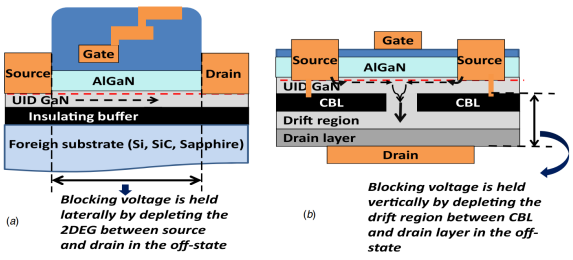


Fig. 4. (a) A lateral AlGaN/GaN power HEMT (b) A vertical transistor using AlGaN/GaN layer structure on bulk GaN drift layer and substrate [22].

vertical vs. lateral topology and the MOS vs. 2DEG channel: 1) p-GaN gate HEMT (lateral [L], 2DEG channel), 2) MOS channel HEMT (L, MOS channel), 3) current aperture vertical electron transistor (CAVET) (vertical [V], 2DEG channel), and 4) U-shaped trench-gate MOS field-effect transistor (UMOSFET) (V, MOS channel). The work in [23], compared the performances by analytical calculation and numerical simulation.

A. Based on Specific On Resistance

Fig. 5 and Fig. 6 summarize R_{ONsp} for 600 V and 1200 V GaN power FETs of different channel lengths [23]. Out of the two cases, CAVETs have the smallest R_{ONsp}. R_{ONsp} of the UMOSFET and the p-GaN gate HEMT at a rating of 600 V are close. When the length of the channel is (L_{ch}) >1.2 μm, the UMOSFET is better than p-GaN. At 1200 V though, the UMOSFET has half the R_{ONsp} than that of the p-GaN gate HEMT. R_{ONsp} of the MOS channel HEMTs is the highest for both voltage ratings. With the decrease in the channel length, R_{ONsp} of MOS channel HEMTs reaches that of the 2DEG channel HEMTs, while for UMOSFETs, it approaches the CAVETs. This is caused by the relatively smaller significance of the (R_{ch}) component resistance. At 600 V, the major component of the MOS channel is R_{ch}. At 1200 V, it is the R_{drift}. Drift region resistance dominates for both voltages in 2DEG channel HEMT and CAVET. In the case of MOS channel HEMT, the channel resistance is dominant.

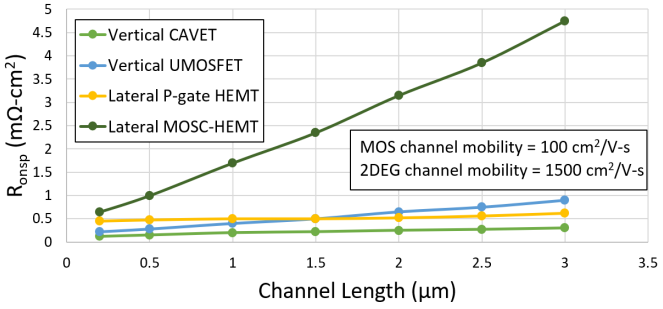


Fig. 5. R_{ONsp} versus L_{ch} for 600 V GaN power FETs [23].

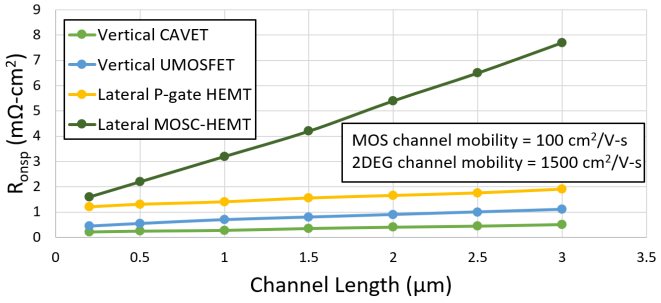


Fig. 6. R_{ONsp} versus L_{ch} for 1200 V GaN power FETs [23].

B. Based on Switching waveforms

In [23], a detailed comparison of switching waveforms for devices was presented at 600 V and 1200 V. The R_{ON}*Q_G

figure-of-merit as well as gate voltage (VGS) swing are observed to be related to switching time and energy loss. Smaller R_{ON}*Q_G and greater VGS swings, in general, imply higher switching performance, i.e., smaller E_{total} and shorter T_{off}. Due to tighter coupling between gate and drain terminals that face each other, vertical HEMTs have a bigger Q_{Gsp} than lateral HEMTs. MOS channel HEMTs have competitive Q_{Gsp} values, but their performance is inferior to the other devices due to significantly greater R_{ONsp}. P-GaN gate HEMTs (lower Q_{Gsp}) and CAVETs have a smaller R_{ON}*Q_G and thus greater switching performance than UMOSFETs at 600 V. Because of the improvement in the UMOSFET's R_{ON}*Q_G compared to the 600 V rating, the UMOSFET outperforms the p-GaN gate HEMT at 1200 V. Although the R_{ON}*Q_G of the p-GaN gate HEMT is lower than that of the UMOSFET, its VGS swing is smaller. Because all devices have the same R_G in switching circuits, a larger VGS swing results in a higher IG and hence a faster switching speed and lower energy loss.

V. Challenges and Drawbacks of Vertical GaN Devices

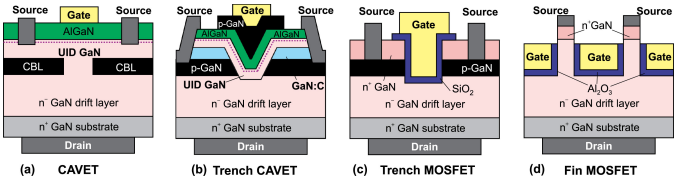


Fig. 7. Schematic of representative vertical GaN transistors: (a) current aperture vertical electron transistor (CAVET), (b) trench CAVET with V-shaped regrowth channel, (c) trench metal-oxide-semiconductor field effect transistor (MOSFET), (d) vertical fin MOSFET [24].

A. Current aperture vertical electron transistor (CAVET)

Designing a robust CBL to decrease leakage current in the vertical direction is one of the CAVET's major challenges. The aperture region in this situation necessitates dry-etching of the Mg-doped GaN and regrowth of the n-GaN channel. It has been reported that etching the aperture region exposes non-c planes, which can result in a non-planar surface if regrowth occurs on such facets. This causes a lot of gate leakage current, according to [13]. Other research has found that significant concentrations of n-type impurities tend to settle in the aperture region, causing an increase in the channel's peak electric field and a decrease in the device V_{BR}. To avoid these problems, a selectively Mg-ion-implanted GaN layer was employed as the CBL, with no GaN regrowth in the aperture region.

Another drawback with the original CAVET structure [Fig. 7.(a)] is its normally-on operation and weak pinch-off functionality, according to [15]. Using a trench CAVET structure, Shibata et al. proposed a new vertical GaN transistor [16]. Regrowth of p-type gate/AlGaN/GaN epitaxial layers across the V-groove is present in the device [Fig 7.(b)].

B. Vertical trench MOSFET

The vertical trench MOSFET [Fig 7.(c)] is the second form of vertical GaN transistor. Obtaining a stable positive V_{th} for usually off operation is one of the key problems of trench MOSFETs. The GaN MOSFET channel is built on the (0001) plane, which has both spontaneous and piezoelectric polarization effects. Because of the charges produced at the GaN/dielectric contact, this causes the MOSFET to have poor pinch-off performance. As a result, trench etching on (0001) planar GaN to create a non-polar surface seems appealing. Cl_2/BCl_3 gases are used to execute dry etching for trench creation. The dry-etched surface, however, is unsuitable for the gate channel region because of increased surface roughness and ion bombardment damage. As a result, post-treatment of the etched surface is required to smooth the surface and remove the damage, according to [24]. Using Tetramethylammonium hydroxide(TMAH) wet etching and piranha clean, a damage-free corner rounding method was recently developed. The gate stability of GaN-based MOSFETs is always an issue due to the lack of sufficient native oxide. A MOCVD-grown AlN/SiN dielectric stack was used. In [10], atomic layer deposited (ALD) SiO_2 was used as the gate dielectric to guarantee better step coverage.

C. Vertical Fin MOSFET

A general device architecture is shown in [Fig. 7.(d)]. A Cl_2/BCl_3-based dry etch is used to provide smooth sidewalls in the fin, followed by a wet etch using hot TMAH [25] [19]. When the fins were aligned to the m-plane (<>1100), it was discovered that the sidewalls are quite rough, because wet etching of GaN is reliant on crystal orientation. A strong peak electric field is present at the edge of the gate and at the bottom of the fins in the GaN vertical fin transistors while they are in off-state operation. The V_{BR} is limited to 100 V without an edge termination arrangement. The V_{BR} is boosted to 400 V with the addition of a gate field plate. The failure of the gate dielectric due to a concentrated electric field at the edge is thought to be the breakdown mechanism in these systems. A 100 nm thick oxide layer was created in the etched trench to lower the electric field in the ALD Al_2O_3 gate dielectric, preventing premature deterioration.

Conclusion

In this review, several key aspects of vertical GaN power transistors, such as device architectures, fabrication methods, and processing issues, are summarized, proving vertical GaN devices are promising candidates for use in not only high-power but also high-frequency applications. To date, vertical GaN transistors have been mostly fabricated on Si substrates due to their low cost. Further research is needed for improvement of the blocking voltage and high-purity epitaxial layers, indispensable for the drift layers. In the future, the superiority of GaN material combined with device processing optimization will allow the commercialization of vertical GaN-based power devices on a wider scale.

REFERENCES

[1] C. Gupta et al., "Impact of Trench Dimensions on the Device Performance of GaN Vertical Trench MOSFETs", IEEE Electron Device Letters, vol. 38, no. 11, pp. 1559-1562, Nov. 2017.

[2] E. A. Jones, F. F. Wang and D. Costinett, "Review of Commercial GaN Power Devices and GaN-Based Converter Design Challenges", IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 4, no. 3, pp. 707-719, Sept. 2016.

[3] R. A. Khadar, C. Liu, R. Soleimanzadeh and E. Matioli, "Fully Vertical GaN-on-Si power MOSFETs", IEEE Electron Device Letters, vol. 40, no. 3, pp. 443-446, March 2019.

[4] T. Mizutani, Y. Ohno, M. Akita, S. Kishimoto, and K. Maezawa, "A study on current collapse in AlGaN/GaN HEMTs induced by bias stress", IEEE Trans. Electron Devices, vol. 50, no. 10, pp. 2015–2020, Oct. 2003.

[5] G. Meneghesso et al., "Reliability of GaN high-electron-mobility transistors: State of the art and perspectives", IEEE Trans. Device Mater. Rel., vol. 8, no. 2, pp. 332–343, Jun. 2008.

[6] B. Baliga, "Trends in power semiconductor devices", IEEE Trans. Electron Devices, vol. 43, no. 10, pp. 1717–1731, Oct. 1996.

[7] Y. Zhang et al., "1200 V GaN vertical fin power field-effect transistors", 2017 IEEE International Electron Devices Meeting (IEDM), pp. 9.2.1-9.2.4, 2017.

[8] D. Biswas, N. Torii, K. Yamamoto and T. Egawa, "Demonstration of a Fully-Vertical GaN MOSFET on Si", Compound Semiconductor Week (CSW) 2019, pp. 1-2, 2019.

[9] T. Oka, T. Ina, Y. Ueno and J. Nishii, "Over 10 A operation with switching characteristics of 1.2 kV-class vertical GaN trench MOSFETs on a bulk GaN substrate", 28th International Symposium on Power Semiconductor Devices and ICs (ISPSD), pp. 459-462, 2016.

[10] T. Oka, Y. Ueno, T. Ina, and K. Hasegawa, "Vertical GaN-based trench metal oxide semiconductor field-effect transistors on a free-standing GaN substrate with blocking voltage of 1.6 kV", Appl. Phys. Express, vol. 7, 021022, 2014.

[11] T. Oka, T. Ina, Y. Ueno, and J. Nishii, "1.8 mΩcm^2 Vertical GaN-based trench metal oxide semiconductor field-effect transistors on a free standing GaN substrate for 1.2 kV-class operation", Appl. Phys. Express, vol. 8, 054101, 2015.

[12] R. Zhu, H. Jiang, C. W. Tang and K. M. Lau, "Enhancing ON- and OFF-State Performance of Quasi-Vertical GaN Trench MOSFETs on Sapphire With Reduced Interface Charges and a Thick Bottom Dielectric", in IEEE Electron Device Letters, vol. 43, no. 3, pp. 346-349, March 2022.

[13] I. Yaacov, Y. Seck, and U. Mishra, "AlGaN/GaN current aperture vertical electron transistors with regrown channels", Journal of Applied Physics,vol. 95, pp. 2073-2078, 2004.

[14] M. Kanechika et al, "A Vertical Insulated Gate AlGaN/GaN Heterojunction Field-Effect Transistor", Jpn. J. Appl. Phys., vol. 46, L503, 2007.

[15] H. Nie et al., "1.5-kV and 2.2-m Ω -cm ^{2} Vertical GaN Transistors on Bulk-GaN Substrates", IEEE Electron Device Letters, vol. 35, no. 9, pp. 939-941, Sept. 2014.

[16] D. Shibata et al., "1.7 kV/1.0 mΩcm^2 normally-off vertical GaN transistor on GaN substrate with regrown p-GaN/AlGaN/GaN semi-polar gate structure", IEEE International Electron Devices Meeting (IEDM),pp. 10.1.1-10.1.4, 2016.

[17] P. Bai et al., "A Novel Trench-Gated Vertical GaN Transistor With Dual-Current-Aperture by Electric-Field Engineering for High Breakdown Voltage", IEEE Transactions on Electron Devices, vol. 69, no. 3, pp. 1219-1225, March 2022.

[18] H. Otake, K. Chikamatsu, A. Yamaguchi, T. Fujishima, and H., "Vertical GaN-Based Trench Gate Metal Oxide Semiconductor Field-Effect Transistors on GaN Bulk Substrates", Ohta, Appl. Phys. Express 1, 011105, 2008.

[19] M. Sun, Y. Zhang, X. Gao and T. Palacios, "High-Performance GaN Vertical Fin Power Transistors on Bulk GaN Substrates", IEEE Electron Device Letters, vol. 38, no. 4, pp. 509-512, April 2017.

[20] Y. Zhang et al., "Large-Area 1.2-kV GaN Vertical Power FinFETs With a Record Switching Figure of Merit", IEEE Electron Device Letters, vol. 40, no. 1, pp. 75-78, January 2019.

[21] H. Wang, M. Xiao, K. Sheng, T. Palacios and Y. Zhang, "Switching Performance Analysis of Vertical GaN FinFETs: Impact of Interfin Designs", IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 9, no. 2, pp. 2235-2246, April 2021.

[22] S. Chowdhury, B. L. Swenson, M. H. Wong and U. K Mishra, "Current status and scope of gallium nitride-based vertical transistors for high-power electronics application", Semiconductor Science and Technology, vol. 28 074014, 2013.

[23] Z Guo et al, "Comparative performance evaluation of lateral and vertical GaN high-voltage power field-effect transistors", Jpn. J. Appl. Phys. Vol. 58, SCCD09, 2019.

[24] J. Hu, Y Zhang, M Sun, D Piedra, N Chowdhury and T Palacios, "Materials and processing issues in vertical GaN power electronics", Materials Science in Semiconductor Processing, Vol. 78, pp. 75-84, May 2018.

[25] Y. Zhang et al., "Origin and Control of OFF-State Leakage Current in GaN-on-Si Vertical Diodes", IEEE Transactions on Electron Devices, vol. 62, no. 7, pp. 2155-2161, July 2015.

Analysis of Dielectric Materials in Insulated Metal Substrates

Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—With the continuous downscaling of electronic devices, all existing components require new perspectives and modifications to work efficiently at the new sizes. One of the major issues in power electronics devices is the heat dissipation due large output power in tiny spaces. Insulated metal substrates (IMS) provide significant benefits in overcoming this challenge. This paper looks at the various materials used in the fabrication of dielectric layers in IMS for power electronics. Analysis with respect to thermal, electrical and mechanical properties and evaluation of state-of-the-art IMS dielectrics is provided. The effects of using anodic aluminum oxide (AAO) in the base layer are also briefly discussed. Reducing the thickness of the dielectric layer resulted in higher thermal conductivity meanwhile using a layer of AAO reported better protection from environmental factors such as friction and abrasion. Ways to minimize heat losses in the substrate due to oxidized aluminum layer are topics of future scope.**

***Index Terms*—Insulated Metal Substrates, Anodic Aluminum Oxide, dielectrics, thermal conductivity**

I. Introduction

Power density is one of the most important terms under consideration at present times, especially since the aim is to produce complex, efficient and lightweight systems. According to the literature survey in [1], the power density is expected to increase seven fold compared with 2020 targets. The area of applications such as electric vehicles (EV), high speed H-bridge converters [2] and electro-hydrostatic actuators [3] show great prospects for improved power electronic devices.

There have been many developments in the design and fabrication technologies of power electronic modules to reduce the heat dissipation caused due to various factors such as high density of semiconductor chips due to miniscule size and high input power required to drive the load as highlighted in [1]. Currently several substrates are available in the market which have shown remarkable results to improve heat transfer inside the module. A few examples are direct bonded copper (DBC) [1], FR4 printed circuit boards (PCB) [4][5], active metal brazed (AMB) [4] and IMS [6]. A typical metal substrate comprises of a metal layer at the top, mostly made of copper and aluminum, and a base metal layer which conducts the heat from the above surfaces to the environment outside. The base layer can be made of any metal however considering that copper and aluminum provide high conductivities and are easily available for cost effective manufacturing, they are usually the best choice [7]. The space in between these two metal layers is filled with a dielectric which can be ceramic or epoxy made from composite materials.

A figurative comparison between a conventional power module and IMS structure is shown in Fig.1. DBCs have a ceramic layer as the insulator as opposed to a polymer-based film in IMS. The former consists of fixed layouts whereas IMS allows for flexible layer thickness for conductive and dielectric layer. Although the former provides higher internal performance with respect to the latter, they have limited heat extraction capabilities versus the optimized multilayer layout in IMS [1]. Therefore, from the literature survey in [1][8][6], it can be assumed that dielectric layer, which is sandwiched between the aluminum base and circuit layer, refer Fig.2, plays an important role in the increasing module's heat dissipation capabilities.

The purpose of this paper is to provide a brief review of state-of-the-art IMS dielectrics and to evaluate the environmental and boundary conditions crucial for IMS substrates. A brief description of test methods for thermal, electric and mechanical properties of the dielectrics is also provided. Lastly, the effects of using anodized aluminum in IMS, are discussed.

II. IMS dielectrics

Thermal conductivity can be defined as the ability of any material, solid, liquid or gas, to transfer heat through a material using conduction [9]. It is denoted by λ, k or κ and is measured in Watts per meter Kelvin (W/mK) [4]. The isotropic thermal conductivities of copper and aluminum used for the IMS boards can be easily retrieved, as will be explained in section III, and are usually ranging from 160 to 380 W/mK [8]. Since the circuit layer is bonded to this thin, thermally conductive and electrically insulating layer, it is of essence that investigation has to be done to understand the effect of dielectric thermal conductivity on the substrate in general.

Also as observed from the data in [1]-[10], the thermal conductivity of the insulator layer is responsible for the effective heat dissipation qualities of the substrate. Unlike in DBC substates, IMS have a thin dielectric layer attached above the base, which acts as the heat sink. This eliminates the requirement for an additional cooling system which would have been needed to extract the internal heat. The thermal conductivity of an IMS generally fluctuates between 3 to 7 W/mK, depending on

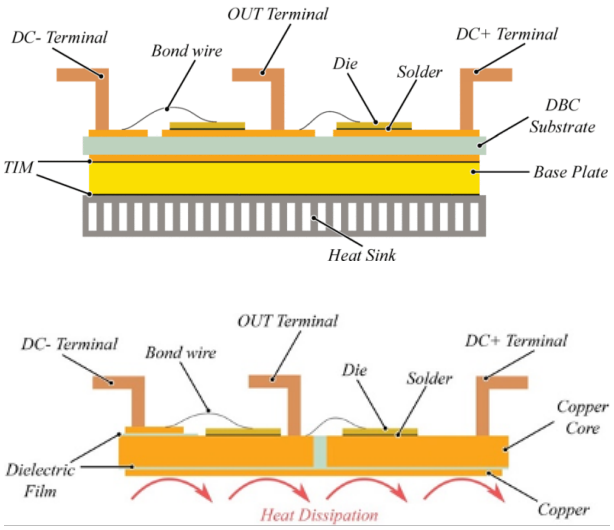


Fig. 1. Illustration of conventional power module(top) and IMS power module(bottom) [1]

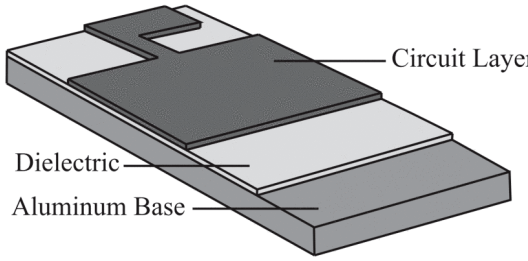


Fig. 2. Cross-section of an IMS structure [6]

the thermal properties of the dielectric material incorporated in the substrate [6][11]. Higher the thermal conductivity of the dielectric material, the lower the junction temperature of the die. However apart from the thickness of the dielectric, which ranges between 40 to 125 µm [6][12], the conductor layer thickness also has an impact on the thermal conductivity of the dielectric. According to experimental data in [11] the thermal conductivity values range between 1 to 5 W/mK, depending on the dielectric layer thickness, thermal conductivity, thickness and material composition of the base layer.

Table I presents few state-of-the-art dielectric materials currently used for IMS in power electronics for high voltage applications. It can be observed that by differing the composition of dielectric materials and the layer thickness, the thermal conductivity varies from 2.2 to 4 W/mK. It can also be of interest to see how different thicknesses leads to same value of thermal conductivity as is evident in case of HT-07006 and Ventec V- 4A2. Hence, it can be inferred from the results of [1][13][12][14], that a thin dielectric can increase the thermal conductivity values to as high as 4 W/mK.

III. Experimental and Boundary Conditions for IMS

In order to modify the thickness of the dielectric layer to obtain application-specific results, it is important to understand

TABLE I
State-of-the-art dielectrics used in IMS

Dielectric Material	Standard thickness (µm)	Thermal conductivity(W/mK)
HT-07006[1]	152	2.2
GPS[13]	150	4
HT04503[12]	100	3
Ventec V- 4A2[14]	177.8	2.2

the parameters that are involved in the calculations. Out of all the factors listed in [7], it can be inferred that thermal conductivity of the substrate materials is one of the most important selection factors. The thermal conductivity of a material in one dimensional system can be calculated using Fourier's law as given in (1)

$$Q = -kA\frac{dT}{dx} \tag{1}$$

Although approximate values of thermal resistance for simple thermal models can be calculated using the Fourier's law, for real-time systems, three-dimensional model is needed to be considered [8]. Hence it is of utmost importance that we use methods that measure this precise multi-directional heat flow from one-layer to another. Multiple approaches have been tested for the measurement of thermal resistance and conductivity of the dielectric material used in the metal substrates [1][6][13][15]. One way to look at this is by considering that the total thermal resistances of substrate is sum of all the resistances from different layers [6]. It can be summarized as

$$R_{total} = R_{top} + R_{dielectric} + R_{bottom} \tag{2}$$

Where, R_{top} is the resistance of top circuit layer, R_{bottom} is the resistance of base layer and $R_{dielectric}$ is the resistance of dielectric layer

Also, resistance in the layer can be expressed in terms of thickness of the layer and its thermal conductivity as

$$\frac{t_{total}}{\lambda_{total}} = \frac{t_{top}}{\lambda_{top}} + \frac{t_{dielectric}}{\lambda_{dielectric}} + \frac{t_{bottom}}{\lambda_{bottom}} \tag{3}$$

Where, t is the thickness and λ is the thermal conductivity of the respective layer.

The contact resistances of copper-aluminum can be figured out using a stack of multiple copper-aluminum pairs and by calculating the slope of thermal impedances versus number of pairs. In IMS, the base layer defines the thickness of the substrate as the thickness of other two layers combined is very small as compared to base layer. Hence it is safe to assume that the modulus of elasticity for an IMS is approximately equal to that of the base layer [6]. According to the derivations in [6] the boundary condition for choosing copper-aluminum pairs is expressed as,

$$E_{Cu} \cdot t^3 = E_{Al} \cdot t^3 \tag{4}$$

As the dielectric material is usually a composite mixture of epoxy and ceramic particles [15], it is usually difficult to estimate accurate thermal conductivity values however the

in-situ measurement method [15], which requires soldering thermal test chip (TTC) on a substrate, provides a way to evaluate an almost-precise value. Using the above-mentioned approach, it resulted in approximately a 7% error in the IMS under test, which gives good estimations of temperature distributions of real power systems. A total power dissipation range of 10W -20W is considered as the boundary condition in different thermal test setups [12][15]. For IMS the thermal resistance range is $5 \cdot 10^4$ square meters Kelvin per Watt $\mathrm{m^2K/W}$ [15], hence optimization of thermal impedance and reduction of thickness plays a vital role in improving the performance of an IM substrate.

Another approach considers using multiple SiC MOSFET dies having different designs, for the testing. The maximum junction temperature in IMS in this method is 175°C meanwhile the base temperature of the substrates is kept at 65°C which is a typical temperature in standard cooling systems [1]. The steady state analysis considered each die as an individual heat source with a constant heat source on the top surface. The specifications of the substrates are discussed in detail in [1]. One interesting finding is the presence of a dielectric separator in one of the tested dies which verifies to improved heat dissipation and consequently lower junction temperature of the substrate. Upon comparing the steady-state and transient state analysis of DBC substrates with IMS, the thermal impedance in the latter improved by up to 40% [1]. These are some significant results, however there is one drawback of this approach, due to the requirement of high torque at low speeds, SiC MOSFETs cannot be utilized. Although within the frequency range of 100 Hz range, the IMS is reported to provide lower junction temperatures which will eventually result in improved stability and lifetime of the module.

Hence, choosing the right boundary conditions for the test setup plays an important role in the overall results achieved. However, to validate the results and for a precise analysis of the test assemblies, it is important to do a modelling simulation of the same. The next section describes some test methods and tools that are useful for evaluation of thermal, mechanical and electrical properties of different materials used in power electronics.

IV. Test methods for steady state and transient state analysis

As described in section III. various test setups have been used to study the thermal properties of the dielectric materials. Here we try to provide a summary (see Table II of all the different analysis models and tools that can be useful for evaluating the thermal characteristics of the substrate and its materials. Most of the studies provided a comparison of steady and transient state analysis between IMS and other substrates like DBC,FR4 etc. which used finite element analysis simulation software like COMSOL Multiphysics [1], Cauer model [4] and ElectroFlo® (TES International) [12]. Thermal impedance and resistance of the individual layers of the substrates can be evaluated using techniques based on TO-220 method [8], ASTM D5470 and D149 [6][16]. CFD software Flotherm™ [8] can be used to calculate three dimensional heat transfer.

During calculation of state analysis using ICEPAK™ computational fluid dynamics [17], two new environments were considered instead of isothermal boundary conditions which resulted in the conclusion that changes in thermal resistance of the substrate do not have intense effect when the heat transfer methods such as conduction or natural convection are used. Thermal conductivity of solids can also be determined using steady state methods such as Guarded Heat Flow which is reported to provide the most reliable and consistent data while transient state can be obtained using Laser Flash Diffusivity and Hot Disk™ method [8].

Therefore, it is evident from the test results, that there are always some discrepancies between the results of simulation and that of a test setup. One reason for it can be the parasitic capacitances between the layers and another is the assumption that the layers are smooth and matte whereas in reality there are some rough edges between the layers, that are ignored in test bench calculation. Hence in order to reduce losses due to roughness of the layers is through the use of nano-particles, which is briefly discussed in next section.

V. IMS with anodised aluminum

As discussed in the sections prior, it is certain that thermal properties such as resistance and conductivity of the substrate are critical parameters for selection of substrates used for high power applications. A study of variety of substrate materials has been provided in literature [7], out of which aluminum stands out as the most feasible and machinable among better competitors like quartz and sapphire. Among the recent developments in IMS, another one is the process of anodizing the aluminum layer with a thin layer of porous aluminum (see Fig. 3). The process involves submerging aluminum discs in acid based electrolytic solutions which result in a layer of oxide on the outer surface of the metal. There are two processes namely hard anodization (HA) and mild anodization (MA) [18]. The former produces a porous layer while the latter leads to formation of a non-porous layer. Results in [18] show that the layers formed using HA resulted in higher values of hardness and elastic modulus as compared to the layers formed using MA. It is also reported that at relatively higher loads (0.1N and 1N), friction coefficient was heavily influenced by the pore size. Smaller pores resulted in lower friction coefficient. Although AAO did not display any significant tribological performance, they can still be utilized to form a self-lubricating constitute. Hence using oxidized aluminum leads to a thin insulating layer which protects the layer against environmental factors such as abrasion due to friction, corrosion etc. [4][18]. Although usage of anodized oxide layer of aluminum on the base-layer has higher breakdown voltage and provides insulation to the surface layer, it also increases resistance to heat flow in the substrate due to the low conductivity of the porous films [6]. However, the author in [4] provides some suggestions which can help to compensate for the low conductivity issues.

TABLE II
TEST METHODS TO CALCULATE THERMAL CONDUCTIVITY OF MATERIALS

Approach	Top		Bottom		Dielectric		
	Material	*Thickness(μm)*	*Material*	*Thickness (mm)*	*Material*	*Thickness (μm)*	*κ (W/mK)*
TTC method [15]	Cu	70	Al(6082)	2	IMS1	100	1.35
TTC method [15]	Cu	70	Al(6082)	2	IMS2	75	1.37
ASTM D5470 based [6]	Cu	35	Al(5052)	1.5	A	117	1.78
ASTM D5470 based [6]	Cu	70	Al(5052)	1.5	B	84	1.29
Half-bridge structure [1]	Cu	1600	Cu	0.07	HT-07006	152	2.2
IMS based GaN HEMT [12]	Cu	140	Al	1.8	HT-04503	100	3

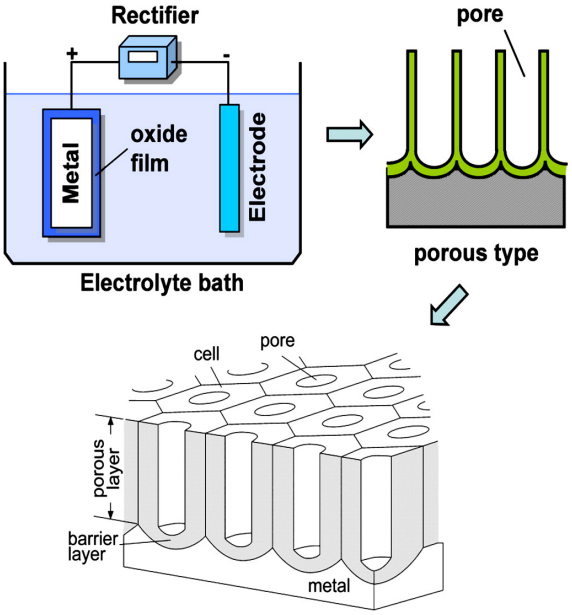


Fig. 3. Process of development of anodized aluminum layer [6]

VI. CONCLUSION

The heat dissipation inside a power module is an important factor that can improve the overall efficiency and power consumption of a system. There are several ways to increase the heat transfer from the substrate to the environment such as adding cooling systems to the power module [19] which increases the surface area for heat flow. Another way is by improving the heat conductivity of the substrate, which results in compact and cheaper systems. As discussed in previous sections, higher thermal conductivity of the dielectric leads to better heat flow throughout the surface and significantly reduces the junction temperature of the die. Currently, IMS are the cheapest and easiest substrates that can be manufactured while simultaneously providing best thermal, electrical and mechanical characteristics among existing technology such as DBC and PCB. An addition of anodized aluminum layer has shown to improve the insulation capabilities of the substrate however it further has its limitations [6][18]. A further topic of research in this area can be about compensating the limitations added by the anodized aluminum layer by varying the physical and thermal properties of aluminum such as porosity or using some other material for insulation.

REFERENCES

[1] E. Gurpinar, B. Ozpineci, and S. Chowdhury, "Design, Analysis, Comparison, and Experimental Validation of Insulated Metal Substrates for High-Power Wide-Bandgap Power Modules," J. Electron. Packag., vol. 142, no. 4, p. 041107, Dec. 2020, doi: 10.1115/1.4047409.

[2] D. Dong, M. Agamy, G. Mandrusiak, and Q. Chen, "Design of high-speed H-bridge converter using discrete SiC MOSFETs for solid-state transformer applications," in 2017 IEEE Energy Conversion Congress and Exposition (ECCE), Cincinnati, OH, Oct. 2017, pp. 1379–1386. doi: 10.1109/ECCE.2017.8095951.

[3] W. Lee, S. Li, D. Han, B. Sarlioglu, T. A. Minav, and M. Pietola, "A Review of Integrated Motor Drive and Wide-Bandgap Power Electronics for High-Performance Electro-Hydrostatic Actuators," IEEE Trans. Transp. Electrification, vol. 4, no. 3, pp. 684–693, Sep. 2018, doi: 10.1109/TTE.2018.2853994.

[4] S. Klarmann, Y. Vagapov, and H. Gotzig, "Modelling and Thermal Analysis of Advanced Insulating Layer of Electronic Applications," in 2018 24rd International Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Stockholm, Sep. 2018, pp. 1–4. doi: 10.1109/THERMINIC.2018.8593297.

[5] R. Polanský, P. Prosr, and M. Čermák, "Determination of the thermal endurance of PCB FR4 epoxy laminates via thermal analyses," Polym. Degrad. Stab., vol. 105, pp. 107–115, Jul. 2014, doi: 10.1016/j.polymdegradstab.2014.03.043.

[6] G. Liu, Y. T. Shi, Y. Z. Tao and S. Cao, "Measurement and significance of the effective thermal conductivity of the dielectric in an IMS™," 2010 5th International Microsystems Packaging Assembly and Circuits Technology Conference, 2010, pp. 1-4, doi: 10.1109/IMPACT.2010.5699481.

[7] S. M. Vaezi-Nejad, "Primary investigation of anodized aluminium as a substrate for hybrid microelectronics: Part 1. Electrical and mechanical properties," Int. J. Electron., vol. 68, no. 1, pp. 59–68, Jan. 1990, doi: 10.1080/00207219008921146.

[8] B. P. Gundale and S. Misra, "Thermal performance of IMS/spl trade/ dielectrics: data and prediction," in ITherm 2002. Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (Cat. No.02CH37258), San Diego, CA, USA, 2002, pp. 1083–1088. doi: 10.1109/ITHERM.2002.1012578.

[9] Tritt, Terry M. Thermal Conductivity : Theory, Properties, and Applications. Vol. 1, New York ; London, Springer, 2011, pp. XXI, 290, doi.org/10.1007/b136496. Accessed 22 June 2022.

[10] M. P. Foster, J. D. Holmes, and D. A. Stone, "Effect of inter-layer interface quality on electrical and thermal characteristics of IMS," in 5th IET International Conference on Power Electronics, Machines and Drives (PEMD 2010), Brighton, UK, 2010, pp. 4–4. doi: 10.1049/cp.2010.0075.

[11] S. Klarmann, B. Manesh, T. Hoenle, and Y. Vagapov, "Analysis of insulated-metal-substrates structures in the context of heat dissipation enhancement," in 2017 Internet Technologies and Applications (ITA), Wrexham, Sep. 2017, pp. 161–164. doi: 10.1109/ITECHA.2017.8101929.

[12] J. L. Lu, D. Chen, and L. Yushyna, "A high power-density and high efficiency insulated metal substrate based GaN HEMT power module," in 2017 IEEE Energy Conversion Congress and Exposition (ECCE), Cincinnati, OH, Oct. 2017, pp. 3654–3658. doi: 10.1109/ECCE.2017.8096647.

[13] N. Yonemura, M. Fukuda and T. Shigi, "Heat analysis on insulated metal substrates [power electronics]," IAS '96. Conference Record of the 1996 IEEE Industry Applications Conference Thirty-First IAS Annual Meeting, 1996, pp. 1407-1410 vol.3, doi: 10.1109/IAS.1996.559249.

[14] F. W. Y. Lau et al., "Programmable High Voltage Distribution for Photodetectors in a 1 mm Resolution Clinical PET System," IEEE Trans. Nucl. Sci., vol. 62, no. 5, pp. 1989–1994, Oct. 2015, doi: 10.1109/TNS.2015.2440437.

[15] X. Jorda, X. Perpina, M. Vellvehi, and J. Coleto, "Power-Substrate Static Thermal Characterization Based on a Test Chip," IEEE Trans. Device Mater. Reliab., vol. 8, no. 4, pp. 671–679, Dec. 2008, doi: 10.1109/TDMR.2008.2005299.

[16] H. Fick, "Design Constraints for the Use of Insulated Metal Substrates®," Microelectron. Int., vol. 10, no. 1, pp. 58–60, Jan. 1993, doi: 10.1108/eb044491.

[17] S. E. Larson and J. Slaby, "Comparison of various substrate technologies under steady state and transient conditions," in The Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena In Electronic Systems (IEEE Cat. No.04CH37543), Las Vegas, NV, USA, 2004, pp. 648–654. doi: 10.1109/ITHERM.2004.1318347.

[18] H. Kim, D. Kim, W. Lee, S. J. Cho, J.-H. Hahn, and H.-S. Ahn, "Tribological properties of nanoporous anodic aluminum oxide film," Surf. Coat. Technol., vol. 205, no. 5, pp. 1431–1437, Nov. 2010, doi: 10.1016/j.surfcoat.2010.07.056.

[19] D. P. Hamilton, S. Riches, M. Meisser, L. Mills, and P. Mawby, "High temperature thermal cycling performance of DBA, AMB and thick film power module substrates," p. 5, 2016.

Temperature Measurement Methods in WBG Power Semiconductors

Satabdi Bastia
Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—Owing to their superior electrical and thermal characteristics, wide band gap devices work in high temperature, high voltage and high frequency environments with better efficiency than Silicon devices. In order to prevent the device from being damaged and to get optimal performance, temperature monitoring is essential since wide band gap devices work at high voltages and currents. These devices' temperatures can be assessed in a variety of ways: either by the use of temperature transducers in direct contact with the device or through the use of a correlation between the device's electrical or optical properties and temperature. The need for quasi threshold voltage and quasi turn-on delay for wide band gap devices is discussed. In this paper, several of these most recent methods are reviewed, categorized and compared on the basis of key performance indicators such as response time, linearity, cost, invasiveness and accuracy.**

***Index Terms*—WBG, thermal shutdown, over-temperature protection, PTAT sensor, calorimetry, overload, thermistor, TSEP**

I. Introduction

Wide band gap (WBG) materials such as Silicon Carbide (SiC) and Gallium Nitride (GaN) have risen into prominence because of their superior electrical and thermal characteristics with respect to Silicon (Si). WBG devices have a higher breakdown electric field. With this property, the device can be highly doped and can be made thinner at the same time. The resulting WBG power semiconductor devices have thinner dimensions than their Si-based counterparts due to smaller drift regions. A semiconductor device's ability to switch at high frequencies is determined by its drift velocity. The drift velocity of the WBG materials is two times more than that of Si as shown in Fig. 1. This allows for higher switching frequencies in WBG devices which reduces the volume of passive components. Higher band gap energy and higher breakdown field allow operation at high voltages. For high voltage ($> 1.2\,\mathrm{kV}$) and high temperature ($> 200^{\circ}\mathrm{C}$) applications, SiC is the ideal choice [1]. GaN devices can perform operations that require switching frequencies in the range of $10\,\mathrm{MHz}$ for power electronic applications [2], [3].

However, due to their high cost, WBG devices require efficient implementation for full exploitation of their performance. WBG devices operate at high voltages and currents which are responsible for significant temperature changes that should be monitored. As temperature increases, thermo-mechanical stress between the device and different layers of the package is introduced due to different thermal expansion coefficients which can result in the malfunctioning of the device. To ensure the reliability of the device, the junction temperature (T_j) has to be monitored carefully. Temperature measurement can be used to design over-temperature protection circuitry and ensure a safe operation of the device by giving feedback to the cooling system so that the cooling system adjusts itself to keep the device temperature within limits [4].

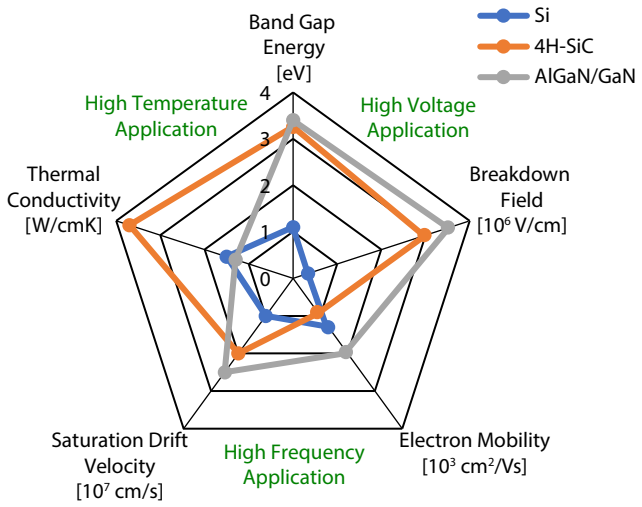


Fig. 1. Si, 4H-SiC, GaN material attributes' comparison (adapted from [3])

Temperature measurement methods can be broadly categorized into three types: physically contacting, optical and electrical methods.

II. Physically Contacting Methods

Physical methods require a sensor to be placed in close proximity to the heat-generating point whose parameters show changes according to the change in the temperature. Calorimetric measurements are required to establish a relationship between power loss due to heat and the change in temperature caused by the power loss. It is also used to develop thermal models for the electrical setup. These require temperature measurements [5]–[8]. The temperature measurements are done via Pt100 sensor [9], [10], thermocouple [5], [11] or PTAT sensor [12]. In addition to Pt100, there are also Pt500, Pt1000, Cu100 and Ni100.

A. Resistance Temperature Detector (RTD)

Platinum resistance thermometers (PRTs) provide good accuracy across a large temperature range starting from $-200\,^{\circ}\mathrm{C}$ to $+850^{\circ}\mathrm{C}$. The technique is based on determining the resistance of a platinum element. At $0^{\circ}\mathrm{C}$, the most common kind of Pt100 has a resistance of $100\,\Omega$ and $138.5\,\Omega$ at $100^{\circ}\mathrm{C}$. The resistance of Pt1000 sensor is $1000\,\Omega$ at $0^{\circ}\mathrm{C}$. The resistance of a Pt100 sensor and a Pt1000 sensor change by $0.385\,\Omega$ and $3.85\,\Omega$ respectively for each $1\,\mathrm{K}$ change in temperature which implies that Pt1000 is more sensitive than Pt100. So, even a slight error in resistance measurement can significantly impact the determination of the temperature. The four-point measuring method is used to get accurate measurements. Again, the sensor is placed very close to the device to be measured and it can only measure the surface temperature. The temperature measurement in this method is affected by the thermal impedance between the junction of the device and the RTD [9], [10], [13].

B. Thermocouple

In thermocouples, two distinct metals are linked at one end and used to measure the temperature. Temperature dependent voltage is generated when the junction between two metals is heated or cooled. These metals form a thermoelectric circuit in which a continuous current flows if they are heated at the joint. A broken circuit at the center will produce a net open circuit voltage known as Seebeck voltage that depends on the temperature at the metal junction and the metal composition. This method is invasive and has a latency of $1\,\mathrm{s}$ [14].

C. Proportional to Absolute Temperature (PTAT) Sensor

These sensors are used for precision temperature measurement but they are invasive. A PTAT sensor generates a voltage value in proportion to the absolute temperature. In [12], a PTAT sensor, shown in Fig. 2, is realized using two MOSFETs (M1 and M2) and two identical pnp BJTs (Q1 and Q2). The aspect ratio (W/L) of M2 is n times of M1 which means if reference current (I_{ref}) flows through M1 then n times I_{ref} will flow through M2 as the two MOSFETs are operated in current mirror mode. This will produce a voltage difference in the BJTs (ΔV_{BE}) as shown in Fig. 2. This voltage is then fed into a differential amplifier which gives the PTAT voltage output (V_{PTAT}). This sensor is then used to measure the temperature of the device under test (DUT) by placing it in between the heat spreaders beneath the DUT. A linear dependency of $19.1\,\mathrm{mV/K}$ is shown by V_{PTAT}.

D. Measurement based on Calorimetry

Though the primary goal of calorimetry is for physical characterization of materials, it can be also used for the calculation of energy loss in the form of heat and correlating it with the change in temperature. A WBG device is placed in the calorimetric chamber. A known DC power is supplied to the device and thereby the thermal capacity of the setup is measured. The thermocouple is used to gauge temperature changes [8]. In [7], a transient calorimetric test is carried out

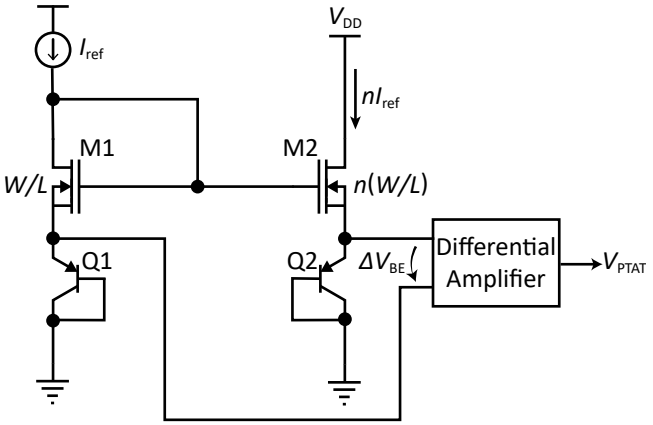


Fig. 2. Schematic of PTAT Sensor (adapted from [12])

on GaN devices that enables measurements (1 minute to 15 minutes, depending on the measurement) with $\pm 15\,\%$ error. This test is also verified for soft-switching. To record the temperature change, a conventional temperature dependent resistor is installed inside the calorimetric chamber as mentioned in [6]. To measure the temperature inside the calorimeter, two thermocouples are mounted on each side of the heatsink. In [5], the calorimetric method is proposed to be of several steps. The first step is to place the thermocouples at the desired location of the setup. Then, the thermal impedance model is derived. It is matched with the simulation. If it matches then the steady-state temperature is measured during operation. The power loss due to the transistor and other components is calculated. Then a correlation between power loss and temperature can be determined.

E. Monolithically Integrated Temperature Sensor

In [15], a method has been proposed that relies on a temperature sensor embedded into the device monolithically by the process of manufacturing itself. The gold metallization chain within two passivation layers between high electron mobility transistor (HEMT) cells and an interconnect metallization is used as a temperature sensor as shown in Fig. 3. A four-point measurement is done where a force current of $10\,\mu\mathrm{A}$ is applied across this temperature sensor and the sensor voltage is sensed at the temperature sensor pads. Then the resistance of the sensor is found using Ohm's law and this process is repeated at different temperatures. It is found that the resistance and temperature share a linear dependency [15].

In [16], a diode sensor is realized in a GaN-on-SiC heterojunction structure. The temperature of the DUT (heterojunction device) is measured by varying temperature from $200\,\mathrm{K}$ to $650\,\mathrm{K}$. This temperature is then measured using forward voltage drop of the integrated diode sensor which shows linear dependency with the increase of temperature at a constant input current through the diode sensor. The diode sensor is realized using devices of different diameters ranging from $400\,\mu\mathrm{m}$ to $1000\,\mu\mathrm{m}$. A sensitivity of $1.81\,\mathrm{mV/K}$ to $2.19\,\mathrm{mV/K}$ is obtained for the sensor. In [17], the integrated

diode temperature sensor which is a string of diodes is embedded into the SiC device module. It can give accurate measurements of the area-related mean junction temperature. A force voltage (current) can be used to sense current (voltage) and this data can be utilized to measure the temperature from the forward characteristics curve of the diode temperature sensor. An example of the dependence of forward voltage (V_f) with temperature (T) with different forward current (I_f) for diode sensor in PMF75120S014 SiC device (1200 V, 75 A) is shown in Fig. 4. The measurement shows a linear sensitivity of V_f with respect to T at different constant I_f.

However, only local temperature is measured in the above two approaches, necessitating the establishment of a link between area-related mean temperature and local temperature [18]. Investigations in the direction of using other physical methods such as liquid crystals, thermographic phosphors and scanning thermal probes in WBG devices lie in the future scope [5], [11].

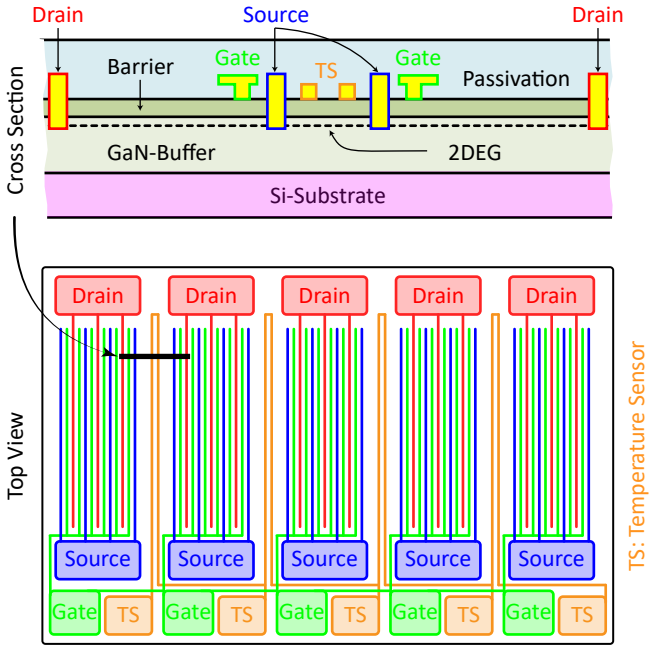


Fig. 3. GaN HEMT with integrated temperature sensor (adapted from [15])

III. Optical Methods

The most prominent optical temperature measurement methods are done using IR sensors, thermography and electroluminescence. Non-contact, compact size and great accuracy are key advantages of IR sensors. The physically contacting methods rely on heat transfer from DUT to the sensors which are impacted by the thermal impedance between them which makes them slower in comparison to IR sensors [19], [20].

The electroluminescence in WBG devices can be categorized as a temperature sensitive optical parameter (TSOP). The spectral emission from SiC devices shows three TSOPs: peak wavelength's red-shift, minor peak's positive temperature coefficient and major peak's negative temperature coefficient.

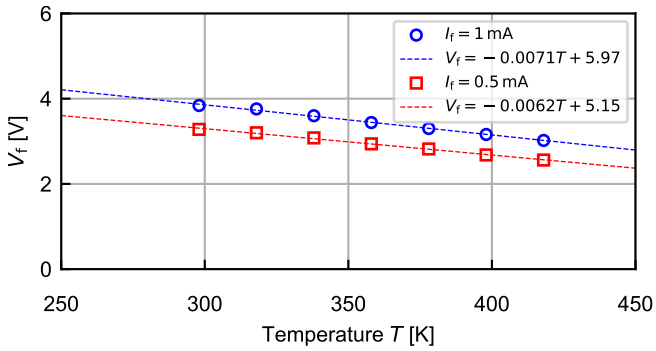


Fig. 4. Measured forward voltage versus temperature curve of integrated diode sensor in PMF75120S014 device

Red-shift cannot be used in the case of power electronic applications because the wavelength shift is small enough (in the range of a few nanometers) for detection. With the use of a couple of bandpass filters and Si photomultipliers, these light emissions can be detected and processed to evaluate the temperature dependency [21]. The reason for electroluminescence in GaN devices is still unknown but it can be used as an indicator for hot electron concentrations and temperature [22]. This can be further investigated to implement the electroluminescence in GaN as a TSOP.

Other TSOPs such as Raman scattering, reflectance and thermo-optic effects can be ventured into WBG devices as in Si devices [11].

IV. Electrical Methods

Electrical parameters vary with temperature and hence measuring the temperature using the temperature sensitive electrical parameters (TSEPs) method requires calibration or a look-up table of the electrical parameters with respect to temperature. One of the biggest challenges in TSEPs is that the electrical parameters not only depend on the temperature but also on other parameters, viz on drain current (I_D), total gate resistance (R_G) and DC link voltage (V_DC), as shown in Table I [19], [23].

TABLE I
TSEPs And Their Dependencies [19]

TSEPs	Dependencies
Forward Voltage (V_DS)	T_j, I_D
Threshold Voltage (V_th)	T_j
Turn-on/off Delay ($t_\mathrm{d,on}$, $t_\mathrm{d,off}$)	T_j, I_D, V_DC, R_G
Gate Current Peak ($I_\mathrm{G,Peak}$)	T_j, V_DC, R_G
Gate Charge (Q_G)	T_j, I_D, R_G
On-state Resistance ($R_\mathrm{DS(ON)}$)	T_j, I_D, V_DC
Diode Forward Voltage (V_f)	T_j, I_f

TSEPs' monitoring in WBG power semiconductor devices has its challenges. They have higher current and voltage slew rates which are very difficult to measure and an uneven miller plateau duration which depends on other parameters as mentioned in [26]. $I_\mathrm{G,Peak}$ shows less temperature sensitivity [27]. The measurement of $R_\mathrm{DS(ON)}$ as a TSEP which requires

TABLE II
COMPARISON OF DIFFERENT TEMPERATURE MEASUREMENT METHODS

Methods	Response	Linearity	Cost	Accuracy	Comments	References
RTD	740 ms	Linear	Medium	0.03°C	Response limited by thermal impedance	[13], [16]
Thermocouple	< 1 s	Non-Linear	Medium	> 1°C	Slowest measurement method	[14], [16]
Monolithically Integrated Sensor	24 ms	Linear	Low	0.05°C	Only local temperature is measured	[13], [16]
Infrared Sensor / Camera	500 ms / 40 ms	—	Medium / High	0.5°C / 2°C	Device has to be opened	[11], [20], [23], [24]
Electroluminescence	—	Non-Linear	Medium	—	Device has to be opened	[21]
Quasi Threshold Voltage	600 μs	Linear	Medium	< 4°C	Varies with different operating points	[23]
Quasi Turn-on Delay	—	Linear	Medium	—	Varies with different operating points	[19]
Internal Gate Resistance	—	Non-Linear	Medium	< 5°C	More optimisation required	[25]

measurement of V_{DS} and I_D is not preferred because of degradation in $R_{DS(ON)}$ due to electron trapping effects in gate oxide layers after 1 h of application of gate voltage. $R_{DS(ON)}$ is also affected by chip metallization degradation and solder degradation. So, it is difficult to measure $R_{DS(ON)}$ as a function of temperature only. Similarly, V_{th} changes during active power cycling due to the trapping effects which makes it unfavourable to implement it as a TSEP [28]. The $R_{DS(ON)}$ in GaN devices has a non-linear relationship with temperature. Under low V_{DS} (< 100 V), $R_{DS(ON)}$ rises and reaches saturation. This saturation time varies with different operating frequencies [29]. As a TSEP, the reverse voltage (V_{SD}) in WBG semiconductors can only be used if sufficient negative gate bias is applied across the device and no additional anti-parallel Schottky barrier diode (SBD) is connected [17]. So, [4] and [19], have come up with quasi threshold voltage ($V_{th,q}$) and quasi turn-on delay ($t_{d,on,q}$) which provide accurate junction temperature measurements. The circuit adapted from [26] is simulated and it is found that $V_{th,q}$ shows a linear sensitivity of -7.5 mV/K and $t_{d,on,q}$ shows a linear sensitivity of -0.1 ns/K as shown in Fig. 5 for BSM120D12P2C005 SiC device (1200 V, 134 A) [30]. Also, internal gate resistance ($R_{G,int}$), shows a non-linear dependency with an increase in temperature with a positive sensitivity [25].

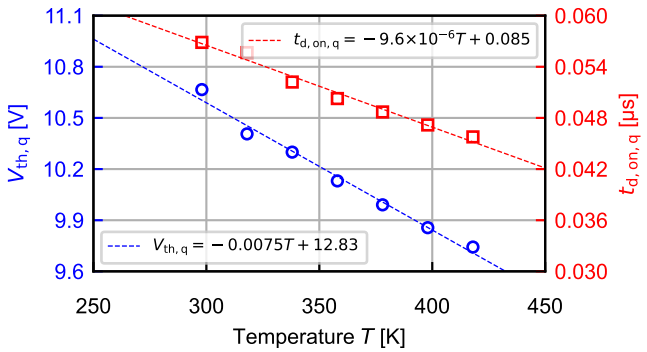


Fig. 5. Simulated results of $V_{th,q}$ and $t_{d,on,q}$

V. COMPARISON AMONG DIFFERENT MEASUREMENT METHODS

The comparison is performed by taking different performance indicators such as response time, linearity, cost, accuracy and certain comments are given in Table II as found in different pieces of literature. The cost is categorized by low, medium and high as less than 10 €, 10 € to 100 € and greater than 100 € respectively. TSEPs and monolithically integrated sensor based measurement methods are non-invasive whereas other methods given in Table II are invasive [31].

VI. CONCLUSION

Several temperature measurement methods for WBG devices have been presented. Among all other methods, optical methods provide the opportunity for easy execution with a wide temperature range but it can get affected by the emissivity of the surface. Fibre optical temperature measurement methods can be explored in the field of semiconductors as they can work in environments having strong electromagnetic field [32]. A comparison among different physically contacting temperature measurement methods shows that diode sensors and temperature sensors based on resistance are the most suitable ones because they are fabricated onto the chip itself which reduces cost and increases accuracy. RTDs are prone to self-heating still they show better long-term stability than thermocouples [16]. Temperature measurement of WBG devices using electrical methods provides fast measurement [23] and it is possible to measure the temperature of the device when it is enclosed. But it is not possible to create thermal maps. Some other ways to characterize junction temperature can be quasi threshold voltage during turn-off and quasi turn-off delay. Further, multiple TSEPs can be combined to accurately monitor the temperature and reduce measurement errors.

REFERENCES

[1] "SiC Power Devices and Modules," 2013. [Online]. Available: https://www.rohm.com/documents/11303/2861707/sic_app-note.pdf

[2] C. T. Ma and Z. H. Gu, "Review of GaN HEMT Applications in Power Converters over 500 W," *Electronics 2019, Vol. 8, Page 1401*, vol. 8, no. 12, p. 1401, nov 2019. [Online]. Available: https://www.mdpi.com/2079-9292/8/12/1401/htmhttps://www.mdpi.com/2079-9292/8/12/1401

[3] L. F. S. Alves, R. C. M. Gomes, P. Lefranc, R. De A. Pegado, P.-O. Jeannin, B. Luciano, and F. V. Rocha, "SIC power devices in power electronics: An overview," in *2017 Brazilian Power Electronics Conference (COBEP)*, Nov. 2017, pp. 1–8.

[4] K. Sharma, D. Dayanand, K. M. Barón, J. Ruthardt, F. Münzenmayer, J. Hückelheim, and I. Kallfass, "A Robust Approach for Characterization of Junction Temperature of SiC Power Devices via Quasi-Threshold Voltage as Temperature Sensitive Electrical Parameter," in *2020 IEEE Applied Power Electronics Conference and Exposition (APEC)*, Mar. 2020, pp. 1532–1536, iSSN: 2470-6647.

[5] H. Li, X. Li, Z. Zhang, J. Wang, L. Liu, and S. Bala, "A simple calorimetric technique for high-efficiency GaN inverter transistor loss measurement," in *2017 IEEE 5th Workshop on Wide Bandgap Power Devices and Applications (WiPDA)*, Oct. 2017, pp. 251–256.

[6] A. Anurag, S. Acharya, and S. Bhattacharya, "An Accurate Calorimetric Loss Measurement Method for SiC MOSFETs," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 2, pp. 1644–1656, Jun. 2020.

[7] D. Neumayr, M. Guacci, D. Bortis, and J. W. Kolar, "New calorimetric power transistor soft-switching loss measurement based on accurate temperature rise monitoring," in *2017 29th International Symposium on Power Semiconductor Devices and IC's (ISPSD)*, May 2017, pp. 447–450, iSSN: 1946-0201.

[8] K. Kroics and V. S. Mainyani, "Application of Calorimetric Technique for Wide Bandgap Transistor Loss Assessment," in *2020 2nd Global Power, Energy and Communication Conference (GPECOM)*, Oct. 2020, pp. 81–86.

[9] D. Koch, S. Araujo, J. Weimer, and I. Kallfass, "Automated Calorimetric Measurement with a Peltier Element for Switching Loss Characterization," in *PCIM Europe digital days 2021; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management*, May 2021, pp. 1–8.

[10] Texas Instruments, "Four-wire PT100 RTD measurement circuit with low-side reference." [Online]. Available: https://www.ti.com/lit/an/sbaa336a/sbaa336a.pdf?ts=1655222026895

[11] D. Blackburn, "Temperature measurements of semiconductor devices - a review," in *Twentieth Annual IEEE Semiconductor Thermal Measurement and Management Symposium (IEEE Cat. No.04CH37545)*, Mar. 2004, pp. 70–80, iSSN: 1065-2221.

[12] D. Iero, F. G. D. Corte, M. Merenda, and C. Felini, "A PTAT-based Heat-flux Sensor for the Measurement of Power Losses through a Calorimetric Apparatus," *Procedia Engineering*, vol. 168, pp. 1617–1620, Jan. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877705816337961

[13] D. Koch, S. Moench, R. Reiner, J. Hueckelheim, K. M. Baron, P. Waltereit, and I. Kallfass, "Static and dynamic characterization of a monolithic integrated temperature sensor in a 600 v gan power ic," in *PCIM Europe digital days 2020; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management*, 2020, pp. 1–7.

[14] "RTDs Thermocouples Frequently Asked Questions — TE Connectivity." [Online]. Available: https://www.te.com/usa-en/faqs/rtds-thermocouples-faqs.html

[15] R. Reiner, P. Waltereit, B. Weiss, M. Wespel, D. Meder, M. Mikulla, R. Quay, and O. Ambacher, "Linear temperature sensors in high-voltage GaN-HEMT power devices," in *2016 IEEE Applied Power Electronics Conference and Exposition (APEC)*, Mar. 2016, pp. 2083–2086.

[16] S. Madhusoodhanan, S. Koukourinkova, T. White, Z. Chen, Y. Zhao, and M. E. Ware, "Highly linear temperature sensor using GaN-on-SiC heterojunction diode for Harsh environment applications," in *2016 IEEE 4th Workshop on Wide Bandgap Power Devices and Applications (WiPDA)*, Nov. 2016, pp. 171–175.

[17] C. Kempiak, A. Lindemann, S. Idaka, and E. Thal, "Investigation of an Integrated Sensor to Determine Junction Temperature of SiC MOSFETs During Power Cycling Tests," in *2019 10th International Conference on Power Electronics and ECCE Asia (ICPE 2019 - ECCE Asia)*, May 2019, pp. 3084–3089, iSSN: 2150-6086.

[18] C. Kempiak, A. Lindemann, S. Idaka, and E. Thal, "Comparative study of determining junction temperature of SiC MOSFETs during power cycling tests by a Tj sensor and the $V_{\mathrm{SD}}(T)$-method," in *CIPS 2020; 11th International Conference on Integrated Power Electronics Systems*, Mar. 2020, pp. 1–6.

[19] K. Sharma, K. M. Barón, J. Ruthardt, and I. Kallfass, "Online Junction Temperature Monitoring of Wide Bandgap Power Transistors using Quasi Turn-on Delay as TSEP," in *2021 IEEE 8th Workshop on Wide Bandgap Power Devices and Applications (WiPDA)*, Nov. 2021, pp. 129–134, iSSN: 2641-8274.

[20] "Datasheet for MLX90614 I Melexis." [Online]. Available: https://www.melexis.com/en/documents/documentation/datasheets/datasheet-mlx90614

[21] J. Winkler, J. Homoth, and I. Kallfass, "Electroluminescence-Based Junction Temperature Measurement Approach for SiC Power MOSFETs," *IEEE Transactions on Power Electronics*, vol. 35, no. 3, pp. 2990–2998, Mar. 2020.

[22] J. Pankove, E. Miller, and J. Berkeyheiser, "GaN electroluminescent diodes," in *1971 International Electron Devices Meeting*, Oct. 1971, pp. 78–78.

[23] K. Sharma, J. Ruthardt, I. Kallfass, and E. Drivers, "Comparison of Different Methods for the Characterization of Online Junction Temperature of a Gallium-Nitride Power Transistor," unpublished.

[24] "UTi260B - UNI-T Thermal Imaging |Thermal Cameras,Thermal Monoculars." [Online]. Available: https://thermal.uni-trend.com/product/uti260b/#Specifications9eb2-57af

[25] J. Ruthardt, D. Hirning, K. Sharma, M. Nitzsche, P. Ziegler, M. Fischer, and J. Roth-Stielow, "Investigations on Online Junction Temperature Measurement for SiC-MOSFETs Using the Gate-Signal Injection Method," in *2021 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2021, pp. 5354–5359.

[26] Z. Ni, S. Zheng, M. S. Chinthavali, and D. Cao, "Investigation of Dynamic Temperature-Sensitive Electrical Parameters for Medium-Voltage SiC and Si Devices," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 5, pp. 6408–6423, Oct. 2021.

[27] K. M. Barón, K. Sharma, M. Nitzsche, and I. Kallfass, "Characterization of Electrical Parameters for Health Monitoring in SiC MOSFETs during AC Power Cycling," in *2021 IEEE 8th Workshop on Wide Bandgap Power Devices and Applications (WiPDA)*, Nov. 2021, pp. 316–321, iSSN: 2641-8274.

[28] C. Herold, J. Sun, P. Seidel, L. Tinschert, and J. Lutz, "Power cycling methods for SiC MOSFETs," in *2017 29th International Symposium on Power Semiconductor Devices and IC's (ISPSD)*, May 2017, pp. 367–370, iSSN: 1946-0201.

[29] W. Wang, Y. Liang, M. Zhang, F. Lin, F. Wen, and H. Wang, "Mechanism Analysis of Dynamic On-State Resistance Degradation for a Commercial GaN HEMT Using Double Pulse Test," *Electronics*, vol. 10, no. 10, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/10/1202

[30] ROHM Semiconductor, "SiC Power Module BSM120D12P2C005 Datasheet." [Online]. Available: https://fscdn.rohm.com/en/products/databook/datasheet/discrete/sic/power_module/bsm120d12p2c005-e.pdf

[31] A. Griffo, J. Wang, K. Colombage, and T. Kamel, "Real-Time Measurement of Temperature Sensitive Electrical Parameters in SiC Power MOSFETs," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2663–2671, Mar. 2018.

[32] L. Hoffmann, M. S. Müller, S. Krämer, M. Giebel, G. Schwotzer, and T. Wieduwilt, "Applications of fibre optic temperature measurement," *Estonian Journal of Engineering*, vol. 13, no. 4, 2007.

An Approach to Radar Processing using OFDM Signals- Advantages and Limitations

MSc. Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract* – In this paper, radar processing such as calculation of range profile and ability of a radar to carry and demodulate the information using Orthogonal Frequency Division Multiplexing (OFDM) technology in contrast with Frequency modulated continuous wave (FMCW) radar is synthesized and reviewed based on literature research of other scientific papers. The OFDM radar processing also comes with the limitations in its performance due to its own properties like disorientation in orthogonality of subcarriers. This paper also discusses the challenges involved and limitations in OFDM radar processing.**

Index Terms – OFDM, MIMO, range, FMCW, interference, chirp, doppler shift

I. Introduction

Orthogonal frequency division multiplexing (OFDM) signals are well harnessed in the field of wireless communication systems such as telecommunication in 5G, where they are a standard for wireless communication. Researchers have studied and implemented its properties in radar processing and have drawn some significant conclusions and advantages in comparison with FMCW radars. Firstly, OFDM signals are reported to be advantageous regarding Doppler shift [1]. Besides providing high tolerance against Doppler shift they also do not experience range- Doppler coupling, which allows for independent and unambiguous range and Doppler processing. Secondly, due to their origin from digital communications, OFDM signals are designed to carry information, which allows for simultaneous information transmission in parallel to radar sensing [2].

In this paper the concept and an approach to OFDM radar processing will be presented, that compensates the drawbacks involved in correlation-based approach and directly operates on the complex modulation symbols that compose the OFDM signal as mentioned in [3]. It will be shown that with this approach it will be possible to overcome the drawbacks of the correlation-based approach and still the possibility of transmitting arbitrary information in parallel to the radar operation will be given. In the following pages, first the structure of OFDM signals will be discussed and then its ability of radar processing while carrying information is compared in contrast with the classical FMCW radars will be presented. Then, finally functionality and feasibility of both OFDM based radar and FMCW radar will be comprehended in respect to the state of the art along with their limitations.

II. OFDM Signal Generation

OFDM is a specialised FDM in which main signal is divided, and placed orthogonal to each other. Orthogonal signals are signals that are perpendicular to each other in its frequency domain, that means wherever one subcarrier has its peak, all other subcarriers have nulls and hence do not overlap each other. An OFDM signal is generated by multiplexing a series of complex modulation symbols onto multiple orthogonal subcarriers. The OFDM time domain signal of one OFDM symbol can be expressed as

$$x(t) = \sum_{n=0}^{N-1} I(n) \exp(\, j2\pi f_n t), \quad 0 \le t \le T \qquad (1)$$

with N denoting the number of orthogonal subcarriers, f_n being the individual subcarrier frequencies, T being the elementary OFDM symbol duration, and $I(n)$ representing an arbitrary information series consisting of complex modulation symbols obtained through a discrete phase modulation technique, e.g., Quadrature phase shift keying (QPSK). In order to avoid interference between the single subcarriers, the subcarriers have to be orthogonal, which is fulfilled in case of

$$\Delta f = \frac{1}{T}. \qquad (2)$$

Here Δf represents the frequency difference between two adjacent subcarriers. In the frequency domain, the spectrum of one OFDM symbol corresponds to

$$X(f) = \sum_{n=0}^{N-1} I(n) \frac{\sqrt{T}\sin(\pi(f-f_n)T)}{(\pi(f-f_n)T)}. \qquad (3)$$

As noted from above equation, OFDM transmits a large number of narrowband carriers, in the form of Sinc waves. The OFDM signal can be described as a set of closely spaced FDM subcarriers. We can also observe that by encoding data $I(n)$, we can transmit some arbitrary data for the communication as indicated by in the frequency domain equation (3). We can observe OFDM spectrum with subcarriers in Fig.1.

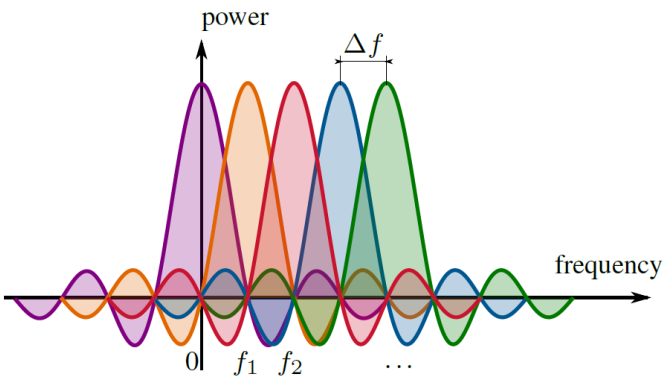


Fig. 1. Spectrum of an OFDM signal with orthogonal subcarriers (from [11])

III. OFDM SYSTEM PARAMETERIZATION

In order to check if the generated OFDM signal is suitable for the purpose of communication, several criteria have to be considered. In the following an example system parameterization for operation in the 24 GHz will be calculated.

A. Doppler Shift

Doppler shift in radar processing arises due to the relative motion between the radar and the target. This relative motion results in shift in the operational frequency and is given by the following equation

$$f_D = 2v_{rel} / \lambda \quad (4)$$

with f_D denoting the Doppler frequency shift and v_{rel} being the relative velocity between the radar and the target. Assuming a maximum relative velocity of v_{rel} = 200 km/h corresponding to $v_{rel,max}$ = 55.6 m/s this results at the carrier frequency of 24 GHz in a maximum Doppler shift of $f_{D,max}$ = 8.9 kHz. In order to always maintain the orthogonality of the subcarriers, as shown in Fig.1., the subcarrier spacing Δf has to be chosen much larger than the maximum Doppler shift $f_{D,max}$. It can be assumed that the limit $\Delta f > 10\ f_{D,max.}$. By having this amount of spacing we can avoid doppler shift.

B. Multi-path Propagation

In order to avoid inter-symbol interference between adjacent OFDM symbols, each OFDM symbol must be extended at the transmitter with a prefix containing a partial cyclic repetition of itself of duration T_P. This is the maximum time difference of the transmitted (Tx) and received (Rx) signals. Generally, this time ranges from milliseconds to microseconds. If the distance is 200 m, we obtain $T_P > 1.33$ µs.

C. Radar Range

Range is the ability of a radar system to distinguish between two or more targets at different ranges. Range of a radar is given by the following equation

$$\Delta r = C_0 / 2B \quad (5)$$

with Δr denoting the range resolution and B is the total signal bandwidth. Range is found to be 1.5m in this case. We need to note here that range is divided by 2 because the signal travels twice before reaching the receiver.

D. Number of Subcarriers

In OFDM systems, first we have the data in frequency domain modulated with symbols with the help of schemes such as QPSK mapped on each subcarrier. These symbols have to be converted to time domain before transmitting through antenna. For this, IFFT is implemented at transmitter side and FFT is implanted at receiver side. This can be done by choosing the number of subcarriers having power of 2.

For the purpose of demonstration, the scenario of OFDM radar with 4 antennas is considered. A Matlab waveform of OFDM transmitter signal is generated using FFT length of 64 and the data has been modulated with 64QAM. We can observe the 100 OFDM symbols and the pilot symbols with guard bands in the following Fig.2.

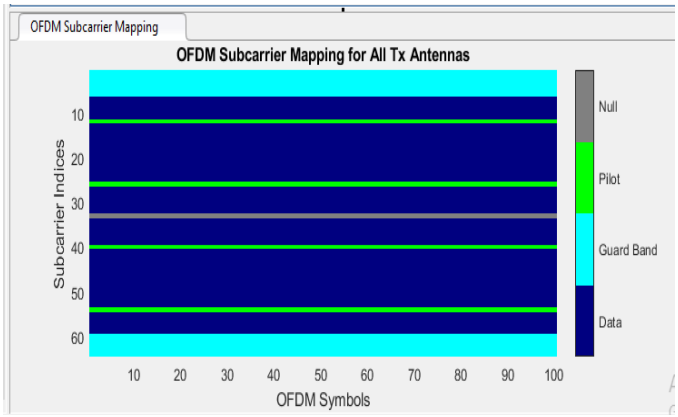


Fig. 2. OFDM mapping with 4 antennas of 100 symbols generated in Matlab

IV. STATE OF THE ART OF OFDM AND FMCW RADARS

The use of signals with multiple carriers for radar was introduced in [4] and [5]. In this paper, a method of distance estimation is proposed in which we need to correlate/compare the transmitted and received OFDM signals for radar processing. However, in order to get very better distance estimation this method imposes a perfect autocorrelation requirement. For this purpose, researcher have developed another method of OFDM radar processing. The basic idea of this approach consists in comparing the transmitted information *I(n)* and the received soft-state information *Ir(n)}*. If in case distortion happens due to some obstacle in the environment or multipath propagation of OFDM signal, this information is obviously stored in the received arbitrary information. We already discussed that we need to keep the minimum subcarrier spacing in equation (2), with the help of this information we can exactly determine in which symbol, and which subcarrier the distortion is happening and correct that before post radar

processing. Results of correlation-based approach and newly proposed method in OFDM is shown in Fig.3. and Fig. 4.

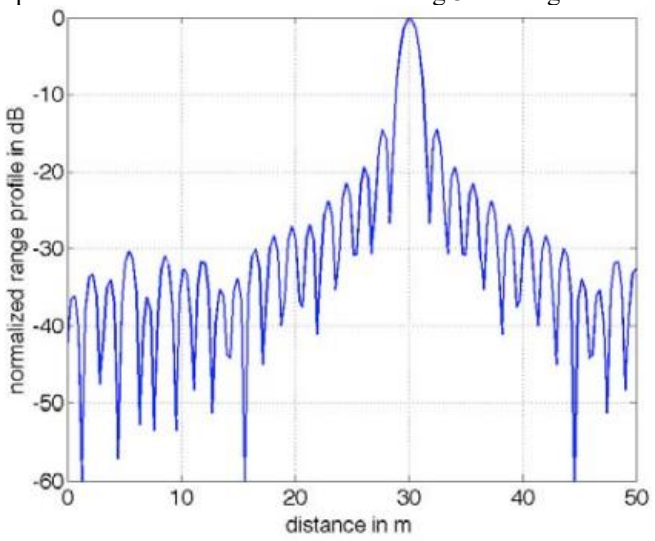


Fig. 3: Radar range profile calculated with the correlation based approach for one single scatterer with σ = 10 dBm^2 , v_{rel} = 0, at d = 30 m (from [3])

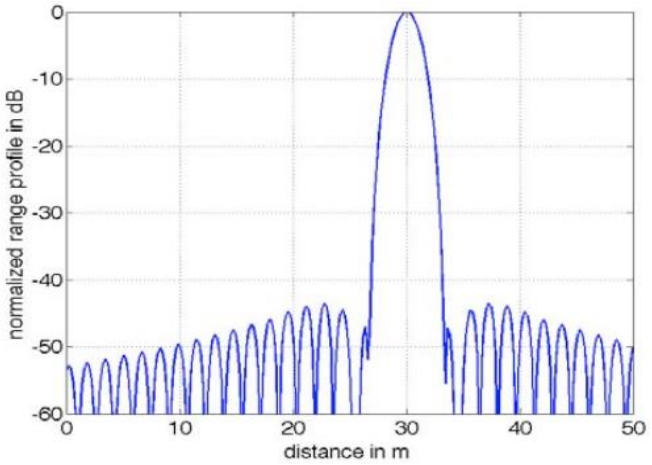


Fig. 4. Radar range profile calculated with the novel approach for one single scatterer with σ = 10 dBm^2 , v_{rel} = 0, at d = 30 m (from [3])

In the above figures Fig.3 and Fig.4, we can observe that by comparing the received arbitrary information and correcting the errors before post radar processing, lower sideband levels are possible. This is possible due to the special ability of OFDM to carry arbitrary information as stated in equation (1). Another signal processing method for distance-velocity estimation with OFDM signals has been proposed in [6]. In these works, the researchers have proposed a new idea of separation of multiple users for communication purposes using spread spectrum approach which also paves a way for spectral efficient radar sensing.

FMCW radar, on the other hand, operates with linear frequency modulated signals. In an FMCW system, the transmitter antenna emits frequency modulated continuous radio waves, also known as chirps because it resembles birds' chirp, and the reflected signal from the target is received by the receiving antenna. Later, the output of the receiving antenna is given to the mixer. In the mixer circuit, a fraction of the frequency-modulated transmitted signal is mixed with the received signal, producing a new signal, which can be used to determine the distance (R) and/or velocity of the moving object. The frequency of the new signal is the difference between the frequency of the sent signal at transmitter and received signal. Then, the signal from the mixer output passes through a lowpass filter, in this stage noise arise from hills or any other objects are filtered out. Finally, the signal passes via an amplifier, A/D converter, and is then proceeded for processing to calculate the distance and velocity of the object. Thus, the mixing of two chirps yields their frequency difference1 that is determined both by the delay and Doppler shift [7]. For this purpose, a FMCW signal of 100kHz bandwidth and sweep time around 100 µs is generated in Matlab and depicted in following Fig.5.

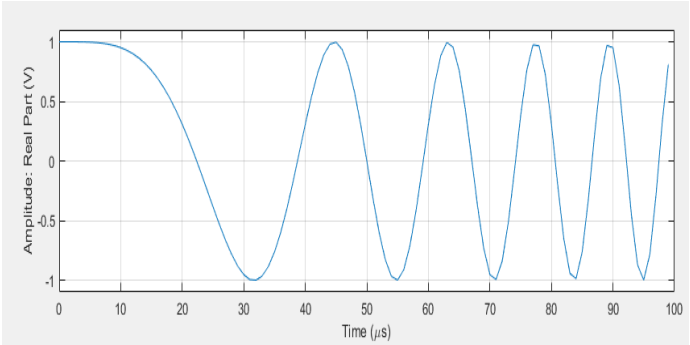


Fig. 5. FMCW waveform generated in Matlab

In the case of multiple targets, a number of parallel lines representing possible solutions in the distance-velocity space occur for each chirp. For example, if there are two different ramps with different slopes and if there are two targets, the radar detects four intersections points on range profile. Two of them do not actually represent physical targets, i.e., they are ghost targets. Thus, we require another chirp to resolve the ambiguity. In general, to resolve this ambiguity we need to keep n+1 chirp signals for n number of targets this was discussed in [8]. For several targets and chirps, there is a study that shows number of intersections grows disproportionately.

V. Limitations of State of the Art of OFDM Radar

A. Sensitivity to Doppler shift

A fundamental issue of OFDM radar is that, it is sensitive to Doppler shift. If the OFDM subcarriers are Doppler shifted due to the relative motion between system and target, the points where the orthogonality is given are shifted in frequency, and at the initial frequency points the subcarrier orthogonality. This will lead to intercarrier interference (ICI), i.e., reduced orthogonality between the OFDM subcarriers, when this happens radar processing will not be accurate.

B. High peak-to-average-power ratio (PAPR)

Another basic drawback of OFDM radar is the potentially high Peak to average power ratio (PAPR). It is ratio of the peak power to the average signal power. It imposes high linearity on power amplifiers (PAs) and resolution of DAC

C. Time and frequency synchronization.

The OFDM subcarriers will be orthogonal only if they are synchronized in time and frequency respectively. Otherwise, it leads to Inter symbol interference (ISI). The errors occurred in time domain synchronization will be appeared in cyclic prefix and causes delay in radar operation. If there is error in frequency domain subcarriers will be shifted, this type of error leads to doppler shift.

D. Internal coupling and crosstalk between antennas.

When OFDM radars operating in monostatic mode, they face the problem of crosstalk. That is the presence of strong input at the receiver leads to coupling problems. For signals such as OFDM radar, the issue is exacerbated because duplex systems need more sophisticated filtering techniques to avoid internal crosstalk such as software-configurable filter described in [9].

VI. Comparison between OFDM and FMCW radar

Based on functionalities of both FMCW and OFDM radar, we can observe that these two radars differ in i) how their signals are generated and ii) how the demodulation at the receiver is performed.

In OFDM radar, there are three available methods. i) Cyclic prefix OFDM (CP-OFDM) is robust against doppler shift but there is a wastage of signal power due to cyclic prefix removal. In order to solve this problem a new approach is proposed called ii) repeated symbol OFDM (RS-OFDM) in [10] in which researchers propose to remove cyclic prefix at other symbols and append only at first symbol. This will improve SNR at the receiver. Another approach called iii) all-cell doppler correction OFDM (ACDC-OFDM) works independent of doppler shift since it operates by shifting entire symbols which have undergone intercarrier interference and has better dynamic range compared to other two approaches. We can observe simulation results of all three methods in Fig.6.

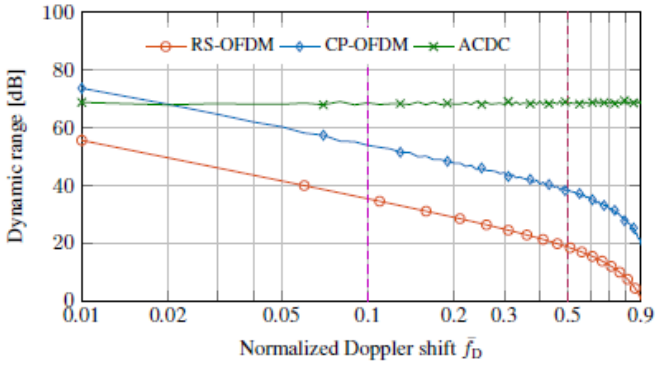


Fig. 6. Dynamic range as a function of the normalized Doppler shift for different OFDM radar approaches, (from [11])

Clearly RS-OFDM has less dynamic range because of not having cyclic prefix at every symbol and CP-OFDM methods has better dynamic range than RS-OFDM. Finally, ACDC-OFDM has constant dynamic range because it is independent of doppler shift.

In FMCW, radar transmits a chirped signal and receive a time delayed version of it gets reflected from the target. After mixing this signal and filtering, the resulting signal is a sinusoid at a frequency that is a function of the target's range. This frequency is known as the "beat" frequency f_b. Thus, the dechirped signal will have a form of

$$x(t)= e^{j(2\pi fbt+\phi)} = e^{j2\pi fbt}\, e^{j\phi} \tag{6}$$

where $e^{j\phi}$ is additional phase shift. For demonstrating FMCW range profile generation method, a Matlab code is written and executed for target at 40 meters and velocity of 20 m/s. It is shown in below in Fig. 7.

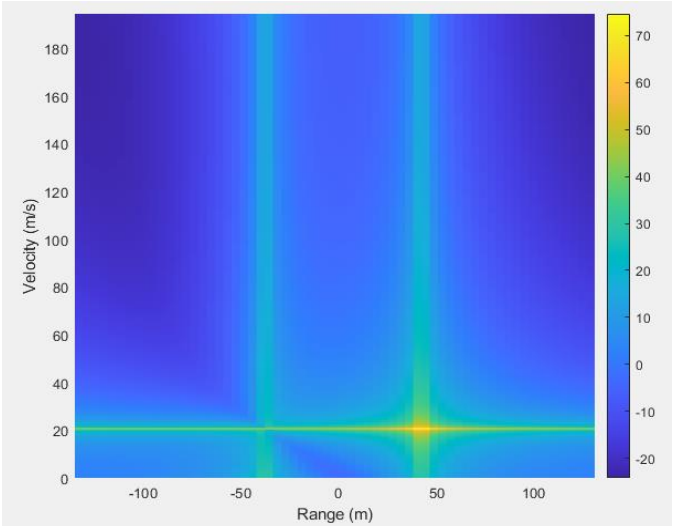


Fig.7. Radar Range profile of FMCW radar generated in Matlab

VII. Conclusion

Radar processing using OFDM is definitely a promising feature as it has better distance velocity profile estimations and can have range profile having sideband levels of around -45 dB is achieved which is a significant improvement in contrast to classical autocorrelation method in which sidelobes cannot be decreased using signal processing such as windowing. By combining several available approaches like ACDC and RS-OFDM in OFDM radars, significant SNR is achieved at the receivers making OFDM based radars power efficient while also having better distance-velocity ranges and better dynamic ranges. Because of its robustness against interference due to orthogonal signals and its ability to carry signals without much noise even in MIMO configurations, will make it a better substitute for FMCW radars where multipath propagation is dominating.

REFERENCES

[1] G. E. A. Franken, H. Nikookar, P. van Genderen "Doppler tolerance of OFDM coded Radar Signals, *Proc. 3rd European Radar Conference*, pp. 108 – 111, Manchester, Sept. 2006.

[2] A.Garmatyuk, J. Schuerger, Y. T. Morton, K. Binns, M. Durbin, J.Kimani, "Feasibility study of a multi-carrier dual-use imaging radar and communication system," in *Proc. 37th European Microwave Conference*, pp. 1473 – 1476, Munich, Oct. 2007.

[3] C. Sturm, E. Pancera, T. Zwick and W. Wiesbeck, "A novel approach to OFDM radar processing," *2009 IEEE Radar Conference, Pasadena, CA,USA,* 2009, pp. 1-4, doi: 10.1109/RADAR. 2009.4977002

[4] Levanon, N. Multifrequency radar signals. In Radar Conference, 2000. *The Record of the IEEE 2000 International*, pages 683–688, 2000a.

[5] Levanon, N. Multifrequency complementary phase-coded radar signal. IEEE Proceedings - Radar, Sonar and Navigation, 147(6):276–284, Dec2000b.

[6] Sturm, Christian & Wiesbeck, W.. (2011). Waveform Design and Signal Processing Aspects for Fusion of Wireless Communications and Radar Sensing. *Proceedings of the IEEE*. 99. 1236 - 1259. 10.1109/JPROC.2011.2131110.

[7] Schoor, M. Hochauflösende Winkelschätzung für automobile Radarsysteme. PhD thesis, University of Stuttgart, 2010.

[8] Reiher, M. and Yang, B. On the occurrence of ghost targets in linea FMCW radar: A worst case study. In *2008 International Radar Symposium*, pages 1–4, May 2008.

[9] Hong, S., Brand, J., Choi, J. I., Jain, M., Mehlman, J., Katti, S., andLevis P. Applications of self-interference cancellation *in 5G and beyond. IEEE Communications Magazine*, 52(2): 114–121, February 2014.

[10] Hakobyan, G., Girma, M., Li, X., Tammireddy, N., and Yang, B. Repeated symbols OFDM MIMO radar at 24 GHz. In *2016 European Radar Conference (EuRAD)*, pages 249–252, Oct 2016f.

[11] Gor Hakobyan, Orthogonal Frequency Division Multiplexing Multiple-Input Multiple-Output Automotive radar with Novel Signal Processing Algorithms. Dissertation Universität Stuttgart (2018).

State-of-the-art SiGe BiCMOS Technologies

Burak Özat
Electrical Engineering
Universität Stuttgart
Stuttgart, Germany

***Abstract*—This paper reviews state-of-the-art SiGe BiCMOS technologies and power & low-noise amplifier designs in state-of-the-art SiGe BiCMOS technologies. Advantages of the SiGe BiCMOS technologies compared to CMOS and III-V technologies and future targets are discussed in this paper. Current SiGe BiCMOS technologies in production offer transit frequency (f_T) and maximum oscillation frequency (f_{MAX}) 300 GHz and 400 GHz on an average, respectively. European and U.S. SiGe BiCMOS manufacturers have made a great effort to push the limits of SiGe BiCMOS technologies. IHP reported the best technology featuring highest f_T/f_{MAX} 470/700 GHz. The breakthroughs in SiGe BiCMOS technology will pave the way for high data rate applications at mm-wave, submm-wave and even THz frequencies.**

Keywords—SiGe BiCMOS technology, Heterojunction Bipolar Transistors (HBT), transit frequency (f_T), maximum oscillation frequency (f_{MAX}), millimeter-wave, THz.

I. Introduction

Millimeter-wave circuits and systems covering the frequency range 30 GHz - 300 GHz in various fields of application such as telecommunications, and automotive radars have been dominated by silicon-germanium (SiGe) devices. BiCMOS technologies combine SiGe Heterojunction Bipolar Transistors (HBT) and CMOS in single silicon die. While HBTs provide high-speed operation and superior RF performance, CMOS offers high computing capability and their outstanding features make the BiCMOS technologies attractive. As a result of continuous scaling of CMOS devices over time, their transit frequency (f_T) and maximum oscillation frequency (f_{MAX}) are improved notably and they find a place in the millimeter-wave applications. Next generation automotive radar and communication systems will perform high-resolution detection and high data rate communication. Therefore, they require hardware having very large operation bandwidth. The upper end of mm-wave and sub-mm-wave (> 300 GHz) provide broad available frequency band to be used in these systems. However, f_T and f_{MAX} of CMOS devices are not sufficient to design circuits at such high frequencies. For this reason, SiGe HBTs will keep the leading position in high frequency applications.

In this paper, advantages of the SiGe BiCMOS technology over its competitors e.g. CMOS, III-V technologies and future directions of the SiGe BiCMOS are discussed and the current state-of-the-art SiGe technologies of leading SiGe manufacturers such as IHP, STMicroelectronics, Infineon and GlobalFoundries and some RF circuit blocks designed in these technologies are reviewed.

TABLE I. The State of the Art SiGe BiCMOS Tecnologies

Manufacturer	*Technology*	*f_T/f_{MAX}*	*Reference*
IHP	SG13G3	470/700 GHz	[11]
Infineon	130-nm	305/537 GHz	[7]
GlobalFoundries	45-nm	375/510 GHz	[14]
Infineon	B12HFC 90-nm	287/495 GHz	[9]
IHP	SG13G2	350/450 GHz	[17]
NXP	90-nm	230/400 GHz	[23]
ST	55-nm	320/370 GHz	[6]
Infineon	B11HFC 130-nm	250/370 GHz	[22]
GlobalFoundries	9HP 90-nm	300/360 GHz	[13]
TowerJazz	SBC18H5	285/310 GHz	[20]

II. Advantages of SiGe Technology

A. SiGe HBT and CMOS RF Performance Comparison

The SiGe HBT and III-V technologies have dominated RF and Analog Mixed Signal (AMS) fields since they offer greater f_T and larger breakdown voltages (BV) compared to CMOS technology. As a result of sustained scaling of the CMOS technology in the last decades, f_T of the CMOS devices increased remarkably which makes the CMOS technology offering greater integration more attractive. Thus, the advanced CMOS has found more use in RF and AMS applications nowadays and it was even expected to take the place of the BiCMOS technologies. However, the BiCMOS technologies will continue dominating the CMOS in many of these applications due to the following advantages over the CMOS [19].

1) Device models and RF performance of the devices: A standard way to extract the device models and report their RF performances is de-embedding the parasitics such as effects of resistive and capacitive metallization layers. Thus, these are not included in the device models. However, the CMOS devices' RF performance, and especially their speed (f_T) worsen due to the parasitics, if the metallization layers are taken into account. In contrast to CMOS, the de-embedding does not have significant impact on HBTs performance [19]. For this reason, it is worth noting that SiGe BiCMOS outperforms CMOS against the parasitics and they must be considered while comparing the high-frequency performance of BiCMOS and CMOS [19].

2) Outstanding RF performance of the SiGe HBTs: SiGe HBTs feature large transconductance g_m, high f_T and f_{MAX} compared to the CMOS devices. Currently, 55-nm BiCMOS technology of STMicroelectronics offers commercial and research bipolar transistors with a transconductance of 200 mS/μm^2 and 600 mS/μm^2 with f_T/f_{MAX} 320 GHz/370 GHz [21]. Thus, the SiGe BiCMOS bipolar transistors provide impressive RF performance e.g. gain, bandwidth (BW), and noise figure even with low collector currents. HBTs are a good choice for the low-power applications as well because they show still good performance under power-efficient biasing conditions. Furthermore, HBTs have lower 1/f noise corner frequency and larger breakdown voltages (BV) compared to the CMOS transistors [6]. The CMOS devices' corner frequencies are several orders of magnitude greater [21]. Moreover, higher output powers are achieved thank to higher BV of HBTs compared to CMOS devices.

3) Sustainable improvement of f_T and f_{MAX} of SiGe HBTs in BiCMOS technology: f_T of the CMOS devices is improved by continuous transistor scaling. Similarly, f_T of SiGe HBTs technology is increased by improving processes e.g. doping profiles, vertical and lateral scaling resulting in lower diffusion & junction capacitances and emitter & collector resistances. To obtain high f_{MAX}, f_T must be increased, base resistance R_B and base-collector C_{BC} must be minimized for HBTs [17]. By scaling down the effective emitter length, f_T and f_{MAX} can be improved. Furthermore, high f_T, low gate-resistance R_G and gate-drain capacitance C_{GD} result in high f_{MAX}. While scaling the CMOS device's channel length improves f_T, gate resistance R_G increases. Consequently, after some scaling, f_{MAX} for the CMOS devices cannot be increased by enhancing the f_T. Due to the increase of the gate resistance R_G which is caused by the scaling, f_{MAX} drops despite f_T increase in technologies below 65-nm node [4]. Thus, scaling of the CMOS devices which is the main driving force for the development of the CMOS technology does not enhance f_{MAX} below a certain gate length.

$$f_T = \frac{g_m}{2\pi\left(C_{diff}+C_{BE}+C_{BC}(1+g_m R_E+g_m R_C)\right)} \quad (1)$$

$$f_{MAX} \cong \sqrt{\frac{f_T}{8\pi R_B C_{BC}}} \quad (2)$$

$$f_{MAX} \cong \sqrt{\frac{f_T}{8\pi R_G C_{GD}}} \quad (3)$$

$$R_G \propto \frac{1}{L} \quad (4)$$

B. SiGe HBT and III/V Transistors RF Performance Comparison

III-V compound devices such as GaN, InP, InAs, show better RF performance e.g. larger f_T, and greater breakdown voltages compared to the SiGe HBTs. HBTs' breakdown voltages are around half of GaAs HEMTs' breakdown voltages and HBTs' transit frequencies are 1/3 of HEMTs' when their BVs are the same or very close [8]. However, the foundry process of the III-V devices is less maturate and hence their circuit complexity is restricted [21]. Moreover, unlike III-V devices, in BiCMOS technology, HBTs can be easily integrated with low power digital CMOS circuits on a single chip. Thus, power consumption, performance and total area of the hardware can be optimized by avoiding external components.

III. SiGe BiCMOS State of the Art

Today, leading advanced SiGe BiCMOS manufacturers are mainly IHP, STMicroelectronics, Infineon, Tower Semiconductors, Global Foundries (IBM), and NXP Semiconductors. Typically, industrial processes in production offer f_T and f_{MAX} around 300 GHz and 400 GHz. Due to the continuous developments in the technology, there is a constant demand for higher speed and higher data rate communication systems.

To meet the demand, the TARANTO project aiming to develop advanced HBTs in BiCMOS technology, integrated circuits and systems for telecommunication and radar applications was launched in 2017 by the European Union. The project ended in 2021. The TARANTO project targeted HBTs with maximum frequencies 600 GHz for STMicroelectronics / Infineon HBTs and 700 GHz for IHP [10]. In the project's scope, IHP reported a new advanced SiGe BiCMOS process SG13G3 which features f_T/f_{MAX} 470/700 GHz. Also, Infineon demonstrated two new BiCMOS processes which are with 90nm and 130nm CMOS nodes. HBTs in these processes offer f_T/f_{MAX} 287/495 GHz and 300/537 GHz, respectively [9].

A. The Advanced IHP SiGe HBTs

IHP developed the fastest HBTs by employing a new process with non-selective epitaxial growth (NSEG) of the base and elevated extrinsic base (EEB) regions generated with selective epitaxial growth (SEG) [22]. In Figure 1, cross section of HBT in the new process is given. The efforts on optimization of the process paid off and IHP could demonstrate the best experimental bipolar-only HBTs in the literature having peak f_T/f_{MAX} 505/720 GHz.

The adaptation of this process to the BiCMOS HBTs is challenged by the thermal budget of the CMOS process. After innovations in the process flow, the NSEG process was adapted to the BiCMOS HBTs. Thus, advanced SG13G3 technology having f_T/f_{MAX} 470/700 GHz was introduced by IHP. Compared to the previous SG13G2 technology, the new process features much greater f_T/f_{MAX} in return for a drop approximately 0.3 V in BV_{CEs}. However, BV_{CEo} remained as 1.6 V [9]. The SG13G3 technology is one of the best candidates for the future BiCMOS technologies.

The second process was introduced to develop BiCMOS HBTs by a collective run of IHP and Infineon. Selective epitaxial growth (SEG) is realized with an epitaxial base link (EBL). EBL is used to increase f_{MAX} of the double polysilicon self-aligned architecture (DPSA) by reducing the resistance of the base link. Several innovations in the process are made. Finally, a new process with f_T/f_{MAX} close to 330/600 GHz was reported by IHP and Infineon. [22].

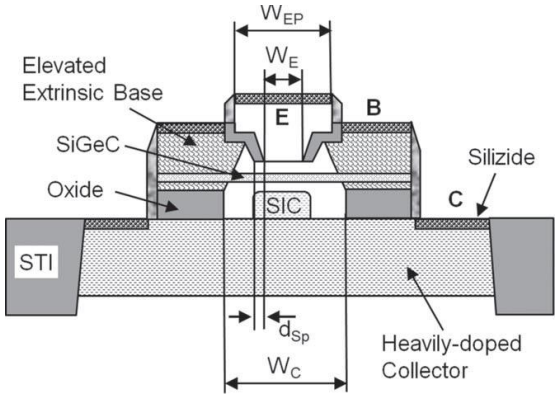


Fig. 1. Cross section of the IHP HBT [11].

B. *Industrial SiGe BiCMOS Process of Infineon*

Infineon developed a new generation SiGe BiCMOS B12HFC 90-nm process on the previous B11HFC 130-nm process. To improve the performance, the SEG EBL architecture presented by IHP was transferred and Infineon used this architecture instead of DPSA used in the older process. Eventually, Infineon's industrial B12HFC 90-nm BiCMOS technology offers f_T/f_{MAX} close to around 300/480 GHz. To achieve the same performance obtained after IHP and Infineon's collective work, need for more lateral scaling and enhanced base link epitaxy doping profile is considered for the future [9].

C. *Industrial SiGe BiCMOS Process of STMicroelectronics*

Main challenges complicating the improvement of STMicroelectronics' (ST) 55-nm SiGe BiCMOS process (BiCMOS055) were reported. Firstly, f_T is limited by the CMOS node's thermal budget similar to the IHP technology and ST targets optimization of the vertical profile to overcome this limitation. Secondly, f_{MAX} is restricted by the extrinsic-to-intrinsic base link. The new emitter-base architecture will be introduced to solve this issue [9]. Furthermore, ST worked on technology modifications to make the technology cost-effective. A highly doped buried layer at the collector part of their HBTs was not cost-effective. Therefore, ST aimed a new fully implanted collector architecture [3]. ST could develop an experimental HBT by solving the first and third challenges and could improve 55-nm HBT's f_T to 450 GHz. However, the HBTs f_{MAX} was degraded to 160 GHz because of very high base resistance at the end [2]. ST projects the development of advanced BiCMOS055X technology with f_T around 400 GHz and f_{MAX} greater than 600 GHz for the future [9].

D. *SiGe BiCMOS Process of GlobalFoundries*

U.S. Defense Advanced Research Projects Agency (DARPA) launched T-MUSIC (Technologies for Mixed-mode Ultra Scaled Integrated Circuits) project aiming the development of advanced SiGe BiCMOS process similar to the European TARANTO project. As part of the T-MUSIC project, GlobalFoundries (GF) demonstrated SiGe HBTs integrated with 45-nm partially depleted SOI CMOS node and presenting f_T/f_{MAX} 375/510 GHz. Remarkably, GFs HBTs exhibit a current gain (β) greater than 2500 which is a very significant current gain. All the details of the process are discussed in [14].

IV. SiGe BiCMOS State of Art Circuit Blocks

The performance of these technologies can be evaluated according to RF performance criteria such as gain, power handling capacity (output power), linearity, and minimum noise figure (NF_{min}).

A. *A 300 GHz Low-Noise Amplifier in 130-nm SiGe SG13G3 Technology*

One of the first published papers on IHP's new 130-nm SiGe SG13G3 technology is on a 300 GHz low-noise amplifier (LNA). The proposed LNA features a 68 GHz 3dB BW, 10.8 dB peak gain and notable noise performance at frequencies around 300 GHz with NF_{MIN} 11 dB. The details of the LNA are reported in [1].

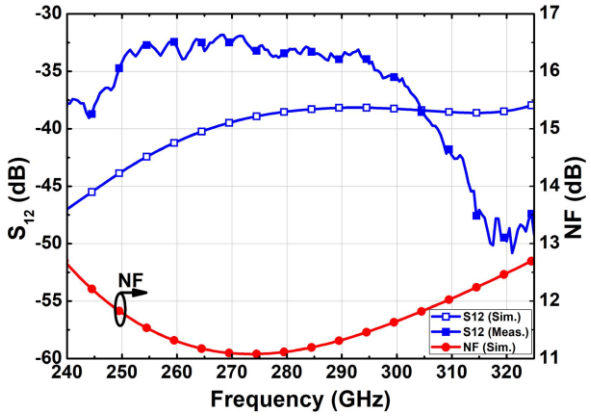


Fig. 2. Isolation and noise figure of proposed LNA [1].

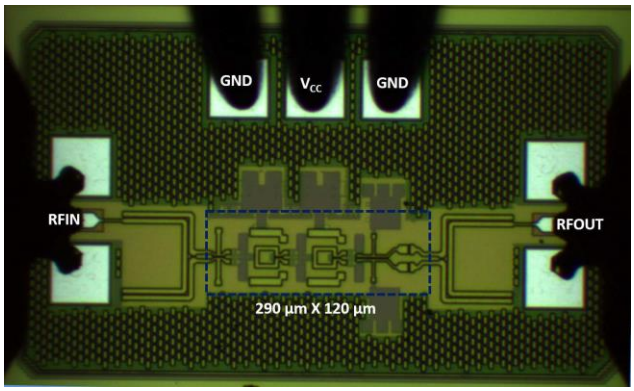


Fig. 3 Micrograph of the chip manufactured with SG13G3 technology [1].

B. *Performance Comparison of Broadband Traveling Wave Amplifiers in 130-nm SiGe:C SG13G2 and SG13G3 BiCMOS Technologies*

Another research done on the SG13G3 technology is the design of broadband Travelling Wave Amplifiers (TWA) in both predecessor SG13G2 and successor SG13G3 technologies. The research gives a significant idea about how IHP HBTs were improved. The TWA architectures used in both designs are identical for fair performance comparison. Differential topology was employed for the TWAs which consist of 6 Variable Gain Cells (VGC). The design steps of the TWAs are discussed in

[15]. According to the measurement, there is a remarkable performance improvement. TWA in SG13G3 technology features 16.9 dB DC gain and BW greater than 110 GHz while TWA in SG13G2 technology offers 15.3 dB DC gain and 87.4 GHz bandwidth. Moreover, the TWA in SG13G3 provides 14.1 dBm OP1dB while the SG13G2 TWA provides 13.3 dBm OP1dB. Thus, the linear power efficiency of the SG13G2 was improved from 4.1% to 5.1% with the new technology [15].

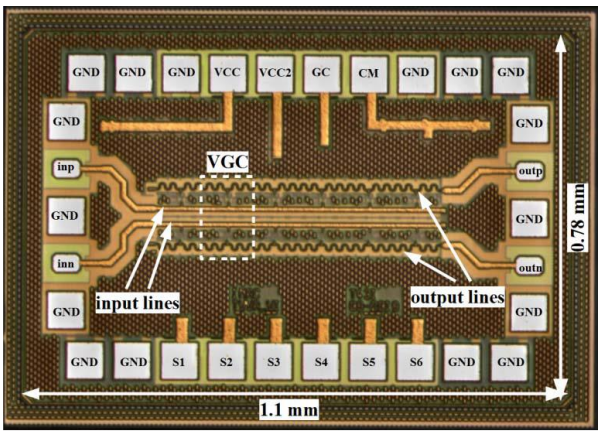


Fig. 4 Micrograph of TWA manufactured with SG13G3 technology [15].

C. D-Band SiGe BiCMOS Power Amplifier

A D-Band SiGe BiCMOS Power Amplifier (PA) was designed in STMicroelectronics' current 55-nm SiGe BiCMOS process with f_T/f_{MAX} 320/370 GHz. The power amplifier operating at 135 GHz center frequency with 34 GHz 3 dB BW is composed of a cascade stage and three cascaded common base stages. It exhibits notable power added efficiency (PAE) 17.1 % with 16.8 dBm P1dB. The power efficiency of the PA is three times better than other state-of-art PAs [12].

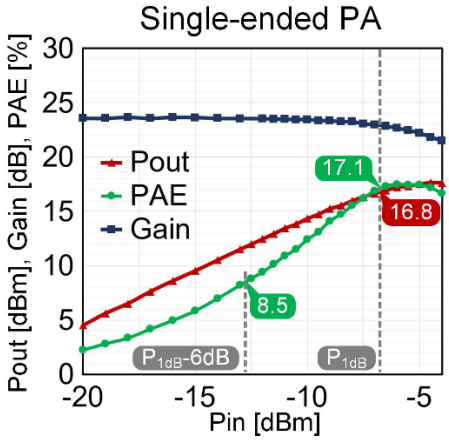


Fig. 5 Pout, Gain, PAE vs Pin graph of proposed PA [12].

V. Future Directions

Sub-millimeter-wave frequency band offers vast available frequency band for future electronics applications. As frequency bands below the terahertz gap are already utilized, the need for advanced device technologies to develop systems that can operate in the terahertz gap has increased. The improvements in the SiGe BiCMOS technologies pave the way for these systems. As mentioned before, the existing SiGe BiCMOS technologies in production feature f_T/f_{MAX} around 300/400 GHz and they are not sufficient to develop systems operating at high frequencies of the terahertz gap. They provide limited gain and output power at high frequencies. Therefore, typical methods to overcome this issue such as frequency multipliers, subharmonic mixers and parallel connections of the frequency multipliers are required. However, these methods are not efficient in terms of power and chip area. Therefore, f_T/f_{MAX} must be pushed further together [6].

In order to increase f_T and f_{MAX}, high collector current density J_C, low base-resistance R_B and base-collector capacitance C_{BC} are required. The effect of resistive and capacitive transistor parasitics limiting f_T/f_{MAX} can be minimized by enhancing metallization schemes featuring different metal layers which are thick and properly distanced to the substrate for the connection of different device terminals. HBT's performance is degraded by thermal budget of CMOS process as well. The SiGe HBT process steps must be improved to reduce effect of thermal budget on HBTs [16]. Additionally, the contact resistivity contributing resistances at terminals can be reduced by developing new contact material compositions. Also, collector current density J_C is limited since high J_C can result in electro migration. The enhanced metallization scheme and new contact material compositions can allow higher collector current density [16].

Furthermore, CMOS nodes integrated with the BiCMOS process are behind state of the art compared to the digital CMOS-only process. For this reason, European and the U.S. the SiGe BiCMOS manufacturers target the development of more advanced CMOS nodes [8].

Moreover, silicon photonic devices are promising in high-speed and long-distance communication. Integrating the silicon photonic and SiGe BiCMOS devices will reduce the parasitics due to interfacing and manufacturing costs [6].

VI. Conclusion

As a result of intense effort on development of next generation SiGe BiCMOS technologies, the state-of-the-art technology featuring f_T/f_{MAX} 470/700 GHz was reported by IHP. Foremost SiGe BiCMOS device manufacturers such as IHP, Infineon, ST and GF carry out research projects to push the limits of the technology. In the future, HBTs with high f_T and f_{MAX} will allow designing integrated circuits and systems operating at frequencies close to terahertz. Terahertz gap presents vast available frequency band. Thus, integration of HBTs with CMOS nodes promises high data rate communications. The first published studies on design of RF circuit blocks in the next generation SiGe BiCMOS technologies demonstrated that high gain & output powers and low noise performance can be obtained at high frequencies. The improvements in the technology processes, metallization layers, and transistor architectures will open the way for the SiGe BiCMOS.

REFERENCES

[1] A. Gadallah, M. H. Eissa, T. Mausolf, D. Kissinger and A. Malignaggi, "A 300-GHz Low-Noise Amplifier in 130-nm SiGe SG13G3 Technology", in IEEE Microwave and Wireless Components Letters, vol. 32, no. 4, pp. 331-334, April 2022, doi: 10.1109/LMWC.2021.3128762.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] A. Gauthier, W. Aouimeur, E. Okada, N. Guitard, P. Chevalier and C. Gaquière, "A 30.1 mW / µm2 SiGe:C HBT Featuring an Implanted Collector in a 55-nm CMOS Node", in IEEE Electron Device Letters, vol. 41, no. 1, pp. 12-14, Jan. 2020, doi: 10.1109/LED.2019.2954600.

[3] A. Gauthier et al., "450 GHz fT SiGe:C HBT Featuring an Implanted Collector in a 55-nm CMOS Node", in 2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), 2018, pp. 72-75, doi: 10.1109/BCICTS.2018.8551057.

[4] A. Mai and M. Kaynak, "SiGe-BiCMOS based technology platforms for mm-wave and radar applications", in 2016 21st International Conference on Microwave, Radar and Wireless Communications (MIKON), 2016, pp. 1-4, doi: 10.1109/MIKON.2016.7492062.

[5] B. Heinemann et al., "SiGe HBT with fx/fmax of 505 GHz/720 GHz", in 2016 IEEE International Electron Devices Meeting (IEDM), 2016, pp. 3.1.1-3.1.4, doi: 10.1109/IEDM.2016.7838335.

[6] D. Kissinger, G. Kahmen and R. Weigel, "Millimeter-Wave and Terahertz Transceivers in SiGe BiCMOS Technologies", in IEEE Transactions on Microwave Theory and Techniques, vol. 69, no. 10, pp. 4541-4560, Oct. 2021, doi: 10.1109/TMTT.2021.3095235.

[7] D. Manger et al., "Integration of SiGe HBT with fT = 305 GHz, fMAX = 537 GHz in 130nm and 90nm CMOS" 2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), 2018, pp. 76-79, doi: 10.1109/BCICTS.2018.8550922.

[8] E. Preisler, "A Commercial Foundry Perspective of SiGe BiCMOS Process Technologies" in 2020 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), 2020, pp. 1-5, doi: 10.1109/BCICTS48439.2020.9392971.

[9] ECSEL JU, "TowARds Advanced bicmos NanoTechnology platforms for rf and thz applicatiOns 1.2.2 Final report on device optimizations and BiCMOS integration", Accessed: Jun. 22, 2022. [Online]. Available: https://cordis.europa.eu/project/id/737454/results

[10] ECSEL JU, "TowARds Advanced bicmos NanoTechnology platforms for rf and thz applicatiOns Final Publishable Summary", Accessed: Jun. 22, 2022. [Online]. Available: https://cordis.europa.eu/project/id/737454/results

[11] H. Rücker, "High-performance SiGe HBTs for next generation BiCMOS technology", in Semiconductor Science and Technology, vol. 33, no. 11, Oct. 2018, doi: https://doi.org/10.1088/1361-6641/aade64.

[12] I. Petricli, D. Riccardi and A. Mazzanti, "D-Band SiGe BiCMOS Power Amplifier With 16.8dBm PıdB and 17.1% PAE Enhanced by Current-Clamping in Multiple Common-Base Stages," in IEEE Microwave and Wireless Components Letters, vol. 31, no. 3, pp. 288-291, March 2021, doi: 10.1109/LMWC.2021.3049458.

[13] J. J. Pekarik et al., "A 90nm SiGe BiCMOS technology for mm-wave and high-performance analog applications" , in 2014 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2014, pp. 92-95, doi: 10.1109/BCTM.2014.6981293.

[14] J. Pekarik et al., "SiGe HBTs with fT/fMAX ~ 375/510GHz Integrated in 45nm PDSOI CMOS", in 2021 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), 2021, pp. 1-4, doi: 10.1109/BCICTS50416.2021.9682454.

[15] M. Inac, A. Fatemi, F. Korndörfer, H. Rücker, F. Gerfers and A. Malignaggi, "Performance Comparison of Broadband Traveling Wave Amplifiers in 130-nm SiGe:C SG13G2 and SG13G3 BiCMOS Technologies", in IEEE Microwave and Wireless Components Letters, vol. 31, no. 6, pp. 744-747, June 2021, doi: 10.1109/LMWC.2021.3067099.

[16] M. Schröter et al., "SiGe HBT Technology: Future Trends and TCAD-Based Roadmap", in Proceedings of the IEEE, vol. 105, no. 6, pp. 1068-1086, June 2017, doi: 10.1109/JPROC.2015.2500024.

[17] N. Rinaldi, M. Schröter, "Silicon-germanium heterojunction bipolar transistors for MM-wave systems: Technology, Modeling and Circuit Applications", River Publishers, 2018.

[18] P. Chevalier et al., " s", in Proceedings of the IEEE, vol. 105, no. 6, pp. 1035-1050, June 2017, doi: 10.1109/JPROC.2017.2669087.

[19] P. Chevalier et al., "SiGe BiCMOS Current Status and Future Trends in Europe", in 2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), 2018, pp. 64-71, doi: 10.1109/BCICTS.2018.8550963.

[20] S. Phillips et al., "Advances in foundry SiGe HBT BiCMOS processes through modeling and device scaling for ultra-high speed applications", in 2021 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), 2021, pp. 1-5, doi: 10.1109/BCICTS50416.2021.9682485.

[21] S. P. Voinigescu, S. Shopov, J. Bateman, H. Farooq, J. Hoffman and K. Vasilakopoulos, "Silicon Millimeter-Wave, Terahertz, and High-Speed Fiber-Optic Device and Benchmark Circuit Scaling Through the 2030 ITRS Horizon," in Proceedings of the IEEE, vol. 105, no. 6, pp. 1087-1104, June 2017, doi: 10.1109/JPROC.2017.2672721.

[22] T. Zimmer et al., "SiGe HBTs and BiCMOS Technology for Present and Future Millimeter-Wave Systems" in IEEE Journal of Microwaves, vol. 1, no. 1, pp. 288-298, Jan. 2021, doi: 10.1109/JMW.2020.3031831.

[23] V. P. Trivedi et al., "A 90nm BiCMOS technology featuring 400GHz fMAX SiGe:C HBT", in 2016 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2016, pp. 60-63, doi: 10.1109/BCTM.2016.7738951.

[24] X. Li et al., "A 250–310 GHz Power Amplifier With 15-dB Peak Gain in 130-nm SiGe BiCMOS Process for Terahertz Wireless System", in IEEE Transactions on Terahertz Science and Technology, vol. 12, no. 1, pp. 1-12, Jan. 2022, doi: 10.1109/TTHZ.2021.3099057.

Degradation Sensitive Electrical Parameters in SiC-MOSFETs

Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—The pursuit of high-efficiency converters leads to an increasing use of SiC-MOSFETs. To ensure reliability of SiC-based power converters and prevent device failures, the early detection of degradation is required. Electrical parameter changes caused by degradation are observed by condition monitoring and their relevance to the predictability of the device's lifetime is assessed. This paper presents state-of-the-art condition monitoring methods for SiC-MOSFETs based on Degradation Sensitive Electrical Parameters. As aging precursors, parameters such as the device's turn-on time, the on-state voltage and the drain-source resistance are examined. The SiC-MOSFETs are aged by accelerated aging tests and the resulting changes in electrical parameters are studied to assess their correlation with the device's lifetime.**

The results show that the increase of the turn-on time reflects the increasing degradation of the device. The rise in on-state voltage is caused by both die and package degradation. Specifically, oxide degradation can be detected by an increase in the drain-source resistance in saturation region and a rise in the MOSFET on-state resistance indicates package-related degradation.

***Index Terms*—condition monitoring, silicon carbide, power cycling, lifetime, temperature sensitive electrical parameter, electron traps**

I. Introduction

The superior performance of silicon carbide (SiC) power metal-oxide semiconductor field-effect transistors (MOSFETs) compared to silicon counterparts, facilitate high-frequency, high-efficiency, and thus high-power-density applications [1]. However, the lack of information on the long-term use of SiC-MOSFETs due to its relatively new technology causes concern with regard to their reliability [2]. Remedy can be provided by using condition monitoring systems, which surveil variations of Degradation Sensitive Electrical Parameters (DSEPs) to predict the device's lifetime and degradation mechanisms [2].

Power cycling demonstrates aging by repeatedly turning a device on and off, leading to alternate heating and cooling of the device. The resulting temperature fluctuations cause thermo-mechanical stress between surfaces of different materials since the layers of a power module consist of materials with different thermal expansion coefficients. This leads to material deformations during heating and cooling of the module, which results in its degradation and can then cause system failures. Condition monitoring represents an approach for the maintenance of a device by observing aging precursors to predict its lifetime.

II. DSEPs in SiC-MOSFETs

Parameters that demonstrate significant changes with aging are the gate threshold voltage and the gate leakage current, but these are difficult to measure and are therefore not further mentioned [3]. The variation in switching transients with aging is indicative of the reliability of the device by using the device turn-on delay as an aging precursor for SiC-MOSFETs [4]. In addition, the change in package resistance and channel resistance with increasing stress time effectuates a change in the on-state voltage [5]. Moreover, near interface oxide traps at the SiC/SiO_2 interface of the SiC-MOSFET trap electrons and increase the MOSFET drain-source resistance in saturation region, resulting in degraded reliability, indicating die-related degradation [3]. Package-related degradation can be detected by the variation of device on-state drain-source resistance [3].

III. State-of-the-art implementations of DSEP measurements in SiC transistors

A. Using the turn-on time

Reference [4] implements a condition monitoring system using the device turn-on time delay as a DSEP. The turn-on time is measured from a gate-source voltage V_{gs} of 0 V to a drain-source voltage V_{ds} of 10% of the DC-link voltage [4].

Fig. 1 presents the schematic of the circuit for the turn-on time monitoring method.

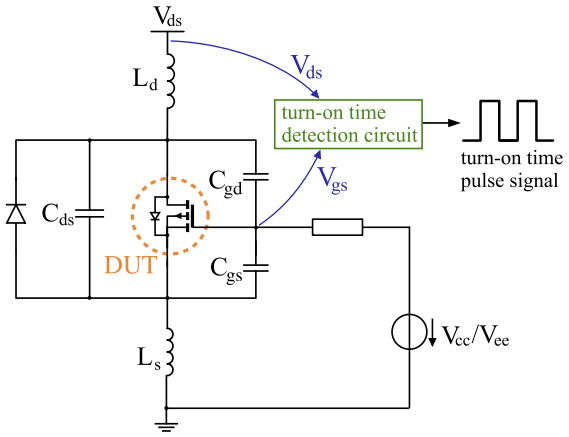


Fig. 1. Schematic of the circuit for turn-on time monitoring method (adapted from [4])

The turn-on time detection circuit is designed to determine the exact turn-on time of the Device Under Test (DUT). It consists of voltage dividers to downscale the input voltages V_{ds} and V_{gs}. A high-resolution capture module detects the rising edge of V_{gs} and the falling edge of V_{ds} and measures the time between the two edges. Voltage pulses are then generated whose width corresponds to the turn-on time of the DUT [4].

The resulting measured waveform of the turn-on pulse is shown in Fig. 2.

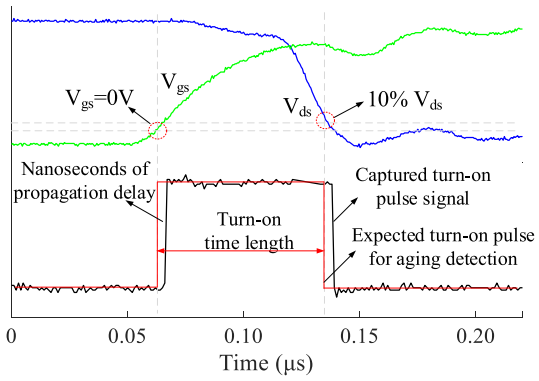


Fig. 2. Measurement of the turn-on pulse [4]

The delay of the captured turn-on pulse signal can be neglected since it does not depend on the aging process. However, the junction temperature T_j affects the switching transient and thus the measurement is executed at system startup when the device has not yet warmed up [4].

As shown in Fig. 3, the turn-on time increases with advancing age.

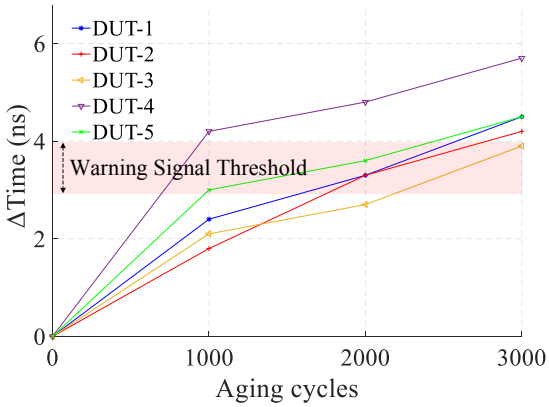


Fig. 3. Measured turn-on time with advancing age [4]

The same stress test was carried out with 5 supposedly identical DUTs but whose turn-on times differ from one another. This highlights the necessity to monitor each device individually. To avoid system failure, a warning can be communicated to the user as soon as the turn-on time exceeds a defined value, meaning that the device is expected to fail shortly [4].

B. Using the on-state voltage

Reference [5] investigates the effects of stress on SiC-MOSFETs in an actual converter. For this purpose, the on-state voltage V_{ds} of the SiC-MOSFET is continuously monitored under accelerated stress in a Power Factor Correction (PFC) boost converter [5].

The selected PFC boost converter consists of a diode rectifier bridge and a boost converter, as depicted in Fig. 4.

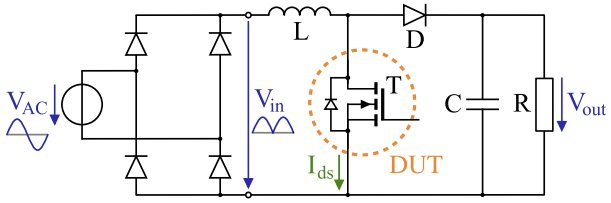


Fig. 4. Schematic of a PFC boost converter circuit (adapted from [5])

The AC-DC conversion in steady-state operation causes periodic fundamental power losses on the SiC-MOSFET, causing thermal cycling stress. By changing the input current, output voltage and switching frequency, the thermal stress applied to the DUT can be controlled [5].

The aging test is performed with TO-247 packaged SiC-MOSFETs under accelerated stress by selecting a stressed operating point. Therefore, the switching frequency is increased to 200 kHz so that switching losses are higher and the junction temperature exceeds 150°C. In addition, a high operation current of 12 A is chosen. The selected high-temperature stress conditions significantly accelerate gate oxide degradation [5].

The on-state voltage V_{ds} is selected as an aging precursor for continuously monitoring the degradation of the SiC-MOSFET during converter operation. The V_{ds} measurement principle used is based on the desaturation protection measurement and extracts V_{ds} at I_{ds} = 12 A [5].

Fig. 5 depicts the measured V_{ds} of two samples of the same SiC-MOSFET with increasing stress time.

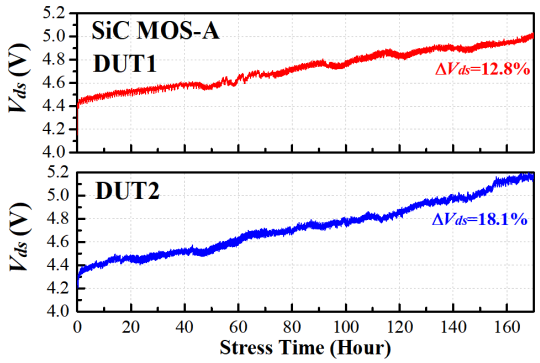


Fig. 5. V_{ds} measurement of two DUTs at 12 A over stress time [5]

V_{ds} gradually rises over time, reaching an increase of 12.8% for DUT1 and 18.1% for DUT2 from its initial value at 168 hours. The slightly different amplitude of the variations may be due to manufacturing differences. The cause of V_{ds} variation can be attributed to both die and package degradation [5].

C. Using the drain-source resistance

The cause of the variation in drain-source resistance can be differentiated between die-related and package-related aging, as shown in [3]. The increase in the MOSFET on-state resistance $R_{\mathrm{ds-on}}$ in case of bond wire lift-off or die-attach solder delamination makes it an effective precursor for package-related aging detection. A suitable aging precursor for detection of oxide degradation is the drain-source resistance in saturation region $R_{\mathrm{ds-sat}}$ due to its dependency on the threshold voltage V_{th} [3].

The drain-source resistances are measured with system current and dc-link voltage sensors at system startup, enabling a relatively slow and simple V_{ds} measurement circuit since V_{ds} is not measured during normal converter operation with high switching frequencies. The schematic of the circuit for $R_{\mathrm{ds-sat}}$ and $R_{\mathrm{ds-on}}$ measurement is illustrated in Fig. 6 (a) and (b).

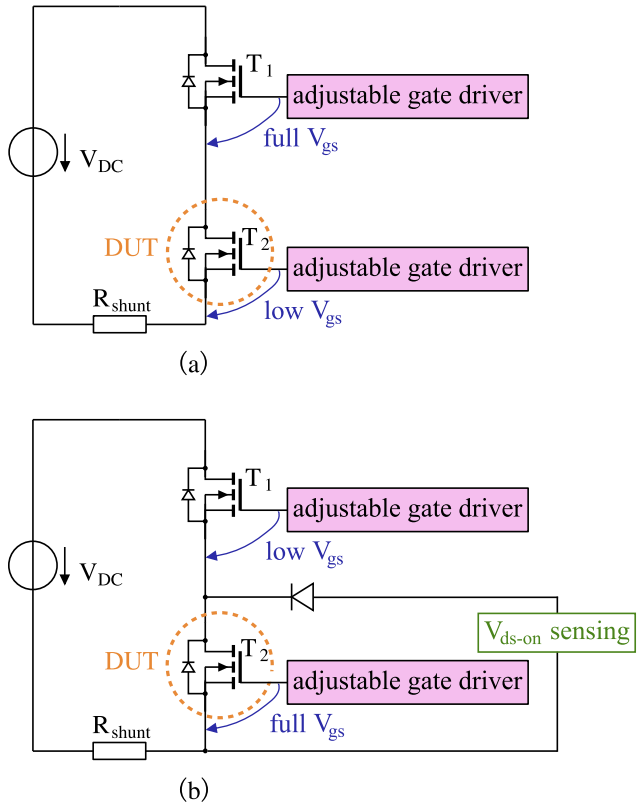


Fig. 6. Schematic of the circuit for (a) $R_{\mathrm{ds-sat}}$ and (b) $R_{\mathrm{ds-on}}$ monitoring method in phase-leg configuration (adapted from [3])

For the implementation, one switch of a phase leg is turned on at low gate voltage so it operates in saturation region and the other switch operates in on-state with full gate voltage, as shown in Fig. 6 (a). An adjustable gate driver is used for the switches to actively vary the applied gate-source voltage V_{gs}. Both MOSFETs of the leg are turned on and its current is limited by the switch operating in saturation region due to the high $R_{\mathrm{ds-sat}}$ of the DUT. The low on-state resistance of the MOSFET operating in ohmic region is neglected, meaning that the resistance of the leg corresponds to $R_{\mathrm{ds-sat}}$ of the DUT. This means $R_{\mathrm{ds-sat}}$ is obtained by dividing the measured voltage across the leg from the DC-link voltage sensor by the measured current through the leg from the current sensor R_{shunt} [3].

For the determination of $R_{\mathrm{ds-on}}$, the DUT is turned on at full gate voltage and the other switch operates at low gate voltage, as illustrated in Fig. 6 (b). This again limits the current in the leg, ensuring safe operation. For $R_{\mathrm{ds-on}}$ calculation, the voltage drop across the switch in on-state $V_{\mathrm{ds-on}}$ is divided by the measured current of the sensor [3].

To ensure that verification can take place in a short time, the measurement is carried out using accelerated methods that demonstrate aging. An effective and well-known method is power cycling, using repetitive self-heating and forced cooling to expose the device to large thermal fluctuations. Thus, an aging cycle of such a power cycling test consists of a heat-up and a cool-down phase, as shown in Fig. 7.

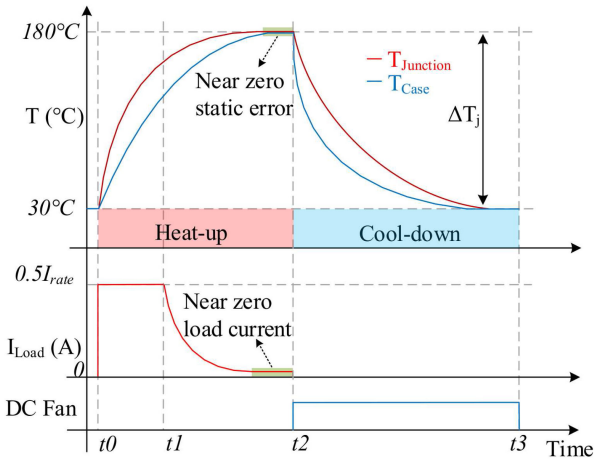


Fig. 7. Aging cycle of the used power cycling test [3]

The junction temperature T_{Junction} as well as the temperature of the device case T_{Case} of the examined SiC-MOSFET is measured with a thermocouple. For the power cycling test used in Reference [3], the device is exposed to a temperature fluctuation ΔT_{j} of 150°C during each aging cycle. Self-heating is achieved by passing half the rated current through the DUT in on-state. When the temperature almost reaches the upper limit of 180°C, a microcontroller decreases the current until it approaches near zero. At time t_2, both temperatures reach the upper limit and are nearly equal. The cool-down phase then begins, in which the DUT is switched off and a DC fan runs until the device has cooled down. This completes the aging cycle, which lasts around 3 minutes [3].

Since thermally triggered degradation processes in SiC components lead to defect charges being trapped in the gate oxide and being attracted by an applied gate bias, a higher gate voltage V_{gs} and thus a higher threshold voltage V_{th} is required. That means the drift of V_{th} is a DSEP since it indicates oxide degradation. However, its measurement for

determining device degradation is not advisable due to the high-speed, high-resolution circuit required [3].

With increasing device degradation, the formation of the conduction channel is weakened so that V_{th} increases and a greater channel resistance in saturation region $R_{\mathrm{ds-sat}}$ is expected. $R_{\mathrm{ds-sat}}$ variation indirectly shows changes of V_{th} and can therefore be used as gate oxide degradation detection. The changes of V_{th} with aging also effect the drain-source resistance in on-state $R_{\mathrm{ds-on}}$, indicating package-related degradation [3].

Fig. 8 (a) and (b) illustrate the measured drain-source resistances in saturation and ohmic regions, with the current being sensed through shunt resistors.

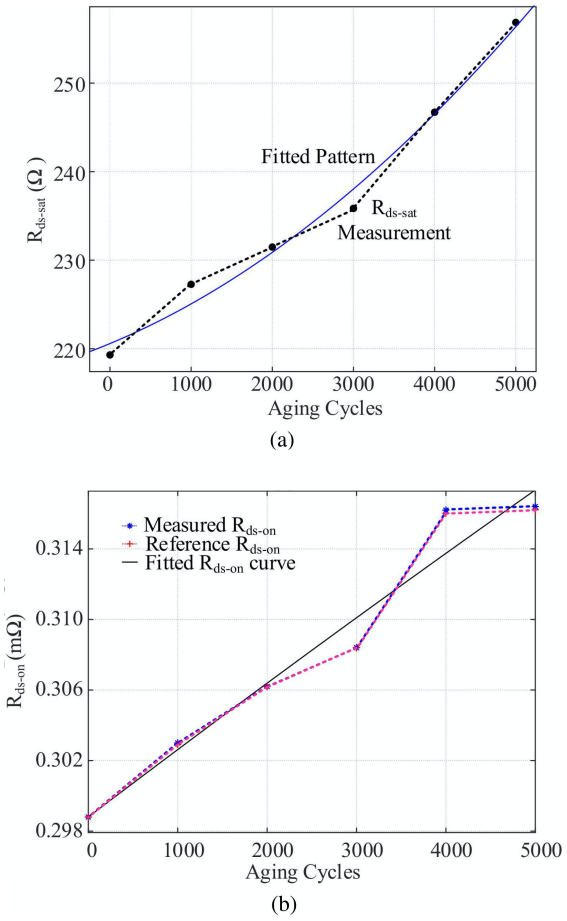


Fig. 8. Measured (a) $R_{\mathrm{ds-sat}}$ and (b) $R_{\mathrm{ds-on}}$ over aging cycles [3]

Fig. 8 (a) depicts the measured $R_{\mathrm{ds-sat}}$ of the DUT. After 5000 aging cycles, $R_{\mathrm{ds-sat}}$ has increased by 38 Ω compared to when the device was new. The curve fitted to the measurement points in blue increases steadily over lifetime.

Fig. 8 (b) shows an increase of $R_{\mathrm{ds-on}}$ over the aging cycles. Up to 3000 cycles, this can be assigned to gate oxide level degradation. A rapid rise of $R_{\mathrm{ds-on}}$ is visible after 3000 cycles and is due to package-related degradation. This includes cracks in the bond-wire and bond-wire lift-offs from die attachment that were detected after 3000 cycles [3].

IV. Comparison of Measurement Approaches

The measurement approaches using the turn-on time as well as the drain-source resistance conduct the DSEP measurement at system startup. This is mandatory since DSEPs are also TSEPs, resulting in different measurement results for warmed up devices. In addition, carrying out the measurement at startup rather than during normal converter operation allows measurement at low switching frequency.

The on-state voltage of the SiC-MOSFET in a converter is monitored during converter operation, allowing continuous knowledge of its state of health.

The determination of the turn-on time is realized with a high-resolution capture module for edge detection. On the other hand, the measurement of the drain-source resistance requires current and voltage sensors. The discussed degradation monitoring methods for SiC-MOSFETs are all applicable for degradation detection. However, the advantage of using the drain-source resistance as a DSEP is its ability to indicate die-related and package-related degradation separately. This allows conclusions to be drawn as to what fails first. And thus provides relevant long-term experience for manufacturers, who can make corresponding improvements to the device accordingly.

V. Conclusion

Overall, condition monitoring methods are essential for lifetime prediction of SiC-MOSFETs. While there are many parameters that reflect degradation, not all are equally well suited for condition monitoring since some require too high resolution for their precise measurement. This means that the monitored DSEPs have to be well-considered.

This paper presented three state-of-the-art implementations of DSEP measurements in SiC-MOSFETs. By monitoring the turn-on time based on the device switching transients in power cycling, the on-state voltage in stressed converter operation and the drain-source resistance in phase-leg configuration over the aging cycles, inferences can be drawn about the state of health of the device. The verification of the presented implementations of DSEP measurements was based on the results of stress tests that accelerate aging mechanisms. In the experimental results, an increase in the turn-on time was observed for aged devices. Reliability can be ensured by setting a turn-on time limit above which a warning signal is given to the user, so that the device can be replaced in time to avoid system failure. A continuous rise in the on-state voltage of the SiC-MOSFET with increasing stress time was detected. The measurement results turned out to be different despite using the same components, meaning that individual monitoring of the converter's SiC-MOSFETs is mandatory for reliable operation. An increase in the drain-source resistance in saturation region was detected and can be attributed to gate oxide degradation. Whereas a rise in the drain-source resistance in ohmic region indicates package-related degradation of the DUT. The investigated DSEPs, the device turn-on time, the on-state voltage and its drain-source resistance were assessed as reliable aging precursors.

ACKNOWLEDGMENT

Thanks to the subject matter expert Kevin Muñoz Barón at the ILH for submitting and supervising this interesting topic.

REFERENCES

[1] J. Wang and X. Jiang, "Review and analysis of SiC MOSFETs' ruggedness and reliability," 2020 IET Power Electronics, 13: 445-455. https://doi.org/10.1049/iet-pel.2019.0587.

[2] E. Ugur, C. Xu, F. Yang, S. Pu and B. Akin, "A New Complete Condition Monitoring Method for SiC Power MOSFETs," in IEEE Transactions on Industrial Electronics, vol. 68, no. 2, pp. 1654-1664, Feb. 2021, doi: 10.1109/TIE.2020.2970668.

[3] S. Pu, F. Yang, B. T. Vankayalapati, E. Ugur, C. Xu and B. Akin, "A Practical On-Board SiC MOSFET Condition Monitoring Technique for Aging Detection," in IEEE Transactions on Industry Applications, vol. 56, no. 3, pp. 2828-2839, May-June 2020, doi: 10.1109/TIA.2020.2980220.

[4] S. Pu, E. Ugur, F. Yang and B. Akin, "In situ Degradation Monitoring of SiC MOSFET Based on Switching Transient Measurement," in IEEE Transactions on Industrial Electronics, vol. 67, no. 6, pp. 5092-5100, June 2020, doi: 10.1109/TIE.2019.2924600.

[5] X. Jiang, J. Wang, J. Chen, H. Yu, Z. Li and Z. John Shen, "Investigation on Degradation of SiC MOSFET Under Accelerated Stress in a PFC Converter," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 9, no. 4, pp. 4299-4310, Aug. 2021, doi: 10.1109/JESTPE.2020.2988447.

Anomaly Detection using Embedded Machine Learning

Electrical Engineering
Universität Stuttgart
Stuttgart, Germany

***Abstract*—Anomaly Detection in Power electronics is important for safety and dependability. One of the primary objectives for Anomaly detection in Power electronics is to extend the lifespan and MTBF (Mean Time Between Failures). In order to understand how ML (Machine Learning) implementations on embedded devices might be used to extend the lifespan of power electronic equipments, this study focuses on a general survey of embedded ML implementations. Additionally, anomaly detection, particularly on microcontrollers, necessitates taking into account some crucial elements like memory use, computation power, and so forth. A significant amount of data from monitored variables under various nominal and erroneous conditions is needed for anomaly detection in order to increase MTBF. As a result, an efficient strategy for the application of ML is required. The State-of-the-Art frameworks and techniques for embedded ML are also covered in this article.**

***Index Terms*—machine learning, anomaly detection, embedded systems.**

I. Introduction

Embedded systems, especially those that are safety-critical, require a higher level of dependability. Many fault tolerance solutions like the Corrective Maintenance (CM) and Preventive Maintenance (PM) are not as effective and efficient as the Predictive Maintenance (PdM) [10]. In CM, devices are only repaired after a failure takes place. Due to the scheduled maintenance schedule in PM, more regular maintenance is necessary. Thus, PdM is preferred as it is reliable by automatic planning of the maintenance prior to the failure which also reduces the frequency of maintenance.

Anomaly detection (AD) is the process of identifying events or observations that are marked differently from the rest of the data being evaluated. If a facility's temperature deviates from the projected value for an extended period of time, a smart HVAC system, for instance, may use anomaly detection to monitor building temperatures and send out an alarm. System or device health monitoring by studying the device's features (like voltage, current, temperature and so on) are further scenarios.

II. Anomaly Detection using Machine Learning

A. Strategies for Anomaly Detection

Several AD strategies have been developed for embedded devices. These methods can be divided into three groups: condition-based, model-based, and data-driven methods [3]. Condition-based methods identify, estimate, or measure the electrical parameters, which show the power device's deterioration. Model-based techniques create an analytical or simulation model that links physical parameters to the observed behavior of the device using knowledge of the physics of the device's normal operation and anticipated physics of failure mechanisms. When physics-based models are unavailable, too complex for real-time application, or difficult to construct, a data-driven method for anomaly detection is recommended. Using a data-driven strategy for AD is somewhat effective because condition based and model based techniques are challenging to implement for real-time application. As a result, AD based on ML is preferred for embedded devices as ML itself is data driven. As technology, techniques, and algorithms advance, ML models are increasingly incorporated into computationally limited devices, such as microcontrollers [1]. In order to successfully implement AD on embedded systems using ML, it is necessary to research several implementation strategies.

B. Machine Learning concepts in Power electronics

Under the traditional cloud-based Internet of Things (IoT) architecture design, all data must be sent to the cloud, which causes unpredictable latency, privacy issues, and non-real-time features like the necessity for time and resources to send all the data to the cloud which makes the IoT system unable to act real time. It is widely observed in IoT systems that computing may become efficient if network edge devices (gateways, nodes, and so on) could be made intelligent enough to analyze data for AD effectively [4]. In order to incorporate such models at the network's end, where severely resource-constrained microcontroller units (MCUs) are located, the following study discusses the optimizations, methodologies, and platforms that are used.

The way ML is implemented into Power electronic devices depends on the device being used, the application, the amount of power needed, and the amount of storage needed. The model needs to be adjusted to fit and function well on such devices, especially microcontrollers, which have limited memory and computation capacity. The resource-constrained MCUs are the main topic of discussion in this study. In accordance with [1], specifications of a resource-scarce device are clock frequency under 200 MHz, on-board flash of no more than 3

MB, power consumption below 20 mA and (S)RAM memory no greater than 500 KB. This paper uses this specification for referring to the term resource constrained devices.

III. State-of-the-Art Implementations of Anomaly Detection in Power Electronics

Due to the scarcity of resources in most of the edge devices, the implementation of ML at the edge is quite challenging. This paper gives an overall review on AD using embedded ML in the context of three main factors to be addressed.

1.Data Preprocessing
2.Automated feature selection
3.Classification

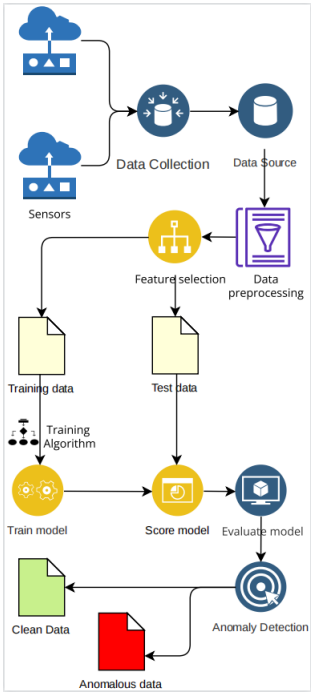


Fig. 1. Overview of Anomaly detection in IoT architecture (adapted from [7]

A. Data Preprocessing

Imbalanced classes are a significant barrier to AD's effectiveness, particularly in IoT systems [1]. The reason is because IoT systems handle enormous amounts of text, image, sensor, and other data. Imbalanced classes mean that one particular class of data having a high occurence than another class of data. For example, faults in real time are greatly exceeded by instances of good performance (faultless samples) [13]. This imbalanced dataset would make it difficult to generalize the model to new data, despite the model appearing to generate high accuracy [1]. As a result, the test error is greater than the training error for the model. Pre-processing is therefore a crucial step in developing the ML model for AD. Over-representation of classes is reduced by using Auto Encoders (AE) and Neural Networks (NN) at the edge side [1].

Data analysis-based approaches can aid in the efficient pre-processing of the large amount of data. A correlation-based AD methodology was given in [1]. The concept stood out because it went beyond the conventional use of sensor data for Predictive Maintenance (PdM) to incorporate correlation between sensors. This tactic works well because an anomaly causes a sudden change in the correlation. Using a real-world dataset, the correlation between each pair of sensors was identified and converted into a latent correlation vector. After that, a probabilistic model was applied to the vector in order to look for anomalies. This method demonstrates how prior data processing to produce the correlation vector allowed the AD to be improved.

B. Automated feature selection

In order to have an effective AD, it is crucial to select the features with the high relevance.

For data-driven analysis to be successful, a sizable volume of data under different nominal and incorrect scenarios is required. Furthermore, it is important to understand which signals or transformations of the signals (referred to as "features") must be used in order to detect malfunctions at an early stage of failure. For tracking the degradation, there are a number of potential signatures and precursors known as condition indicators. For example, according to [5], for AD in MOSFET, these comprise the gate leakage current, I_{DS}, V_{DS}, variations in the resistance of the ON-state (R_{DS}(ON)), gate voltage, V_{GS}, and others. A method for detecting anomalous states in a DC/DC converter is proposed in [5] by statistical feature estimation utilizing Gaussian Process regression (GPR) and Genetic algorithm (GA). However, rather than an automatic feature selection, which is widely desirable, this type of feature selection might result in a converter-specific approach. The sequential forward selection and sequential backward selection techniques are automatic feature selection techniques suggested in [3].

Sequential Forward Selection (SFS): In sequential forward selection, an empty set 'F_0' is first created. In each iteration the most significant single feature x^+ is added to the feature set 'F_k', until the desired number of features 'D' is reached. This is done based on the approach of minimizing the cost function which is the average loss over the entire training dataset.

Sequential Backward Selection (SBS): In the sequential backward selection, at first, we have the set of all features and then the least significant feature is knocked out until the desired number of features is reached.

Sequential Floating Forward Selection (SFFS), an extension of the SFS approach, is a more effective feature selection technique that is detailed in this study. The main disadvanatge of SFS and SBS is that the significant features are not always disitinct. For example, in case of MOSFETS, I_{DS} and V_{DS} might be significant but they are not distinct which adds some redundancy in the feature selection. So an improved feature selection method is required which selects significant as well as distinct features. Assuming 'J' as the cost function, 'd' as the number of original features and 'D' as the desired number of features. The process is started with a null set, or an empty set, making k = 0. (where k is the size of the subset)

Step1: From the feature space, the feature that improves our feature subset's performance the most (assessed by the cost function) is included. Next, we go on to step 2.

Step2: A feature is deleted only if the performance of the resultant subset improves. Return to step 1 if k=2 or an improvement cannot be made (i.e., a feature like x^+ cannot be found); otherwise, this step is repeated.

Steps 1 and 2 are repeated until k reaches the value of D.

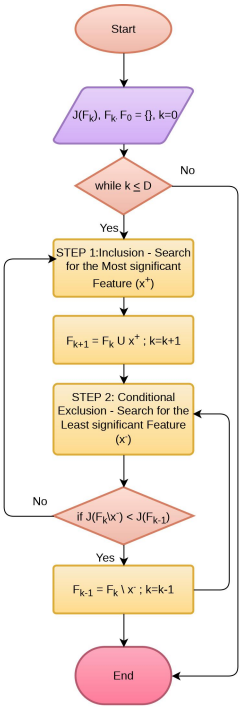


Fig. 2. Sequential Floating Forward Selection (adapted from [6])

By implementing the SFFS, you can prevent the addition of features that share the same data to the significant feature set. This stage of the AD is critical since effective preprocessing is one of the primary areas of research [1]. Floating search need to be the first tool tried for many Feature Selection problems [6].

C. Classification

In IoT AD, classification is a crucial study area. Neural networks can be used to handle anomaly classification problems, and most of these methods make use of deep neural networks. The model size should be kept as small as possible when using Deep Neural Networks by using only a few trainable parameters, and the inference time and energy should be kept as low as possible by limiting the amount of calculations [7].

A Deep Learning-based AD technique was introduced in [4], where network edge devices are made intelligent enough to handle sufficient data. The process involves LSTM (Long Short-Term Memory) implementation at the edge and efficiency testing on a Raspberry Pi. Data from real datasets are utilized to build an anomaly detector, which only uses one feature. There were three layers of LSTM, a 1D convolutional layer, and a 1D Max pooling layer. With 92 percent accuracy, a True Positive Rate (TPR) of 0.81, and a False Positive Rate (FPR) of 0.50, it was possible to distinguish anomalous data from regular data. The accurate classification of anomalous data as such, such as labeling as "abnormal" data that is actually anomalous, is known as a True Positive.The inaccurate classification of normal data as "abnormal" is known as a False Positive. However, a memory reduction technique was not covered in full in the paper [4].

IV. Algorithms for Embedded Machine Learning

The scope of this work does not extend to the wide literature on ML algorithms, their history, and their impact on various fields. However, some algorithms which made it possible to implement conventional ML algorithms in embedded devices are explained in the following sections.

A. ProtoNN

k-NN (k-Nearest Neighbors)-based algorithms are general-purpose algorithms with accuracy based on the selection of the hyperparameter "k". But k-NN can't be successfully integrated into MCUs as its requirement to calculate the distance of a given sample from each training example hinders real-time prediction on resource-constrained devices due to their low processing power.

A machine learning algorithm with a smaller storage demand is called ProtoNN, which is based on k-NN. ProtoNN chooses just a few prototype vectors/samples from this low-dimensional data after projecting it into a lower-dimensional space using a sparse-projection matrix. Due to this projection and the use of only a few representative prototypes, accuracy is lost; however, this is made up for by the joint learning process [7], [8]. In Resource Scarce MCUs, this makes it easier to integrate this method.

B. Bonsai

Prediction models benefit greatly from the accuracy, stability, and interpretability of tree-based methods. In contrast to linear models, they map non-linear interactions quite well. They can adapt to any circumstance and overcome any difficulty (classification or regression).

A ML algorithm built on decision trees that is resource-constrained is called Bonsai. The low dimensional projection is learnt in addition to the non-linear mapping. The low-dimensional projected data is used to train a single tree rather than a whole forest, and a sparse projection matrix is used to keep the model size in check [8].

C. Direct Convolutions

The ability of CNNs to process picture data makes them widely used. The use of picture data processing for AD may seem off-topic given the theme of our discussion. In contrast, the article in [11] lays out a CNN-PdM framework to draw out abnormal patterns from the data and provide more specific advice regarding the importance of executing PdM on assets.In addition, a method for converting one-dimensional

data into two-dimensional data has been proposed. It outperformed conventional ML strategies when paired with the feature representation from CNNs. Therefore, CNNs cannot be excluded from the conversation when it comes to AD.

The primary drawback of CNNs is that they consume a lot of memory. The Direct Convolution [8] method significantly reduces the memory overhead associated with using CNNs by utilizing clever memory reuse. The herringbone approach, which provably consumes optimal space [8] in addition to the space consumed by the input image, is used to scan the pixels in alternate rows- and columns-major order.

D. Fast GRNN

Processing temporal data or data that is presented in sequences is particularly effective when using a recurrent neural network (RNN). The discussion of RNN implementation in microcontrollers is inevitable given that sensor data is primarily sequential data. However, vanishing gradients and inaccurate prediction plague RNNs. Gated RNNs were therefore developed to address these problems [7]. However, using Gated RNNs in Microcontrollers would not be an effective move because they need more parameters than regular RNNs do.

The new architectures FRNN (Fast RNN) and FGRNN (Fast Gated RNN) are capable of providing good accuracies and reliable training while maintaining a modest model size. In FGRNN, the model's size was reduced by quantizing the weights such that they could only have integer values.FastGRNN uses 2-4 times less parameters than other top gated RNN models like LSTM and GRU and achieves the same or occasionally higher accuracy as gated RNN models.

TABLE I
EMBEDDED MACHINE LEARNING ALGORITHM BENCHMARKING

Algorithm	Counterpart	Test Accuracy	Model Size
ProtoNN	k-NN	<15%	787.02 KB
Bonsai	Decision Tree	37.7%	94.52KB
Direct Convolution	CNN	65.7%	58.23 KB
Fast GRNN	RNN	58.9%	63.56 KB

TABLE 1 gives an estimation of the efficacy of the aforementioned ML algorithms in resource constrained devices. The table shows that the Direct convolution algorithm outperformed the other algorithms achieving an accuracy of 65.7 percent. The ML algorithms were tested for their capability to classify images as it is one of the complex tasks. This could serve as a reference for the algorithms' efficiency when they are implemented in microcontrollers. The dataset CIFAR-10 has been used.

V. FRAMEWORKS FOR EMBEDDED MACHINE LEARNING

Numerous ML frameworks have been created and have evolved along with the adoption of ML models as de facto solutions in an ever-growing range of applications. These improvements have been paralleled in the resource-constrained end-edge-cloud architecture by tailored ML frameworks that make it easier to design lightweight models. This section discusses popular frameworks for deploying ML models on devices with resource limitations [7].

A. TensorFlow Lite (TFL)

It is a lightweight TensorFlow implementation created by Google for embedded systems and edge devices. It is a deep learning framework that is open-source and supports edge-aware learning inference. It is popular because it makes use of the five major restrictions of latency, privacy, connection, size, and power consumption, which are the main issues to be solved in the deployment of embedded ML. It is simple to enable a variety of AI applications for object detection, anomaly detection, and classification (e.g., text, image) using TFL [7], [14].

B. uTensor

uTensor is a free embedded learning environment that supports rapid deployment to IoT edge devices and prototype development. The uTensor is a compact module that occupies just 2 KB of storage space. A model is first built, and then it is defined using a quantization effect. The creation of code for appropriate edge devices is the following stage [14].

C. Apache MXNet17

Deep neural networks can be taught and deployed on edge devices with limited resources using Apache MXNet17, a lightweight, scalable, open-source framework. To speed up DNN training and deployment, MXNet offers distributed ecosystems and public cloud interface. It includes tools for tracking, troubleshooting, preserving checkpoints, and adjusting the hyperparameters of DNN models [7].

D. DeepIoT

DeepIoT condenses a neural network into smaller dense matrices while essentially maintaining the algorithm's performance. By reducing layer redundancy, this framework determines the bare minimum of filters and dimensions each layer needs. DeepIoT can reduce energy consumption by 72.2 percent to 95.7 percent, execution time by more than 71 percent, and compression of deep neural networks by more than 90 percent [7].

VI. CONCLUSION

This paper provides an overview of the necessary procedures and processes for ML implementation in embedded devices. Additionally, the State-of-the-Art frameworks and methods supporting embedded ML are also covered. The analysis of time series data from the sensors is the main method for anomaly identification in power electronics. However, other methods are also being employed, such as CNNs, which are specifically designed for picture input and produce accurate detection rates. Therefore, in order to broaden the scope of this discipline, it would be necessary to build an embedded ML approach that can interpret data types other than time series data. This article has attempted to explore this issue as well. Though, this topic has a lot of scope for new approaches both holistically and atomistically.

REFERENCES

[1] C. Lübben and M. -O. Pahl, "Advances in ML-Based Anomaly Detection for the IoT," 2021 5th Cyber Security in Networking Conference (CSNet), 2021, pp. 18-22, doi: 10.1109/CSNet52717.2021.9614280

[2] Branco, S.; Ferreira, A.G.; Cabral, J. Machine Learning in Resource-Scarce Embedded Systems, FPGAs, and End-Devices: A Survey. Electronics 2019, 8, 1289. https://doi.org/10.3390/electronics8111289

[3] Q. Yang et al., "Incipient Residual-Based Anomaly Detection in Power Electronic Devices," in IEEE Transactions on Power Electronics, vol. 37, no. 6, pp. 7315-7332, June 2022, doi: 10.1109/TPEL.2022.3140721

[4] D. Utomo and P. -A. Hsiung, "Anomaly Detection at the IoT Edge using Deep Learning," 2019 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW), 2019, pp. 1-2, doi: 10.1109/ICCE-TW46550.2019.8991929

[5] Y. Jiang, Y. Yu and X. Peng, "Online Anomaly Detection in DC/DC Converters by Statistical Feature Estimation Using GPR and GA," in IEEE Transactions on Power Electronics, vol. 35, no. 10, pp. 10945-10957, Oct. 2020, doi: 10.1109/TPEL.2020.2981500

[6] D. Setiawan, W. A. Kusuma and A. H. Wigena, "Sequential forward floating selection with two selection criteria," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2017, pp. 395-400, doi: 10.1109/ICACSIS.2017.8355064.

[7] Murshed, M. G. Sarwar Murphy, Christopher Hou, Daqing Khan, Nazar Ananthanarayanan, Ganesh Hussain, Faraz. (2019). Machine Learning at the Network Edge: A Survey. 10.1145/3469029.

[8] S. Müksch, T. Olausson, J. Wilhelm and P. Andreadis, "Benchmarking the Accuracy of Algorithms for Memory-Constrained Image Classification," 2020 IEEE/ACM Symposium on Edge Computing (SEC), 2020, pp. 400-404, doi: 10.1109/SEC50012.2020.00059.

[9] J. Yu et al., "CONGO2: Scalable Online Anomaly Detection and Localization in Power Electronics Networks," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2022.3143123

[10] P. Zhao, M. Kurihara, J. Tanaka, T. Noda, S. Chikuma and T. Suzuki, "Advanced correlation-based anomaly detection method for predictive maintenance," 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), 2017, pp. 78-83, doi: 10.1109/ICPHM.2017.7998309

[11] W. Silva and M. Capretz, "Assets Predictive Maintenance Using Convolutional Neural Networks," 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2019, pp. 59-66, doi: 10.1109/SNPD.2019.8935752.

[12] D. Kim et al., "Squeezed Convolutional Variational AutoEncoder for unsupervised anomaly detection in edge device industrial Internet of Things," 2018 International Conference on Information and Computer Technologies (ICICT), 2018, pp. 67-71, doi: 10.1109/INFOCT.2018.8356842.

[13] Fathy, Yasmin Jaber, Mona Brintrup, Alexandra. (2020). Learning With Imbalanced Data in Smart Manufacturing: A Comparative Analysis. IEEE Access. PP. 10.1109/ACCESS.2020.3047838.

[14] Partha Pratim Ray, A review on TinyML: State-of-the-art and prospects, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 4, 2022, Pages 1595-1623,ISSN 1319-1578,

Control of Power Electronics Devices for Optimal Short-Term Power Systems Planning – A Survey

Institute of Robust Power Semiconductor Systems
University of Stuttgart
Stuttgart, Germany

***Abstract*— In this paper, approaches control of power electronics devices, through the applications of artificial neural networks and condition monitoring are presented. The principle of model predictive control using artificial neural networks as predictors is described in this paper. In addition, different machine learning algorithms are compared with each other. Two relevant applications that are used in practice are explained. One is the applications for smart homes and the other is for predicting the remaining mileage of electric vehicles. The relevance of monitoring to ensure that the results remain reliable is also presented.**

Keywords—model predictive control, monitoring, neural network, deep learning, Smart-Homes, TSEP, DSEP

I. INTRODUCTION

The power electronic devices plays major role in the real time industries which is used to control and convert the electrical energy. The silicon control rectifiers (SCRS), Thyristors find many uses in electronics, and in particular for power control. These devices have even been called the pillar of high power electronics. Thyristors are able to switch a large amount of power and are accordingly used in a wide variety of different applications. In thyristors (power electronics), the firing angle is a type of control mechanism. It is the phase angle of the voltage at which the SCR is turned on.

Condition monitoring using temperature sensitive electrical parameters (TSEPs) and degradation sensitive electrical parameters (DSEPS) are widely recognized as an enabler for health management of power modules.

Another control method is predictive models, which can be used to predict future events. Predictive models use data from the past and present to predict future behavior. They are accurate calculations of probabilities in different scenarios based on processing large amounts of data. This advanced method uses machine learning and artificial intelligence to improve predictive power. Machine learning is a system that autonomously changes its behavior based on patterns found in data sets. This can be used to predict future trends and behavior based on past data. The challenge here is to provide and prepare large amounts of data. There are different possible approaches, for example by regression. For this reason, algorithms of this type are often developed or adapted to be used in predictive models [1–4].

II. MODEL PREDICTIVE CONTROL (MPC)

A. What is Model Predictive Control?

Model predictive control (MPC) is a very successful optimization-based control approach. It consists of repeating the following procedure: At each sampling time, solve a finite-horizon optimal control problem, applying the first part of the corresponding optimal solution of the input trajectory until the next sampling time [5]. The main advantages of predictive control and the reasons for its widespread success in many application areas are:

- The guaranteed compliance with hard state and input conditions.
- The direct optimization of a quality criterion in controller design.
- Applicability to general nonlinear systems with multiple inputs and outputs.
- Controls complex multivariable systems with complicated dynamics.

Model predictive controls were originally developed for multivariable control tasks in process engineering [6]. However, because of their practical advantages, these controls are increasingly being used successfully in other areas, such as energy systems [7].

B. Functional principle

The operating principle of model predictive control can be explained by Fig. 1. Starting from the process states at time k, a finite number of control values (u(k)...u(k + p)) is calculated in such a way that the controlled predicted output follows a given course "as well as possible". The course of the controlled variable is also generally considered only for a finite number of sampling steps (up to time k + p). Hp is the prediction horizon. A suitable mathematical quality functional must be selected to calculate the optimum manipulated variable curve. In the simplest case, scalar cost functions are used for this purpose, in which a mostly quadratic evaluation of the control deviation and the manipulated variable change is performed over the

prediction or manipulation horizon. During control, a process model is used to calculate the best possible values for u(k)...u(k + p) over the horizon in each sampling step in terms of the quality functional. The first value of this sequence is output to the process as manipulated variable. The remaining manipulated variable values are discarded and the calculation is performed again in the next scanning step. When calculating the optimal manipulated variable sequence u(k)...u(k + p), the free movement of the system resulting from the current process state and past manipulated variables must be taken into account. Since the optimization is based on the prediction of the process behavior, position or state constraints can be considered very easily. For this purpose, the course of these constraints over the prediction horizon must be specified and included in the calculation. If necessary, (measurable) disturbance variables can also be taken into account in the optimization process [4].

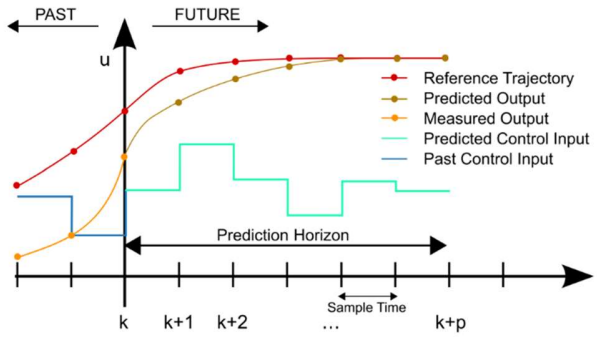


Fig. 1. Basic principle of predictive controls

III. Model Predictive Control- Planning

Before using a model predictive control (MPC), the first step is to determine whether there is any potential for optimization and savings at all. Here, a fundamental distinction must be made between already existing (or possibly expandable) and new energy systems, which will be considered separately below.

A. Existing energy systems

Whether a savings potential can be achieved by model predictive control in already existing energy systems depends very much on the existing general conditions. No general statement can be made about an optimization potential by model predictive control, but systems could be identified in which optimization needs are to be expected. Table 1 lists decision criteria for energy systems where model predictive control is useful or not.

MPC useful if….	MPC not useful if….
Storage tanks are present (thermal, electrical, etc.)	there is no possibility to move loads
generation units are available (solar thermal, photovoltaic, etc.)	plant can only be operated in one meaningful way
variable import and export prices (night tariff, participation in the spot market, etc.)	fixed import and export prices
slow acting or complex systems (biomass furnace, etc.)	only one generation plant is present or these are very flexible
complex sector coupling (CHP-System)	Knowledge about the future can not improve the system behavior at all.

Table 1. Overview of energy systems where MPC is or is not appropriate.

In principle, it can be assumed that an MPC achieves savings in energy systems where at least one point speaks in favor of an MPC, or that the savings potential increases with the number of points speaking in favor of it. In principle, it can be assumed that an MPC achieves savings in energy systems where at least one point speaks in favor of an MPC, or that the savings potential increases with the number of points speaking in favor of it [8].

B. New energy systems

In the case of a new energy system to be planned, the use of an optimization-based planning tool is a good idea. Such tools can take into account the operation of power generation plants by a model predictive control already in the planning stage in order to design an optimally tuned system. See, e.g., [3] for a freely available optimization-based planning tool that uses current market prices for generation plants and typical load profiles for different building types to suggest a technology-neutral investment portfolio for the required energy supply. If the resulting portfolio has a high initial investment but low operating costs, a business model must be used (e.g. contracting) that takes into account the interests of investors and later operators/occupants alike [9].

IV. Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are a subset of machine learning. They are inspired by the human brain and mimic the way biological neurons collect signals from surrounding cells, give a weight to each signal, for importance from the signal, and then decide whether to send a signal themselves.

The special feature of artificial neural networks compared to other algorithms is that artificial neural networks can learn abstract intermediate concepts. They use learning material to learn a general concept that can be used to predict values, for example [10].

A. *Structure and function of ANN*

Neural networks consist of layers of nodes that have an input layer (input), a hidden layer (hidden layer) and an output layer (output). While the input corresponds to the recording of the signals, the middle layer (layer) decides how the network responds to the recording to generate the output. Each node represents an artificial neuron. Each neuron also has edges that connect the individual nodes within the layer. The edges between nodes have weights (w_{ij}) that define how the input is computed across the edge to the next node. This topology of a simple ANN a hidden layer can be seen in Fig. 3 [11].

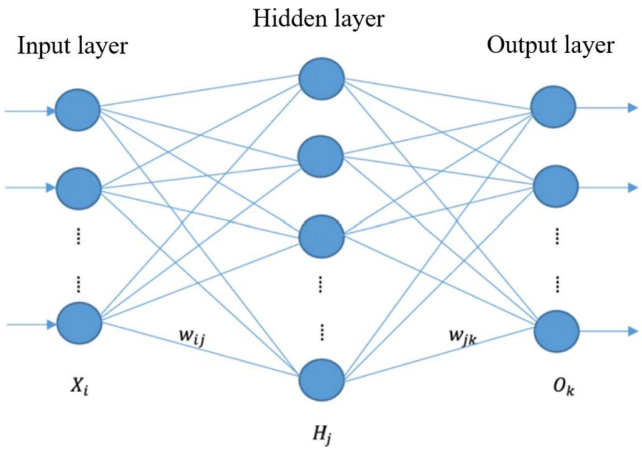


Fig. 3. Structure of a simple ANN with hidden layer

ANN are very well suited for predictive tasks due to their approximation capability. In addition to predicting the future initial value, the networks can also make predictions over a longer period of time by selectively changing the network structure. The fact that manual intervention in the manufacturing process is also possible today ultimately led to the idea of developing a process model that can draw on experience and learned process knowledge. Artificial neural networks (ANN) or fuzzy algorithms can be used as knowledge-based models. The main advantage of ANN over fuzzy logic is that no process knowledge is required in modeling. ANN can assume the behavior of the process from observation alone. In addition, the networks can be adaptively trained so that they continually adapt to changing environmental conditions. The modeling effort, extrapolation capability, and computation times are strongly dependent on the network structure (type of network used, structure, configuration, training data). Generally applicable systematic procedures for the identification of an optimal network topology and configuration do not exist, nevertheless, according to [11] artificial neural networks are now the most commonly used approach for black-box modeling.

B. *Deep Learning*

Deep Learning is not a separate method, but a type of neural networks. The main characteristic of Deep Learning is that an artificial neural network consists of several hidden layers. A neural network that consists of more than one hidden layer can be considered as a Deep Learning algorithm [12].

Fig. 2 shows the distinction between Artificial Intelligence, Machine Learning and Deep Learning.

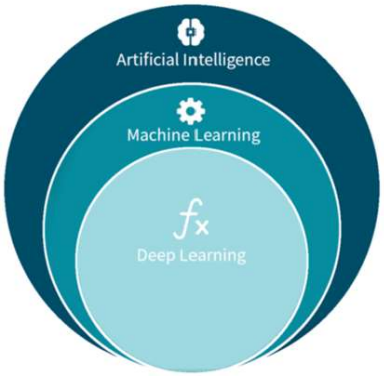


Fig. 2. Differentiation between Artificial Intelligence, Machine Learning, Deep Learning

V. NEURAL PREDICTIVE CONTROL

In addition to classical methods of process description (for example, state space), neural networks enable fast and compact modeling of a multidimensional, dynamic and strongly nonlinear system behavior. Therefore, various applications of this method can be found in the industry. Successful attempts to combine neural networks and predictive controllers are described in [13]. However, neural networks are not only suitable for modeling processes, but also for use as controllers. For example, the implementation of a neural network as a predictive controller is also conceivable [14].

There are no upper limits to the complexity of the control concept to be set up. The MPC approach can be applied for simple single-size control as well as for multivariable control. The simplified basic principle of a simple MPC control is shown in Fig. 4.

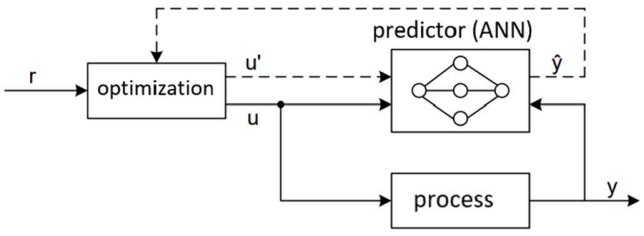


Fig. 4. : Basic principle of neural predictive control ([15])

Based on future manipulated variables u′ to be optimized, the ANN calculates the prediction of the controlled variables yˆ for a defined prediction horizon. The goal of the optimization is to minimize the given quality criterion and to compute the optimal future manipulated variables u′ from it. The result is a sequence of manipulated variables, of which only the first manipulated variable u is output to the process. The considered

time horizons shift continuously and the optimization task is solved anew at each sampling step. This approach is called the principle of the "sliding horizon" in the technical literature and ensures that current information of the controlled and manipulated variables is always included. Without this approach, the calculated sequence of manipulated variables is based on outdated information. A reaction to non-measurable disturbances would only be possible after the calculated prediction horizon (or manipulated variable horizon) has expired.

A. Monitoring

After creating the prediction model, its efficiency should be closely monitored to ensure that the results remain reliable. Condition monitoring refers to the process of monitoring the condition of a system to identify changes that would indicate damage or an incapacitating failure. It allows operators to identify and correct problems (through repair and maintenance procedures) before they lead to equipment failure.

B. TSEP

The use of temperature-sensitive electrical parameters (TSEPs) to determine the junction temperature of power semiconductors is one of the most important techniques for implementing condition monitoring strategies that can be used to assess aging damage, improve module lifetime, and establish operating limitations. The use of TSEPs for condition monitoring has made electrothermal characterization of power devices and modules an essential requirement for effective implementation [2].

C. DSEP

The main objective of a healthy monitoring system is to measure and track degradation sensitive electrical parameters (DSEP) or failure precursors. Several failure precursor parameter-based methods have been published to monitor potential IGBT failures due to degradation or aging, such as the on-state voltage (VCE,on), gate-emitter threshold voltage (VGE,th), gate current (iG), short-circuit current (Isc), turn-off time (toff), and thermal resistance [16].

VI. APPLICATION OF MACHINE -LEARNING ALGORITHMS

A. Method for predicting the remaining mileage of electric vehicles

For electric vehicles, it is important to predict range accurately. Predicting range is a regression problem. In [3], different models and the appropriate machine learning algorithms were compared. The goal was to predict the range of electric vehicles with a very high accuracy. The bagging fusion model, the staking fusion model and the weighted fusion model are considered. The bagging fusion model has the best prediction performance, with a maximum error of 3.5%. The model is based on the model fusion strategy and the dimension expansion, which use the Extreme Gradient Boosting algorithm (XGBoost algorithm). Thus, the remaining mileage of electric vehicles can be predicted directly. Three machine learning algorithms are compared to evaluate the accuracy of prediction. Compared to the Ridge Regression (RR) algorithm or Random Forest (RF) algorithm, the XGBoost model has the highest accuracy. The principles of RR, RF, and XGBoost are described in detail in [17,18].

B. Deep Learning in Smart-Homes

Also the field of smart homes is strongly developed and neural networks are also used and find their applications.
The environment and home systems, such as lighting, appliances, temperature and security systems are automatically controlled via network communication [19].
As smart homes become more personalized, data must be analyzed daily to provide appropriate functions and services.
From this, it can be seen that smart home technology is moving beyond the remote control of the home and is now gradually evolving into a technology that learns user data and provides personalized functions and services with machine learning and deep learning [20].

VII. CONCLUSION

By applying artificial intelligence approaches, the increased amount of data can be processed and fused into a very accurate model. Similarly, the data collected by such a system can be used to train neural networks to extend the life of power electronic systems.
Neural networks are considered to be completely data-based black-box models and have some limitations. Their results are only trustworthy in the domains in which they have been trained. Therefore, for the intended use case, a suitable training dataset should include relevant parts of the system dynamics. It is necessary to test the models before their use in predictive control. A disadvantage of artificial neural networks is the high computational time required for training.
Furthermore, it is important to provide suitable monitoring measures such as monitoring that supervise the model.

REFERENCES

[1] J. Kohler, M. A. Muller, and F. Allgower, “A Nonlinear Model Predictive Control Framework Using Reference Generic Terminal Ingredients,” *IEEE Transactions on Automatic Control*, vol. 65, no. 8, pp. 3576–3583, 2020.

[2] Cutler, C.R. und Ramaker, B.L., “Dynamic Matrix Control - a computer control algorithm. In: Joint American Control Conference, San Francisco (1980,”.

[3] Henrichfreise, H. und Jusseit, J, “Modellprädiktive Zustandsregelung für einen elastischen und reibungsbehafteten Positionierantrieb. In: 4. VDI Mechatronik Tagung 2001, Innovative Produktentwicklung, Frankenthal VDI-Berichte 1631 (2001), September 12-13,”.

[4] C. Steinbrecher, Ein Beitrag zur prädiktiven Regelung verbrennungsmotorischer Prozesse, ARGESIM/ASIM-Verl., Wien, 2011.

[5] M. Grellmann, “Was ist eine prädiktive Analyse, und wie können Sie sie nutzen?,” *Martin Grellmann*, 20.01.2022.

[6] M. Stadler, “et al., Optimizing Distributed Energy Resources and building retrofits with the strategic DER-CAModel, in Applied Energy, Volume

[7] R. Nagyfi, ““The differences between Artificial and Biological Neural Networks,” Towards Data Science, 4 Sep. 2018, URL: https://towardsdatascience.com/thedifferences-between-artificial-and-biological-neural-networks-a8b46db828b7 [retrieved 8 July 2021].,”.

[8] Döbel, I., Leis, M., Molina Vogelsang, ““Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung,” Fraunhofer-Gesellschaft, München, 1 Jan. 2018.,”.

[9] M. A. Nielsen, “, Neural Networks and Deep Learning, Determination Press, 1 Jan. 2015.,”.

[10] A. Bollig, “: Prädiktive Regelung mit Neuronalen Netzen. In: at - Automatisierungstechnik - Oldenbourg Wissenschaftsverlag 51 (2003)

[11] M. Schüler and Hafner, M. und Isermann, R, “: Einsatz schneller neuronaler Netze zur modellbasierten Optimierung von Verbrennungsmotoren - Teil 1: Modellbildung des Motor- und Abgasverhaltens. In: MTZ - Motortechnische Zeitschrift - Vieweg Verlag / GWV Fachverlage GmbH 61 (2000),”.

[12] M.H. Beale, M.T. Hagan und H.B. Demut, “Neural Network Toolbox: User's Guide. Benutzerhandbuch zur Sorftware MATALB R2016a, zuletzt abgerufen unter http://de.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf)

[13] Kessels, J.T.B.A., Rosca, B., Bergveld, H.J., van den Bosch, “P.P.J. On-line battery identification for electric driving range prediction. In: IEEE Vehicle Power and Propulsion Conference, Chicago, IL, USA

[14] B. Pan, “Application of XGBoost algorithm in hourly PM2.5 concentration prediction. IOP Conf. Ser. Earth Environ. Sci. 13(1), 012127 (2018),”.

[15] Ahmad, M.W., Mourshed, M., Rezgui, Y, “Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 147, 77–89 (2017),”.

[16] A. A. Zaidan and B. B. Zaidan, “A Review on Intelligent Process for Smart Home Applications Based on IoT: Coherent Taxonomy, Motivation, Open Challenges, and Recommendations. Artif. Intell. Rev. 2020”.

[17] M. Anti´c, I. Papp, S. Ivanovi´c et al., “M. Learning From Smart Home Data: Methods and Challenges of Data Acquisition and Analysis in Smart Home Solutions. IEEE Consum. Electron. Mag. 2020, 9, 64–71.,”

[18] B. Pan, “Application of XGBoost algorithm in hourly PM2.5 concentration prediction. IOP Conf. Ser. Earth Environ. Sci. 13(1), 012127 (2018),”.

[19] A. A. Zaidan and B. B. Zaidan, “A Review on Intelligent Process for Smart Home Applications Based on IoT: Coherent Taxonomy, Motivation, Open Challenges, and Recommendations. Artif. Intell. Rev. 2020, 53, 141–165.,”.

[20] M. Anti´c, I. Papp, S. Ivanovi´c et al., “M. Learning From Smart Home Data: Methods and Challenges of Data Acquisition and Analysis in Smart Home Solutions. IEEE Consum. Electron. Mag. 2020, 9, 64–71.,”.

Temperature Sensitive Electrical Parameters in SiC-MOSFETs

Institute of Robust Power Semiconductor Systems
University of Stuttgart
Stuttgart, Germany

***Abstract*—This paper presents the secondary literature on different thermosensitive electrical parameters and their potential to measure the junction temperature which is important for both its optimal operation and reliability. The usage of temperature sensitive electrical parameters in temperature measurements are realized and assessed to have a better understanding of the limitations that prevent wide scale implementation. This paper focuses on the Conduction Voltage, Threshold voltage, Internal Gate Resistance, Transient Switching Behaviour and Turn-off delay time. They are compared with respect to sensitivity, linearity, calibration needs, complexity and possibility of on-line measurement.**

***Index Terms*—thermo-sensitive electrical parameter, SiC MOSFET, comparison, junction temperature measurement.**

I. Introduction

In recent years, pushing the limitations of semiconductors' power rating and switching frequency, silicon carbide (SiC) power Metal Oxide Field Effect Transistor (MOSFET) advanced properties has led converters to have a higher power density. Despite these advancements, junction temperature (T_j) variation still needs to be resolved to take full advantage of SiC MOSFET [1]. A critical technique for thermal management, lifetime prediction, and converter reliability enhancement is by monitoring the SiC T_j. The failure and degradation of the SiC device is often attributed to excessive T_j [2] which makes the measurement of T_j important.

The three main types of T_j measuring methods are optical methods, physical contact methods, and temperature sensitive electric parameters (TSEPs) methods. A thermo-reflective technique or optical temperature indicators such as luminescence, Raman effect, refraction index, reflectance or laser deflection gives the possibility of directly obtaining a temperature map of the device. However, optical measurement requires modifying the power module and obtains a non-uniform chip temperature, that is a temperature difference between the center and edges of the chip resulting in a 2-Dimensional map of the heat distribution. The physical contact method is done by incorporating temperature sensors into the package to estimate device temperature [3]. However, the validity of the estimation relies on the accurate device package and tedious analytical analysis [4] because of its poor dynamic response and issues of EMI (Electro Magnetic Interference), voltage isolation and the non-uniform temperature distribution results in the sensor returning either a local or mean average temperature and have a slow response in the measurement [5]. In case of a real-time application, the optical and physical contact methods requires visual or mechanical access to the chip resulting in limitations to implement physically in field applications.

The electrical parameters of the semiconductor which varies with the temperature are called TSEPs. Many electrical properties of the semiconductor devices have a strong relationship with the temperature which can be used to extract the junction temperature. The thermal dependence of electrical properties of SiC is used to determine the temperature. This method is excellent for its relatively high accuracy, fast response and ease to be integrated into the gate driver circuit [5]. The advantage of TSEPs method is that the electrical connections needed for the device to function normally is used for the measurement and the device itself acts as a temperature sensor. The disadvantage is that the temperature maps cannot be made since only the averaged value is extracted.

II. Temperature Sensitive Electrical Parameters

There are many TSEPs which have been utilized in SiC-based devices which extract the information of the T_j indirectly by correlating the T_j with one or multiple electrical parameters. They can be divided into two categories, static TSEPs and dynamic TSEPs. The static TSEPs are extracted during the on-state or off-state while the dynamic TSEPs are extracted during the turn-on or turn-off transition.

A. On-state voltage

The voltage across the device during forward conduction ($V_\mathrm{ds,on}$) have a good sensitivity to temperature [6]. However, the temperature dependence is non-linear and load dependent, therefore its use in online monitoring requires current measurement in order to decouple the load dependency from thermal effects. The main challenge in the practical use of the $V_\mathrm{ds,on}$ for online monitoring resides on the significant voltage overshoot across drain-source terminals during converter switching, ranging from few volts during conduction to well over the DC-link voltage (V_dc) during turn-off transients. These voltage overshoots complicate the design of signal acquisition circuitry with significant challenges in terms of high dynamic

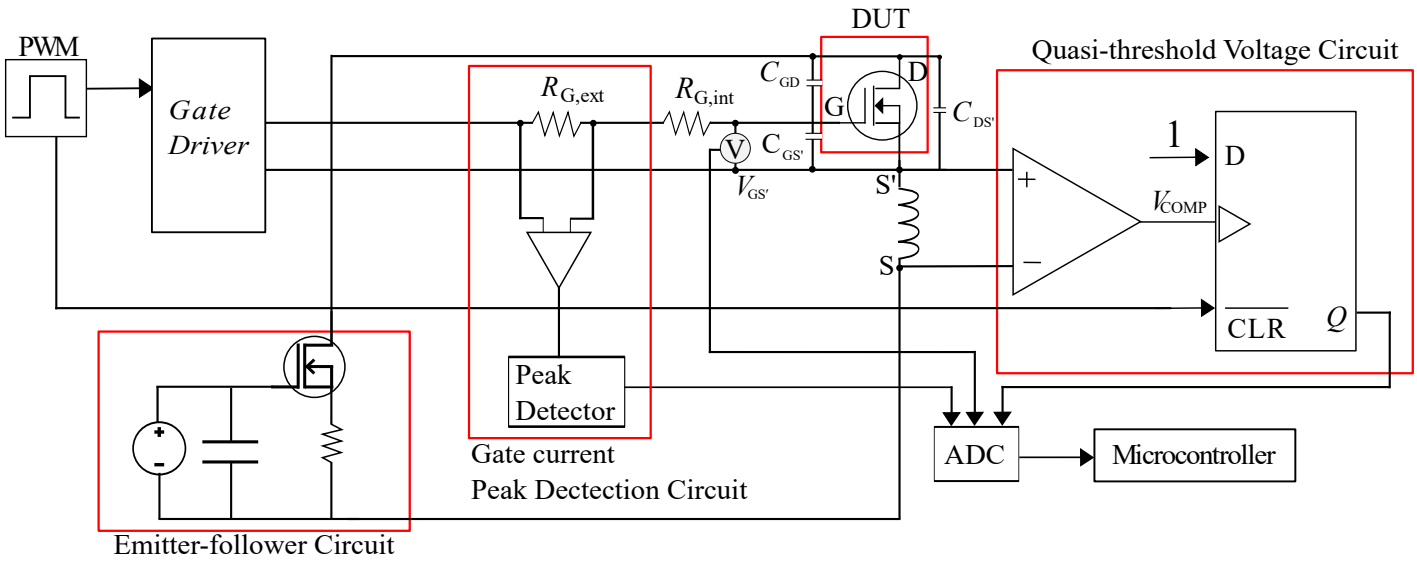


Fig. 1. Schematic circuit diagram for $V_{ds,on}$, gate current and quasi-threshold voltage measurement.

range and isolation between the low-voltage monitoring circuit and the high-voltage power stage. The emitter-follower circuit [6] described in a part of Figure 1 is a practical solution to the measurement of $V_{ds,on}$ of the device under test (DUT). A good linearity and sensitivity with T_j is observed.

B. Threshold voltage

The threshold voltage (V_{th}) is the one of the most critical characteristics of the SiC MOSFET whether static characteristics or switching characteristic [7]. One of the most common V_{th} extraction methods is the measurement static transfer drain current versus gate-source voltage ($I_D - V_{GS}$) of the transistor at constant current. The measurements are carried out using very low drain voltage (around 100 mV) to make sure that the transistor operates in the linear region and to avoid self-heating.

The main challenge in online detection of (V_{th}) during turn-on transients lies in the sampling of the V_{GS} at the correct instance of time since the V_{GS} has a very fast rising times typically up to 1 V/ns in SiC power transistors [6]. A potentially suitable method is the detection of a quasi-threshold voltage [8] described in a part of Figure 1. The V_{th} is observed [8] to have a negative temperature coefficient (NTC). V_{th} is one of the promising temperature sensitive electrical parameters with good sensitivity and linearity.

C. Internal Gate resistance

The total gate resistance (R_G) is the combination of the external resistor ($R_{G,ext}$) which is a part of the driver circuit connected to the power module and internal gate resistance ($R_{G,int}$) which is an intrinsic material property of the gate of the transistor. The time constant τ_{on} of the equivalent gate capacitance charging process during turn-on before V_{th} is defined in (1). During turn-on transition of a SiC MOSFET, for a fixed V_{dc}, the drain–source voltage (V_{ds}) can be assumed to be constant which also results in gate to drain capacitance (C_{GD}) being constant as well. When the gate is connected to the supply voltage, V_{GS} starts to increase until it reaches V_{th}, at which point the I_{DS} starts to flow and the gate to source capacitance (C_{GS}) starts to charge, C_{GS} does not change significantly until V_{th} is reached.

$$\tau_{on} = (R_{G,ext} + R_{G,int})(C_{GS} + C_{GD}(V_{ds})) \tag{1}$$

A simple peak detector circuit [6] described in a part of Figure 1 measures the voltage across the external gate resistor whose peak is directly proportional to the gate current. Similar to V_{th} the $R_{G,int}$ as TSEP is load independent. Heavy doping of polysilicon required for low resistivity gates, results in low temperature coefficient of its resistivity, potentially resulting in low sensitivity of $R_{G,int}$ as a TSEP [6].

Another online T_j measurement is using the Gate-Signal Injection Method [9]. It is a circuit to superimpose a high frequency signal to the gate voltage and a measurement circuit to measure the voltage drop across the $R_{G,ext}$. This method is implemented in a buck converter, consisting of a SiC-MOSFET half bridge and an ohmic-inductive load. There is a biunique characteristic, that is it is having a one-to-one relationship for $T_j < 75°C$ [9]. At higher temperatures, the voltage drop further decreases, which makes it impossible to achieve a reliable temperature measurement resulting in the relationship between the T_j and the voltage drop across the $R_{G,ext}$ not biunique for all temperatures. This injection method has a very low temperature sensitivity at higher temperatures.

D. Switching rate of the drain current

The switching rate of the I_{DS} during turn-on ($\frac{dI_{DS}}{dt}$) could be an effective TSEP for SiC and that its temperature sensitivity is improved when the magnitude of the switching rate is reduced [10]. The temperature sensitivity of the turn-on transient is

determined by the NTC of the V_{th}. The turn-on delay time ($t_{\text{d,on}}$) defined in (2) is the time interval from the start of the V_{GS} rise to the start of the current rise. C_{ISS} is the input capacitance and V_{GG} the output voltage of the gate driver.

$$t_{\text{d,on}} = R_{\text{G}} C_{\text{ISS}} \ln\left(\frac{V_{\text{GG}}}{V_{\text{GG}} - V_{\text{th}}}\right) \tag{2}$$

Switching the devices at lower switching rates with higher R_{G} minimizes the hindering effect of the parasitic inductance on the temperature sensitivity of the turn-on transient. Neglecting parasitic inductance, the switching rate of the I_{DS} of a MOSFET in saturation and its temperature dependency is defined in (3). It can be seen that the technology dependent parameters that determine the switching rate and transconductance (g_{m}) are the V_{th} and the gain factor (β) which depends on chip size and electron mobility (μ_{e}). β reduces with temperature because the μ_{e} reduces with temperature. The temperature coefficient (TC) of β and the V_{th} oppose each other and the Zero-Temperature-Coefficient (ZTC) point is where the effects are counterbalanced resulting in the transconductance being temperature invariant.

$$\frac{dI_{\text{DS}}}{dt} = g_{\text{m}} \frac{dV_{\text{GS}}}{dt} \tag{3}$$

$$\text{where } g_{\text{m}} = \beta(V_{\text{GS}}(t) - V_{\text{th}}) \text{ and } \beta = \frac{W \mu_{\text{e}} C_{\text{OX}}}{L} \ .$$

If the switching rate is to be used as a TSEP for online measurement in SiC power MOSFETs, then its relationship with other parameters like load current and V_{dc}, must be calibrated [11]. The load current and voltage dependency of $\frac{dI_{\text{DS}}}{dt}$ should be decoupled from its temperature dependency so that it can be used for T_{j} sensing. The parasitic source inductance has the effect of slowing the device down and reducing the temperature sensitivity of the $\frac{dI_{\text{DS}}}{dt}$. Hence, either the device is slowed down at the expense of higher switching losses or intelligent gate drivers with the capability of variable gate drive impedance are to be used if $\frac{dI_{\text{DS}}}{dt}$ is to be used as a TSEP.

E. On-state resistance

The on-state resistance ($R_{\text{ds,on}}$) of MOSFET is the resistance between the drain and source terminals when the device is in on-state. It is comprised of different parasitic resistances, which contribute to the $R_{\text{ds,on}}$. In SiC MOSFETs, the channel resistance (R_{CH}) given by (6) has a higher contribution to the total $R_{\text{ds,on}}$ [12].

$$R_{\text{CH}} = \frac{L}{W \mu_{\text{e}} C_{\text{OX}} (V_{\text{GG}} - V_{\text{th}})} \tag{4}$$

Since V_{th} reduces with temperature which means R_{CH} decreases with increasing temperature, so it is a NTC. The drift resistance R_{drift} and Junction Field Effect resistance R_{JFET} increase with temperature due to increased phonon scattering, hence exhibit a PTC. R_{CH} is affected by the V_{GG} used to turn-on the MOSFET. Driving the device with lower V_{GG} will increase the partial contribution of the R_{CH} to the $R_{\text{ds,on}}$, resulting in different temperature coefficients as a function of the V_{GG}. As a result, $R_{\text{ds,on}}$ has different temperature coefficients, it is difficult to determine the global TC of $R_{\text{ds,on}}$ [12]. Therefore using $R_{\text{ds,on}}$ as a TSEP comes with complexity despite having good linearity, sensitivity.

F. Switching times

The $t_{\text{d,on}}$ which is mathematically expressed in (2) has a linear relation to the change in T_{j}. As temperature rises, the $t_{\text{d,on}}$ decreases with temperature due to the decrease of V_{th}. The turn-off delay time ($t_{\text{d,off}}$) is the time interval from V_{GG} falls to 90% to the I_{d} rises to 10% defined in (5).

$$t_{\text{d,off}} = R_{\text{G}} C_{\text{ISS}} \ln\left(\frac{V_{\text{GG}}}{V_{\text{th}} - V_{\text{GG}} + \frac{I_{\text{D}}}{g_{\text{m}}}}\right) \tag{5}$$

For SiC MOSFETs [12], the $t_{\text{d,off}}$ increases with the temperature. This is because the Miller plateau and $R_{\text{G,int}}$ contradict with each other in determining the thermal dependence of the $t_{\text{d,off}}$. The Miller plateau increases with the I_{LOAD}, and the $t_{\text{d,off}}$ is negatively dependent on the Miller plateau. As a result, $t_{\text{d,off}}$ decreases with the I_{LOAD}. The V_{dc} and I_{LOAD} play a significant role in interfering with the temperature sensitivity of the $t_{\text{d,off}}$ for SiC MOSFETs.

III. Comparison of TSEPS based on Key Performance Indicators

The five comparison criteria based on Key Performance Indicators (KPI) of TSEPs will be used in this paper are the sensitivity of the TSEP, its linearity, complexity, the calibration needs and the possibility of doing on-line temperature measurements [3]. An ideal TSEP for use during converter operation should provide a high temperature sensitivity, has a linear temperature dependency and that the sensitivity are comparable between devices of the same type to minimize the calibration effort. It is essential that the TSEP itself can be measured without major modifications to the power converter and without needing to change the power converter is operation. The development of a temperature measurement system using TSEPs is hindered by the dependence of electrical parameters on other variables outside of temperature, parasitic and aging influences throughout a power electronic converter lifetime and the practicality of conducting the measurement of a TSEP without interrupting normal operation and permits to reduce the experimental duration of the calibration [14].

A. Sensitivity

High sensitivity is important as it helps to eliminate the complex and expensive measurement circuit with high resolution. The V_{th} has a good sensitivity. The sensitivity of $\frac{dI_{\text{DS}}}{dt}$ and $t_{\text{d,off}}$ time is improved with slower switching times and intelligent gate driver circuits. The $R_{\text{G,int}}$ and the $R_{\text{ds,on}}$ do not offer a good sensitivity [12] as compared to the earlier mentioned TSEPs.

TABLE I
SUMMARY OF KPI COMPARISON

TSEP [12]		Dependicies [8]	KPI				
			Sensitivity	Linearity	Calibration needs	Online Measurement [13]	Complexity
Static	$R_{\mathrm{G,int}}$	T_j	depends on gate design [13]	depends on gate design [13]	Medium [13]	feasible	High [13]
Static	$R_{\mathrm{ds,on}}$	T_j	Medium [12]	good [12]	Low [12]	feasible	Medium [12]
Static	V_{th}	T_j	good [13]	good [13]	Low [13]	feasible	High [2]
Static	$V_{\mathrm{ds,on}}$	T_j,V_{GS}	good [13]	bad [13]	High [13]	feasible	Low [13]
Dynamic	$\frac{dI_{\mathrm{DS}}}{dt}$	T_j, I_{Load}, R_{G}, V_{GS}	good for slow switching [13]	Medium [13]	Medium [13]	feasible	High
Dynamic	$t_{\mathrm{d,on}}$	T_j, R_{G}, V_{GG}	good for slow switching [13]	good [13]	Low [13]	feasible	Low, with high bandwidth current sensor [2]
Dynamic	$t_{\mathrm{d,off}}$	T_j, R_{G}, V_{GG}	good for slow switching [13]	good [13]	Low [13]	feasible	Low, with high bandwidth current sensor [2]

B. Linearity and Calibration needs

The measure of linearity helps to determine the calibration effort, as non-linear parameters require more calibration points and linearization techniques. The V_{th} and $t_{\mathrm{d,off}}$ provide a linear relationship to T_j while the other TSEPs do not.

C. On-line temperature measurements

State-of-health of the device can be monitored in real-time with the device's T_j, and various lifetime extension strategies can be applied by adjusting the converter operation accordingly by obtaining accurate online T_j measurement. Utilizing the parasitic inductance between kelvin and power sources, the V_{th} is acquired during converter operations thus enabling online measurement. The $t_{\mathrm{d,off}}$ time has also been successfully used for online T_j measurement in SiC MOSFETS. Online measurement can be realized with the peak gate current detection circuit when using $R_{\mathrm{G,int}}$ as a TSEP. The online T_j measurement based on $R_{\mathrm{ds,on}}$ is feasible with the aid of the circuit designs to measure the device's current and voltage simultaneously.

D. Complexity

The $\frac{dI_{\mathrm{DS}}}{dt}$ is the most challenging TSEP in SiC to be detected because of the extremely fast transients resulting in challenging circuit design, but also due to a potentially very low temperature sensitivity and current dependence. The TSEPs $R_{\mathrm{G,int}}$ and V_{th} also have a complex measurement but the measurement data has a better sensitivity and linearity.

IV. CONCLUSION

From the literature study of the $V_{\mathrm{ds,on}}$, V_{th}, $R_{\mathrm{G,int}}$, $\frac{dI_{\mathrm{DS}}}{dt}$ and $t_{\mathrm{d,on}}$ it has been understood that the V_{th} is one of the promising TSEPs. The V_{th} provided a good sensitivity and linearity. A real-time measurement was possible with the quasi-threshold voltage method. This method is better in terms of faster acquisition of the TSEP and higher immunity to the false trigger for the temperature readout [8]. The $R_{\mathrm{ds,on}}$ also appears to be a promising TSEP but it is highly dependent [12] on the SiC MOSFET generation. The $\frac{dI_{\mathrm{DS}}}{dt}$ and switching times as TSEPs in application is highly unsuitable because of the requirement of the TSEPs to be operated at lower frequency which will result in larger size, volume and weight of the measurement setup.

REFERENCES

[1] L. Li, P. Ning, D. Zhang, and X. Wen, "An exploration of thermo-sensitive electrical parameters to estimate the junction temperature of silicon carbide mosfet," in *2017 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific)*, 2017, pp. 1–5.

[2] H. Yu, X. Jiang, J. Chen, Z. J. Shen, and J. Wang, "Comparative study of temperature sensitive electrical parameters for junction temperature monitoring in sic mosfet and si igbt," in *2020 IEEE 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia)*, 2020, pp. 905–909.

[3] Y. Avenas, L. Dupont, and Z. Khatir, "Temperature measurement of power semiconductor devices by thermo-sensitive electrical parameters—a review," *IEEE Transactions on Power Electronics*, vol. 27, no. 6, pp. 3081–3092, 2012.

[4] H.-C. Yang, R. Simanjorang, and K. Y. See, "A method for junction temperature estimation utilizing turn-on saturation current for sic mosfet," in *2018 International Power Electronics Conference (IPEC-Niigata 2018 -ECCE Asia)*, 2018, pp. 2296–2300.

[5] H. Yu, X. Jiang, J. Chen, J. Wang, and Z. J. Shen, "A novel real-time junction temperature monitoring circuit for sic mosfet," in *2020 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2020, pp. 2605–2609.

[6] A. Griffo, J. Wang, K. Colombage, and T. Kamel, "Real-time measurement of temperature sensitive electrical parameters in sic power mosfets," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2663–2671, 2018.

[7] X. Lu, L. Wang, Z. Ma, and Q. Yang, "Static and dynamic characteristics of sic mosfet under extremely high temperature," in *2021 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2021, pp. 5605–5610.

[8] K. Sharma, D. Dayanand, K. M. Barón, J. Ruthardt, F. Münzenmayer, J. Hückelheim, and I. Kallfass, "A robust approach for characterization of junction temperature of sic power devices via quasi-threshold voltage as temperature sensitive electrical parameter," in *2020 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2020, pp. 1532–1536.

[9] J. Ruthardt, D. Hirning, K. Sharma, M. Nitzsche, P. Ziegler, M. Fischer, and J. Roth-Stielow, "Investigations on online junction temperature measurement for sic-mosfets using the gate-signal injection method," in *2021 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2021, pp. 5354–5359.

[10] Z. Ni, S. Zheng, M. S. Chinthavali, and D. Cao, "Investigation of dynamic temperature-sensitive electrical parameters for medium-voltage sic and si devices," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 5, pp. 6408–6423, 2021.

[11] J. O. Gonzalez, O. Alatise, J. Hu, L. Ran, and P. A. Mawby, "An investigation of temperature-sensitive electrical parameters for sic power mosfets," *IEEE Transactions on Power Electronics*, vol. 32, no. 10, pp. 7954–7966, 2017.

[12] J. O. Gonzalez and O. Alatise, "Bias temperature instability and junction temperature measurement using electrical parameters in sic power mosfets," *IEEE Transactions on Industry Applications*, vol. 57, no. 2, pp. 1664–1676, 2021.

[13] F. Yang, E. Ugur, and B. Akin, "Evaluation of aging's effect on temperature-sensitive electrical parameters in sic ¡sc¿mosfets¡/sc¿," *IEEE Transactions on Power Electronics*, vol. 35, no. 6, pp. 6315–6331, 2020.

[14] N. Baker, M. Liserre, L. Dupont, and Y. Avenas, "Junction temperature measurements via thermo-sensitive electrical parameters and their application to condition monitoring and active thermal control of power converters," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, 2013, pp. 942–948.

The future of Gallium-oxide (Ga_2O_3) wide bandgap transistors

Electrical Engineering

University of Stuttgart

Stuttgart, Germany

***Abstract*— Different types of β-based Gallium oxide transistors like MESFETs, MOSFETs, MODFETs, and SBD were demonstrated with the fabrication, advantages and disadvantages of each transistors. The temperature dependencies of β-based Ga_2O_3 Schottky barrier diode are discussed in this paper. With the results such as DC characteristics leads to an indication that Ga_2O_3-based electrical devices can be used for the development of future power device applications. This device can be used in both RF applications as well as in switching technologies.**

Keywords*— *Gallium-oxide (Ga_2O_3), power device, wide bandgap transistor, RF device, Schottky Barrier Diode, thermal conductivity, high breakdown voltage.

I. INTRODUCTION

With excellent properties of ideal semiconductors like wide bandgap, high electron mobility, high critical field strength, high BFOM and also great commercial properties like low cost, ease of fabrication in large volume, Ga_2O_3-based devices might be the future of power switching devices and also for RF applications[3]. The bandgap of Ga_2O_3 semiconductor is between 4.8 to 4.9 eV which is much larger than GaN semiconductor which is only 3.4 eV and is the second largest bandgap after diamonds (see Figure 1)[10]. Diamonds are not used as they are expensive. The advantage of having large bandgap is that it can withstand high electric fields. These properties allow the production of devices with low on-resistance (low conduction losses), high breakdown voltage (high reliability) and fast switching speed (board applications). Although the main disadvantage in Ga_2O_3 when compared to GaN is that it has low thermal conductivity and low carrier mobility as shown in figure 1.

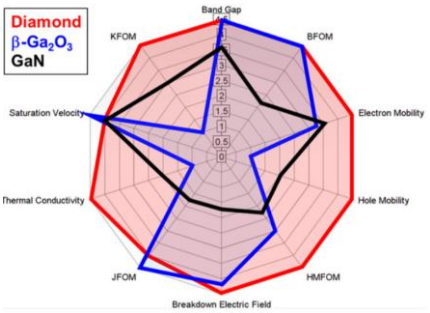


Fig 1 Material properties of power semiconductors[10]

Figure 1 compares different electrical properties of GaN, diamond with Ga_2O_3. It can be found that Ga_2O_3 is a better semiconductor when compared to GaN but not when compared with diamond. Since diamonds are expensive, they are not used in manufacturing process. The major disadvantage which can be found out form figure 1 is that the thermal conductivity is very low when compared to GaN and diamond [10].

II. FABRICATION

The biggest advantage of adopting Ga_2O_3 into the semiconductor technology is that large wafers of crystalline Ga_2O_3 are easy to manufacture. Because Ga_2O_3 is thermally stable, it can be produced using bulk techniques such as the Czochralski method used to produce silicon wafers [11]. In recent days, crystals have been grown by the highly scalable vertical Bridgman-Stockbarger method [11].

A. Gallium oxide metal semiconductor field effect transistor[1]

The fabrication of Ga_2O_3 metal semiconductor field effect transistor, according to M. Higashiwaki, et.al in "Gallium oxide (Ga_2O_3) Metal Semiconductor Field-effect Transistors on Single-crystal β-Ga_2O_3 (010) substrates", was done by doping n-type Ga_2O_3 by tin (Sn) and growing it on β-Ga_2O_3 substrate by molecular beam epitaxy (MBE). Ga and Sn fluxes were introduced by evaporating heated Ga metal and SnO_2 powder in a conventional Knudsen cells. In the oxygen source a gas mixture of ozone and oxygen were used. The substrate temperature was 700°C, with the growth rate of Ga_2O_3 of 0.6 µm/h. The density was 7 x 10^{17} cm^{-3}. Figure 2(a) and 2(b) show the cross-sectional illustration and a micrograph of Ga_2O_3 MESFET under an optical microscope. As device isolation method has not yet been created, circular FET pattern is adopted. Ti (20 nm)/Au (230 nm) was evaporated after a reactive ion etching (RIE) treatment using a gas mixture of BCl_3 and Ar for 1 min, and then lift-off for the development of an ohmic contact. It was found that RIE treatment can significantly reduce contact resistance[1].

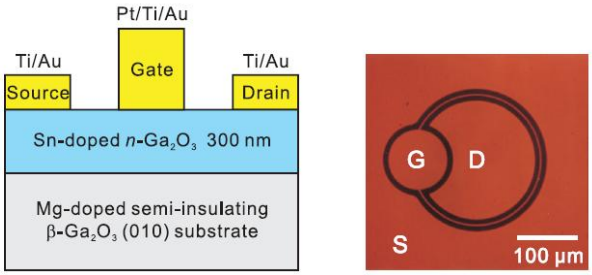


Fig 2 (a) cross-sectional schematic illustration and (b) optical microscope micrograph of Ga_2O_3 MESFET[1]

B. *Gallium oxide metal oxide semiconductor field effect transistor[2]*

The fabrication of Ga_2O_3 MOSFET has the same procedure as that of fabricating Ga_2O_3 MESFET which is proposed by the same author. The device had a high contact resistance at the source and drain electrodes, and the I_d on/off ratio was constrained by the small leakage current through the unpassivated Ga_2O_3 surface. Figure 3(a) and 3(b) shows the cross-sectional illustration and micrograph of Ga_2O_3 under optical microscope. At the source and drain ends, Si^+ ions were implanted which leads to the formation of 150 nm-deep box profile with the silicon density of 5 x 10^{19} cm^{-3}. The reason for using Si+ ions is that is it improves ohmic contacts. It is then annealed for 30 minutes in an atmosphere of N_2 gas at 925°C to reduce internal stresses. By using Ti/Au metal stack which is also annealed in an N_2 atmosphere for 1 minute at 470°C further reduces the contact resistance. An Al_2O_3 passivation layer is formed on the Ga_2O_3.

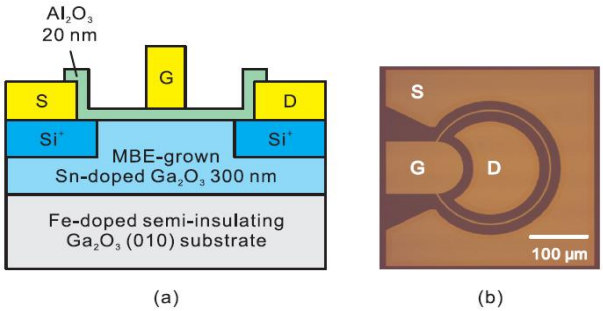


Fig 3(a) cross-sectional illustration (b) optical microscope micrograph of Ga_2O_3 MOSFET[2]

C. *Gallium oxide Schottky barrier diode[4]*

The Schottky barrier diodes (SBD) are manufactured using standard photolithography. Firstly the samples are cleaned to remove organic impurities that are present on the surface with acetone and isopropyl alcohol (IPA). Similar to the other fabrications, metal stacks are used to make the ohmic contacts. Here Ti/Al/Ti/Au metal stacks is annealed in the same way as that of the MOSFETs. This being a Schottky barrier diode the Schottky contacts are made by Pt/Au metal stack.

It is fabricated in such a way that the distance between Schottky and ohmic contact is between 50 to 500 µm[4].

The cross-section and the top view is shown in figure 4.

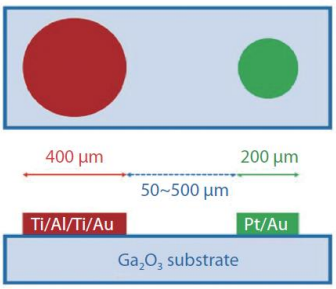


Fig 4 Top and cross-sectional views of fabricated SBDs [4]

III. Gallium Oxide Modulation-Doped Field Effect Transistors

β- Ga_2O_3 has a large bandgap, which makes it attractive for high power, high frequency and optoelectronic applications. It also has high mobility. Although its bulk mobility is reduced at room temperature, it can be improved by phonon scattering[8].

A. *Fabrication of MODFETs*

In the fabrication of MODFETs, Ga_2O_3 was doped with aluminum oxide (AGO) which is a heterostructure. Similar to the fabrication of Schottky barrier diodes, the impurities are removed but here it is heated to 800°C. The substrate is grown by MBE with the flux of Ga being 8 x 10^{-8} Torr which is also assisted by oxygen plasma with the power of 300W. Then Al is grown with the flux of 1.6 x 10^{-8} Torr maintaining the same flux of Ga. In MODFETs, single delta doping was used which lowers the charge density unlike multiple delta doping as it ensures complete modulation in 2DEG. To form the ohmic contacts, Ti/Au/Ni metal stack is which is annealed same as the MOSFETs[8].

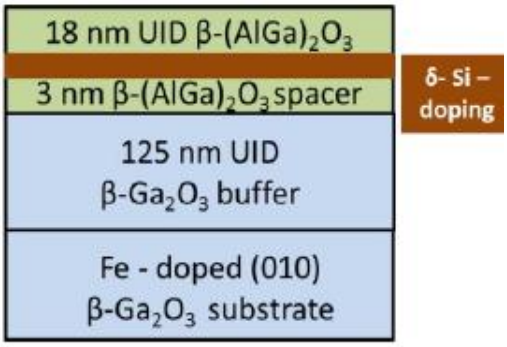


Fig 5 Epitaxial structure [6].

B. *CV characteristics*

The CV characteristics are shown in figure 6. The dielectric constant is chosen to be between 10 and 13 but the dielectric constant along the AGO/ Ga_2O_3 has not been characterized yet. From the graph one can find out that there is charge confinement which means that the mobility is limited. The apparent charge is located at the peak of the CV characteristics which is due to the parasitic capacitance caused due to donor ionization[8].

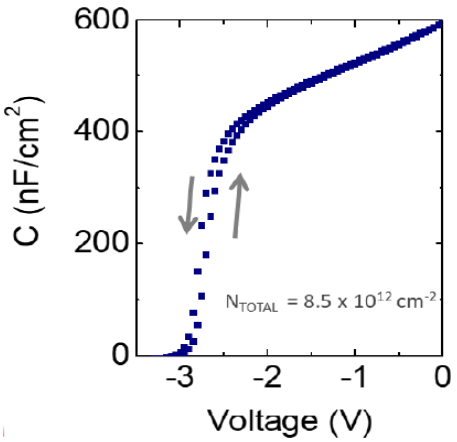


Fig 6 CV measurements for dielectric constant of 10 and 13.

IV. Paper-Thin Gallium Oxide

In the year 2020, the researchers at the University of Buffalo, New York developed a Ga_2O_3 transistor that can withstand upto 8000 V. This can be perfect in the automotive industry which involves the transistors to withstand higher voltages without breakdown. This lateral gallium oxide field-plated transistor yields a breakdown voltage of 1.85 kV. When there is an increase in gate-to-drain distance of the layer (L_{GD}), there will be an increase in breakdown voltage. The significant change from 1.85 kV to 8.032 kV was due to the increase in L_{GD} from 2 µm to 70 µm along with the fact that Ga_2O_3 has a wide bandgap. Semiconductors with a wide bandgap can be thinner, lighter, and stronger to handle more power than semiconductors with a lower bandgap. This thin gallium oxide transistor utilizes the chemical process known as passivation, which involves coating the device to reduce the chemical reactivity of its surface. A layer of SU-8 is added as a passivation layer which is a light-sensitive epoxy-based polymer material. It is hard-baked at 200°C for 10 mins (figure 7)[9].

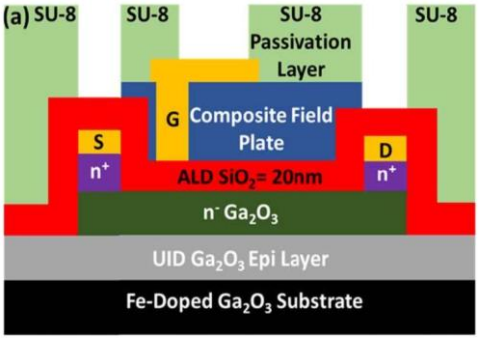


Fig 7 optical microscope image of MOSFET [9].

V. Results and Analysis

Figure 8 shows the DC output characteristics of Ga_2O_3 MESFET. From the figure the maximum drain current I_{DS} can be derived to be 15 mA with V_{GS} of +2V. The breakdown voltage is 257 V when V_{GS} is -30 V. Figure 9 shows the transfer characteristics of Ga_2O3 MESFET. The maximum transconductance is 1.4 mS (red curve). The drain current I_{DS} at off-state is 3 µA[1].

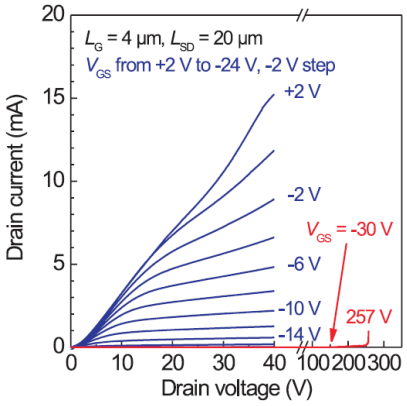


Fig 8 DC output characteristics of Ga_2O_3 MESFET [1].

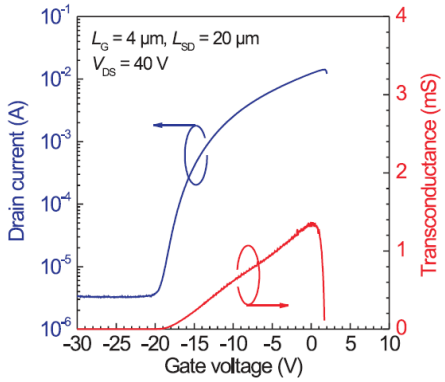


Fig 9 DC transfer characteristics Ga_2O_3 MESFET [1].

Figure 10 shows the output characteristics of Ga_2O_3 MOSFET. The maximum drain current density is 39 mA/mm with V_{GS} = 4V. The breakdown voltage is 370 V at V_{GS} = -20V which is much greater than MESFET. This device has self heating which reduces the thermal conductivity.

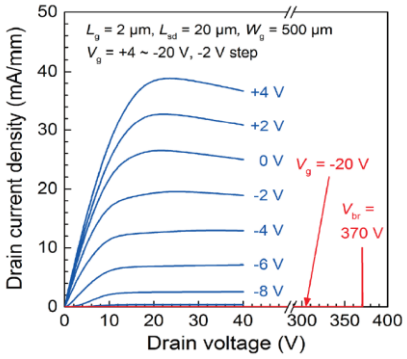


Fig 10 DC output characteristics of Ga_2O_3 MOSFET [2].

Figure 11 shows transfer characteristics of Ga_2O_3 MOSFET with the drain voltage being 25V. It can be found out that I_{on}/I_{off} is greater than 10^{10} which is much greater than MESFET. [2].

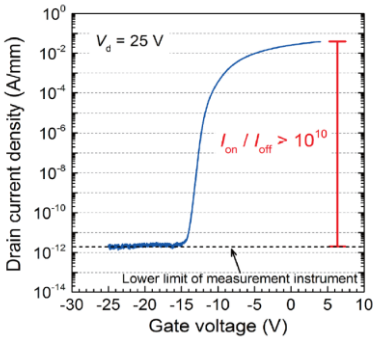


Fig 11 DC transfer characteristics Ga_2O_3 MOSFET [2].

Ga_2O_3 MOSFET is analyzed in temperatures greater than room temperature up to 250°C. The temperature dependent transfer characteristics are depicted in figure 12(a) with the drain voltage being 25V. Figure 12(b) shows the drain current on-off ratio which is the I_d at VG = 0V by I_d at V_G = -25V. It can be observed that there are no kink which means that the device is stable between 25°C to 250°C. [2].

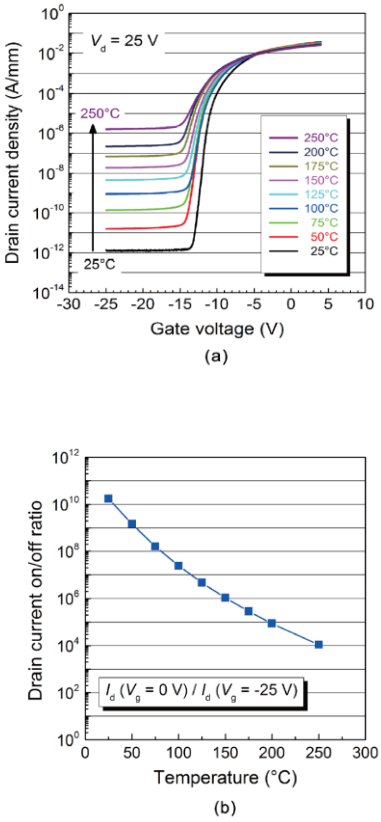


Fig 12(a) Transfer characteristics at V_d = 25V and (b) I_d on/off ratios Ga_2O_3 MOSFET as a function of operating temperature

In Schottky barrier diodes, the height of the diode and the ideality factor is dependent on temperature[4]. There are some non- idealities like the A* which is the Richardson constant, the calculated Richardson constant is A* = 3.28 x 10^{-3} $Acm^{-2}K^{-2}$, which appears much smaller than the theoretical value. There is emission of electrons due to thermal energy in the barrier which causes these non- idealities to occur. It is found that the Schottky barrier height had positive temperature dependence and the ideality factor had negative temperature dependence. From figure 13 it can be found out that the current density is inversely proportional to temperature which indicates that the diode is dominated by bulk leakage current [4].

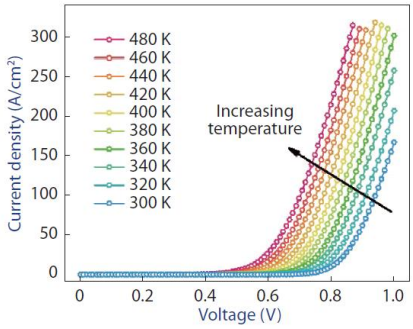


Fig 13 Current density as a function of voltage for different temperatures [4].

VI. Discussion

Almost all the types of Ga_2O_3 technology-based semiconductors have been discussed and each has its advantages and disadvantages. The disadvantages of Ga_2O_3 MESFETs have been overcome by Ga_2O_3 MOSFETs' advantages. Ga_2O_3 technologies can be used in power-switching applications as it has high critical field strength, larger area, substrate availability, low cost and high performance and also can be used in RF applications as Ga_2O_3 has low parasitic resistances and capacitances. The major disadvantages of Ga_2O_3-based semiconductors are that it has very low thermal conductivity around 1/10th of SiC due to self heating.

VII. Conclusion

Even though the research on the Ga_2O_3 semiconductor based is still ongoing, a major part of the findings conclude that Ga_2O_3 has almost all the ideal qualities for a good semiconductor. It is not still used in the industries are due to two reasons. Firstly the MBE process takes a lot of time to manufacture that is why it is used for laboratory purposes. Secondly, it has very low thermal conductivity is 1/10th when compared with Si, i.e, the heat gets trapped inside. The ability of this transistor to withstand over 8000 V, is a boon in the automotive industry. Electric vehicles have high energy density batteries which are usually very heavy. The extreme weight is in correlation with the cars being designed to be able to travel a longer distance before needing to charge. If the car companies were to incorporate gallium oxide transistors for control and power management, the system could potentially double the travelling distance with a lower weight. These are the future of gallium oxide wide bandgap semiconductors.

REFERENCES

[1] M. Higashiwaki, K. Sasaki, A. Kuramata, T. Masuiand S. Yamakoshi, "Gallium oxide(Ga_2O_3) Metal Semiconductor Field-effectTransistors on Single-crystal β-Ga_2O_3 (010) substrates," Applied Physics Letters, vol. 100, issue1, 2012

[2] M. Higashiwaki, K. Sasaki, T. Kamimura, M.H. Wong, D. Krishnamurthy, A. Kuramata, T. Masuiand S. Yamakoshi, "Depletion-Mode Ga_2O_3 Meat-Oxide-Semiconductor Field-Effect Transistor on β-Ga_2O_3 (010) substrates and temperature Dependence of their Device Characteristics," Applied Physics Letters, vol. 103, issue12, 2013

[3] G. Jessen, K. Chabak, A. Green, N. Moser, J. McCandles, K. Leedy, A. Crespo and S. Tetlak, "Gallium oxide technologies and application," in Compound Semiconductor Integrated Circuit Symposium IEEE, CSISC 2017.

[4] M. Higashiwaki, K. Konishi, K. Sasaki, K. Goto, K. Nomura, Q. T. Thieu, R. Togashi, H. Murakami, Y. Kumagai, B. Monemor, A. Koukitu, A. Kuaramataand S. Yamakoshi, "Temperature-dependent Capacitance-Voltage and Current-Voltage Characteristics of Pt/Ga_2O_3 (001) Schottky Barrier Diodes Fabricated on n-Ga_2O_3 drift layers grown by halide vapor phase epitaxy" in Applied Physics Letters, vol. 108, issue13, March 2016.

[5] M. H. Wong, K. Sasaki, A. Kuramata, S. Yamakoshi, and M. Higashiwaki, "Field-Plated Ga_2O_3 MOSFETs with a Breakdown Voltage of over 750 V," in IEEE Electron Device Letters, Vol. 37, No. 2, February 2016.

[6] Ramachandra Kocheta, Wyatt Metzger, Barry Mather, Sreekant Narumanchi, Andriy Zakutayev, "Modelling and Analysis of Gallium Oxide Vertical Transistor" in ECS Journal of Solid State Science and Technology.

[7] Kelson D Chabak , Kevin D Leedy, Andrew J Green, Shin Mou,Adam T Neal, Thaddeus Asel, Eric R Heller, Nolan S Hendricks,Kyle Liddy , Antonio Crespo, Nicholas C Miller, Miles T Lindquist, Neil A Moser, Robert C Fitch Jr, Dennis E Walker Jr, Donald L Dorsey and Gregg H Jessen, "Lateral β-Ga2O3 field effect transistors" in Semiconductor Science Technology 35.

[8] Sriram Krishnamoorthy , Zhanbo Xia, Chandan Joishi , Yuewei Zhang , Joe McGlone , Jared Johnson , Mark Brenner , Aaron R. Arehart , Jinwoo Hwang , Saurabh Lodha and Siddharth Rajan ," Modulation-doped β-(Al0.2Ga0.8)2O3/ Ga2O3 Field-Effect Transistor" .

[9] Paper thin Gallium Oxide Transistor Developed by Dr. Uttam Singisetti in University of Buffalo.

[10] Md Nazmul Hasan, Edward Swinnich, Jung-Hun Seo, "Recent Progress in Gallium Oxide and Diamond Based High Power and High-Frequency Electronics" International Journal of High Speed Electronics and Systems, March 2019.

[11] Article from IEEE Spectrum online "https://spectrum.ieee.org/gallium-oxide-the-supercharged-semiconductor" by Gregg H Jessen, 24 March 2021.

Distributed Mixer Design for Bandwidth enhancement

Elektro- und Informationstechnik
University of Stuttgart
Stuttgart, Germany

***Abstract*—This paper describes the concepts, use and state of the art of mixers using the distributed design principle. It aims to give a historical overview over distributed design techniques as well as recent developments and achievements.**

***Index Terms*—microwave design, distributed mixer, distributed amplifier, artificial transmission line, travelling wave**

I. Introduction

Mixers are one of the most important components in a wide number of RF and microwave circuits. They are of particular importance in today's communication systems that need to satisfy the ever growing demand of data throughput (often between mobile, battery-powered devices). Modern communication standards already specify the use of frequency bands over very high relative bandwidths, e.g. the 5G Frequency Range 2 that includes channels between 24.25 and 71 GHz, and future standards are expected to follow this trend [17]. Other applications are microwave sensors or instrumentation purposes. Crucial requirements for the mixers as key components in these systems are their operation bandwidth, conversion loss/gain and power consumption.

A promising concept, combining very wide bandwidth and low conversion loss are distributed mixers. While they have a certain DC power consumption, they can be beneficial on system level because the need for additional amplification might be reduced. The principle of a distributed (or travelling wave) circuit was first described in 1936 by Percival [1] who used the input and output capacitance of amplifying cells to construct artificial transmission lines. Since then it found widespread use in amplifier design, with early designs using vacuum tubes as amplifying elements [2]. Newer devices often use GaAs, InP or other high electron mobility FETs while in recent years also other technologies like GaN (for power amplifier applications) or CMOS (due to much lower manufacturing costs) are evaluated. Distributed amplifiers reach exceptional bandwidth performance at high output powers by cascading multiple gain cells while absorbing their parasitic capacitances into the connecting transmission lines instead of using frequency-selective matching networks [3].

Very similar techniques can be used to construct mixers with high performance regarding their operation bandwidth and conversion gain. This concept has first been used in 1984 by Tang and Aitchinson [4].

II. Theory of operation

When considering active or passive mixers using three-port gain cells as nonlinear devices, it is apparent that in many cases they share the same structure with amplifiers, using three port amplifying devices such as transistors. Differences lie in the LO insertion concepts: A mixer can be implemented in an additive way as an transconductance mixer. In this case, the nonlinearity of the transistor transconductance g_m is used to generate mixing products. Other designs are multiplicative, meaning two signals are applied to separate ports of the gain cell.

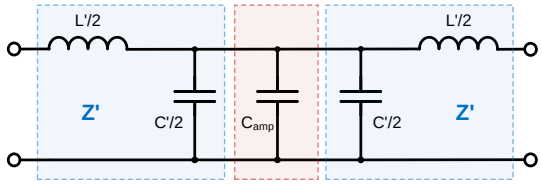


Fig. 1. Embedding of gain cells into inductive transmission line segments to form an artificial transmission line with the impedance Z.

To achieve a high amplification or conversion gain as well in amplifiers as in mixers, it is common to use multiple gain cells. The concept of a distributed circuit overcomes some of the difficulties that other ways of combining multiple gain cells have. When gain cells (regardless of the transistor technology or even vacuum tubes) are cascaded in multiple stages, their in- and outputs need to be impedance matched to the adjacent stage or termination. While there exist different design paradigms for impedance matching such as matching for maximum gain or lowest noise, they all have in common that the matching networks are frequency-selective.

In comparison, travelling wave circuits are serializing multiple amplifier cells by embedding the capacitive input- and output impedances into transmission lines. While these capacitances also change over frequency, the effect on broadband matching is much lower.

Figure 1 shows the concept of an artificial transmission line

consisting of the discrete amplifier capacitance C_{amp} and the distributed line inductance L' and capacitance C' over the length l. A lossless case is assumed. The impedance of the structure is

$$Z = \sqrt{\frac{L'}{C' + \frac{C_{\text{amp}}}{l}}} = \frac{Z'}{\sqrt{1 + \frac{C_{\text{amp}}}{C' \cdot l}}} \tag{1}$$

with the transmission line segments forming the impedance of $Z' = \sqrt{\frac{L'}{C'}}$. Typically Z is designed to be 50 Ω in order to match in- and output circuits. The phase constant in the transmission line segments is

$$\beta' = \omega\sqrt{L'C'} \tag{2}$$

so for in-phase combination of the signals in an amplifier the length of the drain and source transmission line segments between the stages has to be adjusted to fulfill the condition

$$\beta'_{\text{a}} \cdot l_{\text{a}} = \beta'_{\text{b}} \cdot l_{\text{b}}. \tag{3}$$

When considering a mixer, the LO- RF- and IF-signals typically have different frequencies that are not combined with coherent phase. Because the propagation velocity v_p still needs to be the same for all applied signals, equation 3 can be more generally expressed as

$$l_a \cdot v_{p,a} = l_b \cdot v_{p,b} \tag{4}$$

for the proportions in the distributed mixer. Because real transmission lines such as microstrip exhibit dispersion, some compromises have to be used.

The black part in Figure 2 shows the typical topology of a distributed amplifier. The bottom transmission line is the input or gate line. It is terminated with 50 Ω on the output side while the output or drain line on top is terminated on the left side. The coloured parts indicate additions to be made to the circuit to implement a distributed mixer.

For the nonideal case that there is some loss present in the transmission lines and amplifiers, there is a number N of stages where the gain is maximized. This effect is caused by the exponential decay of the gate line voltage as well as the amplified signals generated by the first stages being attenuated towards the output [5] [7].

III. Concepts and Performance of Different Distributed Mixer Topologies

A. LO insertion concepts

As there are different active mixer concepts, there are multiple ways to construct a distributed mixer. Figure 2 shows different ways to implement the insertion of the LO signal into the amplifier structure. The red highlighted part shows option A for the LO port. By inserting the signal into the "unused" end of the drain line, a resistive mixer operating at the transistor threshold voltage (like in [11]) or a drain mixer comparable to [12] can be constructed. The design in [15]

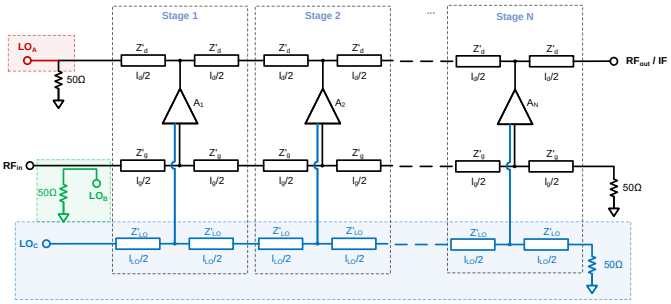


Fig. 2. Distributed Amplifier/Mixer with different concepts of LO insertion. Bias networks are not shown.

is a 3-stage drain mixer, implemented in a 130-nm CMOS technology, demonstrating the use of a widely accessible and cost-effective process. Drain mixers are by definition passive devices despite using amplifier cells, because the power is provided over the LO input via drain-pumping. The circuit in [10] also uses the drain mixing concept, but consists of two independent differentially driven 3-stage drain mixers implemented in a 250 nm GaAs pHEMT technology. This has the advantages of using less LO power for a comparable conversion gain performance as [15], and a much higher isolation.

Option B (green) requires a directional coupler or an other form of power combiner to inject the LO signal at the RF input. This way, a transconductance mixer such as [4] can be implemented. A recent design with this approach is also presented in [16]. Here, the gain cells consist of darlington circuits in a 180-nm SiGe BiCMOS technology. The LO and RF signal are combined using a wide band Wilkinson coupler, followed by the 3-stage distributed mixer. With a moderate DC power consumption and comparably low LO input power, a conversion gain of 3 dB is obtained over a very high bandwidth from 2 GHz to 67 GHz.

If the modulation is done by applying the LO signal at each gain cell, an additional transmission line is needed. This way, topologies such as a source feedback mixer [6] or mixers with dual gate gain cells are built. A high-performance example of this approach is presented in [14]. The most impressive bandwidth performance among the compared designs in table I is exhibited by this design, using 6 half Gilbert cells in a 250-nm InP technology. At the cost of relatively high DC power consumption, it offers relatively low conversion loss and good isolation.

Figure 3 shows the implementation of the 8-stage distributed source-feedback mixer from [6]. The differences between the microstrip transmission lines can be clearly seen. The drain line is the widest, meaning it is the least inductive line as the output capacitances of the cascodes are smaller than the RF and LO input capacitances. Subsequently, the higher phase velocity has to be compensated by meandering routing.

For reference, table I also shows two conventional single-

TABLE I
STATE OF THE ART DISTRIBUTED MIXER DESIGNS

Ref.	***balanced***	***Stages***	***RF BW*** / GHz	P_{LO} / dBm	G_C / dB	**Technology**	**LO-RF Isolation** / dB	**DC power** / mW
[16]	no	3	2 - 67	0	3	180nm SiGe	-10	17.5
[15]	no	3	0.8 - 77.5	10	-5.5	130nm CMOS	-13	N/A
[14]	single	6	0 - 194	2	-3	250nm InP HBT	-	125
[10]	single	2x3	5.4 - 21.8	2	-4	250nm GaAs pHEMT	-33.5	N/A
[18]	double	1	25.3 - 32.5	-3	8.6	65nm CMOS	-48	7
[19]	double	1	3 - 10	16	-12	250nm GaAs pHEMT	-30	N/A

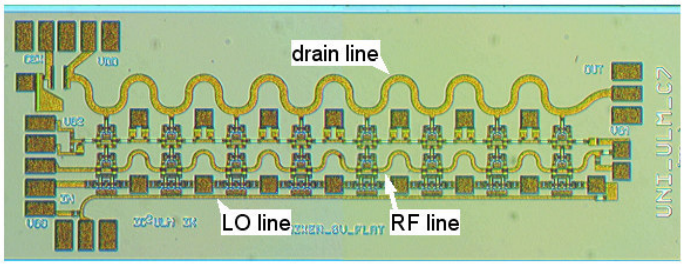


Fig. 3. Chip photograph depicting an 8-stage distributed mixer. The gain cells are GaAn-cascodes with a separate LO transmission line connected to source-feedback transistors [6].

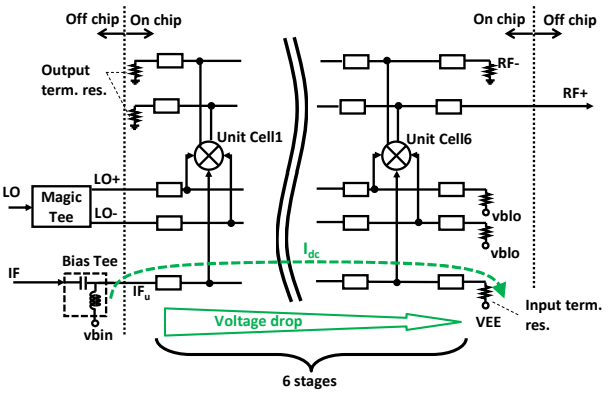


Fig. 4. Single balanced distributed mixer design using 6 stages of half Gilbert cells [14].

stage designs. Reference [18] is an active double balanced mixer in 65-nm CMOS. It offers high conversion gain and good isolation, but is very limited in bandwidth. A passive mixer such as the diode ring mixer in [19] can also reach a wide bandwidth, but has the the major disadvantage of a much higher conversion loss.

B. Balanced mixers

Because of the high circuit complexity, the vast majority of distributed mixers are implemented as single ended circuits. Only a few designs of balanced mixers are proposed, of which most use a series of half Gilbert cells such as the design from [14] shown in figure 4 or implemented in [8], [9]. Obviously this requires five parallel transmission lines in the mixer section (LO+, LO-, IF+, IF-, RF) as well as possibly baluns. Circuit complexity might also in some cases be limited by the technology used. Especially for 3-5 semiconductors like GaN or InP the number of available metal layers is often limited. Another approach that is unique to the author's knowledge is the combination of two single- ended distributed drain mixers for each differential part of the RF signal whose outputs are subsequently combined in an off-chip balun [10].

C. Advantages and disadvantages of the distributed mixer

When distributed mixers are implemented as active mixers, they generally have the disadvantage of introducing more noise than passive components. Active circuits also increase system complexity as they require one or more voltage supplies and consume DC power. However, active and passive distributed mixers exhibit less conversion loss or in some cases even gain and can thus in some cases make a subsequent IF amplifier obsolete. By meeting the problem of matching the active components by using the distributed design principle, very high bandwidths are achieved, but the interconnections and multiple stages require more chip area than single-stage designs.

IV. CONCLUSIONS

Distributed mixers are valuable components when high bandwidth and low conversion loss are prioritized requirements. By using the well known distributed amplification concepts commonly used for power amplifier applications, they overcome the challenge of impedance matching and combining multiple gain stages. Distributed mixers are a topic of current research and seem to have promising future uses.

ACKNOWLEDGMENT

The author would like to acknowledge the support of Prof. Ingmar Kallfass for his valuable lecture on scientific writing and Lukas Gebert for the tutoring and help on the topic of distributed mixer design.

REFERENCES

[1] W. S. Percival, "Thermionic Valve Circuit," US2226255A, USA, Nov 2, 1937.

[2] A. P. Copson, "A distributed power amplifier," Electr. Eng., vol. 69, no. 10, pp. 893–898, 1950, doi: 10.1109/EE.1950.6437036.

[3] G. Nikandish, R. B. Staszewski, and A. Zhu, "The (R)evolution of Distributed Amplifiers: From Vacuum Tubes to Modern CMOS and GaN ICs," IEEE Microwave, vol. 19, no. 4, pp. 66–83, 2018, doi: 10.1109/MMM.2018.2813838.

[4] O. S. A. Tang and C. S. Aitchison, "A Microwave Distributed MESFET Mixer," in 1984 14th European Microwave Conference, 1984, pp. 483–487.

[5] D. M. Pozar, Microwave engineering. Hoboken, NJ: John Wiley & Sons Inc, 2012.
[6] I. Kallfass, T. Purtova, A. Brokmeier, W. Ludwig, and H. Schumacher, "One Single Travelling-Wave MMIC for Highly Linear Broadband Mixers and Variable Gain Amplifiers," in Microwave Symposium Digest, 2005 IEEE MTT-S International, Long Beach, CA, USA, 2005, pp. 97–100.
[7] Ingmar Kallfass, Microwave Engineering. Lecture Script. Stuttgart, 2021.
[8] F. Chang, P. Wu, M. Lei, and H. Wang, "A 4–41 GHz Singly Balanced Distributed Mixer Using GaAs pHEMT Technology," IEEE Microwave and Wireless Components Letters, vol. 17, no. 2, pp. 136–138, 2007, doi: 10.1109/LMWC.2006.890341.
[9] Xiaohua Fan and E. Sanchez-Sinencio, "3-22GHz CMOS distributed single-balanced mixer," in IEEE International SOC Conference, 2004. Proceedings, 2004, pp. 93–96.
[10] H. Lee and S. Jeon, "A GaAs p-HEMT Distributed Drain Mixer With Low LO Drive Power, High Isolation, and Zero Power Consumption," IEEE Access, vol. 9, pp. 158420–158425, 2021, doi: 10.1109/ACCESS.2021.3130423.
[11] K. S. Ang, S. Nam, and I. D. Robertson, "A 2 to 18GHz Monolithic Resistive Distributed Mixer," in 1999 29th European Microwave Conference, 1999, pp. 222–225.
[12] Kuo-Liang Deng and H. Wang, "A 3-33 GHz PHEMT MMIC distributed drain mixer," in 2004 IEE Radio Frequency Integrated Circuits (RFIC) Systems. Digest of Papers, 2004, pp. 151–154.
[13] Ingmar Kallfass, Microwave Analog Frontend Design 1. Lecture Script. Stuttgart, 2020.
[14] T. Jyo et al., "A DC to 194-GHz Distributed Mixer in 250-nm InP DHBT Technology," in 2020 IEEE/MTT-S International Microwave Symposium (IMS), 2020, pp. 771–774.
[15] H. Yang et al., "Design and Analysis of a 0.8–77.5-GHz Ultra-Broadband Distributed Drain Mixer Using 0.13-μm CMOS Technology," IEEE Transactions on Microwave Theory and Techniques, vol. 57, no. 3, pp. 562–572, 2009, doi: 10.1109/TMTT.2009.2013299.
[16] Y. -C. Liu, Yi-Wei Chang, Ya-Che Yeh, S. -H. Weng, J. -H. Tsai, and H. -Y. Chang, "A 2-to-67 GHz 0-dBm LO power broadband distributed NMOS-HBT Darlington mixer in 0.18 μm SiGe process," in 2016 IEEE MTT-S International Microwave Symposium (IMS), 2016, pp. 1–4.
[17] 5G NR User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone: (3GPP TS 38.101-2 version 15.3.0 Release 15), 3GPP.
[18] E. Zheng and D. Zhao, "A Ka-band Up-Conversion Mixer in 65-nm CMOS," in 2021 IEEE International Conference on Integrated Circuits, Technologies and Applications (ICTA), 2021, pp. 147–148.
[19] Y. Pu, Z. Huang, S. Pan, and G. Wang, "A 3 GHz to 10 GHz GaAs double balanced mixer," in Proceedings of 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC 2017): October 3-5, 2017, Chongqing, China, Chongqing, 2017, pp. 1083–1086.

Point-to-Point Outdoor Transmission for back-haul communication links in E-Band

Dept. of Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—Abstract – In this paper, a detailed comparison of achievable data rates for an E-band communication link is reported. Multiple transceivers are either fabricated and or simulated using advanced semiconductor technology suitable for E band communication links. Bit Error Ratio (BER) and Data rates for various modulation schemes are tabulated and presented in this work.**

***Index Terms*—E-Band, High data rates, atmospheric attenuation, satellite communication, BER, millimeter waves.**

I. Introduction

In our global and interconnected world, the demand for higher data rates is increasing drastically. Applications like social networks, live broadcasting telecommunication services, and services of cloud computing are driving engineers to explore higher carrier frequency bands to meet customer requirements. Due to hardware technology advancements communication engineers and scientists are carrying out various experiments on E-band (71-76 GHz and 81-86 GHz). E-band communication link with base-band frequencies from 500 to 2000 MHz [7] can achieve a data rate of several Gigabits per second. In this paper, a detailed study on the influence of hardware characteristics, new circuit components in transceiver architecture [3], [7], and achievable data rates are compared and presented. The paper is organized as: In section 2 detailed explanation of link budget analysis [3], [5], [7] is presented. Based on the conclusions of the link budget analysis. In section 3 influence of atmospheric effects on quality of service (QoS) [7], [8], [10] is explained. In section 4 influence of usage of advanced high-frequency Power amplifiers, and other RF circuits on data rates is presented. This paper is concluded with a comparison table that contains different modulation schemes, hardware diversity, and achievable data rates.

II. Link Budget analysis for E-Band

To characterize any radio transceiver's robustness and performance Link budget analysis is the first step [3], [5], [7].

$$P_{Rx} = P_{Tx} + G_t a + G_r a - FSPL - FM - L_{sys} - L_{rain} \quad (1)$$

where P_{Rx} is the Power at the receiver, G_{ta} and G_{ra} are the gain at the antenna at transmitter and receiver (RX) respectively, and Free space path loss (FSPL) and fading in the channel (FM), L_{sys} losses associated with system properties, and L_{rain} is attenuation of signal due to rain [5].

At higher carrier frequencies signal quality degrades due to adverse weather conditions. In section 3 work related to atmospheric attenuation, and fading mitigation method is introduced in [5], [7]. From eq [7] it suggests that it is possible to achieve good transmission efficiency by using a high gain antenna and or advanced power amplifiers at the transmitter end.

1) Calculation of SNR and Minimum P_{Rx} required for AWGN Channel: When it comes to Qos, most effectively it is quantified by bit error rate (BER) [7]. BER dictates or sets a lower limit on minimum energy to noise power spectral density ratio required for the efficient and reliable communication link. The relation of Signal to noise ratio (SNR) and E_b/N are closely related by the equation (2)

$$SNR = 10 * log(\frac{E_b}{N_o} + \frac{R_b}{B}) \quad (2)$$

The term R_b/B is spectral density expressed as b/s/Hz, Figure 3. illustrates BER as a function of SNR for an AWGN channel. Figure 3, provides the required BER (10^{-3}), and the minimum required SNR is 10dB for a given modulation format.

From a system perspective, the overall sensitivity is estimated by P_Rx, min which is given by [3], [5], [7]

$$P_{Rx,min} = 10log(kTB) + NF_{Rx} + SNR_{min} \quad (3)$$

where k is the Boltzmann constant, T is the antenna temperature and B is the signal bandwidth, the first term of the equation indicates thermal noise floor at low noise amplifier (LNA) input. Figure 3 and Figure 4 provide the required SNR and $P_{Rx,min}$ for a modulation scheme interest.

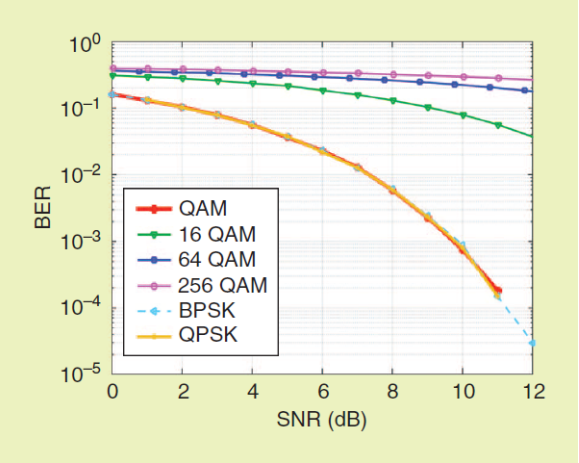


Fig. 1. Fig. 1. [5] The BER v/s SNR for different modulation schemes AWGN channel [7].

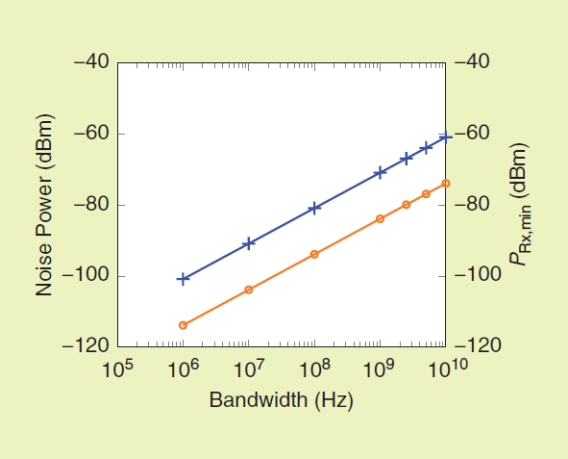


Fig. 2. The noise power v/s Bandwidth and Receiver sensitivity v/s Bandwidth [7].

III. Quality of Service as a Measure of Atmospheric Attenuation

1) Rain Attenuation: Radio propagation of millimeter waves (30–300 GHz) suffer by very high rain attenuation [5], [7]. Rain attenuation models for mm-waves have been studied experimentally and evaluated by experts across the world. International Telecommunication Union $ITU\text{-}R$ specifies the attenuation factor required to consider for various link budget calculations [8], [10]. Equation (2) relates attenuation L_{rain} per km distance and rain rate R (in mm/h unit).

$$L_{rain} = k(R)^{\alpha} \tag{4}$$

The Rain attenuation L_{rain} (dB/Km) is obtained from the rain rate R (mm/h), using values from [8], [10] values for k and α and plotted in figure 3. Rain attenuation impact is recommended by ITU-R based of recorded data are provided in figure 3. It is found that signal degradation associated with rain attenuation effects the link performance significantly while using higher carrier frequencies. For example, during severe monsoons ($150\,\mathrm{mm\,h^{-1}}$), the attenuation can be as high as $80\,\mathrm{dB\,km}$ [5].

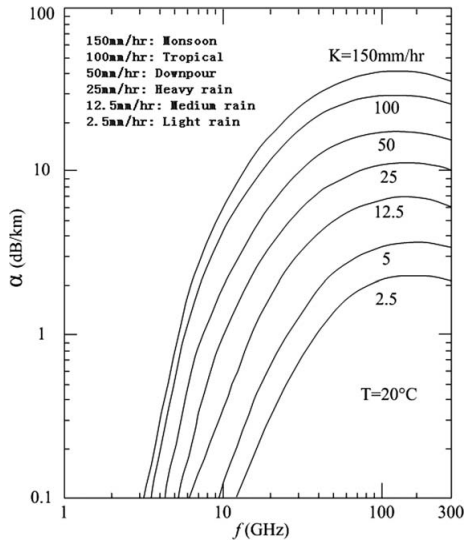


Fig. 3. Rain attenuation for different rainfall rate in mm-wave frequency bands [5]

A. Rain Attenuation model

primarily based on rain facts accumulated for numerous years all the world over, rainfall characteristics are well understood and observe ITU-R recommendations. as an example, in figure four the yearly cumulative distribution of rain attenuation in ITU-R (percentages of time extra than 0.001and rain attenuation models derived from measurements [7] at 58 GHz and 93 GHz frequency bands are given[8].

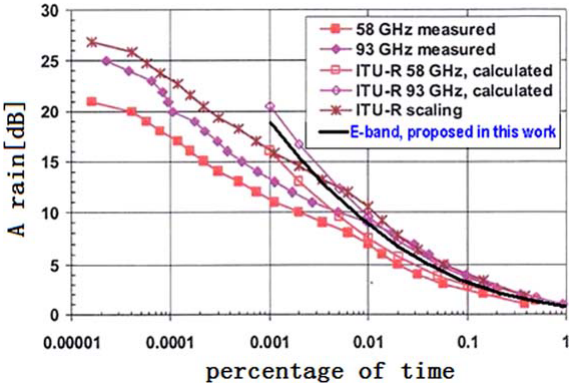


Fig. 4. Rain attenuation model in E-band (80 GHz) proposed based on ITU-R model [5], [7], [8], [10].

IV. E-band Communication Link using Advanced Transceiver Circuits

As described earlier due the advancement in semiconductor technology, several experiments are performed using advanced

RF circuits. These techniques eventually leads to reach required EIPR [7] to maintain the QoS. In this section most promising methods are presented.

A. E-band analog transmit–receive front-end designed, fabricated and packaged in wave-guide technology

In this method, Tx up-converter consists of in-phase/ quadrature (I/Q) mixer, 50 nm mHEMT with 3 dB IF bandwidth of 12 GHz. A 73.5 GHz LO signal is generated from frequency multiplier [1] using a 9.1875 GHz stable input signal. The active frequency multiplier has a conversion gain of 20 dB and output power of 10 dB. A pre amplifier is used in between active mixer (up converting mixer) and Power amplifier module [12] to boot the signal power. The pre amplifier contributes a 15 dB gain and 23 dB m saturated power. To achieve a total saturated out power of 1 W the transmitter module is packaged with a MMIC based power amplifier realized in 100 nm AlGaN/GaN HEMT technology in final stage of amplification [11].

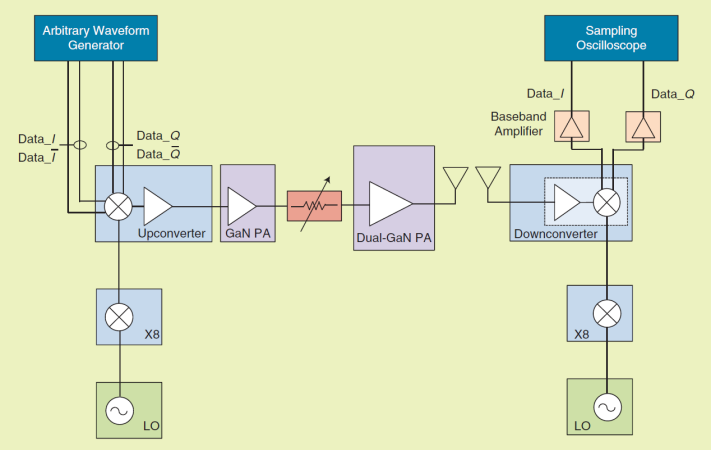


Fig. 5. The measurement setup for the E-band link [7]

1) measurement setup: To evaluate links performance, various bit streams with various bit rates, modulations formats and filters are transmitted. The differential bit are generated as input for balanced mixer, a arbitrary waveform generation is performed with a maximum sampling rate of 65GS/a. At receiver the signal processing is performed offline, the down converted IF signal is amplified using couple of broadband amplifiers realized in MMIC technology. A commercially available 20 GHz analog bandwidth oscilloscope is used to sample and store the output signals which are used in further signal processing steps. The signal processing at receiver is carried out offline. More information on transceiver is in [7].

B. MMIC Based Multi-Gigabit E-band communication link

In [3] wireless data transmission of complex modulated signals are transmitted over E-band using a MMIC based transceiver. Figure 6 represents measurement setup for a 6m incoherent wireless transmission experiment reported.

TABLE I
CALCULATED LINK BUDGET VALUES

Parameter	Value	reference
$P_{Rx,min}$	-62dB	figure 4
L_{rain}	18dB	figure 1
SNR	10dB	figure 3
FSPL	162db	[6]
G_t, G_r	88dBi	[6]

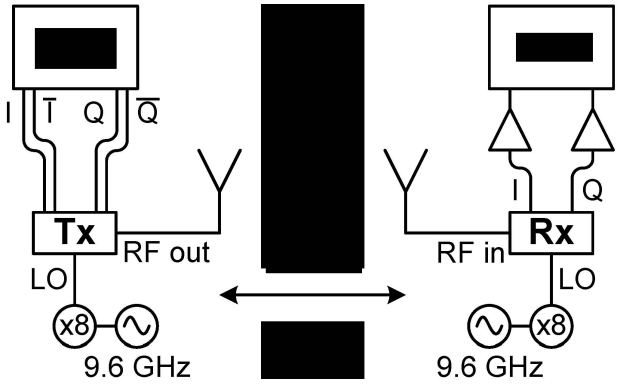


Fig. 6. Measurement setup for the 6 m incoherent wireless transmission experiment [3].

1) Measurement setup: MMIC-based receiver composed of resistive mixer and a low noise amplifier (LNA), implemented using 50 nm mHEMT technology. At transmitter 100 nm mHEMT technology based balanced mixer is used to obtain a high linear, high output power. A 36 dB local oscillator (LO) and frequency multipliers are used to achieve desired carrier frequency. Figure 6 shows a schematics of measurement setup, signal are generated at 9.6 GHz using a free running non synchronized Keysight synthesizer MXG N5183B, frequency multipliers are used to generate 77 GHz LO signals. Antenna front ends are equipped with horn antennas which provides a gain of 23 dBi, input power of -5 dB m is used [5], here atmospheric attenuation of 0.5dB/km can be ignored as the link is only for short range so, we only consider free space path loss (FSPL) of 85.74 dB at the signal center frequency of 77GHz.For this experiment Keysight arbitrary wave form generator(M8195A) is used at sampling frequency of 65GSa/s signals are generated. More details are transceiver components are in [2]

C. Direct conversion transmitter in 40 nm CMOS

A fully integrated E-band transmitter (TX) is developed in 40 nm CMOS [13]. Unique calibration techniques are employed to suppress the Local oscillator feed-through (LOFT) and I/Q imbalance over both 71-76 GHz and 81-86 GHz bands. A millimetres-wave poly-phase filter (PPF) having lowest I/Q imbalance with minimum EM simulations is developed and integrated. The 40 nm E-band transmitter achieves a measured output power of 12 dB m and TX efficiency of 15% with about 15 GHz bandwidth. Measured from 3 chips, the transmitter features an un-calibrated I/Q imbalance of less than

30 dB from 62.5-85.5 GHz. The calibration circuits further reduce the I/Q imbalance by 3-5 dB and ensure the LOFT less than -30 dB, over more than 30 dB output dynamic range. The presented TX achieves 3 dB s 64-QAM across the complete E-band.

D. Real-Time Full-Duplex E-band Link

In [4] a full duplex E band communication link is characterized. The transceiver consists of four base-band differential amplifiers, four differential amplifiers with gain control, and analog wavefronts operating in E-band frequencies are used. The link in [2] uses an uplink frequency from 81-86 GHz, and downlink frequency from 73 GHz - 76 GHz. A local oscillator operating at 9.81 GHz and 10.4375 GHz is used as input of frequency multiplier of factor eight is used for uplink and downlink respectively. Power splitters/combiners, and duplexer are realized as wave guides [9].

Several advanced transceivers are used to evaluate the transmission efficiency based on BER and Data rates for a given complex digital modulation scheme is provided in table No. (2)

TABLE II
TRANSMITTERS PERFORMANCE COMPARISON

-	Ref1	Ref2	Ref3	Ref4
Frequency	73.5 GHz	77 GHz	78.5 GHz	73.5 GHz
Modulation Data Rates(Gb/s)	64QAM 4.5	16QAM 80	64QAM 3	64QAM 7.5

V. Conclusion

Here, a comparison of various technology platforms for semiconductor-based wireless communication likes is documented. The transceivers developed by the latest CMOS technologies are used to evaluate the achievable data rates by employing complex modulation schemes that are tabulated and compared. A brief description of novel circuit design, simulation, and optimization methods are documented.

References

[1] J. Antes, U. J. Lewark, A. Tessmann, S. Wagner, A. Leuther, T. Zwick, and I. Kallfass. Mmic-based chipset for multi-gigabit satellite links in e-band. In *2012 IEEE International Conference on Wireless Information Technology and Systems (ICWITS)*, pages 1–4, 2012.

[2] J. Antes, U. J. Lewark, A. Tessmann, S. Wagner, A. Leuther, T. Zwick, and I. Kallfass. Mmic-based chipset for multi-gigabit satellite links in e-band. In *2012 IEEE International Conference on Wireless Information Technology and Systems (ICWITS)*, pages 1–4, 2012.

[3] F. Boes, J. Antes, T. Messinger, D. Meier, R. Henneberger, A. Tessmann, and I. Kallfass. Multi-gigabit e-band wireless data transmission. pages 1–4, 2015.

[4] Seyyid M. Dilek, Benjamin Schoch, and Ingmar Kallfass. Performance analysis of real-time full-duplex e-band link. In *2019 49th European Microwave Conference (EuMC)*, pages 976–979, 2019.

[5] Suiyan Geng and Xiongwen Zhao. Feasibility study of e-band mm-wave for gigabit point-to-point wireless communications. *Microwave and Optical Technology Letters*, 55, 08 2013.

[6] Salim Hanna, Zhihong Hunter Hong, Donald McLachlan, Jafar Shaker, André Brandao, Wayne Brett, and Matt Basar. Simulations of point-to-point outdoor transmission in 95-110 ghz. In *2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet)*, pages 1–4, 2021.

[7] Parisa Harati, Benjamin Schoch, Axel Tessmann, Dirk Schwantuschke, R. Henneberger, Harald Czekala, Thomas Zwick, and Ingmar Kallfass. Is e-band satellite communication viable?: Advances in modern solid-state technology open up the next frequency band for satcom. *IEEE Microwave Magazine*, 18:64–76, 11 2017.

[8] R Itu. Recommendation itu-r p. 838-2 specific attenuation model for rain for use in prediction methods; 1992199920032. *ITU: Geneva, Switzerland*, pages 1–8, 2005.

[9] Huiliang Liu, Yao Chu, Yulong Zhang, Weiguo Hou, Yinqiao Li, Yuan Yao, and Yaxing Cai. Error tolerance analysis and multipaction suppression of micromachined e-band duplexer. In *2021 IEEE International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale (3M-NANO)*, pages 218–221, 2021.

[10] Eric Regonesi, Lorenzo Luini, and Carlo Riva. Limitations of the itu-r p.838-3 model for rain specific attenuation. In *2019 13th European Conference on Antennas and Propagation (EuCAP)*, pages 1–4, 2019.

[11] D. Schwantuschke, R. Henneberger, S. Wagner, A. Tessmann, I. Kallfass, P. Brückner, R. Quay, and O. Ambacher. Gan-based e-band power amplifier modules. In *2016 46th European Microwave Conference (EuMC)*, pages 564–567, 2016.

[12] Dirk Schwantuschke, Beatriz Aja, Matthias Seelmann-Eggebert, Rüdiger Quay, Arnulf Leuther, Peter Brückner, Michael Schlechtweg, Michael Mikulla, Ingmar Kallfass, and Oliver Ambacher. Q- and e-band amplifier mmics for satellite communication. In *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, pages 1–4, 2014.

[13] Dixian Zhao and Patrick Reynaert. A 40 nm cmos e-band transmitter with compact and symmetrical layout floor-plans. *IEEE Journal of Solid-State Circuits*, 50(11):2560–2571, 2015.

Review of the state of the art satellite communication technologies

Institute of Robust Power Semiconductor Systems
University of Stuttgart
Stuttgart, Germany

***Abstract*—The use of satellite technologies has rapidly grown over the last years in different fields such as communication, research, the military, etc., which resulted in a rapid improvement of the used technology followed by an immense increase in the captured, processed, or transmitted data on the satellite. This paper focuses on the state of the art communication systems meeting the requirements of these new challenges.**

***Index Terms*—Satellite communication, microwave/millimeter-wave systems, optical communication**

I. Introduction

The capability and the quality of the used equipment on the satellites developed quickly in recent years. For example, high-resolution infrared/optical cameras or new radar systems generate data at the rate of gigabits per second. These high data volumes result in hardware challenges storing and transmitting the data considering that these satellites have a limited power source generally supplied from solar panels. Another example is the STARLINK project. Thousands of satellites are sent into a range of 500 km to 600 km and work in Ku- and Ka-bands. They provide their users access to the internet with up to 250 Mb/s download speed [1]. As it is understood from these two examples, the demand for high capacity channels with high data rates increases rapidly. Therefore, new approaches such as optical carrier systems instead of conventional millimeter-wave carrier systems, new methods of optimizing the millimeter-wave systems, or/and higher carrier frequencies are researched and tested. This paper discusses the applications and methods through some examples in satellite communications, and it analyses the state of the art technologies to fulfill the new requirements of the increasing data rate demand.

II. Different Approaches

One of the trends to meet up the high capacity demands is to move away from traditional carrier frequencies in the L-, S-, C-, X-, and Ku-bands in favor of higher frequencies in the Ka- , Q-, and V-bands to meet the growing demand for channels with higher data rates. Even though the use of extremely high frequency bands is not new, more resources are put into its research. The higher frequency bands offer wider bandwidth and higher carrier capacities. This approach is further explained in chapter III and IV.

Another solution to this ever growing requirement for an increase in data rate and multimedia services can be the shift from the traditionally employed millimeter-wave carrier to the optical carrier or a combination of both. Unlike conventional carriers, which are limited in spectrum usage, optical carriers do not require any spectrum licensing, making them an appealing option for large bandwidth and capacity applications [2]. The larger spectrum range results in a higher carrier capacity followed by a higher data rate [3]. An example regarding this approach is given in IV-A.

In addition to these trends, dynamic satellite systems are also becoming increasingly popular. Dynamic systems with supporting algorithms allow the allocation of the limited resources such as bandwidth, capacity, or data rate in a more efficient way in high traffic situations [4]. This approach is further explained in chapter V.

III. Radio Frequency (RF) Spectrum

The RF spectrum is an important part of space operations. Almost every satellite communicates with the earth or other satellites using some component of the RF spectrum. The most common frequencies for the satellite related technologies are the super high frequency (SHF) and the extremely high frequency (EHF) bands [5]. The SHF and EHF bands are given in table I.

TABLE I
Conventional and uncommon frequency bands

	Super high frequency				
	L	***S***	***C***	***X***	***Ku***
Frequency [GHz]	1–2	2–4	4–8	8–12	12–18
	Extremely high frequency				
	Ka	***Q***	***V***	***E***	
Frequency [GHz]	26–40	33–50	50–75	71–76 and 81–86	

Since the RF spectrum is a finite natural resource, increasing the number of terrestrial and space users causes RF congestion, which leads to unintentional radio frequency interference (RFI) [5]. Especially with the increase in the number of commercial satellites such as STARLINK, the possibility of RFI becomes higher. Because of that, the used frequencies expand out of the conventional bands, that are the SHF bands, into the relatively lesser used frequency bands, such as the EHF bands, to open

physical space for the many new satellite systems that are being developed. The use of the EHF bands also allows a more optimal allocation of the limited sources, avoiding RFI. Moreover, EHF bands offer a higher bandwidth, which equates to a higher channel capacity. That allows the transmission of higher volumes of data compared to the SHF bands. Therefore, EHF bands such as E-, Q-, or V-bands offer a more reasonable and flexible environment for satellites that work with high amounts of information in comparison to SHF bands [6].

IV. Possible E-Band Optical Carrier and Wave Communication Systems

E-band provides a total RF bandwidth of 10 GHz that allows a wide range of possible channel capacities depending on how much of the bandwidth is used. A larger bandwidth equates to a higher theoretical channel capacity which also translates into higher data rates. Moreover, the data rate is also directly related to the number of levels in the signal regarding its modulation and the noise on the channel. Therefore, even though higher channel capacity generates theoretically higher data rates, it is a false assumption to expect a higher data rate observing only the channel capacity. Commercial wireless point-to-point E-band networks use the regulated frequency bands of 71–76 and 81–86 GHz, with baseband bandwidths ranging from 500 to $2,000$ MHz depending on their licensing and modulation formats as sophisticated as 256 quadrature amplitude modulation (QAM), to reach data speeds of up to 10 Gb/s in the conducted tests [6].

A. An Optical Carrier System

Despite all the benefits of the E-band, bad weather has a significant impact on the quality of service (QoS) of satellite communication systems that use optical carriers in contrast to those that use micro-/millimeter-wave carriers at lower frequencies. [6]. To have a better understanding of the QoS, a link-budget calculation is necessary. It is calculated as given in formal 1 [7].

$$P_{Rx} = P_{Tx} + G_{A,TX} + G_{A,RX} - FSPL - L_{atm} \quad (1)$$

P_{Rx} is the received power, and the P_{Tx} is the output power. The $G_{A,TX}$ and the $G_{A,RX}$ are the transmit and receive gains. Free-space path loss (FSPL) and cumulative atmospheric loss L_{atm} represent the atmospheric effects such as rain, fog, or water evaporation. The challenge is to minimize the effects of the L_{atm} and to achieve the optimal FSPL as much as possible without exceeding the power limitations of the satellite itself. Therefore, improved antennas or ground and satellite hardware with higher gain, better transmission power, receiver noise gain, etc, and/or less power consumption are essential to maximizing the QoS.

Such an E-band satellite-to-ground communication system is demonstrated in Ref. [6]. This system combines low-noise blocks in indium gallium arsenide (InGaAs) metamorphic high-electron-mobility transistor technology (mHEMT), high-power devices realized in gallium nitride (GaN)-based HEMT technology, and high-gain parabolic antennas to achieve both high equivalent isotropic radiated power (EIRP) at the transmitter Tx and low sensitivity at the receiver Rx to compensate for FSPL and L_{atm} [6]. InGaAs mHEMT is a compound semiconductor technology reaching the f_T and f_{max} values in the subterahertz region. 50 nm and 100 nm metamorphic HEMT technologies are suited for the realization of low-noise amplifiers up to frequencies beyond 200 GHz that can be used in E-band communication systems [8]. Furthermore, GaN-based HEMT technology achieves saturated output power levels of more than 1 W to be gained on a single chip which is relatively high considering that satellites can only supply a limited amount of power [9]. This system achieved a maximum data rate of 6 Gb/s having the best error vector magnitude value of -11.48 dB, which was achieved with the modulation formats of QPSK and 8-PSK wherein the Tx chain was working near saturation [6]. Further research is necessary to improve the data rate, which is possible by solving the saturation problem of the Tx chain and then using a more efficient modulation method like 16 QAM or even 256 QAM.

B. A mm-Wave Communication System

The high capacity mm-wave communication link system between a fixed ground location and an aircraft in the Ref. [10] is a different approach compared to the system in chapter IV. Considering that the aircraft moves at a lower level and a different speed. This difference changes the effective noise in the system. It is demonstrated that it achieved a 40 Gb/s capability both in the uplink and downlink direction with a slant range of up to 12 km.

The system can be divided into the Communication (COM) and the Pointing, Acquisition, and Tracking (PAT) subsystems. The baseband modem, optical fiber interface, RF transceiver, multiplexer and polarizer, antenna, and data stream interfaces are all part of the COM system. The baseband modem is based on four commercially available high capacity mm-Wave modems (HCMWM), which reduces the cost immensely. Wideband single carrier QAM modulations are used in each HCMWM (BPSK to 128QAM). The greatest data rate mode employs 128QAM with a symbol rate of 1.6 Gb/s, giving each HCMWM an information output of around 10 Gb/s. To achieve the 40 Gb/s capabilities in both the uplink and downlink directions, the transceiver system multiplexes four of these 10 Gb/s modulated carriers. Furthermore, the PAT system is in charge of the antenna beam's pointing, acquisition, and tracking to and from the moving aircraft. A pointing accuracy of better than 0.05 is required for the ground antenna, which equates to a 0.5 dB loss. Because the initial null of the antenna pattern is positioned at 0.2, a 0.2 miss-pointing would be equivalent to 32 dB of loss. This sensitivity caused calibration problems during the tests. However, data rates of 40 Gb/s downlink was achieved and bit-error-rate of 10^{-5} to 10^{-7} was observed [10].

V. Dynamic Approach

A dynamic approach generally consists of different sections aimed to optimize the data transmission as much as possible.

Next to hardware and physical methods of increasing the data rate and data capacity, such as utilizing E-, Q-, or V-bands and employing an optical carrier, software solutions and algorithms also provide efficient ways to meet the requirements. Such a dynamic system can consist of many different parts, like more than one ground station, more than one satellite, or different aircrafts in between. Different possible systems are visualized in figure 1 and 2. These examples are just two of many possibilities.

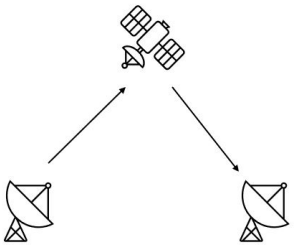


Fig. 1. Example of a direct satellite to ground communication system

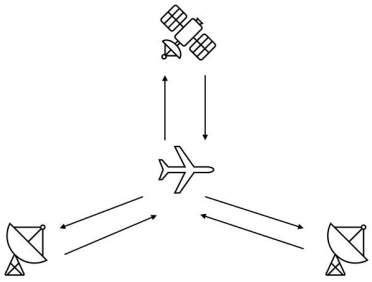


Fig. 2. Example of a satellite to aircraft to ground communication system

A. Different Optimization Methods

Design of power and beam allocation over satellite downlinks in terms of Shannon capacity, jointly based on traffic distribution and channel conditions and maximizing system performance while providing reasonable fairness among users, is one of the well-known optimization strategies. The onboard transmission power is divided and shared across transmitters to allocate capacities to cells to create the necessary model. Besides, a multiple beam antenna or a phased array antenna is used to broadcast several signals. This method allows using the coverage area of the cells more flexibly and dynamically in the case of high-demand situations. When the traffic at each cell increases or backs off, the cells can be scanned sequentially across the coverage area to satisfy the demand. As a result, relatively lower data rates would be sufficient to provide the necessary information flow [11].

A different optimization method is an optimal dynamic capacity allocation scheme based on Monge arrays minimizing the losses and rate matching performance metrics [12]. This algorithm reorganizes the offered capacity and requested capacity in scenarios with future High Throughput Satellites (HTS) systems to achieve gateway-user equipment pairing. For the

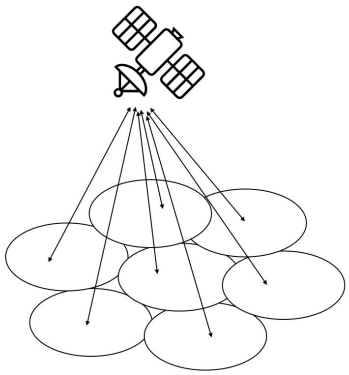


Fig. 3. Multi-beam allocation

offered capacity, realistic spatial-temporal total atmospheric attenuation channels are implemented. Its low complexity provides a highly promising general, fast and optimal solution for dynamic satellite environments in regards to multi-beam HTS systems [12].

B. A Dynamic System: QV-LIFT Project

Q- and V-bands offer a higher capacity solution with bigger bandwidths to the higher data demands similar to the E-band, as explained in chapter III. The concept of HTS, which operate in the EHF bands, is supported to achieve the so-called Terabit connectivity. The QV-LIFT project is based on a cutting-edge and adaptable architecture that combines a high performance satellite IP-based network with a system based upon the idea of smart gateways, which has been pointed out as a potential technique to mitigate the propagation impairments in the Q/V band [13]. A simplified model of the system is shown in 4.

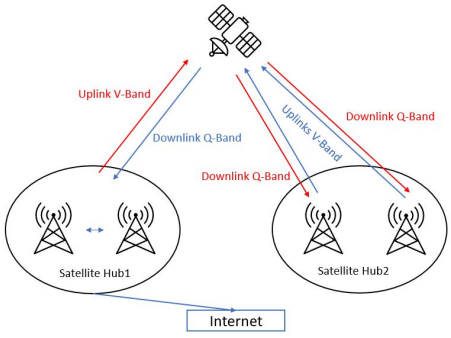


Fig. 4. Simplified Model of QV-LIFT Project

New hardware is developed to satisfy the need of the ground segment of the project. A GaN power amplifier Monolithic microwave integrated circuit (MMIC) at V band, a power combining High Power Amplifier based on GaN MMIC power amplifiers, a BUC (Block Up Converter) operating in the V band, an LNB (Low Noise Block down converter) operating in the Q band are the main focus of the hardware research [13].

A smart Gateway Management System (SGMS) is added to the system, compatible with the developed hardware. SGMS

is employed to manage propagation impairments, gateways handover, and to ensure bi-directional IP connectivity to the QV-LIFT network. Besides that, the integration of the existing systems into the QV-LIFT system is possible with the help of the SGMS [13].

VI. Comparison of Optical and mm-wave Carriers

The requirements of a satellite communication system highly depend on its tasks and its used environment. Therefore, a compromise cannot be avoided considering what is required from the developed system. Optical carrier systems offer an undeniable theoretical advantage in the aspects of data rate and carrier capacity compared to the mm-wave carriers. Nevertheless, this can easily be undermined with a low QoS if the optical carrier system isn‘t supported by the necessary hardware considering the high sensitivity of the optical carrier against weather and atmospheric effects. On the other hand, the mm-wave carrier systems offer a more stable system with a lower capacity, which can be more suitable for low information density systems.

In the work of Schulz [14] a point-to-point optical wireless link (OWC) based on infrared LEDs is described and compared with a 60 GHz mm-wave radio link in a 100 m outdoor deployment. Even though this does not translate directly into the satellite communication systems, this works provides a clear understanding of the differences between the optical and mm-wave carriers. The observed data during the conducted tests are given in table II.

TABLE II
Comparison of Optical and mm-wave Carriers in a 100 m outdoor deployment

	Data Rate [Mbps]	
	Minimum	***Maximum***
OWC link	20	560
mm-wave link	30	350

The minimum data rate for the optical link, as determined by a low visibility event with more than 160 dB/km attenuation, was 20 Mbps. The minimum rate for the mm-wave link is higher at 30 Mbps. The optical link, which is more than twice as fast as the mm-wave link on average, consistently achieves data rates of 360 Mbps. Nevertheless, the optical link shows higher fluctuations in case of bad weather conditions [14]. As a result, while the mm-wave link has higher minimum data rates in bad weather conditions and is more stable, the optical link has higher data rates most of the time, which makes it a reasonable alternative.

VII. Conclusion

This paper analyzes the state of the art technologies in satellite communication systems and optimization methods to improve the architecture of these systems in different aspects such as data rate and channel capacity. Examples of mm-wave, optical, and dynamic systems are examined and discussed. Even though some of these systems offer better solutions to different problems, one system is not better than the other one. The use of the methods depends on the requirement of the mission on hand. Therefore, a trade-off between different aspects cannot be avoided. Nonetheless, recent Projects, like the QV-LIFT project, demonstrates that a dynamic approach provides a solution that improves many different aspects simultaneously, creating a more adaptable system. Consequently, more dynamic systems with new components subsidized with different algorithms combining both hardware solutions and diverse optimization methods, enhancing flexibility, can be expected in the upcoming years considering the demands and requirements of the market and users change and increase rapidly.

References

[1] STARLINK official website, accessed 20 June 2022, https://www.starlink.com/

[2] H. Kaushal1 and G. Kaddoum, “Free Space Optical Communication: Challenges and Mitigation Techniques” 1Department of Electrical, Electronics and Communication Engineering, ITM University, Gurgaon, Haryana, India-122017.

[3] V. W. S. Chan, “Free-space optical communications,” J. Lightwave Tech., vol. 24, no. 12, pp. 4750–4762, 2006.

[4] D. W. Young, H. H. Hurt, J. E. Sluz and J. C. Juarez, “Development and Demonstration of Laser Communications Systems” , Johns Hopkins APL Technical Digest, Vol. 33, No. 2, 2015

[5] Brian Weeden, “Radio Frequency Spectrum, Interference and Satellites Fact Sheet”, Secure World Foundation, 2013.

[6] P. Harati et al., ”Is E-Band Satellite Communication Viable?: Advances in Modern Solid-State Technology Open Up the Next Frequency Band for SatCom,” in IEEE Microwave Magazine, vol. 18, no. 7, pp. 64-76, Nov.-Dec. 2017, doi: 10.1109/MMM.2017.2738898.

[7] U. J. Lewark et al., “Link budget analysis for future E-band gigabit satellite communication links (71–76 and 81–84 Ghz)” CEAS Space J., vol. 4, no. 1, pp. 41, June 2013.

[8] A. Leuther, A. Tessmann, I. Kallfass, R. Losch, M. Seelmann-Eggebert, N. Wadefalk, F. Schafer, J. D. Gallego, M. Schlechtweg, M. Mikulla, and O. Ambacher, “Metamorphic HEMT technology for low-noise applications,” in Proc. IEEE Int. Conf. Indium Phosphide and Related Material, Newport Beach, CA, 2009, pp. 188–191.

[9] C. Haupt, S. Maroidt, R. Quay, W. Pletschen, A. Leuther, and O. Ambacher, “Development of transconductance GaN MMIC technology for millimeter-wave applications,” Phys. Stat. Sol. C, vol. 8, no. 2, pp. 297–299, 2011.

[10] Q. Tang et al., ”Demonstration of a 40Gbps Bi-directional Air-to-Ground Millimeter Wave Communication Link,” 2019 IEEE MTT-S International Microwave Symposium (IMS), Boston, MA, USA, 2019, pp. 746-749, doi: 10.1109/MWSYM.2019.8700741.

[11] Jihwan P. Choi, Vincent W. S. Chan, “Optimum Power and Beam Allocation Based on Traffic Demands and Channel Conditions Over Satellite Downlinks” , IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 4, NO. 6, NOVEMBER 2005.

[12] Anargyros J. Roumeliotis , Charilaos I. Kourogiorgas, Athanasios D. Panagopoulos, “Optimal Dynamic Capacity Allocation for High Throughput Satellite Communications Systems”, IEEE WIRELESS COMMUNICATIONS LETTERS, VOL. 8, NO. 2, APRIL 2019.

[13] Giuseppe Codispoti, Giorgia Parca, et al. “Validation of ground technologies for future Q/V band satellite systems: the QV – LIFT project” , 018 Aerospace Conference, Big Sky, Montana.

[14] D. Schulz, J. Hohmann, J. Hilt, P. Hellwig, V. Jungnickel “Outdoor Performance Comparison of Optical Wireless and Millimeter Wave” , Fraunhofer Heinrich Hertz Institute, 2018, Berlin.

Techniques to Actively Control Electromagnetic Interference in Gallium Nitride Power Circuits

Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*—Due to the introduction of high-speed switching power components like Gallium Nitride HEMTs, tackeling electromagnetic interference becomes serious. Since Passive Filters are large and difficult to handle at high frequencies Active Control Techniques are a promesing approach. This paper brings together and compares different concepts to actively control EMI emissions in Gallium Nitride based power circuits. Therefore strategies how to prevent the problem on its origin and how to supress the propagation path are shown. By comparing their electrical performance, complexity, variability and costs an overview of active electromagnetic interference control mechanisms is given.**

***Index Terms*—Power Devices, Gallium Nitride, Electromagnetic Interference, Active control**

I. Introduction

In order to make more electrial efficient, lighter and smaller power devices increasing the switching frequency (f_{SW}) is an appropriate way. With modern Wide Bandgap Semiconductors like Gallium Nidride (GaN) f_{SW} in the MHz range are no exception. However, this comes along with the challenge of Electromagentic Interference (EMI) [2]–[5]. Conventional Methods to reduce the effect of common mode (CM) and differential mode (DM) detraction in power devices can be devided into three categories:

1) Passiv EMI filters: Different condensators and inductivities can be utilized to block CM and DM interferences. Though, compromises with in the package size, efficiency and performance in high frequency ranges have to be made [5].
2) Reducing parasitic inductivities and capacities: The current and voltage transients generated by fast switching power devices are causing parasitics to resonate. The induced oscillations, overshoots and ringing are considered as the main cause for EMI. Optimizing the PCB design to decrase the parasitics and therfore raising the resonance frequency can also ensure Electromagnetic Compability (EMC) [4].
3) Limiting f_{SW} and voltage slew rate: A decrease in the Drain-Source Voltage Slope (dV_{DS}/dt) is achieved with a higher gate resistance. Lowering f_{SW} also results in less emitted distortions. Although, the EMI problem can be reduced with this methods power loss during the swiching event is increasing due to larger cross times [4], [5].

Improving EMC preformance and power loss while cuncurrently maintaining high f_{SW} is a goal which is hardly compatible with the above mentioned approaches. Thus, active EMI control (AEC) is a promising alternative to master this. Such AEC techniques are active EMI filters (AEF) and active gate driving (AGD). The essential difference between these is the conceptual point of attack. With AGD the basic idea is preventing the occurence of EMI at its root. Whereas, applying AEF tackels the problem at the propagation path. This paper is organized as follows: first, an overview of EMI generation is given. Section III then presents the concept of AGD and how it is applied. In Section IV the concept of AEF is presented. After this a Comparison of the two methods is given and summed up. Finally, in Section V the Result is discussed.

II. EMI in Power Circuits

In general, EMI in power circuits is generated by the fast voltage or current switching transients [1]–[7]. Parasitic Inductivities and capacities which occure in wires or the junction of a free wheeling diode are resulting in resonant circuits which are triggered by the high voltage slopes (dv/dt) and current slopes (di/dt). The induced oscillations, overshoots and ringing (Fig. 1) can damage the device itself or lead to conducted and in case of very high frequency oscillations also radiated EMI.

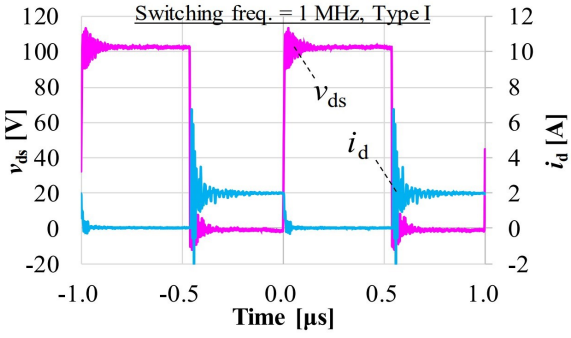


Fig. 1. Vds and Id of a GaN DC/DC-Boost Converter at 1MHz [12]

III. Active Gate Driving

AGD already proved to be an efficient way for reducing EMI and simultaneously lowering the switching losses in silicon carbide (SiC) MOSFET power devices [3]. However, with higher f_{SW} and even faster switching transients EMC challenges on AGD in GaN power devices are quite demanding. The main idea of AGD is to prevent the occurance of oscillations, overshoots and ringing while turn-on and turn-off by manipulating parameters of the gate circuit. Fig. 2 shows the conceptual control diagram of an AGD. Since the transconductance (g_m), Drain-Source Resistance (R_{DS}) and internal Gate Resitance RG, int can not be controlled externally, only f_{SW}, the Gate Current (I_G) and the external Gate Resistance $R_{G,ext}$ can be utilized to manage this concept. In case of AGD for GaN-FETs one major challenge are the very fast rising (t_r) and fall time (t_f) requiring control techniques with sub-nanosecond response and actuating times. Whereby the technique of f_{SW}-modulation has no direct influence on the switching event.

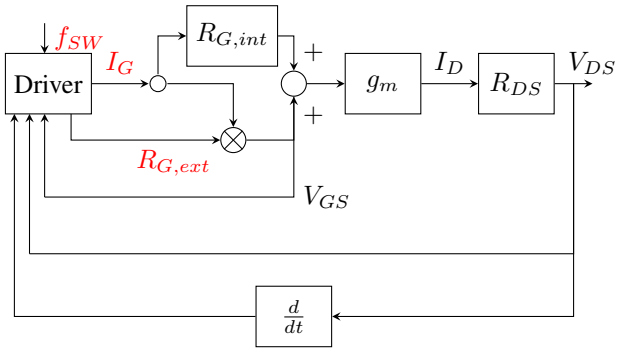


Fig. 2. Conceptual Control Diagram of Active Gate Driving

A. Controlling the Voltage Slope

1) Active Gate Resistance Control: As mentioned in the introduction, increasing R_G successfully decreases dv/dt and therefore EMI but also due to higher cross times during the switching event power loss increases. To solve this issue [4] proposing the control of $R_{G,ext}$. By changing the value $R_{G,ext}$ multiple times during the turn-on and turn-off event the Drain-Source Voltage (V_{DS}) waveform can be shaped. [4] utilizes a programmable high speed AGD that can change the output pull-up and pull-down resistance every 150 ps. With this open loop technique a predefined profile of R_G, ext is applied. The turn-on event is split in 4 subsequences where different $R_{G,ext}$ are operating. First, during the turn-on delay phase a large pull-up resistance (R_{PU}) is choosen. A strong I_G is the result. Thus, time till Gate-Source Voltage (V_{GS}) reaches the Threshold-Voltage (V_{TH}) is minimized. Then, during current commutation time R_{PU} is reduced so di/dt does not exed a certain limit. To prevent current ringing at the end of this phase R_{PU} is increased. Here, also the trade off between the di/dt-limits and overlap-loss of V_{DS} and V_{DS} has to be considered. While V_{DS} commutates during the miller plateau (MP) R_{PU} is reduced step by step to control dv/dt and reduce oscillations. At the end of this segment the driving strength then is increased again slightly to hold the overlap-loss to a minimum. In the last segment of the turn-on sequence R_{PU} is not changed because here V_{DS} does not have an effect on overshoots and oscillations. Similar to the turn-on sequence the turn-off sequence consists also of four segments. Here, the delay time at the beginning also is kept shot by allowing a higher I_G out of the gate by means of a high pull-down resistance (R_{PD}). When V_{GS} reaches the MP the same trade off as during the turn off transition has to be made and a resitor sequence which restricts dv/dt and keeps the overlap-loss small has to be applied. After the voltage commutated current starts to commutate. Here, a smaller R_{PD} is choosen to decrease di/dt and the resulting voltage overshoot induced by the parasitic inductance. Finally, R_{PD} does not affect the switching transition after V_{GS} drops below V_{TH}, therefore a high R_{PD} is applied keeping the gate in off-state [4].

2) Active Gate Current Control: To directly control the voltage slope via I_G [5]–[7] each suggesting closed loop techniques. While [6] relies on a complete digital approach [5], [7] utilizing an analog approach. Similar to AGD with $R_{G,ext}$ a manipulated I_G during the current and voltage commutation is occured therfore dv/dt, oscillations and overshoots can be controlled. [6] proposes the digital control of I_G by means of the first and second derivative of V_{DS} as feedback signals. The control unit generates a profil of I_G which is derived from the given limits for both derivatives. Doing so helps in case of the first derivative obviously to limit dv/dt during the switching transitions and in case of the second derivative also the oscillations are getting reduced. Cutting down the control meachnism to a mathematical model with two parameters has the advantage of scalability and independence of load conditions. The feedback loop and current source can also be implemented in an analog manner which eliminates a complicated digital logic. [7] suggests a sensing condensator that transfers the measured voltage in a current proportional to the derivative of V_{DS} and the capactor itself. This measured current then is fed into an inverse current mirror that reduces I_G during the MP and therefore slows down dv/dt. The main difference between a simple increas of $R_{G,ext}$ and this method is that the current commutation phase is not influenced thus the overlap time can be hold to a minimum. So at similar dv/dt as with a constant $R_{G,ext}$ the switching losses can be decreased. [5] expands the thought of reducing I_G during the MP by increasing I_G before and after MP to decrase the overlap time and simultaniously kepping dv/dt and oscillations under control. By means of adaptive MP detection and compensating the feedback delay it is possible to not comprimise the swiching speed, reducing EMI and switchig losses compared to a constant driving current.

B. Modulating the f_{SW}

Another technique to actively reduce EMI in switching devices is based on the spreading of the inference power spectrum [8]. This is attained by modulating f_{SW}, therefore

power peaks in the power spectrum are getting split Fig. 3. [8] proposes three different spread spectrum modulation

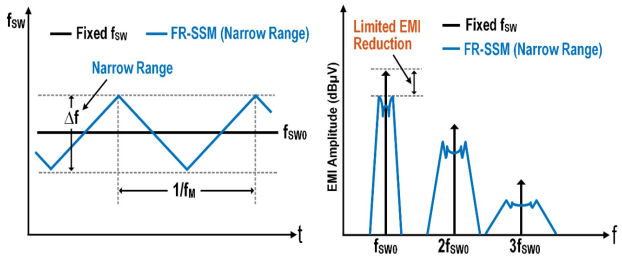


Fig. 3. Time domain and power spectrum of FR-SSM [8]

(SSM) techniques . First, a fixed rate SSM (FR-SSM) then two methods for a randomized SSM (RSSM) and then a multi rate SSM (MR-SSM) are described. FR-SSM, as the name suggest, modulates f_{SW} linearly with a fixed rate and modulation frequency (f_M) arround f_{SW}. Althrough, a solid peak EMI reduction can be achieved, the potential of FR-SSM is limited by the maximum frequency range (Δf). A to large Δf causes the overlap of the splited frequency bands and results in new peaks in the power spectrum. To reduce the EMI peaks even further RSSM is proposed. The implemantation can be realized both, in discrete manner (D-RSSM) as well as in a more cost friendly analog continous manner (C-RSSM). C-RSSM utilizes the avalanche noise of a Zener diode to randomize the modulation. On the backside C-RSSM makes the precise prediction of the duty ratio difficult. Even though [8] suggests a closed loop methode for C-RSSM the exact voltage regualtion remains complex and can couse problems in PWM driven power circuits. A solid compromise between the good EMC performance of C-RSSM and the better controllable FR-SSM is the Multi Rate SSM (MR-SSM). Here, the rate of frequency modulation is cloosed loop controlled. As a result the duty cycle is predictable and no overlaps in the power spectrum occure.

IV. Active EMI Filters

In contrast to AGD, AEF deals with the emitted EMI and mitigates the impact [9]–[11]. Fig. 4 shows the conceptual model of and an AEF. It distinguishes two types of compensation filter. First, the AEF is classified wheter it suppresses CM or DM [11]. Also, a differentiation between the feedback (FB) type, which senses the current or voltage before the Line Impedance Stabilization Network (LISN) and feeds back the compensation signal after the Device Unter Test (DUT) and the feedforward (FF) type senses the current or voltage at the DUT and inject the compensation before LISN has to be made. The major benefit of a FF concept is that the disturbance does not effect a receiving device, since it is canceled out before. In contrast, the FB concept has the advantage that it is more robust since it is a cloosed system. Furthermore, for a good EMI rejection an op-amp with a constant gain at a large bandwidth is required [9].

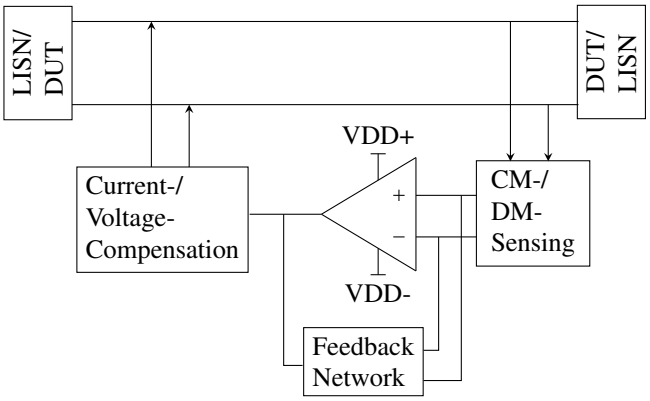


Fig. 4. Conceptual model of an Active EMI Filter

A. Active CM Filters

To sense the CM entity a CM current transformer is utilized. Taking advantage of the Low-pass characteristics only the high frequency parts are getting transformed and the desired signal is not significantly disturbed. Then, an op-amp inverts and amplifies the entity. The injection back into the phases can be realized as a current or voltage compensation. Fig 5 shows the current-compensation realized with two condensators. Whereas, the voltage-compensation is implemented with a drive choke in Fig. 6. In [11] the best EMI rejection result for CM appear with the current-canceling concept.

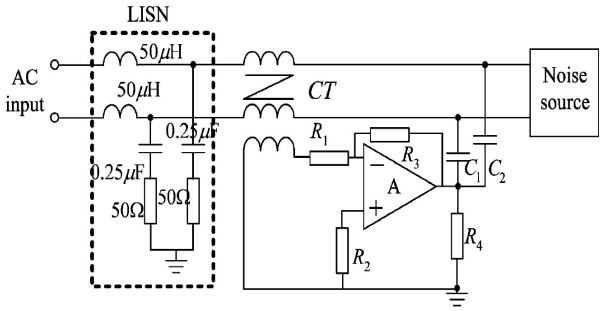


Fig. 5. Scheme of CM active filter with current-canceling [11]

B. Active DM Filters

The Sensing of DM distortion in AEF is implemented with two high-pass filters Fig 6. In contrast to the CM AEF, the compensation signal results from the differential sum of two phase signals. In [11] the sensing path is also built with two high frequency transformers, which have the benefit of a pre amplification and, even more important, a high CMRR. Similar to the CM AEF the injection back into the phases can be executed for voltage-canceling and for current-canceling. In [11] the best EMI rejection result for DM appear with the Voltage-canceling concept.

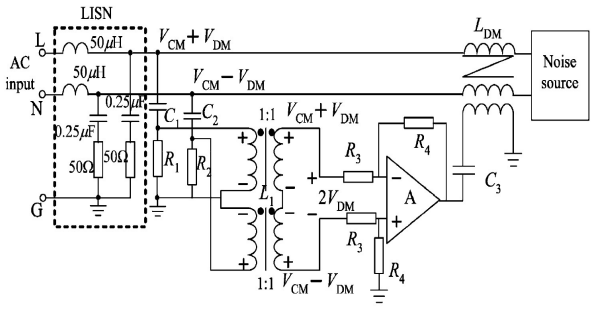


Fig. 6. Scheme of DM active filter with voltage-canceling [11]

V. Comparison

The active EMI controlling methods are campared based on their high frequency capability, adaptability, complexity of the concept, its package size and their cost. GaN based power devices often reache f_{SW} in the MHz range so the high frequency (HF) capability of an effectiv EMI control has to be given. In this point of view the AGD techniques of activly controlling $R_{G,ext}$, f_{SW}-modulation with an open loop concepts have their advantages. Since no feedback is needed the time for this path does limit the maximal frequency. A closer inspection in this matter as to be done on the closed loop AGD concepts. The described digital control approchach of dv/dt via I_G [6] is too slow since the minimal f_{SW} mentioned were 100Hz and therefore not be considered any further. On the contrary, the analog approch of [5], [7] is suitable for HF purposes. AEF HF capability is limited by the bandwith of the op-amp. In [10], [11] the upper cut off frequeny lays at arround 30 MHz, therefore HF capability is given.

When comparing the active control methods in terms of their adaptibility to changing conditions like a different load or voltage closed loop techniques like AEG via I_G and AEF benefit from their feedback loop. In case of SSM it has to be mentioned that the investigated methods [8] are only tested at voltages up to 40 V and currents up to 5A. So the adaptivity is not researched for higher powers. Also, the most promising methode MR-SSM is not suitable for PWM driven applications like an electric motor, since the variance in frequency can cause mechanical problems.

The aspect of complexity often plays a role when it comes to feasability. Here, AEF and AGD via I_G profits from an analog structure. Whereas the 6.7 GHz digital AGD [4] is very complex in terms of architecture and tuning.

Another influence factor on the feasability is the size of the EMI reduction method. Compared with each other, AGD takes up less space in contrast to AEF. AGD can be integrated on a chip [5], [7] and no current transformer or bigger capacity is needed less volume on a chip is used.

Even more important for the feasability are costs. No exact numbers are given in the references. Under the assumption that an integrated self developed AGD is more costly than AEF consisting of non integrated parts, the AEF is superior in this respect. Tab. I sums up all the above mentioned Tab. I

Concept	AGD			AEF
	$R_{G,ext}$	I_G	f_{SW}	+
HF-capability	+	0	+	+
Adaptability	-	0	+	+
Complexity	- -	+	+	++
Size	-	+	+	-
Cost	-	+	+	++

TABLE I
Comparison of the active EMI control methods

does not take in account the electrical efficiency of AGD and AEF. Here, AGD brings in a big advantage. Due to the active shaping of the switching transients also the loss power gets reduces compared to a constant $R_{G,ext}$ or passiv EMI filters. Less oszillations and shortened overlap times incresing the efficiency due to decreasing switching losses [4], [5], [7]. Whereas, AEF additionaly needs a power supply which and does not activly increase the efficiency.

VI. Conclusion and Discussion

This paper presents and compares different approaches to activly control EMI in Gan power circuits. With an AGD concepts to shape the switching transients in the time and frequency domain are presented. Whereas AEF suppresses the conduction path of EMI activly. AGD is an effective, adaptable, space-saving but more costly and complex way to decrease EMI and simultanously increasing the efficiency compared to a constant $R_{G,ext}$ or passiv EMI filters. AEF, in contrast prove to be a more cost friendly way for low power devices to mitigate EMI. So, AGD and AEF both are a reasonable alternatives to passiv EMI filters in GaN power circuits.

Futher research topics on active EMI control techniques could be the high power application of AEF, a closed loop control for $R_{G,ext}$, a more HF suitable approache for the digital slew rate control and a plausibility check for soft switching devices

References

[1] Sheng Ye, W. Eberle and Yan-Fei Liu, "A novel EMI filter design method for switching power supplies," in IEEE Transactions on Power Electronics, vol. 19, no. 6, pp. 1668-1678, Nov. 2004.

[2] M. Ali, J. Friebe and A. Mertens, "Design and Optimization of Input and Output EMI Filters under the Influence of Parasitic Couplings," 2021 23rd European Conference on Power Electronics and Applications (EPE'21 ECCE Europe), 2021, pp. 1-10.

[3] A. P. Camacho, V. Sala, H. Ghorbani and J. L. R. Martinez, "A Novel Active Gate Driver for Improving SiC MOSFET Switching Trajectory," in IEEE Transactions on Industrial Electronics, vol. 64, no. 11, pp. 9032-9042, Nov. 2017.

[4] H. C. P. Dymond et al., "A 6.7-GHz Active Gate Driver for GaN FETs to Combat Overshoot, Ringing, and EMI," in IEEE Transactions on Power Electronics, vol. 33, no. 1, pp. 581-594, Jan. 2018.

[5] J. Zhu et al., "33.2 A 600V GaN Active Gate Driver with Dynamic Feedback Delay Compensation Technique Achieving 22.5 % Turn-On Energy Saving," 2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 462-464.

[6] M. Blank, T. Glück, A. Kugi and H. Kreuter, "Digital Slew Rate and S-Shape Control for Smart Power Switches to Reduce EMI Generation," in IEEE Transactions on Power Electronics, vol. 30, no. 9, pp. 5170-5180, Sept. 2015.

[7] P. Bau et al., "Sub-Nanosecond Delay CMOS Active Gate Driver for Closed-Loop $\mathrm{d}v/\mathrm{d}t$ Control of GaN Transistors," 2019 31st International Symposium on Power Semiconductor Devices and ICs (ISPSD), 2019, pp. 75-78.
[8] D. B. Ma, D. Yan and L. Du, "Active Conducted EMI Suppression in GaN Switching Power Circuits," 2021 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), 2021, pp. 1-6.
[9] D. Shin, S. Jeong, Y. Baek, C. Park, G. Park and J. Kim, "A Balanced Feedforward Current-Sense Current-Compensation Active EMI Filter for Common-Mode Noise Reduction," in IEEE Transactions on Electromagnetic Compatibility, vol. 62, no. 2, pp. 386-397, April 2020.
[10] Yilin Sha, Wenjie Chen, Heyuan Qi, Yaqiang Han, Xu Yang and Yuan Hao, "Research of active EMI filter for Gallium Nitride based high frequency resonant converter," 2016 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC), 2016, pp. 491-494.
[11] W. Chen, W. Zhang, X. Yang, Z. Sheng and Z. Wang, "An Experimental Study of Common- and Differential-Mode Active EMI Filter Compensation Characteristics," in IEEE Transactions on Electromagnetic Compatibility, vol. 51, no. 3, pp. 683-691, Aug. 2009.
[12] T. Ibuchi and T. Funaki, "A study on parasitic inductance reduction design in GaN-based power converter for high-frequency switching operation," 2017 International Symposium on Electromagnetic Compatibility - EMC EUROPE, 2017, pp. 1-5,

Low Inductance Power Loop Design for High Frequency, High Power-Density GaN DCDC Converters

Institute for Robust Power Semiconductor Systems
University of Stuttgart
Stuttgart, Germany

***Abstract*—Modern wide bandgap semiconductors such as Gallium Nitride offer the possibility to use high frequencies in the range of MHz for high power density in power cell designs.With this high frequency and high power density, the influence of parasitics during the switching process is significant, which leads to high losses and device failure. For these reasons, the parasitic inductance should be reduced in both the power circuit and the driver circuit. This paper will explain the influence of parasitic inductance in high-frequency GaN power cells and different methods to reduce the inductance of power loops and then compare the solutions in terms of complexity , utility and feasibility. It will also review the state-of-the-art in low inductance power cells for power converters .**

***Index Terms*—High frequency switching, power loop inductance, field cancellation, snubber circuit, EMI, PCB-Design**

I. Introduction

In recent years, demand for high efficiency and high power density power devices has been increasing rapidly. This is possible with high switching frequencies. At higher frequencies, the size of passive components decreases as they have to store energy for a shorter period of time. Gallium Nitride high-electron-mobility transistors (GaN HEMTs) are promoted to be enablers for those requirements with low on-state resistance, high carrier density and minimum input-output capacitance [1]. Increasing the switching frequency can increase power density.Although at high frequencies, small parasitics of the circuit lead to severe voltage overshoot, ringing, which induces Electromagnetic interference(EMI) and sometimes even breakdown of the devices [2]–[7]. Because of GaN HEMTs' faster switching speed than silicon MOSFETs, high frequency GaN HEMTs are more sensitive to parasitic inductance. This is the reason why low inductive PCB design is necessary to reduce the influence of parasitic inductance [8].

In this section, main parasitic inductance and their influence on high-frequency GaN power devices' switching operation will be discussed. There are three main parasitic inductances which affect the most during the switching process : gate loop inductance L_{g} , common source inductance L_{s} and power loop inductance L_{p}. which is shown in Fig 1.

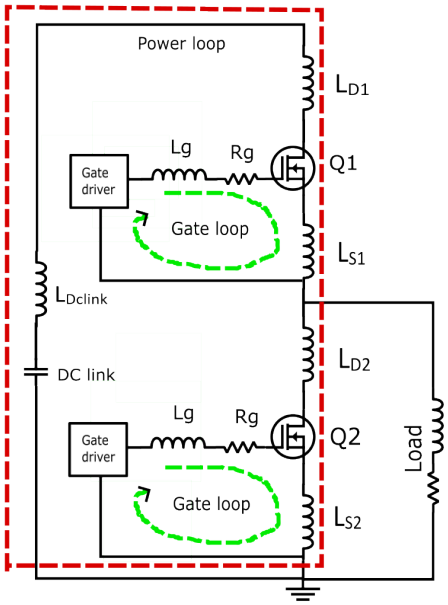


Fig. 1. Parasitic Inductance in Halfbridge

A. Common Source Inductance

Common source inductance is the inductance shared by the drain to source power current path and gate driving loop [9].It is mainly depends on package of GaN device and mostly impact on driving speed of device.Gate current can be determined by the equation (1) where V_{Drive} is Gate driver supply voltage,V_{Gs} is GaN transistor threshold voltage, as per below equations of gate current I_{g}, V_{LS} is voltage across common source inductance and R_{g} is input gate resistance.

$$I_{\mathrm{g}} = \frac{V_{\mathrm{Drive}} - V_{\mathrm{Gs}} - V_{\mathrm{LS}}}{R_{\mathrm{g}}} = \frac{V_{\mathrm{Drive}} - V_{\mathrm{Gs}} - L_{\mathrm{s}}\frac{dI_{\mathrm{D}}}{dt}}{R_{\mathrm{g}}} \quad (1)$$

During the turn on process, rising drain current across the common source inductance will decrease the overall value of gate current I_{g} than expected as per the equation 1. So a lower gate current value leads to a slower gate driving speed. While turn off process drain current decreases and overall gate current increases which make again negative impact on switching process [10].

Reduction of parasitic in GaN HEMT is utmost priority from the manufacturing process to the end process of PCB layout. So, designers of the GaN transistor made the reduction of package parasitic with the lateral design so that the die can be soldered on the same side of the PCB. This kind of package design reduces the parasitic due to internal vias connection. To further decrease parasitic, the drain and source connections are arranged in an interleaved land grid array(LGA), providing multiple parallel connections to the PCB from the die. Interleaved structure with opposite current flow will lead to self field cancellation and reduces the inductance. The result is a device package inductance of Advanced package such as land grid array, GaNPX ,PQFN which have very low parasitic inductance in the range of a couple hundred picohenry [1] With this advanced package at higher density, our only constraints are gate loop inductance and power loop inductance, which is why methods of PCB layout of power cells are the most crucial part . Although the common source inductance is very small with respect to gate and power loop inductance, they require more optimization [8] .

B. Gate Loop Inductance

Gate loop inductance mostly depends on gate driver circuit design , how the gate driver is placed, and the length of the gate signal trace. In [8] an optimized double layer layout to reduce the gate loop inductance is investigated. Experimental results support those methods.

C. Power Loop Inductance

The high frequency power loop inductance L_{p} contains parasitic inductance of dc link capacitor L_{Dclink}, package inductance of switching devices L_{switch} and power loop PCB trace's inductance $L_{\mathrm{PCB\ trace}}$ as shown in Fig 1. where Q_1 is a high-side transistor and Q_2 is a low-side transistor [8].

$$L_{\mathrm{Loop}} = L_{\mathrm{Dclink}} + L_{\mathrm{PCB\ trace}} + L_{\mathrm{switch}} \quad (2)$$

Power loop inductance has a direct impact on the switching speed and peak drain to source voltage V_{DS} spike during turn off. Drain current I_{d} increases during the turn on transition. And the result is that voltage across power loop inductance L_{p}increases (as di/dt increases), which leads to overall voltage across the half bridge being low . Therefore, switching losses decrease. So it helps during turn-on time.During the turn off process, drain current decreases. Negative di/dt across the loop inductance induces the negative voltage. The loop inductance increases the effective voltage across the device. That causes a peak voltage spike and increasing switching loss [11].While the common source inductance is controlled by the advanced packaging of the GaN, the loop inductance has a large effect on switching losses.

II. Different PCB Layouts and Techniques to Reduce Power Loop Parasitic Inductance

With fast switching speed and low inductive package inductance of GaN HEMT layout design have a huge impact on total parasitic inductance of power loop which can degrade the performance of the devices.The following section discusses different state-of-the-art PCB layout strategies for GaN converters [10].

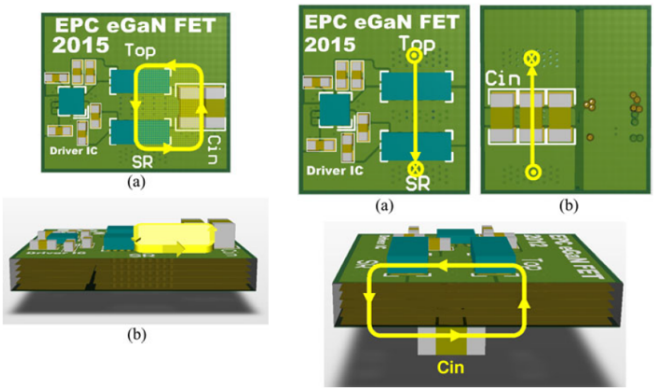


Fig. 2. Left:Lateral structure (a) Top view (b) Side view ; Right: Vertical structure (a) Top view(b) Side view ,adapted from [10]

Conventional way is lateral layout as shown in fig 2(Left) , where DC link capacitor , all the switching devices are placed on one side of the PCB by keeping them in close proximity to reduce the physical size of power loop.Another most common way is vertical layout as shown in fig 2(Right), where all the switching devices are placed on one side of the PCB and DC link capacitor is placed directly opposite side of the PCB underneath the main switching devices. Another way will be keeping One switch on top side and another switch on the bottom side of the PCB. Any how the target is always to reduce physical loop size.One optimized method would be to place all the power loop components on one side of the PCB as near as possible without any of the shield layer,as shown in Fig 3. Here, via is used to make the current return path from the exact next layer of the top layer, which helps in the reduction of loop size.

As described above, all state-of-the-art layouts have all the components in close proximity to reduce the physical size of the power loop. In addition, employ the following techniques to decrease power loop inductance.

A. Field Cancellation

Current flowing through a conductor or PCB trace will generate a magnetic field. Due to that magnetic field, another

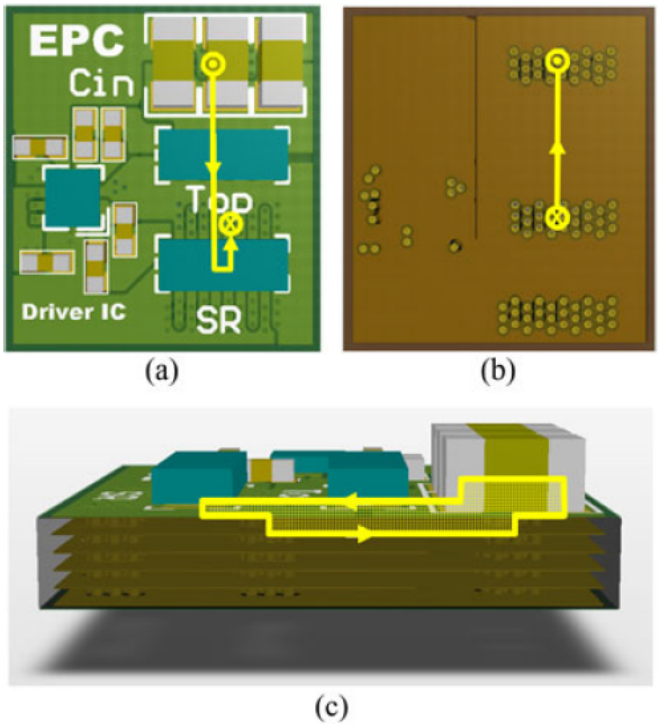


Fig. 3. Optimized layout (a) Top view (b) Bottom view (c) Side view ,adapted from [10].

current will be induced in a near-by conductive medium, which is in the opposite direction to the original current. This induced current also now generates a magnetic field, which nullifies the very cause of it. This is the basic principle of the field cancellation technique to reduce parasitic inductance.

As shown in Fig.2, the vertical structure has a current return path that is directly beneath the power loop current. So self-field cancellation is used to reduce the parasitic inductance. In Fig.3 the optimized structure also uses this technique to minimize the parasitic inductance. In this technique, both the distance between PCB layers and the thickness of the PCB have a significant impact on parasitic. However, the impact totally depends on the way the power loop is designed. In Fig.2, the vertical structure has more impact due to PCB thickness than the distance between PCB layers. But in Fig.3, the optimized structure has more impact on parasitics due to distance between PCB layers than overall PCB thickness [10].

B. Shielding Layer

The shield layer is an important layer to reduce the field generated by the high frequency power loop. The field cancellation technique's basic operating principle also applies to this. Inside the shield layer, a current is induced due to the magnetic field of the power loop. This induced current in the shield layer is in the opposite direction to the power loop current and it also generates the magnetic field. This magnetic field counteracts the power loop magnetic field, which helps to reduce the parasitic inductance. Fig.2, The Lateral Structure: the first inner layer of the PCB acts as a shield layer to reduce parasitics. The shield layer is used when self-field cancellation is not possible due to the layout of the power loop as shown in fig. 2 Lateral Structure . In this technique, the distance between the power loop and the shield layer(first inner layer) has a significant impact on the parasitics of the power loop [10].Also, some shield layer losses will add up. That is one disadvantage of using the shield layer.

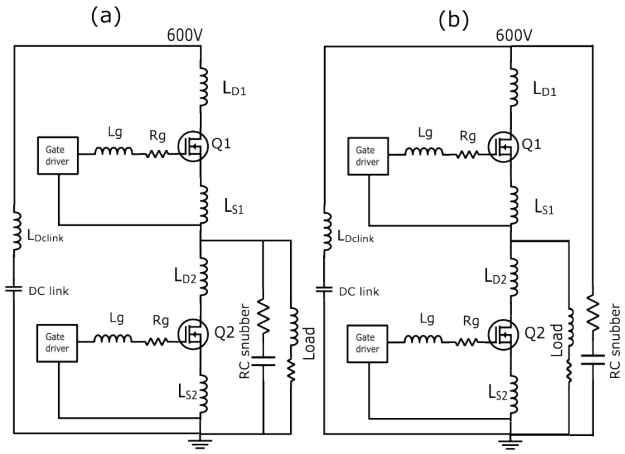


Fig. 4. RC snubber circuit

C. Snubber Circuit

An RC snubber circuit can be used to reduce the ringing and overvoltage during the switching process. Keep the RC circuit parallel to the respective switching device as shown in the Fig.4 (a) [13].Arranging the RC circuit in parallel to the switching device helps to reduce voltage overshoot and damp the ringing due to parasitic inductance, as shown in the Fig.4 (a) . But the disadvantage of this circuit is that the fast changing voltage across the switch will increase the losses of the RC snubber circuit [14].So instead of connecting them across the switch, we can connect the RC snubber circuit across the DC link capacitor as shown in Fig.4 (b) . This will damp the ringing at lower snubber circuit loss.

D. Interleaved Structure

The advanced package of GaN transistors mostly uses an interleaved structure. In this structure, more than one current path is connected in parallel to each other. And this can be achieved by having multiple drain and source contacts on a single transistor, as shown in Fig. 5.These parallel current paths are loosely coupled so that mutual inductance between them is almost zero. All parallel paths have the same current value, so overall loop inductance can be described as per equation3 [8] . where L is the total inductance of the loop, L_o is the inductance of the single current path, and N is the number of parallel current paths.

$$L = \frac{L_o}{N} \tag{3}$$

The multiple number of parallel current paths decreses the loop inductance .Fig 5 shows the advanced package GaN transistors are used to design a half bridge using the interleaved structure to minimize the parasitic inductance.

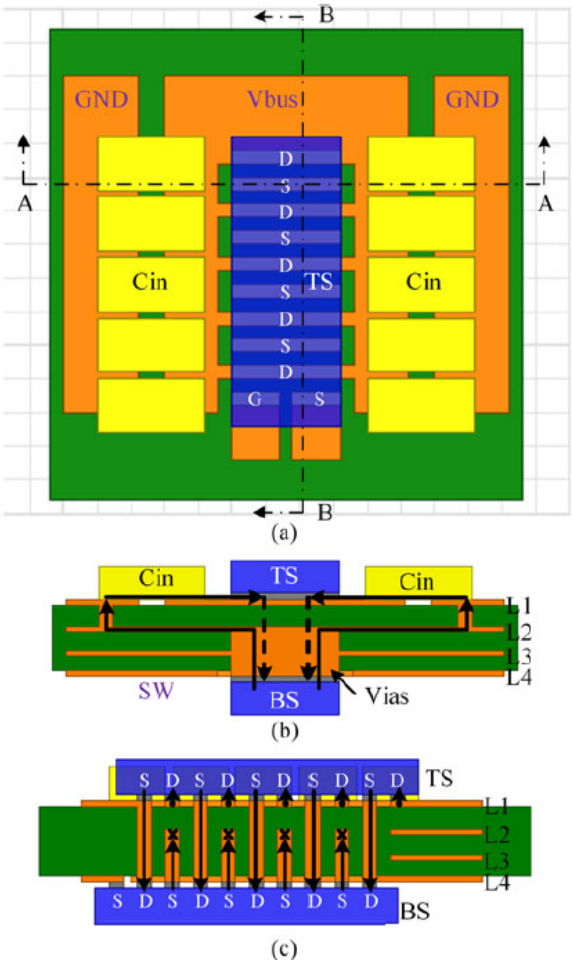


Fig. 5. Double sided layout TS:High side GaN BS:Low side GaN (a)Top view (b)side view (c)A-A cross section, adapted from [1].

Above all three explained layouts ,the optimized layout is more effective and also not very complex. Shield layer requirements are reduced as a result of structure design, but thermal dissipation must also be considered during the design process. DC link capacitor placement helps to reduce inductance.Using small ESL(Equivalent series inductance) capacitors placed in parallel in place of single dc link capacitor also help us to reduce the parasitics due to dc link package as shown in the Fig.5 [12].Also connection between all the component and traces can be done in a such way that field cancellation due to return current path is possible which leads to reduce parasitic inductance too.This is only possible if the structure is vertical as per Fig.2 (Right) and Fig.3. In other cases use of shied layer is beneficial for field cancellation but it add the losses from the shield layer.

TABLE I
FACTORS AFFECTING THE DIFFERENT PCB LAYOUTS

Factor Affecting the Parasitics	Lateral Structure	Vertical Structure	Optimized Structure
DC link capacitor placement on PCB	Top	Bottom	Top
Distance between PCB Layers	Less	Less	More
Thickness of PCB	More	Less	Very less
Shield layer requirement	Yes	No	No

Other than that, the ultra-low parasitic power module is the latest solution for the minimization of power loop inductance. where the DC link capacitor is also manufactured on a monolithic GaN half-bridge IC . Due to this, the connection of the DC link to the switch and the connection between the switches will have the least parasitic component. In this technology, they use a bus structure for DC link connection to switch. This is based on two phenomena: one, we have a wide conductor structure, which creates a long path for magnetic flux to close around it. The result is that magnetic resistance increases and inductance decreases. Second, the return path of the current is brought close in order to cancel the magnetic field,So we're employing the field cancellation technique [15].

III. CONCLUSIONS

High-performance GaN FETs have the ability to switch with greater efficiency and at higher frequencies than conventional Si MOSFET technology. GaN FETs need a low parasitic PCB layout in addition to better figure of merits and low parasitic packaging to operate to their full potential. State-of-the-art layouts and techniques to reduce power loop PCB parasitics have been discussed. From that, an optimized layout is more efficient to reduce the physical size of the power loop without a shielding layer. Moreover,the use of an advanced GaN transistor is a better solution for the lowest package parasitic inductance. Layout optimization is also improving as a result of the LGA interleaved structure. Field cancellation and a shield layer can be used to cut down on parasitics brought on by the magnetic field of a power loop current. To reduce DC link capacitor internal parasitic inductance, parallel connection of multiple small-size capacitors can be used, but only up to some voltage level. Along with thick and short PCB traces, the snubber circuit helps to reduce over-voltage spikes by filtering out the high frequencies.Use of the ultra-low parasitic power module is also recommended.Furthermore, gate loop inductance needs to be removed from the half bridge to get smoother and faster lossless switching of GaN power converters.

REFERENCES

[1] A. Ball, M. Lim, D. Gilham and F. C. Lee, "System design of a 3D integrated non-isolated Point Of Load converter," 2008 Twenty-Third Annual IEEE Applied Power Electronics Conference and Exposition, 2008, pp. 181-186, doi: 10.1109/APEC.2008.4522719.

[2] Y. Xiao, H. Shah, T. P. Chow and R. J. Gutmann, "Analytical modeling and experimental evaluation of interconnect parasitic inductance on MOSFET switching characteristics," Nineteenth Annual IEEE Applied Power Electronics Conference and Exposition, 2004. APEC '04., 2004, pp. 516-521 Vol.1, doi: 10.1109/APEC.2004.1295856.

[3] M. Shiraishi, T. Iwasaki, N. Akiyama, T. Kawashima, N. Matsuura, and S. Chiba, "Low loss and small SiP for DC-DC converters," in Proc. 17thM. Shiraishi, T. Iwasaki, N. Akiyama, T. Kawashima, N. Matsuura and S. Chiba, "Low loss and small SiP for DC-DC converters," Proceedings. ISPSD '05. The 17th International Symposium on Power Semiconductor Devices and ICs, 2005., 2005, pp. 175-178, doi: 10.1109/ISPSD.2005.1487979. Int. Symp. Power Semicond. Devices IC's, 2005, pp. 175–178.

[4] K.W. Kam, D. Pommerenke, L. Cheung-Wei, andR. Steinfeld, "EMI analysis methods for synchronous buck converter EMI root cause analysis," in Proc. IEEE Int. Symp. Electromagn. Compat., 2008, pp. 1–7.

[5] M. Pavier, A. Sawle, A. Woodworth, R. Monteiro, J. Chiu, and C. Blake, "High frequency DC:DC power conversion: The influence of package parasitics," in Proc. 18th Annu. IEEE Appl. Power Electron. Conf. Expo., 2003, pp. 699–704.

[6] J. Wang, H. S. Chung, and R. T. Li, "Characterization and experimental assessment of the effects of parasitic elements on the MOSFET switching performance," IEEE Trans. Power Electron., vol. 28, no. 1, pp. 573–590, Jan. 2013.

[7] K. Wang, X. Yang, H. Li, H. Ma, X. Zeng, and W. Chen, "An analytical switching process model of low-voltage eGaN HEMTs for loss calculation," IEEE Trans. Power Electron., vol. 31, no. 1, pp. 635–647, Jan. 2016.

[8] K. Wang, L. Wang, X. Yang, X. Zeng, W. Chen and H. Li, "A Multiloop Method for Minimization of Parasitic Inductance in GaN-Based High-Frequency DC–DC Converter," in IEEE Transactions on Power Electronics, vol. 32, no. 6, pp. 4728-4740, June 2017, doi: 10.1109/TPEL.2016.2597183.

[9] J. Noppakunkajorn, D. Han and B. Sarlioglu, "Analysis of High-Speed PCB With SiC Devices by Investigating Turn-Off Overvoltage and Interconnection Inductance Influence," in IEEE Transactions on Transportation Electrification, vol. 1, no. 2, pp. 118-125, Aug. 2015, doi: 10.1109/TTE.2015.2426503.

[10] D. Reusch and J. Strydom, "Understanding the Effect of PCB Layout on Circuit Performance in a High-Frequency Gallium-Nitride-Based Point of Load Converter," in IEEE Transactions on Power Electronics, vol. 29, no. 4, pp. 2008-2015, April 2014, doi: 10.1109/TPEL.2013.2266103.

[11] D. Reusch, F. C. Lee, D. Gilham and Y. Su, "Optimization of a high density gallium nitride based non-isolated point of load module," 2012 IEEE Energy Conversion Congress and Exposition (ECCE), 2012, pp. 2914-2920, doi: 10.1109/ECCE.2012.6342367.

[12] T. Ibuchi and T. Funaki, "A study on parasitic inductance reduction design in GaN-based power converter for high-frequency switching operation," 2017 International Symposium on Electromagnetic Compatibility - EMC EUROPE, 2017, pp. 1-5, doi: 10.1109/EMCEurope.2017.8094824.

[13] Zaman H, Wu X, Zheng X, Khan S, Ali H. Suppression of Switching Crosstalk and Voltage Oscillations in a SiC MOSFET Based Half-Bridge Converter. Energies. 2018; 11(11):3111. https://doi.org/10.3390/en11113111

[14] I. Josifović, J. Popović-Gerber and J. A. Ferreira, "Improving SiC JFET Switching Behavior Under Influence of Circuit Parasitics," in IEEE Transactions on Power Electronics, vol. 27, no. 8, pp. 3843-3854, Aug. 2012, doi: 10.1109/TPEL.2012.2185951.

[15] E. Hoene, A. Ostmann and C. Marczok, "Packaging Very Fast Switching Semiconductors," CIPS 2014; 8th International Conference on Integrated Power Electronics Systems, 2014, pp. 1-7.

Gallium-Nitride based FinFETs

Electrical Engineering
Universität Stuttgart
Stuttgart, Germany

***Abstract*— Gallium nitride (GaN) is becoming a mainstream semiconductor for power and radio-frequency. In the last few years, the fin field-effect transistor (FinFET) and m-gate architectures have been leveraged to develop a new generation of GaN power and RF devices, which have continuously advanced the state-of-the-art in the area of microwave and power electronics. This review paper presents an overview of the reported GaN FinFET device technologies for RF and power use cases, as well as includes details about device design parameters and performance analysis. The paper concludes with a summary of current challenges and outlook on the FinFET technologies.**

Keywords—Break Down Voltage (BV), SiC, GaN, AlGaN, Si, CMOS, Drain Induced-Barrier Lowering (DIBL), FET, Double Gate (DG).

I. INTRODUCTION

The Si-based power systems are fast approaching their theoretical limit. But we need transistors that can handle high voltage (high current) for next generation applications. We should note that the global power device market that is expected to reach $32.35 billion by 2027 is growing fast [1]**.** So, to meet the expectation, science community and the industry start to look at the alternative ways to achieve further improvement in device performance. One possible alternative is to use other semiconducting materials. Because of the material advantages like higher thermal conductivity, higher breakdown voltage, larger energy bandgap, comparable carrier mobility, and saturation electron velocity, wide bandgap compound semiconductors are the most attractive alternative [Table 1][2]**.** We have a term "specific ON-resistance" expressed as:

$$R_{on,sp} = R_{on} \cdot A \quad (1)$$

Where R_{on} is ON-resistance and A is the device area. Our goal is to achieve a low specific ON-resistance which then means smaller area (capacitances) and small switching losses. We observe that power devices based on wide bandgap materials can potentially achieve lower specific ON-resistance for the same BV. Therefore, compared to similarly rated Si devices, we get higher operating frequency. FETs based on GaN materials are now strong candidate for applications in high-power, high-frequency electronics.

	Si	*4H-SiC*	*GaN*	*Al(Ga)N*	*Diamond*
Eg [eV]	1.1	3.3	3.4	5-6	5.5
Ec [MV/cm]	0.3	2.5	3.4	8-12	10
n-type doping	Yes	Yes	Yes	Difficult	Difficult
p-type doping	Yes	Yes	Yes	Difficult	Yes
Electron/Hole Mobility [cm^2/Vs]	1300	800	1200	300	2000
Thermal Conductivity [W/cmK]	1.3	4.3	1.5-2.5	2.5-3.2	10-30
Baliga's FOM	1	340	1450	10067	24661

Table 1: Physical Properties of major Semiconductors

The other possible alternative which has been widely used to get better performance is to scale (reducing the lateral and vertical dimensions of) the transistors. But then, to avoid the oxide breakdown and to maintain the device performance, the V_{dd} and V_t need to be scaled down as well. The combination of narrow oxide thickness and low V_t gives rise to gate leakage and sub-threshold leakage currents, respectively. These effects (and few others like, DIBL) in short are known as short channel effects (SCEs). So, there is a need to find replacement for classical structure at nanoscale level. One solution is to introduce Multi-Gate MOSFETs. A type of MG-MOSFETs which has 2 gates are called as double-gate MOSFET or FinFET. The term 'Fin' refers to the fin-shaped channel (size is in the regime of nanometer). FinFET provides better gate control of channel and therefore significantly reduces the SCEs. To enable scaling for technology nodes below 22nm, FinFETs are widely used [14]. The FinFET technology has been widely adopted in digital and memory applications leading to an expected global market of around $196.85 billion by the end of 2026 [11]**.**

In this paper we discuss and summarize how we can combine GaN and FinFET layout. The paper is organized as below. In Section 2 we discuss the basics of Si based FinFET, Section 3 discusses the how the FinFET concept can be applied to GaN transistors, their advantages and disadvantages. Section 4 deals with the outlook for future studies and conclusion.

II. FinFET based on Si

In 1989 Hisamato et al. fabricated a double-gate SOI structure which they called a fully-depleted lean channel transistor (DELTA) is known to be the first reported fabrication of a FinFET-like structure [3]**.** As we discussed before when we scale (like, reduce the channel length) an FET, we observe the SCEs. To prevent one such short channel effect, we can use the thin gate oxide or the thin depletion depth below the channel to

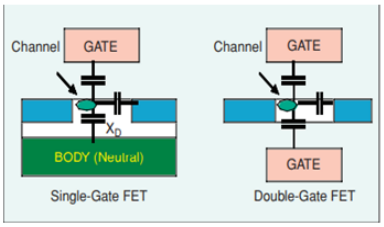


Fig. 1: The double-gate device architecture escapes the intrinsic compromise presented by conventional FETs [4]

the substrate. But reducing both/either of these after a point is not wise and it will result in unreasonable power increase and/or slower turn on of the channel region respectively. But in double-gate FETs, the presence of second gate results in reducing short channel effects, particularly reducing the DIBL and improving the subthreshold swing. Figure 1 schematically illustrates the above points and shows the advantage of double gate layout[4].

One point to note here is, in single-gate design we can use some method (like, increased body doping) to decrease the DIBL, but, after some point it will increase the subthreshold swing and if we chose to improve the subthreshold swing (e.g., decreased body doping) this will worsen the DIBL. So, a compromise

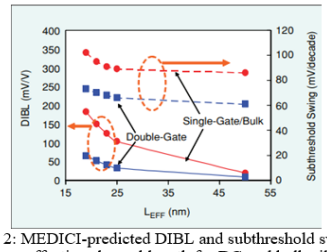


Fig. 2: MEDICI-predicted DIBL and subthreshold swing versus effective channel length for DG and bulk-silicon nFETs. The DG device is designed with an undoped body and a near-mid-gap gate material [4]

between the two is necessary. Figure 2 illustrates the above discussed points. In the graph we draw the MEDICI-predicted DIBL and subthreshold swing values versus the effective channel length. Here, we can see that with the double-gate FETs both the parameters can be improved. This shows the difference between the single-gate and double-gate layout in a greater detail.

Next, let's look at few of the different types of FinFETs. We look at two main types, namely, Shorted-Gate FinFET and Independent-Gate FinFET. As the name suggests in SG-FinFETs the front and back gates are shorted and in IG-FinFETs they are isolated from each other. In Figure 3 we can see the

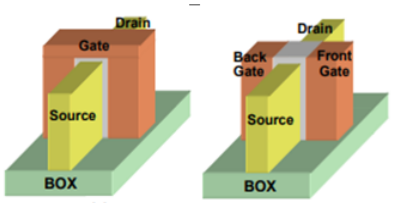


Fig. 3: FinFET device layouts. Shorted-gate FinFET (lhs). Independent-gate FinFET (rhs) [6]

layout structure of both SG and IG-FinFETs. Few of the simple differences between the two are, In SG-FinFET there is no external control of the V_{th} and channel is controlled jointly by both the gates. But, in IG-FinFET one gate is used for switching and the other to control V_{th}. This increases the flexibility. Relation between front gate threshold voltage (V_{tG1}) to the back gate voltage (V_{G2}) using the charge sheet approximation [7] is:

$$\frac{V_{tG1}}{V_{G2}} = -\frac{C_{Fin}C_{ox2}}{C_{ox1}(C_{Fin}+C_{ox2})} \cong -\frac{3t_{ox1}}{3t_{ox2}+Fin_{WIDTH}} \quad (2)$$

Each design has few advantages and disadvantages. One main point to note is that, IG-FinFET requires more area [6].

III. GaN based FinFETs

In this section let's look at how GaN transistor will take advantage of the FinFET design which we discussed in the section before. For e.g., in the case of RF applications, the suppression of SCEs can potentially lead to a high unity current-gate frequency (f_T) [9]. Also, we can demonstrate superior DC and RF performance of AlGaN/GaN FinFETs featuring high extrinsic G_m linearity as shown here [13] or we can achieve high performance with GaN/AlGaN FinFET Heterostructure design as shown here [12]. In this paper, we will discuss switching performance and figure of merit of 1.2kV vertical GaN Power FinFETs[8]**.**

A. Device Design and Capacitance

Note that, FinFETs can be realized in either lateral or vertical layout. Typically, GaN devices have a lateral layout. But, recently the vertical layout FinFETs have been designed. Since, the vertical structure can be used in medium voltage range, we will be using these layout structures. First, let's look at the capacitances that can be observed in our design structure. We can divide the capacitances into two groups. One, capacitances in active region and the other is capacitances in pad/edge region.

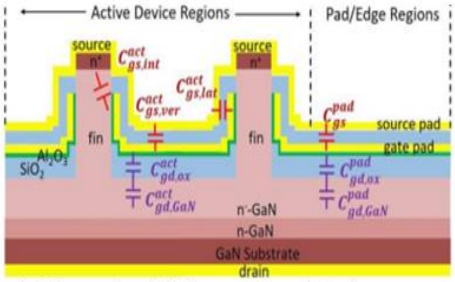


Fig. 5: Schematics of different Cgs and Cgdcomponents in the active and pad/edge regions of a GaN vertical FinFET [8]

Figure 5 illustrates the vertical layout with main components of device junction capacitances in both of these regions. All the capacitances (includes active and pad/edge region) shown in the figure belong to two main capacitances, that is, C_{gs} and C_{gd}. Note that, C_{ds} is very small because the fin region is fully depleted at zero gate bias. We can optimize the device layout to reduce the capacitances in pad/edge region. Now, to calculate the switching performance and FOM, we need input capacitance C_{iss} (= C_{gs} + C_{gd}), output capacitance C_{oss} (= C_{gd} + C_{ds}), and reverse capacitance C_{rss} (= C_{gd}). Here we observe that compared to Si technology, the capacitances are low in our

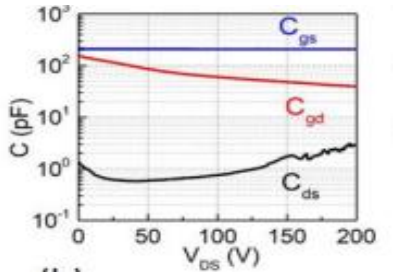


Fig. 6: Simulated junction capacitance as a function of drain voltage [8]

device(see Table 2). Figure 6 shows the simulated junction capacitances as a function of voltage V_{ds}. Here, we can see that C_{gd} and C_{gs} are the main sources of the device capacitance and C_{ds} is very less compared to other capacitances.

B. Performance comparison

Let's look at the switching performance and figure of merit of our device and compare that with the similar rated devices with other technologies. Dimensioning of components and detailed calculations on how to find the switching frequency and figure of merit are presented in the original paper in a great detail. Here, I will briefly discuss the procedure and give the equations, but mainly will try to summarize the result and draw some conclusion with the results that we have.

First, let us look at how switching frequency and figure of merit can be expressed and calculated. In general, device power loss under switching operations consists of the conduction losses, switching losses and the losses related to device/diode reverse conduction:

$$P = P_{cond} + \left(P_{gate} + P_{qoss}\right) + P_{rr} \quad (3a)$$

$$= I^2 R_{on} D + Q_G V_G f_{sw} + \frac{Q_{gd} V_{ds} f_{sw} I_{ds}}{I_G} + P_{rr} \quad (3b)$$

$$= I^2 R_{on} D + Q_G V_G f_{sw} + Q_{oss} V_{ds} f_{sw} + P_{rr} \quad (3c)$$

where D is the duty cycle and fsw is the switching frequency. Also, here equation 3b is in the case of hard switching and the equation 3c is in the case of soft switching. Equation 3a, 3b and 3c are used to calculate the switching frequency.

To evaluate the performance, we use the following power switching FOM (out of many FOMs, considering all possible conduction and switching losses):

$$FOM_{sw} = R_{on}(Q_G + Q_{gd} + Q_{rr}) \quad (4)$$

This FOM is independent of device area or current ratings, and is suitable for evaluating these devices for both hard and soft switching. Now if we substitute values for $Q_{G,}$ Q_{gd} and Q_{rr}:

$$FOM_{sw}^{FinFETs} = R_{on}(C_{iss} V_G + q N_D t_D A) \quad (5)$$

Now, if we use the below values for V_G (5V); N_D (~4×1015 cm^{-3}) and t_D (8.5 µm) is the net donor concentration and thickness of the drift region between the gate and drain, respectively; and A is the total device area (~0.45mm^2), we get the figure of merit to be equal to 3.3nΩ.C and a practical f_{sw} of ~3.5 MHz is calculated.

Now that we have calculated the switching frequency and FOM in our case, let's look at where does this result stand when compared to the similar rated devices from other technologies. Table 2 summarizes the key device metrics of our 1.2 kV GaN vertical FinFETs, benchmarked against state-of-the-art commercial 0.9-1.2 kV Si and SiC power transistors. Other large-area 0.9-1.2 kV GaN devices at the R&D level are also included. Here, we will test the different properties and parameters like, the chip area, R_{on}, BV, V_{th}, capacitances,

Device Technology	Maker	$R_{on,sp}$ (mΩcm^2)	BV (kV)	I_{rate} (A)	Chip Area (mm^2)	V_{th} (V)	C_{iss} (pF)	C_{oss} (pF)	Q_G (nC)	Q_{GD} (nC)	Q_{rr} (nC)	Switching FOM (nΩ·C)
GaN vertical FinFET[8]	MIT	2.1	1.2	5	0.45	1.3	248.3	42.2	1.24	2.42	~0	3.3
SiC MOSFET(CPM2-1200-0160B)	Cree	2.7[17]	1.2	20	6.28	2.5	525	47	34	14	105	7.68
Si IGBT (NGTB15N120FLWG)	ON semi	~20	1.2	15	320	2	3600	110	150	68	1500	54.5
GaN HEMT Cascode [15]	Transphorm	R_{on} = 0.19Ω	1.2	>20	-	2.3	-	43.1	10	-	-	-
Vertical GaN CAVET [16]	Avogy	2.2	1.5	2.3	0.17	0.5						

Table 2: Summary and benchmark of device technologies and their key device metrics for 900-1200 V power switching applications

switching charges and FOMs. As shown, the above discussed device exhibited the best power switching FOMs among all 0.9-1.2 kV power transistors. It's very difficult to get the values for R&D level devices, and we see that SiC is somewhat closer to our device, but Si based device is well far of. Also, note the difference in chip area. As we discussed, low specific ON-resistance means small area, the same can be seen here. All these improvements would mean nothing if the breakdown voltage is bad. But as we discussed before, we see that these improvements can be achieved with similar breakdown voltage compared to the other technologies. These improvements bring the threshold voltage down and the device capacitances compared to the other technology are very low in our device. Finally, the switching frequency that we obtained (~3.5 MHz) is much higher than the equivalent Si IGBTs (10-20 kHZ).

C. Challenges and Improvements

The design discussed here has its own shortcomings and few improvements can be done both in the manufacturing and design point of view. The original paper explains these in a great detail. Here, I would like to discuss more general points. A common gap for many GaN FinFETs and trigate HEMTs (irrespective of

the layout design) is their thermal management, GaN material has a thermal conductivity of 1.3W/cmK whereas Si has a thermal conductivity of 1.5W/cmK. Manufacturing process is another short coming of GaN technology when it's compared to that of Si technology. The number of defects per square centimetre is very high compared to that of Si. This in turn means that the cost to produce same amount of GaN wafers is very high compared to Si. Next, regarding vertical GaN FinFETs, an immediate need for advancing the performance of vertical GaN Fin-MOSFETs is to understand the physics, material science, and processing technologies regarding the electronic transport properties in the fin channel and the sidewall accumulation-mode MOS channel. To overcome few of the shortcomings we can look into alternatives. GaN-on-Si could offer an alternative to lower the device costs, but significant progress has to be made on the GaN quality with low defect density and on the thickness of the GaN layers grown on Si substrates, in order to enable larger voltage devices.

IV. Conclusion

In summary, we reviewed the advantages of GaN semiconductor material and the technology and characteristics of GaN based transistors. FinFET and trigate architectures based on GaN can play a vital role in improving the state-of-the-art in the area of RF technologies and power electronics. For e.g., the device mentioned in this work shows superior power switching FOMs when compared to commercial 0.9-1.2 kV Si and SiC power transistors. This demonstrates the great potential of vertical GaN FinFETs. Extensive research both at academic and industrial level will help achieve these advancements. As a final thought, we should note that GaN as a technology is still very young. So, the future is very promising for GaN technology and the end users like us.

References

[1] (Jun. 2022). Global Power Semiconductors Market, by Material, by Application, by Component-Forecast 2027-Market Research Insitute. WebLink:https://www.marketresearchfuture.com/reports/power-semiconductor-market-1178.

[2] Chow, T.P., and Tyagi, R.: '*Wide bandgap compound semiconductors for superior high-voltage unipolar power devices'*, Trans. Electron. Devices, 1994, 41, (8), pp. 1481–1483, doi:10.1109/16.297751

[3] D. Hisamoto, T. Kaga, Y. Kawamoto, and E. Takeda, "*A fully depleted lean-channel transistor (DELTA)—a novel vertical ultra thin SOI MOSFET*," in Proceedings of the International Electron Devices Meeting (IEDM '89), pp. 833–836, Washington, DC, USA, December 1989.

[4] E. Nowak, T. Ludwig, I. Aller, J. Kedzierski, M. Ieong, B. Rainey, M. Breitwisch, V. Gernhoefer, J. Keinert, D.M. Fried, "Scaling beyond the 65 nm node with FinFETDGCMOS," in Proc. IEEE CICC, San Jose, CA, pp. 339–342.

[5] A. N. Bhoj and N. K. Jha, "*Design of logic gates and flip-flops in high-performance FinFET technology*," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 21, no. 11, pp. 1975–1988, 2013.

[6] Meng-Chou Chang, Kai-Lun He and Yu-Chieh Wang, "*Design of asymmetric TCAM (ternary content-addressable memory) cells using FinFET*" 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), 2014, pp. 358-359, doi: 10.1109/GCCE.2014.7031172.

[7] Lim H, Fossum JG (1983) Threshold voltage of thin-film siliconon-insulator MOSFETs. IEEE Trans Electron Devices 30–10:1244.

[8] Zhang, Yuhao et al. "Large Area 1.2 kV GaN Vertical Power FinFETs with a Record Switching Figure-of-Merit." IEEE Electron Device Letters 40, 1 (January 2019): 75 - 78 © 2018 IEEE.

[9] Jessen G H, Fitch R C, Gillespie J K, Via G, Crespo A, Langley D, Denninghoff D J, Trejo M and Heller E R 2007 Short-channel effect limitations on high-frequency operation of AlGaN/GaN HEMTs for T-gate devices IEEE Trans. Electron Devices 54 2589–97.

[10] Y. Zhang et al., "Electrothermal simulation and thermal performance study of GaN vertical and lateral power transistors," IEEE Trans.Electron Devices, vol. 60, no. 7, pp. 2224–2230, Jul. 2013, doi:10.1109/TED.2013.2261072.

[11] (Jun. 2022). "FinFET Technology Market Analysis" by Market Data Forecast 2026. WebLink: https://www.marketdataforecast.com/market-reports/finfet-technology-market.

[12] Im, Ki-Sik, et al. "High-performance GaN-based nanochannel FinFETs with/without AlGaN/GaN heterostructure." IEEE Transactions on Electron Devices 60.10 (2013): 3012-3018.

[13] Zhang, Kai, et al. "High-linearity AlGaN/GaN FinFETs for microwave power applications." IEEE Electron Device Letters 38.5 (2017): 615-618.

[14] Bhattacharya D and Jha N K 2014 FinFETs: from devices to architectures Adv. Electron. 2014 1–21.

[15] T. Oka, T. Ina, Y. Ueno, and J. Nishii, "Over 10 A operation with switching characteristics of 1.2 kV-class vertical GaN trench MOSFETs on a bulk GaN substrate," in 2016 28th International Symposium on Power Semiconductor Devices and ICs (ISPSD), 2016, pp. 459–462. DOI: 10.1109/ISPSD.2016.75208.

[16] [16] H. Nie, Q. Diduck, B. Alvarez, A. P. Edwards, B. M. Kayes, M. Zhang, G. Ye, T. Prunty, D. Bour, and I. C. Kizilyalli, "1.5-kV and 2.2-mΩ·cm2 Vertical GaN Transistors on Bulk-GaN Substrates," IEEE Electron Device Lett., vol. 35, no. 9, pp. 939–941, Sep. 2014. DOI: 10.1109/LED.2014.2339197.

[17] J. W. Palmour, L. Cheng, V. Pala, E. V. Brunt, D. J. Lichtenwalner, G.-Y. Wang, J. Richmond, M. O'Loughlin, S. Ryu, S. T. Allen, A. A. Burk, and C. Scozzie, "Silicon carbide power MOSFETs: Breakthrough performance from 900 V up to 15 kV," in 2014 IEEE 26th International Symposium on Power Semiconductor Devices & IC's (ISPSD), Waikoloa, HI, USA, 2014, pp. 79–82. DOI: 10.1109/ISPSD.2014.6855980.

Wide-Bandgap Superjunction Transistors

Electrical Engineering
Universitat Stuttgart
Stuttgart, Germany

Abstract— An analysis of wide-bandgap superjunction transistors is conducted to emphasize the significance and benefits of such devices in high voltage power electronics applications. This paper outlines the characteristics of wide-bandgap superjunction semiconductor materials Silicon Carbide (SiC) and Gallium Nitride (GaN), their fabrication and design process, and the challenges associated with such devices. It further explains how different parameters of the device affect the properties of wide-bandgap superjunction power devices.

Keywords—Superjunction, wide-bandgap, SiC superjunction, GaN superjunction.

I. Introduction

Silicon (Si) based power devices have been prevalent in power electronic applications for decades, and to improve and push the theoretical limit of unipolar Si power devices, the superjunction concept was introduced. The superjunction structure incorporates multiple highly doped regions in the drift region instead of using a single lowly doped region [7]. By vertically stacking pillars, the depletion layer is extended to the entire drift region. This allows a higher doping concentration in the drift region, which reduces the drift resistance [4] further reducing the on-resistance for a device in the same voltage class.

Wide-bandgap semiconductors have a high critical electric field allowing them to have a smaller area-specific on-resistance ($R_{ON,SP}$) than conventional silicon power devices at the same voltage class. Furthermore, they have high thermal conductivity, allowing for high temperature operation and a correspondingly high saturation velocity that allows for a high level of current capability in the same package. Table I gives a comparison of important semiconductor properties.

Property	Si	4H-SiC	GaN
Band gap [eV]	1.12	3.23	3.42
Electric Breakdown [10^5 V/cm]	3	28	33
Thermal Conductivity [W/cmK]	1.5	4.9	2/0
Saturation Velocity [10^7 cm/s]	1	2.2	2.7

Table I Comparison of semiconductor properties

The SJ structure that benefits silicon could be used for wide-bandgap material's unipolar devices to attain superior performance with a similar approach. The superjunction concept can be applied in various methods like using vertical variation doping profile in SiC SJ MOSFET [2], vertical SJ MOSFET with a different number of n-type and p-type pillars [3] by multi-epitaxial growth, fabrication of vertical GaN SJ using lateral polar junction [8] and more. Section II and III gives an analysis on fabrication, design, and electrical characteristics of SiC and GaN superjunction devices respectively to the author's best knowledge.

II. Silicon Carbide Superjunction

A. Fabrication and Design

A multi-epitaxial growth method is used to achieve a 1.2 kV-class SJ UMOSFET [1][4] and a 3.3 kV-class SiC SJ MOSFET [9]. The thickness of each layer and the number of repetitions used for the device can vary depending on the requirements although a common thickness of 0.65 µm in epitaxial film was found in 1.2 kV-class devices [1][4]. A donator and acceptor concentration of $3*10^{16}$ cm^{-3} and $6*10^{16}$ cm^{-3} respectively, and a p-pillar depth of 5.2 µm is achieved [1]. A comparison of normal and narrow width pillars is given where the characteristics of the device are altered by changing the width. Fig. 1 depicts the schematics used in [1] and [4].

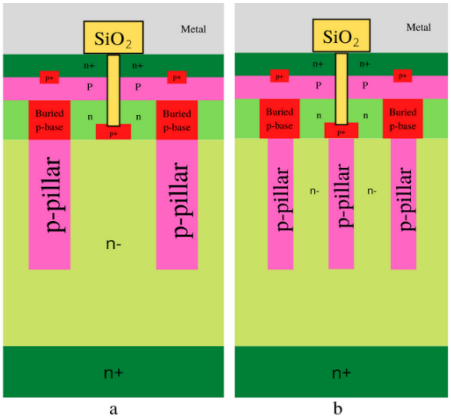


Fig. 1 Schematic diagrams of super junction SiC devices. (a) Normal width (b) Narrow width (adapted from [1]and [4])

With the epitaxial growth of 3C-SiC on Si, a device can be fabricated that has both an n-doped Silicon pillar and a p-doped 3C-SiC pillar, which is uncommon in a conventional SJ MOSFET. However, building the p-doped 3C-SiC pillar

requires a separate challenging chemical deposition process at low pressures[3]. A variation vertical doping (VVD) profile with epitaxial growth is an alternative method of forming SJ structures [2]. In this design, instead of using a single highly doped n-pillar, it is partitioned into three columns with increasing doping concentration as they approach the substrate layer. When the p-pillar is divided into smaller parts, however, the doping concentration decreases with increasing proximity to the substrate layer. Apart from epitaxial growth, a method of trench etching with sidewall implantation can be implemented [5]. The method has been applied to devise a junction field effect transistor (JFET). The manufacturing flow for this device is given in Fig. 2. The cost of manufacturing was found to be reduced when the multi-epitaxial growth is not implemented.

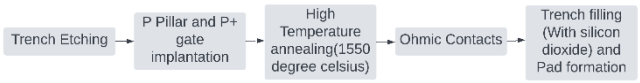


Fig. 2 Fabrication process of JFET using trench etching with sidewall implantation (adapted from [5])

The electrical characteristics (static and dynamic) of the device are affected due to changes made to the geometry, doping concentration, and thickness of the pillars/layers inside the device. To observe these effects, simulations and experiments are conducted.

B. Static Characteristics

The trade-off between $R_{ON,SP}$, and BV is very important for any power device, especially reducing on-resistance while maintaining the voltage class and size is desirable. Device characteristics of the non-SJ structure, normal pillar width SJ structure, narrow pillar width SJ structure, SJ VVD structure, and n-Si and p-3C-SiC pillar structure have been discussed in [1-4] and [9]. The normal pillar width SJ structure has shown a decrease of approximately 13% (from 3.1 mΩcm^2 to 2.7 mΩcm^2) in $R_{ON,SP}$ at a threshold voltage of 4.0 V, and 1.2 kV breakdown voltage when compared to a non-SJ structure [1]. Meanwhile when n-Si and p-3C-SiC structures are compared to the conventional SJ structure, the $R_{ON,SP}$ vs BV is distinct when the width *a* is larger which is undesirable since a smaller device with better balanced trade-offs is preferred [3]. The BV was found to be highest when $N_{NET} < 0$, where $N_{NET} = N_D-(N_A-D_I/a)$, D_I is the interface charge density, *a* is the width of the device and N_A and N_D are p/n pillars doping concentration, respectively [3]. The normal pillar width SJ structure shows good results for $R_{ON,SP}$ at 25°C, however, it decreases with increasing temperature while the narrow pillar width SJ structure has a better $R_{ON,SP}$ at 175°C [4]. Nevertheless, the overall characteristics of the normal width SJ structure outperforms the narrow SJ structure. The I-V characteristics can be used to infer the transconductance and subthreshold slope. The SJ VVD shows higher transconductance and lower subthreshold slope at 300 K. This implies lower losses with faster switching [2]. The gate charge (Q_G) is reduced by 10% by using VVD while compared to a planar MOSFET [2]. Table II compares the $R_{ON,SP}$ of SiC SJ MOSFET at room temperature (approximately 25°C).

Device	$R_{ON,SP}$ (mΩcm^2)
Non-SJ Structure at 1.2 kV	3.1
Normal pillar Width SJ Structure at 1.2 kV	2.7
Narrow pillar Width SJ Structure at 1.2 kV	≈2.9
SJ VVD MOSFET at 6.6 kV	6.6
n-Si and p-3C-SiC structure at 600 V	≈4.7
4H-SiC SJ JFET at 1000 V	1.3

Table II $R_{ON,SP}$ of various SiC SJ power device

C. Dynamic characteristics

Switching characteristics mainly concerned with the various SiC SJ power devices have been evaluated. The turn-on and turn-off switching waveforms are similar in non-SJ and SJ devices with normal/narrow pillar widths when the Schottky barrier diode is used as a freewheeling diode. Alternatively, when an intrinsic body diode is used, turn-on losses in SJ are higher than in non-SJ due to a high reverse recovery charge (Q_{RR}) [1][4]. Meanwhile, an increase in output capacitance (C_{OSS}) is observed in the normal width SJ structure and it's higher in the narrow width SJ structure; this is due to the high doping concentration of n-channel and large p-n junction area. The Miller Capacitance or the Reverse Transfer Capacitance (C_{RSS}) is an important factor in switching speed. By reducing these parameters, better transient characteristics can be achieved. The SJ VVD MOSFET shows a turn-off speed of 350 ns and C_{RSS} of 0.734 aF/m^2 compared to 390 ns and 1.853 aF/m^2 of conventional SJ structure respectively, and 765 ns and 2.856 aF/m^2 of non-SJ structure MOSFET respectively [2]. The $R_{ON,SP} * Q_G$ figure of merit was observed to be the least 5813nC-mΩ compared to the 13641 nC-mΩ of a planar MOSFET. Figure 3 demonstrates the turn-off voltage transient of SJ VVD MOSFET in conjunction to conventional SJ and planar MOSFET.

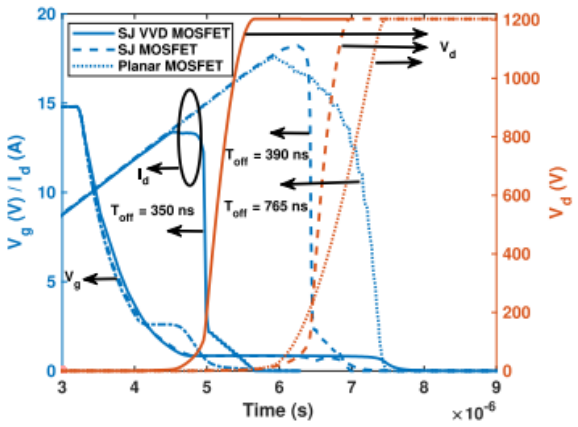


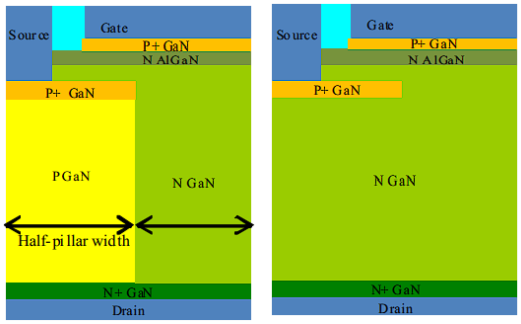
Fig. 3 Turn off switching characteristics [2]

The proposed n-Si and p-3C-SiC structures show improvement in electrical oscillations [3]. The reverse recovery charge is found to be 10% lower than in a conventional SJ. Additionally, the low hole mobility of the 3C-SiC pillar contributes to the lower oscillations. A low leakage current is observed due to a negligible generation current in the p-3C-SiC structure

III. Gallium Nitride Superjunction

A. Fabrication and Design

The common variations in GaN power devices are lateral high electron mobility transistor (HEMT), vertical HEMTs, and MOSFETs. A GaN device without an AlGaN layer is a MOSFET. This paper discusses vertical SJ GaN power devices, with a focus on HEMTs. A selective area growth (SAG) and in situ dopant implantation during crystal growth can be used to fabricate a vertical SJ structure. The device follows the conventional SJ structure by replacing the n-GaN pillar with a combination of p-pillar and n-pillar of the same doping concentration and width [6]. To achieve a positive threshold voltage, a p+ GaN layer is stacked on top of the AlGaN layer. A composite structure (CS) SJ HEMT can be realized by using vertical variation doping in the n-pillar of the SJ structure [11]. Figure 4 illustrates the difference between a SJ vertical GaN HEMT and a vertical GaN HEMT.



(a) (b)

Fig. 4 Comparision between SJ structure and non-SJ structure[6]

By combining an SJ drift region, p+ GaN current blocking layers, and 2-DEG carrier channels, a GaN SJ current aperture vertical electron transistor (CAVETs) can be formed [7]. A normally-off operation without a p-GaN layer stacked on AlGaN, is achieved by implementing a dual-gate structure; a normally-off MOS channel and a normally-on 2-DEG channel. The p+ GaN blocking layer makes sure that the vertical electron conduction is confined below the gate electrode. The n+ GaN source region provides electrons to the 2DEG channels. Alternatively, a GaN vertical SJ fin field effect transistor (FinFET) can also be used [7]. The device is accommodated with multiple fin channels. These fin channels deplete electrons at zero gate bias due to the work function difference between gate and GaN. They deplete when shrunk below 500 nm and become normally off. On the other hand, when the gate bias is increased, the depletion width decreases, and the current conduction is dominated by accumulation type MOS channels at the fin sidewalls [7]. However, this presents fabrication challenges like the etching damage to GaN in regrowth technique leading to higher reverse bias leakage and large Mg drive-in in ion implantation; this limits the reduction of on-resistance and impedes on the benefit of small pillar (p and n) widths. A lateral polar junction to fabricate a vertical GaN SJ device [8][10] can be used to overcome these difficulties. Fig. 5 depicts the fabrication process of a GaN SJ using a lateral polar junction approach.

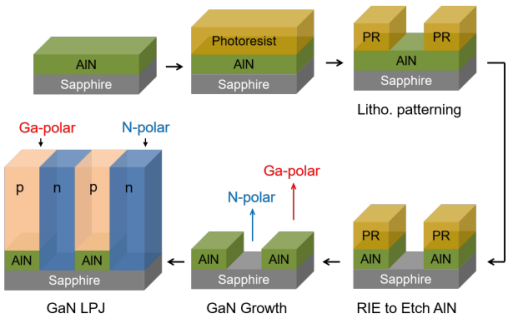


Fig. 5 process flow for GaN SJ using lateral polar junction (LPJ) [8]

B. Static Characteristics

The $R_{ON,SP}$ vs BV trade-off, as expected is improved by implementing a SJ structure in the existing GaN power device. This is done by keeping the same half pillar widths, constant $1*10^{16}$ cm^{-3} n-pillar doping and changing the p-pillar doping which changes the $R_{ON,SP}$. An increase in the p-pillar doping would slightly increase the $R_{ON,SP}$[6]. The BV is dependent on p-pillar doping; if doping concentration is higher or lower than the n-pillar, the BV decreases. However, a small pillar width is suitable for a better trade-off [8]. This can be derived from equation (1).

$$N_D \cdot W_N = N_A \cdot W_P = 1.08 * 10^{14} \cdot BV^{(-1/6)}\ [cm^{-2}] \quad (1)$$

Where N_A and N_D are the doping concentration of n and p-pillars respectively, and W_N and W_P are the widths of n and p-pillars respectively. Equation (2) states the equation to calculate an ideal $R_{ON,SP}$:

$$R_{ON,SP-IDEAL} = \left(\frac{L_D}{q\mu_n N_D}\right) * \left(\frac{W_N+W_P}{W_N}\right) \quad (2)$$

When the $R_{ON,SP}$ of CAVET and FinFETs is compared, it is observed that the FinFETS have a better $R_{ON,SP}$. This is because they accommodate more dense carrier channels in the chip area when compared to CAVET [7]. The $R_{ON,SP}$ can be further improved by reducing fin spacing, thus increasing the number of carrier channels. This although comes at a higher manufacturing cost. In the context of SJs, other resistances like substrate resistances must be very low. Failure to achieve this would render the reduction of $R_{ON,SP}$ useless. Table III shows a comparison of $R_{ON,SP}$ of various GaN SJ concepts.

Device	Ron,sp ($m\Omega cm^2$)
GaN SJ HEMT at 12.4 kV [6]	4.2
GaN SJ FinFET at 1000 V [7]	0.35
GaN SJ CAVET at 1000 V [7]	0.38
GaN CS-SJ HEMT at 2570 V [11]	3.47
GaN special p-pillar SJ HEMT at 2570 V [11]	3.61
GaN doping pillar SJ HEMT at 2570 V [11]	3.86

Table III $R_{ON,SP}$ comparison of various GaN SJ concepts

The robustness of the SJ structure in GaN vertical HEMT for higher breakdown voltage 13.6 kV has also been evaluated. The SJ structure manages a $R_{ON,SP}$ of 12.24 $m\Omega cm^2$ as opposed to

55.1 mΩcm^2 of a conventional HEMT [6]. However, a decrease in on-state from its off-state BV by 4.5% is observed as compared to a decrease of 1.7% in conventional vertical HEMT. The charge imbalance during avalanche breakdown showed an unfavorable negative dynamic on-resistance [6].

C. Dynamic Characteristics

The switching performance of the GaN SJ structure is evaluated. The SJ-FinFET and SJ-CAVET show considerably lower miller capacitance (C_{RSS}) than their counter non-SJ structures. Since, C_{RSS} influence switching losses, it can be reduced for better performance. The C_{RSS} is reduced by 23 times in SJ-CAVET and by 18 times in SJ-FinFET at a drain to source voltage of 1000 V [7]. The overall power loss at 1 MHz, 1.7 kV, and 50 A is 400 W for SJ-CAVET and approximately 475 W for SJ-FinFET. Table IV shows the time associated with turning the devices on and off.

Time and Device	SJ CAVET	SJ FinFET	CAVET	FINFET
Total Turn-on (ns)	2.26	2.91	4.41	9.09
Total Turn-off (ns)	1.63	2.89	5.2	11.35
Rise time (ns)	1.06	1.47	2.50	6.55
Fall time (ns)	0.95	1.87	1.10	4.35
Turn-on delay (ns)	1.2	1.44	1.91	2.54
Turn-off delay (ns)	0.68	1.02	4.10	7.00

Table IV. Switching time of devices (adapted from [7])

The C_{RSS} for the SJ-CAVET is relatively high at low voltages, which could be due to the domination of 2-DEG at C_{GD} at low drain-source voltages. However, by using a thick gate dielectric the C_{GD} can be potentially reduced, but at a cost of higher channel resistance and increased gate operation voltage. The drain-source voltage is approximately 200 V when the C_{RSS} is minimum in GaN SJ HEMT, which is much larger than in Si CoolMOS [7]. The gate charges, total gate charge, gate-source charge (Q_{GS}), and gate-drain charge (Q_{GD}) are lower in SJ-CAVET as compared to SJ-FINFET due to the smaller gate area.

IV. Challenges

The super junction concept is efficient in terms of performance, achieving a good $R_{ON,SP}$ - BV trade-off, but it also poses some challenges. The cost of all proposed methods has been one of the biggest concerns. The complex and intricate doping, and structure of the power devices require high-cost manufacturing. Further, both the SiC SJ structure and GaN SJ structure exhibit several undesirable effects. The carriers injected from the p-n column of SiC SJ MOSFETs have a longer life which leads to a higher Q_{RR}, which is undesirable. This can be observed when the intrinsic body diode of the SiC SJ device is used as a freewheeling diode in the third quadrant [1]. Semi-superjunction designs can potentially improve the robustness of GaN SJ HEMTs as well [6]. Additionally, due to charge imbalance during avalanche breakdown, GaN SJ HEMTs exhibit an unfavorable dynamic on-resistance. If the charge balance is not maintained the breakdown voltage in-turn is not maintained. This imbalance would cause the BV to drop drastically.

To the author's best knowledge, the SiC and GaN based superjunction devices are not available commercially yet. Therefore, analysis of these devices is only done through simulations and laboratory experiments. Although, there are industrial companies working to make wide-bandgap superjunction devices commercially available. Fuji Electrics have worked on manufacturing and testing a 4H-SiC superjunction MOSFETs as an application in inverter circuits [13].

The literature discussed in this brief suggests that GaN SJ devices are likely to perform better than SiC SJ devices for higher breakdown voltages (greater than 3.3 kV-class BV). There is $R_{ON,SP}$ of 4.2 mΩcm^2 for GaN SJ HEMTs at 12.4 kV voltage class when compared to the $R_{ON,SP}$ of 3.3 mΩcm^2 for SiC SJ MOSFETs at 3.3 kV voltage class[6][12]. Furthermore, the advantages of superjunction structure are distinct in higher temperature operation of the device.

V. Conclusion

This paper gives a detailed review of the wide-bandgap semiconductor superjunction power devices. The fabrication process, electrical characteristics, and dependence of various parameters on device properties have been shown. The SiC and GaN-based power devices can benefit massively from the superjunction concept. The advantages can be seen more prominently in higher temperature applications. They have some major challenges such as sensitivity to charge imbalances, reverse recovery charges, and costly manufacturing. Although, the improved $R_{ON,\ SP}$ vs BV trade-off, faster switching, and reduced switching losses make such power devices a compelling argument for manufacturing and making them commercially available.

References

[1] S. Harada et al., "First Demonstration of Dynamic Characteristics for SiC Superjunction MOSFET Realized using Multi-epitaxial Growth Method," 2018 IEEE International Electron Devices Meeting (IEDM), 2018, pp. 8.2.1-8.2.4.

[2] P. Vudumula and S. Kotamraju, "Design and Optimization of SiC Super-Junction MOSFET Using Vertical Variation Doping Profile," in IEEE Transactions on Electron Devices, vol. 66, no. 3, pp. 1402-1408, March 2019, doi: 10.1109/TED.2019.2894650 I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] M. Huang, Y. Deng, L. Lai, Z. Yang, B. Gao and M. Gong, "A Vertical Superjunction MOSFET With n-Si and p-3C-SiC Pillars," in IEEE Transactions on Electron Devices, vol. 66, no. 9, pp. 3922-3928, Sept. 2019, doi: 10.1109/TED.2019.2929831. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] Y. Kobayashi et al., "High-temperature Performance of 1.2 kV-class SiC Super Junction MOSFET," 2019 31st International Symposium on Power Semiconductor Devices and ICs (ISPSD), 2019, pp. 31-34.

[5] H. Wang, C. Wang, B. Wang, N. Ren and K. Sheng, "4H-SiC Super-Junction JFET: Design and Experimental Demonstration," in IEEE Electron Device Letters, vol. 41, no. 3, pp. 445-448, March 2020.

[6] Z. Li and T. P. Chow, "Design and Simulation of 5–20-kV GaN Enhancement-Mode Vertical Superjunction HEMT," in IEEE Transactions on Electron Devices, vol. 60, no. 10, pp. 3230-3237, Oct. 2013.

[7] M. Xiao, R. Zhang, D. Dong, H. Wang and Y. Zhang, "Design and Simulation of GaN Superjunction Transistors With 2-DEG Channels and

Fin Channels," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 7, no. 3, pp. 1475-1484, Sept. 2019.

[8] Khachariya, D., Szymanski, D., Reddy, P., Kohn, E., Sitar, Z., Collazo, R. and Pavlidis, S., 2020. A Path Toward Vertical GaN Superjunction Devices. ECS Transactions, 98(6), p.69.

[9] M. Baba, T. Tawara, T. Morimoto, S. Harada, M. Takei and H. Kimura, "Ultra-Low Specific on-Resistance Achieved in 3.3 kV-Class SiC Superjunction MOSFET," 2021 33rd International Symposium on Power Semiconductor Devices and ICs (ISPSD), 2021, pp. 83-86.

[10] Szymanski, D., Khachariya, D., Eldred, T.B., Bagheri, P., Washiyama, S., Chang, A., Pavlidis, S., Kirste, R., Reddy, P., Kohn, E. and Lauhon, L., 2022. GaN lateral polar junction arrays with 3D control of doping by supersaturation modulated growth: A path toward III-nitride superjunctions. Journal of Applied Physics, 131(1), p.015703.

[11] M. Zhang et al., "Study of AlGaN/GaN Vertical Superjunction HEMT for Improvement of Breakdown Voltage and Specific On-Resistance," in IEEE Access, vol. 9, pp. 9895-9902, 2021.

[12] M. Baba, T. Tawara, T. Morimoto, S. Harada, M. Takei and H. Kimura, "Ultra-Low Specific on-Resistance Achieved in 3.3 kV-Class SiC Superjunction MOSFET," 2021 33rd International Symposium on Power Semiconductor Devices and ICs (ISPSD), 2021, pp. 83-86.

[13] TAWARA, T., BABA, M. and TAKENAKA, K., 2020. 1.2-kV SiC Superjunction MOSFETs. Power Semiconductors Contributing to Energy Management, 66(4), p.237.

A Review on Metaheuristic Design Optimization of Power Converters using Artificial Intelligence

Msc in Electrical Engineering (Smart Systems)
Universität Stuttgart)
Stuttgart, Germany

***Abstract*—Efficiency and power density are two substantially at odds critical performance indicators that must be simultaneously maximized in power converter design and layout. A stimulative technique is preferred to achieve design goals within the provided boundary conditions, such as the maximum junction and case temperature, maximum volume, because an optimization with hardware iterations is quite laborious and expensive. Artificial intelligence has recently been utilized to significantly reduce simulation effort and accomplish optimization with realistic time-effort through the application of genetic algorithms and ant swarm populations.**

This paper outlines utilization of genetic algorithms and ant colony optimization in design optimization, in addition to that Comparison of different algorithms and their advantages with respect to classical methods and Reviewing the different subtopics of power converters in which the use of artificial intelligence.

***Index Terms*—Metaheuristic methods, optimization methods, power converter, genetic algorithm, ant colony optimization.**

I. Introduction

Optimization is a procedure that is essential to day-to-day existence. In its most basic form, it can be understood as the method of determining how to use resources most effectively while also abiding by any potential limits. There are various steps in the optimization process: Mathematically define the system, list the variables and the requirements they must meet, list the system's properties, and then look for the system of state that is, the variable values—that produces the desired properties, either maximum or minimum.

Many methods have been carried out put out over the years for optimization. The majority of strategies have basis of established techniques, including sequential unconstrained minimization, augmented Lagrangian, Newton-Raphson (NR), successive quadratic programming, steepest descent, dynamic and integer programming, and stochastic Newton optimization.

Traditional techniques, such nonlinear and linear programming, are effective strategies that can be applied to address specific optimization issues in power system applications. These methods do not work well to handle complex optimization issues, which is a disadvantage. More advanced optimization strategies that get around the drawbacks of traditional approaches must be utilized as the complexity of the problem rises, especially when the addition of uncertainties on the system. This objective has guided the development of metaheuristic techniques.

Natural selection and social adaptation drive metaheuristic strategies to mimic the best aspects of nature. It is basic properties and advantage, Roli and Blum, Boussa¨ıd et al., and Dreo´ et al., among many others. [1] [2] [3] [4] respectively.

Alternatively, a collection of solutions are used in population-based algorithms (that is, a population of solutions). An iterative procedure is used to increase the initial population after it is formed randomly. A new generation is created after each iteration when some members of the population are replaced with freshly created individuals (typically those whose traits are more suited to the issue at hand). These methods are known as exploration-oriented approaches because their primary strength lies in the diversity of the search domain. [1]

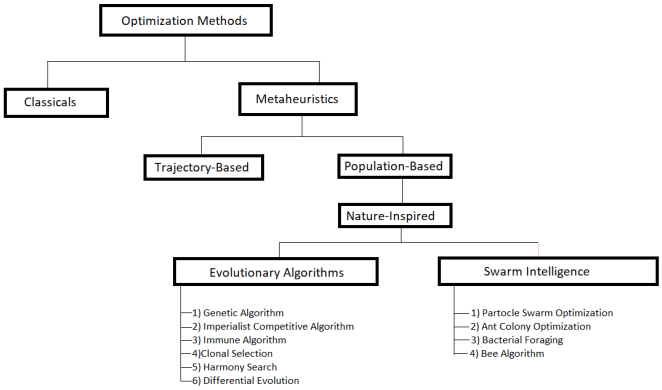


Fig. 1. (Metaheuristics optimization methods classification. [5])

On basis of earlier classifications that have been put forth in the literature, a condensed classification of the optimization techniques explored in this study is presented in the Fig. 1.

The goal of a design optimization procedure from the perspective of power electronics is to fully specify the components that will be used to construct a converter in such a way that its values, dimensions, operating frequency, etc., will result in the minimization (or maximization, depending on their nature) of a given characteristic previously defined by the designer (for example, power losses, power density, THD, etc.), while simultaneously satisfying the design-specific requirements. [6].

Table (1) Advantages and Limitations						
Type	Algorithms	Implementation Simplicity	Global Convergence	Convergence Speed	Parallel Capability	Exemplary Applications
Population based method	Praticle swarm optimization(PSO)	Inferior (+)	Superior (+++)	Inferior (+)	YES	Design, Control Maintenance
	Genetic Algorithm (GO)	Inferior (+)	Superior (+++)	Inferior (+)	YES	Design, Control Maintenance
	Ant colony optimization(ACO)	Intermediate (++)	Intermediate (++)	Intermediate (++)	YES	Design, Control
	Differential evolutionary(DE)	Intermediate (++)	Superior (+++)	Intermediate (++)	YES	Control
	Immune Algorithm (IO)	Intermediate (++)	Intermediate (++)	Intermediate (++)	YES	Control
Trajectory based method	Tabu search method	Superior (+++)	Inferior (+)	Superior (+++)	NO	Design
	Simulated annealing method	Superior (+++)	Superior (+++)	Superior (+++)	NO	Control

II. METAHEURISTICS METHODS

Once the optimization objective for a given application has been stated, the best solution can be found using either a nondeterministic programming method, such as the metaheuristic method, or a deterministic programming method (such as linear or quadratic programming). The intricacy of the majority of the optimization jobs in power electronics makes it difficult for deterministic programming techniques to calculate the gradient and Hessian matrices [9]. For a variety of optimization tasks, metaheuristic approaches act as a general end-to-end tool that requires less specialized knowledge and is effective and scalable. [8]

The metaheuristic techniques can be divided into trajectory-based techniques (such as tabu search method [13], simulated annealing method [14], etc.) and population-based techniques (such as genetic algorithm (GA), particle swarm optimization (PSO) [15], ACO, differential evolution [16], immune algorithm (IA) [17], etc.). For the trajectory-based techniques, there is only one candidate solution included in each exploration step, and it develops into another solution in accordance with a set of rules. The effectiveness of the approach is mostly determined by the standard and effectiveness of the rule. As a result, for non-convex optimization tasks, the ultimate solution is frequently a local rather than a global solution, and the convergence speed of the trajectory-based approaches is typically slow. Multiple candidate solutions are created at random for the population-based techniques. They are more effective for large-scale optimization tasks than trajectory-based approaches and have faster convergence rates and global searching capabilities. However, population-based approaches have a heavier computing requirement. For online application scenarios where effectiveness and speed are crucial, this difficulty must be taken into account.

A list of power electronics-related metaheuristic techniques, together with their benefits and drawbacks, is presented in Table I. In terms of a number of crucial characteristics, such as implementation ease, global convergence, convergence speed, and parallelism, these metaheuristic algorithms are qualitatively compared. It is important to note that choosing the optimum strategy is a difficult undertaking that depends on the application. Since they serve as the foundation and exemplars for evolutionary algorithms and swarm intelligence algorithms, respectively, on which several versions are constructed, GA and PSO are the two metaheuristic techniques most frequently used in power electronics. According to Table 1, practitioners might select a method based on its superiority.

A. *Genetic Algorithm*

Genetic algorithm (GA), one of the metaheuristic optimization methods, has been proved to be an effective way to identify solutions close to the global optimum. GA was invented by J. H. Holland in the early 1970s as a class of search strategies inspired by evolutionary biology. It has been used to optimize particular components in power electronic systems, including system controllers [11], modulation schemes [17], component value optimization [18], and battery chargers [18]. It is less reliant on the initial guess.

The GA-optimized values for the circuits, however, are occasionally not immediately accessible and call for post-fabrication. For instance, discrete values are used in the manufacturing of resistors. To create the values that GAs have optimized, it is occasionally required to link a number of resistors in series and/or parallel. Additionally, the real component value would inevitably deviate somewhat from its nominal values due to component tolerance. A GA optimizer works on a coding of the parameters rather than the parameters themselves. It carries out the ensuing actions in order.

B. *Ant colony optimization (ACO)*

Paths via graphs are often searched using ant colony optimization (ACO). The idea is based on modeling how ants navigate their way from the colony to food. Its searching algorithm can be used to optimize electrical circuits that include discrete-valued components like resistors and capacitors. However, inductors and other components with continuous

values are typically found in power electronic circuits (PECs). As a result, straight use of the conventional ACO algorithm is not possible.Combinatorial optimization issues like traveling salesman problems, data mining, network routing, and controller design are solved using a multiagent approach. ACO is increasingly being used in electrical engineering projects like load-flow studies and developing power distribution systems [18].

III. Different subtopics of power converters in which people used artificial intelligence

A. Control Tuning

Constrained optimization techniques based on evolutionary computation algorithms have been used to tackle the estimation of self-tuning parameters for inverter controllers. These techniques can offer real-time response, a crucial component that enhances the quality of the power in grid-connected systems. [19]. The system load current is evenly distributed among the converters, and the system efficiency is low, especially at low- and medium-load circumstances. Lexuan et al. [19] offer an optimization approach for enhancing operating efficiency of a paralleled dc-dc converter system.

The system's virtual resistances serve as decision variables, and the aim function is to minimize the system's overall conversion losses. To change the ratio of converters exchanging power, a virtual resistance shifting technique is suggested. In order to increase system efficiency, a hierarchical control concept is used, where the droop method (" The so-called droop control is to select the frequency droop characteristic curve (Droop Character) similar to the traditional generator as the micro-source control method, that is, to obtain stable frequency and voltage through P/f droop control and Q/V droop control respectively ") is used on top of the primary control level, voltage secondary control is in charge of voltage deviation restoration, smoothing higher level regulation, and a GA is implemented in the tertiary level to search for the optimal sharing ratio. The best adjustment of virtual resistances can be made online because the bandwidth of optimization control is typically significantly smaller than that of inner loops and droop controller. [5]

anabria and Garzon [20] describe the multiobjective optimization of a fuzzy controller in a buck converter. The objective functions are the steady-state error e_ss , and the ripple factor F_r. The goals is to achieve a step response satisfying the constraints $(e_ss \leq 3\%) and (\mathrm{F}_r \leq 3\%)$.

The duty cycle, which is regarded as a decision variable, is what the controller architecture aims to ascertain. In order to ensure that the buck converter is set at the required value or a value that is close to the set point, the fuzzy controller modifies the duty cycle.

The non-dominated sorting genetic algorithm II was used for the optimization since it uses fewer computer resources and is appropriate for online applications. Fig.4 shows the step response of the converter, before and after tuning (with e_ss = 2.0684% and F_r = 1.784%). A grid-connected inverter must exhibit the fewest frequency fluctuations possible in order

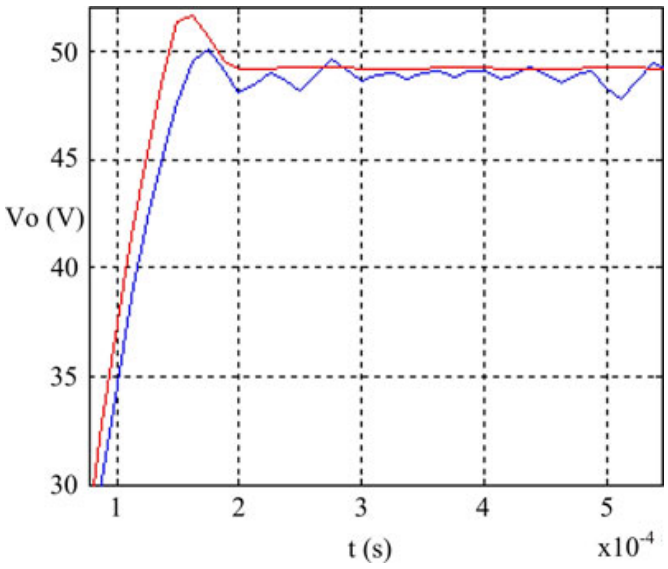


Fig. 2. (Comparison of the buck converter to step response. Blue: before tuning. Red: after tuning. [19])

to avoid affecting the active power transferred, the control system's equilibrium point, or the power flow equilibrium.

Applying a differential evolution-based optimization approach, as described in [20], can significantly improve the system's dynamic reactivity. It is possible to assure system stability with a short settling time by adjusting the slopes of the P_w and Q-V curves. The two decision variables, kv and kp, respectively, determine the slope of the curves Q-V and P_w. Differential Evolution satisfies a number of criteria, which makes it a desirable method. First, regardless of the original population decision, the solutions eventually reach a global maximum. Second, the algorithm's convergence rate enables its implementation in online applications. According to [20], the PSO methodology has also been used to find the best controller values. Additionally, the suggested control solutions efficiently reduce current harmonic distortion, resulting in THD values for grid-connected inverters that are substantially below the 5 percent limit imposed by standard IEEE 1547..

The optimization of a PI voltage controller (decision variable being the proportional and integral gains k_{pv} and k_i respectively) as well as a PI current controller (with k_{pc} and k_{Ic} as decision variables) in a microgrid application is described by Hassan and Abido in [21]. As a constrained problem, the optimization is handled and includes limitations for the choice variables' upper and lower limits.

Several runs in the autonomous and grid-connected modes were conducted in order to evaluate the resilience of the suggested controllers under various scenarios. To show the efficiency of the suggested design method, several disruptions were used. The step change in the reference power was employed in the grid-connected mode to evaluate the system's capacity to follow this reference power. Nonlinear time-domain simulations and eigenvalue analysis were both used to examine the stability of the system. Step change and fault

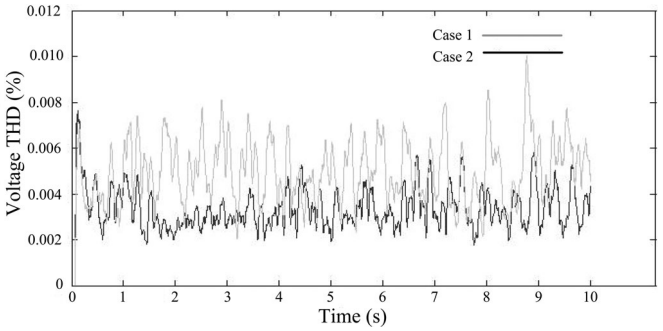


Fig. 3. (Voltage THD obtained. Case 1: Based on trial and error method, and case 2: based on the BFO algorithm. [22])

disturbances were employed in the autonomous mode to test the system's stability. The outcomes support the usefulness of the suggested PSO-based method for PI controller parameter optimization. The proposed PSO technique's robustness in relation to its initial hypothesis has been proven.

A BFO algorithm is used in [22] to tune the controllers while adhering to grid interconnection regulations and reducing voltage and current harmonic distortions. When a trial-and-error tuning technique is compared to the BFO algorithm, the later one offers a lower voltage THD, as seen in Fig. 03.

A metaheuristic algorithm will typically struggle to converge to an appropriate value in a reasonable amount of time given the computing capacity of today's computers. The metaheuristic algorithm must offer a suitable control signal for each control moment in the case of a controller. The population's behavior might still be subpar if evolution has just taken place over a few generations. Parallel computing is a very helpful tool for overcoming these difficulties.

B. Automated Heatsink Optimization for Air-Cooled Power Semiconductor Modules

For improving power density and reliability of power semiconductor modules, heat-sink design is essential. In this letter, a genetic algorithm- and finite element-based automated design and optimization methodology for air-cooled heatsinks is given. The fitness function of each heatsink, or junction temperature of semiconductor devices, is evaluated using finite element analysis while the genetic algorithm creates a population of candidates with complicated heatsink cross-section shape in each iteration. For the purpose of evaluating thermal performance, the optimal heatsink is constructed using three-dimensional printing technology. Based on a 50-kW three-phase air-cooled inverter with the manufactured heatsinks, simulation and experimental assessments have been carried out. The comparative evaluation results show that the optimized heat-sink is superior over a customized solution by 27% less in size and 6% lower in junction temperature. [23]

The simulation results using the suggested design methodology converge to the heatsink illustrated in Fig. 4 with the highest junction temperature of 106.29 °C after a total computation time of 135 h. The second-stage optimized heatsink is produced using the first stage's findings. The image illustrates

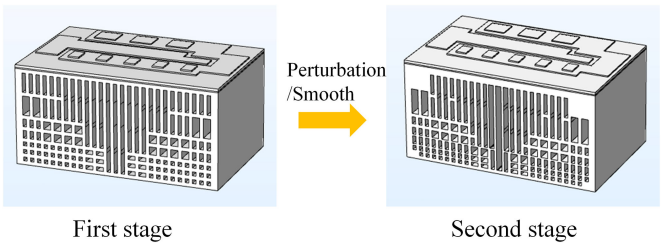


Fig. 4. (Optimization results using the proposed approach. [23])

how further minor adjustments are made as well as a reduction in wall thickness in the center. A reduced maximum junction temperature of 102.1 °C is consequently attained. [23]

IV. CONCLUSION

The metaheuristic algorithms that are relevant to the study of power converter development have been introduced in this survey, along with a discussion of their benefits over more traditional approaches. The survey only represents a portion of the available literature in terms of the number of citations. However, the optimization cases discussed in this work are representative of the wide range of power converter-related subjects where metaheuristic algorithms have demonstrated their effectiveness and profitability.

THD reduction, harmonics elimination, and efficiency maximization have all gotten greater attention than other areas, according to a survey of the literature on power converter optimization. Other topics, like dependability optimization, haven't been fully investigated despite their importance. The bulk of publications describe offline optimization techniques that produce positive outcomes. Due to the considerable computational effort frequently needed to ensure that the parameters are available when needed, online optimization employing metaheuristic algorithms is typically fairly uncommon to date.

In numerous power electronics applications, the benefits of metaheuristic optimization methods—including GA, PSO, ACO, BFO, and BA—over traditional optimization techniques have been well shown. It should be noted, however, that choosing the appropriate metaheuristic approach is not an easy task because the outcomes depend on the objective function chosen and the specific application needs. Numerous optimization issues in the field of power electronics are inherently complicated, and finding answers necessitates significant computational effort.

The augmented Lagrangian, dynamic programming, linear and integer programming, among other traditional optimization techniques, can be used to find optimal solutions. Although these methods are exceedingly time-consuming when handling complicated and dynamic issues (such as large-dimension search spaces, scarcely confined problems, multimodal and time-varying problems), they are frequently ineffective in practice. As a result, they offer numerical precision of the findings and computational efficiency. In contrast, metaheuristic optimization methods enable the computation of suboptimal or optimal solutions in a suitable execution time.

Metaheuristic approaches also have the benefit of being generic, so they may be utilized as ”black boxes” and don't need to be fine-tuned for every problem.

V. FUTURE PERSPECTIVES

The use of metaheuristic algorithms in power electronics is fundamentally linked to broader advancements in the study of the theory and applications of metaheuristic algorithms. As a result of metaheuristics' general character, advancements in industries including banking, engineering, logistics, and transportation have been put to the test and have produced positive outcomes in power electronics applications. But even though it is a developing field of study, the research on metaheuristic algorithms used in power converter design is still quite young.

The availability of low-cost, high-performance computer architectures is essential for maximizing the potential of these algorithms because they will improve the effectiveness and caliber of the results obtained for offline optimization and help to overcome the limitations of online applications that involve instantaneous decisions (e.g., fractions of a second). The extensive application of metaheuristic optimization techniques in online optimization has been significantly hampered by their high computing cost. Parallelization can speed up computations for optimization utilizing the GA, however as suggested in [24], the computational cost can only be decreased by limiting the number of function evaluations.

Numerous optimization issues in the field of power electronics sometimes involve several objectives, necessitating the use of sophisticated models to faithfully capture the case's actual circumstances. Therefore, a common method for solving problems is the hybridization of metaheuristic algorithms, which combines the advantageous aspects of many algorithms to lessen their particular flaws. A significant topic for future research is the hybridization of metaheuristic algorithms.

REFERENCES

[1] E. Alba, Parallel Metaheuristics: A New Class of Algorithms. Hoboken, NJ, USA: Wiley, 2005.

[2] C. Blum and A. Roli, “Metaheuristics in combinatorial optimization: Overview and conceptual comparison,” ACM Comput. Surv., vol. 35, pp. 268–308, 2003.

[3] Boussa¨ıd, J. Lepagnot, and P. Siarry, “A survey on optimization metaheuristics,” Inf. Sci., vol. 237, pp. 82–117, 2013.

[4] J. Dreo, A. P ´ etrowski, P. Siarry, and E. Taillard, ´ Metaheuristics for Hard Optimization. Berlin, Germany: Springer, 2006.

[5] Metaheuristic Optimization Methods Applied to Power Converters: A Review S. D. De Léon-Aldaco, H. Calleja, Jesús Aguayo Alquicira 2015, IEEE Transactions on Power Electronics

[6] S. Balachandran and F. C. Y. Lee, “Algorithms for power converter design optimization,” IEEE Trans. Aerosp. Electron. Syst., vol. AES-17, no. 3, pp. 422–432, May 1981.

[7] E. I. Amoiralis, M. A. Tsili, and A. G. Kladas, “Global transformer design optimization using deterministic and non-deterministic algorithms,” in Proc. 20th Int. Conf. Electr. Mach., 2012, pp. 2323–2331.

[8] An Overview of Artificial Intelligence Applications for Power Electronics Shuai Zhao , Member, IEEE, Frede Blaabjerg , Fellow, IEEE, and Huai Wang , Senior Member, IEEE

[9] Matlab, “Manual of neural network toolbox,” 2005. [Online]. Available: http://matlab.izmiran.ru/help/toolbox/nnet/

[10] S. E. De Leon-Aldaco, H. Calleja, and J. A. Alquicira, “Metaheuristic optimization methods applied to power converters: A review,” IEEE Trans. Power Electron., vol. 30, no. 12, pp. 6791–6803, Dec. 2015.

[11] J. Zhang, H. S. H. Chung, W. L. Lo, S. Y. Hui, and A. K. M. Wu, “Implementation of a decoupled optimization technique for design of switching regulators using genetic algorithms,” IEEE Trans. Power Electron., vol. 16, no. 6, pp. 752–763, Nov. 2001.

[12] L. L. Jiang, D. L. Maskell, and J. C. Patra, “A novel ant colony optimization-based maximum power point tracking for photovoltaic systems under partially shaded conditions,” Energy Buildings, vol. 58, pp. 227–236, Mar. 2013.

[13] Q. Wang and S. Niu, “Design, modeling, and control of a novel hybridexcited flux-bidirectional-modulated generator-based wind power generation system,” IEEE Trans. Power Electron., vol. 33, no. 4, pp. 3086–3096, Apr. 2018.

[14] S. Lyden and M. E. Haque, “A simulated annealing global maximum power point tracking approach for PV modules under partial shading conditions,” IEEE Trans. Power Electron., vol. 31, no. 6, pp. 4171–4181, Jun. 2016.

[15] J. Zhang, Y. Shi, and Z.-H. Zhan, “Power electronic circuits design: A particle swarm optimization approach,” in Proc. Asia-Pac. Conf. Simulated Evolution Learn., 2008, pp. 605–614.

[16] Y. C. Hung, F. J. Lin, J. C. Hwang, J. K. Chang, and K. C. Ruan, “Wavelet fuzzy neural network with asymmetric membership function controller for electric power steering system via improved differential evolution,” IEEE Trans. Power Electron., vol. 30, no. 4, pp. 2350–2362, Apr. 2015.

[17] J. Yuan, B. Chen, B. Rao, C. Tian, W. Wang, and X. Xu, “Possible analogy between the optimal digital pulse width modulation technology and the equivalent optimisation problem,” IET Power Electron., vol. 5, no. 7, pp. 1026–1033, Aug. 2012.

[18] Extended Ant Colony Optimization Algorithm for Power Electronic Circuit Design Jun Zhang, Senior Member, IEEE, Henry Shu-Hung Chung, Senior Member, IEEE, Alan Wai-Lun Lo, Member, IEEE, and Tao Huang

[19] M. Lexuan, T. Dragicevic, J. M. Guerrero, and J. C. Vasquez, “Optimization with system damping restoration for droop controlled DC-DC converters,” in Proc. IEEE Energy Convers. Congr. Expo., 2013, pp. 65–72

[20] L. A. Carniato, R. B. Godoy, J. O. P. Pinto, C. A. Canesin, and R. C. Garcia, “Power flow optimization for grid connected inverter using evolutionary algorithm and additional control loop,” in Proc. Braz. Power Electron. Conf., 2011, pp. 422–427.

[21] M. A. Hassan and M. A. Abido, “Optimal design of microgrids in autonomous and grid-connected modes using particle swarm optimization,” IEEE Trans. Power Electron., vol. 26, no. 3, pp. 755–769, Mar. 2011.

[22] A. G. Agamy, A. H. Kasem Alaboudy, H. E. Mostafa, and M. Y. Fekry, “Bacterial foraging-based PI controller of inverter-based distributed generators,” in Proc. IEEE Trondheim PowerTech Conf., 2011, pp. 1–7.

[23] Automated Heatsink Optimization for Air-Cooled Power Semiconductor Modules Tong Wu , Student Member, IEEE, Zhiqiang Wang , Member, IEEE, Burak Ozpineci, Senior Member, IEEE, Madhu Chinthavali, Senior Member, IEEE, and Steven Campbell , Member, IEEE

[24] T. Goel and N. Stander, “A study on the convergence of multiobjective evolutionary algorithms,” in Proc. Conf. Multidisciplinary Anal. Optim., May 3, 2010, pp. 1–18.

Time-Efficient Multi-Domain Simulations for Virtual Prototyping of Power Converters

Electrical Engineering
Universität Stuttgart
Stuttgart, Germany

***Abstract*— To simulate a power converter, various aspects like its electrical and thermal properties have to be considered. This paper gives an overview of the virtual prototyping process of a power converter. Also, modelling approaches like numerical methods and equivalent circuit methods are discussed briefly. It is seen that Model Order Reduction (MOR) techniques reduce the mesh size of a model without reducing its accuracy which helps in faster simulation. Finally, the ρ-η Pareto optimization technique is discussed which presents the need of finding a trade-off between power density and efficiency of a power module.**

Keywords—Virtual Prototyping, Multi-Physics, Multi-Domain, Simulation tools, Power converters, Model Order Reduction, Pareto Optimization

I. INTRODUCTION

A power electronic device is usually made up of materials like Si, SiC or GaN. The converter circuit along with its driving circuit comprises elements like a switching device, capacitors, inductors, a double-bonded copper layer, and a heatsink. When designing a physical prototype of a structure like this, parameters like material properties, geometrical design and components have to be considered. The prototype is then tested for different conditions and specifications. If a specific requirement is not getting fulfilled, the respective changes are made to the design and the whole process is repeated. This process drains a lot of time, money and human efforts. Virtual prototyping is an easy and affordable solution where a power electronic system can be built on a simulation tool. This tool would, ideally, evaluate the performance of the prospective design and analyze its behaviour in electrical, thermal, magnetic, and mechanical domains without any errors [1]. This process is often referred to as 'Multi-Domain Simulation'. The near-perfect virtual prototype, in reality, is obtained by an iterative process with design changes made in each iteration so that the obtained simulation results are much closer to the desired results. Figure 1 shows the flowchart of a virtual prototyping process.

However, there are a few barriers to achieving the above-mentioned ideal conditions. The time taken in electrical domain simulation – nanoseconds, microseconds – is very less when compared to the time taken in the thermal domain simulations (seconds and minutes) because the thermal model takes longer to reach the steady state operation [2]. Moreover, not every simulation tool is perfect for evaluating all the parameters. Some are better at calculating the electrical parasitic, while others focus more on thermal or electromagnetic characteristics. The predominant way to find a solution to both problems is interfacing. To explain the concept of interfacing in a simpler way, consider X, Y, and Z to be the three different simulation tools.

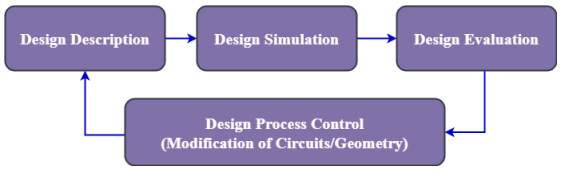


Fig.1 – The basic structure of the virtual prototyping process [7]

Let us assume that electrical properties are analyzed in X, thermal properties are analyzed in Y and magnetic properties are analyzed in Z. Now, to get the combination of all the results in a single simulation tool, results obtained in X and Z are coupled into Y and finally, all the performance and characteristics of the device in each domain can be observed in Y. This method has proven to be very helpful in designing and evaluating the power electronic structures. Though the error between the virtual and physical prototype has reduced compared to the past, the high simulation time and accuracy are still a concern.

The other barrier to virtual prototyping of power electronic modules is finding the optimum trade-off between the power density of the module and its efficiency [3]. The increasing power density of a power module is very important as research today is focusing more on reducing the size of electronic components while maintaining the power handling capacity of the module. However, losses due to various aspects (switching and conduction losses, parasitic losses and thermal losses) should also be reduced to increase efficiency. However, these two parameters are conflicting with each other as one gets decreased if we try to increase the other parameter. Hence, to maintain the balance between the two of them, Pareto optimization is done. This gives a range of efficient models among which a user can select a model according to his application.

The paper is divided in such a way that first, the virtual prototyping process of the power electronic module is discussed. This paper majorly focuses on electrical and thermal models. Simultaneously, model order reduction techniques are discussed which are used to make the models lighter in size and

faster to simulate. Finally, the concept of Pareto optimization is understood and a few examples are discussed.

II. Virtual Prototyping Process

A. Modelling Approaches

For any power module to be modelled, an initially desired design geometry is prepared. This design geometry can be modularized in various ways. A classification of all the methods is done as shown in figure 2.

A design geometry can be of two ways: physical model and behavioural model. The physical model represents the internal structure and physical processes of the device. The drawbacks of this model are it is very complex and its simulation time is very high. On the other hand, a behavioural model can be used to represent the electrical and thermal behaviour of the device. It is very flexible and easy to implement in a simulation tool. Moreover, this model can be used to represent passive and active device driver circuits.

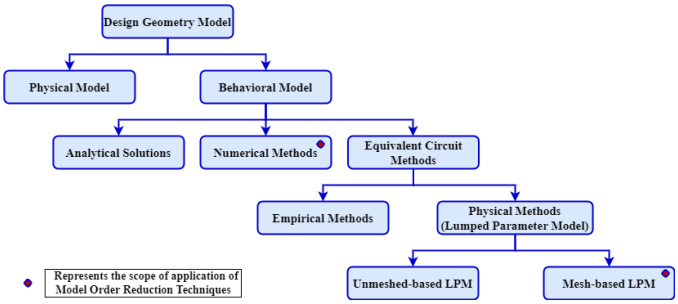


Fig.2 – Classification of modelling methods

The behavioural models can be divided into three types: analytical solutions, numerical methods and equivalent circuit methods [4]. In general, analytical solutions are possible with only a limited set of simple geometries, like cubes and cuboids. Whereas numerical methods can analyse any geometry. Additionally, it can reflect information on any geometry and layout. A few examples of this method are the Finite Difference Method (FDM), Finite Element Method (FEM), Partial Element Equivalent Circuit (PEEC), Computational Fluid Dynamics (CFD) and many more. These methods, to start with, translates the basic parameters of the device such as geometry and construction materials into an equivalent mathematical model. This is done by the concept called spatial discretization (meshing) which generates a huge number of equations, each for a discretized element. These equations when solved give the behaviour of the respective element. The advantage of this process is we tend to get more accurate results as we increase the mesh size. However, solving these large number of equations along with providing accurate detailed results is very challenging. The simulation time and computational costs are very high while modelling a large system with an iterative design and optimization process.

The third type of modelling is equivalent circuit methods. This method is used to represent the design in terms of basic electrical elements such as resistances, capacitances, ideal voltage source, ideal current source, etc. This can be used to represent electrical as well as thermal models. Equivalent circuit methods are of two types: empirical and physical. In empirical methods, the values of resistances and capacitances are obtained by applying curve fitting to the waveforms which are generated from an experimental test performed on the power device or from the waveforms provided in the datasheet of the specific semiconductor device or power module. This method is only helpful in analysing the fixed systems and hence cannot be used for iterative design optimization. Whereas, in the physical equivalent circuit method, usually called as Lumped Parameter Model (LPM), the values of resistances and capacitances are obtained from the geometrical data and material properties. The LPM also has two classifications: the unmeshed-based and meshed-based LPM. The unmeshed-based method is fast to calculate but this model can't be used for describing the geometry and layout of complex power electronic systems, their hotspot temperature and location. By using the mesh-based LPM (MLPM), the above drawbacks can be removed. This method is also a numerical method where the meshing concept is applied to the research object. Later, every element is represented by a subcircuit which is then connected to form the network for the whole system.

B. Model Order Reduction (MOR) Method

Although the models obtained from the above methods can be accurate according to the application, these models are very time-consuming to simulate due to their bulky size. This poses a difficulty for simulating static as well as dynamic simulations. Initially, while simulating a static model, finding the solution for each equation is very time-consuming. In addition to that, when the dynamic model is considered, the same simulation done for the static model is done here but for each time stamp. This not only increases the time but also the size of the total data being recorded as data is being recorded for each time stamp. Moreover, the computational cost and effort are huge.

Model order reduction is a technique used to reduce the size of the simulated model so that the duration of simulation reduces along with maintaining the accuracy of the obtained model [5]. This technique is applied to a model which is generated by a numerical method using the process of discretisation. Considering the obtained model contains n equations, the reduced order model will contain m equations, where the value of m will be very less than the value of n. This reduction does not affect the accuracy of the system and full spatial post-processing is maintained. To achieve this reduced order model, Krylov subspace-based algorithms are used because these algorithms ensure the speed and reliability of the system.

The above concept can be explained in a better way using the thermal aspect of a power module [6]. Once the thermal model is designed using a suitable numerical method, its respective equations can be reduced to the given following equation 1.

$$\begin{bmatrix} T1 \\ T2 \\ \dots \\ Tn \end{bmatrix} = \begin{bmatrix} A11 & A12 & \dots & A1n \\ A21 & A22 & \dots & A2n \\ \dots & \dots & \dots & \dots \\ An1 & An2 & \dots & Ann \end{bmatrix} . \begin{bmatrix} P1 \\ P2 \\ \dots \\ Pn \end{bmatrix} + Tamb \qquad (1)$$

where the elements Aij represents the thermal transfer function between i and j, Pi represents the loss of power at point i, Tamb represents the ambient temperature and Ti represents the temperature at selected point i. This expression, which evaluates the temperature only at selected points, would still give similar results compared to that of temperature values obtained through an unreduced model.

The Krylov subspace-based algorithms are of various types. Block Arnoldi algorithm is one of them [1]. This algorithm reduces the number of states drastically but retains the same inputs and outputs. In this method, the original matrix is available to allow re-interpolation of any original nodal value, giving us the approximation for values like temperature or heat flux from the reduced model. One of the main advantages is that the volume of the data that is stored for each time step is reduced by a factor of order $10^3 – 10^4$. PRIMA is another type of algorithm which is based on the Block Arnoldi algorithm but it is modified such that it can assure the stable and passive reduced order model.

After the reduced model is generated for electrical and thermal models, it is then solved at a different time constant and then later coupled with each other. Here, the concept of a solver comes into focus. There are various types of solvers used for different scenarios and models. Some of the solvers include Successive Order Relaxation (SOR), and the KLU method. This paper does not delve into the concepts of the solver.

C. Electrical and Thermal Models

In this section, we are going to see a few electro-parasitic models and thermal models which have been designed based on the concepts mentioned above. To obtain the desired model out of the design geometry, there are numerous methods through which this can be done. One such method is discussed in [7], where a modelling hierarchy is followed to obtain the electrical and thermal models. Later, FDM and PEEC methods were used to generate the state space model for the electrical and thermal domains, respectively. The essential point to be noted is that this method uses the Block Arnoldi algorithm for obtaining reduced order thermal model which reduced the mesh size by 1,750 times. Similarly, PRIMA was used to obtain the reduced order electrical model and the mesh size was reduced by approximately 88 times. Moreover, despite the drastic reduction in the size of the model, the simulated values of the load current and temperature were found to be in good agreement with the measured values. Figure 3 shows an example of the layout of mesh generated for thermal analysis of the design.

Simulation of power electronic systems means simulation of not just the parasitic element effect and thermal effect, but also the simulation of a power semiconductor device. In [5], the modelling of SiC-JFET is discussed along with MOR. The modelling of a unipolar power transistor is modelled with equivalent circuit parameters as shown in figure 4. To model the static behaviour, forward static characteristics are represented by current source Gch which is controlled by both voltages Vds and Vgs. A non-linear resistor Rbd and a constant resistor Rbds are used to represent the body diode. For

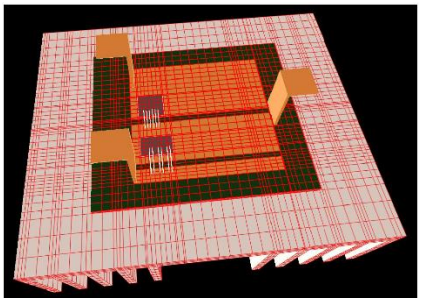


Fig. 3 – Typical mesh generated for thermal design containing 40,932 elements [1]

modelling dynamic behaviour, the three non-linear capacitances C_{gd}, C_{ds}, and C_{gs} are used and these capacitances are dependent on the voltage across them. The MOR technique reduces 723 equations down to 23 equations and hence improves the simulation speed.

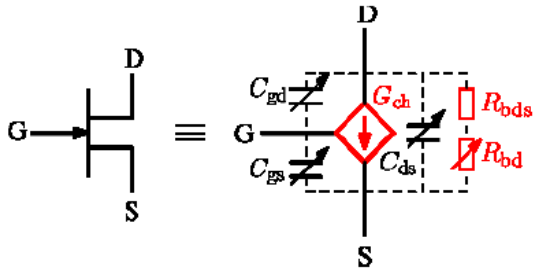


Fig. 4 – Modelling of a transistor [5]

In [8], a T-type LPM is used to model the 3D structure of a power module. In the given T network, the connection of thermal resistances as shown in figure 5 represents the heat propagation path. Only conduction and convection heat transfer mechanisms are being replicated through this model. Along with the modelling of conduction and convection parameters, few conclusions can be drawn in the favour of LPM from this paper.

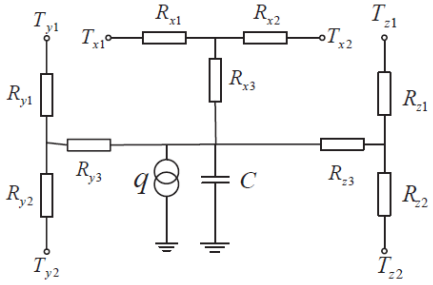


Fig. 5 – LPM for 3- dimensional heat transfer [8]

The duration for obtaining analytical solutions of LPM does not depend on the mesh size. Moreover, when this method is compared with CFD, the simulation speed is found to be much faster along with the value of convection heat transfer coefficient under acceptable error limits.

In [4], a similar method as mentioned above is used for thermal analysis and modelling. However, this paper compares two and three resistors' thermal networks and states that the two resistors' thermal network is accurate only when there is no

internal heat generation, i.e., q=0. The capacitor in figure 5 predicts transient state temperature response which shows the energy storing ability of the solid materials. The author then states MLPM is better than FDM as its approach is more defined for thermal applications along with more accurate results with a smaller number of meshes.

III. Pareto optimization

The transistors and diodes made of semiconductor material constitute the part of the power converter module along with its circuit which has other elements like capacitors, inductors, a heat sink, and the board which accommodates all these elements. If we arrange these elements spatially, we can ensure minimum electromagnetic interference between the devices. This arrangement provides an extra benefit as all the components and the power device are properly ventilated, reducing the heating tendency of the device and tending to reduce losses. However, this leads to more volume of the structure and in turn reduces the power density (ρ) i.e., the power per unit volume. Along with that, the overall cost of manufacturing also increases.

Now, if we try to increase the power density of the module by using smaller components of better material and stacking up the elements close to each other, this will increase the power handling capacity per unit volume. However, losses will rise due to electromagnetic interference between the components due to switching transients and overheating of the device due to non-uniform heat propagation. This leads to poor efficiency(η). So, ρ and η are the two parameters, of which neither can be increased without making the other parameter worse.

Pareto optimization is an optimizing technique which is used to find an optimal solution when there are two or more conflicting parameters, as mentioned above. This optimization technique is very prominently used for ρ – η optimization. For example, in a given particular application, we can determine if we can use a single, high current rating switching device which may need forced heat propagation methods and consumes less space or multiple devices with the lower current rating, used in parallel, with less power density and natural heat propagation. As mentioned in [9], different values of three parameters namely, switching frequency, current zero interval duration and transformer volume, are used to obtain various models using a flow chart. The efficiency and power density are plotted for each obtained model. The Pareto front labelled in figure 6 shows the set of the most optimized design models where neither entity (efficiency or power density) can be increased at the cost of the other.

Similarly, there are various other constraints like input and output capacitor thermal constraint, heatsink constraint, system dynamic constraint [10], inductance, switching frequency, and battery voltage [3] which can be used to obtain the Pareto optimal solutions between efficiency and power density of a power converter. It is also possible to obtain a three-dimensional Pareto analysis where total power loss is reduced, volume is reduced and reliability is improved [11].

This paper deduced the few conflicting trade-off relationships like efficiency and reliability are positively correlated,

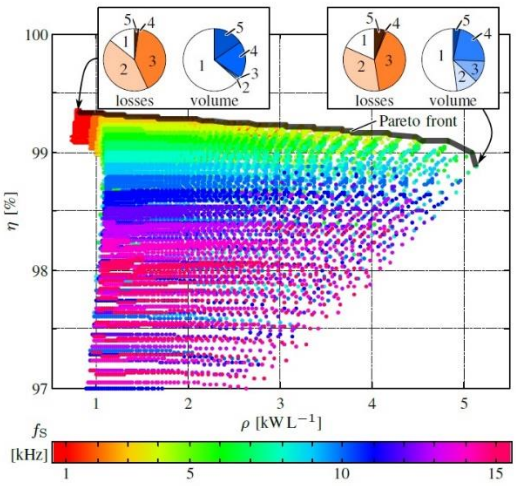


Fig. 6 – Pareto front highlighted in the ηρ-plane [9]

efficiency and power density are negatively correlated and reliability and power density are negatively correlated.

IV. Discussion

As we have seen various modelling and optimizing methods for simulating a power converter, these methods will need deeper understanding and application when it comes to the simulation of power converters involving wide band-gap semiconductors. This is because the WBG devices have higher temperature handling capacity and their switching transients are very fast. Hence, the modelling of dV/dt, dI/dt characteristics and thermo-mechanical characteristics require a deeper insight. In addition to that, it is also equally important to make interfacing of models among different simulation tools user-friendly. In future, we can expect a single tool to be capable of performing a simulation of multiple domains such that the need for interfacing is reduced.

V. Conclusion

In this paper, we learnt about power converters modelling approaches like numerical methods and equivalent circuit methods have been discussed. We identified and understood briefly about making the above-mentioned methods time-efficient and lighter in size through model order reduction techniques. We further observed that the LPM method is more accurate and faster than FDM and CFD methods for thermal domain simulations. Lastly, we found that it is very important to find the right trade-off between power density and efficiency to reduce losses and costs. The study about modelling in WBG devices can be extended in future.

REFERENCES

[1] P. L. Evans, A. Castellazzi and C. M. Johnson, "A design tool for rapid, multi-domain virtual prototyping of power electronic systems," *2015 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2015, pp. 3069-3076, doi: 10.1109/ECCE.2015.7310089.

[2] M. O. Faruque et al., "Interfacing Issues in Multi-Domain Simulation Tools," in IEEE Transactions on Power Delivery, vol. 27, no. 1, pp. 439-448, Jan. 2012, doi: 10.1109/TPWRD.2011.2170861.

[3] B. Strothmann, F. Schafmeister and J. Böcker, "Pareto Design and Switching Frequencies for SiC MOSFETs Applied in an 11 kW Buck Converter for EV-Charging," 2019 IEEE Applied Power Electronics Conference and Exposition (APEC), 2019, pp. 2234-2240, doi: 10.1109/APEC.2019.8721850.

[4] Wenbo Wang, Xibo Yuan, P. L. Evans and P. H. Mellor, "Mesh-based lumped parameter model with MOR for thermal analysis of virtual prototyping for power electronics systems with comparison to FDM," 2017 IEEE 3rd International Future Energy Electronics Conference and ECCE Asia (IFEEC 2017 - ECCE Asia), 2017, pp. 1037-1042, doi: 10.1109/IFEEC.2017.7992184.

[5] K. Li, P. Evans and M. Johnson, "Developing Power Semiconductor Device Model for Virtual Prototyping of Power Electronics Systems," 2016 IEEE Vehicle Power and Propulsion Conference (VPPC), 2016, pp. 1-6, doi: 10.1109/VPPC.2016.7791664.

[6] M. Musallam and C. M. Johnson, "Real-Time Compact Thermal Models for Health Management of Power Electronics," in IEEE Transactions on Power Electronics, vol. 25, no. 6, pp. 1416-1425, June 2010, doi: 10.1109/TPEL.2010.2040634.

[7] P. L. Evans, A. Castellazzi and C. M. Johnson, "A Multi-Disciplinary Virtual Prototyping Design Tool for Power Electronics," CIPS 2014; 8th International Conference on Integrated Power Electronics Systems, 2014, pp. 1-7..

[8] W. Wang and X. Yuan, "Lumped-parameter-based thermal analysis for virtual prototyping of power electronics systems," 8th IET International Conference on Power Electronics, Machines and Drives (PEMD 2016), 2016, pp. 1-6, doi: 10.1049/cp.2016.0244.

[9] J. Huber, G. Ortiz, F. Krismer, N. Widmer and J. W. Kolar, "η-ρ Pareto optimization of bidirectional half-cycle discontinuous-conduction-mode series-resonant DC/DC converter with fixed voltage transfer ratio," 2013 Twenty-Eighth Annual IEEE Applied Power Electronics Conference and Exposition (APEC), 2013, pp. 1413-1420, doi: 10.1109/APEC.2013.6520484.

[10] B. Zhang et al., "Prediction of Pareto-optimal performance improvements in a power conversion system using GaN devices," 2017 IEEE 5th Workshop on Wide Bandgap Power Devices and Applications (WiPDA), 2017, pp. 80-86, doi: 10.1109/WiPDA.2017.8170526.

[11] X. Li, X. Zhang and F. Lin, "Pareto-Frontier-Based Multi-Objective Design of Output LC Filter for High Efficiency, High Reliability, and High Power-Density Buck Converter," 2020 IEEE 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia), 2020, pp. 1579-1582, doi: 10.1109/IPEMC-ECCEAsia48364.2020.9367986.

Parasitic Influences of Wide Bandgap Semiconductors on Soft-Switching

Institute of Robust Power Semicondutors Systems
University of Stuttgart
Stuttgart, Germany

***Abstract*—Soft switching is utilized in order to reduce the switching losses to a great extent as well as to increase the energy density of the power electronic system. Wide bandgap semiconductors require new experimental methods to characterize soft switching losses as the conventional double pulse test fails to measure soft switching losses accurately because of measurement limitations. Calorimetric methods are very accurate for measuring soft switching losses as compared to the conventional double pulse test. But in order to measure the switching losses accurately using calorimetric methods, the thermal system with all the parasitic influences should be properly modeled to get the most accurate switching losses. This work gives a review of state-of-the-art calorimetric characterization techniques; furthermore, different parasitic influences and their impacts on soft switching are discussed.**

***Index Terms*—wide band gap semiconductors, soft-switching, zero voltage switching, calorimetric measurements.**

I. Introduction

The wide bandgap semiconductors are gaining more and more importance due to their high switching frequencies, slew rates, lower losses, and very compact device configurations. The Jhonson figure of merit of wide bandgap semiconductors is high compared to silicon, which shows that the semiconductor can be made fast and can still maintain its breakdown voltage. Soft switching techniques have greatly reduced losses (ideally to zero) as only the charging and discharging of the output capacitance C_{oss} contributes to switching losses [1]. Soft switching losses are challenging to model mathematically as they are very much dependent upon the employed gate driver and the PCB layout. The well-known and widely accepted double pulse test fails to measure switching losses because the accuracy and frequency behavior of current and voltage probes prevent accurate measurement [2]. Minor deskew causes a significant error in the measurement results. The calorimetric measurement methods [2]–[11] are the alternatives to the electrical measurement methods. They measure the switching losses with high accuracy with only 12-15 % [9] error depending upon the setup. The major challenge is the high time requirement to perform the test. Other challenges include the separation of conduction and switching losses as on-resistance ($R_{\mathrm{ds,on}}$) varies with temperature; the losses must be symmetrically distributed between the low and high sides; and proper temperature monitoring as the ambient temperature changes over time.

Calorimetric methods very much depend upon the thermal model of the system, so it becomes very crucial to design such a thermal model in which every heat dissipated parasitic element is included. The contribution of various parasitic elements in soft switching depends upon the specific setup, but the main common and major influences which can be present in most of the setups are kink current [12], [13], Gate resistance, switchnode capacitance [13], parasitic capacitance [14], Dead time shift [14]. The analytical relationship of these influences with respect to the temperature is to be modeled so that the loss measurement can be accurate. Apart from this, the temperature measuring sensor should be as close as possible, but this can add more parasitic influences. For example, it can sense heat from the driver circuit. And also, the errors from the sensor amplifier and analog to digital converter used during measurements should also be considered.

II. State-of-the-art calorimetric characterization techniques

The calorimetric measurement method essentially focuses on measuring the total power losses, which the device under test(DUT) dissipates as heat during the operation. This heat results in a temperature rise surrounding the DUT, and this temperature rise is an indication of the total power losses of the DUT if the heat is completely absorbed by the coolant, such as metal in the case of the C_{th} calibrated method. In general, calorimetric methods can be classified into two subcategories, i.e., steady state methods [3]–[7] and Transient methods [1], [8]–[11] and transient methods are further classified as Z_{th} [10] and C_{th} [8], [9] methods as shown in Fig.1. In the steady state methods such as the Air-flow method [3], the setup is operated until the thermal steady state is reached, which takes more time as compared to the transient calorimetric methods. On the other hand, transient methods analyzes the

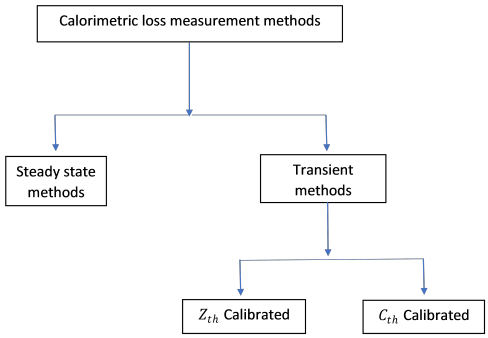


Fig. 1. Classification of Calorimetric methods

thermal dynamics such as thermal step response during turn on and turn off pulse, this increases the level of complexity for transient methods but offers reduced times and almost the same accuracy in comparison. The transient calorimetric methods have evolved over the years due to their shorter measurement times. The C_{th} calibrated method is one of the most widely used method.

A. Comparision of Different Calorimetric Methods

Different calorimetric methods can be compared based on various properties such as measurement time, cooling time, and accuracy of the measured losses. The Steady State Air flow method [3] has overall shown good accuracy, but the measuring times are almost 20–25 minutes, which is way more compared to the Z_{th} calibrated method. The C_{th} method(as described in [8], [9]) takes into account the heating of the heat capacity caused by the switching operation. A large metal block is used in the C_{th} method, as shown in the cross-section in Fig. 2(a), in order to linearize with the known caliberation curves, which increases the measurement time. In the Z_{th} method (as described in [10]), the heat capacity is smaller in comparison and also the placement of the temperature sensor is not in the metal block but it is placed on the silicon pad of the PCB as shown in the cross-section in Fig. 2(b). The C_{th} method has a comparitively smaller cooling time as compared to the Z_{th} method. The smaller measurement time of Z_{th} calibrated methods reduces the non-linear effects of various temperature-dependent effects such as $R_{\mathrm{ds,on}}$.

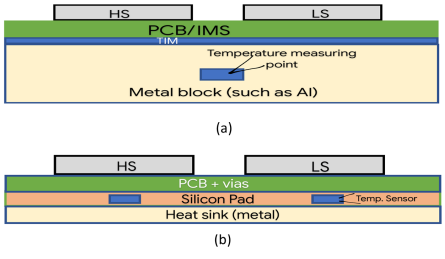


Fig. 2. (a) shows C_{th} cross-section, (b) shows Z_{th} cross-section (HS = high side, LS = low side

But due to the placement of the temperature sensor, electromagnetic interference effects the measured results, and in order to minimize this effect, a separate power source is given to the temperature sensor.

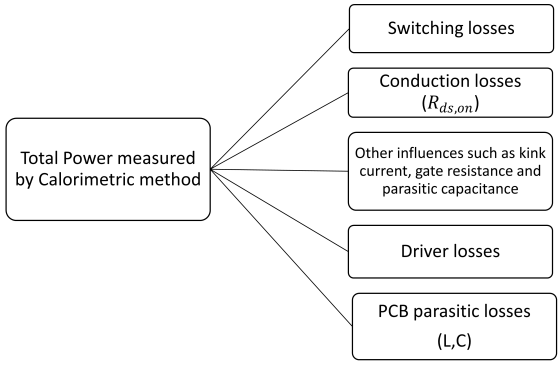


Fig. 3. power measured via calorimetric techniques

B. Power Loss Measurement

The energy measured using the calorimetric methods is the sum of switching energy, conduction energy losses as well as various other losses, including parasitic losses, as shown in the Fig. 3. The conduction losses are temperature dependent as $R_{\mathrm{ds,on}}$ is temperature dependent entity. PCB(Printed Circuit Board) parasitic losses and other losses totally depend upon the specific setup. All these losses need to be mathematically modelled in such a way that accurate switching losses can be extracted from the total losses, which are measured by a calorimetric setup. In the next section, an overview of all possible influences on soft switching identified in various setups are described in brief.

III. Various Parasitic Influences on Soft Switching

A. C_{oss} losses

Ideally, soft switching losses should be zero, but in practice, there are some losses, and these losses need to be calculated accurately in order to use the system in an efficient manner. The charging and discharging of output capacitance C_{oss} contributes significantly to soft switching losses. To validate this, soft switching measurements are done by adding external capacitance in parallel to the DUT as explained in [1]. Since the presence of external capacitance reduces the fraction of switching current flowing through the output capacitance, which in turn reduces the voltage slew rate for the same switching current. As shown in the Fig. 4. , the graph of switching energy V/s switched voltage slope shows the results with and without connecting external capacitance in parallel to the DUT. Surprisingly, the graph aligns (dots and triangles), indicating that the switching energy is proportional to the current flowing through the output capacitance (or to voltage slope), exactly as if the charging and discharging process were lossy, indicating that the charging and discharging of output capacitance C_{oss} accounts for the majority of soft switching losses.

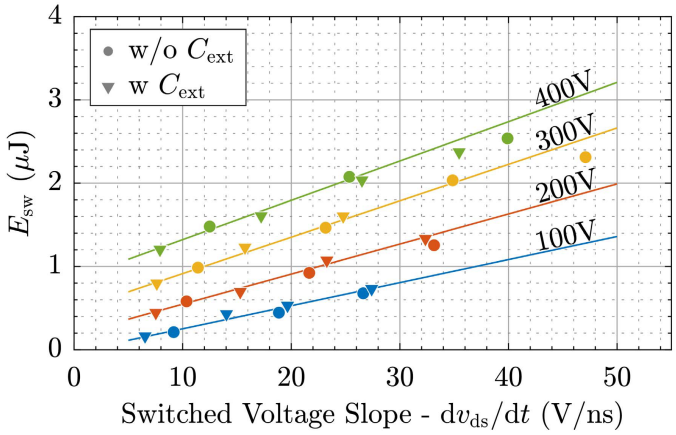


Fig. 4. Switching energy E_{sw} as a function of switched voltage slope with different switched currents(5A...20A for different DC link voltages (100V...400V), adapted from [1]

B. Kink Current(Residual ZVS Losses)

To explain this phenomenon, we consider a zero voltage switching(ZVS) half-bridge converter as shown in Fig. 5. The thermal analytical models predict losses with almost 12-15% deviation at nominal load, including residual ZVS losses and external gate resistance. Switching losses are higher than expected and also increase with current. This shows that there are additional loss mechanisms apart from C_{oss}. The equivalent circuit diagram and the turn off transition waveform of switching of the low side are adapted from [12].

Here, we are assuming that the inductor current is flowing during the whole switching transition and the dead time is large enough to ensure a zero voltage switching (ZVS). The capacitance between drain and source shows the equivalent capacitance including parasitic capacitance($C_{\text{parasitic}}$) and the output capacitance of the high side($C_{\text{oss,Th}}$) and low side($C_{\text{oss,Tl}}$).

$$C_{\text{ds}} = C_{\text{ds,Tl}} + C_{\text{oss,Th}} + C_{\text{parasitic}} \tag{1}$$

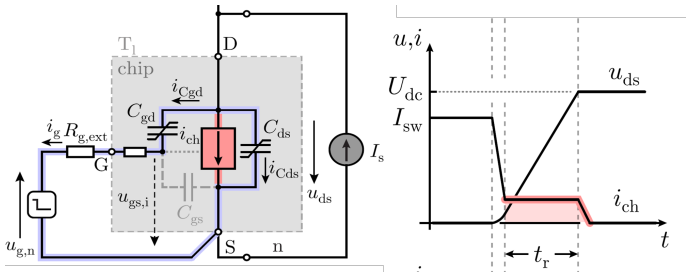


Fig. 5. current is more than the kink current, overlap happen as shown and residual ZVS losses are added, adapted from [12]

During the turn-off transition, the current (I_{sw})which is constant and commutates from channel current (i_{ch})to charge drain-to-source capacitor and miller capacitor. As gradually the current (I_{sw}) increases, the switching voltage transition is expected to occur more rapidly due to higher capacitor charging current. This increases the voltage slew rate upto a certain level. As current (I_{sw})increases miller current ($I_{\text{C}_{\text{gd}}}$)increases according to the current division between miller capacitance and gate-to-drain capacitance.

$$i_{\text{g}} = \frac{U_{\text{gs,i}} + U_{\text{g,n}}}{R_{\text{g}}}; \tag{2}$$

$$R_{\text{g}} = R_{\text{g,internal}} + R_{\text{g,external}} \tag{3}$$

$U_{\text{gs,i}}$ = Internal gate-to-source voltage, $U_{\text{g,n}}$= negative gate driver voltage, R_{g}= total gate resistance, i_{g} = gate current.

As long as $i_{\text{Cgd}} \leq i_{\text{g}}$, gate-to-source capacitance (C_{gs})discharges and channel is closed before drain-to-source voltage (U_{ds}) starts rising. This is the ideal case where there are no residual turn off losses apart from C_{oss} losses. The slew rate is given as:

$$\frac{dU_{\text{ds}}}{dt} = \frac{I_{\text{sw}}}{C_{\text{ds}} + C_{\text{gd}}} \tag{4}$$

When $i_{\text{Cgd}} = i_{\text{g}}$, $U_{\text{gs,i}} = U_{\text{th}}$, then the current is called kink current. As the gate-to-source voltage is clipped at the threshold voltage, no gate charge is removed from the gate source capacitor. If $I_{\text{sw}} > I_{\text{kink}}$, the gate source voltage is discharged up to a voltage slightly higher than the threshold voltage and the channel remains turned on and conducts the switching current. This voltage current overlap is shown in the Fig. 5 causes the ZVS residual losses.

As soon as the voltage transient is completed, the gate-source voltage drops further, the device turns off fully, the remaining channel current commutates to the high-side device, and then the transition is completed. The gate resistance should be as low as possible to keep the kink current high. As gate resistance is low, gate current will be high, so to achieve $i_{\text{Cgd}} = i_{\text{g}}$, kink current will be high. The gate resistance is inversely proportional to the kink current.The level of kink current is to be as high as possible in order to avoid the overlap and thereby the residual ZVS losses. So a low gate resistance and a large negative gate driver voltage magnitude push out the kink current, but by increasing the negative gate driver voltage, gate driver losses may increase.

While this derived model predicts the behaviour accurately, the loss estimation is difficult to implement analytically as there is a discontinuity at kink current. The residual ZVS losses can instead be modeled as a quadratic loss model:

$$E_{\text{sw}}(I_{\text{sw}}) = a + bI_{\text{sw}} + cI_{\text{sw}}{}^2 \tag{5}$$

The kink current derivation, which can be understood in the context of the current paths during switching, helps us arrive at a critical finding for the broader field: that switching losses in a half bridge are not quadratic with current, but are rather piecewise linear on current.

C. Miller Capacitance

This is to show the effect of miller capacitance in switching rather than the parasitic influence. The next generation Variable speed drives(VSD) with SiC MOSFETS has created complications due to its faster switching speeds, i.e., slew rates [13]. High voltage slew rate at the output of the VSD can cause partial discharge phenomena and it can lead to the gradual destruction of power cables as well as the winding of the motor. Surge voltage formed due to reflections can have a huge impact on the life of the motor and power cables. So the VSD system must meet critical standards to protect industrial motors from deleterious effects. One of the methods to reduce the voltage slew rate by a small factor is to control the gate driver circuit directly. In order to do so, gate resistance and Miller capacitance can be varied to reduce the voltage slew rates [13]. by increasing the gate resistance, leading to slower switching transients but higher switching losses. An improvement to this approach is to increase the miller capacitance, which influences the transient of switchnode voltage but, unlike the gate resistor, it does not have any meaningful effect on gate time constant. An additional miller capacitance approach prevents fast voltage transients instead of filtering them out, which linearizes the gate voltage. And the size of the capacitance is small and can be adjusted within the circuit. In such cases, the miller capacitor concept is simple and compact to implement.

D. Parasitic Capacitance and Dead Time

These parasitic influences mainly impact measurement accuracy due to the small changes in the switching current [14]. The switching frequency (f_{sw}) results from the load inductance (L) and the desired switching current (I_{Δ}) according to:

$$f_{\mathrm{sw}} = \frac{U_{dc}}{8LI_{\Delta}} \tag{6}$$

For the easy separation of conduction losses and switching losses, conduction losses are kept constant by changing the load inductance according to the switching frequency. The change in load inductance causes different switching behaviors due to the change in winding capacitances. To understand this phenomenon in more detail, calorimetric dead time dependent power loss measurements were done for different additional switch node capacitances. The soft switching duration increases as shown in Fig. 6.

At high frequencies, the switching duration becomes large as compared to the conduction period. This actually lowers soft switching losses with more influence of parasitic capacitance, and the dead time is shifted [14]. By taking this frequency-dependent shift into account, switching energy measurement can be done over the entire range of frequencies as shown in the equation 7 :

$$f_{\mathrm{sw}} = \frac{U_{\mathrm{dc}}(1 - D_{\mathrm{eff}}(t_{\mathrm{dead}}))}{4LI_{\Delta}} \tag{7}$$

Effective duty cycle (D_{eff}) according to the switching frequency by keeping the switching current constant with high

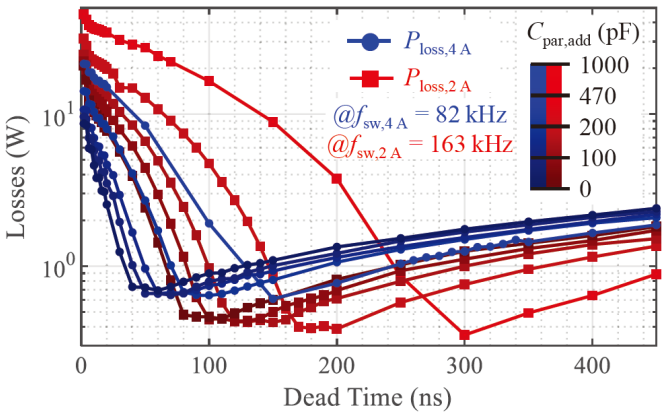


Fig. 6. Different total losses for different values of switchnode capacitance, adapted from [14]

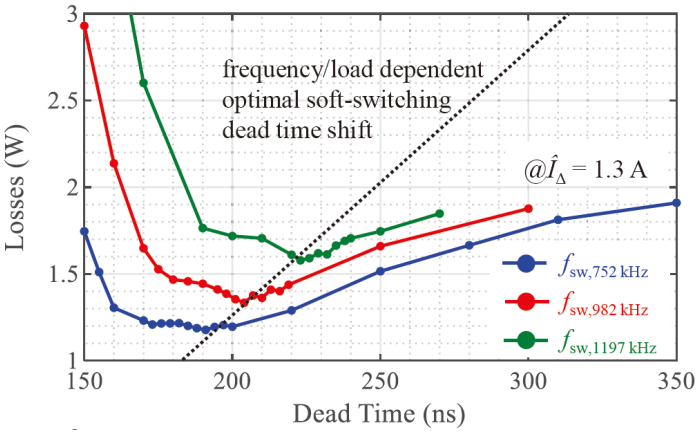


Fig. 7. Frequency dependent soft switching dead time shift, adapted from [14]

accuracy can be achieved. A frequency-dependent optimal dead time (t_{dead}) shift is shown in Fig. 7.

IV. Conclusions

The calorimetric methods mainly rely on the accuracy of the thermal model in order to get the accurate switching loss measurements. The transient calorimetric measurement calibrated by thermal impedance promises an extremely fast and non-invasive loss characterization in a wide power range without any modification to the circuit. By measuring losses with calorimetric methods, the switching losses have an error of almost 12-15%. This error is due to the various parasitic influences at such high frequencies of switching. The proposed paper examines the state-of-the-art calorimetric methods. Furthermore, it describes the major parasitic influences and their analytical relationship, which can be modeled thermally to calculate accurate switching losses. The parasitic influences described in Section III are proven in various setups and analytical relations are given, but apart from these, there can be some more parasitic elements, which is an ongoing topic of research.

REFERENCES

[1] M. Guacci et al., "On the Origin of the C_{oss} -Losses in Soft-Switching GaN-on-Si Power HEMTs," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 7, no. 2, pp. 679-694, June 2019, doi: 10.1109/JESTPE.2018.2885442.

[2] D. Rothmund, D. Bortis and J. W. Kolar, "Accurate Transient Calorimetric Measurement of Soft-Switching Losses of 10-kV SiC mosfets and Diodes," in IEEE Transactions on Power Electronics, vol. 33, no. 6, pp. 5240-5250, June 2018, doi: 10.1109/TPEL.2017.2729892.

[3] J. A. Anderson, C. Gammeter, L. Schrittwieser and J. W. Kolar, "Accurate Calorimetric Switching Loss Measurement for 900 V 10 m Ω SiC mosfets," in IEEE Transactions on Power Electronics, vol. 32, no. 12, pp. 8963-8968, Dec. 2017, doi: 10.1109/TPEL.2017.2701558.

[4] D. Bortis, O. Knecht, D. Neumayr and J. W. Kolar, "Comprehensive evaluation of GaN GIT in low- and high-frequency bridge leg applications," 2016 IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia), 2016, pp. 21-30, doi: 10.1109/IPEMC.2016.7512256.

[5] K. Surakitbovorn and J. R. Davila, "Evaluation of GaN transistor losses at MHz frequencies in soft switching converters," 2017 IEEE 18th Workshop on Control and Modeling for Power Electronics (COMPEL), 2017, pp. 1-6, doi: 10.1109/COMPEL.2017.8013330.

[6] H. Li, X. Li, Z. Zhang, J. Wang, L. Liu and S. Bala, "A simple calorimetric technique for high-efficiency GaN inverter transistor loss measurement," 2017 IEEE 5th Workshop on Wide Bandgap Power Devices and Applications (WiPDA), 2017, pp. 251-256, doi: 10.1109/WiPDA.2017.8170556.

[7] S. Bolte, L. Keuck, J. K. Afridi, N. Fröhleke and J. Böcker, "Calorimetric measurements with compensating temperature control," 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), 2017, pp. 636-639, doi: 10.1109/ISIE.2017.8001320.

[8] D. Neumayr, M. Guacci, D. Bortis and J. W. Kolar, "New calorimetric power transistor soft-switching loss measurement based on accurate temperature rise monitoring," 2017 29th International Symposium on Power Semiconductor Devices and IC's (ISPSD), 2017, pp. 447-450, doi: 10.23919/ISPSD.2017.7988914.

[9] D. Koch, S. Araujo and I. Kallfass, "Accuracy Analysis of Calorimetric Loss Measurement for Benchmarking Wide Bandgap Power Transistors under Soft-Switching Operation," 2019 IEEE Workshop on Wide Bandgap Power Devices and Applications in Asia (WiPDA Asia), 2019, pp. 1-6, doi: 10.1109/WiPDAAsia.2019.8760332.

[10] J. Weimer and I. Kallfass, "Soft-Switching Losses in GaN and SiC Power Transistors Based on New Calorimetric Measurements," 2019 31st International Symposium on Power Semiconductor Devices and ICs (ISPSD), 2019, pp. 455-458, doi: 10.1109/ISPSD.2019.8757650.

[11] A. Anurag, S. Acharya and S. Bhattacharya, "An Accurate Calorimetric Loss Measurement Method for SiC MOSFETs," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 8, no. 2, pp. 1644-1656, June 2020, doi: 10.1109/JESTPE.2019.2920935.

[12] Haider, M., Anderson, J. A., Nain, N., Zulauf, G., Kolar, J. W., Xu, D., & Deboy, G. (2021). "Analytical calculation of the residual ZVS losses of tcm operated single-phase PFC rectifiers." IEEE Open Journal of Power Electronics, 2, 250-264

[13] Haider, M., Fuchs, S., Zulauf, G., Bortis, D., Kolar, J. W., & Ono, Y. (2022). "Analytical Loss Model for Three-Phase 1200V SiC MOSFET Inverter Drive System Utilizing Miller Capacitor-Based dv/dt-Limitation." IEEE Open Journal of Power Electronics.

[14] J. Weimer, D. Koch, R. Schnitzler and I. Kallfass, "Determination of Hard- and Soft-Switching Losses for Wide Bandgap Power Transistors with Noninvasive and Fast Calorimetric Measurements," 2021 33rd International Symposium on Power Semiconductor Devices and ICs (ISPSD), 2021, pp. 327-330, doi: 10.23919/ISPSD50666.2021.9452230.

Active vs. Passive Power Detectors for mmW Applications

Electrical Engineering
Universität Stuttgart
Stuttgart, Germany

Abstract— A detailed analysis of Active and Passive Power Detectors in Telecommunication System. This paper outlines advantages and disadvantages of Active and Passive approaches. This paper further extends to capture the state-of-the-art detectors for mmW applications in HEMT-technology.

Keywords— Active and Passive Power Detectors, mmW, HEMT, RF

I. INTRODUCTION

The Major parts of wireless communication systems that transform an RF signal into an output DC voltage that is proportionate to the RF input Power are called Power Detectors. Many microwaves and millimeter wave (MMW) applications, including imaging, target detection, testing and the shielding of delicate circuits from spikes and pulses, require power detectors. Bipolar, Field Effect transistor and Schottky-diode technologies can all be used as PDs. Active and Passive PDs are two common varieties. Diodes are one of the easiest ways to detect power because they take advantage of their non-linear properties to produce an output signal with numerous frequency components. With the right filtering, the DC component can be separated and can be detected as a part of power monitoring [8]. Due to its shorter wavelength and wide frequency bands, which enable smaller components with better capabilities, the millimeter-wave (mmW) band (30-300GHz) is currently experiencing an increase in attention. The need to operate equipment at higher frequencies is a result of the development of wireless communication technology. High Electron Mobility Transistor or HEMT, is one technique that is utilized to get around this restriction. High frequency operation requires the usage of HEMTs, which also provide a very efficient and low noise performance. The Key component of a HEMT, which is a type of FET or Field Effect Transistor that uses several materials on either side of the junction, is the PN junction. Aluminium gallium arsenide (AlGaAs) and gallium arsenide are the most often used materials (GaAs) [6].

II. RF POWER DETECTORS

Controlling transmitted power efficiently to reduce power consumption and RF signal interference with other electronic equipment is one of the key tasks of RF devices. In order to measure the power level of a signal, an RF power detector transforms a sinusoidal RF signal into a DC voltage [10]. Since power management is one of the crucial elements in automated gain control and automatic level control to maintain optimal output levels, RF power detectors must be a component of a transmitter.

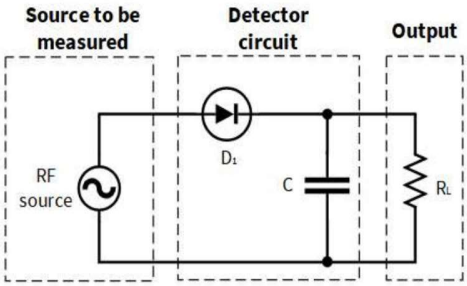


Fig. 1: Schematic diagram of a diode-based power detector [10].

The output power of equipment that operates in the RF or microwave frequency band has a big impact on the circuit's design and functionality. Signal level is crucial for comprehending circuit components and assessing system performance, therefore components in a system must receive an undistorted signal level from their predecessors and transmit the appropriate signal level to the component in the following row. If an output signal is too low, it cannot be identified and blends in with the background noise. If the signal level becomes excessive, the performance becomes nonlinear and produces distortion [3].

III. PASSIVE POWER DETECTORS

A diode's ability to correct is used by passive power detectors. Diodes use an integrator to transfer the rectified signal after rectifying the AC signal in order to retrieve the DC component. Figure 1 depicts the single diode detector's schematic. The capacitor C is selected so that it is big enough and that its capacitive reactance is negligibly low in comparison to the diode's impedance. The capacitor ensures that the diode has a strong RF short-circuit and that all RF voltages are visible across the diode terminals. The circuit's detecting speed is controlled by the load resistors R, L, and C [4].

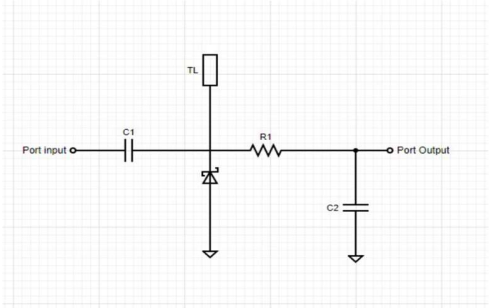


Fig. 2; Schematic diagram of Schottky diode based power detector

A Schottky diode serves as the primary component of the power detector circuit, and its properties are comparable to those of a normal PN diode in terms of both current and voltage behaviour. Schottky diodes have a lower forward voltage drop (0.15 V to 0.45 V) than PN diodes, which is advantageous (0.7 V to 1.7 V). A Schottky diode conducts via charge transfer from the majority carriers across the barrier, resulting in rapid switching action and making them extremely beneficial for RF and microwave rectification. The low recombination velocity of these minority carriers in the space charge region causes problems for PN junction diodes, which also operate on minority charge carriers. Due to its silicon low barrier N-type construction, small forward voltage, low junction capacitance, and low barrier height, RF Schottky diodes are an ideal option for power sensing and mixing tasks at high frequencies [5].

IV. ADVANTAGES

- Passive Circuits are simple in design.
- Passive Detectors operate at very high frequencies.
- They operate at large dynamic range (-70 to +20dBm).
- More diodes are used in rectification.

DIS ADVANTAGES

- The operation exploits the diodes non-linear characteristic, but this is not the same in entire dynamic range.
- The Sensitivity (that is how much DC voltage generated for a fixed RF input power) is limited.

V. ACTIVE POWER DETECTORS

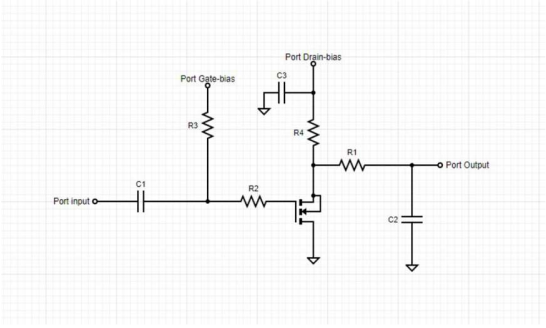


Fig. 3: Schematic diagram of Transistor based power detector

Active Power Detectors rectify using an analog amplifier. To conduct an equal or less than one halfwave of the RF signal, the transistor must have a class B or C bias point. A resistor is used to load the transistor and bias the drain to the same voltage as the amplifier. The transistor's gate receives the same RF signal, and this signal's rectified form appears as the drain current waveform. The RF power level is indicated by the DC component of the drain voltage. Due to the fact that the detector output is exactly proportional to the RF signal power for low input power levels, drain current is reliant on the square of gate voltage [3].

VI. ADVANTAGES

- Active PDs have the advantage of a higher sensitivity.
- Active PDs are used for monitoring power levels.

DISADVANTAGES

- These type of PDs needs more area on the chip
- Active PDs tends to increase in size as the operating frequency decreases.

VII. HEMT TECHNOLOGY

Unlike most field-effect transistors (FETs), which typically have a doped region as the conducting channel, high electron mobility transistors (HEMTs) have a junction between two materials with different band gaps (i.e., a heterojunction) as the conducting channel. In ICs, HEMTs are utilized for digital switching. HEMT transistors are capable of operating at frequencies greater than that of conventional transistors, which are millimetre wave frequencies. Due to their minority charge carriers, HEMTs offer quick switching speeds, high gain, and low current fluctuation characteristics [11]. A form of HEMTs utilized in wireless communications are called pHEMTs devices. High power efficiency and exceptional low noise value and performance are provided by pHEMT transistors. Another variety of pHEMT that finds usage in numerous applications is the metamorphic high electron mobility transistor (mHEMT).

VIII.STATE-OF-THE-ART DETECTORS FOR MMW APPLICATIONS IN

A. Two types of microwave and millimetre power detectors realized considering commercial metamorphic high electron mobility transistor (mHEMT) monolithic microwave integrated circuit (MMIC) process.

1. Schottky Diode (Passive) Detector

The simplified schematic representation of the Schottky detector is shown in Figure 1. The RF is applied to the diode through a 170 fF MIM coupling capacitor, an open stub is used for minimizing the reflection coefficient at 60 GHz [2]. Bias current can be supplied to the diode through a 900-ohm resistor. This resistor forms a simple low pass filter with capacitor C2 (220 fF), which is connected to ground. The Schottky diode consist of two paralleled gate fingers, each 20-um wide. The diode is modelled by the foundry model, based on the enhancement high electron mobility transistor (EEHEMT) model.The output voltage was simulated as a function of input power at different frequencies and bias currents, and optimized for frequencies above 50 GHz[5]. In Figure 5, the measured output voltage is depicted for a constant bias current of 1 uA. The bias current is supplied through the output port by an external current source. At no input power the output voltage is of the order -710mV which is the voltage drop across the diode and R1. A sensitivity of 500 V/W is obtained at 60 GHz at -20 dBm. The incremental detector voltage, defined as the voltage difference with and without a certain RF-power, is plotted in Fig 6.

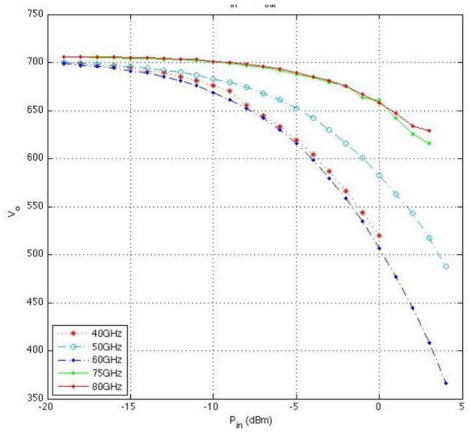


Fig. 4: Measured detector output voltage at 1 uA biascurrent as a function of Frequency [3].

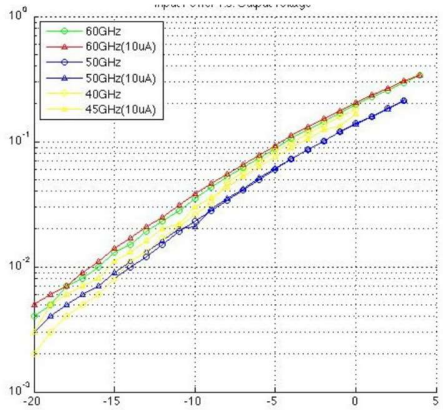


Fig. 5: Measured incremental detector output voltage at 1 and 10 uA bias current as a function of input power in dBm and frequency [3].

2. Active Detector

Figure 3 depicts the active power detector's simplified schematic form.

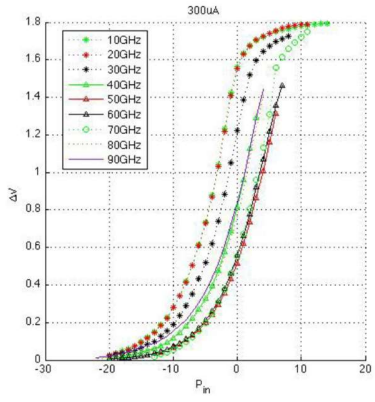


Fig.6: a [5]

A MIM-capacitor C1 in series with a resistor R2 couples the gate of the 2x20 um gate width mHEMT device to the drain quiescent current IDD, which is controlled by the gate bias voltage, VGG (37 Ohm). The drain is connected to VDD through resistor R4 (750 ohm). to bring VGG as near as feasible to the pinch off voltage[1]. Capacitor C2 (0.66pF) is a part of the lowpass filter. As a function of frequency and bias settings, the characteristic of the active detector was tuned. The incremental output voltage was measured and plotted in Fig. 6 a, b at VDD=2V, IDD=15, and 300 uA. In comparison to 15 uA, the 300 uA bias has a better sensitivity, but a far wider variance with frequency[3]. At 20 GHz, the greatest sensitivity is 2800V/W. 200 uA or so was chosen as the bias current. The outcome is displayed in Fig 7. It is possible to achieve linear operation from lowest power up to -10dBm [6].

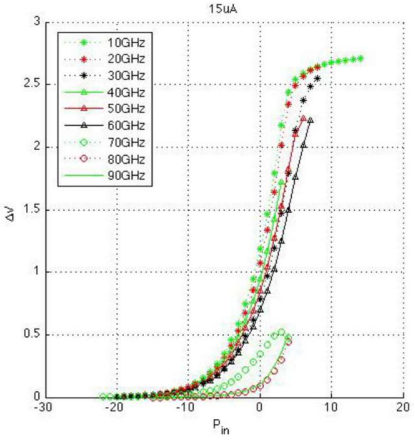


Fig 6:b

Fig. 6: Measured incremental detector output voltage at 15 and 300 uA drain current and VDD=2V. The frequency is 10 to 90 GHz in 10 GHz steps [5].

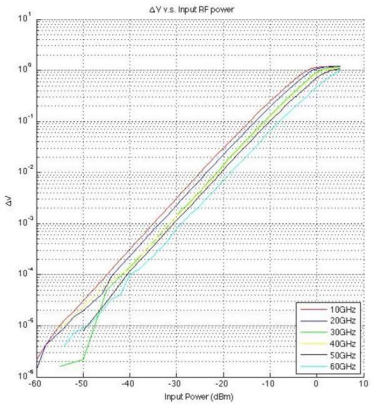


Fig. 7: Output incremental voltage as a function of input power and Frequency [7].

B. AlN/GaN-based material designs for mm-Wave applications

GaN has a 3.4 eV greater energy gap than InP, GaAs, or Si, which allows for higher breakdown voltages and operating voltages. GaN also has a high-saturated electrical velocity of 2.5107 cm/s. Due to the relationship between the electron velocity and current density, GaN generates high current at high voltage. Due to the energy gap and the high electron velocity, high power devices can be practically optimal since power is a function of voltage and current. GaN has a thermal conductivity that is significantly higher (0.4761-1.4285 W/cm.K for GaN)

than GaAs and InP[6]. High power and frequency applications benefit from GaN, but it also reduces chip cost and size. GaN MMICs now outperform GaAs MMICs in size and power density by a factor of five. However, two important phenomena continue to afflict GaN HEMTs, especially when the size of the device is reduced [4]. These phenomena include kink effects, which reduces device performance, and entrapment effects, which can directly lead to current collapse. See Figure 8.

C. GaAs-based material designs for mm-Wave aapplications.

High power amplifiers based on III-V materials like gallium arsenide are primarily developed in the mmW range (GaAs). Due to their high electron mobility and potential for the development of a two-dimensional electron gas layer (2DEG), these materials show promise for use in high frequency devices. This results in lesser effects of collisions and makes switching quicker. The power detectors would have a large dynamic range, and can be fabricated as a chip on the semiconductors. Schottky diodes are known for its fast-rectifying property and they can be used for RF detection, GaAs HEMTs are used for its higher electron mobility[1].

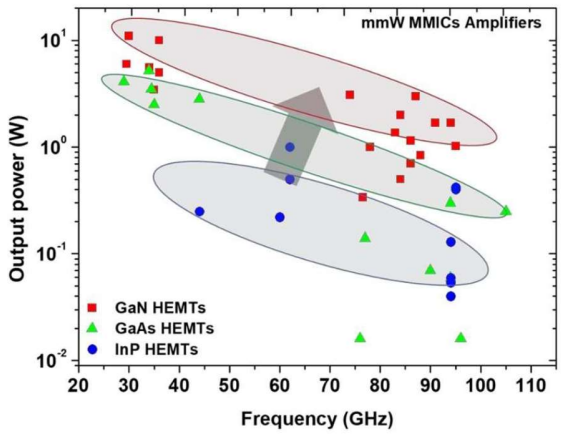


Fig. 8. MMIC amplifier output power in CW mode using several semiconductor transistor technologies [2].

X. CONCLUSION

The paper presents a detailed review on active and passive power detectors, a brief overview of their working principle, also the advantages and dis-advantages. Passive detectors mainly deal with high frequency and are used in applications like radar and radio astronomy, output power sensing and levelling. Active power detectors are used predominantly in low powered frequency applications like cell phones, electronic warfare systems and envelope detectors.

REFERENCES

[1]. Microwave Integrated Circuits Conference, 2010, pp. 353-356 F. Díaz Canales and M. Abbasi, "A 75–90 GHz high linearity MMIC power amplifier with integrated output power detector," 2013 IEEE MTT-S International Microwave Symposium Digest (MTT), 2013, pp. 1-4, doi: 10.1109/MWSYM.2013.6697600.

[2]. Ayssar Serhan, Estelle Lauga-Larroze, Jean-Michel Fournier."Sub-threshold based powerdetector for low-cost millimeter-wave applications", 2014 10th Conference on Ph.D.Research in Microelectronics and Electronics(PRIME), 2014.

[3]. Navid Piri Yengijeh, Hassan Moradi CheshmehBeigi, Amin Hajizadeh. "Inertia emulation with the concept of virtual supercapacitor based on SOC for distributed storage systems in islanded DC microgrid", IET Renewable Power Generation, 2022.

[4]. D. Lopez-Diaz. "Monolithic integrated 210 GHzcouplers for balanced mixers and image rejection mixers", 2010 Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits, 04/2010.

[5]. P. Colantonio, E. Cipriani, F. Giannini. "Microwave power amplifiers", Elsevier BV, 2016

[6]. "Nitride Semiconductor Technology", Wiley,2020.

[7]. Inder J. Bahl. "Fundamentals of RF and Microwave Transistor Amplifiers", Wiley, 2009.

[8]. HAL Open Science, URL:https://hal.archivesouvertes.fr/?lang=en

[9]. Chalmers Publication, URL:https://www.chalmers.se/en/research/publications/Pages/publications.aspx

[10]. InfineonTechnologies, URL: https://www.infineon.com/dgdl/Infineon

[11]. Wikipedia, URL: https://en.wikipedia.org/wiki/High-electron mobility_transistor

Novel Microstripline Topology for High-Frequency Wide Bandgap Semiconductor Technologies

Institute of Robust Power Semiconductor Systems, University of Stuttgart, Germany

***Abstract*—This paper presents various structures and arrangements that cause the slow wave effect. A slotted ground plane between 0-20 GHz, a helix structure between 5-45 GHz, two different meander lines between 12-45 GHz and 68-105 GHz and four different arrangements of metal layers and microstrips between 200-400 GHz are investigated. The insertion losses can be reduced by the shift of the magnetic field under the signal line, and this results in a reduction of the phase velocity [1]. The frequencies investigated are presented and their slow-wave factors and S-parameters are compared. The helix structure has the smallest insertion loss of -28 dB and a return loss of -0.1 with a slow wave factor of 5.93 at 43 GHz. (*Abstract*)**

Keywords—microstrip-line, slow-wafe effect, defected ground plane, ground plane pattern

I. INTRODUCTION

Currently, the focus is on making structures on chips and devices smaller and smaller. This is because the area required by circuits is increasingly growing as the number of functions and applications also increases. Since a large part of the area must be used for lines to fulfil certain phase relationships between the connections, the structures can be made smaller by reducing the size of the microstrip conductors. To be able to achieve this, slow-wave structures are used [2, 3, 1]. Slow-wave structures as a slotted ground plane, a helix structure, a polyline structure, two different meander line structures as a rectangular or a V-shaped meander line and an interdigital planar structure can cause the slow wave effect. They influence the slow wave effect to different degrees at different frequencies.

II. DIFFERENT IMPLEMENTATIONS OF SIGNAL LINES

In the following, various methods that cause this effect are discussed [4, 5]. In this section, we first look at the different parameters that favour the slow wave effect. Both the materials used, and the parametric of signal lines cause the slow wave effect [4].

A. Permeability

The magnetic field below the signal line is displaced by the slow wave structure, resulting in a reduction of the phase velocity and a lower insertion loss compared to conventional microstrip lines. The permeability in the upper part is higher than in the lower part. By introducing a slow-wave structure, the return currents cannot flow directly under the signal line but must flow laterally towards the continuous ground plane. Since the structures have different impedances, the resulting phase velocity v_{ph} can be calculated with capacitance C' and inductance L' per unit length as follows [4, 1]:

$$v_{ph} = \sqrt{\frac{1}{L'C'}} = \frac{c}{\sqrt{\mu_r \varepsilon_r}} \qquad [2]$$

In Figure 1 a slow wave structure as a substrate integrated patterned with microstrip polyline is shown.

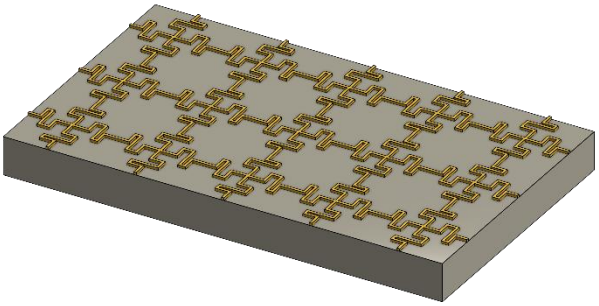


Figure 1: Microstrip Polyline Pattern

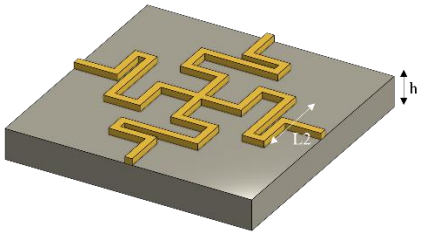


Figure 2: Single Polyline Cell

B. Parametric

The most important parameters of a polyline pattern that influence the slow wave effect are the thickness of h and the meander line *L2* shown in Figure 2. It is observed that the

thinner h becomes, the lower the cut the phase velocity becomes. This effect is caused by the change in permeability. If the upper layer of the substrate, which has a high permeability, is not changed with a constant pattern, while the lower substrate with a lower permeability is reduced, the average permeability is increased. By compressing the meander lines, the induction in the upper surface layer can also be increased, thus increasing the average induction. Thus, the phase velocity and the cut off frequency can be reduced [4].

III. Slow wave Structures

First, different arrangements of the microstrip line patterns are considered. Depending on the arrangement, the strength of the slow wave effect differs. The slow-wave factor n indicates the relationship between the speed of light c_0 and the phase velocity v_{ph} and is calculated with

$$n = \frac{c_0}{v_{ph}} \quad (1) \text{ [2].}$$

In Figure 3. all the different structures that are examined are shown in their respective operating frequencies.

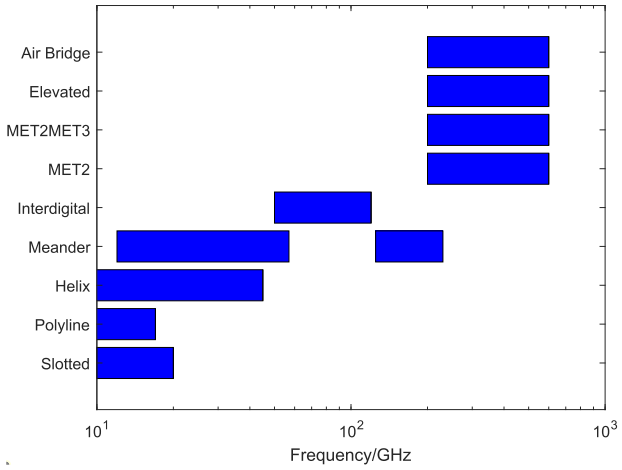


Figure 3: Investigated frequencies in the range 0 GHz to 400 GHz

A. *Interdigital Planar Slow Wave Structure*

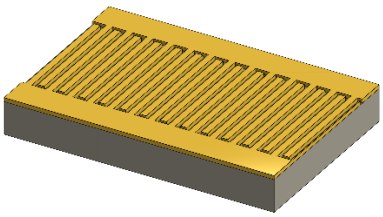


Figure 4: Interdigital Planar Slow Wave Structure

For the simulations in the V-band, the frequencies for 50 GHz up to 70 GHz are investigated with quartz and polycor as substrate material. The dimensions of the structure remain the same and the structure itself is shown in Figure 4. For the quartz substrate, the pin length is 825 µm and is longer than for the polycor substrate, which has a length of 575 µm. The cold test results in a VSWR of less than 2.0 in the range from 53.56 GHz to 78.33 GHz and insertion losses of less than -5 dB. It is also found that the slow wave factor with increasing frequency increases exponentially to the frequency [2].

B. *Meander Slow Wave Structure*

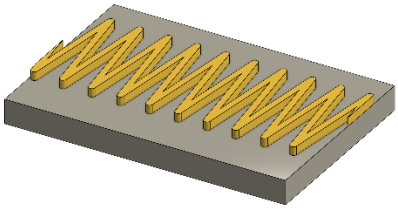


Figure 5: V-shaped Meander Line

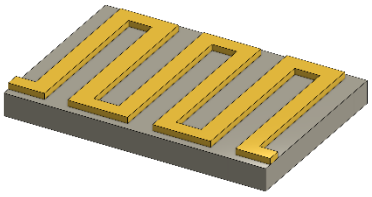


Figure 6: Rectangular Meander Line

Meander slow wave structure are investigated in the G-band from 50 GHz to 70 GHz which is seen in Figure 6 and from 180 GHz to 230 GHz with a structure which is seen in Figure 5 . From 180 GHz to 230 GHz results in a slow wave factor of 7 to 11 [6]. A wide passband at 50 GHz and a centre frequency of about 210 GHz is achieved from 50 GHz to 70 GHz. Furthermore, the cold test results in a VSWR of less than 2.0 in the range from 53.56 GHz to 78.33 GHz and transmission losses of less than -5 dB. For the frequency from 50 GHz to 70 GHz, a slow wave factor of 7 to 10 is obtained [2].

For a V-shaped meander slow wave structure, the range from 125 GHz to 155 GHz has been investigated. Here, the return loss S_{11} is less than -22 dB and the insertion loss S_{21} is greater than -0.4 dB [3].

C. Helix Slow Wave Structure

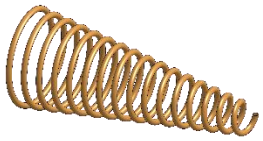


Figure 7: Pseudoperiodic Helix SWS

The slow wave structure of a helix has many advantages: a low operation voltage, a wide bandwidth, and a high RF efficiency. However, above a frequency in the W-band, the tolerances have more influence on the performance [3]. So far, various frequencies used at helix structures are being investigated, which were in the 33 GHz to 36 GHz, X-band, 17.3 GHz to 18.4 GHz and 12.75 GHz to 14.8 GHz ranges. In addition, a pseudo-periodic helix slow wave structure is shown in Figure 7 and implemented in a broadband helix travelling-wave tube and is investigated in the 8 GHz to 18 GHz. The radius of the helix becomes smaller with each revolution. The electromagnetic field is strictly periodic. This results in a VSWR of less than 1.85 over the range 7 - 18 GHz [7].

D. Slotted Ground Planes

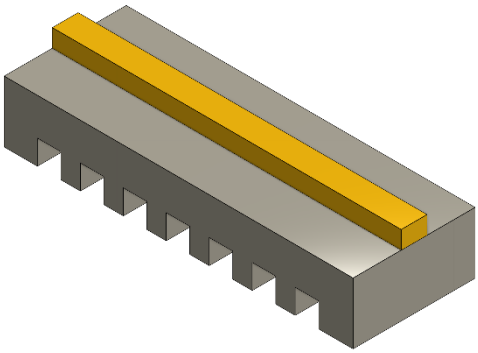


Figure 8: Slotted Ground Plane with Transmission Line

A slotted ground plane can be used to induce the slow wave factor. Frequency up to 20 GHz are investigated. Its structure is shown in Figure 8 [8]. The loss per quarter wavelength can be reduced by up to 54 % with such a structure. The length of a transmission line can be reduced by over 20% [1].

A. Different Arrangements of Metal Layers and Microstrip lines

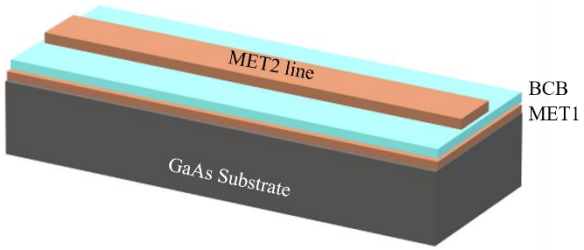


Figure 9: MET Line on a BCB MET TFMSL

Different arrangements of metal layers (MET) and thin film microstrip lines (TFMSL), a benzocycobulane layer (BCB) and a coplanar waveguide (CPW) are investigated. First, a MET2 thin film microstrip line (TFMSL) is investigated. It has a BCB2 layer with a thickness of 1.4 µm and is shown in Figure 9. With a resistance of 50 Ω, the MET2 TFMSL must have a length of 3.5 µm. From 200 GHz to 400 GHz over a length of λ/4, it has a maximum loss of a good 0.92 dB and a minimum loss of about 0.7 dB with a line impedance of 50 Ω [9].

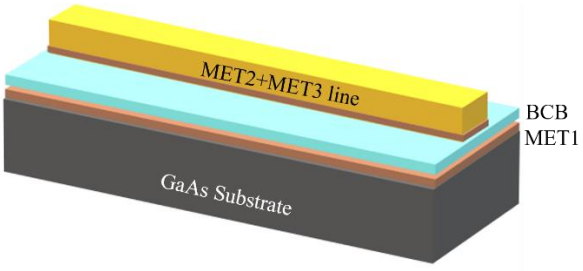


Figure 10: Met2+Met3 TFMSL

A MET2MET3 TFMSL is shown in Figure 10Tabl. MET3 has the thickness of 2.7 µm and consists of Au. It also has a BCB2 layer with a thickness of 1.4 µm. With a resistance of 50 Ω, the MET2 TFMSL must have a length of 2.7 µm and is thus shorter than MET2 TFMSL with the same impedance. From 200 GHz to 400 GHz, it also has a lower maximum loss than MET2 TFMSL of a good 0.7 dB over a length of λ/4 and a minimum loss of about 0.55 dB at a line impedance of 50 Ω [9].

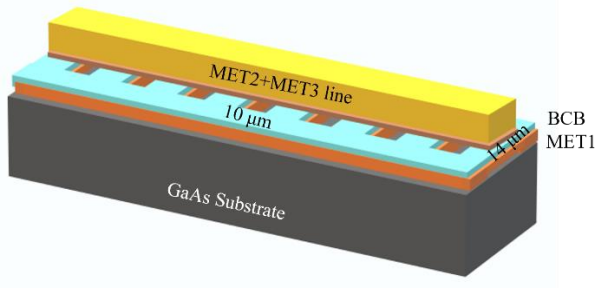


Figure 11: Elevated CPW

An elevated coplanar waveguide (CPW) with GaAs substrate is investigated. It also has a BCB2 layer and an air-

bridge of 10 µm. It is shown in Figure 11. With a resistance of 50 Ω, the elevated CPW must have a length of 8.6 µm and is thus shorter than both lines before with the same impedance. From 200 GHz to 400 GHz, it also has a lower maximum loss of about 0.35 dB over a length of λ/4 than both lines at 200 GHz and a minimum loss of just under 0.3 dB at 400 GHz with a line impedance of 50 Ω. The loss curve has a smaller drop than both lines before [9].

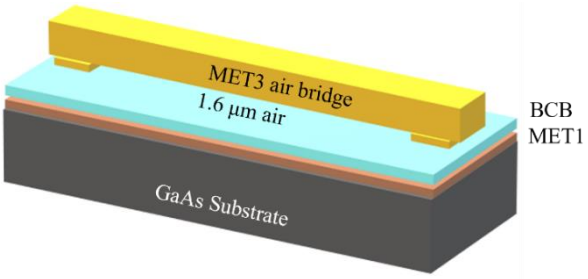


Figure 12: Air Bridge

An air-bridge TFMSL has a BCB2 layer and an air substrate of 11.6 µm thickness. It is shown in Figure 12. The substrate thickness is thus increased. With a resistance of 50 Ω, the elevated CPW must have a length of 10 µm and is thus shorter than MET2 TFMSL and MET2MET3 TFMSL but longer than the elevated CPW. From 200 GHz to 400 GHz over a length of λ/4, it also has a lower maximum loss than MET2 TFMSL and MET2MET3 TFMSL at 200 GHz with a similar curve to the elevated CPW of about 0.35 dB an [1]d a minimum loss of just under 0.3 dB at 400 GHz with a line impedance of 50 Ω [9].

IV. Recommended Pattern

Table 1: Comparison of the Insertion loss S11, return loss S21 and the slow wave factor n in dependency of the frequency f

Slow Wave Structure	Insertion loss S21/dB	Return loss S11/dB	Slow Wave Factor n (f /GHz)
Interdigital	-5	-0.33	37 (50)
Rectangular Meander	- 15	-0.33	10 (50)
V-Shaped Meander	-22	-0.4	5.6 (60)
Helix	-28	-0.1	5.93 (43) [10]
Polyline	-1.8	-0.025	3.0 (15)
Slotted Ground Plane	-22.6 [11]	-0.1	0.75 (10)
Elevated CPW	-5.9 [11]	-6.8 [11]	10 (0.1) [12]

With a MET2MET3 TFMSL, the shortest possible transmission distance can be achieved, whereby the loss becomes smaller as the frequency increases [9]. In addition, the slow wave effect can be improved by a suitable slow-wave pattern. According to Table 1, a helix slow wave structure, for example, would be suitable as a pattern, since particularly good values for S21 are obtained with a high slow wave factor of 5.93. However, the frequencies that occur must be considered when making the selection, as otherwise undesirable effects such as a band-pass filter can occur. [4, 2, 7].

V. Conclusion

Since there are many different slow wave structures, this paper compares them with each other. Some structures have a larger slow wave factor. In addition, the losses of the individual structures also vary among themselves. Overall, slow wave structures can cause a strong slow wave effect and thus come closer to the goal reducing the structures on chips.

VI. Acknowledgement

The author wants to thank Athanasios Gatzastras and Ingmar Kallfass for their help and advice.

VII. Bibliography

[1] A. Gatzastras, H. Massler und A. Leuther, „Implementation of Slow-Wave Thin-Film Microstrip Transmission Lines in a 35nm InGaAs Technology; Sebastien Chartier; Ingmar Kallfass,“ *IEEE Xplore*, pp. 121-124, 3 April 2021.

[2] N. Ryskin, A. Benedik, A. Rozhnev, N. Sinitsyn, R. Torgashov und G. Torgashov, „Study of Low-Voltage Millimeter-Wave Tubes with Planar Slow-Wave Structures on Dielectric Substrates,“ *10th UK-Europe-China Workshop on Millimetre Waves and Terahertz Technologies (UCMMT)*, pp. 1-2, 2017.

[3] F. Shen, Y.-Y. Wei, X. Xu, Y. Liu, H.-R. Yin, Y.-B. Gong und W.-X. Wang, „140-GHz V-Shaped Microstrip Meander-Line Traveling Wave Tube,“ *Journal of Electromagnetic Waves and Applications*, pp. 89-98, 13 April 2012.

[4] H. Jin, K. Wang, J. Guo, S. Ding und K. Wu, „Slow-Wave Effect of Substrate Integrated Waveguide Patterned With Microstrip Polyline,“ *IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, VOL. 64, NO. 6*, pp. 1717-1725, June 2016.

[5] H. Johnson, „EDN,“ 08 November 2001. [Online]. Available: https://www.edn.com/slow-wave-mode/. [Zugriff am 12 Juni 2022].

[6] T. Karetnikova, A. Benedik, A. Rozhnev, G. T. N.M. Ryskin, N. Sinitsyn und a. P. Shalaev, „Developement and modeling of G-band vacuum tube power amplifiers with sheet elevtron beam,“ *41st International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, 2016.

[7] G. Wu, H. Yin, Z. Xu, R. Yang, X. Lei, Q. Li, S. Fang, L. Yue, J. Xu, G. Zhao, W. Wang und Y. Wei, „Design of a Pseudoperiodic Slow Wave Structure for a 6-kW-Level Broadband Helix Traveling-Wave Tube Amplifier," *IEEE TRANSACTIONS ON PLASMA SCIENCE, VOL. 48, NO. 6,,* pp. 1910-1916, June 2020.

[8] C. Caloz, H. Okabe, T. Iwai und T. Itoh, „A Simple and Accurate Model for Microstrip Structures With Slotted Ground Plane," *IEEE MICROWAVE AND WIRELESS COMPONENTS LETTERS; VOL. 14, NO. 4,* pp. 133-135, 4 APRIL 2004.

[9] L. John, A. Tessmann, ArnulfLeuther, P. Neininger, T. Merkle und T. Zwick, „Broadband 300-GHz Power Amplifier MMICs in InGaAs mHEMT Technology," *IEEE TRANSACTIONS ON TERAHERTZ SCIENCE AND TECHNOLOGY, VOL. 10, NO.,* pp. 309-320, 3 Mai 2020.

[10] C. L. Kory und J. A. Dayton, „Computational Investigation of Experimental Interaction Impedance Obtained by Perturbation for Helical Traveling-Wave Tube Structures," *IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 45, NO. 9,* pp. 2036-2071, September 1998.

[11] T.-Y. Yun und K. Chang, „Uniplanar One-Dimensional Photonic-Bandgap Structures," *IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, VOL. 49, NO. 3,* pp. 549-553, March 2001.

[12] G. Plaza, R. Marqués und F. Medina, „Quasi-TM MoL/MoM Approach for Computing theTransmission-Line Parameters of Lossy Lines," *IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, VOL. 54, NO. 1, JANUARY,* pp. 198-208, January 2006.

[13] M. Pontón, F. Ramírez, A. Herrera und A. Suárez, „Oscillator stabilization through feedback with slow-wave structures," *IEEE Transactions on Microwave Theory and Techniques,* pp. 1-16, September 2019.

[14] G. X. SHU, G. LIU und Z. F. QIAN, „Simulation study of a high-order mode terahertz radiation source based on an orthogonal grating waveguide and multiple sheet electron beams," *OPTICS EXPRESS 8040 Vol. 26, No. 7,* pp. 1-9, April 2018.

[15] A. V. Starodubov, A. A. Serdobintsev, A. M. Pavlov, V. V. Galushka, D. M. Mitin und N. M. Ryskin, „A Novel Microfabrication Technology of Planar Microstrip Slow-Wave Structures for Millimeter-Band Traveling-Wave Tubes," *ResearchGate,* April 2018.

[16] D. Dutta, S. Rafidul, D. Guha und C. Kumar, „Suppression of Cross-Polarized Fields of Microstrip Patch Across All Skewed and Orthogonal Radiation Planes," *IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 19, NO. 1,* pp. 99-103, January 2020.

[17] L. Liu, Y. Wei, F. Shen, G. Zhao, L. Yue, Z. Duan, W. Wang, Y. Gong, L. Li und J. Feng, „A Novel Winding Microstrip Meander-Line Slow-Wave Structure for V-Band TWT," *IEEE ELECTRON DEVICE LETTERS, VOL. 34, NO. 10,* pp. 1325-1327, October 2013.

[18] L. Liu, Y. Wei, H. Yin, J. Xu, G. Zhao und Y. Gong, „A Modified Slotted Helix Slow-Wave Structure for High-Power Millimeter-Wave TWT," *2013 IEEE 14th International Vacuum Electronics Conference (IVEC),* May 2013.

The concept of Substrate Integrated Waveguides and associated Challenges

Electrical Engineering
University of Stuttgart
Stuttgart, Germany

***Abstract*— A detailed discussion on the concept of Substrate Integrated Waveguides (SIW) and the need for efficient transitions between SIWs and planar components as well as other waveguides. This paper touches upon the new technology and design of SIWs as well as their advantages over traditional microwave devices. The concept and challenges of transitions between SIWs and other waveguides as well as chips has been discussed in brief. Various newer and advanced methods of making these transitions between SIWs and different microwave components have been discussed and their improvements have been analyzed.**

***Keywords*— SIC, SIW, Rectangular waveguides (RWG), microstrip line (MSP), coplanar waveguides (CPW), wire bonding, E-band, transition,**

I. INTRODUCTION

The Substrate integrated Waveguide (SIW) is a relatively new form of millimetre wave technology developed on a single substrate. It is a modified rectangular waveguide (RWG) filled with dielectric. Substrate integrated circuits (SIC) enable a number of planar components like microstrip and coplanar waveguides and non-planar components like metal waveguides to be connected on same substrate. This helps achieve better characteristics in terms of performance, design, size and ease of manufacture of components over classical waveguide technology. While the developments can be promising, the integration of components requires transition between these components which can be challenging in terms of added circuit size and losses. The transitions should have minimum losses and higher capacity while adhering to miniature nature of circuits. Different methods to make these transitions have been introduced and more improved transitions have been discussed.

II. BASICS OF SIWS

A. Losses in transmission

The total loss (insertion loss) in a transmission line can be given as in [1]

$$\alpha_t = \alpha_c + \alpha_d + \alpha_r + \alpha_l + (\text{other losses}) \quad (1)$$

Where (α_c), (α_d), (α_r) and (α_l) represent conduction, dielectric, radiation and leakage losses respectively. In traditional metal waveguides, the losses are dominant due to bulkiness of the non-planar structure. However, they have better power handling capacities. The planar components like Microstripline (MSL) and coplanar waveguides (CPW) have better advantages in terms of small structures. These are however affected by insertion and radiation losses.

RWG are the earliest type of transmission lines for microwave signals. They have the ability to transmit at higher frequencies (E-band) with higher immunity unlike coplanar technologies. The need for integrating miniature scaled circuits cannot be achieved by RWG due to non-planar structure. Hence planar lines like Microstrip lines (MSL) can fulfil this requirement. Microstrip line are a popular type of planar transmission line which are easy to fabricate and integrate with other passive and active devices.

The radiation loss in MSL is given as [10]

$$\alpha_r = 60\left(\frac{2\pi h}{\lambda o}\right)^2 \times F(eff) \quad (2)$$

Where h is thickness of substrate, λ_o is wavelength in free space and F(eff) is a function of the effective dielectric constant ε_{eff} .. The loss will also be higher at higher frequencies (smaller λ). In our E-band frequency (60-90GHz) which is the band used for RF/Microwave backhaul links, the radiation loss will be higher. So, at higher frequencies, MSL introduces more radiation and hence insertion losses as given in equations (1) and (2).

B. Advantages of SIWs

Substrate integrated circuits were first designed to combine the desirable characteristics of planar and non-planar components. A single substrate is used for easy manufacture with a rectangular cross section to ensure low radiation and insertion losses. Since SIW is designed as a modified RWG, it retains properties of RWG while improving the Q factor due

to dielectric filling and volume miniaturization with planar structure.

C. Design specifications

In [2] and [3], the authors have introduced the concept and design specifications for fabrication of a SIW. They are designed as a thin dielectric substrate with rectangular cross section. Fig. 1 shows the parameters for a simple design of SIW. There are two parallel rows of via holes in the substrate. The diameter of each hole is given as post diameter *d*. The pitch p is defined as the spacing between two consecutive holes in a row. *W* and *h* are defined as the distance between two rows and thickness of substrate respectively.
The ratio *d/p* is critical for designing as diameter and pitch are correlated as given in [3]. The radiation loss decreases with decrease in post diameter for a constant value of ratio *d/p*. Two design rules for design of a good SIW are critical as given by [2]

$$d < \lambda_g/5 \quad (3)$$

$$p \leq 2d \quad (4)$$

where, λ_g is the guided wavelength

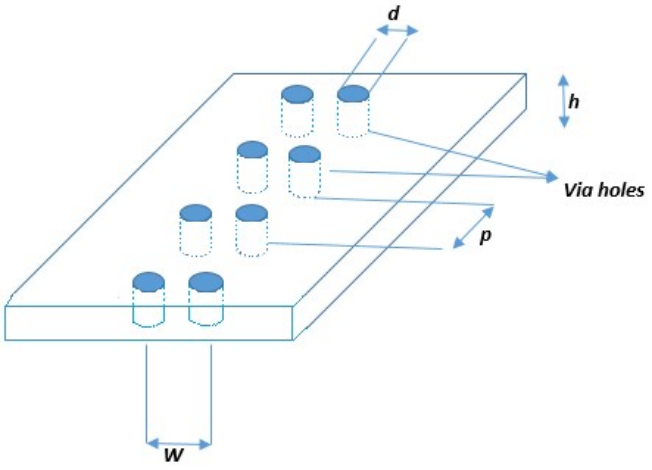


Fig. 1. Design of an SIW adapted from [2])

D. Challenges

The integration of all components on single substrate requires efficient transitions between the planar and non-planar components. Transitions can however add more losses and add to the bulkiness of circuits which is not desired.

General losses in a waveguide are described by its leakage characteristics which depend on the physical parameters of substrate. Due to the structure of SIW, a certain leakage wave loss is introduced due to periodic gaps in the whole structure [2]-[3].

In subsequent sections, a detailed analysis on the transitions between SIW and other microwave components will be done. Various design modifications have been reviewed to improve upon these transitions. Further sections will discuss the losses and leakage characteristics of a SIW.

III. TRANSITIONS

Transitions from SIW to both classical waveguides and planar waveguides have to be considered. While introduction of the initial transition designs were themselves promising as they enabled the design of all components on a substrate, they had to be made more efficient in terms of sizing and losses. Advancements and design modifications have been done over the last decade to achieve this better.

A. SIW to RWG

While the future scope of SIW stands to replace RWG in most circuits, the existing commercially available circuits mostly make use of classical RWGs.
Moreover, RWGs still provide better power handling, more rigidity and easier fabrication. This generates the need for developing good transitions between SIWs and these waveguides. The conduction loss in a RWG from (1) can be calculated from [3]

$$\alpha_c = \frac{R_s}{a^3 b \beta k \eta}\left(2b\pi^2 + a^3 k^2\right) \quad (5)$$

Where R_s is the surface resistance of the metallic walls, *a* and *b* are the width and the height of the waveguide, η is the intrinsic impedance of the material-filled waveguide, β is the propagation constant and $k = \omega\sqrt{\mu\varepsilon}$ the wavenumber of the material-filled waveguide region. The conduction loss will be significant at lower frequencies and if substrate is thin. So, a thicker substrate is preferred to minimize loss.

The earliest designs focused on making RWG transitions to planar forms. In [5], the authors have proposed a simple and novel method for this transition on the SIW substrate using PCB. A longitudinal slot window has been used on the wall of SIW to couple energy between SIW and RWG. There is a limitation on the frequency bandwidth due to added resonance of the slot window. Paper [6] has countered this resonance issue by employing an optimization algorithm to find best dimensions with max return loss and minimum insertion loss. Adding miniaturization of ohmic losses has shown to make the design more resistant to resonance of added transition structure. The paper has made a thorough analysis of the different transition structures over the years to counter this resonance of added structures. The recent designs have achieved the desired miniaturization while keeping the losses minimum with higher power handling characteristics. Table I shows the comparison in characteristics of the two discussed designs.

TABLE I.

BW(%)	Transition type	Substrate height	Reference
>37	Single layer slot	5.05	[4]
>43	Multilayer planar	0.15	[5]

B. SIW to MSL

1) Single layer

Initial transitions shown in paper [7] employed a simple taper design. They were efficient but bulky. Paper [8] proposed an innovative modification to the taper design by utilizing stepped impedance transformers. The final design simulations have shown that a bandwidth improvement of 10% and dimensions decrease by 60% have been achieved as compared to the simple taper.

The radiation loss in equation (1) shows a direct dependence on the substrate thickness. The loss will also be higher at higher frequencies (smaller λ). So, in our E-band frequency range (60-90GHz) which is the primary band of interest for RF/Microwave backhaul links, the loss can be more.

Equations (1) and (2) show a contradiction for design specifications. While a thin substrate is desired to minimize α_c a thicker substrate is needed to reduce α_r. A solution to this has been proposed in paper [9] by using a multilayer substrate which is discussed in the next section.

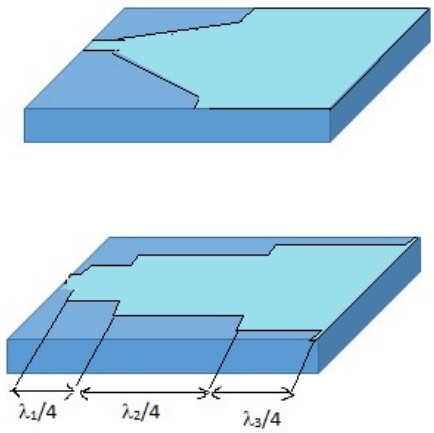


Fig. 2. (a)Tapered transition for single layer MSL adapted from[7]. (b)Modified design for MSL transition using stepped impedance transformers adapted from [8]

2) Multilayer

Authors in paper [9] have proposed transition using a multilayer substrate to achieve broadband response. Transition consisted of a tapered ridged SIW and tapered MSL by employing both impedance and field matching. This design opened up the possibilities of integrating SICs on multilayer substrate for better transition circuits.

C. SIW to CPW

The attenuation due to conduction α_c can also be reduced by using CPW.[3]

While radiation loss depends on substrate thickness and a thinner substrate is desired in MSL, CPWs show a different type of behaviour. This can be attributed to a lot of factors. The effective dielectric constant of CPW increases with thickness and a higher dielectric constant in turn reduces the overall radiation loss as shown in papers [11]-[12]. Overall, this enables CPWs to have a higher sensitivity to thicker substrates.

D. Single layer CPW

Initial transition designs started with design of complete CPW with SIW on single substrate without mechanical assembling or tuning [13]-[14]. The earlier designs had higher manufacturing costs due to back side patterning. Later advancements have been made with no step discontinuity or backside patterning. In paper [15], the authors have made use of a U-shaped slot antenna to achieve transition between SIW and ultra-wideband CPW. The slot dimensions variation on input impedance has been analyzed and band of operation in 79-110 GHz was realized.

E. SIW to IC Chip

Initial transitions from SIW to IC chip involved transition via a microstrip. This additional introduction of microstrip adds losses as discussed in SIW to MSL transition. To counter this problem, authors in [16] proposed an innovative technique of a direct wire bonding between the SIW and IC chip without the usage of microstrip. The structure is designed in the band 75-78 GHz. The inductive effect of wire bond was compensated using a low pass filter. Two different designs, one including the MSL transition and one with direct wire bonding have been compared. The later model had achieved improved insertion loss over the conventional design due to elimination of microstrip.

IV. LEAKAGE LOSS ANALYSIS

Considering the periodic gaps, SIW structures are subject to a potential leakage loss thus resulting in wave attenuation or leakage as shown in papers [2]-[3]. Therefore, it is pertinent to

inquire into what type of design models may exist in SIW and how losses are ranging in such models.

In paper [16], a thorough analysis of guided wave and leakage loss of SIW has been presented. The propagation of electromagnetic waves can be similar to that of a conventional RWG. The cosine of the incident angle θ can be given as

$$cos\theta = \frac{n\lambda}{2W} \quad (3)$$

Where n is the waveguide mode order, λ is the wavelength and W is the distance between rows of via holes.

And the the leakage part ratio can be expressed as

$$R_{leak} = 1 - \frac{d}{pcos\theta}, \ \theta<\theta_c \quad (4)$$

where d and p are the diameter of via holes and pitch length respectively.
From equation (3), when frequency increases, the incident angle also increases. From equation (4), the leakage loss also increases with increase in frequency at constant incident angle. And the loss additionally increases for higher modes when n is higher. Except when the pitch p is small, the leakage losses tend to increment significantly with increasing frequency. Therefore, for designing SIWs with low leakages, pitch p is kept small.
To obtain a good results in terms of return loss when the number of via holes are limited, another design rule is given by[]

$$d/W<0.4 \quad (7)$$

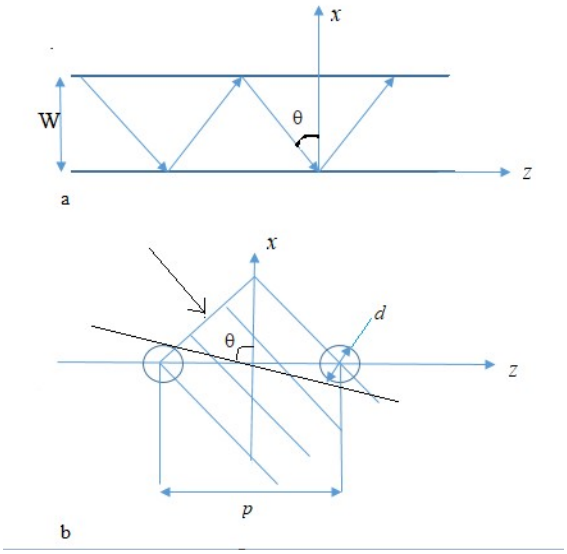


Fig. 3. (a)Guided wave propagation in a SIW adapted from []. (b)Wave incidence to via holes adapted from []

V. Analysis of Losses

A comparison can be made with the simplest models and the later advanced ones for analyzing improvements in the transitions and loss characteristics.

TABLE II.

Device	Design	BW (GHz)	Insertion loss(dB)	Return loss(dB)	Reference
RWG	SL	34-35	-1.5	-11	[4]
	ML	18-28	-1	-10	[5]
MSL	SL	10-19	-0.3	-20	[6]-[7]
	ML	23-27	-1.5	-15	[8]
CPW		19-38	-0.6	-14.5	[13]
	WB-CPW	79-110	-0.73	-20	[14]
IC Chip	Wire bonding	75-78	better	-12	[15]

VI. Conclusion

SIW first started off as a promising solution for integration of planar and non-planar components at in mm wave ranges. The technology came with its own challenges of leakages between holes. With design optimization, the size parameters have been tuned to reduce this loss by some extent. The technology further introduced the need for transitions with existing microwave components. These transition structures introduced additional losses and bulkiness. Over the years, advancements in designs have ensured the losses to be minimized while keeping the circuits small. The transitions of SIW with other waveguide components as well as with chip has been shown to be efficient enough to give a boost to microwave circuits whether they are concerned with structural as well as functional aspects..

References

[1] D. Pozar, Microwave engineering. New York: John Wiley & Sons, ., 2012.

[2] Ke Wu; Deslandes, D.; Cassivi, Y.; , "The substrate integrated circuits – a new concept for high-frequency electronics and optoelectronics," Telecommunications in Modern Satellite, Cable and Broadcasting Service, 2003. TELSIKS 2003. 6th International Conference on , vol.1, no., pp. P- III-P-X vol.1, 1-3 Oct. 2003.

[3] D. Deslandes and K. Wu, "Single-substrate integration technique of planar circuits and waveguide filters," in IEEE Transactions on Microwave Theory and Techniques, vol. 51, no. 2, pp. 593-596, Feb. 2003, doi: 10.1109/TMTT.2002.807820.

[4] L. Li, X. Chen, R. Khazaka and K. Wu, "A transition from substrate integrated waveguide (SIW) to rectangular waveguide," 2009 Asia Pacific Microwave Conference, 2009, pp. 2605-2608, doi: 10.1109/APMC.2009.5385245..

[5] E. Hassan et al., "Multilayer Topology Optimization of Wideband SIW-to-Waveguide Transitions", IEEE Transactions on Microwave Theory and Techniques, pp. 1-14, 2020.

[6] D. Deslandes and K. Wu, "Integrated microstrip and rectangular waveguide in planar form," in IEEE Microwave and Wireless Components Letters, vol. 11, no. 2, pp. 68-70, Feb. 2001, doi: 10.1109/7260.914305.

[7] Zheng Liu, GaoBiao Xiao " A New Transition for SIW and Microstrip Line" , 2013 Asia-Pacific Microwave Conference Proceedings.

[8] Y. Ding and K. Wu, "Substrate integrated waveguide-to-microstrip transition in multilayer substrate," in IEEE MTT-S Int. Microw. Symp. Dig., 2007, pp. 1555–1558.

[9] E. J. Denlinger, "Losses of Microstrip Lines," in IEEE Transactions on Microwave Theory and Techniques, vol. 28, no. 6, pp. 513-522, Jun. 1980, doi: 10.1109/TMTT.1980.1130112

[10] Rastogi, A.K., Jakal, A.Z., McEwan, N.J. et al. Loss minimisation and sensitivity analysis for coplanar waveguide. Int J Infrared Milli Waves 16, 1789–1810 (1995). https://doi.org/10.1007/BF02068690

[11] John Coonrod, "Insertion Loss Comparisons of Common High Frequency PCB Constructions", IPC APEX EXPO 2013, Feb. 2013

[12] D. Deslandes and K. Wu, "Integrated transition of coplanar to rectangular waveguides," 2001 IEEE MTT-S International Microwave Sympsoium Digest (Cat. No.01CH37157), 2001, pp. 619-622 vol.2, doi: 10.1109/MWSYM.2001.966971

[13] Taringou, F.; Bornemann, J.; , "New substrate-integrated to coplanar waveguide transition," Microwave Conference (EuMC), 2011 41st European , vol., no., pp.428-431, 10-13 Oct. 2011.

[14] G. Gentile, B. Rejaei, V. Jovanov ć, L. K. Nanver, L. C. N. de Vreede and M. Spirito, "Ultra-wide band CPW to substrate integrated waveguide (SIW) transition based on a U-shaped slot antenna," 2013 European Microwave Integrated Circuit Conference, Nuremberg, 2013, pp. 25-28.

[15] H. Zhang and W. Hong, "A Wire Bonding Structure Directly Based on Substrate Integrated Waveguide Technology", IEEE Microwave and Wireless Components Letters, vol. 24, no. 11, pp. 757-759, 2014.

[16] F. Xu and K. Wu, "Guided-wave and leakage characteristics of substrate integrated waveguide", IEEE Transactions on Microwave Theory and Techniques, vol. 53, no. 1, pp. 66-73, 2005

[17] D. Deslandes and K. Wu, "Design Considerations and Performance Analysis of Substrate Integrated Waveguide Components", Milano, European Microwave Conference, 23-27 Sept. 2002, pp. 881-884

[18] K. Wu, M. Bozzi and N. J. G. Fonseca, "Substrate Integrated Transmission Lines: Review and Applications," in IEEE Journal of Microwaves, vol. 1, no. 1, pp. 345-363, Jan. 2021, doi: 10.1109/JMW.2020.3034379.