

Nanoscale air channel devices- inheritance and breakthrough of vacuum tube

Baihong Chen^{a,c}, Linjie Fan^{a,c}, Jinshun Bi^{b,c,*}, Zhiqiang Li^a, Ziming Xu^c, Sandip Majumdar^d

^a Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, 100029, China

^b School of Integrated Circuits, Guizhou Normal University, Guiyang, Guizhou Province, 550025, China

^c School of Integrated Circuits, University of Chinese Academy of Sciences, Beijing, 100049, China

^d Serampore Girls College, Hooghly, West Bengal 712201, India

ARTICLE INFO

Keywords:

Vacuum electronics
Nanogap
Simulation
Emission mechanism
Temperature resistance
Radiation endurance

ABSTRACT

The nanoscale air channel device (NACD) has recently gained significant attention as a novel vacuum electronic that can be fabricated through nanofabrication technologies. Here, the research and progress of the NACD since it was reviewed, with a focus on working mechanism analysis, nanofabrication technologies, device structure optimization, electrode materials and simulation approach. Furthermore, the application fields and future development of NACD were summarized and prospected. The NACDs are expected to surpass the physical limits of traditional solid transistors due to its advantages such as smaller heat loss, high-speed, resistance to harsh environments.

1. Introduction

The integrated circuit technology roadmap as shown in Fig. 1, since the invention of vacuum tubes in the early 20th century, the ability to amplify signals without distortion has made it possible for devices such as radios and televisions to transmit or receive electronic signals. Subsequently, the first general-purpose electronic computer “ENIAC”, which could perform 5000 additions or 400 multiplications per second using 18,000 vacuum electronic tubes was introduced. Its computing speed was 1000 times faster than that of electromechanical computers [1]. However, traditional vacuum tubes rely on thermal cathodes for thermionic emission, which exposes various problems such as high-power consumption, short lifespan, large volume, and difficulty in integration.

In the middle of the last century, the transistor was invented and gradually became the mainstream of electronics. According to Moore's law [2], the number of transistors per unit area in integrated circuits are doubled every year. This push for miniaturization brought complexity to the next generation of transistors from microelectronics to nanoelectronics. In recent years, with advanced nano process technologies were proposed, multiple new materials such as two-dimensional ultra-thin monolayer materials [3,4], high mobility materials of the III-V group semiconductors [5], and new structures such as fin field-effect transistor (FinFETs) [6], gate-all-around transistor (GAAFET) [7] have been developed, the critical dimension (CD) of transistors is continuously

reduced, resulting in a higher level of integration for the chip. At present, solid-state electronics have gradually approached their physical limits, resulting traditional silicon-based transistors facing increasingly serious problems such as short channel effects, parasitic resistors, parasitic capacitors, and quantum tunneling effects. This leads to the entry of the integrated industry into the post Moore era.

T K S Wong fabricated a lateral tunnelling junction with the sub 50 nm vacuum channel in 1993 [8], A A. G. Driskill Smith developed a nanoscale tip field emission system in 1997 [9], J W Han et al. proposed a nanoscale vacuum channel transistor in 2012 [10], which are all equipped with nanoscale vacuum channels and exhibit excellent performance similar to vacuum tubes. However, the requirement of vacuum packaging and complex nanofabrication processes make it more difficult to prepare the nanoscale vacuum channel. This situation was not changed until the vacuum channel metal oxide semiconductor (MOS) field effect transistor [11], the single grain boundary (GB) structure [12] and the lateral quasi-vacuum channel field emission transistor (VFET) [13] were proposed. Although, with quite different names and structures, the devices have the common features: the air channel of the devices are all in nanoscale, which are closed to or less than the mean free path (MFP) (≈ 68 nm) [14] of electrons at room temperature and pressure (RTP). This enables the electrons are transported through ballistics without scattering in the air channel, which can be explained by the Space Charge Limited (SCL), the Schottky effect and the Fowler-Nordheim (F–N)

* Corresponding author. School of Integrated Circuits, Guizhou Normal University, Guiyang, Guizhou Province, 550025, China.

E-mail address: bijinshun@gznu.edu.cn (J. Bi).

<https://doi.org/10.1016/j.nanoms.2024.01.002>

Received 7 September 2023; Accepted 27 December 2023

Available online 19 February 2024

2589-9651/© 2024 Chongqing University. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

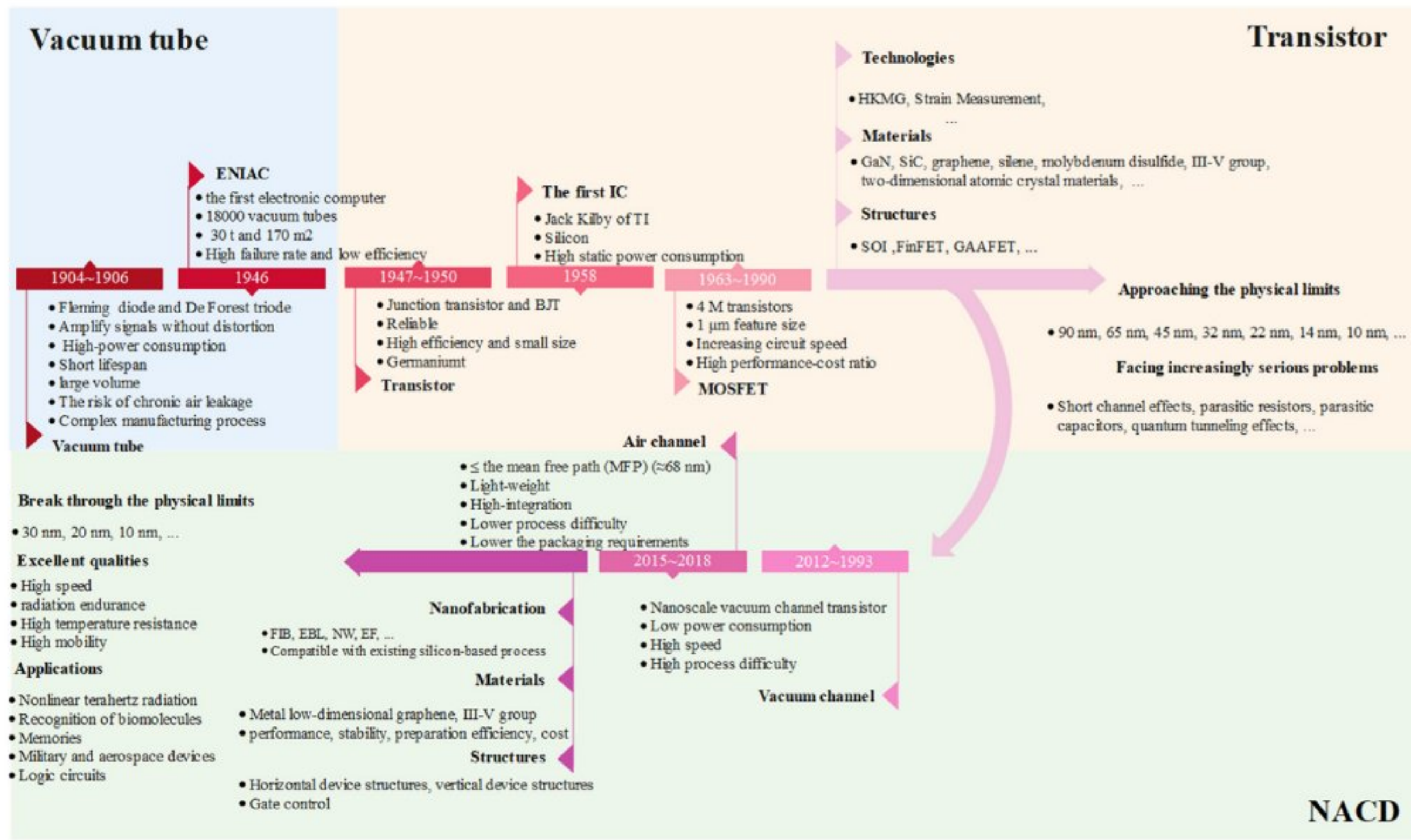


Fig. 1. The integrated circuit technology roadmap. The invention of vacuum tube has promoted the electronic industry, but large volume, large power consumption and difficulty in integration have hindered its development. Since the transistor was put forward in 1947, the “silicon age” has been opened. With the evolution of technologies, transistors have experienced several changes, the CD is continuously reduced, but it gradually approached physical limits. The excellent-performance NACD is expected to break through the physical limits and become another new scheme after transistors.

tunneling. For simplicity, Nanoscale channel devices with different vacuum levels and field emission nanodevices are uniformly called the nanoscale air channel devices (NACDs) next. The NACDs have basic functions similar to traditional vacuum tubes, but the NACDs’ fabrication process such as focused ion beam (FIB) etching, electron beam lithography (EBL), nanowire (NW) and electro-forming (EF), etc. can be compatible with existing silicon-based process, which are suitable for the

multiple device structures, electrode morphology and electrode materials. Ballistic transport without scattering allows the NACD to achieve extremely fast response speed, show the radiation endurance and high temperature resistance, avoid the short channel effects and thermal losses, and the NACD is expected to break through the physical limits of solid-state devices and achieve the smaller CD. The NACD has demonstrated competitiveness in fields such as THz technologies, national

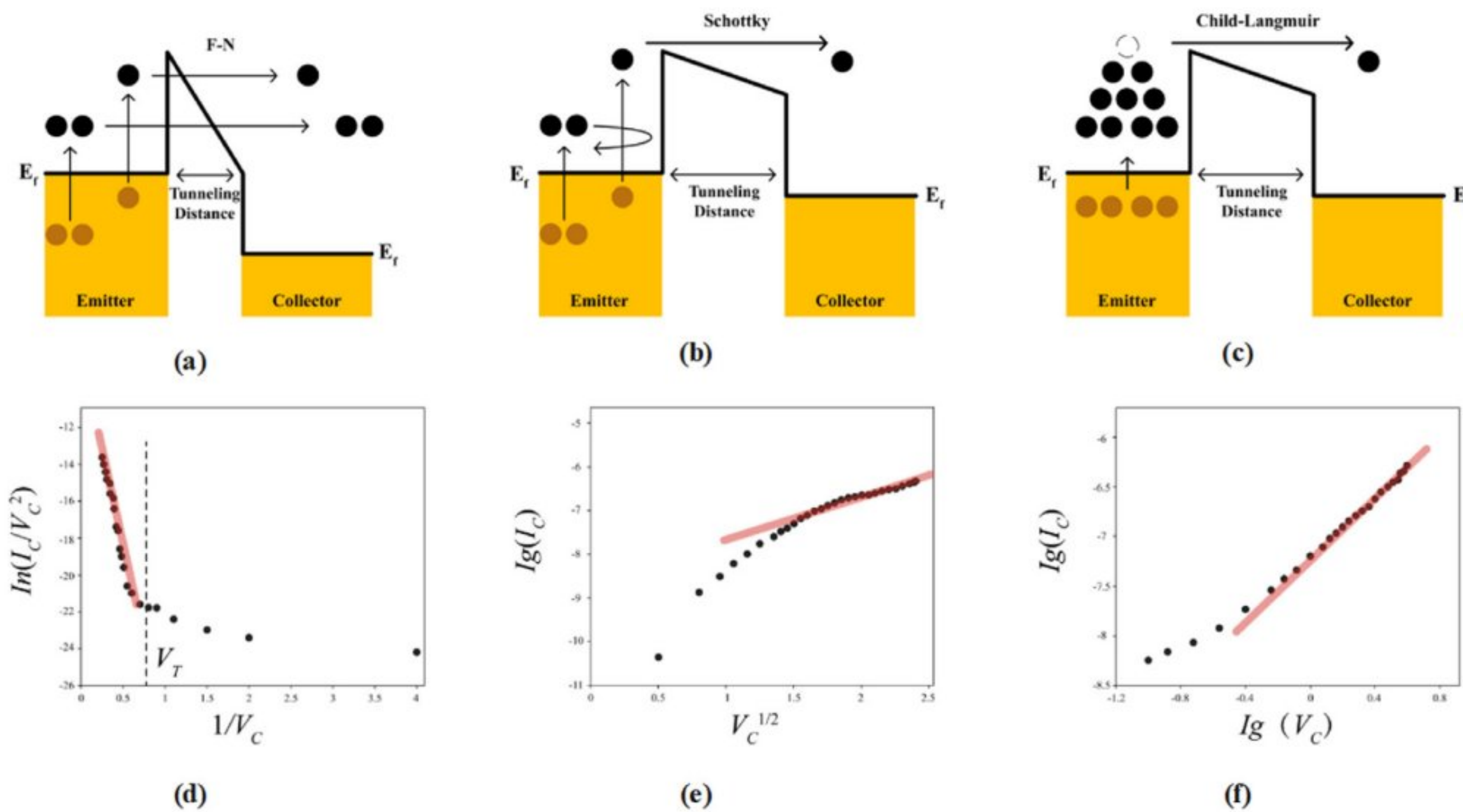


Fig. 2. The band diagram and fitting curve of the three basic emission mechanisms. (a) Fowler-Nordheim (F–N) band diagram of the NACD. (b) The Schottky effect can be explained as image force assisted field emission of electron. (c) According to Space Charge Limited (SCL), the electrons will move towards the space charge zone. The fitting of the actual current emission curve (I–V) and emission mechanisms curve of the (d) F–N, (e) the Schottky effect and (f) the SCL [20,21].

defense, space exploration. Therefore, how to shrink the air channel CD, control the electrical performance, optimize the device structures and fabrication process and grasp the electron emission and transport mechanism more accurately have become more important.

2. Emission and transport mechanisms of the NACD

The field emission theory is the basic working mechanism for electron transport through the channel in the NACD. In this section, we introduce Fowler-Nordheim (F–N) tunneling model [15] and Schottky effect to explain the electron emission mechanisms, and Space Charge Limited (SCL) to explain the current limitation theory.

2.1. F–N tunneling emission mechanism

Fowler-Nordheim is the most commonly used theory for modeling field-induced electron emission of the NACD. Fig. 2(a) shows the band diagram of the NACD. When the operating voltage is high, the width of the barrier can be compressed. Thus, electrons can tunnel through triangular air channel barriers experiencing F–N emission, which can be modeled analytically as [16–18]:

$$J = A \frac{\beta^2 V^2}{\varphi d^2} \exp\left(-B \frac{\varphi^{3/2} d}{\beta V}\right) \quad (1)$$

Where J is field emission current density, A , B is the field emission constants, β is an enhancement factor which is usually related to the morphology and material of the emission electrode, φ is the work function of the emission electrode material, the channel size (d) and V is the working voltage. β can be modeled [16–19] as show in equation (2), where l is the emission electrode length and r is the emission electrode tip curvature. We can find from equations (1) and (2) that the electrode morphology and the work function of the electrode material will affect the current density.

$$\beta = \frac{2l/r}{\ln(4l/4) - 2} \cdot \frac{1}{d} \quad (2)$$

The actual current emission curve (I – V) and the corresponding mechanisms curve of field emission devices are usually drawn in the following introduction of field emission experiments, as shown in Fig. 2(d–f), so that the electron emission performance of cathode materials or emission structures can be effectively estimated. In Fig. 2(d), the linearity tendency of $\ln(I/V^2)$ versus $1/V$ testifies that F–N tunneling mechanism is dominated for electron transportation at high voltage; F–N emission would appear linear with negative slope. By taking the logarithm of both side, equation (1) turns into

$$\ln\left(\frac{I}{V^2}\right) = -B \frac{\varphi^{3/2} d}{\beta} \frac{1}{V} + \ln\left(A \frac{\beta^2 A}{d^2 \varphi}\right) \quad (3)$$

This means at higher operating voltage; the electron transport mechanism is fitted to Fowler Nordheim (F–N).

The electron emission mechanism of the planar vacuum field emission triode (VFET) prepared by Xiao Wang et al. [20], the nanoscale vacuum channel transistor (NVCT) reported by Jin-Woo Han et al. [22], the nano vacuum channel (NVC) electronics showed by Marco Turchetti et al. [23] and the sub-100 nm air channel proposed by Meng Liu et al. [24] conforms to F–N tunneling emission mechanism at higher voltage. However, F–N tunneling emission mechanism is no longer applicable at lower voltage because of the larger scattering probability of electrons in air channel, the wider barrier and the lower energy of electrons. Shrutu Nirantar [25] believed that at the low operating voltage is low, electrons tunnel through a rectangular air channel barrier experiencing direct tunneling (DT), which can be expressed as [26].

$$J \propto \frac{CV(m_{eff}\varphi)^{1/2}}{d} \exp\left(-D(m_{eff}\varphi)^{1/2}d\right) \quad (4)$$

where C and D are the field emission constants, m_{eff} is the tunneling effective mass. The current density of DT is independent of temperature and has a linear relationship with voltage. DT is suitable for small voltage ranges, while F–N is suitable for large voltage ranges.

2.2. Schottky effect

At the metal-air interface, electrons outside the metal will induce a positive charge on the metal surface, and electrons will be attracted by this positive charge, this attraction is called image force. The metal-air interface barrier caused by the image force becomes lower with the increase of the reverse voltage and electrons will overcome the energy barrier, as shown in Fig. 2(b). The barrier-lowering effect due to the image force is called as Schottky effect. Under the influence of an electric field, the temperature required for purely thermally excited electron emission can be significantly reduced through the Schottky effect. However, during the Schottky effect process, the electrode remains hot. Schottky emission is a thermally activated electron emission process enhanced by an electric field [27]. This is the Schottky effect emission dominant mechanism in low voltage operating mode, and the Schottky effect can be expressed as [28,29].

$$I \propto T^2 \exp\left(\frac{-\varphi + \frac{q}{2} \sqrt{\frac{q\gamma V}{d\pi\epsilon_0}}}{k_B T}\right) \quad (5)$$

Where γ is the field enhancement factor, ϵ_0 is the vacuum dielectric constant, φ is the work function of the emission electrode material, q is the basic positive charge, k_B is the Boltzmann constant. Schottky emission depends strongly on temperature, while FN emission does not, Marco Turchetti et al. [23] extracted electrical characteristic curves at different temperatures and found that Schottky effect is strongly correlated with temperature. In Fig. 2(e), the linearity tendency of $\lg I$ versus $V^{1/2}$ testifies that Schottky effect is dominated for electron transportation at low voltage. By taking the logarithm of both side, equation (5) turns into

$$\lg I \propto \frac{\sqrt{V}}{T} \quad (6)$$

$\lg I$ and $V^{1/2}$ satisfy a positive slope linear relationship, Xiao Wang et al. [20] fitted the Schottky effect with the actual current emission curve (I – V) of the planar lateral vacuum field emission triode (VFET) drain voltage at 0 V–4 V, and the fitting was good. Ji Xu et al. [21] found that using Schottky emission by thermal effect is the most suitable method for fitting high quality nanoscale vacuum channel transistors (NVCTs) under low voltage operating conditions, at lower operating voltages, only a few electrons could overcome the broad barrier due to the thermal energy.

2.3. Space charge limited current theory

The electron transport of the NACD is influenced by the Space Charge Limited (SCL) current theory. The electrons move towards the surface of the electrode and accumulate, forming a space charge zone between the two electrodes. As the electric field at the two electrodes gradually increases, the accumulated charge obtains enough energy to cross the potential barrier and form a small emission current, as shown in Fig. 2(c). Due to the repulsion of space charges, which are accumulated near the electrode, some emitted electrons will not reach the collector electrode, and only restricted electrons can reach it [23]. Before the SCL threshold, all electrons reach the collector. After the SCL threshold, only a part of the restricted electrons reaches the collector [27]. The traditional SCL can be modeled as [30,31].

$$J = K\varepsilon_0 \sqrt{\frac{2e}{m}} \times \frac{V^{3/2}}{D^2} \quad (7)$$

Where ε_0 is the permittivity of free space, K is the correction coefficient, which is usually taken as $4/9$ and is related to the characteristics and structure of the electrode material. D is the distance between the two electrodes, m and e are the electron's mass and charge. In Fig. 2(f), the linearity tendency of $\lg I$ versus $\lg V$ testifies that SCL is dominated for electron transportation. By taking the logarithm of both side, equation (7) turns into

$$\lg I = \frac{3}{2} \lg V + \lg \left(\frac{KA\varepsilon_0}{D^2} \sqrt{\frac{2e}{m}} \right) \quad (8)$$

$\lg I$ versus $\lg V$ satisfy a linear relationship with a positive slope of $3/2$.

In summary, the various transport mechanisms of the NACD may become very complex due to factors such as operating voltage range, electrode morphology, temperature, and electrode material, and cannot be explained by a unified model. In general, the F–N emission mechanism can better fit the electron transport of the NACD at high operating voltage, and the electron emission state is relatively ideal. When the working voltage is lower, the electron emission state becomes complex. A smaller operating voltage is insufficient to provide sufficient energy for electron to cross the potential barrier, the electron transport can be explained by the space charge zone between the electrodes or the thermal assisted field emission of electron. Therefore, understanding the electron transport mechanisms in air channels is of great significance for the

NACD. These theoretical models provide guidance for researchers to explore the working principle, the better emission electrode solutions and device design of the NACD.

3. Nanofabrication of the nanoscale air channel

Nanofabrication is used to achieve channel less than MFP of electrons at RTP, which allows the NACD to transport carriers without scattering like vacuum tubes in air. Recent efforts in this direction have addressed channel scale nanofabrication of the NACD by FIB etching, EBL, deep reactive ion etching (DRIE) and buffered oxide etching (BOE).

3.1. FIB etching for nanoscale air channel

Under the excitation of strong electric field, a gallium ion beam is formed by liquid metal ion source (LMIS) of the FIB etching system and injected into the first lens group. At this time, the gallium ion beam will be focused and the injection energy will be increased and sent to the primary deflector to reduce astigmatism. Separation deflector will purify ion beam according to different ion charge. The secondary deflection of the purified gallium ion beam further corrects astigmatism and is sent into the secondary lens to refocus the spot with higher resolution. In this way, a gallium ion beam with higher energy than photons and electrons with a specific incident angle is obtained. Gallium ion beam sputtering on the target will transfer energy to the atoms of the target so that they can break away from the surface of the target, at the same time, a large number of rebound atoms will continue to collide with other atoms on

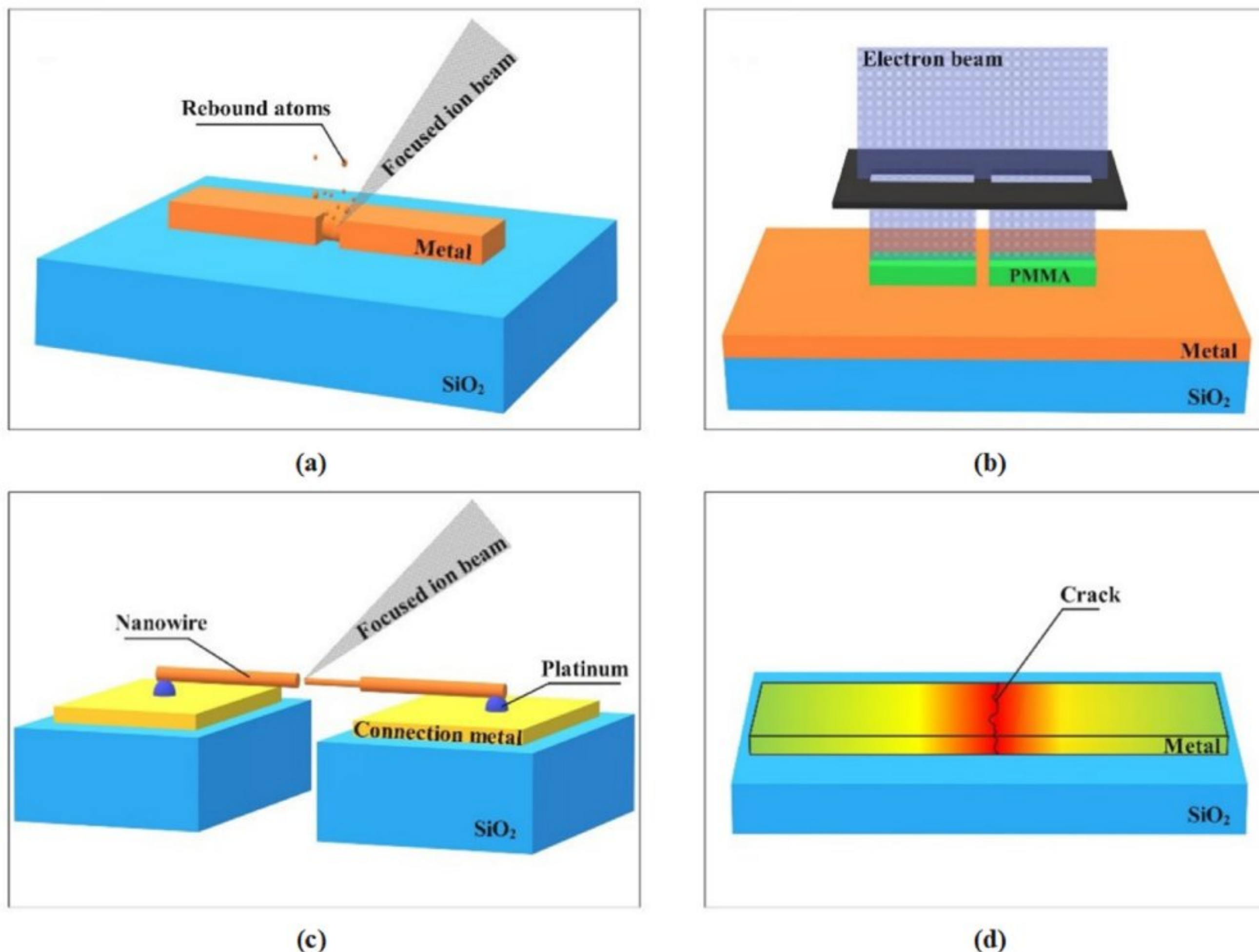


Fig. 3. The four main nanofabrication processes for achieving the NACD. (a) Focused gallium ion beam sputters on the target will transfer energy to the atoms of the metal so that they can break away from the surface of the metal, at the same time, a large number of rebound atoms are produced. (b) The electron beam reacts with the PMMA photoresist to break chemical bonds and change the solubility of the photoresist, after development, the exposure part is left to form a pattern. (c) FIB etching cuts off the thinned nanowire to achieve nanogap. (d) when a large current flows through a metal film, Joule heat is generated, resulting in the formation of cracks in the metal film to achieve nanogap.

the surface of the target, so that more atoms will leave the surface [32, 33], as shown in Fig. 3(a).

Siwapon Srisonphan et al. [11] utilized FIB etching technology to achieve a vertical structure of 70 nm nanogap. The nanogap structure is a three-terminal device, which with a thin layer of indium tin oxide gate, Si cathode and Al anode. The three electrodes are divided by SiO₂ layers, then a deep well is etched using FIB, allowing field emission electrons to reach the top electrode through the deep well. The sub-5 nm co-planar metal–insulator–metal junctions fabricated by M. Saifullah in 2002 [34], the metallic electrodes with nanometer gaps showed by Takashi Nagase [35], the suspended tungsten electrodes with a gap of less than 2 nm prepared by K. Shigeto in 2006 [36] and the planar four-electrode nanogap devices proposed by Linjie Fan [37] in 2022 are all processed by FIB etching technology.

FIB etching technology, with its high-energy ion beam, can directly sputter and etch the target material without the masks and photoresists, so that there is no photoresist residue during the fabricating process, the reliability of the NACD is enhanced and equipment pollution is prevented. Additionally, the three-axis movable FIB stage and the flexible zoom of LMIS can achieve very complex three-dimensional structures and modify the diverse electrode morphology of the NACD. However, issues such as redeposition, poor homogeneity, non-mass production and slightly low resolution cannot be ignored.

3.2. EBL for nanoscale air channel

EBL utilizes a focused electron beam with extremely short wavelengths to act on the photoresist to achieve predetermined patterns, which is expected to break through higher resolution. Control the electron optical device to select scanning such as raster scanning, vector scanning, Gaussian beam scanning and variable beam spot scanning. PMMA photoresist exposure as shown in Fig. 3(b), when reaching the photoresist surface, the focused electron beam will scatter at a small angle, and then it will cross the photoresist and reach the substrate, and the electrons will backscatter at a large angle. During the whole lithography, a series of secondary electrons will be produced because of the inelastic collision scattering deceleration between the incident electrons and the photoresist. The electrons react with the PMMA photoresist to break chemical bonds and change the solubility of the photoresist [35]. Compared to traditional lithography techniques, EBL has become another mainstream technology for achieving the NACD due to its advantages of achieving higher resolution, short fabrication cycle, mass production, excellent homogeneity and simple fabrication process.

Shruti Nirantar et al. [13] utilized EBL technology to achieve a planar structure nanogap transistor with a back gate. Ji Xu et al. [38] proposed a gold nanogap array, which was processed by EBL and FIB etching technologies. The team also used EBL to implement high-quality nano vacuum channel devices [21], nano vacuum channel transistors [39] and asymmetric tantalum electrode planar field-effect diodes [40]. W. Max Jones [41] produced sub-20 nm gaps between deposited emitter and collector metals in symmetric two-terminal field emission devices by EBL.

Although chip-level integrated devices can be produced in mass, there is still a serious problem-proximity effect. The secondary electrons can also break the chemical bonds of photoresist leads to proximity effect, which will be more serious, when the gap size reaches 10 nm. However, compensating exposure and reducing photoresist thickness can optimize proximity effects [42]. In addition, the residue of photoresist in manufacturing will increase the risk of equipment pollution.

3.3. Nanowire (NW) for nanoscale air channel

In order to achieve a smaller gap distance and increase the field enhancement factor to achieve low turn-on voltage and high emission current effects in the NACD, The Meng Liu [43–45] used the nanowire fabrication process to manufacture nanogap, and then examine the

emission characteristics of the NACD. Meng Liu et al. synthesized single crystal nanowire clusters by non-aqueous synthesis [43] or chemical vapor deposition (CVD) [44]. Next, selected a relatively complete nanowire and adhered one end of it to a blunt tungsten tip, and then separated the selected nanowire from the nanowire cluster by FIB etching and assembled the selected nanowire between two gold electrodes by platinum deposition. The process of nanogap was divided into narrow down and cut off, for narrow down, a resistance measuring device was set at both ends of the nanowire, the device is a two-port device with one port set high level and the other port grounded. Meng Liu et al. has been monitoring the resistance value in FIB etching process. Finally, the target channel width is cut off at the thinned position, as shown in Fig. 3 (c). The thickness of nanowires and the size of nanogap can be controlled by changing the etching time and etching strength of FIB etching.

The NACD implemented by nanowire technology can be controlled around 10 nm, but the growth, selection and transfer of nanowires are too complex to be integrated and massively produced.

3.4. Electro-forming (EF) for nanoscale air channel

Electro-forming is shown in Fig. 3 (d), when a large current flows through a conductive film, Joule heat is generated, resulting in the formation of cracks in the conductive film and the achievement of nanogap [46–49]. Compared to other nanofabrication process, the electro-forming process has higher fabrication efficiency, lower cost, and lower fabrication difficulty. However, the position of the fuse crack, the width of the fuse crack and the shape of the fracture surface cannot be controlled. These are the serious problems of electro-forming in the processing of nanogap. Xiao Wang et al. [20,50] showed a layer of PdO conductive film with symmetrical notches, they found that when a current flowed through the PdO film, a large amount of Joule heat would accumulate at the tips of the symmetrical notches. A channel gap of 80–90 nm will be formed by inward fusing between two symmetrical notches. They proposed the uneven fracture surface will change the enhancement factor of the NACD, which is formed by the EF process.

3.5. Other nanofabrication technologies

In addition to the four main nanofabrication processes for achieving nanogap mentioned above, other process methods have also been proposed: Alexandre R et al. [51] used EBL technology to obtain narrower connected electrodes, and then set pushing screw at the bottom of the Si sheet to squeeze the Si sheet upwards, causing the connected electrodes to fracture from the middle. As the pushing screw pushes upwards, the width of the fracture will also change, resulting in a nano channel transistor with adjustable width. S A Guerrero and A I Akinwande [52] obtained Si nanowire emission arrays using the DRIE process. Meng Liu et al. [24] utilized the BOE process to etch SiO₂ between the metal and Si to achieve an 80 nm vertical nanogap, so that electron can be transported between metal electrode and Si electrode. Yazhou Wei [53] proposed a vertical GaN NACD prepared by BOE with a 27 nm air channel.

In summary, EBL, FIB, NW, BOE and EF can be used to fabricate nanoscale gaps. As shown in Table 1, NW needs to be combined with FIB to fabricate the NACD, but due to the difficulty of nanowire preparation technology and significant process fluctuations, NW is not easy to achieve large-scale integration, but special nanodevice structures can be customized. The difficulties in grasping the location and size of the nanoscale air channel, the significant process fluctuations, and the incompatibility with traditional silicon-based processes make EF unsuitable for practice. For the NACD produced by BOE, a large current density can be obtained by reducing the thickness of the oxide film between the two electrodes to shrink the CD, or by increasing the relative area size of the two electrodes significantly increase the current density. Selective etching is also expected to miniaturize the CD while maintaining process consistency of BOE. However, BOE is limited to the vertical structure NACD, which makes it difficult to achieve large-scale integration goals.

Table 1

Comparison table of several fabrication processes for the NACD. Comparisons were made on the minimum channel size (CD) of different fabrication processes, whether mass production is possible, and the structure and process consistency of NACD devices that can be prepared (the more+, the better the process consistency).

Technology	Critical dimension (nm)	Mass production	Device structures	Consistency
FIB	10	Can	Horizontal/ vertical	+++
EBL	15	Can	Horizontal	+++
NW + FIB	10	Cannot	Horizontal/ vertical	++
EF	80–90	Cannot	Horizontal	+
BOE	27	Cannot	vertical	++++

FIB and EBL have high flexibility and can process various device structures. Advances in nanofabrication technology are advancing to reduce the CD and enhance process consistency. FIB is more suitable for modifying electrode morphology, EBL has higher production efficiency.

4. NACD structures

There are two typical structures of the NACD: 1) Horizontal device structures. The two terminals are on the same plane and the current flows in the horizontal direction. 2) vertical device structures. The two terminals are located on different horizontal plane, so the current flows in the vertical direction. In this section, the two kinds of structures are summarized, with introduction of the different gate morphology and the gate control effect of the NACD.

4.1. Horizontal device structure

The horizontal NACD structures can be formed by a pattern process which is compatible with existing silicon-based process technologies. In 2012, Jin-Woo Han proposed a horizontal structure vacuum nanogap transistor with back gate structure [10]. On a SiO_2/Si substrate, a continuous array of Au nanoribbons is formed by EBL, the nanogaps with

a width of 10–30 nm by the FIB process. Two years later [21], the team showed a high-quality NVCT with a 60 nm Au layer deposited by electron beam evaporation with 5 nm Cr as adhesion layer. The NVCT showed remarkable performance of high drive current ($>10 \mu\text{A}$), low work voltage ($<10 \text{V}$) and high on/off current ratio ($>10^4$). The NVCT with four terminals (one emitter and three collectors) improved the collection efficiency of emitted electrons and electrical performance. Subsequently [39], the team prepared a three-terminal (emitter, collector, and gate) NVCT with a sub-100nm gap by EBL, which has high-speed, high-frequency electrical properties, high switching ratio ($>10^3$), large working current ($>100 \text{nA}$), and small driving voltage ($<20 \text{V}$) performance. In the same year [40], a metal based asymmetric electrodes for field emission (FE) diode horizontal structure was proposed by EBL. The diode has a tip cathode and blunt semicircular anodes, the current FE diodes can be effectually different the forward current from reverse current. Aniello Pelella et al. [54] proposed a horizontal NACD with back gate as shown in Fig. 4 (a). Jin-Woo Han et al. of NASA Ames Research Center [22] showed the horizontal NACD with surround gate, as shown in Fig. 4 (b). Xiao Wang [55,56] proposed a 110–200 nm horizontal nanogap structure with a top gate and a bottom gate, as shown in Fig. 4 (c).

In order to increase the controllability of conductive channels, transistors have evolved from a single gate to FinFETs, which have since been replaced by GAAFET [57]. The gate between the two terminals of the NACD can effectively control the bandwidth and adjust the band height, non-thermally excited field electron emission. In addition, the control effect of the gate on the field emission current can be improved by changing the gate topography, changing the thickness of the gate dielectric and changing the distance between the gate and the emitter electrode. When the gate of the NACD [54] is set a positive gate voltage, the doping level of MoS_2 channel will change, so the higher conductivity makes it easier for electrons to pass through the channel and emit large current, Aniello Pelella attributed this phenomenon of back gate enhancing emission current to the increase of enhancement factor. The surround gate of the NACD [22] increased the electric field by 2.2 times. The top gate is twice as effective as the bottom gate in regulating the current, the closer the distance between the gate and the electrode, the stronger the control effect of the gate on electron emission [55,56].

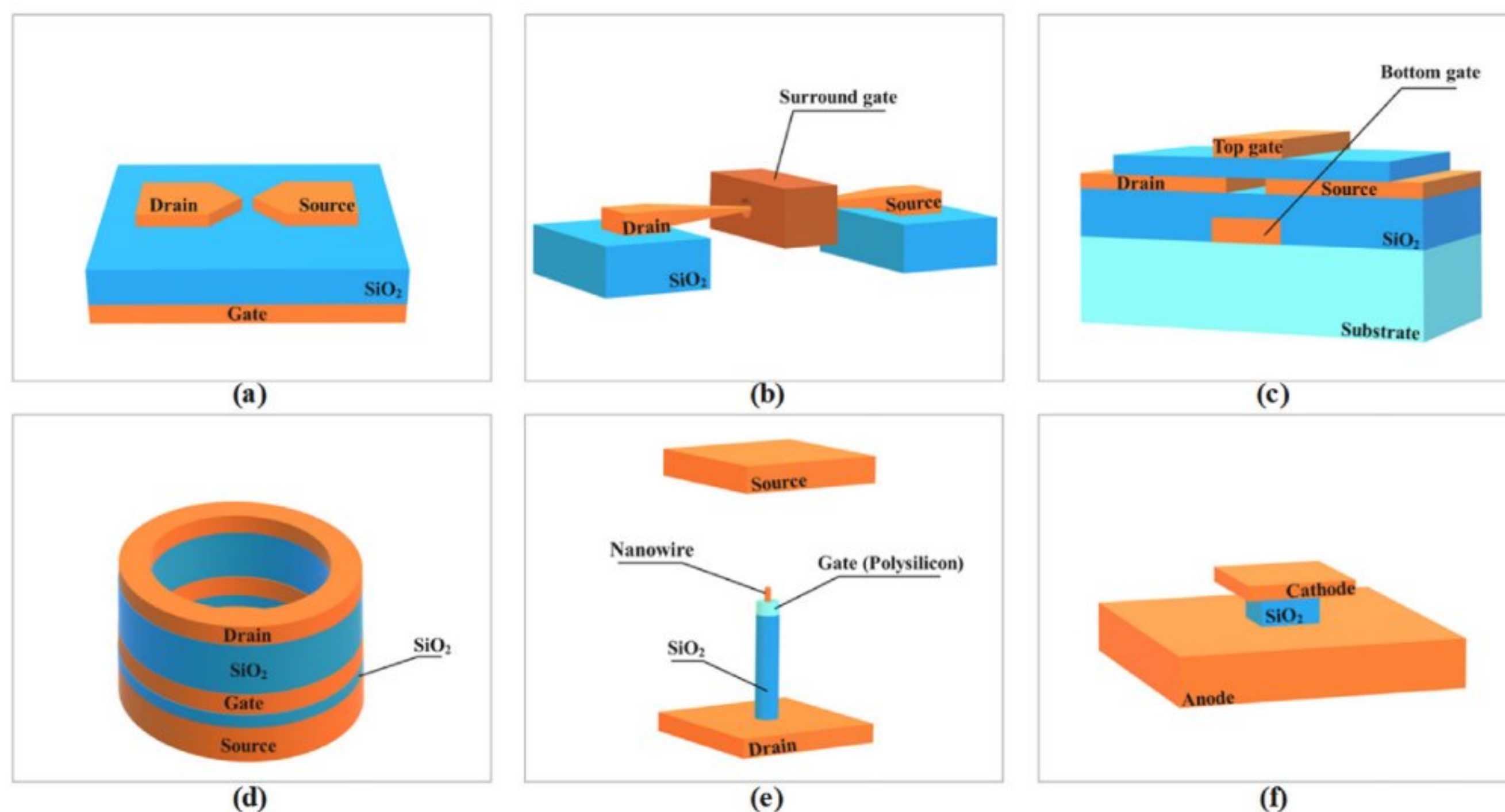


Fig. 4. The horizontal structure and vertical structure NACDs. (a) The horizontal NACD with single back gate. (b) The horizontal NACD with surround gate. (c) The horizontal NACD with a top gate and a bottom gate. (d) The three-terminal vertical NACD. (e) The vertical nanowire array with surround gate. (f) The two-terminal vertical NACD.

4.2. Vertical device structure

Horizontal structure has gradually become the mainstream NACD structure choice because of its compatibility with traditional silicon-based device process lines, excellent electrical performance and mass production. However, in recent years, many novel structures and fabrication have emerged, such as the vertical structure of the NACD, the terminals of vertical structure are on the different planes, and the current flows in the vertical direction.

Siwapon Srisophon et al. [11] proposed a vertical structure of 70 nm nanogap. The nanogap structure is a three-terminal device, as shown in Fig. 4 (d). Firstly, stack the five-layer structure of drain-gate insulator-gate-gate insulator-source, and then use FIB technology to etch a deep pit running through the five-layer structure, exposing the drain, gate, and source. This allows the electrons emitted from the bottom source to reach the top drain vertically, and the gate surrounds the two electrode channels to regulate the emission current. In 2022, Infineon [58] proposed a vertical nanogap structure capacitor with breakdown voltages up to 6×10^9 V/m. Description conforming to Paschen theorem [59]. S A Guerrero and A I Akinwande [52] obtained Si nanowire emission arrays using the DRIE process as shown in Fig. 4 (e), An electrode plate is arranged above the nanowire to receive electrons emitted from the tip of the nanowire, and the gate around the tip of the nanowire can control the emission current. Meng Liu et al. [24] used the BOE wet etching process to remove SiO₂ between the metal electrode and Si substrate, as shown in Fig. 4(f), allowing electrons to emit in the vertical direction.

In summary, the vertical nanogap structures can achieve more device types (diodes, transistors ...), has a high design freedom, and the distance between electrodes can be reduced by adjusting the planar structure and optimizing the electrode morphology. The horizontal structure NACD being more potential in the field of circuit integration compared to the vertical structure NACD. The minimum electrode spacing is constrained by the capabilities of the lithography tool used, the horizontal structure NACD is prepared by EBL, which can achieve smaller CD and better process consistency compared to the other NACD fabrication technologies. However, compared to lithography of silicon-based integrated circuits, EBL exhibits higher cost and lower efficiency. In the future, with the development of lithography technology, it is expected to find a low-cost and high-efficient NACD fabrication technology to achieve large-scale integration. Before that, the NACD still faces many challenges. In order to further reduce the distance between the two terminals in the vertical direction, the thin film thickness can only be continuously reduced. The reduced thickness is not only limited by the ability of lithography tools, but also by the risk of electrical performance degradation caused by thin film thinning. Of course, for the NACD prepared by BOE process, if the vertical structure continuously thins the film to reduce electrode distance and increases the top metal electrode area, it can actually increase current density to optimize electrical performance, or if more convenient integration processes for vertical structures are proposed, then the vertical structures is still very attractive.

5. Electrode materials and morphology

Regardless of the fabrication process and structure of the NACD adopted, we must carefully select electrode materials, which require both excellent electrical performance and sufficient stability to avoid electrical performance fluctuations caused by strong field evaporation. Electrode materials and morphology will affect the performance, stability, fabrication efficiency and fabrication cost of the NACD. In recent years, except for metal, some new electrode materials, such as high melting point noble metal low-dimensional graphene or III-V group semiconductors have emerged, which makes the emission performance of the NACD better, while the fabrication cost and efficiency are not satisfactory. The tip electrode becomes sharper, so that the emission current density will gradually increase, but the working stability will deteriorate. In addition,

the emission current density of the NACD is related to the work function of the electrode material. Therefore, we need to weigh the needs to choose electrode materials and electrode morphology.

5.1. Metals for electrode materials

The metal electrode with high melting point/work function are shown in Table 2, can endure strong field evaporation, which is the choice of electrode material for the NACD. There is a negative correlation between the enhancement factor of metal materials used in metal electrode and the current density of the NACD, so we prefer to choose metals with small work function for higher emission current density.

Linjie Fan [37] demonstrated a two-terminal Au electrode NACD prepared by FIB process. Shruti Nirantar [13] showed a NACD with tip gold electrode with back gate structure by electron beam etching. Compared with the performance of W and Pt electrode materials, it is found that gold electrode can improve the field electron emission performance without technological process limitation. Cu electrode with 20 nm channel gap can control the turn-on voltage at about 1.75 V [43]. Tantalum has high melting point and can be used as capacitor electrode to increase the electric field [60], Or asymmetric metal cathode of the field effect diode [40].

5.2. Semiconductor for electrode materials

Relying on the mature Si-based integration process, Si is a common choice for the NACD electrode materials for integration. S A Guerrero and A I Akinwande [52] obtained Si nanowire emission arrays. Jin-Woo Han et al. [22] proposed a sub-50nm NACD with surround gate structure and Si electrode, which has high driving current ($>3 \mu\text{A}$) at a low opening voltage (<5 V). The team [10] tried to soften and polish the tip of Si electrode by using photoresist reflow process to continuously reduce the nanogap distance. In addition to Si, some metal oxide semiconductors such as ZnO are also used as electrode materials for NACD, Masahiro Ohara [61] presented a vertical structure NACD with ITO glass anode and ZnO cathode prepared by the simple solution and hydraulic process.

Compared with Si, III-V group semiconductors exhibit excellent electrical properties such as radiation resistance, thermal stability and high breakdown voltage [62]. The electron affinity of semiconductor materials is also negatively correlated with the current density. According to related reports, the electron affinity of GaN is 3.1–3.6 eV [63,64], and that of Si is 4.05 eV, therefore, the NACD with GaN electrode can get higher current density than Si electrode in theory. In fact, the actual current density of nanoscale air channel devices using GaN as electrode material is influenced by electrode morphology, electrode processing technology and doping concentration [65,66]. However, heteroepitaxial process compatibility is still a problem to be considered in selecting III-V group semiconductors. Sapkota KR team [67] used metalorganic chemical vapor deposition (MOCVD) to epitaxial a Si-doped N-type GaN layer on a C-doped GaN layer/c-sapphire layer, and then used electron beam etching process to form a 26 nm gap as shown in Fig. 5 (a) GaN has better thermal inertness and chemical inertness [68], which improves the stability of emitter electrode. Smaller electron affinity can increase emission current and lower turn-on voltage (0.24 V).

Table 2
Summary of melting points (°C) and work functions (eV) of some metals.

Metal	Melting point (°C)	Work function(eV)
Pt	1772	5.7
Au	1064	5.1
Cu	1083	4.65
W	3410	4.5
Al	658	4.28
Ti	1800	4.33
TiN	2930	4.66

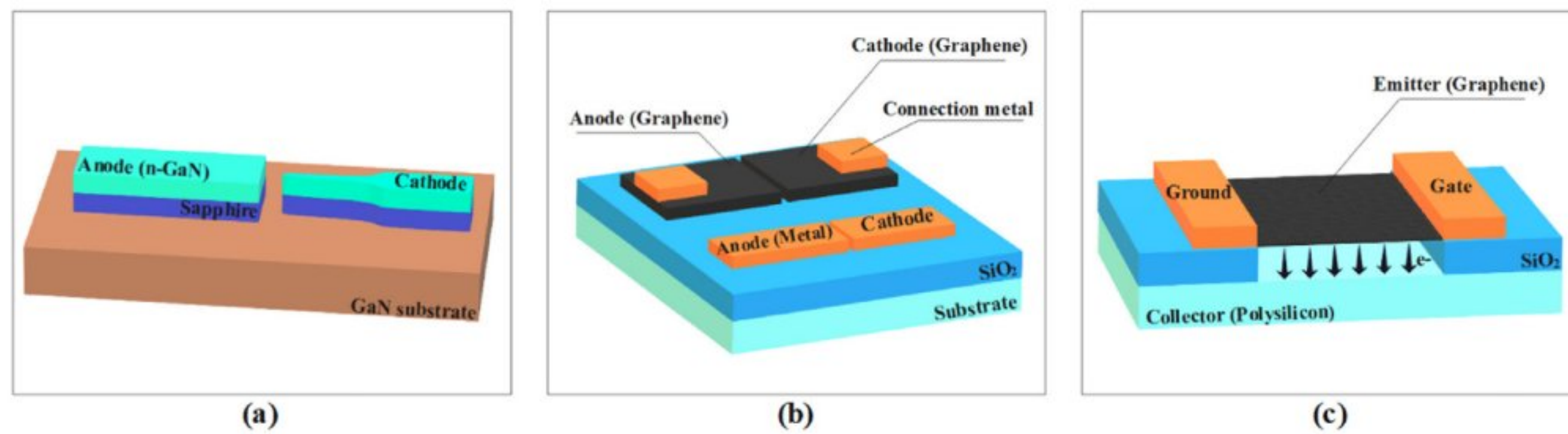


Fig. 5. The new electrode materials of the NACD. (a) N-type GaN layer/c-sapphire layer/C-doped GaN layer heteroepitaxial structure of GaN. (b) The horizontal Graphene and common metal electrode of the NVCT. (c) The vertical suspended Graphene electrode of the GVT.

5.3. Low-dimensional for electrode materials

Low-dimensional materials [69–71] such as graphene with high carrier mobility [72], compatible with existing silicon-based process technologies, and ultra-high specific surface area [73] are also the choice of the NACD electrode materials in recent years. Xu Ji team [74] proposed a 90 nm horizontal NVCT structure with graphene electrode as shown in Fig. 5 (b), which has large switching ratio (≥ 100) with low working voltages (< 20 V), they used a complete high graphene film transferred by double-layer PMMA and then a nanoscale graphene gap etched by electron beam etching and oxygen plasma. The influence of graphene gap notch on the mobility of graphene film, which also makes graphene as the electrode material of the horizontal structure NACD have higher requirements for fabrication process, while metal materials are better. It is worth noting that the thin graphene film electrode with vertical structure can reduce the device performance degradation difference caused by process fluctuation, and Gongtao Wu team [75] designed a graphene-based vacuum transistor (GVT) with vertical suspended graphene thin film cathode and N-type Si anode as shown in Fig. 5 (c). Phonon-assisted electron emission of graphene thin films [76] can increase emission current and switch ratio (10^6). Low-dimensional graphene is expected to achieve the low cost, high speed and high reliability NACD. However, Graphene lacks band gap, so that the switching ratio of the prepared devices is small even cannot be completely turned off [77,78]. The complex fabrication process and large fluctuations make it still full of challenges to achieve the NACD with graphene.

In summary, the Si and metal electrodes show the advantages of having mature process, low manufacturing cost and small fabrication difficulty, but the electrodes made of them are difficult to work reliably for a long time because of strong field evaporation. III-V group semiconductors have higher emission current, high frequency, high reliability and can adapt to extreme work environment, but need to be epitaxially grown on Si-based substrate to achieve compatibility with silicon-based process [79], epitaxial defects in fabrication will affect device performance. The fabrication process of horizontal graphene structure devices with low-dimensional materials is complex, the fabrication cost is high, and the actual electrical properties are poor.

5.4. Electrode morphology

There is a positive correlation between the enhancement factor/affinity of materials used in electrode and the current density of the NACD. It is worth mentioning that this negative correlation is based on only changing the electrode material. If the electrode morphology changes, the relationship between enhancement factor and current density becomes complicated. Enhancement factor is affected by electrode morphology (electrode length and electrode tip angle), the current density will change with the change of enhancement factor, thus changing the electrical performance of the NACD, as show in equations (1) and (2).

Ji Xu et al. [39] proposed that the sharper the tip electrode, the higher

the electric field intensity, the higher the current density and lower the turn-on voltage. However, it cannot be ignored that the sharp tip electrode is very likely to have strong field evaporation, which will lead to the change of emitter morphology in operation and deteriorate the working stability of devices. Therefore, the electrical properties and working stability of the electrode should be comprehensively considered in the electrode morphology. Shruti Nirantar [13] thinks that decreasing the angle of the tip and the distance between the two electrodes can increase the γ factor so that higher field emission current can be achieved at lower working voltage. In 2022 [25], the team proposed four electrode morphologies: PF (point-flat) FF (flat-flat) PP (point-point) PT (point-triangular). This is for the emission and collection electrodes of the NACD. The P-type electrode is a tip electrode, the F-type electrode is a flat electrode, the T-type electrode is a clamp electrode that complements the shape of the tip electrode, and the two letters represent the shape equipped by two electrodes of a NACD. For example, the PP type NACD has both emission and collection tip electrodes. PF type has the best electron emission performance, and PT structure has the least process fluctuation, which provides more basis for the selection of electrode geometry.

6. Simulation methodologies of the NACD

Device simulation take advantages of mathematical physics models to describe device structures, select electrical eigenvalues, and predict physical laws and emission mechanism. By utilizing various simulation platforms to explore the internal working mechanisms of semiconductor devices, we can deepen our understanding of the target devices and provide predictions for their actual working states in the later stages. To achieve the goal of reducing research and development costs, improving product quality, shortening research and development cycles, and seizing market opportunities. Different from traditional silicon-based devices, the NACD has unique electron emission mechanism and device structure. Using simulation tools to simulate the device reliability, particle trajectory, electrical performance and emission mechanism of the NACD helps us to understand the NACD and its working mechanism more deeply.

6.1. Electrical performance simulation

Sentaurus TCAD is mainly composed of Sde and Sprocess modules, Sde module is for device structure editing through code or visual tools, the electrode definition and grid division. The Sprocess module build functional layers according to the NACD fabrication process. Jungsik Kim [80] simulated the single-gate, double-gate surround gate of the NACD by Sentaurus TCAD, constructed the nonlocal mesh of nanogap and used discretize the FN equation to fit the emission process. Jungsik Kim found that the surround gate of the NACD shows high symmetry in structure and electron transport as shown in Fig. 6 (a) higher picture.

Silvaco TCAD simulator is mainly composed of Atlas and Athena. Atlas is similar to Sde module of Sentaurus TCAD. The difference is that

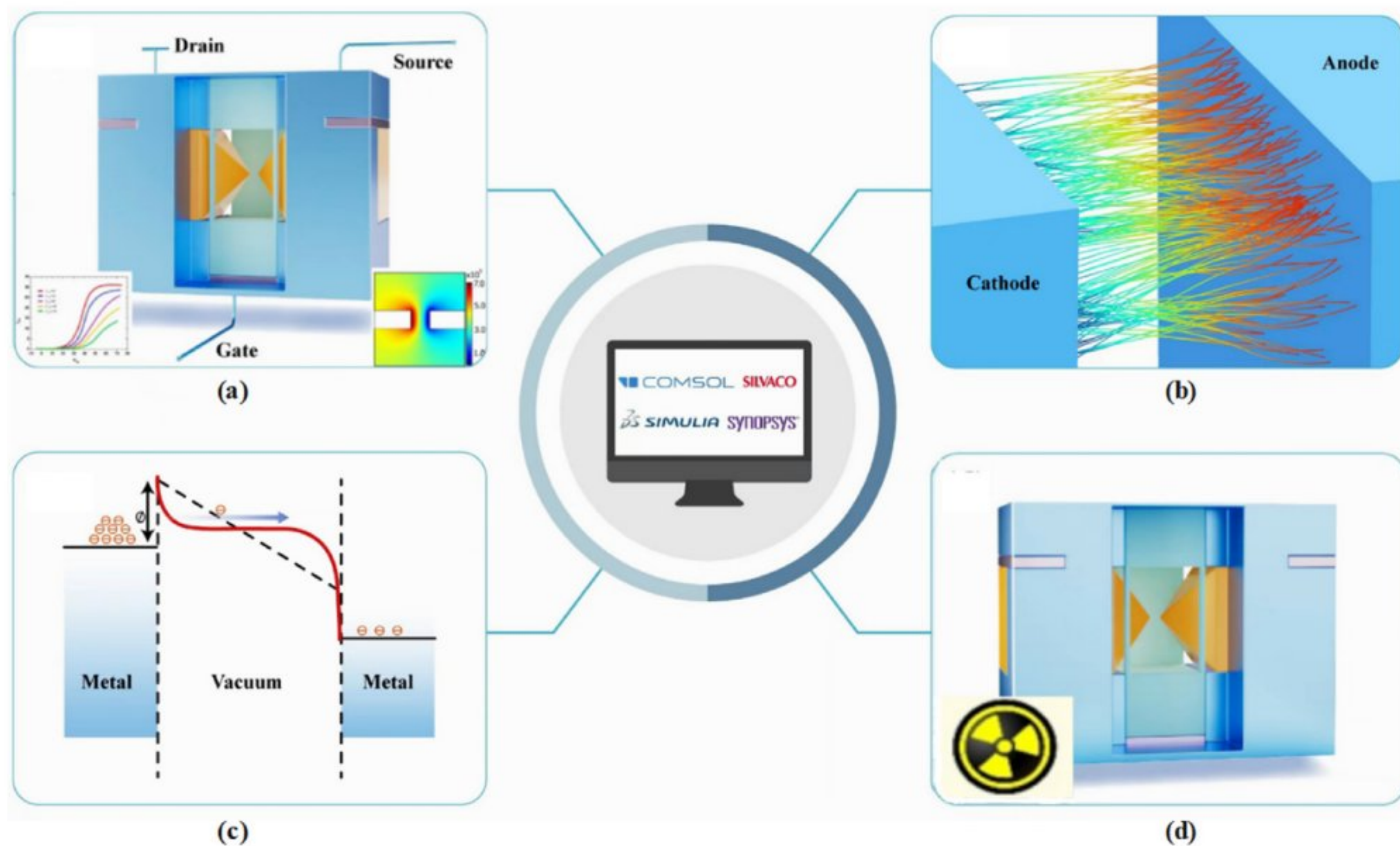


Fig. 6. The simulation methodologies of the NACD. (a) Extracting working state electrical parameters of the NACD, the lower pictures show I–V curve and electric field distribution of the NACD. (b) The NACD's electron trajectory tracking figure between anode and cathode, trajectory line color indicates electron velocity. (c) The F–N emission mechanism of the NACD, when the gate voltage increases beyond the threshold voltage, electrons could overcome the barrier. (d) Reliability irradiation simulation of the NACD.

Altas first meshes structure regions, then defines materials, electrodes and dopants for different regions, while Sde module call mesh engine at the end. Athena, like Sprocess module, builds the device structure according to the process logic. Ravariu and Cristian [81] evaluated the process of nothing on insulator (NOI) structure by Athena simulation tool of Silvaco TCAD and provide guidance for manufacture process.

After device structure modeling [80–82], variable voltage can be added to defined electrodes to detect relevant electrical parameters through mathematical models, Sentaurus Svisual module and Sentaurus Atlas module can monitor electrical parameters, DC characteristics, AC characteristics, transient characteristics, etc. The electric field distribution of the NACD plays important role, which should be deep insight.

Both CST studio suite and COMSOL Multiphysics can perform electromagnetic simulations of the NACD structures [83]. Specific flow can be summarized as geometric modeling, definition of vacuum degree, the NACD gridding, grid refinement for the key monitoring area of electric field, and definition of electric field solution model, such as electrode voltage application area and large and small electric field distribution function, etc. Hattacharya R et al. [84] simulated the electric field distribution of the CVFET by CST studio suite to verify that the influence of electrode topography on electron emission performance is shown in Fig. 6 (a) lower pictures. Siwapon Srisonphan [85] extracts the electric field between two electrodes by COMSOL Multiphysics.

Except for device simulation, Adina R. Bechhofer et al. [86]. established a lumped component circuit model for the development of resonant vacuum nanoscale antennas driven by light excitation exceeding 100 THz, and used LTspice to complete the circuit model simulation work.

6.2. Particle trajectory tracking simulation

COMSOL Multiphysics calls the particle tracking module to calculate discrete particle trajectories as an alternative to the finite element method to simulate real physical problems [87]. The NACD's electron trajectory tracking simulation process is roughly defined as the particle

trajectory as the cylindrical aperture according to the geometry of the electron emission channel. The space filling material is defined as the vacuum degree of the nanochannel. The nanogap is gridded, while the particle type is set to electron. The entrance and exit for electrons are defined as cathode and anode. The internal electrostatic field of the particle trajectory space is defined, while the space is defined as the state after the occurrence of electrons and apertures (bounce, adhesion, freezing, disappearance, and crossing). The setting of particle release time and particle initial velocity must comply with electron emission. Jin-Woo Han's team [88] used COMSOL Multiphysics to call the particle tracking module to simulate the trajectory of electrons in the nanogap in the vertical three-electrode NACD structure, as shown in Fig. 6 (b). Increasing the gate voltage can make the electron trajectory concentrate towards the central axis of the nanogap.

CST Studio Suite is a comprehensive solution for the global electromagnetic compatibility (EMC) simulation software market, providing complete time-domain and frequency-domain full-wave algorithms and high-frequency algorithms for designing, simulating and optimizing electromagnetic systems [89] CST studio suite provides Particle Tracking Solver to simulate the NACD electronic trajectory tracking. Hattacharya R et al. [84] studied the electron emission trajectory of the emitter electrode of complementary vacuum field emission transistor (CVFET) through Particle Tracking Solver in CST Studio, so as to explore the influence of electrode distance and electrode morphology on the electron emission trajectory.

6.3. Reliability simulation

Radiation has three main effects on devices: single event effects, total ionization dose (TID), and displacement damage [90,91]. High energy particles or heavy ions in CMOS pass through semiconductor materials and release sufficient induced charges in the depletion region, leading to device failure. TID damage mainly affects the insulator regions such as gate oxide and field oxide layers. The radiation induced charges in these insulators are often trapped at the interface or bulk region and

accumulate continuously, ultimately leading to device performance degradation. However, the NACD equip air channel in which there is no depletion zone. At the same time, the vacuum channel prevents any contact with the insulation material. The effect of total ionizing dose is negligible.

For TID simulation as shown in Fig. 6 (d), the Sentaurus TCAD provides the Radiation model [92], which users can directly call in the Sdevice module, and define dose rate and irradiation duration. Before the Radiation model takes effect, defects are added to some structures of the NACD to simulate the process of radiation-induced charge capture. For single event effect simulation, corresponding to the HeavyIon model [92], this model is directly called in Sdevice module, and the position, angle, time, energy, and path length of heavy ion incidence are defined. We can control these parameters to determine the sensitive area of heavy ion incidence in the NACD. After the simulation is completed, we can retrieve the trajectory map and probability distribution map of radiation induced charges after heavy ion incidence in NACD.

6.4. Electronic emission mechanism simulation

The electronic transport mechanisms of the NACD are complex and difficult to accurately grasp due to various factors such as electrode materials, temperature, operating voltage, and electrode morphology. The existing research on the electron transport mechanism of the NACD mostly uses the theoretical model introduced in the Section 1 to fit the I–V curve of the NACD at different operating voltage. Simulators can integrate material properties and construct the NACD models based on mathematical models to simulate the fitting process of emission mechanisms. Georgios Doundoulakis et al. [65] simulated vertical gallium nitride (GaN) nanowire (NW) vacuum field emission diodes (VFEDs) and GaN NW VFETs by Silvaco Atlas TCAD. The electrode-vacuum interface region is encrypted by grid to extract I–V characteristics and fit the F–N mechanism. Siwapon Srisonphan et al. [85] simulated the I–V curve of the space charge limited emission triode with nanogap channel to fit the space charge limiting current and F–N mechanism by the COMSOL Multiphysics.

In summary, we can use simulators to model the NACD and study the impact of variables on the NACD by controlling the variables in simulation results. At the same time, simulators can help us observe particle trajectories under ideal conditions and establish a deep understanding of the NACD electron emission process. Simulators provide fast and low-cost irradiation degradation research solutions for the NACD, providing guidance for real reliability testing.

However, due to the complexity of the electronic emission mechanism of the NACD, which is influenced by multiple variables such as temperature, electrode material properties, electrode morphology, nanoscale channel vacuum degree, and process fluctuations, the results of the simulators deviate significantly from the actual results, and there is no proposed NACD integrated simulation model in the industry. The simulation schemes of NACD are mostly limited to studying the influence of a single variable, and the simulators still face challenges in coupling variable effects of the NACD.

7. Applications

With the deepening of research on nanogap devices and the development of micro/nano processing technology, the dimension of nanogap devices is gradually shrinking, with higher integration level and lower packaging requirements. The potential for high performance and low power consumption is gradually being discovered, and more application fields are explored.

At present, in most common applications, the NACD still cannot achieve the performance of solid-state devices, but in some special applications as shown in the Fig. 7, they show excellent qualities.

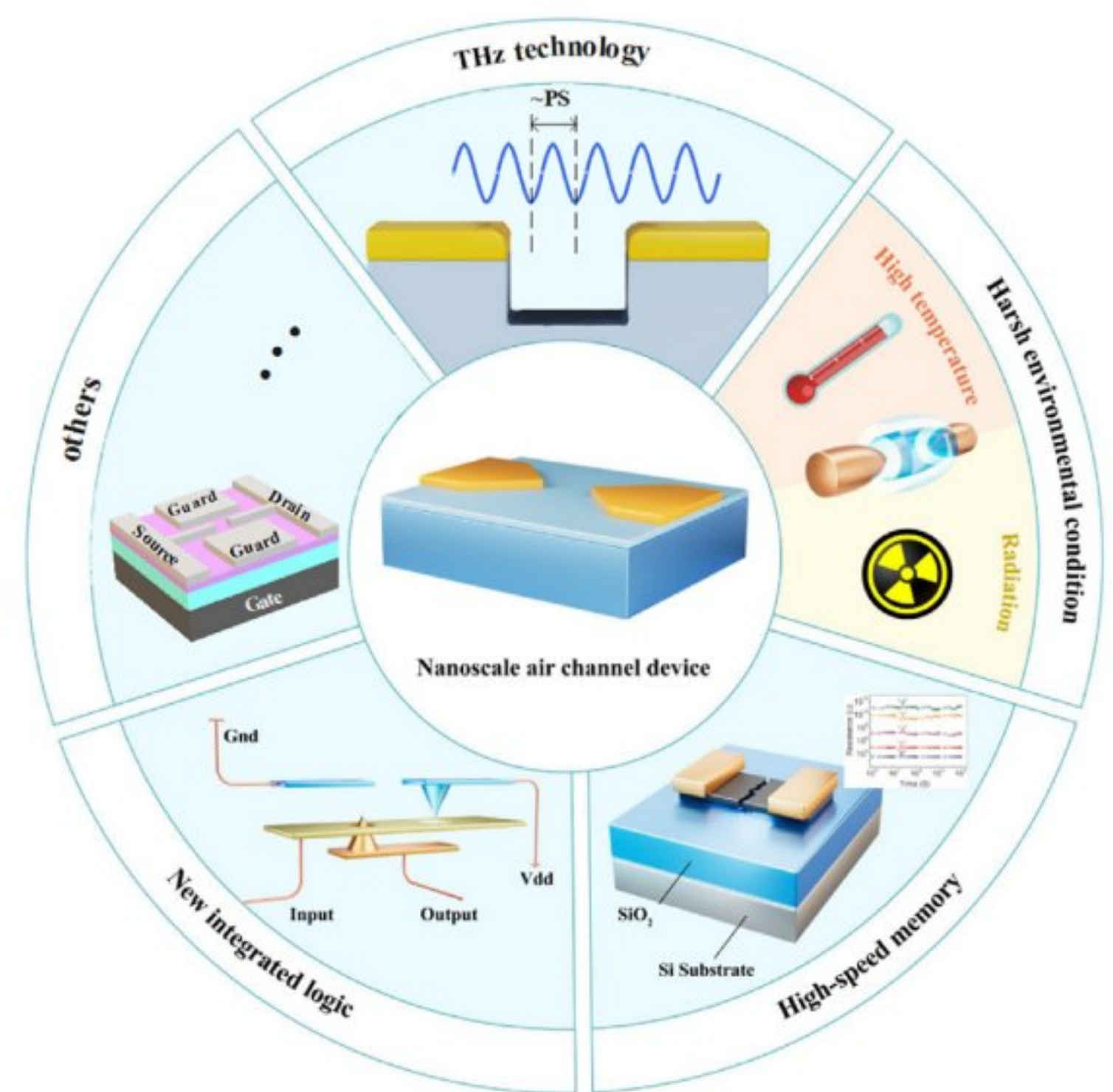


Fig. 7. The applications of the NACD in THz technology, new integrated logic, memory, harsh environmental condition and others.

- 1) Terahertz waves contain electromagnetic waves with frequencies ranging from 0.1 to 10 THz. As a new and unique radiation source, terahertz technology has broad application prospects in many fields such as communication, security, sensing, and national security. Its extensive engineering applications rely on a variety of terahertz devices. As the frequency of the device increases, the characteristic size of the device gradually decreases, and the short channel effect, quantum effect, heat generation and strong field effect of the device gradually become severe, seriously restricting the performance improvement of the device. The non scattering ballistic transport mechanism of the NACD is expected to achieve faster response speed than solid-state devices. Elison Matioli et al. [93] proposed a nano-device consisting of two closely spaced metal plates, with a distance as low as 20 nm between them. When a voltage is applied, electrons rush towards one of the plates, forming a nano plasma there. Once the voltage reaches a certain threshold, electrons almost immediately shoot towards the second plate. The rapid movement achieved through such rapid switching creates a signal that can generate extremely high power within picoseconds, thereby generating high-power terahertz waves.

However, the issues of high operating voltage (10–100 V) and weak rectification characteristics in nano plasma devices [93] cannot be ignored, which is one of the limitations in achieving the high-frequency NACD. There is no ultrafast sub-picosecond nanodevice with high rectification characteristics at low operating voltage matches modern electronic ICs in the industry. Nannan Li et al. [94] proposed a special fabrication process to control the nano air channel below 10 nm while obtaining the tip-to-edge electrode morphology. The tip electrode accumulates larger electric field, and electrons are easier to reach the edge electrode without scattering, resulting in a sub-picosecond response time and a lower turn-on voltage (0.7 V) for this nanostructure. It must be noted that due to limitations in sampling capability and bandwidth, the industry has not yet reported direct measurement of femtosecond response using time-domain electrical devices. Although it may be feasible to measure the femtosecond response using optical methods, this method is not suitable for directly measuring the response time of the NACD due to the significant difficulties in converting complex