COVER STORY

The Toughest

BY LESTER F. EASTMAN
Cornell University

& UMESH K. MISHRA
University of California
at Santa Barbara

Broadband wireless networks, ubiquitous hybrid electric vehicles, sophisticated controllers for electric grids, and compact, rugged radars: prognosticators have made plenty of promises in recent years. To that group, add gallium nitride transistors. All those technologies—and many more—are either much easier to envision, or could be hugely improved, with these devices.

The transistors withstand extreme heat and are capable of handling frequencies and power levels well beyond those possible with silicon, gallium arsenide, silicon carbide, or essentially any other semiconductor yet fabricated. And frequency and power-handling capabilities of this caliber could make all the difference in the amplifiers, modulators, and other key components of the advanced communications networks many are counting on to help revitalize the technology sector. The base stations of future wireless networks are a good illustration. The hope is that they will let people tap into high-speed streams of data, using their cell phones, personal digital assistants, or some other pocket console to capture video or high-quality sound.

Still, it's hard to imagine how cellular base-station amplifiers will deal with the data deluge implied by a world in which anyone can download full-motion video any time, anywhere. These amplifiers are already being pushed to their limits. They use a silicon technology that is only about 10 percent efficient, meaning that 90 percent of the power that goes into the transistors is wasted as heat. Powerful fans must continually blast this heat away from the amplifiers, which must also be outfitted with complex circuitry that corrects for the effects of harmonic and other distortions.

GaN transistors could double or triple the efficiency of base-station amplifiers, so that a given area could be covered by fewer base stations or, more likely, be flooded with more data at much higher rates. Freed from the powerful fans and correction circuitry, it might even be possible to shrink an entire base station to the size of a smallish dormitory refrigerator, something that would fit on a utility pole, rather than taking up expensive space in a telephone company central office.

Those same characteristics of speed, high-power handling, and heat resistance would also suit the transistors for countless other uses. Hybrid electric vehicles, for example, depend on circuits that convert direct current from their bat-

Transistor Yet

From broadband wireless to compact radars, countless future scenarios depend on the high power and high frequencies that only gallium nitride can deliver

teries to an alternating current capable of running an electric motor. GaN transistors would be ideal in these circuits.

Plenty in the picture

With so many potential uses, gallium nitride has become one of the most active and robust thrusts in semiconductor R&D. At least four U.S. companies, RF Micro Devices (Greensboro, N.C.), Cree Inc. (Durham, N.C.), Sensor Electronic Technology (Latham, N.Y.), and ATMI (Phoenix, Ariz.) already sell semiconductor wafers coated with gallium nitride that can be turned into transistors. (RF Micro Devices, a leading maker of power amplifiers for cellphone handsets, got into gallium nitride last October by buying RF Nitro Communications, a spinoff from Cornell University, for US $30 million.)

Other labs, including those at HRL (formerly Hughes Research Laboratories), Lucent Technologies' Bell Labs, TRW, and NEC and Sumitomo in Japan are making GaN-coated wafers for their own internal experiments and are fabricating transistors and amplifiers for testing. There are also sizable research efforts at many universities, notably Cornell, the University of California at Santa Barbara [see photo, p. 32], Rensselaer Polytechnic Institute, the University of Texas, the University of South Carolina, and North Carolina State University in the United States; Sweden's Chalmers University of Technology and Linkopings University; andJapan's Meijo University.

Nitronex, a start-up company (Raleigh, N.C.), has developed a method of coating relatively inexpensive silicon wafers with gallium nitride. Several other organizations, mostly small start-ups, are chasing the Holy Grail of large, perfect, so-called bulk crystals of the semiconductor, which could be used to make transistors of exceptional performance. These companies include Kyma Technologies (also in Raleigh), ATMI, Sumitomo in Japan, Astralux (Boulder, Colo.), Technologies and Devices International (Gaithersburg, Md.), and Crystal IS (Latham, N.Y.).

Not surprisingly, military electronics specialists are among the technology's most devoted backers. This past October the Defense Advanced Research Projects Agency in the United States announced it would award tens of millions of dollars in grants to developers of gallium nitride and related technologies. Many military or contracting labs, including ones at Raytheon, General Electric, Boeing, Rockwell, TRW, Northrop Grumman, and BAE Systems North America, are fabricating or testing GaN transistors.

Unique claim to fame

The military's interest isn't hard to fathom. Radar and satellite-communications links, which operate at frequencies ranging from hundreds of megahertz to tens of gigahertz, often have high power-amplification requirements, and would therefore benefit tremendously from gallium nitride. Many of the amplifiers in the transmitters of these radar systems and satellite-communications links still use traveling-wave tubes, a World War II–era technology and one of the few remainingbastions of vacuum-tube electronics.

The reason is that no semiconductor widely available today can cope with the frequencies and power levels involved. GaN transistors, however, would work in many of these units, conferring on them the solid-state advantages of ruggedness and portability. Potential benefactors exist outside the military, too: a DaimlerChrysler laboratory in Ulm, Germany, is investigating the use of GaN transistors in compact radar units for collision avoidance in automobiles.

Green, blue, and beyond

Gallium nitride has a colorful recent history. In the 1990s, Shuji Nakamura, then at Nichia Chemical Industries (Tokushima, Japan), used the semiconductor to fabricate the first green, blue, violet, and white light-emitting diodes (LEDs) and the first blue-light semiconductor laser. Working alone and with modest resources, Nakamura had succeeded where some of the world's top corporations, General Electric, NEC, Sony, 3M, and RCA among them, had failed.

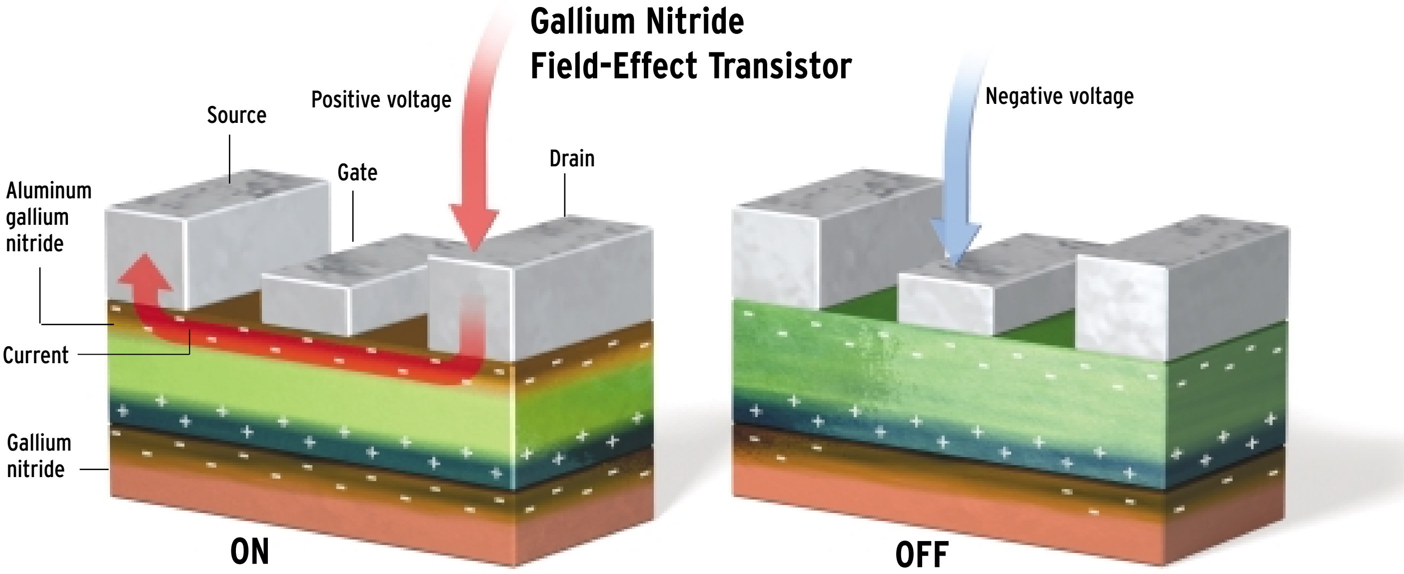
His achievement turned Nakamura, now a professor of electrical engineering at the University of California in Santa Barbara, into a minor folk hero and Nichia into the world's largest manufacturer of LEDs, with annual sales of several hundred million dollars. Today, nitride LEDs are so ubiquitous that you probably see them every day, in traffic lights, colorful video billboards, children's toys, even flashlights.

Nakamura's success did more than trigger a revolution in optoelectronics. It also helped direct attention—and funding—to work on high-power, high-frequency GaN transistors. Most of this development was done at U.S. universities and at a handful of research laboratories specializing in military technology. Much of the funding came from the Office of Naval Research and the Ballistic Missile Defense Organization.

GETTY IMAGES

A New and Improved Field-Effect Transistor

In a field-effect transistor, a voltage signal applied to a gate controls the flow of charge between a source and a drain. Small voltage changes on the gate are thus amplified into much larger changes in an external circuit connected to the source and drain. In the nitride FET, sharply polarized regions of charge exist naturally in the aluminum gallium nitride underneath the source, gate, and drain; in a sil-



The projects had some dazzling results. The power-handling capabilities of a transistor are usually specified as the number of watts the device can tolerate per unit width of its gate, the gate being the part of the transistor that controls the flow of current between its source and drain. For high power and frequency, a wide, short gate is desirable. A gate that is wide (in the direction perpendicular to electron flow) lets through lots of current and therefore high power; a gate that is short (in the direction of electron flow) decreases the transit time for electrons. Ultimately, the gate width is limited by the degradation of the signal from one side of the gate, where the signal is applied, to the other.

Last year our groups at Cornell University and the University of California at Santa Barbara both fabricated GaN transistors capable of sustaining power densities above 10 W/mm of gate width, while amplifying signals at 10 GHz—a few of the transistors exceeded 11 W/mm at that frequency. For comparison, ordinary silicon-based transistors can efficiently amplify signals only up to 2–3 GHz. As for silicon carbide, experimental devices at Cree recently achieved 7.2 W/mm, but at frequencies no higher than 3.5 GHz. Gallium arsenide transistors can handle 10 GHz but withstand a power density of less than 1 W/mm at that frequency. Silicon germanium devices can handle even higher frequencies, but, like gallium arsenide, cannot withstand high power.

Why not just connect many silicon devices in parallel to get higher power-handling capabilities? For one thing, the paralleled silicon devices still couldn't deliver both high power and high frequency. For another, the more transistors are connected together, the lower the circuit's impedance would be. Ultimately, when the impedance became low enough, the circuit would be useless because connecting it to any other electronics would require intervening circuitry that would sap whatever additional power could be coaxed out of the paralleled transistors.

Nor does heat bother gallium nitride much. In experiments at Astralux, researchers tested a GaN transistor at a sizzling ambient temperature of 300 °C and found that it amplified very well, with a gain of about 100. In contrast, silicon transistors stop working at about 140 °C. The result bodes well for the use of GaN electronics in engines, satellites, and other hot environments.

Flaw in the ointment

With so many advantages, why aren't GaN transistors already ubiquitous? Fundamentally, no inexpensive substrate material for gallium nitride exists. A substrate is a semiconductor's foundation, upon which engineers build devices or circuits in layers. Ideally, the substrate and the lower layers of the device are made of the same material. This uniformity prevents structural discontinuities, called mismatches, between the crystalline lattices of the substrate and the device above it. These lattice mismatches can cause damage, like microscopic fault lines, which intrude into the device and can seriously degrade its performance.

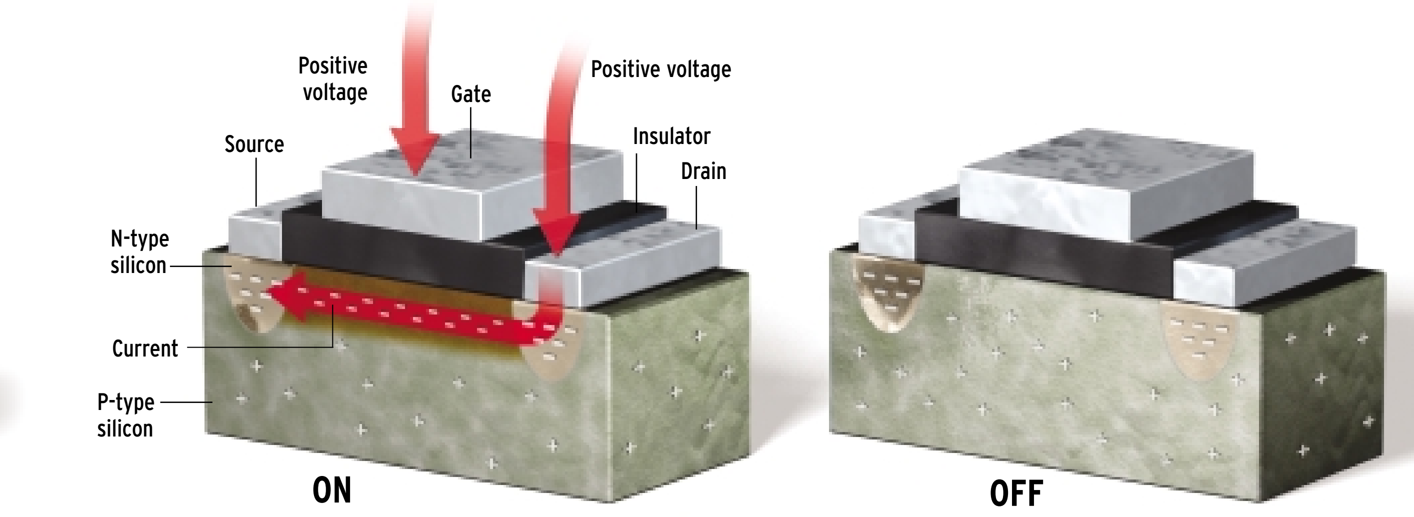
Substrate wafers are sliced from huge crystal cylinders called boules. Unfortunately, no one has yet figured out how to grow boules of gallium nitride. So most experimenters today fabricate GaN devices on substrates of sapphire or silicon carbide.

As it turns out, silicon carbide is a fairly good match for gallium nitride—the crystal lattices of the two compounds are mismatched by only 3.3 percent (the figure for sapphire and gallium nitride is 14.8 percent). Also, silicon carbide is an even better conductor of heat than gallium nitride.

BRYAN CHRISTIE

icon FET, these carriers must be evoked by a voltage applied to its gate. For both, charges flow in response to a positive voltage on the drain. In the nitride transistor, the layer of positive charges at the bottom of the aluminum gallium nitride gives rise to a complementary layer of mobile electrons in the gallium nitride. A negative voltage on the gate interferes with this layer, halting current flow.

Silicon Field-Effect Transistor



Naturally, all these advantages do not come cheaply. A 50-mm wafer of gallium nitride on silicon carbide, which might yield a couple thousand transistors, can cost in excess of $10 000, far more than prepared silicon wafers in much larger sizes.

Clear as crystal

Besides a high thermal conductivity about seven times that of gallium arsenide, gallium nitride has a high breakdown field—hence its ability to withstand high power levels—and excellent electron-transport properties, which enable it to operate at high frequencies.

The avalanche breakdown field of a semiconductor is the strength of the field, in megavolts per centimeter, that is needed to trigger what is known as avalanche breakdown. The condition occurs when the field is strong enough to liberate more and more electrons and electron deficiencies (called holes) from the atoms of the semiconductor crystal. Finally, when huge numbers of electrons and holes have been freed from the atoms, the current through the semiconductor surges, burning it out.

● Where Gallium Nitride Outstrips Other Semiconductor Materials

Semiconductor (commonly used compounds)		Silicon	Gallium arsenide (AlGaAs/InGaAs)	Indium phosphide (InAlAs/InGaAs)[a]	Silicon carbide	Gallium nitride (AlGaN/GaN)
Characteristic	**Unit**					
Bandgap	**eV**	1.1	1.42	1.35	3.26	3.49
Electron mobility at 300 K	**cm2/Vs**	1500	8500	5400	700	1000-2000
Saturated (peak) electron velocity	**$X10^7$ cm/s**	1.0 (1.0)	1.3 (2.1)	1.0 (2.3)	2.0 (2.0)	1.3 (2.1)
Critical breakdown field	**MV/cm**	0.3	0.4	0.5	3.0	3.0
Thermal conductivity	**W/cm•K**	1.5	0.5	0.7	4.5	>1.5
Relative dielectric constant	ε_τ	11.8	12.8	12.5	10.0	9.0

[a] The compounds are loosely known as indium-based.

A high breakdown field is desirable because it means that a semiconductor device made out of the material can withstand higher voltages over smaller dimensions. A field-effect transistor (FET), for example, could tolerate a higher voltage, and that higher voltage, along with higher current, would translate into greater power density for the device.

Gallium nitride has a breakdown field of about 3 MV/cm, as opposed to 0.4 MV/cm for gallium arsenide [see table, left]. Silicon carbide also has a breakdown field of about 3 MV/cm, but lacks gallium nitride's favorable electron-transport characteristics.

A high breakdown field, by the way, is an offshoot of wide bandgap. The bandgap of a semiconductor is a measure of the amount of energy it takes to move an electron from the so-called valence band, in which the electrons are not free to conduct electricity, to the conduction band, in which they are. In some semiconductors, an electron falling back from the higher-energy conduction band to the valence band causes the emission of a photon of light. The wider the bandgap, the higher the energy, and therefore frequency, of this emitted photon. This fact explains why gallium nitride, with one of the highest bandgaps of any semiconductor, can emit photons of green, blue, purple, and even ultraviolet light.

So wide is the bandgap of gallium nitride, in fact, that the material is transparent, much like diamond. Photons at all frequencies of the visible spectrum have energies that are less than the GaN bandgap, so they merely slip through without being absorbed by the crystal. That is why blue and green GaN LEDs are water-clear when they are off. Someday, when GaN devices and circuits are fabricated on GaN substrates, the resulting dice will also be perfectly clear.

Electron transport: what a gas

Gallium nitride's great boon is that it combines the high breakdown field of silicon carbide with the high-frequency characteristics of gallium arsenide, silicon-germanium, or indium phosphide. And while the high breakdown field is relatively straightforward, the electron-transport characteristics that permit high-frequency operation are more subtle.

How freely electrons move in a semiconductor typically depends on two factors, known as electron mobility and saturation velocity.

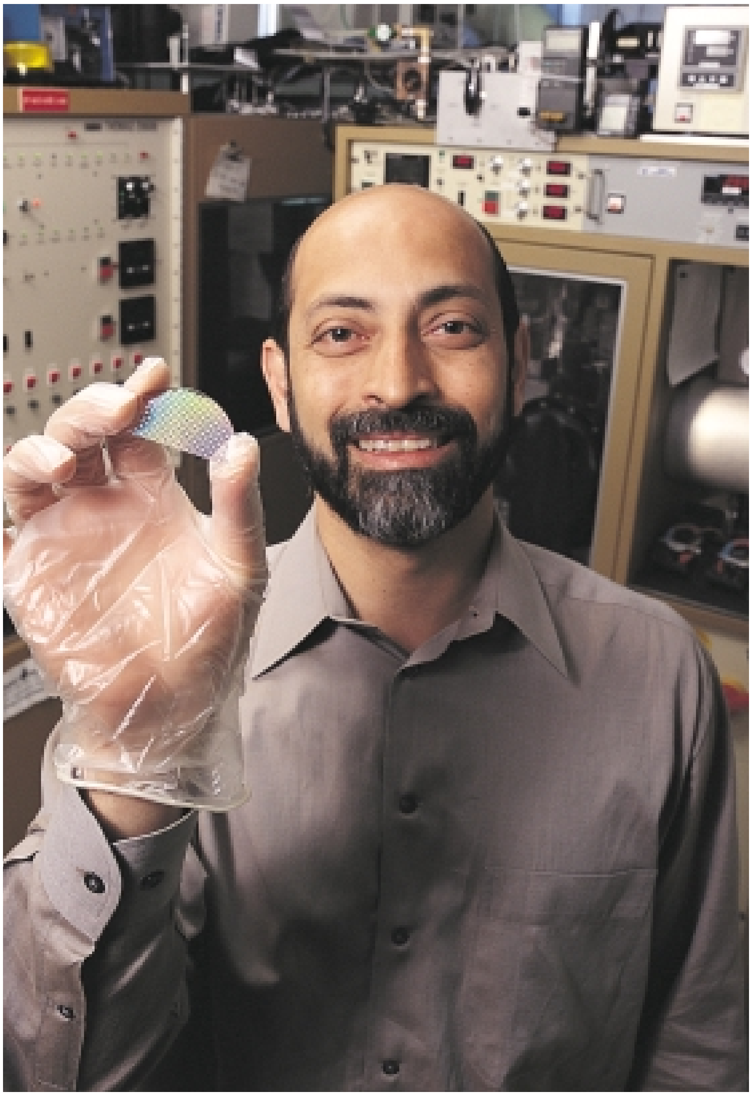
The mobility is determined by how fast electrons move in the material under the influence of relatively weak electric fields (for example, caused by an externally applied voltage). The saturation velocity, on the other hand, refers to the maximum speed the electrons are capable of reaching under the influence of a relatively strong field. The electron mobility in gallium nitride is lower than that in gallium arsenide, but the saturation velocity, at about 1.3×10^7 cm/s, is equivalent and should reach 2×10^7 cm/s in the next few years.

But in the unusual case of gallium nitride, those numbers do not begin to tell the whole story about electron movement. When aluminum gallium nitride is grown on top of a layer of a similar crystal, such as gallium nitride, a rather remarkable phenomenon occurs at the boundary, known as a heterojunction, between the two different crystals. This phenomenon is a major contributor to gallium nitride's outstanding high-frequency characteristics, so it is worth exploring in a bit more detail.

Inside the gallium nitride crystal, as in any similar crystal, the individual atoms are electronically charged, or ionized, and the large gallium and small nitrogen atoms are arranged somewhat irregularly with respect to each other, because of the difference in size. This combination of ionization and irregularity leads to a spontaneous electrical polarization within the crystal, or a separation of charge into countless, regularly spaced negative and positive atoms.

In an ordinary GaN crystal, the polarization does not accumulate, because the oppositely charged regions cancel out overall. But that canceling doesn't occur where the GaN crystal ends suddenly, for example, at a heterojunction with another crystal, such as aluminum gallium nitride. In that case, the abrupt change at the interface gives rise to an electrically charged region in the immediate vicinity of the boundary. That charged, polarized region is augmented, furthermore, by a piezoelectric polarization, which arises from the strain caused by the coming together of the two different crystal lattices.

These combined polarizations, in turn, induce an excess of free-moving electrons in the gallium nitride. The electrons



Author Umesh Mishra holds half a semiconductor wafer containing several hundred state-of-the-art gallium-nitride FETs. Mishra is in the laboratory that produced the devices, at the University of California at Santa Barbara.

MARCELO COELHO

concentrate near the polarization region, hard against the aluminum gallium nitride but without straying into it because the material's higher bandgap acts as a barrier.

Thus a dense two-dimensional "gas" of charge-carrying electrons spontaneously forms in the gallium nitride, very close to the boundary with the aluminum gallium nitride. Happily enough, this is basically the same configuration that all FETs exploit to amplify power.

FET accompli

The ordinary silicon FET has three terminals, called the source, gate, and drain [see illustrations, p. 30–31]. The gate is separated from the underlying substrate by a thin insulator. In the most common kind of device, a positive voltage applied to the gate creates a region under the insulator with many mobile electrons. This region, in turn, permits current between the source and the drain. The higher the gate voltage, the higher the current can be.

In the GaN FET, on the other hand, the two-dimensional electron gas already exists naturally. So a positive voltage applied to the drain immediately pushes current from source to drain. Thus the amount of current is varied by applying a negative voltage to the gate, which restricts the number of electrons available to flow from source to drain. A large enough negative voltage turns off the flow altogether. Thus in contrast to a silicon FET, which is normally off, a GaN FET is normally on.


TO PROBE FURTHER, SEE PAGE 79

By now, sharp-eyed readers may have spotted another unique feature of the GaN FET: it requires no "doping" with impurities. Gases of mobile charge-carriers can be created in most other semiconductors only by doping them with impurities to enable them to support an excess of either electrons or holes.

Doping of gallium nitride is needed, nevertheless, for the other major type of transistor, known as bipolar. This device requires two different kinds of materials, known as n-type (which has excess electrons) and p-type (excess holes). The bipolar transistor is the classic semiconductor device, with a history stretching back almost to the very beginning, in 1947 at Bell Telephone Laboratories. It would be great to have a bipolar GaN transistor because bipolar devices are exceptionally linear in their amplification across different frequencies. That characteristic would be very desirable for wireless base stations, for example, where that linearity would cut down on interference, or cross talk, among different communications channels.

One of us (Mishra) has succeeded in making bipolar GaN transistors. But they are not yet as reliable as the FETs because at the moment it is very difficult to make p-type material good enough to use in a bipolar transistor. Applying electrical contacts to the material, as is necessary to connect the device into a circuit, often wiped out the semiconductor's p-type character.

Nevertheless, Mishra's group at Santa Barbara did manage recently to make some experimental bipolar transistors that performed quite well. The transistors have current gain greater than 10, and a breakdown voltage of about 500 V. Silicon devices with comparable dimensions have a breakdown voltage of less than 100 V.

Gallium arsenide redux?

Some day, probably years from now, some bright researchers will figure out how to economically produce gallium nitride in bulk. That breakthrough will make it possible to build devices with truly astounding capabilities. Well before then, though, GaN transistors will begin carving out a niche of the market for high-performance transistors. Indeed, we foresee quite a race in coming years between the various organizations producing gallium nitride on substrates of silicon carbide, sapphire, and possibly even silicon.

Performance and cost will decide the winner. And there is plenty of room for improvement in both. The electron mobility of gallium nitride at 300 K is now approaching 2000 cm^2/V•s. However, it pales in comparison with gallium arsenide, at 8500 cm^2/V•s. Encouragingly, the mobility figures for gallium nitride are going up all the time, just as they did for gallium arsenide in the 1980s. The improvements follow from researchers' ability to hone the techniques of growing gallium nitride on substrates and improve the substrates themselves.

At the device level, the FET's weak link is the electrode that electrically connects its gate to the external circuit. It has a tendency to degrade, especially when the transistor is driven hard or subjected to high heat. One possible solution would be to electrically insulate the gate, as suggested by Asif Khan of the University of South Carolina, to prevent electrons from passing from it into the semiconductor below.

The price of GaN devices will come down, too. A big contributor to the high price of the best devices is the cost of silicon carbide wafers. A wafer measuring 50 mm across, the only size now available, currently goes for around $4000. But Cree is preparing to introduce 75-mm wafers. The larger wafers should help bring the costs down because they will make it possible for a much larger group of foundries to process the GaN-coated wafers with standard equipment and turn them into transistors.

None of this is to suggest that gallium nitride will be the next silicon. But it just might be the next gallium arsenide. Just as the growth of wireless telephony in the 1990s turned gallium arsenide into a multibillion-dollar industry, the ever-escalating demands placed on power devices by future generations of wireless and satellite communications, hybrid-electric cars, electric power grids, and military systems could propel gallium nitride through a succession of increasingly large niches.

If there is an elixir for the go-anywhere, do-anything future envisioned for broadband wireless, it is this continually surprising semiconductor, with a bandgap big enough to see through. ●

Glenn Zorpette, *Editor*

Acknowledgment: the authors wish to thank John C. Zolper of the Office of Naval Research for his support and suggestions.