Unlocking The Potential of Gallium Nitride (Gan) For Third-Generation Semiconductors: Advantages, Applications, And Future Prospects

Sihan Chen *

Department of Electronic & Electrical Engineering, University College London, London, UK

* Corresponding Author Email: zceesc0@ucl.ac.uk

Abstract. Wide-bandgap materials, exemplified by gallium nitride (GaN), are poised to elevate electronic performance to new heights. This paper is focus on the third-generation semiconductor which is considered as the most promising semiconductor. The first section will discuss the fundamentals of CMOS and HEMT, including their construction, operation, benefits, and properties. The two are next be compared, and it is revealed that although GaN has some limitations on its cost and manufacture, it also having higher energy gap, compact size, higher efficiency, lower switching losses. Meeting the increasingly rigorous requirements associated with high power and high-density applications requires the use of these advantages. Then, using examples from the aerospace, mobile phone chargers, and LiDAR fields, applications of GaN HEMTs are discussed. The issues and prospects for third generation semiconductors are finally went over. It shows significant potential for energy efficiency benefits in infrastructure, and it is intriguing to see how the new semiconductor generation develops in the next few years.

Keywords: CMOS, HEMT, GaN, semiconductor.

1. Introduction

Starting from the 1960s, the initial generation of semiconductors, like silicon and germanium, found extensive application in information processing and automation control systems. This significantly fueled the expansion of the IT sector and propelled the evolution of the microelectronics industry, centered around integrated circuits. In the 1990s, Indium phosphide (InP), representing the second generation of semiconductor materials, began to gain prominence. This coincided with the rapid advancement of mobile communications, the establishment of optical fiber-based communication networks as the backbone of the information superhighway, and the advent of the Internet. With the development of Internet technology, it is also used in satellite communications and GPS, etc. In 1993, the United States invented the first gallium nitride-based electronic devices, which is the third generation of semiconductor materials.

This article begins with the basis of CMOS and HEMT including their structure, working principle, advantages, and characteristics. It then compares the two, showing that third generation semiconductors have high bandgap, high conductivity and more stable. It also has low switching losses and high efficiency, but more costly. Next, applications of GaN HEMTs are then discussed, taking aerospace, mobile phone chargers and LiDAR as examples. Finally, the challenges and future directions of third generation semiconductors are discussed. Third-generation semiconductors also hold great promise for energy efficiency savings, especially for future infrastructure, data centers and 5G base stations.

2. CMOS and HEMT

2.1. Basis of CMOS

A complementary metal-oxide semiconductor (CMOS) is a fabrication technology widely used in integrated circuits (ICs). The primary objective of developing it was to achieve high speed and low power consumption of logic gates in digital circuits. Electronic watches were the only products that

used CMOS technology in the 1970s. In the 1980s, CMOS overtook NMOS logic and replaced earlier transistor-transistor logic (TTL) technology. Since CMOS has advantages including lower consumption, larger noise margin, wider temperature and voltage working range, overall circuit simplicity, and ease of layout, very large-scale integration (VLSI) technology has started to transition to it. The development of VLSI in the 1980s drove integration densities up to millions of transistors on a single chip. In 2011, CMOS technology was adopted by 99% of IC chips, including most digital, analog, and mixed-signal ICs. VLSI digital and mixed-signal designs are currently dominated by CMOS technology [1].

2.1.1 Anatomy of CMOS

There are two types of MOSFETs, which are p-channel MOS (PMOS) and n-channel MOS (NMOS). Both PMOS and NOMS are made of both p-type and n-type semiconductors. The type of MOSFET shown in the figure1 is fabricated using a p-type substrate, i.e. a wafer of single-crystal silicon with a background concentration of some acceptor, such as boron. Into this substrate, two heavily n-doped regions (n+) are created (either by implanting donor atoms or diffusing them into the substrate). These regions are termed source and drain, though as is evident from the symmetry of the device, this has more to do with how the FET is operated in a circuit, rather than any inherent difference between the two regions. The area between the n+ source and drain regions is known as the channel.

A thin insulating layer is grown on top of the channel. The insulator is typically silicon oxide (SiO2) which can be made by heating the silicon in an oxygen- or water-rich atmosphere so that it begins to oxidize. Metallic contacts (e.g., made of aluminum) are then deposited over the source (S) and drain (D) regions, contacting them, and over the insulating layer. The contact that lies over the channel is known as the gate (G). The substrate is also contacted with a fourth terminal, known as the body (B). The three-dimensional figure of MOSFET is shown in Fig.1.

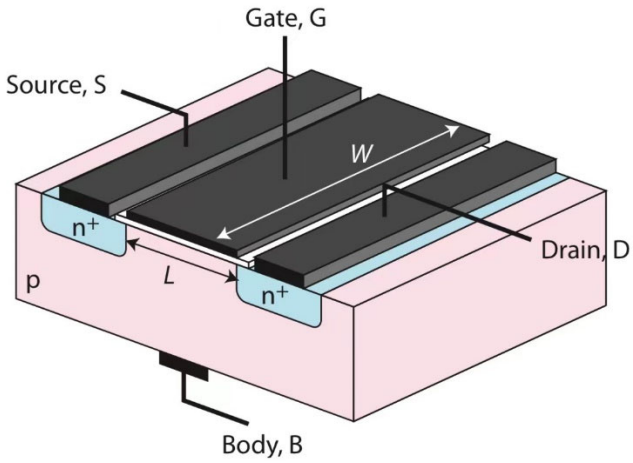
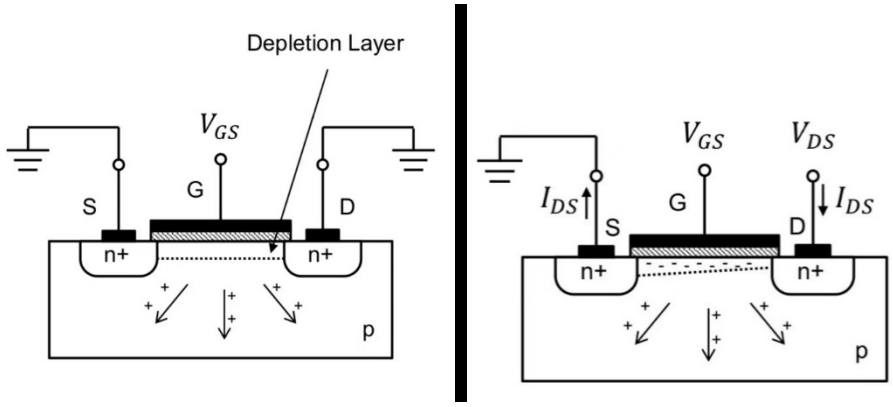


Fig. 1 A dimensional illustration of MOSFET [1]

2.1.2 I-V feature of CMOS

Three distinct operation regions can be identified: cut-off, linear (triode), and saturation. The application of a minor positive gate-source voltage results in the displacement of free holes from beneath the oxide layer, leading to the formation of a depletion layer. This region becomes a reservoir for accumulating negative charges as holes evacuate the fixed acceptor atoms. Despite the presence of a positive V_{DS}, the absence of free carriers within the depletion layer, as illustrated in Figure 2, prevents any current from flowing.



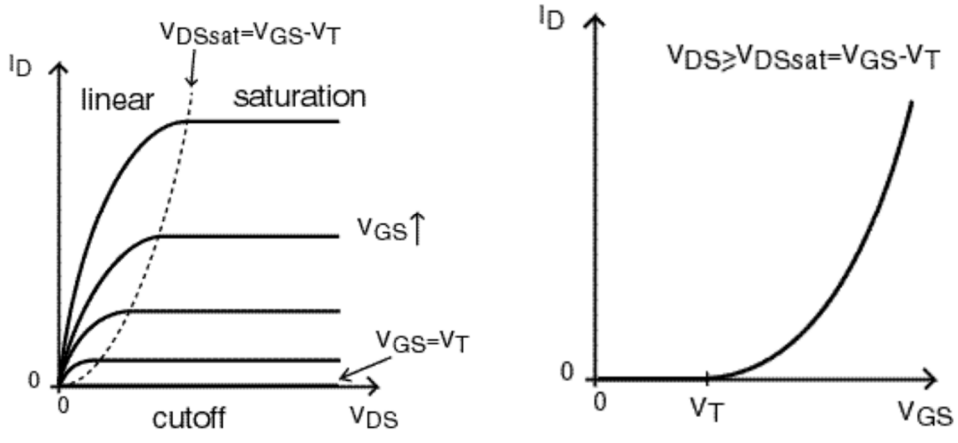
(a) (b)

Fig. 2. Schematic of NMOS in (a)cut-off region and (b)triode region [2]

In the case where $V_{GS} \geq V_{thr}$ and a small positive drain-source voltage V_{DS} is applied, electrons flow from the source to the drain, resulting in a current I_{DS} moving from drain to source. As V_{DS} increases, I_{DS} also increases concurrently, causing a reduction in the potential difference between the gate and the drain. The drain-source current, expressed as I_{DS}, follows the formula $I_{DS} = g_n(V_{GS} - V_{thr} - \frac{1}{2}V_{DS})V_{DS}$.

Conversely, when $V_{GS} \geq V_{thr}$ and a larger positive drain-source voltage V_{DS} is applied, the n-channel's depth diminishes to zero at the drain end as V_{DS} continues to rise beyond $V_{GS} - V_{thr}$, a phenomenon known as pinch-off. Although I_{DS} persists through the pinched-off channel, further increases in V_{DS} no longer influence I_{DS}. In this scenario, the drain-source current is represented by $I_{DS} = \frac{1}{2}g_n(V_{GS} - V_{thr})^2$, corresponding to the saturation region.

Therefore, the I-V relationship of n-channel MOSFET is shown in Figure3. In the linear region, $I_{DS} = \frac{1}{2}g_n(V_{GS} - V_{thr})^2$ while in saturation region, with the increase in V_{DS}, I_{DS} remains constant. And when V_{GS} higher, I_{DS} becomes higher as well. And I_{DS} will only be generated when V_{GS} is greater than Threshold voltage. As shown in Fig.3.



(a) (b)

Fig. 3. (a) I-V characteristic of an n-channel MOSFET **(b)**Relationship between Drain Current and Gate-Source Voltage [2]

2.1.3 Advantages of CMOS

CMOS technology is known for two primary features: its exceptional power efficiency and robust noise resistance. The power consumption is notably low, especially during state transitions, as one transistor in the MOSFET pair remains off. Consequently, CMOS devices generate significantly less heat compared to other technologies like TTL and NMOS, which often exhibit some level of standby

current even when not switching states. CMOS operates efficiently on a single Vcc battery, boasting a high input impedance and excellent temperature stability. These attributes have positioned CMOS as the dominant technology in VLSI chips, making it the most widely adopted choice for integrated circuits.

2.2. Basis of HEMT

The high-electron-mobility transistor (HEMT), also referred to as heterostructure FET (HFET) or modulation-doped FET (MODFET), belongs to the category of field-effect transistors. Its origins are often credited to physicist Takashi Mimura, who embarked on GaAs (gallium arsenide) MOSFET research as a potential alternative to conventional silicon (Si) MOSFETs, laying the foundation for HEMTs starting in 1977. In May 1980, the inaugural D-HEMT device was successfully demonstrated, followed by the introduction of the first E-HEMT device three months later [2]. A significant development occurred in 2004 when GaN (Gallium Nitride) Metal Oxide Semiconductor HEMTs (MOS-HEMT) were showcased. These devices employed atomic oxide deposition (ALD) with an aluminum oxide (Al2O3) film for gate dielectric and surface passivation layer purposes [3].

2.2.1 Anatomy of GaN FET

Typically, the cross-section of GaN HEMT device is shown in Fig.4. This structure has a substrate (usually made of silicon), a buffer made of gallium nitride (GaN), and a barrier made of aluminum gallium nitride (AlGaN). Three terminals—source (S), gate (G), and drain (D)—as well as a passivation layer (protection dielectric) and a field plate are present. To lessen the impact of surface electron traps between the gate and drain contacts, GaN HEMTs require a passivation layer. Additionally, the field plate can lower the surface electron traps and raise the breakdown voltage. A 2-dimensional electron gas channel (2DEG) is created by the heterojunction (a sort of junction between two distinct semiconductors) of an aluminum nitride (AlGaN) barrier and gallium nitride (GaN) buffer. Both the charge density and mobility of this channel are extraordinarily high. The cross-section of a GaN FET based on a Si substrate is shown in fig.4.

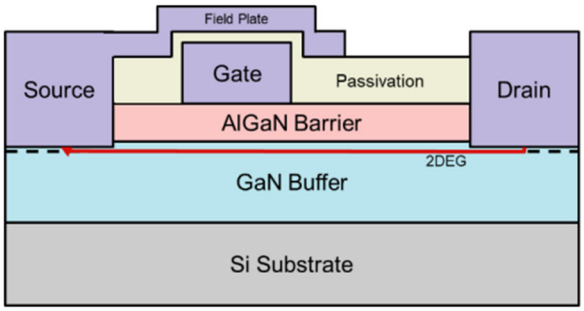


Fig. 4. Cross-section of a GaN FET based on a Si substrate [4]

2.2.2 Electrical Characteristics of GaN

GaN naturally adopts a hexagonal structure known as "wurtzite," which exhibits remarkable chemical stability, rendering it mechanically robust and capable of enduring high temperatures without decomposition [4].

The leakage current between the drain and source terminals fluctuates with temperature. R_{on} represents the cumulative resistance components within the transistor. To establish a connection between 2DEG and the source and drain metals, we encounter contact resistance ((R_C), which is directly proportional to electron mobility, the quantity of electrons generated by 2DEG (N_{2DEG}), electron travel distance (L_{2DEG}), the width of 2DEG (W_{2DEG}), and the universal charge constant, q (1.6×10^{-19} C)). This relationship is mathematically expressed as:

$$R_{2DEG} = L_{2DEG}/(q \cdot \mu_{2DEG} \cdot N_{2DEG} \cdot W_{2DEG}) \quad \text{.....(1)}$$

The electron counts within 2DEG hinges on the strain induced by the AlGaN barrier. However, the 2DEG concentration beneath the gate may differ from the region between the gate and drain. This variance is influenced by the gate type, specific fabrication processes, and the heterostructure used. It also correlates with the gate voltage. A fully enhanced gate exhibits a higher electron concentration than a partially enhanced gate. Thus, an approximate value for transistor resistance can be calculated as:

$$R_{HEMT} = 2 \cdot R_C + R_{2DEG} + R_{2DEG(gate)} \cdots\cdots\cdots\cdots \quad \cdots\cdots\cdots\cdots (2)$$

Additional parasitic resistance (($R_{\text{parasitic}}$) arises from the metal resistance of multiple metal buses conducting current from a single source and drain to the transistor's terminals. In power conversion circuits, transistors incur substantial conduction losses, prompting the device to operate either in an on (ohmic) or off state. Consequently, a crucial parameter for a power transistor is the on-resistance ($R_{\text{DS(on)}}$), defined as:

$$R_{DS(on)} = R_{HEMT(fully\ enhanced)} + R_{parasitic} \cdots\cdots\cdots\cdots \quad \cdots\cdots\cdots\cdots (3)$$

2.2.3 Advantages of GaN FET

Since GaN has a wide bandgap, it enables support of higher voltages. Its higher switching frequencies permit use of smaller passives. HEMTs have high gain, so they can used as amplifiers and fast switching speeds. Since the dominant charge carriers in MOSFETs are majority carriers, with minor participation of minority carriers and extremely low noise figures as compared to other and the current changes is very small [5]. It has low switching energy loss and reverse recovery energy loss; thus, the efficiency will be higher. It has low R_{on} and low cost. Furthermore, in contrast to silicon (Si), GaN material exhibits elevated electron mobility, saturation electron velocity, and breakdown electric field characteristics.

2.3. Comparison

To start, let's examine the five crucial electrical properties, as detailed in Table 1 for three semiconductor materials: GaN, Si, and SiC, all vying for dominance in the power management market. It's evident that GaN possesses a wider bandgap. The wurtzite crystal structure imparts GaN with piezoelectric characteristics, contributing to its exceptional conductivity when contrasted with silicon.

Table 1. Some key properties of GaN and SiC—from Si at 300Kelvin.

Properties	GaN	Si	SiC
Band gap (eV)	3.4	1.12	3.3
E_C-Critical electric field (MV/cm)	3.3	0.3	3.5
V_S-Saturated drift velocity ($\times 10^7 cm/s$)	2.5	1.0	2.0
μ-Electron mobility (cm^2/V_S)	990-2200	1500	650

On-resistance and breakdown voltage have a closer link as a result of the greater critical electric field strength. This makes it possible for devices to be more compact and to have their electrical terminals closer together for a certain breakdown voltage requirement. Higher electron mobility in 2D electronic components is another benefit of gallium nitride over silicon carbide. GaN devices are therefore smaller when the breakdown voltage and on-resistance are constant [6]. GaN FET consumes less power and is more efficient compared to MOSFET.

Last, this compared their costs. Take GaN, SiC and SJ MOSFET as examples. Fig.5 compares projected cost trends for the three technologies. These devices are at different stages of maturity and have different depreciation models. In fact, GaN offers the greatest opportunity for cost savings going forward. Strong market adoption increases manufacturing utilization and accelerates capital depreciation, reducing total cost over time. It obviously that GaN cost the least in the future.

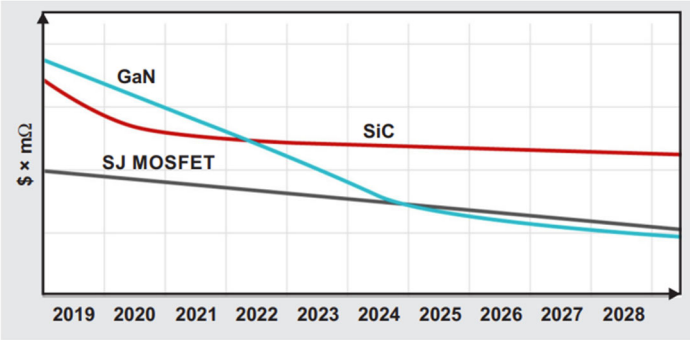


Fig. 5. Relative cost projection by FET technology [6]

3. Application

3.1. Smartphone charger

Traditional chargers are made of silicon, but now GaN can be a replacement. Gan is more thermally and power efficient and has a higher temperature limit and electron mobility as well. In addition, GaN has a bandgap nearly three times that of Silicon, which enables it to conduct electrons at higher voltages.

Therefore, GaN gradually comes into its own in situations where large amounts of power need to transfer quickly. Because of these advantages, GaN chargers have a significantly merit than the traditional silicon-based chargers. GaN chargers also provide more power to devices and can provide all the power to multiple devices at the same time. This makes the multi-device charger even more versatile, allowing it to quickly charge multiple mobile phones and laptops in the meanwhile. Which is impossible of in silicon charger. As Table 2 and content mentioned above, compared with silicon chargers, GaN has many advantages, but its main disadvantage is temporarily limited to high cost. Ultimately, as production unit economics improve, GaN chargers will be cheaper than silicon chargers [7].

Table 2. Some key properties of GaN and Si Chargers

	Silicon chargers	GaN chargers
Bandgap of base materials where no electrons can exist	Silicon:1.14eV	Gallium Nitride:3.4eV
Size	Large sized charges	Smaller sized chargers with compact designs
Efficiency	Not very power and heat efficient	Great power and heat efficiency
Multidevice charging	Chargers usually charge one or two devices simultaneously	Chargers can charge anywhere between 1 to 7 devices
Price	Low price as it is established technology	High price currently as it is new technology. Price should go down as a adoption increases over the years

3.2. Light Detecting and Ranging (LiDAR)

Lidar, a derivative of radar, functions within the optical segment of the electromagnetic spectrum. In a standard lidar setup, light is emitted into the atmosphere or space, and afterward, the system captures the returning light that has interacted with one or more targets positioned at a specific distance from the lidar device [8].

Recent years, manufacturers and technology companies developing self-driving cars are looking to a combination of sensors and cameras which can provides maximum awareness and visibility of surroundings a at a affordable cost for mass production. The new designs using GaN integrated circuits for novel monoliths create true solid-state LiDAR sensors. GaN is blazingly fast, 100 or maybe even 1000 times faster than silicon in a LiDAR application and it is very efficient.

3.3. Aerospace

Advanced GaN HEMTs with exceptional power and efficiency characteristics are now available for applications in communication satellites and navigation satellites operating in the P/L/S bands. Moreover, there are ongoing developments of high-power and high-gain GaN IM HEMTs tailored for X-band applications, such as satellites used in deep space exploration and earth observation. These GaN HEMT devices are engineered to deliver both reliability for space missions and resilience against radiation, making them well-suited for demanding space applications.[9]

4. Development and future

4.1. Challenges

The two biggest obstacles that need to be removed for GaN technology to gain wider acceptance are reliability and cost. With commercial devices guaranteed to work at junction temperatures greater than 200 C and a mean time to failure of more than 1 million hours, the major reliability-related difficulties have mostly been handled. GaN devices have been manufactured initially utilizing 2- to 4-inch wafers, to 6-inch wafers, and more recently to 8-inch (200mm) wafers. Even though early GaN devices were substantially more expensive than competing technologies like silicon, the price gap has significantly shrunk. GaN devices will become more cost competitive as a result of recent innovations and ongoing process advancements that lower production costs [10].

Apart from the price challenges, GaN faces many challenges in terms of application. The fields of signal processing and visualization have not been explored, despite their widespread application. To take a step back and consider the larger 3-D image, though, is very crucial. A key advance is the ability to map the environment inexpensively and correctly on a human size in real time [11].

4.2. Future Direction

GaN technology is just getting started, the journey is in its early years, the amount of integrated functionality will accelerate for years to come. Lateral GaN HEMT technology is a natural fit for power Integrated Circuits due to the ability to reduce parasitic and increase functionality. The inherent capability of GaN technology to combine power-level and signal-level devices on a single substrate could offer the most significant potential for enhancing the efficiency and cost-effectiveness of power conversion systems. A forthcoming initiative entails the development of complete power system-on-a-chip integrated circuits [4]. Conversely, for more advanced GaN HEMT applications, GaN-on-Silicon technology may be the only practical choice.

5. Conclusion

In conclusion, the recent surge of interest in third-generation semiconductor materials, including SiC, GaN, and diamond, has underscored their exceptional performance and the promising prospects they offer in the global semiconductor market. These materials, characterized as wide-band semiconductors with bandgap energies (Eg) equal to or exceeding 2.3 electron volts (eV), distinguish themselves from first-generation elemental semiconductors and second-generation compound semiconductors by their remarkable properties.

Specifically, third-generation semiconductor materials exhibit high thermal conductivity, robust breakdown field strength, impressive, saturated electron drift rates, and strong bonding energies. These attributes make them exceptionally well-suited to meet the evolving demands of modern

electronic technology. In applications requiring resilience under high-temperature, high-power, high-voltage, high-frequency, and radiation-prone conditions, these materials demonstrate immense promise.

In essence, the advent of third-generation semiconductors signifies a significant milestone in the ongoing advancement of human science and technology. As researchers continue to explore their potential and refine their applications, these materials are poised to play a pivotal role in shaping the future of the semiconductor industry, enabling innovations across a wide spectrum of electronic devices and systems. The journey of these third-generation materials is far from over; it is a journey towards a brighter, more efficient, and technologically advanced future.

References

[1] Romli N, Minhad K, Reaz M, et al. An overview of power dissipation and control techniques in cmos technology. J. Eng. Sci. Technol, 2015, 10(3): 364-382.

[2] Mimura T. The early history of the high electron mobility transistor (HEMT). IEEE Transactions on microwave theory and techniques, 2002, 50(3): 780-782.

[3] Ye P D, Yang B, Ng K K, et al. GaN MOS-HEMT using atomic layer deposition Al/sub-2/O/sub-3/as gate dielectric and surface passivation. Proceedings. IEEE Lester Eastman Conference on High Performance Devices, 2004. IEEE, 2004: 167-172.

[4] Butnicu D, Lazăr L C. An Efficiency Comparative Workbench Study of eGaN and Silicon Discrete Transistor based Buck Converters. 2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME). IEEE, 2020: 350-353.

[5] Asif Khan M, Bhattarai A, Kuznia J N, et al. High electron mobility transistor based on a GaN-Al x Ga1− x N heterojunction. Applied Physics Letters, 1993, 63(9): 1214-1215.

[6] Lidow A, De Rooij M, Strydom J, et al. GaN transistors for efficient power conversion. John Wiley & Sons, 2019.

[7] Mishra U K. Gallium Nitride Versus Silicon Carbide: Beyond the Switching Power Supply. Proceedings of the IEEE, 2023, 111(4): 322-328.

[8] Glaser J. How GaN power transistors drive high-performance lidar: Generating ultrafast pulsed power with GaN FETs. IEEE Power Electronics Magazine, 2017, 4(1): 25-35.

[9] Satoh T, Osawa K, Nitta A. GaN HEMT for space applications. 2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS). IEEE, 2018: 136-139.

[10] Bockowski M, Iwinska M, Amilusik M, et al. Challenges and future perspectives in HVPE-GaN growth on ammonothermal GaN seeds. Semiconductor Science and Technology, 2016, 31(9): 093002.

[11] Glaser J. How GaN power transistors drive high-performance lidar: Generating ultrafast pulsed power with GaN FETs. IEEE Power Electronics Magazine, 2017, 4(1): 25-35.