Proceedings of the 4th International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/2026.33563

Normally-off GaN HEMT Device Technology and Performance Enhancement Methods

Haotian Wang

College of Optic and Electronic Technology, China Jiliang University, Hangzhou, China
2200404144@cjlu.edu.cn

Abstract. With the rapid development of high-frequency, high-power electronic systems, traditional silicon-based devices are no longer sufficient to meet the application requirements of high efficiency and high power density. Gallium nitride (GaN) is a wide-bandgap semiconductor material with excellent electrical and thermal properties, and is currently an important research subject for next-generation power electronic devices. Enhancement-mode (E-mode, normally off) GaN high electron mobility transistors (HEMTs) have attracted widespread attention due to their higher security and simpler driving requirements. This paper systematically reviews the main technical approaches to realizing E-mode GaN HEMTs, including fluorine ion implantation, MIS structures, and p-GaN gate technology. Its working mechanism, structural characteristics, advantages and disadvantages were analyzed and compared. The main performance enhancement technologies, including interface state control and field plate structure optimization, were thoroughly examined. In addition, important issues such as current collapse and short-circuit protection were analyzed in the context of practical device applications. Finally, the development trends of GaN devices are discussed from an application perspective. With continued advances in materials, device structures, and packaging technologies, GaN devices are likely to become increasingly important in high-frequency and high-power-density power electronic applications.

Keywords: GaN HEMT, Enhancement-mode, Normally-off

1. Introduction

Currently, the continued growth of global energy consumption has become an increasingly important issue. Electricity is expected to account for a large proportion (approximately 60%) of total energy use in the future, which places higher demands on power electronic devices, especially in terms of frequency, voltage, and power levels. Meanwhile, practical applications are also developing towards lower costs and miniaturized devices. This combination of factors is causing traditional silicon-based devices to approach the limit of their performance. As a result, wide-bandgap semiconductors like gallium nitride (GaN) have received great interest as potential materials used in future high-frequency, high-power devices. GaN devices may be operated with higher electric field strengths and higher temperatures, with less conduction loss compared to silicon

(Si). Not only do these benefits concern the material itself, but also the AlGaN / GaN heterojunction structure.

Based on this heterojunction, GaN field-effect transistors (FETs) utilize the two-dimensional electron gas (2DEG) formed at the interface as the conduction channel, enabling high current density and low on-resistance (Ron). Thus, they have found extensive application in power and radio frequency applications. A majority of traditional GaN FETs are depletion-mode (normally-on) which implies that the devices can carry a current even with zero gate voltage. Although this allows high current capabilities, it also presents a number of practical issues, including poor safety, greater complexity of the gate driver circuitry, and increased probability of failure. In contrast, E-mode GaN FETs remain off at zero gate voltage and turn on under positive gate bias, making them more suitable for modern circuit systems with simplified driving requirements [1].

Several approaches have been suggested to realize E-mode GaN FETs. These include recessed gate structures, p-type GaN gate technology, fluorine ion implantation, cascode configuration, and metal-insulator-semiconductor gate structures. All these techniques are capable of producing a positive threshold voltage (Vth) with the devices obtained having different performances.

The article is devoted to the E-mode GaN high electron mobility transistor (HEMT) technology, which examines the principle of its work, its typical structures, and some important performance indices. Moreover, three exemplary techniques including fluorine ion implantation, gate-recessed MIS-HEMT and p-GaN gate were compared in accordance with their working mechanisms, structural properties, benefits and drawbacks. Besides, the present paper also makes reference to the most important device parameters, which include Vth, Ron, breakdown characteristics, gate reliability, and dynamic performance, and addresses the existing issues and development prospects. The objective is to elucidate the technical disparities among different implementation approaches and offer an orientation on the optimization of devices and associated studies.

2. Main technical approaches to achieving E-mode GaN HEMT

In the research of realizing E-mode GaN HEMTs, different technical routes are essentially about the manipulation of the gate region band and the 2DEG. Based on the different manipulation methods, they can be divided into three categories: first, charge-controlled methods, represented by fluorine ion implantation, which deplete the channel by introducing negative charges; second, gate dielectric manipulation methods, represented by hybrid MIS-HFETs, which achieve E-mode by adjusting the interface potential; and third, structural manipulation methods, represented by p-GaN gates, which achieve channel depletion by raising the barrier through the p-n junction.

2.1. Fluorine ion implantation technology

Fluorine ion implantation is a classic method for achieving normally-off GaN HEMTs. The basic idea is to introduce negatively charged fluorine-related charges near the gate region through plasma etching or ion implantation, thereby weakening the polarization effect in the AlGaN/GaN heterojunction, reducing the concentration of the 2DEG under the gate and shifting the V_{th} toward positive values [2,3]. As illustrated in Figure. 1, fluorine ions introduced beneath the gate region generate negative charges that deplete the 2DEG at the AlGaN/GaN interface.

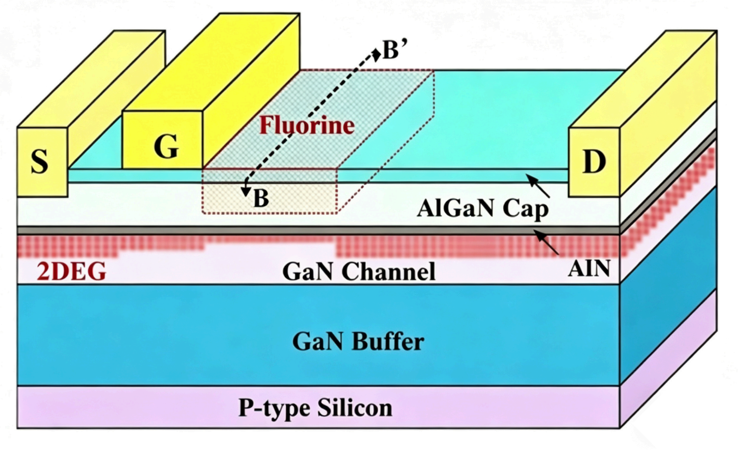

Figure 1. The structure of fluorine ion implanted AlGaN/GaN HEMT

Currently, the mature fluorine ion implantation method involves injecting F^- ions using fluorine-based plasma processing technology. When enough F^- ions are injected, the Vth will shift to a positive value, resulting in E-mode HEMT. In their experiments, Yong Cai et al. demonstrated this method by introducing fluorine ions into the AlGaN barrier layer using CF_4 plasma. This approach shifts the threshold voltage (Vth) of AlGaN/GaN HEMTs from approximately −4 V to +0.9 V, allowing the device to operate in E-mode. The incorporated fluorine ions also exhibited good thermal stability up to 700 °C [2].

Although this method is simple and effective, the stability of Vth of such devices after annealing becomes a problem. Fluorine ion implantation can cause lattice damage to devices, introducing current collapse. Thermal annealing can reduce deep-level traps caused by fluorine ion implantation and repair lattice damage, thereby reducing current collapse. However, this also shifts Vth towards a negative value, weakening the effect of ion implantation [4].

Although fluorine ion implantation technology has advantages such as simple process and flexible threshold control, the lattice damage and V_{th} stability issues it introduces still limit its application in high-reliability devices.

2.2. Hybrid MIS-HFET structure

To overcome the limitations of the above methods and achieve higher-performance E-mode HEMTs, Kambayashi et al. proposed a hybrid MIS-HFET structure. In this particular device, SiO_2 was adopted as the gate dielectric, making it a specific MOS-HFET implementation. In this approach, a recessed MOS channel is formed by chlorine (Cl_2) plasma etching, resulting in an E-mode device with a Vth of 2.7 V [5].

This structure avoids ion implantation and enables the fabrication of uniform, large-area AlGaN/GaN MOS-HFETs. Because silicon dioxide (SiO_2) is inexpensive and has good bandgap alignment with GaN, it is often used as a gate insulating layer [6]. Taking it as an example, the large conduction band offset at the SiO_2/GaN interface can effectively block electron injection under positive gate bias, thereby reducing gate leakage current. With an insulating layer introduced, gate control over the channel is established primarily through the electric field, which enhances the ability to modulate the energy band of the AlGaN barrier. In this structure, when the gate voltage is zero, the gate electric field can deplete the 2DEG at the interface, thereby reducing or even eliminating the carrier concentration in the channel and realizing the normally-off characteristic of the device, as shown in Figure 2. The introduction of an insulating layer enhances gate control and also enables more stable electrostatic modulation of the channel.

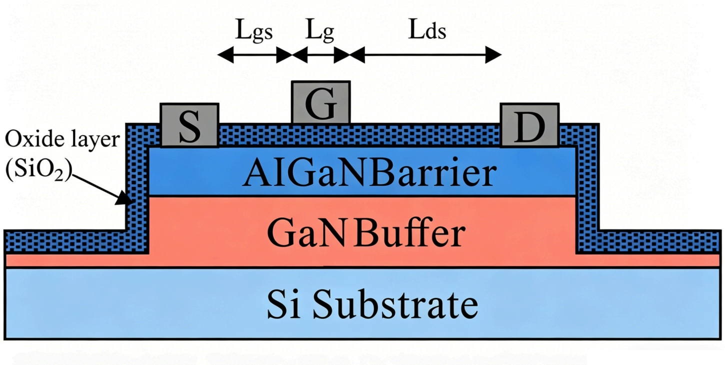

Figure 2. A cross-sectional structure of an AlGaN/GaN MOS-HEMT with a SiO_2 dielectric layer

However, Fiorenza et al. reported that such devices still exhibit instability with respect to the Vth [6]. The reason is that the near-interface traps were not sufficiently passivated after the nitride annealing treatment. Under gate bias stress, these traps can capture and release charge carriers, resulting in a shift in the V_{th} during operation. Since E-mode devices rely on a stable positive V_{th} to ensure safe shutdown, this instability has a significant impact on device reliability.

2.3. P-GaN gate technology

Among the various approaches proposed to realize E-mode GaN HEMTs, the p-GaN gate structure is one of the most widely adopted because it can simultaneously provide a positive Vth, low gate leakage current, and good process compatibility. These advantages make it one of the most widely used solutions.

This approach introduces a p-GaN layer over the AlGaN layer. The p-type layer raises the conduction band of the underlying AlGaN, thus, depleting the 2DEG with zero gate bias. This structure allows normally-off operation, as depicted in Figure 3. Because doped GaN tends to be n-type conductivity, it is important to have enough holes in the p-GaN layer to effectively deplete the channel. The most frequently used p-type dopant in GaN-based materials is magnesium (Mg). Its Vth can be improved by increasing its doping concentration or activation efficiency.

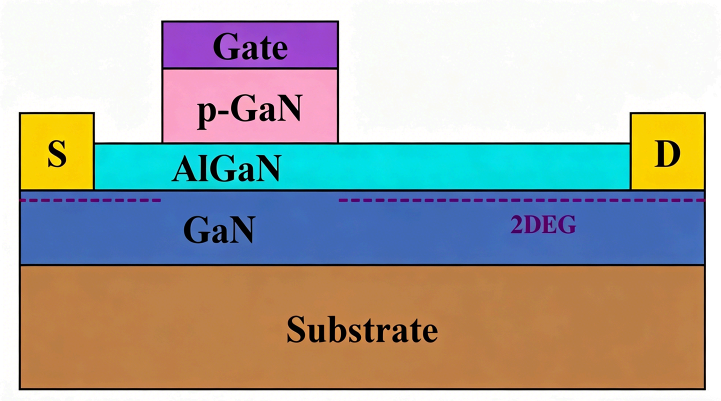

Figure 3. P-GaN gate AlGaN/GaN HEMT structure enables normally-off operation by depleting the 2DEG at zero gate bias

Nevertheless, Mg has certain drawbacks. The ionization energy of Mg is quite high in GaN (which restricts the hole concentration at high doping concentrations). Also, over-doping with Mg negatively affects the crystal quality of p- GaN films and lowers the amounts of electroactive acceptors because of the compensation effect [7]. Thus, it is not possible to continue improving the performance of the devices by increasing the doping concentration indefinitely; the maximum obtainable Vth is inherently constrained by the material.

To address the aforementioned issues, extensive research has been conducted on gate materials. Comparative tests were conducted on various metals (such as nickel, molybdenum, and titanium) for p-GaN gates. The results show that titanium nitride (TiN) exhibits excellent thermal stability, chemical robustness, and compatibility with standard manufacturing processes, and has great application potential. By combining TiN gate electrodes with p-GaN technology, device performance can be further improved, and the V_{th} can be shifted forward by about 2.1 V [1].

Overall, p-GaN gate technology achieves a stable normal turn-off state through its inherent structural design, and has become the mainstream solution for manufacturing E-mode GaN devices. Although its performance is still limited by the trade-off between doping efficiency and material quality, and further improvements require synergistic optimization of doping strategies and gate engineering, it remains the most commercially viable and feasible solution at present.

3. Key technologies for improving device performance

In E-mode GaN HEMT devices, performance improvements primarily depend on the control of interface states and electric field distribution. Interface states are mostly due to the defects created in the process of fabricating the gate insulation layer interface or AlGaN/GaN interface, contamination with oxygen, and surface damage. Electric field distribution is controlled using field plate structures. It redistributes the electric field and suppresses local field peaks, increasing the breakdown voltage and reducing the dynamic performance degradation. This section will focus on introducing the two main technologies listed above that can enhance the performance of GaN devices.

3.1. Interface state control

Interface states are primarily due to defects, oxygen contamination, and surface damage that have been introduced during the preparation of the gate insulating layer and the AlGaN/GaN interface. Such interface states may capture and release charge carriers causing issues such as frequency dispersion in C-V properties, Vth drift, and higher dynamic Ron, which are very limiting on the operation of the devices.

The interface chemical state is strongly associated with the development of interface states. For example, in the Al_2O_3/GaN structure, Ga-O bonds are readily created at the interface and such oxide defects have a major effect on increasing the density of interface states. The use of this approach is also effective in preventing the diffusion of oxygen into the system and suppressing the creation of Ga-O bonds through the introduction of an AlN interface layer between the gate insulating layer and GaN. It is possible to decrease Dit by almost an order of magnitude, the C-V frequency dispersion of the device will be significantly reduced, and the quality of the interface will be greatly improved [8].

Besides the optimization of interface materials, processing technology plays a vital role in interface state as well. Conventional dry etching causes surface degradation and contamination of trenches in the fabrication of trench gates, resulting in a higher interface state density. In order to solve this problem, a self-terminating trench gate process has been proposed that may be able to remove the damaged layer and recover the surface chemical state. This technique will lower the interfacial state density by approximately 10^{13} $eV^{-1}\cdot cm^{-2}$ to 10^{11}–10^{12} $eV^{-1}\cdot cm^{-2}$, and greatly decrease C-V frequency dispersion [9].

To sum up, it is possible to implement interface state control using materials engineering or optimizing processes. One solution would involve adding an interface of high quality to prevent the occurrence of interface defects and optimizing the etching and heat treatment steps in order to minimize damage and impurity entry. Therefore, reducing the interface state density through material engineering and process optimization is essential for improving the overall performance and reliability of E-mode GaN HEMTs.

3.2. Field plate structure optimization

Field plate structure is one of the methods that are frequently applied in modulating the electric field. A metal extension above the gate or source creates an additional potential profile, modifying the electric field distribution. Field plate structures shift the electric field peak at the gate edge toward the drain, thereby reducing the local electric field intensity [10]. This mechanism improves device stability under high-voltage conditions and thus greatly improving its off-state breakdown voltage and decreasing the leakage current.

Besides improving breakdown performance, the field plate structure has a strong influence on the dynamic operation of the device. Under conditions of a high electric field, electrons can be more readily trapped at the interface or the surface, resulting in current collapse. The insertion of a field plate structure can reduce the peak electric field and suppress electron trapping, and alleviate the reduction of dynamic Ron. Experimental results show that, in comparison to devices without field plates, the dynamic Ron is much lower and the device performance is much better when the field plate structure is introduced [10].

Moreover, various field plate structures have performance improvement differences. Typical structures are gate field plates, source field plates, and dual field plate structures which incorporate both the former. The optimum electric field distribution achieved through the dual-field plate structure can be used to lower the peak of the electric field and reduce current collapse even further, thus providing an overall improvement in performance. The rational choice of the field plate structure and its dimensional parameters play a significant role in the practical design of devices as it can be used to achieve high-performance and high-reliability enhancement GaN-FETs.

4. Current challenges and outlook

4.1. Challenges

4.1.1. Current collapse effect

Current collapse is another major constraint that impedes the efficiency and reliability of the E-mode GaN HEMTs. In most cases, when the devices are turned on again after they have been switched off at a high voltage, the drain current decreases significantly with a corresponding increase in dynamic Ron. The degradation is not a permanent failure, but it is strongly associated with charge carrier trapping and release in traps [11].

Current collapse is primarily caused by carrier trapping under high electric field conditions. During the off-state, a high electric field in the gate–drain region promotes electron trapping by surface, interface, and buffer layer states. Upon reactivation of the device, the long detrapping time constant will lead to failure of some electrons to return to the channel in time and thus reduce the carrier concentration in the channel and reduce the turn-on current. It is time-dependent and depends on bias conditions, electric field strength, and device structure.

In order to alleviate current collapse, several strategies have been created, such as surface passivation to decrease interface states, field plate structures to reduce the electric field, and dual heterostructures to enhance carrier confinement. These methods can enhance the performance of the device to a certain degree, yet they do not totally remove this issue. Interface and surface defects remain challenging to eliminate entirely during the production process. At the high voltage operation of the device, the carrier capture process will be continually induced by the electric field, which is very strong, so that the current collapse remains rather serious in high power applications. In addition, current collapse has a multi-timescale nature. The capture and release time constants of different types of traps differ, which results in different levels of performance degradation at different frequencies of operation and pulses. It also makes the modeling and optimization of the devices more challenging and requires greater stability of the devices in high-frequency, high-efficiency operation.

Current collapse remains a key challenge in the development of GaN power devices. The solution to this issue involves combined optimization of various aspects such as material growth, interface engineering and device structure to ensure a balance between high performance and high reliability of the device.

4.1.2. Short circuit protection

Short-circuit faults can significantly affect the reliability of GaN HEMTs in high-power applications. Compared with more conventional Si and silicon carbide (SiC) devices, E-mode GaN HEMTs are not as resilient to short-circuits, usually surviving only a few hundred nanoseconds, and much shorter at high voltages. The requirements on the response time of the protection circuit are also very high [12].

During a short circuit, the device is subjected to extreme high voltage and current conditions leading to an abrupt rise in instantaneous power consumption. It may lead to rapid rise in junction temperature and thermal failure. GaN devices have high carrier density in the channel and high conductivity which may produce large peak currents in the first stage of a short circuit, thus making the device more vulnerable to failure. Thus, short-circuit current detection and suppression within an extremely short time is essential to guarantee the safety of equipment.

Even though several methods of short-circuit protection have been suggested including current detection, voltage detection, and desaturation protection, they still have certain limitations when applied to GaN devices. Since the response time of conventional protection circuits is typically in the microsecond range, it is not able to satisfy the nanosecond-level protection demands of GaN devices. Moreover, the high-dv/dt switching environment can be susceptible to noise interference causing spurious triggering or protection failure.

In order to resolve these problems, several ultra-fast protection methods have been suggested over the past few years. For example, by observing the instantaneous changes in the phase node voltage, short-circuit fault detection is possible in tens of nanoseconds. This combination of gate voltage clamping and soft turn-off is effective in reducing the short-circuit current and the possibility of damage to the device. But they usually require complicated circuit design and exact parameter matching, which increases system implementation complexity. The future studies should aim at improving other areas like fast detection, anti-interference properties, and integration into a system to enable highly reliable protection schemes for GaN devices.

4.2. Outlook

With the ongoing enhancement of E-mode GaN HEMT performance, the research direction is gradually shifting toward system-level applications rather than optimizing individual devices. The GaN devices, which have a low Ron and are capable of extremely high-speed switching, can minimize switching losses and also reduce the size of passive devices, making it possible to create power modules with high power density in the high-frequency and high-efficiency power conversion systems. For example, half-bridge power modules based on high-voltage GaN have demonstrated switching speeds of nanoseconds and low-loss operation at 650 V and thus have potential applications in electric vehicle, motor drive, and new energy inverter applications [13].

GaN devices are also highly advantageous when it comes to high temperature application. Monolithic integrated circuits that use GaN devices can operate at temperatures up to 200 °C, which is very useful in harsh conditions like space or drilling in deep wells [14]. Through approaches like interface charge modulation, the Vth of the device may be raised, thus improving its immunity to false triggering and ensuring safety and stability of the system.

A major future direction of GaN technology lies in GaN integrated circuits (GaN ICs) and high-power modularization [15]. When power devices are combined with drive and protection circuits on a single chip, parasitic parameters can be minimized and system response time can be increased. The potential of GaN devices in high-frequency and high-power applications can be further realized through advanced packaging technologies to realize multi-chip parallel interconnection and low-inductance designs. In actual applications, it is necessary to optimize electromagnetic interference induced by high-speed switching, thermal management, and system reliability in a comprehensive way.

To sum up, the further evolution of GaN devices would not be limited to enhancing the performance of the devices alone but also the practical applications situations, including system-level optimization. As material technologies, device structures, and packaging technologies keep developing, the importance of GaN in high-frequency and high-power-density power electronic systems is expected to continue to grow.

5. Conclusion

With the development of wide-bandgap semiconductor technology, E-mode GaN HEMTs have shown significant application potential in the field of high-frequency, high-efficiency power electronics. This paper reviews the implementation techniques and performance optimization methods of E-mode GaN HEMTs. The study focuses on three typical implementation routes: fluorine ion implantation, MIS structure, and p-GaN gate. The mechanism for improving device performance was also explored from two aspects: interface state control and field plate structure optimization. Furthermore, this paper analyzes the main challenges currently faced by devices in practical applications, including issues such as current collapse and short-circuit protection. Overall, the future development of GaN devices will increasingly rely on synergistic optimization at the material, structure, and system levels to achieve a balance between high performance and high reliability.

References

[1] F. Roccaforte, G. Greco, P. Fiorenza, and F. Iucolano, "An Overview of Normally-Off GaN-Based High Electron Mobility Transistors, " Materials, vol.12, no.10, p.1599, May 2019, doi: 10.3390/ma12101599.

[2] Y. Cai, Y. Zhou, K. M. Lau, and K. J. Chen, "Control of Threshold Voltage of AlGaN/GaN HEMTs by Fluoride-Based Plasma Treatment: From Depletion Mode to Enhancement Mode, " IEEE Trans. Electron Devices, vol.53, no.9, pp.2207–2215, Sep.2006, doi: 10.1109/TED.2006.881054.

[3] G. Greco, F. Iucolano, and F. Roccaforte, "Review of technology for normally-off HEMTs with p-GaN gate, " Mater. Sci. Semicond. Process, vol.78, pp.96–106, May 2018, doi: 10.1016/j.mssp.2017.09.027.

[4] A. Lorenz et al., "Influence of thermal anneal steps on the current collapse of fluorine treated enhancement mode SiN/AlGaN/GaN HEMTs, " Phys. Status Solidi C, vol.6, no.S2, Jun.2009, doi: 10.1002/pssc.200880838.

[5] H. Kambayashi et al., "Over 100A operation normally-off AlGaN/GaN hybrid MOS-HFET on Si substrate with high-breakdown voltage, " Solid-State Electron., vol.54, no.6, pp.660–664, Jun.2010, doi: 10.1016/j.sse.2010.01.001.

[6] P. Fiorenza, G. Greco, F. Giannazzo, F. Iucolano, and F. Roccaforte, "Effects of interface states and near interface traps on the threshold voltage stability of GaN and SiC transistors employing SiO2 as gate dielectric, " J. Vac. Sci. Technol. B Nanotechnol. Microelectron. Mater. Process. Meas. Phenom, vol.35, no.1, p.01A101, Jan.2017, doi: 10.1116/1.4967306.

[7] U. Kaufmann, P. Schlotter, H. Obloh, K. Köhler, and M. Maier, "Hole conductivity and compensation in epitaxial GaN: Mg layers, " Phys. Rev.B, vol.62, no.16, pp.10867–10872, Oct.2000, doi: 10.1103/PhysRevB.62.10867.

[8] S. Liu et al., "Interface/border trap characterization of Al2O3/AlN/GaN metal-oxide-semiconductor structures with an AlN interfacial layer, " Appl. Phys. Lett, vol.106, no.5, p.051605, Feb.2015, doi: 10.1063/1.4907861.

[9] S. Su et al., "Self-terminated Gate Recessing with a Low Density of Interface States and High Uniformity for Enhancement-mode GaN HEMTs, " in 2020 32nd International Symposium on Power Semiconductor Devices and ICs (ISPSD), Vienna, Austria: IEEE, Sep.2020, pp.333–336.doi: 10.1109/ISPSD46842.2020.9170199.

[10] W.-C. Cheng et al., "Technological Development in Pursuit of High-Performance Normally-off GaN-based HEMTs, " in 2020 IEEE 15th International Conference on Solid-State & Integrated Circuit Technology (ICSICT), Kunming, China: IEEE, Nov.2020, pp.1–5.doi: 10.1109/ICSICT49897.2020.9278368.

[11] S.-Y. Ho, C.-H. Lee, A.-J. Tzou, H.-C. Kuo, Y.-R. Wu, and J. Huang, "Suppression of Current Collapse in Enhancement Mode GaN-Based HEMTs Using an AlGaN/GaN/AlGaN Double Heterostructure, " IEEE Trans. Electron Devices, vol.64, no.4, pp.1505–1510, Apr.2017, doi: 10.1109/TED.2017.2657683.

[12] X. Lyu et al., "A Reliable Ultrafast Short-Circuit Protection Method for E-Mode GaN HEMT, " IEEE Trans. Power Electron, vol.35, no.9, pp.8926–8933, Sep.2020, doi: 10.1109/TPEL.2020.2968865.

[13] B. Passmore et al., "A 650 V/150 A enhancement mode GaN-based half-bridge power module for high frequency power conversion systems, " in 2015 IEEE Energy Conversion Congress and Exposition (ECCE), Montreal, QC, Canada: IEEE, Sep.2015, pp.4520–4524.doi: 10.1109/ECCE.2015.7310298.

[14] L.-F. Jia et al., "E/D-Mode GaN Inverter on a 150-mm Si Wafer Based on p-GaN Gate E-Mode HEMT Technology, " Micromachines, vol.12, no.6, p.617, May 2021, doi: 10.3390/mi12060617.

[15] Q. Zhou, Y. Yang, K. Hu, R. Zhu, W. Chen, and B. Zhang, "Device Technologies of GaN-on-Si for Power Electronics: Enhancement-Mode Hybrid MOS-HFET and Lateral Diode, " IEEE Trans. Ind. Electron, vol.64, no.11, pp.8971–8979, Nov.2017, doi: 10.1109/TIE.2017.2652373.