VOLUME 32 ISSUE V 2026

COMPOUNDSEMICONDUCTOR.NET

PROPELLING THE PCSEL INTO THE DEEP-UV

AIXTRON

THE G10 SERIES

G10-GaN

1st fully automated compact GaN MOCVD cluster designed 100% for SI Power fabs

Your **Productivity Solution**
for All Advanced Epitaxy Materials

AIXTRON SE - Dornkaulstr. 2 • 52134 Herzogenrath • Germany - Phone: +49 2407 9030 0 - www.aixtron.com

VIEWPOINT

BY RICHARD STEVENSON
EDITOR

GaN's various vectors

For those that are looking to break new ground with the GaN-on-silicon HEMT, the most common target is to increase its blocking voltage. Success on this front will allow this transistor to compete with the SiC MOSFET and open up revenue streams in the electric vehicle industry and grid infrastructure.

As well as that promising pathway to increasing sales, there are other opportunities for this form of HEMT. One involves introducing products that block just a few tens of volts, and compete with silicon MOSFETs; and another is to improve the RF properties of this device by quashing its parasitic channel. Efforts on both fronts are detailed in this issue (see p. 14 and p. 16).

By going down in voltage, Infineon has expanded its power portfolio with three bi-directional switches that block 40 V. Following the launch of its first product in that range last summer, developed in close collaboration with a smartphone manufacturer, two more have been added this summer.

You might wonder why such a high blocking voltage is needed, given that mobiles run off batteries supplying just a handful of volts. But plug a smartphone into a fast charger and current flows at 20 V or 28 V – and designers also need to consider voltage spikes, due to unplugging and plugging cables. Based on these factors, 40 V provides the protection that's needed.

The silicon MOSFETs dominating this market are deployed in pairs to provide bi-directionality. As well as halving the component count, going to Infineon's GaN-based solution trims gate charge, leading to superior protection, and slashes the leakage current, helping the mobile's battery to retain its power.

One manufacturing challenge for GaN-on-silicon HEMTs is wafer bow, stemming from lattice mismatch. But as strain increases with epilayer thickness, and lower blocking voltages enable thinner epi, Infineon's 40 V switches could be one of the first products to roll off of the company's 300 mm line.

Operating with a markedly different business model that's based on technology licensing, Atomera is offering makers of GaN-on-silicon HEMTs an opportunity to address an unwanted parasitic channel sitting beneath their devices.

During growth of the buffer and stress-relief layers in RF III-N-on-silicon HEMTs, gallium and aluminium atoms descend into the substrate, where they act as unwanted dopants. This creates a parasitic channel, leading to losses and non-linearity.

Atomera's technology eradicates this channel. It involves pausing the growth of silicon, adding a puff of oxygen to the chamber, and then resuming deposition of silicon. Get this right and a stable silicon layer is introduced that's extremely effective at blocking diffusion, the cause of the parasitic channel.

While its early days, initial results are promising. As well as combatting substrate loss, Atomera's technology has been shown to deliver industry-leading levels of linearity.

Atomera and Infineon are by no means alone as innovators within our community. So, it will be fascinating to see what the GaN HEMT is capable of delivering throughout this decade and beyond.

Free weekly e-news roundup go to: compoundsemiconductor.net

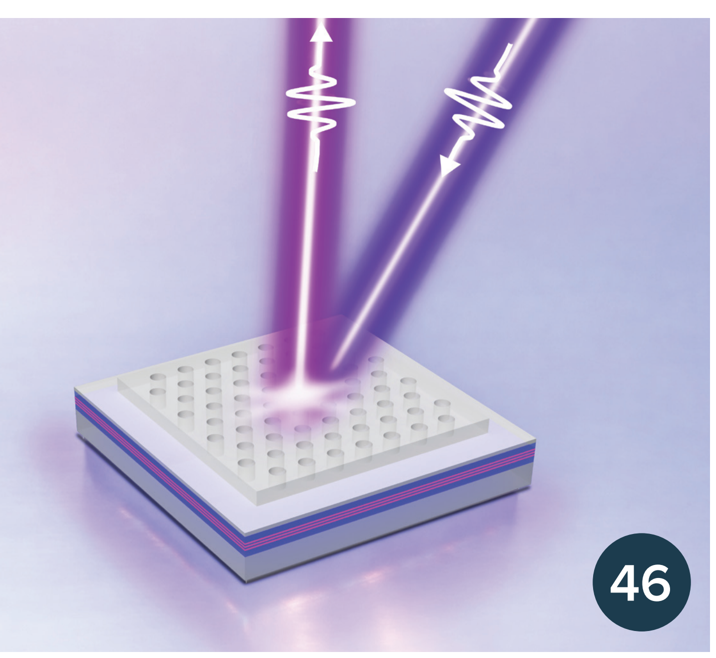
CONTENTS

VOL. 32 ISSUE V 2026

COVER STORY

Propelling the PCSEL into the deep-UV

With edge-emitters and VCSELs stalling, is the PCSEL emerging as the best candidate for delivering laser emission deep into the UV?



18 How to benchmark the UVC LED

A new metric from ams Osram aids the evaluation of UVC sources

22 Gallium oxide MOSFETs: Why interfaces matter

Developers of gallium oxide power devices should draw on lessons learnt from the pioneers of compound semiconductor MOSFETs

26 Improving the performance and production of the InP transistor

Leading producers of InP HEMTs and HBTs have increased the gain, efficiency and yield of their devices through process improvements and refined epitaxial structures

32 An elgant solution for SiC

With edge-emitters and VCSELs stalling, is the PCSEL emerging as the best candidate for delivering laser emission deep into the UV?

38 Monolithic microLEDs eye AI

InGaN nanopyramids, produced with a single epitaxial step, promise to provide red, green and blue emission from one device

42 AI and the future of epitaxy

The challenge of applying AI to semiconductor manufacturing is not technological, but structural – and ultimately human

32

NEWS ANALYSIS

14 Quashing parasitics in RF GaN-on-silicon HEMTs

Atomera's oxygen-based epitaxial technology is addressing problematic parasitic channels in GaN-on-silicon HEMTs

16 Taking GaN in a different direction

Infineon has just released a pair of bi-directional switches for mobile applications that block up to 40 V

RESEARCH REVIEW

51 Combatting compositional pulling in red InGaN microLEDs

52 Scrutinising 50 mm AlN substrates

53 Giving ß-Ga_2O_3 an insulating buffer

NEWS

06 Infineon-Innoscience patent battle continues

07 Power SiC enters the AI age, says Yole

08 Imec unlocks system-level III-V chiplet integration

09 UK CS Catapult to become 'Semiconductor' Catapult

10 Infineon launches EU flagship project Moore4Power

11 Integra announces single-device 10 kW GaN/SiC transistor for L-Band

12 Joint US-Taiwan initiative to scale the future of SiC

11

Editor Richard Stevenson richardstevenson@angelbc.com +44 (0)1291 629640
News Editor Christine Evans-Pughe christine.evans-pughe@angelbc.com
Design & Production Manager
Mitch Gaynor mitch.gaynor@angelbc.com +44 (0)1923 690214
Director of Logistics Sharon Cowley sharon.cowley@angelbc.com +44 (0)1923 690200
Senior Sales and Product Development Executive
Ranjodh Avern ranjodh.avern@angelbc.com +44 (0)2476 718970
Circulation Scott Adams scott.adams@angelbc.com
Publisher Jackie Cannon jackie.cannon@angelbc.com +44 (0)1923 690205

Sales and Product Manager James Cheriton james.cheriton@angelbc.com +44 (0)2476 718970
Chief Executive Officer Sukhi Bhadal sukhi.bhadal@angelbc.com +44 (0)2476 718970
Chief Technical Officer Scott Adams scott.adams@angelbc.com +44 (0)2476 718970
Directors Jackie Cannon, Sharon Cowley

Published by Angel Business Communications Ltd, 6 Bow Court, Fletchworth Gate, Burnsall Road, Coventry CV5 6SP, UK. **T:** +44 (0)2476 718970
E: info@angelbc.com
W: angelbc.com

Compound Semiconductor is published nine times a year on a controlled circulation basis. Non-qualifying individuals can subscribe at: £115.00 per annum (UK), €165 per annum (Europe), $198 per annum (air mail) (USA). Cover price £4.50. All information herein is believed to be correct at time of going to press. The publisher does not accept responsibility for any errors and omissions. The views expressed in this publication are not necessarily those of the publisher. Every effort has been made to obtain copyright permission for the material contained in this publication. Angel Business Communications Ltd will be happy to acknowledge any copyright oversights in a subsequent issue of the publication. Angel Business Communications Ltd © Copyright 2026. All rights reserved. Contents may not be reproduced in whole or part without the written consent of the publishers. The paper used within this magazine is produced by chain of custody certified manufacturers, guaranteeing sustainable sourcing. US mailing information: Compound Semiconductor, ISSN 1096-598X, USPS Permit Number 25366, is published 9 times a year, Combined Jan/Feb, March, Combined April/May, June, July, Combined August/September, October, November, December by Angel Business Communications Ltd, Unit 6, Bow Court, Fletchworth Gate, Burnsall Rd, Coventry CV5 6SP, UK. Airfreight and mailing in the USA by agent named World Container INC 150-15, 183rd St, Jamaica, NY 11413, USA. Periodicals Postage Paid at Brooklyn, NY 11256. POSTMASTER: Send address changes to Compound Semiconductor, Air Business Ltd, c/o World Container INC 150-15, 183rd St, Jamaica, NY 11413, USA. We strive for accuracy in all we publish; readers and contributors are encouraged to contact us if they recognise an error or omission. Once a magazine edition is published [online, in print or both], we do not update previously published articles to align old company names, branding, marketing efforts, taglines, mission statements or other promotional verbiage, images, or logos to newly created or updated names, images, typographic renderings, logos (or similar) when such references/images were accurately stated, rendered or displayed at the time of the original publication. When companies change their names or the images/text used to represent the company, we invite organizations to provide Angel Business Communications with a news release detailing their new business objectives and/or other changes that could impact how customers/prospects might recognise the company, contact the organisation, or engage with them for future commercial enterprise.
Printed by: The Manson Group. ISSN 1096-598X (Print) ISSN 2042-7328 (Online) © Copyright 2026

Infineon-Innoscience patent battle continues

Rulings by native courts go in the favour of Infineon and Innoscience

IN the long-running patents war between Infineon and Innoscience, courts in both countries have come down on the side of their producers of GaN products.

On 18 June Infineon announced that the District Court Munich, Germany, ruled in its favour in two further patent infringement cases – specifically one based on a patent, one based on a utility model – concerning GaN technology between Infineon and the Chinese company Innoscience.

Just prior to this announcement, Innoscience announced that China's Supreme People's Court upheld a sales injunction against certain GaN products manufactured by Infineon Technologies.

Based on the rulings in Munich, Innoscience is prohibited from manufacturing, selling, and marketing additional patent-infringing products in Germany, says Infineon. Furthermore, the court has ordered Innoscience to pay damages to Infineon.

Infineon says the case in Germany now marks Innoscience's third and fourth legal defeat in a series of court cases, each of which found that Innoscience's products infringe Infineon's patents.

Courts and authorities in both Germany and the United States have repeatedly concluded that Innoscience's products infringe Infineon's intellectual property rights.

Prior rulings include the decision against Innoscience from August 1, 2025, in an initial German proceeding. Furthermore, on May 7, the Full Commission of the US International Trade Commission (ITC) found that Innoscience had infringed an Infineon patent in the field of GaN technology.

Additional proceedings regarding the infringement of other Infineon patents are pending in the US and Germany.

"Today's ruling demonstrates the value of our GaN portfolio and underscores our commitment to vigorously defending our intellectual property and promoting fair competition," said Johannes Schoiswohl, SVP and head of Infineon's GaN Systems Business Line.

For the ruling in favour of Innoscience, in China, it follows a May 27, 2026, judgment by the Intermediate People's Court of Suzhou, which found that specified Infineon GaN products infringed an Innoscience patent covering GaN device technology.

The court ordered Infineon to cease the sale, offer for sale, and importation of the infringing products in China and awarded Innoscience damages of $1.4 million (RMB 10 million).

Infineon subsequently sought reconsideration from the Supreme People's Court. The court has now upheld the lower ruling in full, leaving the injunction unchanged and making the decision final and enforceable. With the Supreme Court's decision, the case has reached a definitive conclusion in China, according to the Innoscience press release.

Certain Infineon GaN products covered by the ruling are now prohibited from being sold, offered for sale, or imported into the Chinese market.

$50 million CHIPS Act funding for Coherent InP fab expansion

COHERENT has signed a letter of intent to receive up to $50 million in direct funding under the CHIPS and Science Act from the US Department of Commerce to expand its 6-inch InP semiconductor manufacturing facility in Sherman, Texas. The announcement coincided with a breaking ground ceremony with Nvidia at Coherent's Sherman facility along with federal and state officials, and local community leaders.

The fundiing, building upon $20 million provided through the Texas Semiconductor Innovation Fund and the Sherman Economic Development Corporation, will support growing demand for optical networking technologies that power AI data centres and strengthen Coherent's longstanding and recently expanded partnership with Nvidia.

The expansion project adds advanced wafer fabrication equipment and cleanroom capacity to double manufacturing production space and quadruple InP wafer production capacity. At project completion, the Sherman site is expected to create more than 1,000 jobs, including more than 550 direct advanced manufacturing, engineering, and technical roles.

"InP photonics are essential for enabling high-speed data transmission within AI systems, telecommunications, and advanced networks," said Bill Frauenhofer, executive director for semiconductor livestment and innovation at the US Department of Commerce.

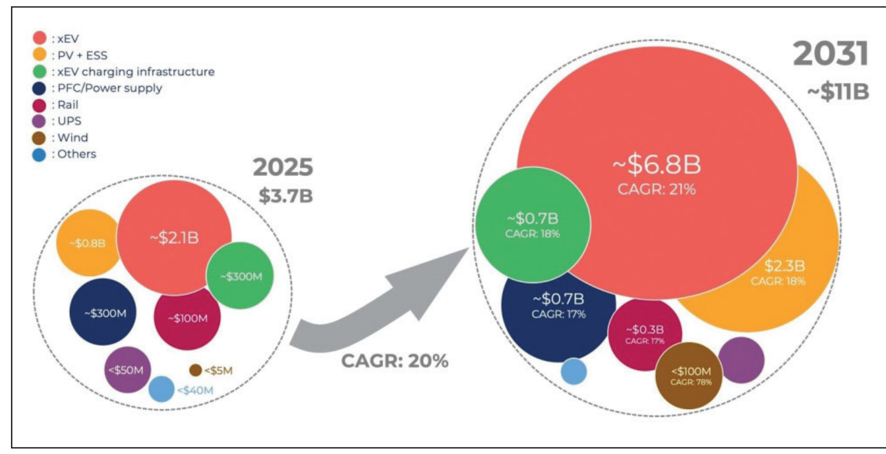
He added: "The CHIPS incentives will expand production capability, strengthen the US semiconductor supply chain, and accelerate the next generation of critical optical technologies."

Power SiC enters the AI age, says Yole

Power SiC market worth $11 billion by 2031, fuelled by AI data centres, 800 V EVs, and renewable energy systems

IN a new report 'Power SiC 2026 – Markets and Applications', Yole Group forecasts a power SiC device market of $11 billion by 2031, growing at a 20 percent compound annual growth rate (CAGR) between 2025-2031.

The growth is supported by strong demand from EVs, renewable energy systems, battery storage installations, EV charging infrastructure, and rail transportation, as well as the rapidly evolving AI data centre ecosystem.



Today, AI is driving an unprecedented demand for power in data centres. Meeting these requirements will require new power architectures, creating significant opportunities for power SiC technologies, says Yole. At the same time, electrification continues to transform mobility and energy systems. In this context, SiC plays the strategic role for power electronics. It enables higher efficiency, higher power density, and lower energy losses. As a result, SiC is becoming a key enabler of next-generation digital and energy infrastructure.

Despite a temporary slowdown in EV demand during 2024 and 2025, the automotive market is entering a new growth phase.

Yole's report highlights the increasing deployment of 800 V battery electric vehicles, which require significantly more SiC content than previous platforms. By 2031, 800 V platforms are expected to represent approximately half of global BEV shipments, making automotive the largest contributor to power SiC revenue over the entire forecast period.

SiC is one of the key enablers for the new generations of power supply, which represents the additional growth momentum across the entire SiC supply chain. With power requirements continuing to increase, the industry is working on improving the efficiency of the power supply unit (PSU).

The power level of PSU is going for 3 kW; the migration from 48/54 V rack distribution to three-phase sidecar HVDC architectures and, ultimately, to 800 VDC/±400, becomes a necessary step for VDC power architectures designed for AI factories. High-voltage DC architectures are considered as milestones to achieve, as well as solid-state transformers in the next generations. SiC is well positioned for some reasons besides the outstanding performance in high voltage and high efficiency.

The investment across the supply chain in the past years also secures the supply to deal with the surge demand.

Imec unlocks system-level III-V chiplet integration

Laser-assisted bonding supports assembly of III-V chiplets on passives-rich interposer stacks

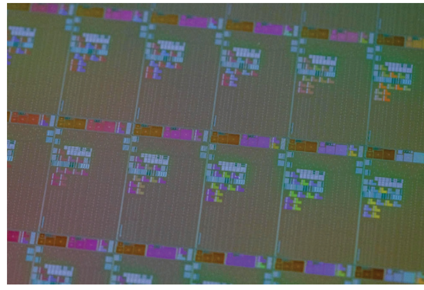
RESEARCH organisation Imec is evolving its 300 mm RF silicon interposer into a system-level platform for the heterogeneous integration of III-V chiplets on silicon CMOS.

By uniquely combining high-density embedded capacitors, a scalable modelling framework for passive components, and laser-assisted bonding for III-V chiplet assembly, the platform lays the foundation for next-generation wireless (millimetre-wave and sub-terahertz) systems, as well as RF-grade signal handling for ultrafast data centre applications.

As wireless systems move into millimetre-wave and sub-terahertz frequencies, and electronic and photonic interfaces in data centres are increasingly reaching their limits, it is becoming more difficult to deliver high-performance signal handling without driving up system integration complexity, cost, power consumption, and footprint.

A promising solution is to combine the superior gain, power, and efficiency of III-V materials – such as InP, GaAs, and GaN – with the scalability and cost efficiency of silicon CMOS technology. Chiplet-based heterogeneous integration on a high-performance RF silicon interposer makes this possible: it keeps performance-critical functions in compact III-V chiplets, while the interposer provides low-loss interconnects and hosts the remaining passive components.

Imec has been steadily advancing such a platform. In 2024, it demonstrated seamless InP chiplet integration on a 300 mm RF silicon interposer with negligible insertion loss at 140 GHz. In 2025, it extended the platform's record-low insertion loss up to 325 GHz. Now, Imec expands this platform with three new, complementary enablers: high-density embedded capacitors,



a scalable modelling framework for passive components, and laser-assisted bonding for III-V chiplet assembly.

"A key lever to reduce III-V chiplet size and cost is the offloading of passive components – such as decoupling capacitors – onto the RF silicon interposer," said Xiao Sun, principal member of technical staff at Imec.

"In a paper presented at this year's IMS/RFIC conference, we demonstrate how combining this offloading approach with a new MIMCAP architecture enables a 10-to-100-fold increase in capacitance density compared to typical on-chip capacitors in III-V technologies. This supports more compact and cost-efficient system designs and improves power delivery for millimetre-wave and sub-terahertz wireless systems as well as high-speed data centre applications."

Imec's new MIM capacitor (MIMCAP) architecture (pictured above) combines a high-κ aluminum-hafnium-oxide dielectric with 3D oxide-stud structures in the back-end-of-line (BEOL).

Imec's new MIM capacitor (MIMCAP) architecture (pictured left) combines a high-κ dielectric with 3D oxide-stud structures in the back-end-of-line (BEOL). Complementing this effort, Imec recently presented a modelling framework for RF interposer passives, validated up to the sub-terahertz regime (about 300 GHz). Imec's model enables designers to accurately predict circuit performance as geometries change, without needing to re-simulate or measure every variation, significantly reducing development time.

To date, Imec's framework has focused on transmission line performance – but lays the foundation for a comprehensive design library that is being extended to other passive components, including inductors and MIMCAPs.

Finally, Imec demonstrated the use of laser-assisted bonding to integrate III-V chiplets onto its RF silicon interposer, enabling assembly of chiplets on a complex, passives-rich stack without compromising thermal budgets, or damaging temperature-sensitive interposer layers.

Imec's approach achieves alignment accuracy below 600 nm, and rotational misalignment below 0.05° across 43 devices. RF measurements confirm preserved performance after assembly, with reflection below −15 dB in the 110-170 GHz range, demonstrating a viable path toward fully assembled high-frequency chiplet-based systems.

Xiao Sun: "With this work, we demonstrate a uniquely integrated platform that brings together performance, scalability, and manufacturability. Our next priority is to further advance the platform's technology readiness, and to enable support for low-volume manufacturing – helping our partners more easily develop and scale next-generation RF systems."

UK CS Catapult to become 'Semiconductor' Catapult

UK's authority on compound semiconductor applications and commercialisation to transition to a more general role

FOLLOWING the publication of the UK Government's AI Hardware Plan, the Compound Semiconductor Applications (CSA) Catapult will transition to become the 'Semiconductor' Catapult. The change of role will be to address a critical gap in UK AI infrastructure.

The Compound Semiconductor Applications (CSA) Catapult was established in 2018 by Innovate UK to help the UK become a global leader in compound semiconductors. CSA Catapult is currently the UK's authority on compound semiconductor applications and commercialisation.

Caroline O'Brien, CEO of CSA Catapult, said: "Improving the performance and energy efficiency of AI data centres remains a significant global and national challenge. UK semiconductor companies have world-leading products and services that can be used in the hardware needed for AI data centres.

The Semiconductor Catapult working with industry will deliver breakthroughs in R&D in energy efficient deployable systems, to reduce risk and ensure even more start-ups can leverage new supply chains anchored in the UK."

She added: "By refocusing, we will broaden our support for UK companies in design, development and commercialisation across the wider UK semiconductor sector. As the Semiconductor Catapult, we will work to increase the adoption of UK-designed semiconductor technologies in AI infrastructure, including data centres across the UK. We look forward to continuing our work with the Department for Science, Innovation and Technology (DSIT) and Innovate UK to support delivery of the UK's AI Hardware Plan."

UK AI Minister Kanishka Narayan said:"Our AI Hardware Plan is about backing British firms at every stage, from early research through to deployment. AI hardware is one of the fastest growing and most strategically important sectors in the global economy, and the UK is well placed to lead.

"One of the biggest challenges for UK chip companies is getting from a brilliant design to a product that is ready for market. The new Semiconductor Catapult will help close that gap, supporting British firms to get their technology out of the lab and into the systems driving the AI revolution."

IQE and Tower sign multi-year InP epiwafer agreement

COMPOUND SEMICONDUCTOR wafer company IQE and Tower Semiconductor have announced a multi-year agreement for the supply of InP epiwafers for optical connectivity solutions serving AI-driven data centre infrastructure.

IQE's InP epiwafers will be used in several of Tower's advanced silicon photonics platforms for next-generation optical technologies, providing a high-quality supply base for Tower's product roadmap.

IQE and Tower's collaboration includes technology for the production of 200 Gbps/lane for pluggable transceivers and the prototyping of next-generation 400 Gbps/lane modulators, as well as other critical applications including optical-circuit-switches for deployment in data centres. The agreement provides for a minimum purchase commitment by Tower in the first year, a reciprocal supply commitment from IQE, and minimum volume commitments thereafter.

Under a separate agreement, Tower will also provide a broad worldwide and royalty-free license to IQE for porous silicon patents, which have been the subject of an IP dispute between the companies, settling all litigation in the matter.

Jutta Meier, CEO of IQE, commented: "This agreement reinforces IQE's position within Tier 1 global hyperscale cloud and AI infrastructure markets. With decades of InP epitaxy expertise and established high-volume manufacturing capability, IQE is primed to support next-generation optical connectivity applications as they scale from innovation to commercial deployment."

Marco Racanelli, president of Tower Semiconductor, said that the combination of InP and in-house silicon will enable high-volume products that deliver the performance required to scale future AI infrastructure capacity.

Infineon launches EU flagship project Moore4Power

Heterogeneous integration project to drive next generation of sustainable power electronics

COORDINATED by Infineon, 62 European partners have joined forces to develop smart power electronics in the Moore4Power project.

They will pioneer a 'More than Moore' approach combining heterogeneous and functional integration involving different semiconductor technologies such as silicon, SiC and GaN, together with sensing, control and communication functions. The idea is that each technology is used where it performs best, enabling higher efficiency, improved reliability and more compact designs.

Power chiplet technology enables scalable architectures and more flexible product variants at competitive cost levels. This modular approach will pave the way for next-generation solutions in a wide range of highly relevant applications. The foundation was laid in the predecessor project PowerizeD, a large Chips JU-funded project completed in 2025, which delivered advances in efficiency and reliability.

"Power electronics are a decisive enabler for energy efficiency and sustainability. With Moore4Power, we are setting the next level of smart integration to achieve significantly higher energy and resource efficiency," says Jochen Koszescha, coordination lead for the Moore4Power project at Infineon Technologies AG. "We are proud to join forces with an outstanding consortium from academia, research and industry to make a decisive contribution to Europe's Clean Industrial Deal."

Moore4Power focuses on high-impact sectors where power conversion drives cost, CO_2 reduction and reliability. In wind energy, advanced power

electronics can operate at the core of turbines to improve energy conversion and increase harvested power; in e-mobility, advanced power electronics will enable up to 99 percent efficiency with near-loss-free, bidirectional charging; and in railway systems, they will cut propulsion losses by at least 30 percent, significantly boosting energy efficiency.

Moore4Power also aims to reinvents the development process itself. AI-assisted models, digital twins and automated workflows will be used to shorten development cycles. Hardware and software will be designed in parallel to reduce simulation time while increasing accuracy. As a result, the aim is to cut time from first fab samples to a validated datasheet release to just one week, compared to several weeks today.

A Digital Product Passport (DPP) will be embedded directly into power modules via wireless access. The DPP will provide lifecycle data, such as operating conditions, state of health and remaining lifetime throughout the product's use.

This transparency will enable smarter maintenance, longer product lifetimes, better reuse and reduced raw-material consumption – delivering substantial CO_2 savings and a concrete contribution to Europe's circular economy and climate goals.

EC approves €66 million for German SiC facility

The European Commission has approved €288 million in German State aid to support the setting up of two new facilities in the semiconductor supply chain.

Part of the aid is €66 million for Zadient Materials Europe GmbH to set up a facility for the manufacturing of SiC source materials in Bitterfeld, Saxony-Anhalt.

Zadient is a venture-backed start-up that has developed an industrial-scale process for efficiently producing tens of thousands of tons per year of granular SiC source material at purities reaching 8N (99.999999 percent). It is using the material to develop new technologies targetting high yield, low defect SiC wafers.

Zadient's 'SiC-Pro' investment project involves a circular system in which process gases are recovered and reintroduced in the production cycle. This is a novel process in the EU and is expected to support high material quality, energy efficiency and long-term cost-effectiveness. The aid will be provided in the form of a direct grant of €66 million.

The other €222 million will be for Carl Zeiss's 'HNA@SCALE' project, which will introduce and industrialise extreme ultraviolet ('EUV') optical columns required for the next generation of EUV lithography machines produced by the Dutch company ASML

Integra announces single-device 10 kW GaN/SiC transistor for L-Band

Company claims significant advance for RF power amplifier system designers

INTEGRA TECHNOLOGIES, a US RF and microwave semiconductor company, has announced the industry's first single high-voltage GaN/SiC 150 V transistor, the IGN1030S10000, that produces 10 kW of pulsed saturated output power at L-Band.

The company says achieving this level of power from a single transistor represents a significant advance for RF power amplifier system designers.

Traditional high-power architectures at L-band typically require power combining from multiple lower-power devices, increasing circuit complexity, insertion loss, thermal management demands and overall system cost. By delivering 10 kW from a single 150 V HV GaN/SiC transistor, the IGN1030S10000 enables substantially simpler amplifier system architectures while reducing the size, cost, and complexity of high-power transmitter architectures. This part simultaneously enables smaller system volumes and increased system power density.

"As the pioneer of HV GaN/SiC Technology, Integra has spent years pushing the boundaries of what's possible for solid-state RF power — and achieving 10 kW of pulsed output power at L-Band from a single transistor marks our most defining milestone yet" said Integra's president and CEO Suja Ramnath.

She added: "Prior to Integra's HV GaN/SiC, the practical solid-state replacement of Vacuum Electron Devices (VEDs) was limited to 10 kW. Now, Integra has extended practical, solid-state replacement of VEDs to the megawatt level in high-power architectures — opening a new chapter for system architects long constrained by the limits of existing solutions. This is not incremental progress. This is a fundamental shift in what is possible, and Integra intends to keep leading it."

The IGN1030S10000 is targeted at L-Band pulsed applications where peak power, efficiency and ruggedness are critical. Integra's 150 V HV GaN/SiC technology can be leveraged at other frequencies for high-power, high-performance applications.

Joint US-Taiwan initiative to scale the future of SiC

Purdue/GCCS initiative targets critical thermal and power bottlenecks in global AI infrastructure

PURDUE UNIVERSITY in the US has signed a five year partnership with Taiwanese company GeChi Compound Semiconductor (GCCS), a specialist in SiC crystal growth, to accelerate the commercialisation of SiC.

The collaboration targets the critical thermal, power and 6G bottlenecks currently constraining the next generation of high-compute infrastructure. Joint research will focus on isolating crystal defects and optimising SiC material growth to accelerate the transition to high-yield 8-inch and 12-inch wafer platforms.

GCCS will serve as a provider of semiconductor materials and Purdue as a hub for the technology. It's the first MoU between Purdue and Taiwan-based GCCS.

"With this agreement, Purdue and GCCS are leveraging research strengths at the academic and industry levels," said Dan DeLaurentis, executive vice president for research. "It continues Purdue's position as a leader in the latest semiconductor research."

"This partnership represents a profound strategic alignment between GeChi Compound Semiconductor and Purdue University," commented Kuan-Ming Hsiung, board chairman.

He added: "By combining our manufacturing scale with America's leading academic institution, we are taking decisive action to secure the domestic supply chain for SiC."

"This collaboration is not merely about advancing materials; it is about establishing the resilient, high-yield manufacturing capacity within the United States that is absolutely essential for national tech security and the future of global critical infrastructure."

Wolfspeed unveils lowest $R_{DS(ON)}$ SiC MOSFETs

SiC specialist Wolfspeed has introduced its fifth SiC technology generation, demonstrating a substantial performance increase in efficiency for next-generation 1200 V and 750 V automotive and industrial applications.

"With Gen 4, Wolfspeed delivered the switching breakthrough our customers needed, and less than two years later we're introducing Gen 5 that gives engineers the most current possible with a 5 by 5 mm SiC footprint," said Cengiz Balkas, Wolfspeed chief business Officer. "What excites me most isn't just the pace of our innovation; it's what this technology unlocks for our customers: an accelerated path to smarter, more efficient, compact systems made for real-world conditions."

According to Wolfspeed, Gen 5 products enable system architects to design more compact traction inverters and improve mileage per charge and right-size costly EV batteries. They also unlock new SiC opportunities by replacing mechanical relays with solid-state circuit breakers and set new efficiency standards for EV charging infrastructure. The benefits also extend beyond automotive to applications like industrial power supplies.

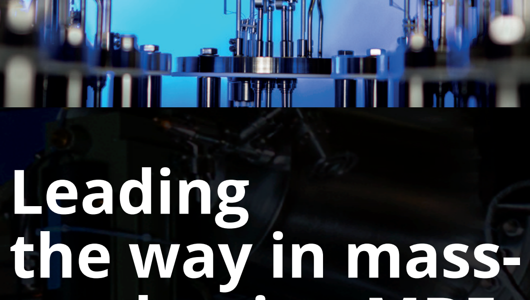
According to the company. Gen 5-based systems can achieve the highest current possible at high temperatures when compared to competitive 5 mm x 5 mm footprint SiC MOSFETs. And its optimiation of $R_{DS(ON)}$ resolves two compelling design challenges. First, it significantly improves system-level conduction losses via an up to 27 percent reduction in specific on-resistance (R_{SP}) over today's commercially available competitive 1200 V solutions. The 1200 V QEM50120-25D10 achieves a 175°C chip-level R_{SP} of 3.4 mΩ cm^2, and the 750 V QEM50075-025D10 achieves a 175°C chip-level R_{SP} of 2.0 mΩ cm^2. Second, it reduces the need for system-level design margin with ultra-low +/- 18 percent $R_{DS(ON)}$ distribution for both voltage nodes.

Wolfspeed Gen 5 includes the same body diode introduced with the Gen 4 technology platform but has an improved junction temperature of 200°C continuous (215°C limited life).

This is the second Wolfspeed MOSFET technology generation to be designed, manufactured, and qualified within Wolfspeed's ramp-ready 200 mm device fabrication facility in Mohawk Valley, NY. New product introductions, sampling, and customer validation will be completed using 200 mm production material. Furthermore, no new manufacturing toolsets are required for volume production.

"Our planar MOSFET technology still has innovation runway. We established Gen 5 on tools and processes our customers are familiar with to create a low-risk upgrade path for next-generation programmes," said Adam Barkley, VP of power device and package development. "For customers facing compressed development timelines, that means faster validation, faster qualification, and faster time to market — without sacrificing the performance they know and trust."

RIBER's MBE 6000



Leading the way in mass-production MBE

LOW RESIDUAL BACKGROUND
LOW ENERGY DEPOSITION
LOW TEMPERATURE
LOW DEFECTS DENSITY

CRYSTALLINE PERFECTION
HIGH UNIFORMITY
CONTROL AT MONOLAYER SCALE
HIGH REPRODUCIBILITY

RIBER

INNOVATIVE SOLUTIONS FOR SEMICONDUCTOR INDUSTRY

info@riber.com
www.riber.com

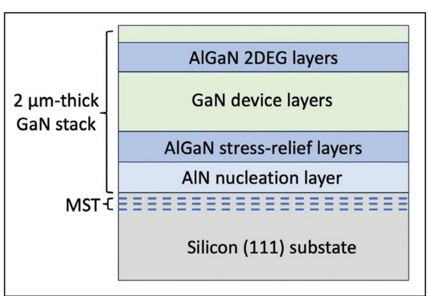
Quashing parasitics in RF GaN-on-silicon HEMTs

Atomera's oxygen-based epitaxial technology is addressing problematic parasitic channels in GaN-on-silicon HEMTs

BY RICHARD STEVENSON, EDITOR, CS MAGAZINE

FACING a dilemma are producers of GaN RF HEMTs, a class of transistor that's widely deployed for mobile communication and defence infrastructure. If performance tops their agenda, they will produce their devices on SiC substrates, which have a high thermal conductivity that aids performance, but adds to cost. And if price is their top priority, they will select a silicon foundation, as this drives down production costs – but at the expense of a parasitic channel that introduces RF losses and drags down linearity.

Fortunately, it is now possible to address parasitic limitations and enjoy lower costs by adopting an approach pioneered by technology-licensing company Atomera. This US outfit is expanding its expertise developed in the silicon industry by demonstrating that its epitaxial technology can incorporate a much-valued oxygen-rich region into the silicon substrate beneath the III-N RF heterostructure – this addition reduces the highly detrimental parasitic charge by more than an order of magnitude.



➤ With Mears Silicon Technology (MST), aluminium and gallium atoms, which act as dopants, are prevented from descending into the silicon substrate. This prevents a parasitic channel from forming.

A parasitic layer plagues GaN-on-silicon HEMTs, because during the growth of the AlN nucleation layer and the AlGaN stress-relief layers, aluminium and gallium atoms, which both act as dopants, descend into the silicon substrate. This creates a layer of charge at the substrate-epilayer interface that's responsible for the parasitic channel, even when III-N heterostructures are grown on high-resistivity silicon.

Atomera's Mears Silicon Technology (MST), which addresses this weakness, has already been shown to enhance a broad swathe of silicon devices, from power switches to leading-edge gate-all-around transistors.

Company's co-founder and CTO, Robert Mears, explains that at the heart of Atomera's trailblazing tech is a pause in silicon epitaxy, when a puff of oxygen is introduced into the growth chamber. After that, the growth of the silicon stack resumes.

"If you get the oxygen right, it's a very stable layer," argues Mears, who adds: "It's not big clusters of silicon dioxide." MST retains the silicon planes and creates a structure that's extremely effective at blocking diffusion. Ultimately, this prevents a parasitic channel from forming.

The degree of effectiveness of MST is highlighted in measurements of epitaxial structures for RF GaN-based HEMTs, grown by Edwin Piner's team at Texas State University.

Spreading resistance probe measurements, which offer a good insight into the local conductivity that results from the parasitic charge, show that MST leads to a reduction in interfacial charge by more than an order of magnitude, compared with the conventional heterostructure. The origin of this reduction is quantified by secondary ion mass spectrometry, which shows a substantial reduction in the penetration of gallium atoms into the substrate.

Mears points out that the blocking of diffusion does not need to be perfect. Once it allows resistivity to exceed around 10 kΩ cm, the substrate is effectively lossless, and other characteristics are more worthy of consideration.

Expounding on that point, he remarks: "What really matters to all RF designers is the linearity. It's how much you can suppress the second and third harmonics for the incoming RF signal. That's where this new data is exceptional."

According to Mears, GaN-on-silicon has traditionally fallen far short of silicon-on-insulator (SOI) technologies for supressing the second harmonic. One leader of SOI technologies, Soitec, has products eSI80, eSI90, and eSI100 – named to reflect values of -80 dBm, -90 dBm and -100 dBm for the second harmonic power level at the standard input power of 15 dBm. For essentially a first iteration of a GaN RF structure with

MST technology, the comparable value is -97 dBm, according to measurements made by Incize, a Belgium-based company that provides characterisation services for the RF industry.

Today, many RF applications involve input powers of more than 15 dBm (that is just over 30 milliwatts). And in this regime, RF GaN structures with MST technology are also performing well, says Mears: “The linearity up to input powers of up to 40 dBm – 10 watts – is extremely good. That’s unique for GaN-on-silicon.”

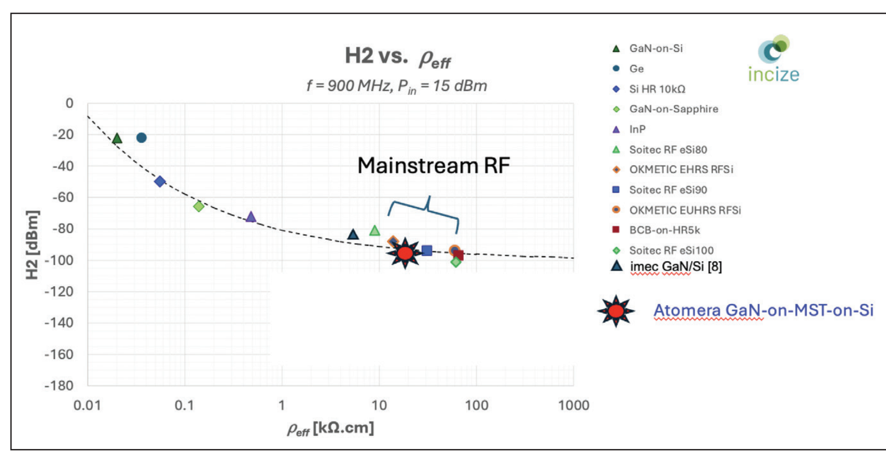
Built on experience

While Atomera may be a new name to many in the compound semiconductor industry, it has a long history, having been founded in the early 2000s. Initially, efforts focused on *ab initio* simulations of the addition of oxygen into silicon lattices, proving the stability of the structure and the epitaxial growth process for realising such structures in silicon devices.

According to company CEO and President Scott Bibaud, the global financial crash of 2008 hampered progress, due to difficulties in raising funds. But within a few years, the team were once again progressing their technology.

“We didn’t have the money to really aggressively go after customers until we went public in 2016,” explains Bibaud. “Public markets have been able to give us funding to continue developing this [technology] and bring it out to customers.”

➤ Scott Bibaud (left), has been Atomera’s CEO and President since 2015. During his career, he has held positions at Broadcom, Conexant and Raytheon. Robert Mears (right)co-founded Atomera in 2001, and is the company’s CTO. As well as developing MST, he has played a key role in the developed of erbium-doped fibre amplifiers.



➤ Benchmarking various RF technologies highlights the capability of GaN-on-silicon, when it includes MST technology.

Bibaud says that Atomera is currently working with handful of customers in the silicon industry, and claims that “virtually all” major semiconductor companies have, or are, evaluating MST technology.

For silicon CMOS processes, standard substrate orientations are (100) and (110), alongside an offcut of around 4°. So, for the work related to GaN RF HEMTs, which are typically grown on silicon (111) with essentially no off-cut, Atomera has modified its MST recipe.

During these efforts, Atomera found that its MST technology provides an extra degree of freedom in the lattice, and helps accommodate lattice mismatch between the III-Ns and silicon.

For companies that adopt MST technology, disruption should be minimal and additional cost marginal. As most firms already have epitaxial tools in their fabs, these just need to be modified to allow the injection of oxygen into the reactor.

“The implementation is roughly similar to the time and effort required to put in strain engineering,” argued Bibaud, who explains that Atomera’s business model is for customers to pay a royalty on their wafers, once they have been shipped. “Generally speaking, that [royalty] is a small fraction of the economic benefit they get from the improvement we bring them.”

In addition to these chipmakers, fabless firms and foundries can deploy MST technology.

“When foundries offer our technology as a standard part of their PDK, we will be out evangelising to their fabless customers, so they understand what benefits they get by using our technology,” enthuses Bibaud.

There’s no doubt that Atomera’s technology has much promise for the production of GaN-on-silicon RF HEMTs. But there’s still a long way to go, beginning with engaging with customers, and getting them to evaluate the technology. The appeal is that significant rewards result, due to the potential to combine low costs with impressive performance figures and compatibility with silicon CMOS technology.

Atomera’s Mears Silicon Technology, which addresses this weakness, has already been shown to enhance a broad swathe of silicon devices, from power switches to leading-edge gate-all-around transistors

Taking GaN in a different direction

Infineon has just released a pair of bi-directional switches for mobile applications that block up to 40 V

BY RICHARD STEVENSON, EDITOR, CS MAGAZINE

FOR MANY MAKERS of GaN power devices, much effort is directed at increasing their blocking voltage from around 600 V to 1 kV and beyond.

By entering territory that's occupied by SiC MOSFETs, there's the opportunity to compete for sales to the electric vehicle industry and to installers of electrical grid infrastructure.

But there is also the promise of ramping revenues by launching products operating at much lower voltages, which can be deployed in mobile devices – note that here, it is also possible to exploit the strengths of GaN, such as high switching frequencies, excellent efficiencies and impressive power densities.

One company seeking success on the latter front is Infineon. This multi-national launched its first 40 V bi-directional switch last summer, and based on its reception, has just expanded the portfolio with two more products providing the same blocking voltage: the IGK048B041S and IGK120B041S. This duo differs in size – they have dimensions of 2.1 mm by 2.1 mm, and 1.7 mm by 1.2 mm – and in drain-to-drain on resistances, which are 4.2 mΩ and 9 mΩ, respectively.

A key asset of Infineon's switches, giving them an edge over the incumbent silicon solution (a pair of MOSFETs), is their bi-directionality. This means that they can conduct in both directions, and also block up to 40 V in both directions.

A collaborative development

For the development of Infineon's first product, the company collaborated closely with a smartphone manufacturer.

"For us, this was also co-developing, learning, understanding the value proposition," reveals Infineon's Senior VP and General Manager of the GaN Business Line, Johannes Schoiswohl. He told *Compound Semiconductor* that they drew on what they had learnt from the launch of the first switch when introducing the two new products, a move that offers customers different values for the likes of resistance and impedance.

Given that smartphone batteries provide just a few volts, deploying devices with a blocking voltage of 40 V suggests excessive over-engineering. But that view overlooks that fast-charging a mobile a via USB-C cable may involve voltages of 20 V or 28 V. It's also possible that when plugging and unplugging cables, voltage spikes may occur, which must be protected from the battery. "So there needs to be enough headroom. That's why it's a 40-volt device," says Schoiswohl.

As well as targeting mobile phones, Infineon is eyeing opportunities in wearables, such as smartwatches, fitness trackers, augmented-reality glasses, wireless earbuds, and smart rings. For all these products, high-power charging is attractive.

Inside every Infineon 40 V bidirectional switch is a pair of HEMTs, united monolithically. Employing a common source or drain ensures bi-directionality, along with control of the current flow and blocking behaviour. Using these small switches as alternatives to pairs of silicon MOSFETs, arranged in a back-to-back configuration, halves component count and trims the footprint by up to 82 percent.

Another advantage over the silicon MOSFETs is a reduction in gate charge by 40 percent.

"So, essentially you can say it turns off with double speed, versus a silicon switch," says Schoiswohl, who explains that superior protection results.

Offering yet another advantage over silicon technology is the minimal drain-to-drain leakage current – it is 85 percent lower. That's a very attractive feature for makers of smartphones that will charge the batteries to a certain level before shipping. It's imperative that their smartphones, which could sit on the shelves of stores for many weeks, will turn on when shoppers compare various models.

As well as improving the performance of mobile devices, Infineon's bidirectional switches can help fulfil safety requirements, needed to reduce the likelihood of electrical fires from lithium-ion batteries. There is much documentation to support an ever-growing number of stringent safety requirements, including a need for

higher overvoltage protection. The 40 V bi-directional switch addresses these specifications.

Thanks to the 40 V capability, Infineon's switches are also ideal for load-switching and power multiplexing in multi-rail power architectures, where there's a need for precise control of current direction across multiple supply rails. This situation is a common one, as the electronics within a smartphone may run off of a 5 V rail, and fast-charging may operate at 20 V or 28 V.

Scaling production

Production of Infineon's switches is not particularly complicated, according to Schoiswohl: "It's just a different layout, versus the unidirectional switch."

Today, the company makes its bidirectional switches on 200 mm lines. "But the target is to be in mass production by end of next year on 300 millimetre."

A downside of moving to the larger format is that it exacerbates issues associated with wafer bow, stemming from lattice and thermal mismatches between GaN and silicon. But within Infineon's GaN portfolio, the 40 V devices should be relatively easy to move to the 300 mm line, given the relatively thin GaN layers required to block 40 V.

Given the relatively thin GaN epilayers in 40 V bi-directional switches, these products are strong contenders for volume production on 300 mm lines.

On the plus side, there is much to motivate a move to the bigger wafers. "Because the tools are so advanced, that makes it really tempting to move as fast as possible to 300 millimetre, especially for us, because we can reuse 80 percent of our silicon 300 millimetre lines."

When Infineon ramps production, the handset will yet again provide a major market for a compound semiconductor device. While smartphones no longer account for significant sales of LEDs, they still provide much business for makers of GaAs RF devices, and they now look to offer a significant revenue stream for some producers of GaN power electronics.

Rear view of a tractor applying UVC irradiation in a vineyard to provide sustainable protection.

How to benchmark the UVC LED

A new metric from ams Osram aids the evaluation of UVC sources

BY ALEXANDER WILM FROM AMS OSRAM

THE DEPLOYMENT of UVC LEDs is on the up, with sources increasingly appearing in disinfection and treatment systems across consumer, professional, and industrial markets. As uptake expands, design engineers face more decisions relating to the specifications of the emitter, such as having to weigh up the importance of optical output, efficiency, wavelength, lifetime behaviour and cost. While datasheets list many of these parameters, designers still lack a single figure-of-merit that consolidates them into an application-relevant basis for comparison. This omission matters, complicating component selection, reducing transparency in technology benchmarking, and delaying optimisation, both at the source and system level.

A multi-dimensional decision

To illustrate the challenge associated with selecting an optimal UVC LED for a particular application, let's consider this situation when trying to identify the best device for disinfection. The process begins by defining disinfection objectives, such as target dose, required reduction level and exposure geometry, and aligning all of them with the system constraints – they may include available electrical power, the thermal environment, form factor, safety and regulatory requirements, and expected operating profile. Only once that task is completed is there a context for evaluating an appropriate balance of technical performance and cost.

Within a typical UVC LED, there's much variety – differences in radiant flux levels, efficiencies, forward voltages, wavelengths, package types, and reliability classes – with each product potentially optimised for a distinct set of system conditions. However, while this variety is beneficial, it hampers comparison. For example, two LEDs may look similar when judged by just one parameter, but behave very differently, reflecting their full set of attributes.

Due to this, when selecting the most appropriate UVC LED for a particular task, trade-offs are inevitable. A common one is that of efficiency versus cost, as higher efficiencies trim running costs, but may carry a higher price tag. Another matter to consider is lifetime versus output. Cranking up the drive current boosts the optical output of a UVC LED in the short-term, but accelerates degradation unless effort is directed at a superior thermal design. There's also the need to consider effectiveness at different wavelengths, as the germicidal effect varies, due to wavelength-dependent action spectra.

One may argue that having a decision-making process that evaluates a range

of characteristics is hardly unique to the UVC LED. However, when selecting this particular device there are significant consequence at the system level, due to differences in delivered dose, spectral weighting, and degradation behaviour impacting performance margins and the cost over the product's life.

Traps for the unwary

In consumer applications, decisions are often dominated by one or two parameters, and it's possible to enjoy an acceptable basic performance while prioritising low component costs. Under these circumstances, it can be enough to simply rank products by their purchase price or nominal radiant power.

It's a different ball game for professional applications. Here, the decision space is more complex. System operation is often either continuous or long-duration, and the source cost, electrical energy consumption and lifetime may govern the total-cost-of-ownership and maintenance strategy. So, it may be misleading to compare LEDs with a single parameter, as the 'best' option depends on how parameters interact at a certain operation point.

Helping the decision-making process are visualization methods, such as a spider (radar) chart. They help to consider multi-parameter data, by communicating differences at a glance. But these charts tend to be descriptive, rather than quantitative: they don't incorporate parameter interactions, they don't reflect how parameters translate into delivered germicidal performance, and they often fail to convert performance into economic outcomes. Due to these limitations, while they may support early-stage screening, they are less suited to providing the primary decision basis for optimising professional or industrial systems.

A new benchmark

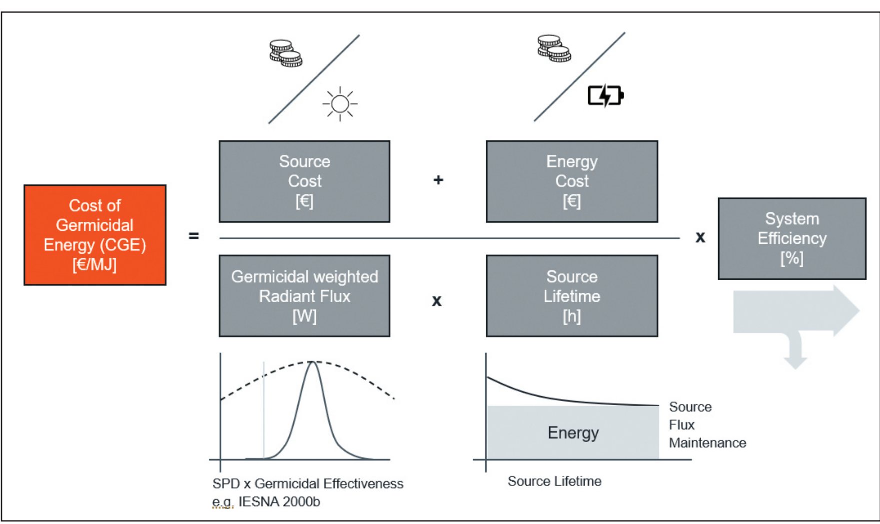
To provide a consolidated basis for comparison, our team at ams Osram has introduced a unified metric for UVC sources: the Cost of Germicidal Energy (CGE). This yardstick expresses the cost required to deliver a defined amount of germicidal-effective UV energy over the usable operating life of a source. It considers the source lifetime, which is the time it takes for the radiant flux to fall to a specified maintenance threshold or lifetime criteria, and ties this to a flux maintenance criterion – it might be an RxxB50 lifetime, which is the lifetime criteria at which 50 percent of LEDs are still above the specified performance limit of xx percent. The key implication is that performance is represented over time rather than at initial operation only.

One of the great strengths of CGE is that it integrates all the key technical and economic parameters into a single indicator. Included in the CGE is the: source cost, which is the purchase cost per UVC source; the energy cost, considered as the cumulative electricity cost over operating life; the lifetime and flux maintenance, defined as the operational duration until output declines to a defined maintenance threshold used in the CGE normalisation; the system efficiency, a figure that accounts for system-level losses, such as optical coupling efficiency, beam shaping and target geometry; and the germicidal-weighted radiant flux. The latter is the radiant output, weighted by an action spectrum (standardised, such as IESNA 2000b, or organism-specific, such as MS2), accounting for wavelength effectiveness and spectral power distribution, and scaled to the flux maintenance limit defined by the RxxB50 lifetime value.

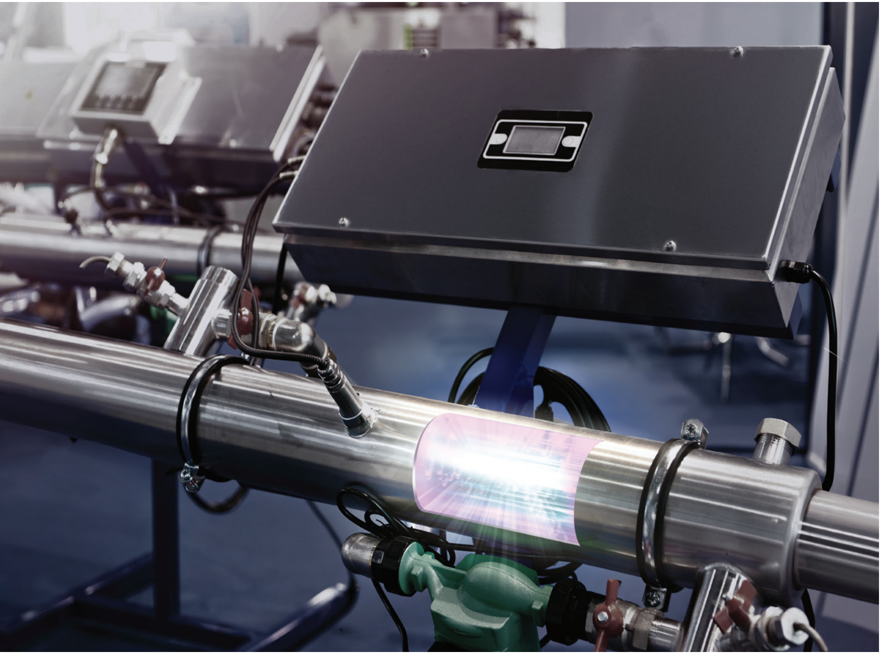
Conceptually, the CGE can be understood as the total cost over life (device + energy), divided by total germicidal-weighted optical energy delivered over that life in a dedicated system setup.

When using our new metric, it's easy to compare LEDs with different wavelengths, efficiencies, and lifetime behaviours, because differences are converted into a common, application-relevant quantity: the cost per germicidal-effective energy. In addition, the CGE enables comparison across technology classes when equivalent assumptions are used, such as electricity price, maintenance threshold, and action spectrum selection.

When using our new metric, it's easy to compare LEDs with different wavelengths, efficiencies, and lifetime behaviours, because differences are converted into a common, application-relevant quantity: the cost per germicidal-effective energy.



➤ The Cost of Germicidal Energy (CGE) may be viewed as the total cost over life (device + energy), divided by total germicidal-weighted optical energy delivered over that life in a dedicated system setup.



➤ Installations for disinfection of water with ultraviolet radiation.

Supporting engineering decisions

While a single metric will never replace detailed engineering analysis, it can improve decision workflow on a number of fronts. They include providing an objective ranking under defined assumptions, with CGE providing a consistent basis for comparing candidates when system requirements are fixed and assumptions are documented. In addition, there's: sensitivity insight, as by evaluating how CGE changes when parameters vary, it becomes clearer whether cost, efficiency, lifetime, or spectral effectiveness is the dominant driver in a given application context; and there's roadmap guidance – CGE can highlight which parameter improvements deliver the largest economic impact, supporting prioritisation in LED development and packaging innovation.

It should be noted that just because one decides to use the CGE metric, that does not prevent them from using additional metrics. When working with a variety of metrics, CGE acts as a unifying 'headline number', complementing detailed specifications and application constraints.

To system-level comparisons

When the CGE also incorporates and considers the system efficiency of UVC fixtures, the comparison naturally extends from source-level to system-level evaluation. In this form, CGE can compare complete UVC solutions – including different source technologies – because it accounts for how effectively each system converts electrical energy and component cost into germicidal-effective energy delivered to the target.

This capability for higher-level evaluation is particularly powerful when the system architecture strongly influences performance – that's the case in air and surface treatment, where optical control and directionality play a significant role; and in water treatment, where key factors include reactor geometry, hydraulic design, and UV utilisation. By including system efficiency, the CGE metric ensures that a number of advantages are accounted for, such as improved photon utilisation through tailored optics, reduced losses, and better irradiation of the treatment target within the same unified framework.

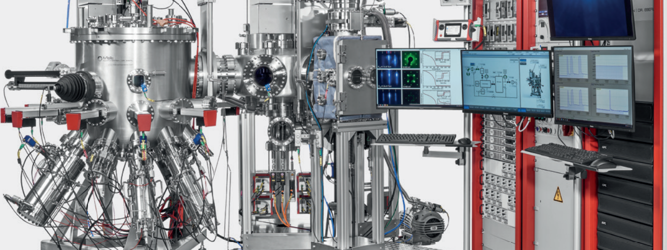
Thanks to this merit, the CGE is not just a tool for component selection. In addition, it's also a basis for comparing end-to-end system concepts on a common economic and performance axis.

FURTHER READING

➤ A. Wilm *et al*. "A unified framework for evaluating UV-C sources: introducing the cost of germicidal energy (CGE) metric," Proc. SPIE 13913, Light-Emitting Devices, Materials, and Applications XXX, 1391302 (March 4, 2026)

OCTOPLUS 800
MBE system with multi-wafer capability

- Engineered for advanced III/V heterostructure growth
- Maximized material utilization for optimal efficiency
- For 200 mm single wafers or up to 4x 100 mm multi-wafer substrate plates
- Up to 12 high-capacity source ports
- Fully automated wafer transfer
- Training and support by MBE experts with extensive application know-how



MBE Systems and Components - Trusted expertise since 1990 www.mbe-komponenten.de

12-14 APRIL 2027

Sheraton Brussels Airport Hotel Belgium

THEMES ANNOUNCED

AI: WHO WILL BE THE BIGGEST WINNERS?

With phenomenal demands for optical sources for data transfer, and the need for efficient power devices for megawatt servers, which companies will thrive in the AI era?

GAINING AN EDGE IN THE POWER INDUSTRY: SCALE AND STRATEGY

How do you time your transition to 200 mm and 300 mm lines to perfection? And what are the most effective strategies for expanding capacity?

MAKING MONEY FROM TRAILBLAZING TECH

Which novel ideas are opening up new, lucrative markets? And where's our industry delivering a step-change in device performance, to create must-have products?

STRATEGIES IN PHOTONICS: IDENTIFYING THE BEST OPPORTUNITIES

Which laser architectures will deliver the best bang-per-buck? Will microLEDs find a killer application in data transfer for data centres? And what are the biggest opportunities for deep-UV LEDs?

If you are interested in speaking at CS International 2027, contact Ranjodh Avern:
ranjodh.avern@angelbc.com
+44 (0)2476 718 975

https://csinternational.net

Angel
BUSINESS COMMUNICATIONS

Gallium oxide MOSFETs:
Why interfaces matter

Developers of gallium oxide power devices should draw on lessons learnt from the pioneers of compound semiconductor MOSFETs

BY IAIN THAYNE FROM THE UNIVERSITY OF GLASGOW

THANKS to its ultra-wide bandgap (UWBG), alongside availability of native substrates, β-Ga_2O_3 is steadily transitioning from an intriguing laboratory material to a serious option for next-generation power electronics. For researchers and technologists pursuing that goal and working across materials, devices, and integration, the central question is no longer whether gallium oxide channel transistors are scientifically interesting, but how – and how quickly – they can be translated into a robust, scalable device platform that aligns with realistic application and qualification requirements.

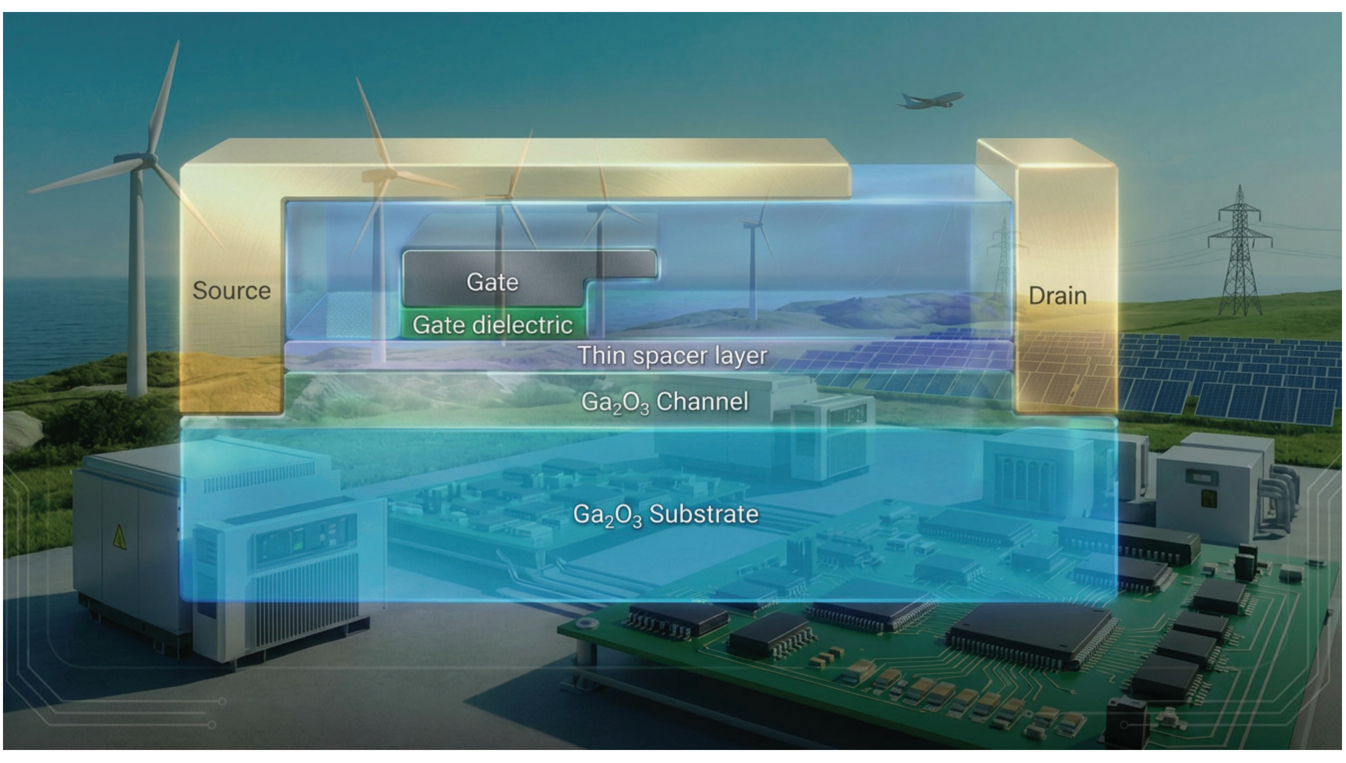
It is evident that today's gallium oxide MOSFETs remain at an early stage of technical maturity. Threshold voltages tend to lie on the wrong side of zero, subthreshold swings are well above the thermal limit, and there's a paucity of published reliability data. However, none of these observations are fundamental showstoppers. Rather, they reflect an emerging MOS technology, where progress in channel material development outpaces refinements in interfaces and gate stacks.

Many of the challenges now confronting gallium oxide are neither unprecedented nor unsolved. As is often the case, delving into the journals of yesteryear can speed development.

For today's developers of Ga_2O_3 MOSFETs, valuable insight can be gained from revisiting earlier work on GaAs siblings, where sustained effort was devoted to interface control, gate-stack chemistry, and charge management. Papers by pioneers of GaAs MOSFETs provide a convincing demonstration that high-performance MOS operation is achievable, even though the industrial demand required for large-scale adoption did not ultimately emerge at the time.

A recurring challenge: the MOS interface

Whenever a semiconductor material is proposed as a MOSFET channel, the focus quickly shifts from bulk properties to the gate interface. This makes much sense. After all, the interface is crucial, as it is where there's a transition from a crystalline semiconductor to an amorphous dielectric, a change that almost inevitably has the potential to introduce electronic defect states that degrade electrostatics, increase variability, and complicate reliability. One need look no further than the historical development of the silicon MOSFET to see that the ability to control this interface will often determine whether a material system advances from demonstration to deployment.



Gallium oxide is at this inflection point. Many reported β-Ga_2O_3 MOSFETs exhibit large subthreshold swings and threshold voltage instability, behaviour consistent with significant charge trapping, occuring at or near the gate dielectric interface. While there have been impressive advances in crystal growth, substrate availability, and doping control, the same cannot be said for interface engineering. Here, progress is slower, mirroring events in earlier compound semiconductor development cycles.

Lessons from GaAs MOSFET research

During the late 1990s and early 2000s, GaAs MOSFETs were intensively investigated for RF and logic applications. The motivation behind this is that GaAs offers outstanding carrier transport. However, this binary also suffers from well-known surface chemistry challenges, historically limiting MOS device operation. Through the focused research of that time, it was shown that these challenges are not fundamental.

One of the milestones of that era was the development of chemically compatible amorphous gate stacks. Introducing ultrathin gallium-oxide-based interlayers and carefully selecting ternary gadolinium-based high-κ dielectrics with related chemistry enabled the fabrication of GaAs MOSFETs with a low interface state density, near-ideal subthreshold swing, and positive threshold voltage. Crucially, these improvements were maintained through realistic fabrication sequences, demonstrating that these interface solutions were not merely laboratory artifacts.

Ultimately, GaAs MOSFETs with these highly engineered gate stacks did not enter large-scale production. But that's not because they failed to meet performance targets – it's because alternative device technologies were a better match to system-level needs, and the cost structures of the time. Nevertheless, the technical lessons learned – particularly regarding interface control – remain directly relevant to today's UWBG devices.

For those currently working on gallium oxide devices, a noteworthy finding from previous GaAs MOSFET development is the potential of appropriately selected ternary gadolinium-based high-κ materials, deposited optimally on Ga_2O_3. More generally, though, there's an underlying philosophy to take onboard – identifying amorphous dielectrics, engineered specifically for chemical compatibility with Ga_2O_3, deserves attention. Unfortunately, much current work involves incremental surface treatments, followed by deposition of conventional dielectrics. While this strategy delivers measurable improvement, it may fail to provide the degree of interface control needed for stable enhancement-mode operation, low variability, and acceptable lifetime under bias.

Switching to more fundamentally engineered gate stacks, combined with appropriate interlayers, promises to suppress interface trap formation systematically, rather than reducing their impact indirectly. For UWBG semiconductors, where deep traps exhibit slow dynamics and long recovery times, this distinction is particularly important from a device-physics and qualification perspective.

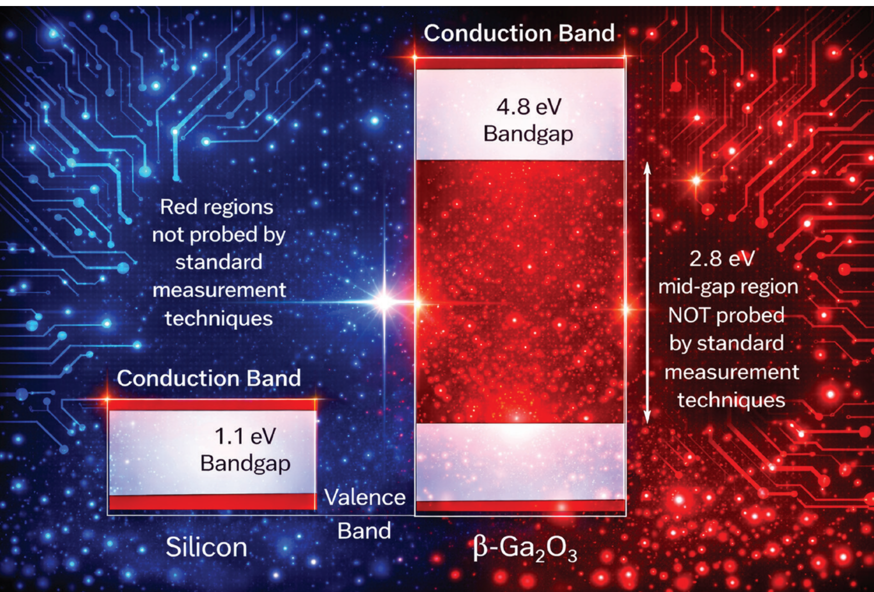


Figure 1. Standard gate dielectric screening methods probe defects/traps with energies across virtually all of the bandgap in silicon-based devices (left). In contrast, for Ga_2O_3-based devices (right), trap/defect states with a wide range of energies around mid-gap go undetected. As a consequence, dielectric trap state densities will be underestimated. This is why transistor sub-threshold swing is a better metric for dielectric screening in Ga_2O_3 MOSFETs.

Subthreshold swing: A practical indicator

In gallium oxide MOSFETs, subthreshold swing is a particularly useful metric for assessing interface quality. In both wide and ultra-wide bandgap materials, many electrically active states lie deep in the bandgap, making them difficult to probe with the conventional capacitance-based techniques, which were developed to screen oxides for silicon MOS devices as illustrated in Figure 1. Subthreshold swing, by contrast, reflects the cumulative impact of these states on channel formation and electrostatic control.

Based on this state of affairs, as gallium oxide devices mature, convergence toward lower and more reproducible subthreshold swing values will be a key indicator that interface-related challenges are truly being addressed. From an application standpoint, improvements in this metric are closely linked to threshold voltage stability,

Switching to more fundamentally engineered gate stacks, combined with appropriate interlayers, promises to suppress interface trap formation systematically, rather than reducing their impact indirectly

device-to-device uniformity, and ultimately manufacturability.

Architecture considerations

Beyond gate materials, device architectures play an important role. During the development of GaAs MOSFETs, heterostructure channels were used to spatially separate carriers from the oxide interface. This led to reduced carrier interaction with interface states. Similar 'thin spacer layer' concepts are now being explored in gallium oxide, through the use of aluminium- and indium-containing oxide alloys.

While one should not underestimate the materials challenges associated with such heterostructures, if successful implementation is realised, this could complement advances in gate stack design. In turn, this could ease interface requirements and expand the available design space for gallium oxide devices.

Towards application readiness

The path from promising research devices to application-ready power transistors is defined by more than headline performance metrics. As technologies move closer to deployment, it becomes increasing important to consider the likes of variability, bias stability, endurance under stress, and reproducibility across a wafer. Tightly coupled to each of these considerations is interface quality.

In the near term, if there is meaningful market penetration for gallium oxide MOSFETs, this will probably be associated with grid-level power infrastructure, solid-state protection, pulsed-power systems, and specialised industrial or aerospace platforms. For these applications, scaling volume is not a standout priority. What also matters is realising sufficient confidence in gatestack reliability, having thermal management strategies, and achieving long-term stability under realistic operating conditions. Success in these early application spaces would represent an important step toward application readiness, establishing gallium oxide as a credible UWBG option for demanding voltage regimes, and laying the groundwork for broader adoption as system requirements evolve.

Development of the GaAs MOSFET shows that it's possible to produce high-quality MOS interfaces on compound semiconductors by addressing chemistry, structure, and processing holistically. Optimisation of the gallium oxide MOSFET may benefit from leaning into these approaches, and leveraging knowledge surrounding gadolinium-based high-κ dielectrics deposited on Ga_2O_3, as well as the judicious use of hetero-interfaces.

FURTHER READING

➤ M. Passlack *et al*. "Gallium oxide in field effect transistors – a perspective," Journal of Vacuum Science & Technology **A 44** 038503 (2026)

III/V-Reclaim

More than just reclaiming...

Sustainable
Innovative
Unique

Materials
GaAs
InP
Ge
GaSb
InSb
GaP
InAs
AlN

Epi polishing
low removal process for epitaxial layers to enable direct bonding

Backside thinning
down to 50 micron

sales@35reclaim.com
www.35reclaim.com

kSA ICE for MOCVD and MBE

Real-time Deposition Monitoring

- Real-time temperature, curvature, growth rate, reflectivity, and more
- Custom configured for your chamber geometry
- Plus the best support in the industry
- Learn more by visiting our website: **k-space.com**

Improving the performance and production of the InP transistor

Leading producers of InP HEMTs and HBTs have increased the gain, efficiency and yield of their devices through process improvements and refined epitaxial structures

BY RICHARD STEVENSON, EDITOR, CS MAGAZINE

WITH INTEREST in AI at fever pitch, it's easy to understand why the supporting technologies are generating eye-catching headlines. There's positive press surrounding the manufacturers of InP lasers, used to transfer data, and the supporting cast, including the producers of InP substrates. Makers of these devices and their foundations are now darlings of the stock market, given their phenomenal gains in share price over the last year or so.

Against this backdrop, the progress of other important classes of InP devices – including HEMTs and HBTs – are attracting less attention. But these gains are valuable, given that the InP transistor is the trailblazer when it comes to delivering high-quality amplification at incredibly high frequencies. Thanks to this, InP transistors are deployed in radar, atmospheric sensing, communication systems, and high-speed test and measurement equipment.

The manufacturers of these devices are continuing to make progress on a number of fronts, including performance, yield and reproducibility. These advances featured at this year's CS Mantech, held in mid-May in Portland, Oregon, with papers on this theme presented by Northrop Grumman, Teledyne Scientific and WIN Semiconductors.

Serving space applications

Northrop Grumman has developed an InP HEMT technology that's claimed to be cutting edge, in terms of its high frequency, low noise and low-DC-power electronics. One key feature of this device is its InAs composite channel (IACC) that enables operation up to 1 THz, when employed in conjunction with aggressive device scaling.

Efforts are being directed at improving the capability of this technology through the shrinking of device dimensions and an increase in the indium content of the channel from $In_{0.6}Ga_{0.4}As$ to InAs (see table 1). Devices with a 100 nm gate length and a $In_{0.6}Ga_{0.4}As$ channel are already serving in a number of applications, while those at more advanced nodes are under development.

There are opportunities for Northrop Grumman's InP HEMT technology in government and commercial sectors, and in particular in communications with high-bandwidth requirements, as well as in radar and atmospheric sensing. The

Specifications	Technology node			
	1 µm InP HEMT	**IACC70**	IACC35	IACC25 (THz)
Gate length (nm)	100	**70**	60	40
$In_xGa_{1-x}As$ channel (%x)	60	**100**	100	100
R_{con} (mΩ mm)	120	**40**	40	40
Peak gm (mS/mm)	1400	**2500**	2500	3000
f_t (GHz)	200	**250**	400	600
f_{max} (GHz)	400	**600**	1100	1500
Wafer diameter (mm)	100	**100**	75	75
MRL	9	**In Qual**	4	3

Table 1. The InP technology nodes at Northrop Grumman.

company's technology is already demonstrating its reliability and high-level performance in two NASA programmes: the Temporal Experiment for Storms and Tropical Systems mission that have deployed InP HEMT technology on cubesats, and provided detailed observations of the microphysics of hurricanes, typhoons and tropical cyclones during three consecutive hurricane seasons; and the Electrojet Zeeman Imaging Explorer, a heliophysics mission that employed three SmallSats to study auroral electrojets – they are the electrical currents that flow 60 to 90 miles above the poles, linking the aurora to Earth's magnetosphere.

At this year's CS Mantech, Micheal Eller and co-workers detailed efforts to improve the yield of Northrop Grumman's 70-nm-gate-length HEMTs, a technology referred to as IACC70.

To produce these transistors, Eller and colleagues produce epistructures with an InAs composite channel (see Figure 1 for details). Wet etching and ion implantation ensures device isolation, prior to the use of electron-beam lithography, which defines openings of around 50 nm in an electron-beam-sensitive resist.

A crucial step in the fabrication process is the wet gate recess etch, undertaken before gate metal deposition. This etch removes the InGaAs cap under the gate, and isotropically etches into the InAlAs barrier to form a semi-circular recess.

Eller and co-workers have found that the gate recess etch is the most critical step in the fabrication of the IACC70 HEMTs. There's significant non-uniformity and poor reproducibility of the epitaxial structure, caused by a within-wafer etch-rate variation and poorly controlled time etching.

Highlighting this variation is a 100 mm wafer map of the gate-source voltage at peak transconductance (see Figure 2 (a)). This metric tends to be more positive as the gate is positioned closer to the channel. The team has also plotted peak transconductance, which is higher when the barrier is thinner (see Figure 2 (b)). However, go too far and the Schottky gate significantly depletes the channel – and forward biasing fails to restore the channel charge and current, due to an excessive gate tunnelling current. The off-state gate-leakage current has also been plotted (see Figure 2 (c)). It increases gradually as the gate recess becomes deeper, and is said to rise dramatically for stronger enhancement-mode devices.

It has been found that the RF performance of IACC70 HEMTs is also very sensitive to the final gate length and the ratio of the gate length to the barrier thickness – its optimal value is about 10, so it is challenging to optimise RF performance and mitigate short-channel effects. There's the concern that if there's no etch-stop process, a wide range of gate-recess depths combine with small variations in gate length to produce a significant spread in aspect ratios.

Looking into this matter, Eller and co-workers have considered two standard MMICs: a Ka-band standard evaluation circuit and a Q-band low-

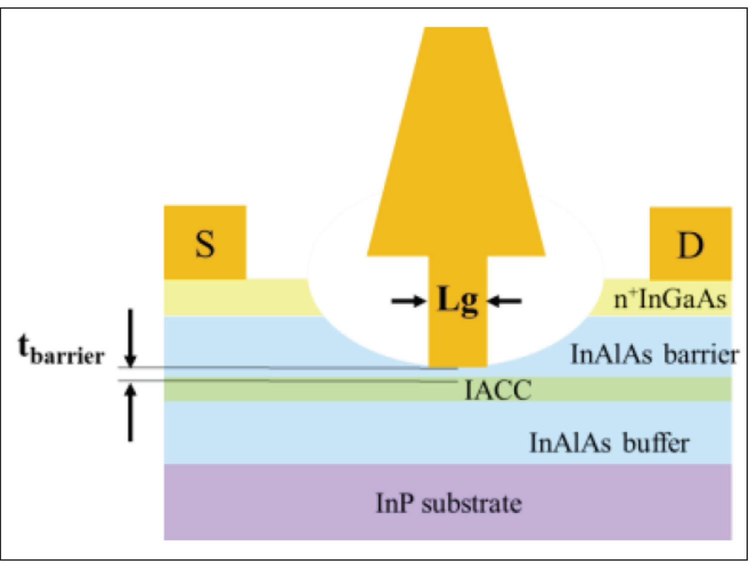


Figure 1. For Northrop Grumman's IACC70 HEMT, the gate recess depth and barrier thickness determine critical characteristics, such as transconductance, threshold voltage and gate leakage.

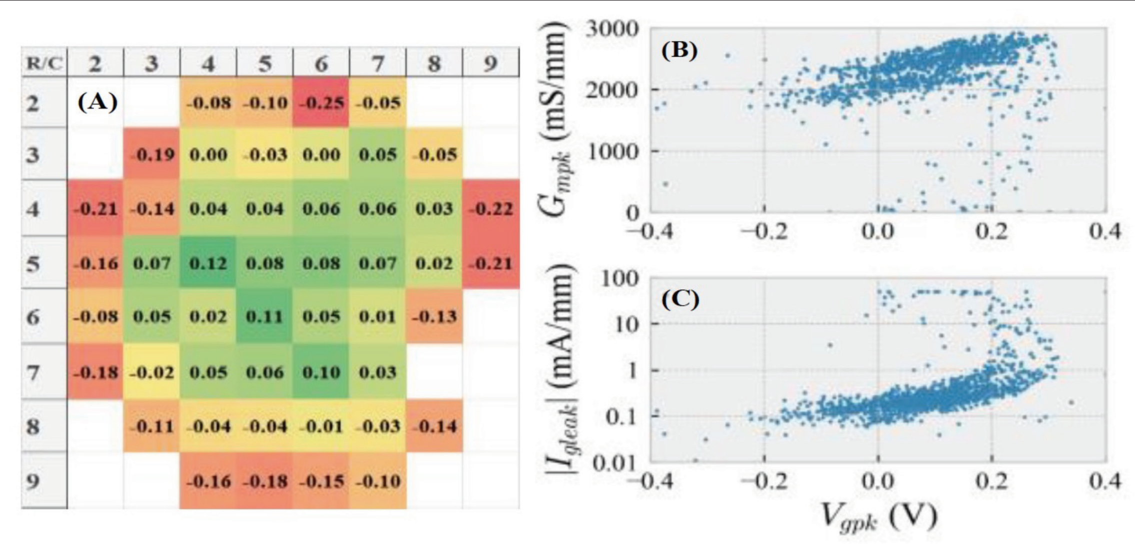


Figure 2. (a) Maps show a significant value of the gate-source voltage of a wafer, for a first generation of Northrop Grumman's IACC70 HEMT. There is also a significant spread in values for (b) peak trans-conductance and (c) gate leakage current.

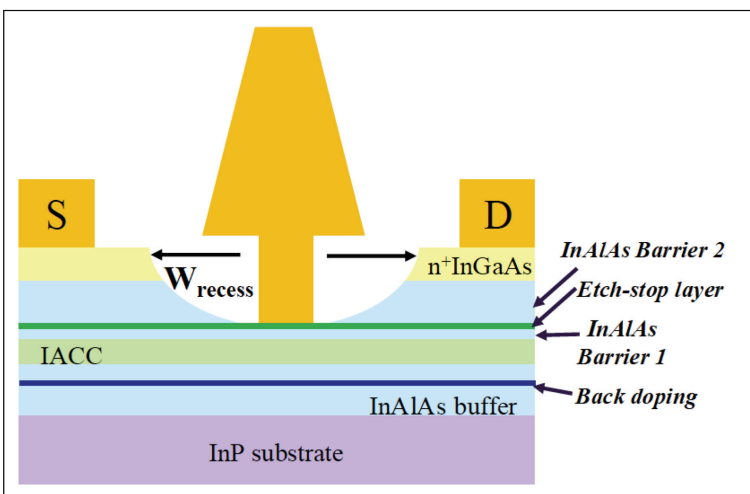


Figure 3. Northrop Grumman modified the epitaxial structure for its IACC70 HEMT, adding an etch-stop layer.

noise amplifier. Based on this study, they found that to maintain good DC and RF performance, it is essential to have precise control of barrier thickness and gate length.

To try and fulfil these objectives, engineers at Northrop Grumman have investigated the incorporation of a thin etch-stop layer, embedded within the InAlAs barrier (see Figure 3). This etch-stop layer has a slower etch rate in the gate recess etch than the adjacent arsenic-containing layer, and enables a uniform gate recess depth across the entire 100 mm wafer.

Another modification to the IACC70 HEMTs is the doping of the back barrier, designed to reduce the impact of parasitic access resistance and deliver an additional boost to drain current and device transconductance.

Plots of transfer sweeps and transconductance (see Figure 4) underscore the improved process, which has the additional benefit of allowing engineers to tune the lateral gate recess dimension while enjoying minimal impact on gate recess depth. Eller and co-workers claim that they can now modulate important device parameters, such as gate leakage, device breakdown voltage and gate-drain capacitance, in a controlled and uniform fashion that enables them to find the best device configuration.

The new design has enabled a clear increase in uniformity, as well as improved performance. There is a reduction of more than 70 percent in the spread of gate-source voltage at the peak transconductance, a fall of more than 80 percent in the spread of peak transconductance, and an

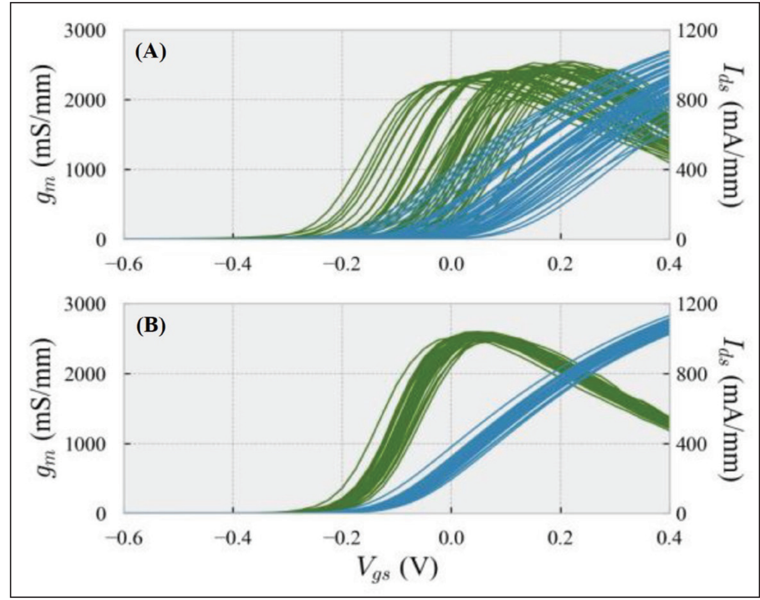


Figure 4. There are reductions in variations associated with the transfer sweep and transconductance in Northrop Grumman's IACC70 HEMT, thanks to improvements in the epistructure that now features an etch-stop.

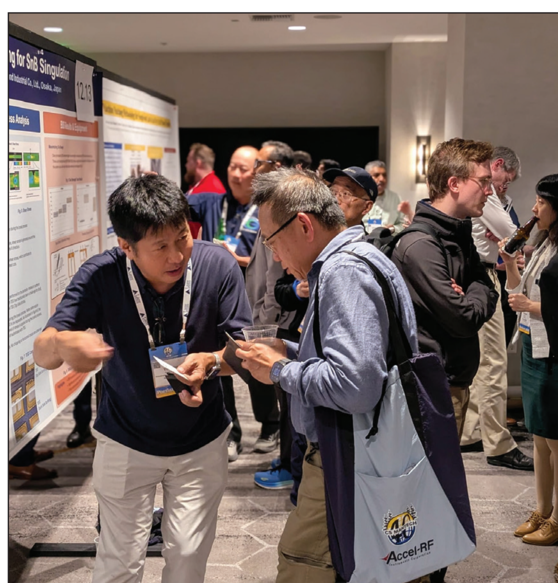
increase of about 0.4 dB in median maximum available gain. Thanks to these improvements, there is a significant improvement in MMIC yield (see Figure 5).

Combining cells, higher powers

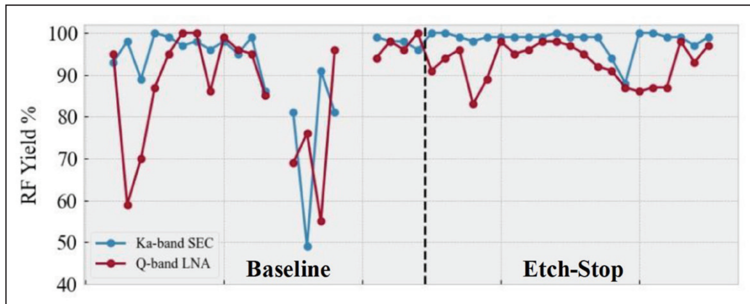
At this year's CS Mantech a team from Teledyne Scientific showcased power amplifier designs. Based on InP HBTs, these amplifiers, operating between 50 GHz and 270 GHz, are claimed to offer a combination of the highest RF output power, power-added efficiency and operational bandwidth.

The designs are constructed using the company's 250 nm and 130 nm HBT technologies, which are claimed to have a number of attributes, including: the definition of all the technology features for the 250 nm HBTs by optical lithography; use of an accurate physics-based scalable large-signal model; the inclusion of four metal interconnect layers, and a low-loss interlayer dielectric; a metal interconnect layer that prevents RF energy from entering the InP substrate; and thanks to the substrate playing no role in MMIC performance, the opportunity to use RF screening for automatic wafer probing and the identification of known good die.

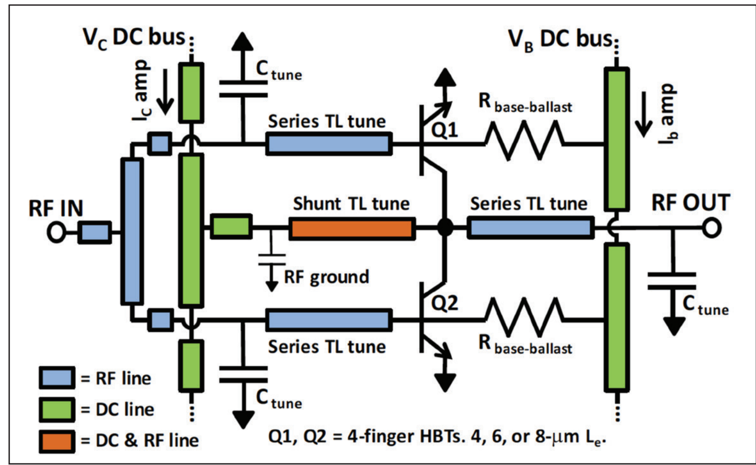
Recently, Teledyne improved its 250 nm HBT technology, trimming the base-collector capacitance by introducing a smaller metal contact to the HBT base terminal. This has enabled an increase in cut-off frequency, f_T, from 350 GHz to 380 GHz, and an increase in the maximum oscillation frequency, f_{max}, from 650 GHz to 700 GHz. In comparison, the 130 nm HBT technology has values for f_T and f_{max} of 500 GHz and 1 THz, respectively, but a slightly lower collector-emitter breakdown voltage – 3.5 V, compared with 4.5 V for the HBTs at the 250 nm node.



➤ The poster session plays a prominent role at CS Mantech, helping promote conversation between delegates.



➤ Figure 5. The refined IACC70 HEMT, featuring an etch-stop, has a much higher RF yield than its predecessor. Missing points indicate wafers that have failed line yield tests. RF yield is computed from DC yielding parts, based on specification limits for gain, input/output return loss, and noise figure.



➤ Figure 6. Teledyne's PA cell topology, employed for 50-200 GHz amplification, addresses the challenges associated with a 250 nm InP HBT common-emitter high-power amplifier design.

Engineers produce power amplifier MMICs from these transistors on 100 mm semi-insulating InP substates, with the designs guided by an established Keysight HBT technology model. To suppress PA parasitics in packaged devices, substrates are thinned from around 635 µm to just 76 µm.

The team from Teledyne say that they have invented a novel PA cell topology to address the challenges associated with a 250 nm InP HBT common-emitter high-power amplifier design (see Figure 6). Features of this topology include the biasing of the collector through a vertical DC bus that runs across the input matching networks – this configuration avoids RF output interconnect crossovers, and allows large currents required in InP HBT amplifiers to be distributed along the vertical DC bus. It is said that by avoiding the flow of DC currents across the output matching network or the on-chip combiner, a matching structure is realised with narrow transmission line traces that have a high impedance.

For PA designs operating above 200 GHz, the team from Teledyne use a cascode topology (see Figure 7). With this design, the base node of the common-base device is at DC and RF ground – a condition that minimises parasitic inductance at this node and broadens PA cell bandwidth.

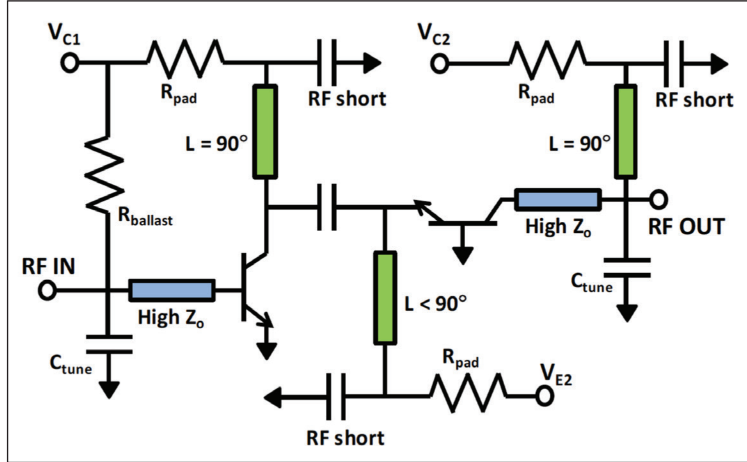


Figure 7. Teledyne's cascode power amplifier cell covers 200-320 GHz.

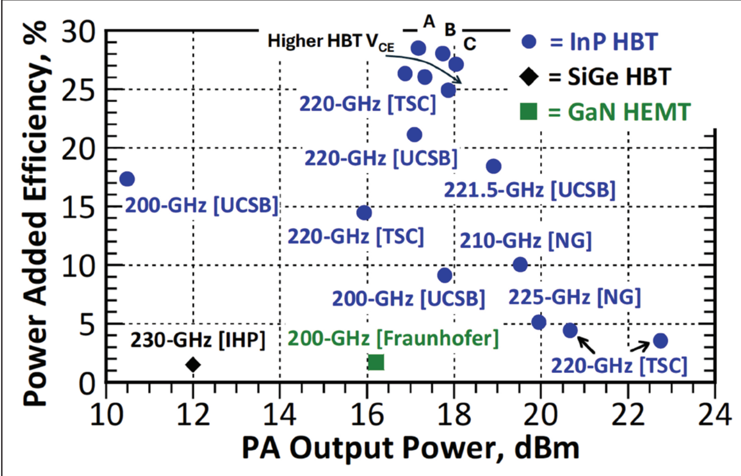


Figure 8. Teledyne's power amplifiers are expanding state-of-the-art capability.

Featuring four gain stages and an 8-way on-chip combiner, the team's 90-140 GHz PA, based on 250 nm HBT technology, produces a large-signal gain of 13.3-12.4 dB and an output power of 350-430 mW. Building on this design, the team aims to develop a 120-140 GHz PA that delivers more than 0.5 W and has a power-added efficiency of 15 percent.

For amplification spanning 140-220 GHz, Teledyne's engineers have produced an amplifier that has five stages, but no on-chip combiners. Large signal gain across this band is 13-16 dB, when using an input of 0 dBm. There are plans to develop two-way and four-way on-chip combiners, delivering 50 mW and 100 mW across the entire band. The team has employed its 130 nm HBT in a 220 GHz PA that has two gain stages and no on-chip combining. Using an input of 6.9 mW, output is 59.6 mW, indicating 9.3 dB of gain, which is realised at a power-added efficiency of 28 percent. One of the next goals with this design is to increase that efficiency to over 30 percent.

The performance of Teledyne's PAs has been benchmarked against other results, claimed to be state-of-the-art (see Figure 8). Zach Griffiths and colleagues stated: Results using InP HBTs show a clear performance advantage compared with other solid-state technologies, due to higher gain and available watts-per-millimetre power density.

Antimony addition

Another producer of InP HBTs is the foundry WIN Semiconductor. At this year's CS Mantech, the Taiwanese chipmaker reported results on D-band (130-175 GHz) devices, fabricated on its 150 mm line.

In the paper from WIN, Lai-Hsiang Kuo and co-workers pointed out that the D-band is a critical frequency domain for 6G wireless transmission and broadband amplification in optoelectronic ICs.

It is claimed that the InP/GaAsSb double-HBT is a compelling technology at these millimetre-wave frequencies, due to its combination of high values for f_T and f_{max} and its robust collector-emitter breakdown voltage. In addition, thanks to a type-II base-collector heterojunction, current-blocking effects are mitigated – that's not the case for conventional type-I InGaAs double-HBT architectures.

Lai-Hsiang Kuo and co-workers are by no means the first to produce InP/GaAsSb double-HBTs, with reports by other teams highlighting the promise of scaling performance to the terahertz regime. However, that body of work has very little to say regarding yield and large-scale manufacturability, with studies tending to be limited to InP substrates with diameters of 100 mm or less.

Addressing those limitations, WIN has introduced the first 150 mm InP double-HBT platform. Its heterostructure consists of an *n*-type InGaAs cap, an *n*-type InP emitter, an emitter stabilisation ledge, a *p*-type GaAsSb base, and an *n*-type InP collector – as well as an InGaAs etch-stop layer and an InP subcollector, included to aid process control and reduce collector series resistance.

Most wafer breakage originated from mechanical collisions, causing microscopic nicks at the wafer edges that led to catastrophic fractures during wafer transfer. By optimising the wafer transport path and standardising manual handling procedures, the team slashed the breakage rate for 150 mm InP wafers from 6 out of 11 to 1 out of 30.

To minimise extrinsic base-collector capacitance without compromising base resistance, the engineers at WIN employ a triple self-aligned process. This begins by using an emitter metal hard mask to define the self-aligned emitter mesa. The next step is the minimisation of emitter-base spacing, realised through precise control of the emitter mesa undercut and a self-aligned base metal deposition. Thanks to these two steps, emitter and base metals can serve as masks, enabling the formation of a self-aligned base-collector mesa.

WIN's engineers use a hybrid etching scheme, involving both dry and wet etching, to provide mesa isolation. Ensuring high anisotropy and etch selectivity, this is used to produce HBTs with a 0.3 µm emitter width and an emitter-base spacing of 0.13 µm.

Fabrication is completed with back-end-of-line integration, involving three levels of interconnect metal, and the use of benzocyclobutene for surface planarization and the reduction of parasitic capacitance. WIN's platform also supports multiple passive components, including 50 Ω/sq thin-film resistors and metal-insulator-metal capacitors with a density of 600 pF mm^{-2} (see Figure 9 for an overview of WIN's technology).

Using small-signal measurements, Kuo and co-workers have determined values of 350 GHz for f_T and 411 GHz for f_{max}. The maximum power output is 7.41 dB at 6 dB gain compression, according to large-signal 60 GHz load-pull measurements.

One of the team's aims is to establish the optimal trade-off between peak device performance and manufacturability with a high fabrication yield. When considering large-signal power performance, they have found that trimming the emitter width from 0.5 µm to 0.3 µm increases the output power density from 2.95 mW $µm^{-2}$ to 3.67 mW $µm^{-2}$. There are also increases in power gain and power-added efficiency, rising from 10.04 dB to 12.37 dB, and 31.4 percent to 37.5 percent. All these gains are attributed to a 14.4 percent reduction in base-collector capacitance.

Kuo and colleagues have revealed that early process iterations suffered from emitter and base metal collapse, leading to an unsustainable low fabrication yield – it is just 15.4 percent for devices with a 0.3 µm emitter width. This figure has been improved to 87 percent through systematic optimisation of lithographic lift-off parameters.

A study of process stability, using four pilot production runs, shows that all wafers have a yield exceeding 70 percent (see Figure 10). This trial confirms repeatability and process uniformity. In parallel, WIN's engineers have been enhancing mechanical stability, by refining wafer-handling methodologies. Initial failure analysis found that most wafer breakage originated from mechanical collisions, causing microscopic nicks at the wafer edges that led to catastrophic fractures during wafer transfer. By optimising the wafer transport path and standardising manual handling procedures, the team slashed the breakage rate for 150 mm InP wafers from 6 out of 11 to 1 out of 30.

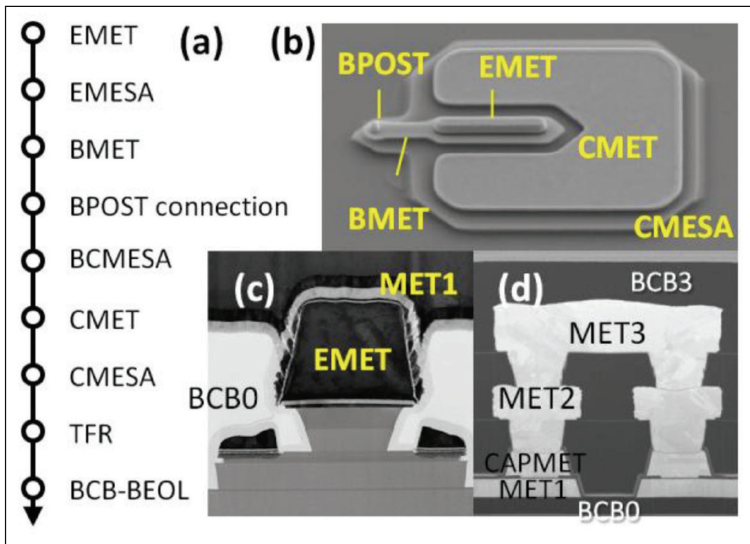


Figure 9. WIN has developed a 150 mm InP double-HBT technology. (a) Fabrication process flow. (b) Tilted-view, scanning electron microscopy micrograph. (c) Cross-sectional view of the device. (d) Benzocyclobutene-based back-end-of-line integration platform.

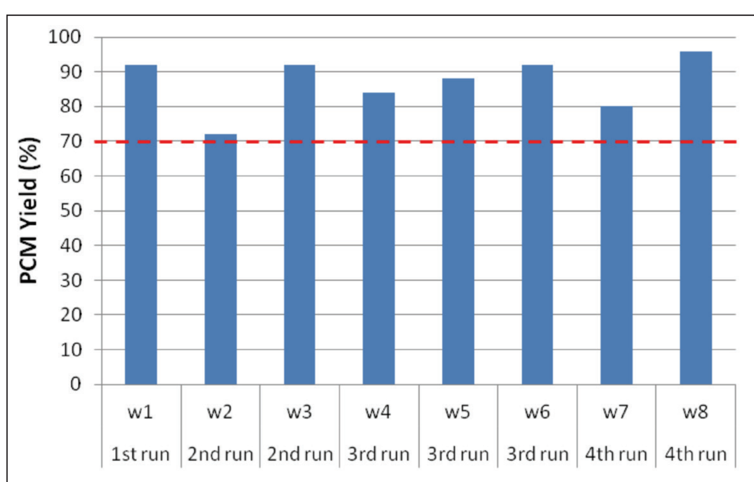


Figure 10. Yield repeatability across four pilot runs (eight wafers in total) at WIN Semiconductor.

Kuo and co-workers have found that increases to power density and power-added efficiency start to saturate when the emitter width is reduced below 0.4 µm. Based on this observation, an emitter width of 0.3 µm is viewed as optimal, balancing peak RF performance manufacturability and process yield.

The efforts by WIN, as well as Teledyne and Northrop Grumman, highlight the increasing capabilities of InP transistors, in terms of both their performance and the maturity of the manufacturing process. Expect further progress to be reported at next year's CS Mantech.

An elegant solution for SiC

By slashing the dislocation density, solution growth promises more reliable power electronics for electric vehicles and renewable energy systems

BY KE ZUNBIN, QIAN HAO, YIN ZURONG, WANG HAN, SHI ZHIJIE, HU JIANTAO AND LU MIN FROM CHANGZHOU PERFECT CRYSTAL SEMICONDUCTOR AND SHI TIANLE FROM THE UNIVERSITY OF SHANGHAI FOR SCIENCE AND TECHNOLOGY

THERE'S LITTLE DEBATE surrounding the darling of the power electronics world. It's SiC, a remarkable material that keeps electric vehicle motors spinning efficiently, ensures a smooth power flow in smart grids, and prevents data centres from overheating under immense computational loads. Its secret lies in its unique combination of assets: the thermal conductivity of copper, the breakdown strength of diamond, and the ability to handle voltages that would fry ordinary silicon chips.

But beneath this glittering reputation lurks a microscopic menace. The crystal defects that chipmakers struggle to eliminate – known as dislocations – hold back device performance and reliability. Today's commercial SiC substrates are riddled with these atomic-scale imperfections, with thousands threading through each square centimetre of material. Every one of them acts as a tiny electron trap, sapping efficiency and providing a pathway for destructive current leakage. For engineers designing the next generation of high-power systems, these defects are a critical bottleneck.

Fortunately, help is at hand. While the mainstream semiconductor industry continues to focus on physical vapour transport – the established but imperfect method for growing SiC crystals – researchers are refining an alternative approach that could leapfrog current limitations. It's solution growth, a technique that crystallises SiC from a molten metal bath rather than from vapour – and it's demonstrating an unprecedented ability to purge these troublesome dislocations.

Leading the way is our team at Changzhou Perfect Crystal Semiconductor, supported by the University of Shanghai for Science and Technology. Working together, we have systematically unpicked the mechanisms of dislocation behaviour during solution growth, revealing how to optimise this process and

Device	BPDs	TEDs/TSDs	MP
MOSFET	R_{on}↑ Gate oxide reliability↓	Gate oxide reliability↓	I_{leak}↑ Limit I_{op}
SBD	I_{leak}↑	I_{leak}↑	V_B↓
p-n diode	R_{on}↑	V_B↓	I_{leak}↑ Limit I_{op}
Bipolar/IGBT	Gate oxide reliability↓	τ↓	Limit I_{op}

Table 1. Effects of SiC dislocations on different devices (for more details, see Nanoscale Research Letters **17** 30 (2022). Note: V_B is blocking/breakdown voltage; R_{on} is on-resistance; I_{leak} is leakage current; I_{op} is operating current; τ is carrier lifetime.

produce crystals of remarkable purity.
Our findings point to a future where SiC wafers contain not thousands of dislocations per square centimetre, but merely hundreds. For device manufacturers, that means more efficient power converters, more reliable motor controllers, and ultimately, cheaper renewable energy.

The enemy within

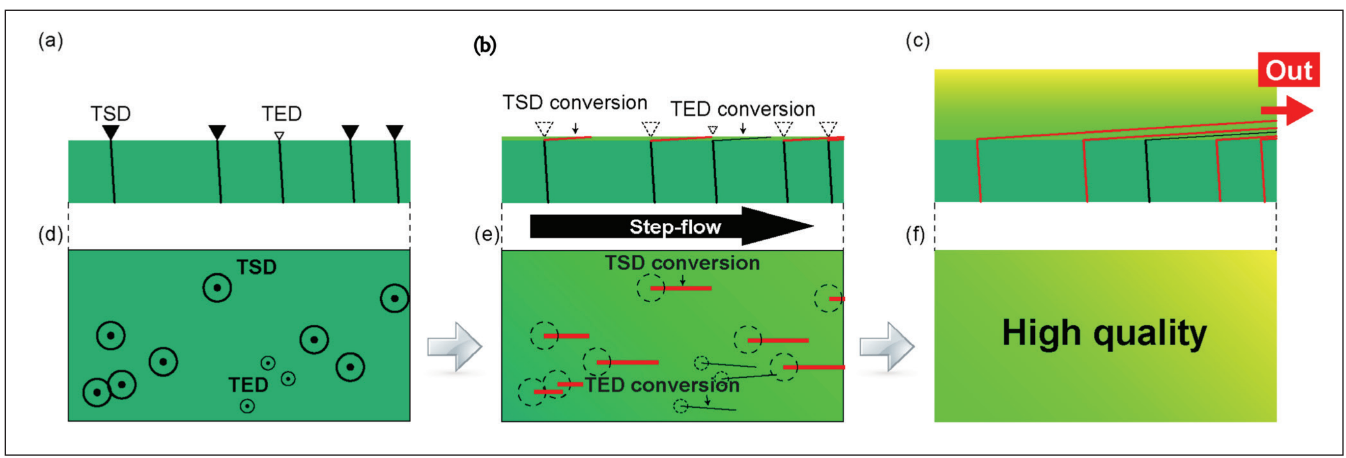
To appreciate the challenge in making a breakthrough, one must first understand the adversary. This comes in three forms, each with a different personality. There are: threading screw dislocations, spiralling through the crystal-like microscopic corkscrews; threading edge dislocations, creating atomic mismatches along their path; and Basal plane dislocations, slipping along the crystal's natural layering planes. All act as trap centres – electronic black holes that capture charge carriers and waste energy, lost as heat.

These dislocations have stark consequences. In MOSFETs, the workhorses of power conversion, these imperfections increase resistance and degrade the fragile gate oxide. In Schottky barrier diodes, they hike leakage currents. And in bipolar devices, they shorten carrier lifetimes. As illustrated in Table 1, every major device type suffers. Since commercial substrates contain between 10^3 and 10^4 dislocations per square centimetre, device manufacturers take a yield hit.

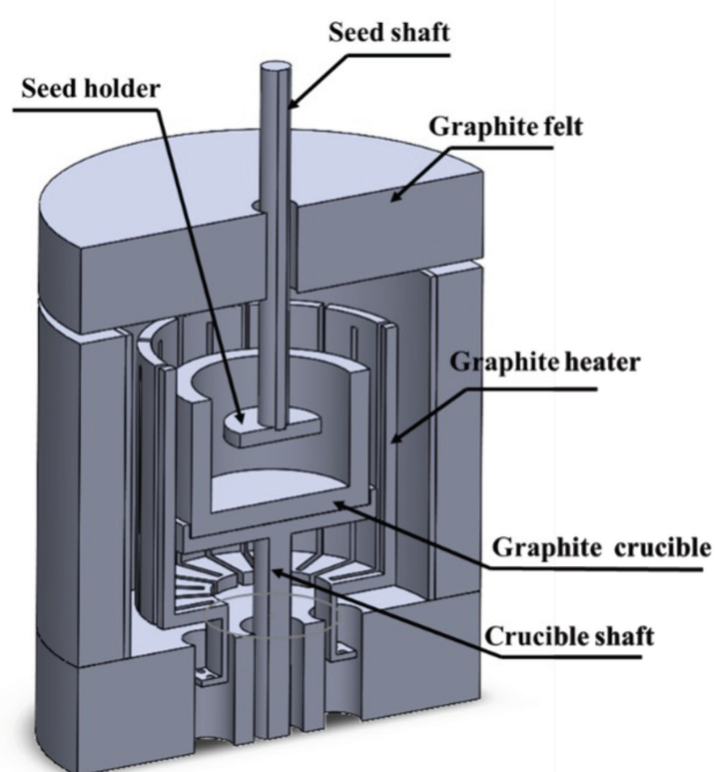
Traditional crystal growth methods struggle to reduce defect densities below this range. Physical vapour transport, which sublimates SiC powder at over 2,300°C before recondenses it as a crystal, inevitably introduces thermal stresses that generate fresh dislocations. Switching to high-temperature CVD ensures better control, but uses hazardous gases, and is prohibitively expensive for bulk crystal production. With both approaches hitting a quality ceiling, the power electronics industry is desperate for a solution – in fact, it has been wanting one for a number of years.

A superior solution

Solution growth takes a fundamentally different approach. It involves immersing a seed crystal in a molten solution of silicon, mixed with metals such as chromium and aluminium. A dash of chromium dramatically boosts carbon solubility, while aluminium smooths the growth surface. Using a modest temperature gradient of around 1,800°C, which is significantly cooler than vapour methods,



➤ Figure 1. The transformation dance, with threading dislocations in an off-axis seed crystal converted into basal plane defects as macrosteps flow across the growth surface. These defects glide sideways and exit the crystal, dramatically reducing defect density in the final wafer.



➤ Figure 2. Inside the furnace. A schematic of the resistance-heated single-crystal furnace used for solution growth. The graphite crucible holds molten silicon-chromium-aluminium flux, with the seed crystal positioned precisely above. Multiple heating zones allow fine control of the temperature gradient that drives crystal growth.

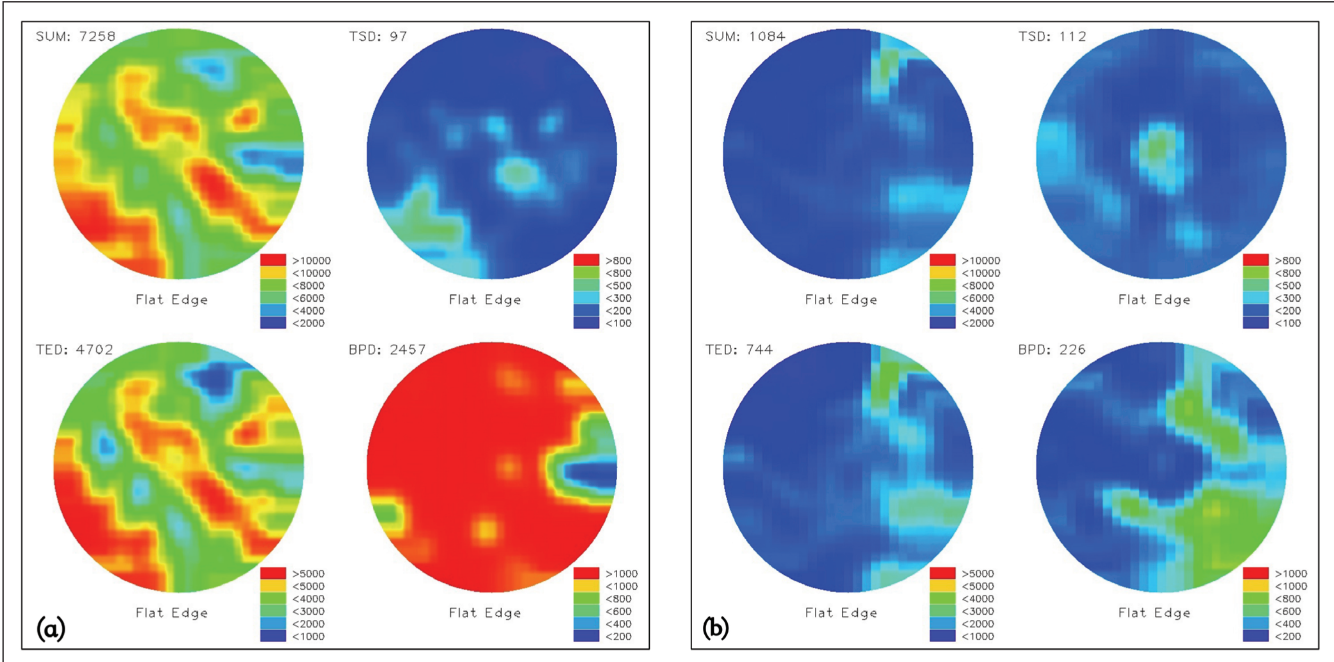
carbon atoms precipitate onto the seed, steadily assembling a high-quality crystal.

The advantages of solution growth are not limited to lower run temperatures. Additional benefits include a natural suppression of micropipes, the giant screw dislocations plaguing vapour-grown crystals. More intriguingly, solution growth actively converts problematic threading dislocations into less harmful forms, and can even sweep them entirely out of the crystal (see Figure 1). It's this transformative capability that we set out to understand and optimise.

Our experiments, conducted in a bespoke resistance furnace (see Figure 2), have focused on the systematic variation in key growth parameters. Using 6-8-inch 4H-SiC seed crystals – that's the industry standard polytype – we have grown crystals for 50 hours under controlled conditions.

Each of these crystals has been sliced, polished, and etched to reveal its defect structure under optical microscopy. Results provide a clear path to defect suppression.

The first breakthrough has come from temperature control. In solution growth, a temperature gradient at the growth interface drives the process, but excessive gradients spell trouble. When we pushed the gradient to 5°C per centimetre, dislocation densities soared. Dialling this gradient back to 2°C per centimetre slashed defects dramatically (see Figure 3). The mechanism behind this improvement is straightforward: thermal stress, induced by steep gradients, generates new dislocations upon relief. Moving to gentler gradients



➤ Figure 3. Temperature matters. Dislocation density maps of wafers grown under different thermal gradients. (a) A steep 5 °C cm^{-1} gradient produces high defect counts. (b) Reducing the gradient to 2 °C cm^{-1} slashes dislocations by minimising thermal stress that generates fresh defects.

allows the crystal to relax, preserving its structural integrity.

We have found that the moment of seeding is even more critical. When a cold seed crystal first touches a hot solution, there's a sudden shock, threatening to generate thousands of new dislocations.

However, by pre-heating the seed just 5 mm above the solution surface – giving it a modest 3°C temperature boost – we quash this thermal shock by half compared with a 10 mm stand-off distance. Thanks to just this simple procedural tweak, we are cutting the dislocation generation at the crystal's foundation, and influencing the quality of everything that grows above (see Figure 4).

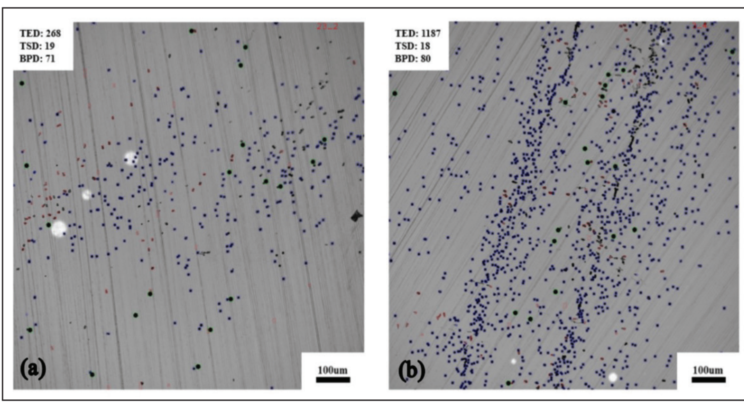
Pores are a second significant defect source. These gas bubbles, trapped between seed and solution, interrupt growth and spawn dislocations. We have identified two culprits: surface misalignment during seeding, and dissolved gases precipitating from solution.

Some of the pores are quite large, with dimensions of up to two millimetres. One of their downsides is that they create circular dead zones, where no growth occurs. But that's not the only issue – the surrounding crystal is rough, with solvent inclusions, causing nearby dislocation counts to skyrocket (see Figure 5). Equally damaging, while less visually dramatic, are smaller pores. We found that in a region with micron-scale pores, there were 628 threading dislocations in the microscope's field of view, versus just 172 in a pore-free region – that's a fourfold increase (see Figure 6).

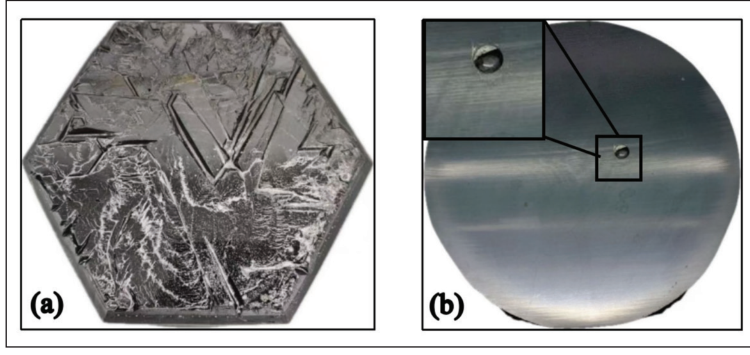
The solution is meticulous process control. It's possible to all but eliminate pore formation by ensuring parallel contact between seed and solution, optimising furnace pressure, and thoroughly dissolving any surface corrosion pits on the seed. Do all of this, and a smooth, uninterrupted growth front follows that doesn't generate fresh dislocations.

The expansion advantage

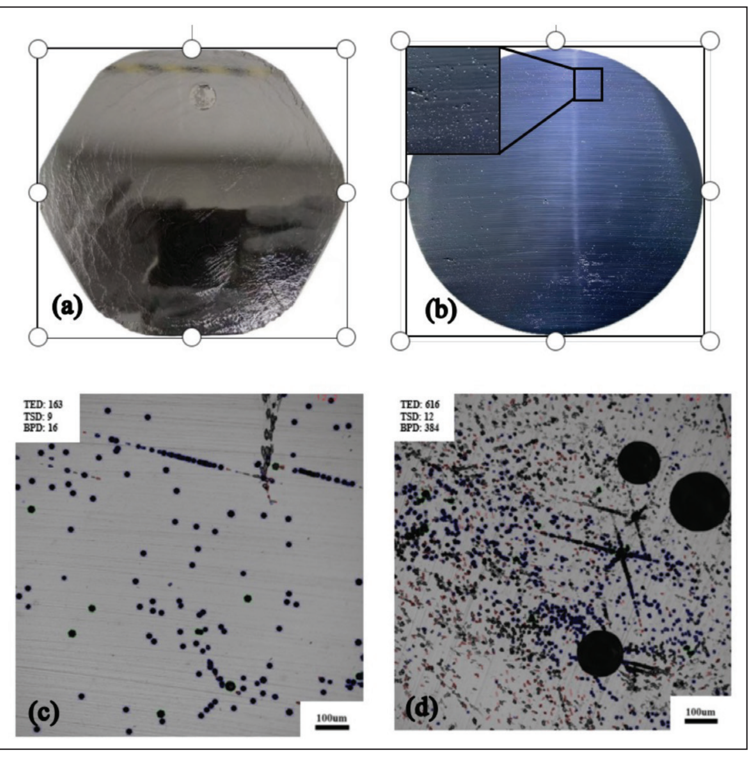
Perhaps our most valuable finding concerns diameter expansion. We have found that a remarkably effective approach to purging defects is to start with a small seed crystal, and gradually widen the crystal as it grows. Note that in the expansion region, the growth direction tilts relative to the threading dislocation lines, encouraging them to bend into the basal plane and exit through the crystal edge. In some experiments, we observed



➤ Figure 4. The shock of contact. Preheating the seed crystal close to the solution surface minimises thermal shock. (a) Baking the seed 5 mm above the solution, creating only a 3 °C temperature jump, yields far fewer dislocations than (b) a 10 mm stand-off with a 6 °C shock.



➤ Figure 5. When pores strike. (a) A crystal containing large millimetre-scale pores that interrupt growth. (b) A wafer sliced through such a pore shows a circular dead zone surrounded by rough, inclusion-rich crystal.



➤ Figure 6. Micropores, macro impact. (a) A crystal with micro-pores that closed early, leaving a smooth surface. (b) Edge micro-pores visible in cross-section. (c) Dislocation density in a clean region: 172 threading defects in view. (d) Near pores, this jumps to 628 – a fourfold increase.

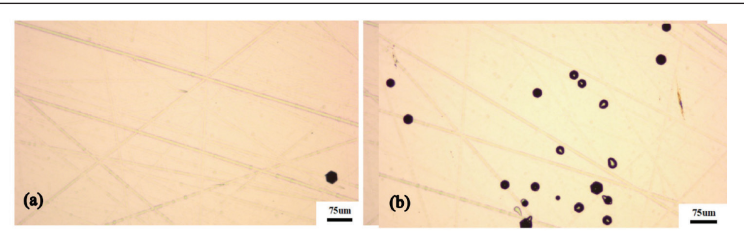


Figure 7. Expansion eliminates defects. (a) In the diameter expansion region, dislocation density drops dramatically as threading defects bend and exit the crystal. (b) The original seed region retains higher defect densities.

complete elimination of dislocations in the expansion region (see Figure 7).

This finding stands in sharp contrast to conventional on-axis growth, where threading dislocations propagate straight upwards, locked parallel to the growth direction. So, we can conclude that the solution method's inherent ability to control crystal shape becomes a powerful defect-filtration mechanism – a trick that vapour-phase methods struggle to replicate.

Another leaver for control is the orientation of the seed crystal. When growth proceeds on a precisely on-axis (0001) seed, this suppresses basal plane dislocations, which cannot easily propagate vertically. However, this geometry does nothing for threading dislocations, which sail straight through (see Figure 8).

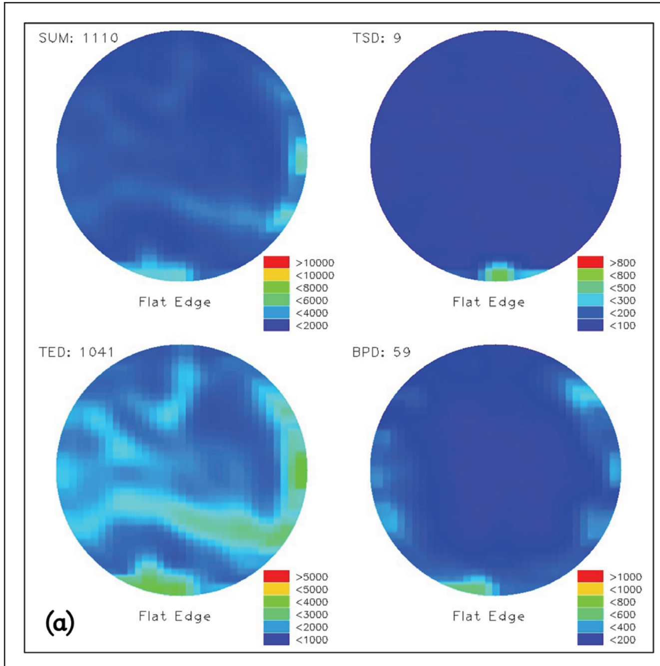
The game-changer is a 4° off-axis seed, tilted toward the <11$\bar{2}$0> direction. A stepped growth surface results, where macrosteps continuously flow across the crystal. Intersecting these steps are threading screw and edge dislocations. Basal plane defect steps result, which then glide laterally out of the crystal. Using this approach, we deliver dramatic reductions in threading dislocation density compared with the original seed crystal (see Figure 9).

From lab to fab

What do our improvements mean in practice? Our optimisation – gentle temperature gradients, pore elimination, diameter expansion, and off-axis seeding – work synergistically. While we have not reported a single combined figure from these approaches, individual improvements are substantial: cutting the temperature gradient from 5°C to 2°C per centimetre reduces thermal stress by more than half; eliminating pores prevents a four-fold spike in dislocations observed near bubble traps; diameter expansion enables local dislocation densities of zero; and off-axis seeding transforms the majority of threading dislocations, realised within the first few millimetres of growth.

Such success will pique the interest of device manufacturers, as it translates into higher production yields and better performance. Transformative gains may result, such as the difference between a power MOSFET that operates reliably at 200°C, and one that fails prematurely. A step change in dislocations could also enable Schottky diodes to block higher voltages without leakage, and open the door to IGBTs with longer carrier lifetimes for more efficient switching.

Implications extend well beyond individual devices. As electric vehicles move toward 800 V architectures for faster charging, demand for ultra-low-defect SiC



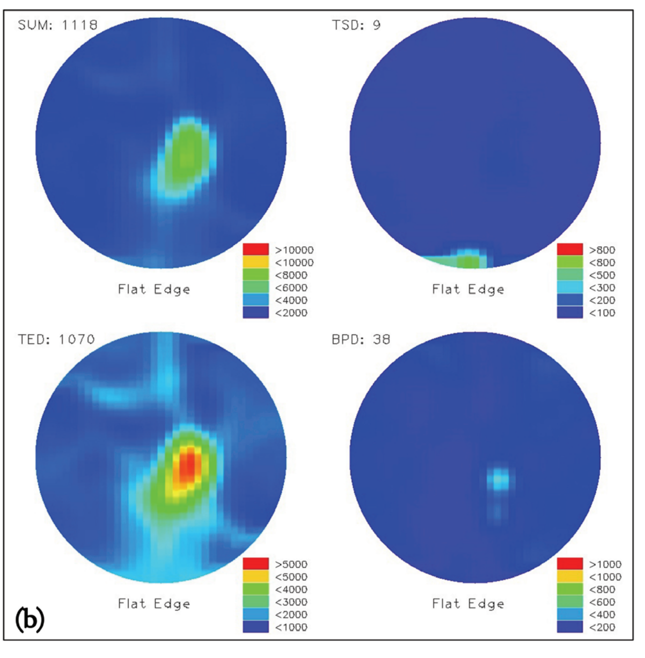
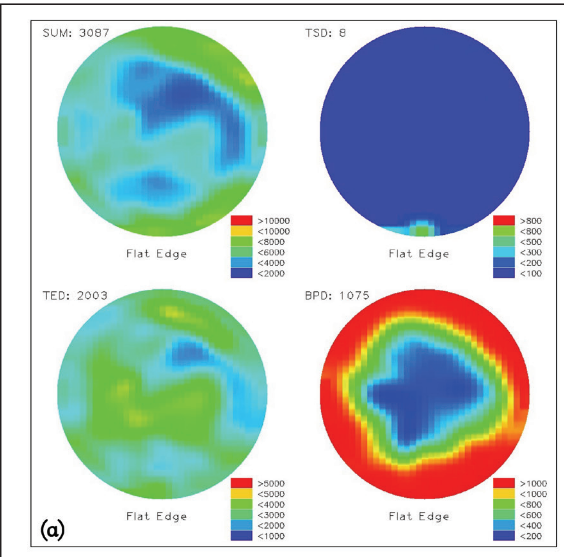
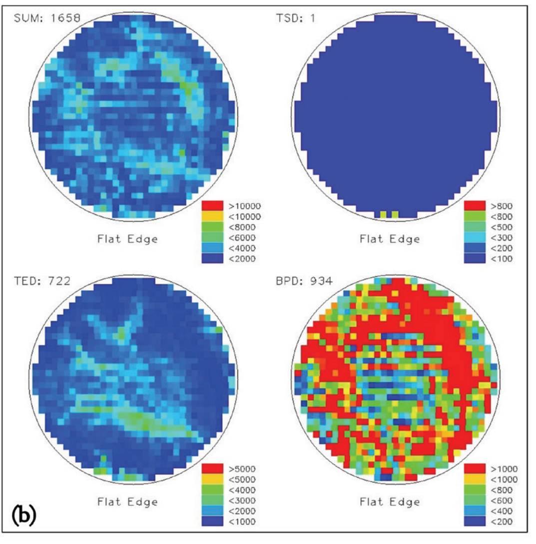


Figure 8. On-axis limits. (a) Dislocation map of an on-axis seed crystal. (b) After growth, threading dislocation density remains similar, though basal plane dislocations are suppressed.





➤ Figure 9. Off-axis advantage. (a) Starting seed with moderate defect density. (b) After growth on a 4° off-axis seed, threading dislocation density plummets as macrosteps convert them to basal plane defects that exit the crystal.

is intensifying. Magnifying this demand are: solar inverters and grid storage systems, operating at megawatt scales, that value incremental efficiency gains; and data centres, where power conversion losses account for a significant fraction of electricity consumption, and superior SiC trims operating costs and cuts carbon footprints.

Like many new technologies, while there's promise, challenges are still to be addressed. One is the growth rate, which remain slower than vapour transport, but the gap is narrowing. Another is scaling to 12-inch wafers – the next industry target – a goal that demands careful engineering of thermal fields and fluid dynamics; and there is also the requirement for recycling, to keep costs competitive. Yet the fundamental advantages are compelling: lower temperatures, lower defect densities, and lower material costs.

Through Changzhou Perfect Crystal Semiconductor, we are now commercialising our technology, offering 6-12-inch solution-growth furnaces and low-resistivity *p*-type substrates – a speciality that remains difficult for physical vapour transport.

The road ahead involves integrating our insights into automated production systems. Through the introduction of real-time monitoring of growth interfaces, AI-driven process control, and advanced seed preparation techniques, we will push dislocation densities to new lows. We anticipate commercial substrates with fewer than 500 dislocations per square centimetre – a quality level that today seems aspirational. For the power electronics industry, our message is clear. The SiC revolution is just beginning, with solution growth poised to deliver a hike in crystal quality that next-generation applications demand.

FURTHER READING

➤ T Narumi *et al*. "3C-, 4H-, and 6H-SiC crystal habitus and interfacial behaviours in high temperature Si-based solvents" Cryst Eng Comm **22** 3489-3496 (2020)

➤ P C Chen *et al*. "Defect inspection techniques in SiC" Nanoscale Res. Lett. **17** 30 (2022)

➤ S Kawanishi *et al*. "Availability of Cr-rich Cr-Si solvent for rapid solution growth of 4H-SiC". J. Cryst. Growth **549** 125877 (2020)

➤ K Hyun *et al*. "Effect of cobalt addition to Si-Cr solvent in top-seeded solution growth" Appl. Surf. Sci. **513** 145798 (2020)

➤ H Daikoku *et al*. "Solution growth on concave surface of 4H-SiC crystal". Cryst. Growth Des. **16** 1256 (2016)

➤ T Mitani *et al*. "Effect of aluminum addition on the surface step morphology of 4H-SiC grown from Si-Cr-C solution" J. Cryst. Growth **423** 45 (2015)

➤ N Komatsu *et al*. "Modification of the surface morphology of 4H-SiC by addition of Sn and Al in solution growth with Si-Cr solvents" J. Cryst. Growth **458** 37 (2017)

➤ M T Ha *et al*. "A review of the simulation studies on the bulk growth of silicon carbide single crystals" Korean Ceram. Soc. **59** 153 (2022)

➤ T Ujihara *et al*. "Crystal quality evaluation of 6H-SiC layers grown by solution phase epitaxy around micropipes using micro-Raman scattering spectroscopy" Mater. Sci. Forum **457/458/459/460** 633 (2004)

➤ T Ujihara *et al*. "Crystal quality of a 6H-SiC layer grown over macrodefects by solution-phase epitaxy: a Raman spectroscopic study" Thin Solid Films **476** 206-209 (2005)

Monolithic microLEDs eye AI

InGaN nanopyramids, produced with a single epitaxial step, promise to provide red, green and blue emission from one device

BY AMELIE DUSSAIGNE AND BERANGERE HYOT FROM CEA-LETI AND ADRIEN MICHON AND BENJAMIN DAMILANO FROM CÔTE D'AZUR UNIVERSITY/CRHEA/CNRS

ONE OF THE emerging, potentially lucrative applications for III-nitride materials is augmented reality (AR). When microLEDs are made from this material systems, they can deliver a high brightness at pixel pitches down to 10 µm and less. Thanks to this accomplished performance, they are compelling candidates for high-resolution micro-displays, which lie at the heart of AR technology.

Much work has already been devoted to devising the optimum approach for manufacturing high-resolution, full-colour micro-displays. These efforts have established three leading contenders: a pick and place process, combining blue and green emitters made from III-nitrides with phosphide materials for red emission; colour conversion, realised by depositing phosphors or quantum dots on top of blue microLEDs; and monolithic integration, with the same material family providing native emission of all three primary colours.

All three solutions are progressing. Those based on picking and placing microLEDs made from different materials, and on colour conversion, are yielding some encouraging results – but that employing monolithic integration is more attractive on a number of fronts, including cost, efficiency and resolution.

For the latter, as InGaN covers the entire visible spectrum by tuning the indium content of this alloy, it is, without doubt, the material of choice for monolithic integration. However, there are issues, predominantly an external quantum efficiency that varies across the spectral range. It is high in the blue, typically 60-80 percent, but falls at longer wavelengths, dropping to 30-40 percent in the green and 10 percent or less in the red. Behind this plummet in efficiency is a degradation in the crystal quality of InGaN, with deterioration beginning at an indium content of more than 15-20 percent. To put that figure in perspective, the indium content in a blue LED is around 15 percent, that for a green LED about 25 percent, and for red LED, depending on the growth orientation, 35-40 percent.

Due to this issue, it is clear that one of the challenges facing the development of full colour

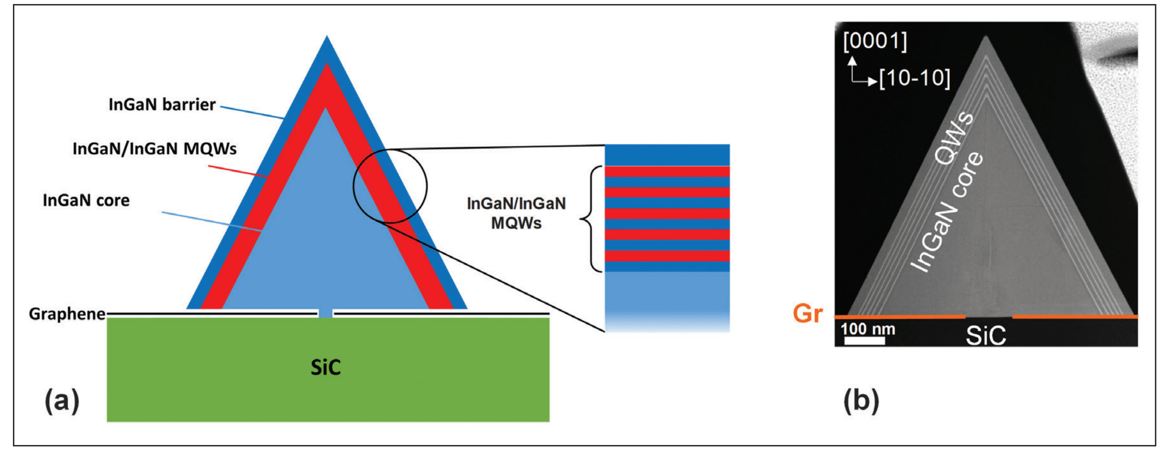


Figure 1. (a) The structure of a single InGaN nanopyramid. (b) Transmission electron microscopy image of this structure for a red-emitting nanopyramid (Gr = graphene layer).

➤ Figure 2. Photos of a 1 cm^2 sample under laser excitation, with emission from blue to red depending on the position of the laser.

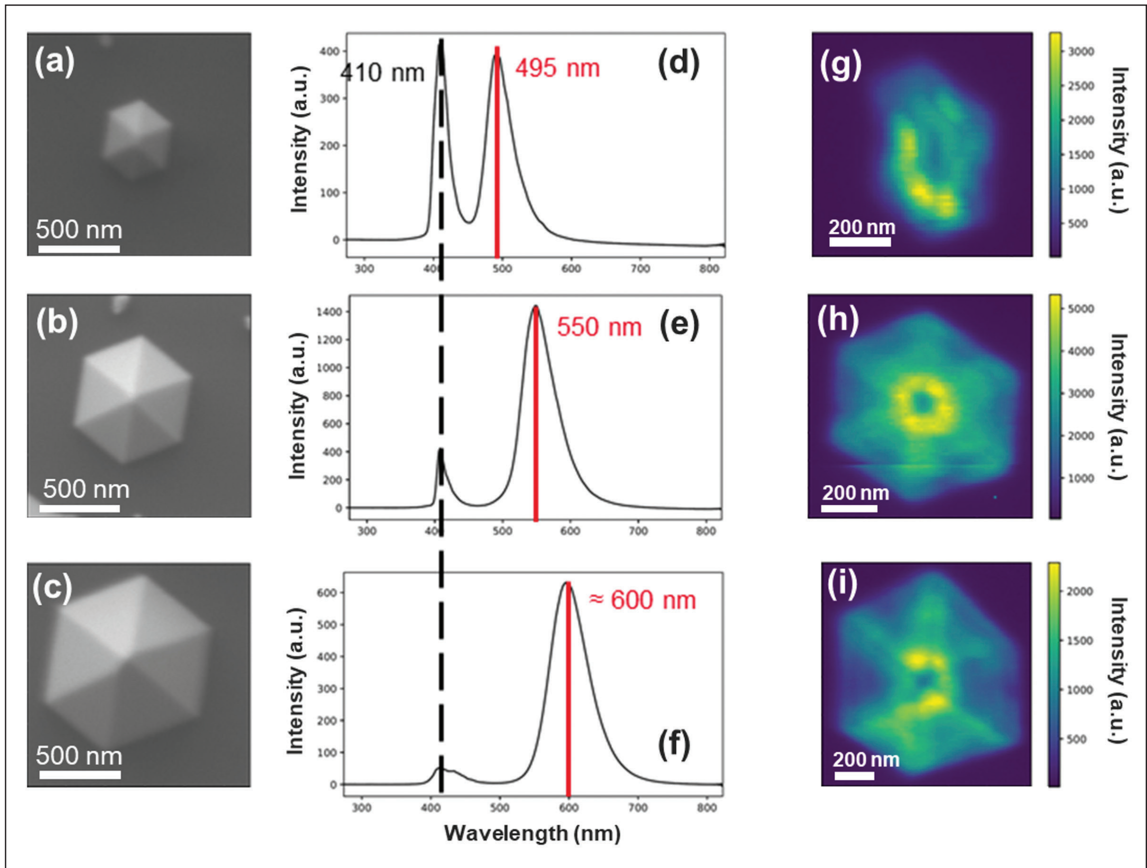
micro-displays that will employ efficient InGaN-based red, green, and blue microLEDs is the realisation of efficient InGaN-based red emission. What's more, if red, green and blue microLEDs are to serve in tomorrow's AR technology, where they provide high resolution, high efficiency and reduced cost, there also needs to be a technology that enables manufacture of three-colour sub-pixels, using a monolithic integration scheme.

At CEA-Leti and CRHEA/CNRS, we are developing an approach that addresses both these matters. Our solution, involving the growth of InGaN nanopyramids through a 2D material, enables three primary colours from one epitaxial growth, using the same substrate. With this approach, we have already demonstrated regular red, green and blue quantum wells with an indium content up to 45 percent.

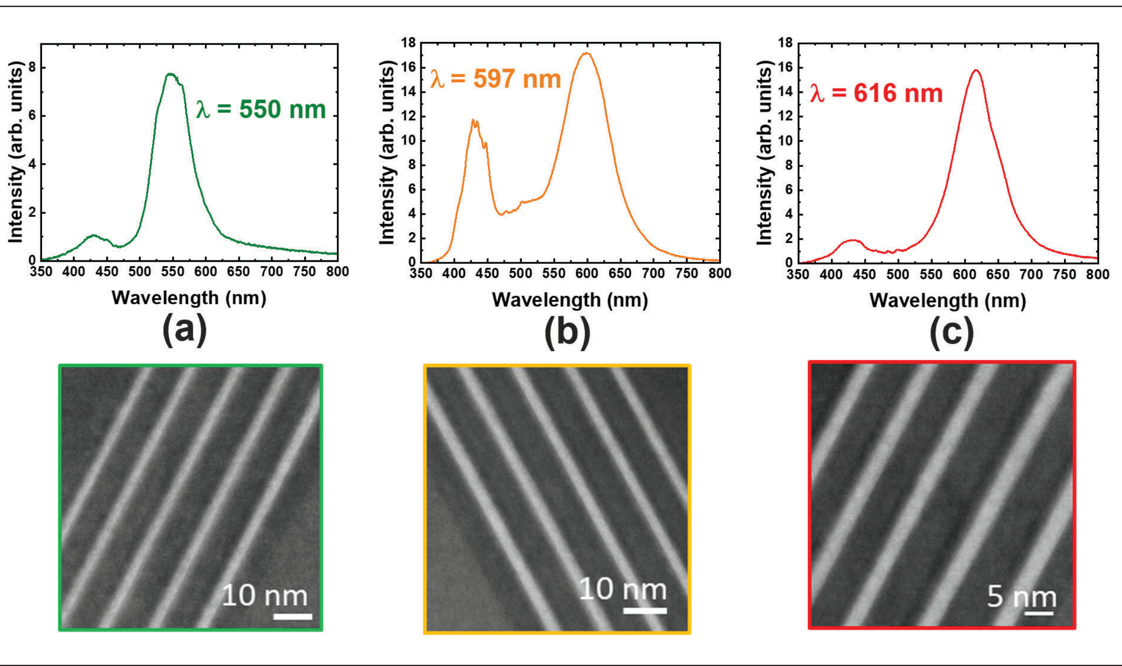
Growth of graphene

Thanks to the absence of dangling bonds on their surface, 2D materials constitute ideal masks for selective-area growth of InGaN nanopyramids. In our case, we employ a graphene mask. The graphene layer is first obtained by CVD on SiC in a hydrogen atmosphere. This method ensures self-limited growth of a single layer of graphene with a high structural quality.

If we were to attempt direct nucleation of InGaN on graphene, by either MBE or MOCVD, this would produce islands with random orientations and a lack



➤ Figure 3. (a)-(c) Scanning electron microscopy images of the studied single nanopyramid with a lateral size of 500 nm, 750 nm and 1000 nm, respectively. (d)-(f) Cathodoluminescence (CL) average spectra acquired on the same single nanopyramid, respectively (emission from the core (dotted black line) and emission from the quantum wells (QWs) (red line)). (g)-(i) CL integrated intensity mapping for QW emission range acquired on the same single nanopyramid, respectively.



➤ Figure 4. High-resolution transmission electron microscopy images performed on a single nanopyramid emitting in the (a) green, (b) orange, and (c) red range. Corresponding CL average spectra of the quantum well areas recorded on the same transmission electron microscopy lamellas.

of clear crystallographic facets. Such a foundation is incompatible with LED manufacturing. So, we locally modify graphene before InGaN growth, creating small openings around 50-100 nm in size by H_2/NH_3 high-temperature treatments. These apertures allow InGaN to nucleate on the SiC substrate, beneath the graphene layer. During InGaN growth, pyramids are formed with a well-oriented crystalline structure.

InGaN nanopyramids

Following the formation of our InGaN hexagonal nanopyramids by selective-area growth, using an *in situ* patterned graphene monolayer on SiC as an embedded mask, we add an InGaN active zone on the nanopyramid sidewalls. This light-emitting region is composed of InGaN/InGaN multiple-quantum wells.

We have scrutinised these structures with transmission electron microscopy. Images reveal an InGaN core directly grown on SiC through small apertures in the graphene layer, and multiple quantum wells on the sidewalls of the pyramids (see Figure 1).

When we use transmission electron microscopy to provide a wider view, we see that the InGaN nuclei appear preferentially organised along specific directions, corresponding to step edges of the SiC substrate. Since the graphene layer on these edges is more defective than on the terraces, the conditions are favourable for nucleation of the InGaN nanopyramids on the SiC substrate.

Red, green and blue

Under UV laser excitation, we observe emission spanning the blue to the red on different areas of our 1 cm^2 sample (see Figure 2). Red emission is centred at 645 nm at room temperature.

Crucially, different emission wavelengths are realised with a single epitaxial growth. This demonstrates it's possible to obtain red, green and blue emission via monolithic integration, using just one epitaxial growth.

Cathodoluminescence mappings on single nanopyramids (see Figure 3) reveals that emission wavelengths are correlated to the lateral sizes of the nanopyramids: cyan emission for small nanopyramids, around 500 nm in size; green-yellow emission for medium-sized nanopyramids, around 750 nm in size; and amber-red emission for larger pyramids, with dimensions of about 1 µm. We have also observed that emission is homogeneous on the facets and edges of the nanopyramids, with a higher intensity at the edges, due to superior light-extraction efficiency.

Another noteworthy observation is that differences in the nanopyramid lateral size are related to differences in nanopyramid densities – a small lateral size is accompanied by a high density, and a large lateral size occurs alongside a low density. This makes sense, given that during epitaxial growth, material is distributed in each area according to the available sidewall total surface, so there will be variations in the thicknesses and indium content in the InGaN wells and barriers of the active region. Note that at this stage, the graphene mask presents a self-organized hole pattern created during the initial *in-situ* annealing step.

To assess the crystalline quality and optical properties of different quantum-well regions, we extracted three transmission-electron-microscopy lamellae, taken from three different areas of the sample. On each lamella – they had lateral sizes of 500 nm, 750 nm and 1000 nm – we conducted transmission electron microscopy and

cathodoluminescence characterisation (see Figure 4). Results revealed that the wells are regular, with an average width from 3.5 nm to 4 nm, and homogeneous in terms of chemical contrast, which is a sign of a homogeneous indium distribution in the InGaN alloy.

To gain greater insight into our red-emitting nanopyramids, we obtained transmission electron microscopy images of the whole nanopyramid, and also zoomed in on the quantum-well area (see Figure 5).

In addition, we conducted cathodoluminescence wavelength mapping, finding red emission homogeneously distributed along the sidewalls (see Figure 5 (b)); and obtained an electron dispersive X-ray spectroscopy profile, taken from the core of the nanopyramid through the quantum wells. The latter revealed an indium content of 13 percent in the core, and 42-45 percent in the wells – that's a record indium content in the quantum-well area, realised in conjunction with homogeneity in this region.

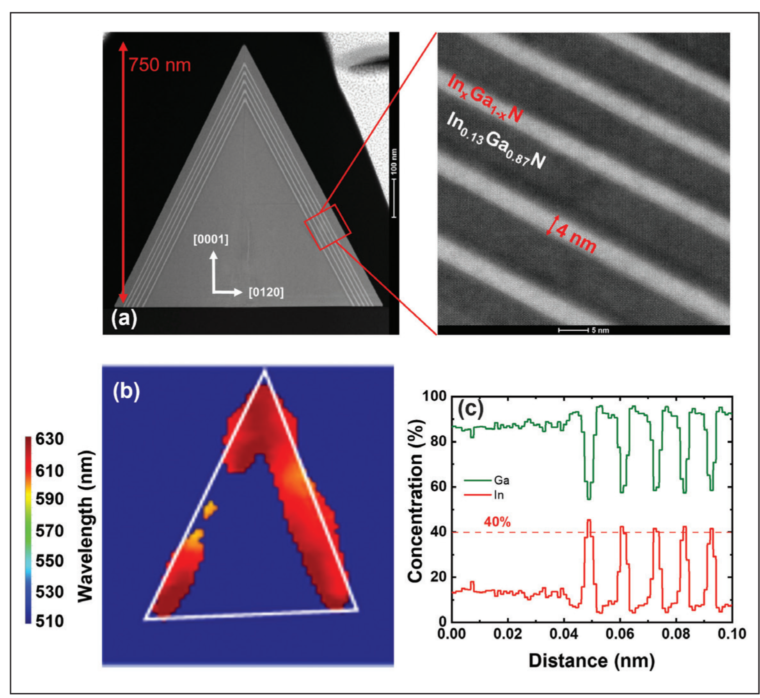
According to strain mapping, the core of our nano-pyramids is at least nearly relaxed, from base to apex. The implication is that a relaxed InGaN pseudo-substrate, in our case the pyramid core with a 13 percent indium content, helps enhance the indium incorporation rate in the InGaN alloy without degrading crystal quality. The advantage of the graphene mask, which benefits from the presence of the Van der Waals gap, is that it provides almost free-standing nanopyramids – they are linked to the SiC substrate by just a hole that's 80-100 nm in diameter, with a height of 2 monolayers.

Our efforts show that it's possible to produce a light-emitting structure on the same substrate, with just a single growth run, that provides red, green and blue emission. Additional key findings are that our approach can produce relaxed InGaN pseudo-substrates that enhance indium incorporation in the InGaN alloy beyond the theoretical limit of 25 percent when grown on GaN, a feat that's accomplished without degrading crystalline quality.

One of our next goals is to define an organized hole pattern in the graphene mask, so that we can form red, green and blue sub-pixels next to each other, by playing on the nanopyramid diameters and densities.

We also plan to advance our process integration of our nanoLEDs, by transfer of the nanopyramids to a conductive substrate. The electrical contacts will be taken from the back side for the *n* contact, and from the nanopyramid sidewalls for the *p* contact.

To conclude, our combination of a graphene mask and a full InGaN structure is an attractive solution to realising red, green and blue emission from a single epitaxial growth while maintaining high crystalline quality in the active area. This is even realised with the red-emitting structure, a breakthrough that promises high efficiency for all three InGaN-based primary colours.



➤ Figure 5. (a) Transmission electron microscopy (TEM) images of the red emitting whole nanopyramid and of the quantum well (QW) area. (b) Cathodoluminescence wavelength mapping on the same TEM lamella. (c) Electron dispersive X-ray spectroscopy profile, starting from the core through the QW area.

Part of this research was supported by the "Recherche Technologique de Base" and "France 2030 - ANR-22-PEEL-0014" programmes of the French National Research Agency (ANR).

The authors wish to thank the contributions provided by C. Paillet, N. Rochat, D. Cooper, and A. Grenier from CEA-Leti, and S. Vezian from CRHEA/CNRS. Note that C. Paillet is now at Aixtron.

FURTHER READING

- A. Dussaigne *et al.* "Regular red-green-blue InGaN quantum wells with In content up to 40% grown on InGaN nanopyramids" Commun. Mater. **5** 280 (2024)
- C. Mastropasqua *et al.* "Self-limited monolayer graphene growth on SiC with propane-hydrogen CVD" npj 2D Mater. Appl. **9** 32 (2025)
- C. Paillet *et al.* "InGaN islands and thin films grown on epitaxial graphene," Nanotech. **31** 405601 (2020)

AI and the future of epitaxy

The challenge of applying AI to semiconductor manufacturing is not technological, but structural – and ultimately human

BY FAEBIAN BASTIMAN AND ISABELLA LORENTE FROM BIZMUTH MBE

A YEAR AGO, we started tackling a problem that many within our industry have been mulling over, but few addressing directly: how to bring AI into epitaxy in a manner that actually delivers.

At that time, we had a plan, along with a strong point of view and an exceptional partner in the Paul-Drude-Institute for Solid State Electronics, Leibniz-Institute in Forschungsverbund Berlin e.V. It's a laboratory that's been performing at the cutting edge of MBE research since 1992. What we did not yet have was proof that our approach would work.

Over the past year, that's changed. Working directly on operational systems under real-world conditions, we have tested our ideas where they matter the most: on the tools, with the data, and within day-to-day workflows.

Our original aim was a modest one: to determine whether it's possible to integrate AI into the process in a way that provides consistent, measurable benefit. But our ambitions did not stay modest for long.

A real-world problem

For our work, we chose GaN deliberately. It's central to modern devices – LEDs, lasers, and power and RF electronics – it's widely studied, and it is still difficult to handle in practice. When combined with MBE, a precise but unforgiving growth technique, we had a combination that provided an ideal proving ground for our AI ambitions.

UnicornOne with MBE Lab Blur IR Camera.

Our aim was clear. Optimise the GaN growth window on one system, then replicate it on others. On paper it's a simple objective. And if we achieved it, we could naturally extend our approach beyond MBE to other deposition techniques, including CVD, ALD, and PVD.

With hindsight, the real difficulty was never going to be GaN or MBE. It was something more intangible.

GaN is hard, but it's not unknowable. The physics is largely understood. Far less defined is how to apply that knowledge. Much of the process remains tacit, embedded in the experience of the operator, rather than captured in the system.

Another impediment to introducing AI is that the systems offer little support. Data is fragmented, units are inconsistent, and tools operate in isolation. While information exists, it is difficult to align, interpret, or reuse in a meaningful way.

This results in a self-reinforcing problem loop: GaN MBE remains difficult, because the knowledge is not fully captured. And that's the case, because systems are not designed to hold it. Instead, the process continues to rely on experienced users, bridging the gaps.

Our solution breaks this cycle. We have built a system that's capable of integrating data, workflows and AI into a coherent whole. It's an approach that breaks new ground, tackling the problem with full visibility of the process and its limitations.

Success has come from having a clear idea of what's there, and equally important, what's missing. Regarding the latter, there's nothing new. For decades, our industry has been held back by omissions quietly undermining progress across systems, workflows, and data.

Integration is key

In a world of vendor-locked software, fragmented data and isolated tools, AI cannot operate effectively. Under this modus operandi, AI only analyses individual datasets, and cannot participate in the process itself.

In the months leading up to the work with PDI, and with this constraint in mind, we started developing our new software and LabOS 'UnicornOne'. We knew that if AI was to play a meaningful role, integration could not be an afterthought. It had to be the foundation. This meant that we would have to unify hardware, data, and workflows into a single, coherent system – one where humans and AI could observe, reason, and act.

As well as providing insight, our collaboration offers proof. Working directly on operational systems, we began by confirming the scale of the problem. Prior to our efforts, legacy tools, modern instruments, and external data sources existed side by side, but rarely in communication. Data was present, but not structured. Control was possible, but not unified. Addressing these issues, we constructed a unified control layer, UnicornOne, sitting across these systems. Rather than replacing them, it joins the dots, creating a consistent interface that captures all actions and measurements in a structured, time-resolved form.

Building the AI system

With a strong foundation in place, we embarked upon the next challenge, data.

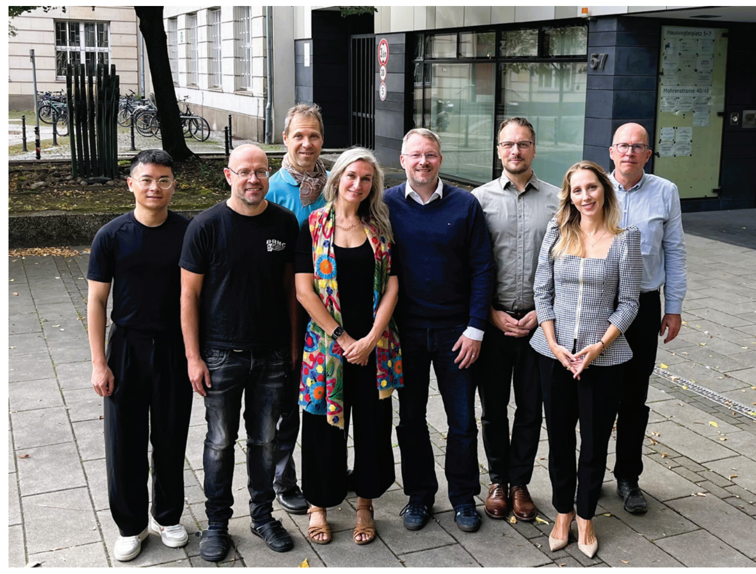
While real experimental data is valuable, it is inherently slow to produce and difficult to scale. A single iteration may take hours, and exploring a parameter space exhaustively is rarely practical in a laboratory setting.

So, rather than waiting for more data to accumulate, we took a different approach, focused on generating it.

Starting from a small, annotated dataset, featuring Reflection High-Energy Electron Diffraction (RHEED) images and videos, we developed a simulation framework capable of reproducing the key features of the growth process under controlled conditions. This real data provided the anchor, with simulation ensuring rapid expansion, allowing datasets to be extended, refined, and reused at speed.

From this point on, progression became relatively straightforward. We combined real and synthetic data into coherent datasets, which supported model development without the need for extensive experimental repetition.

Models naturally followed from this foundation. They were lightweight and targeted, trained not in isolation, but part of a wider system that's designed to reflect how an engineer would approach the same problem. In situations where physics is well understood, it's applied directly; and where interpretation is required, AI provides support. This leads to a hybrid approach, deterministic where possible, stochastic only where necessary.



➤ Project UnicornOne members from PDI and Bizmuth.

The result is not a single AI model. Instead, it's a multi-model system, combining stochastic and deterministic elements, integrated within a software environment supporting human and machine decision-making. To accomplish this, we divided the growth process into a series of discrete subtasks – each one interpretable, executable, and aligned with how a researcher would approach the same problem.

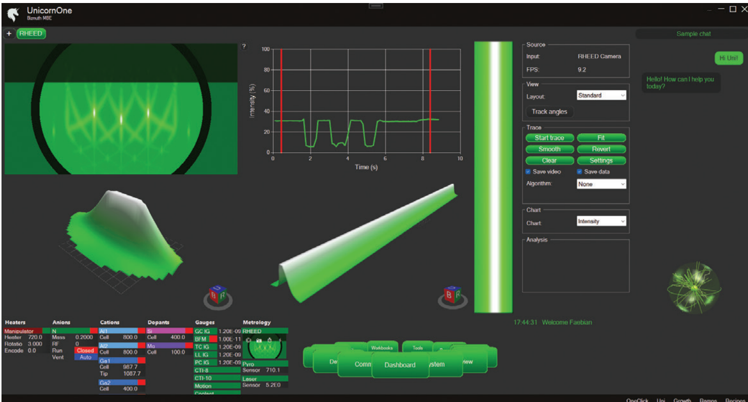
With our multi-model system in place, determining and reproducing the GaN growth window was no longer a theoretical question. It had become operational.

Real world deployments

At this stage, we shifted our focus to proving that the system could be deployed quickly and reliably. Thanks to our vendor-agnostic hardware abstraction layer, we could interface with existing tools without modification, reducing the installation time from weeks to days.

When embarking on this task, long-hidden issues surfaced – configuration inconsistencies, communication faults, and components operating outside intended conditions. These problems were caused by fragmented software and limited visibility, and addressing them became part of the deployment. As we worked through the problems, systems were not just integrated, but improved.

After establishing the interface, we deployed the workflow. Despite emerging real-world constraints, including electrical noise affecting RHEED signals, we had a smooth transition. As the process had already been developed and validated in simulation, and grounded in real conditions, little adaptation



Unicorn One's build in RHEED analysis tool.

was required when applying it to the physical system.

The system, workflows, and subtasks have operated as intended, delivering an immediate impact. A GaN growth calibration, previously requiring around two and a half hours of manual effort, can now be completed in about thirty minutes.

We then deployed the same workflow on a second system, made by a different vendor and featuring a different configuration. Another accomplishment, with successful execution on the first run.

This was the key moment. It demonstrated that, with consistent physical units and a sufficiently abstracted control layer, it's possible to transfer workflows between systems without modification. In effect, the process had become portable.

By now our efforts were paying dividends. Based on Bizmuth's calculations, time savings alone were translating into a trimming of annual operating costs by around €55,000 per system for this single workflow, with reproducibility built in by design.

Compounding interest

As the system matured, its role expanded. Researchers at PDI started developing their own workflows, referred to as 'workbooks', to address specific tasks. They included tools for mapping parameter spaces, calibrating nitrogen flux based on mass flow and plasma conditions, and replaying historical data to generate new training datasets.

Unicorn One's GaN temperature calibration AI/ML workbook.

Further extension to this platform has come from support for Python-based workflows and external APIs. Users have been given the opportunity to integrate custom scripts, connect external tools, and build more complex pipelines on top of the core system.

What began as a solution to a specific problem has now evolved into a more general environment for conducting and managing experiments.

One of the more interesting outcomes of our development has been its effect on how researchers approach their work. For routine automated tasks, such as calibration and monitoring, time has been created for observation and analysis. Now subtle variations in RHEED signals, possibly previously overlooked, are the basis for new lines of investigation. With this superior way of working, the system is not replacing the researcher, but changing the nature of their engagement with the experiment.

By combining structured workflows, historical data, and large language models, it's now possible to interpret high-level intent and translate it directly into execution. Users no longer define every step. Instead, they define the outcome – and the system plans, selects, and acts.

With this way of working in place, AI has moved on from simply helping to solve a problem to taking the reins and growing a whole sample. That's possible, thanks to our strong foundation: integrated systems, structured data, documented processes.

While we are still at an early stage, our approach demonstrates the huge potential for more autonomous operation. Complex tasks can be coordinated across multiple components, with feedback loops ensuring that the process remains aligned with the intended outcome.

The broader reality

Alongside many breakthroughs, the experience at PDI has highlighted a critical limitation: adoption has not been uniform across all users and systems. While some embrace the platform and extend it in new directions, others are more cautious.

This state of affairs is not that surprising. It reflects a broader reality within the semiconductor industry, which is that barriers to adopting AI are rarely technical. Playing a major role are established workflows, individual incentives, and concerns regarding transparency. For some, systems that make processes more explicit and reproducible change how expertise is perceived and shared, even valued.

Judged against backdrop, the challenge of applying AI to semiconductor manufacturing is as much

human as it is technological. The foundations are already in place, as the tools required to build AI-enabled systems already exist. The greater challenge lies in integrating these tools into a coherent framework, and aligning that framework with the needs and expectations of its users.

Our efforts at PDI suggest that meaningful progress is possible. Focusing on integration, rather than isolated optimisation, we have created systems where AI operates effectively. Capturing data with sufficient context allows processes to be reproducible and transparent; and by providing extensible platforms, it's possible to support standard workflows and innovation.

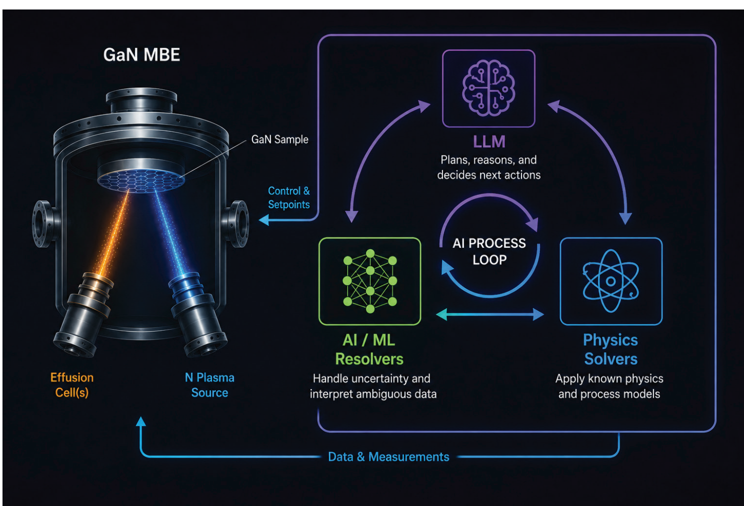
Looking ahead

The approach we have developed extends naturally beyond individual tools. By coordinating systems, integrating *in-situ* and *ex-situ* data, and aligning workflows, our methodology can be adopted by facilities, enabling them to operate as coherent units, rather than collections of independent tools.

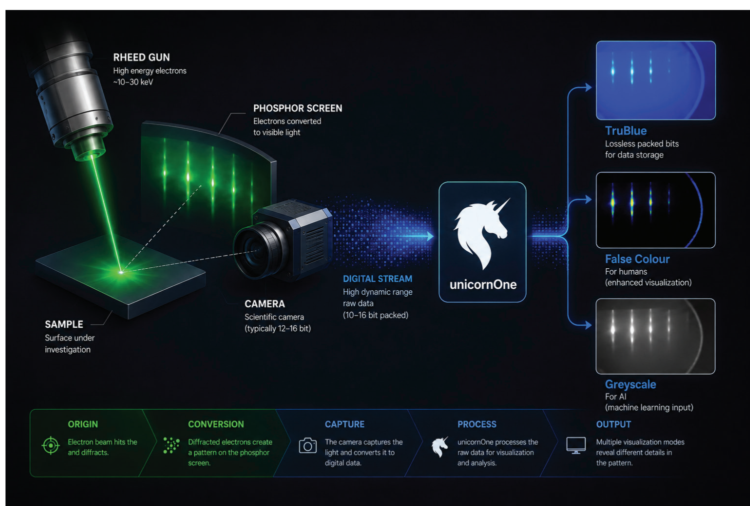
It's a transition that's already underway at PDI, an institution pleased with progress, and we have agreed to expand the collaboration, but we cannot yet announce any details yet. What began with GaN on a single MBE system has expanded to now include oxides, complex materials, and traditional III-V processes, with multiple tools brought under a unified control framework.

Beyond MBE, we are applying our software to PVD and CVD through collaborations with equipment manufacturers, with the aim of establishing a consistent control layer across thin-film technologies.

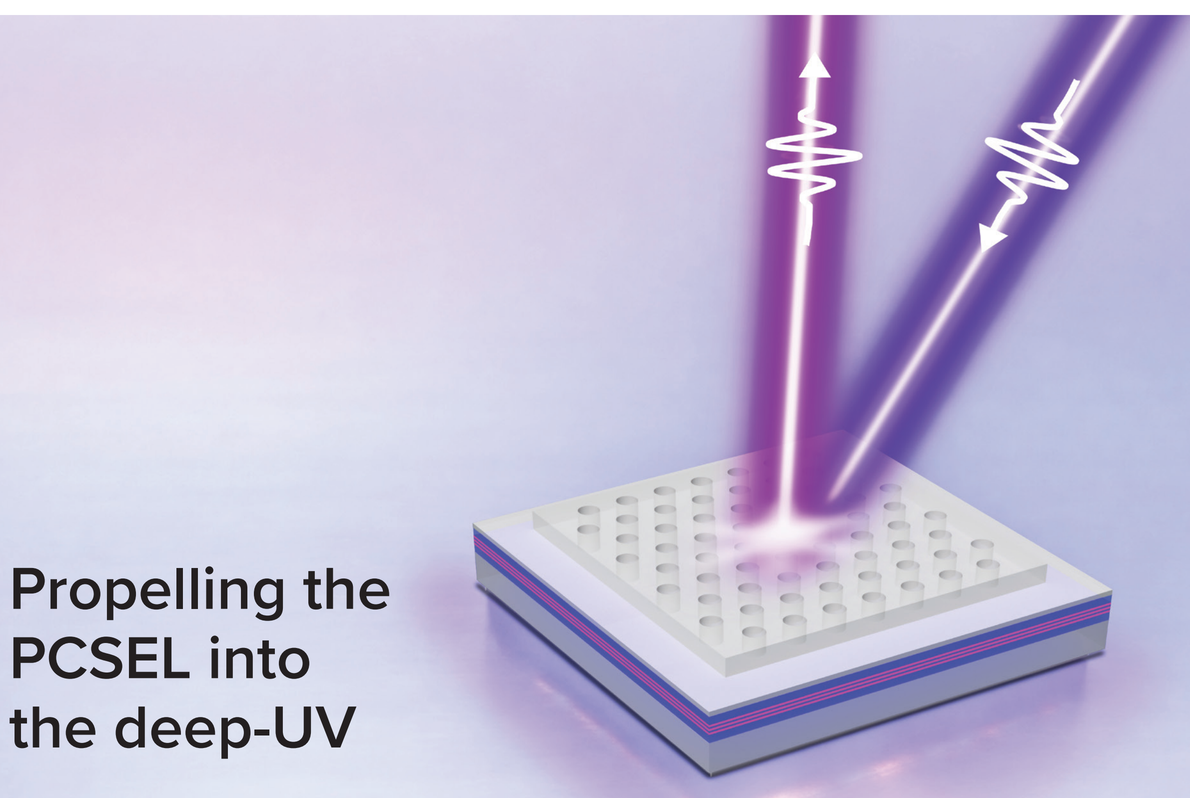
The conclusion is clear. AI is not the limiting factor. What determines its impact is the infrastructure into which it is deployed, and the degree to which systems, data, and workflows are aligned. The future of semiconductor manufacturing will not be defined by AI alone, but by the systems that enable it – and by those prepared to adopt them.



UnicornOne's multi AI mesh and physics solvers.



UnicornOne's adaptive RHEED image pipeline for storage, humans and AI.



Propelling the PCSEL into the deep-UV

With edge-emitters and VCSELs stalling, is the PCSEL emerging as the best candidate for delivering laser emission deep into the UV?

BY DOĞUKAN APAYDIN FROM CHALMERS UNIVERSITY OF TECHNOLOGY

SINCE the birth of the semiconductor laser in the early 1960s, effort has been directed at expanding its spectral range. Much progress has been made at ever longer wavelengths, with success stemming from the introduction of a new architecture, the quantum cascade laser. In contrast, stretching to shorter wavelengths has been far less productive, with the greatest progress over the last few decades involving the introduction of a new material system: GaN and its related alloys, particularly AlGaN. However, while these nitrides undoubtedly offer the best chance of success, maybe they need to be combined with an alternative design, so they can excel when plunging deep into the depths of the UV?

The AlGaN challenge

The III-N family spans the visible to the UV, with shorter wavelengths reached by increasing the aluminium content of AlGaN. All optoelectronic devices made from this material share weaknesses, including a high defect density, low electrical and thermal conductivity, low refractive index contrast and low material gain.

As these issues are yet to been fully solved, despite growing demand for diode lasers throughout the UV, many applications are relying on excimer lasers and frequency-converted solid-state lasers, sources that are far from ideal, being large, expensive and not power efficient. As semiconductor lasers are small, compact and low-cost, if they could perform in the UV, they would be compelling candidates for spectroscopy, lithography and biomedical imaging – and possibly a number of new applications. However, today's devices emitting in the UV fall far short of the performance of their visible cousins, and closing this gap is challenging.

But there are glimmers of hope. In late 2022, a collaboration between Hiroshi Amano's team at Nagoya University and the AlN substrate producer Asahi Kasei Corporation reported room-temperature, continuous-wave lasing at 274 nm from an edge-emitting laser grown on a bulk AlN substrate. Key to this success is combatting low *p*-type conductivity with a distributed polarisation-doped layer featuring a gradual change in material composition. Unlike impurity doping, which yields very low hole concentrations at room temperature, due to the high activation energy of magnesium, distributed polarisation doping reaches higher hole concentrations. By introducing distributed polarisation-doping, it is possible to produce current-injection deep-UV lasers. Building on this success, Ultec, a spin-out of this Japanese startup, introduced the first commercial deep-UV laser diode.

Other noteworthy successes include great progress in the UVB, where, very recently, Meijo University demonstrated the first continuous-wave lasing. And in just the last month or so, Zetian Mi's team from the University of Michigan announced a 299 nm laser produced by MBE. Driven in pulsed mode, this device has a very low threshold current density.

Within the nitride community, efforts have also been directed at VCSELs, which have a number of merits over edge-emitters, including being better suited to high-volume manufacture. However, it is far harder to make a VCSEL from AlGaN than AlGaAs. All VCSELs feature a pair of highly reflective mirrors made from distributed Bragg reflectors. With a GaAs VCSEL, the pairing of GaAs and AlAs is well suited to this task – but there's no good equivalent for the nitrides, due to a lack of candidates that combine an adequate refractive index contrast with lattice matching. Turning to dielectric mirrors addresses this issue, but demands substrate removal, and that's not easy to do. Note that to date, in the UV, lasing has only been demonstrated under pulsed optical pumping.

PCSELs: A potential solution

While the performance of edge-emitters and VCSELs has not plateaued, of all the approaches for a III-N UV source, the photonic crystal surface-emitting laser (PCSEL) has the greatest promise. This class of laser employs a two-dimensional (2D) photonic crystal for in-plane optical feedback. What makes it particularly compelling is its large gain area, enabling high output powers to be combined with single-mode operation. Devices deliver a circularly symmetric beam with a low divergence angle, usually less than 1°.

The inventor of the PCSEL is Susumu Noda from Kyoto University, Japan. In 1999, along with his co-workers, he demonstrated the first PCSEL at infrared wavelengths. Since then, Noda's team have made significant strides, including, in 2008, GaN-based PCSELs delivering blue lasing with current injection. This group has also demonstrated the first green PCSEL, and a PCSEL producing an output of hundreds of watts.

Where Noda has led, a growing number are following. There have now been demonstrations by several groups of PCSELs based on the III-nitride material system emitting in the blue (see Figure 1).

For PCSELs operating at visible wavelengths, performance is now starting to catch up with their infrared counterparts. However, in the UV, the pace has been much slower.

First deep-UV PCSELs

To date, there has been a limited number of reports of lasing in the UV, with none reaching the deep-UV regime until recently.

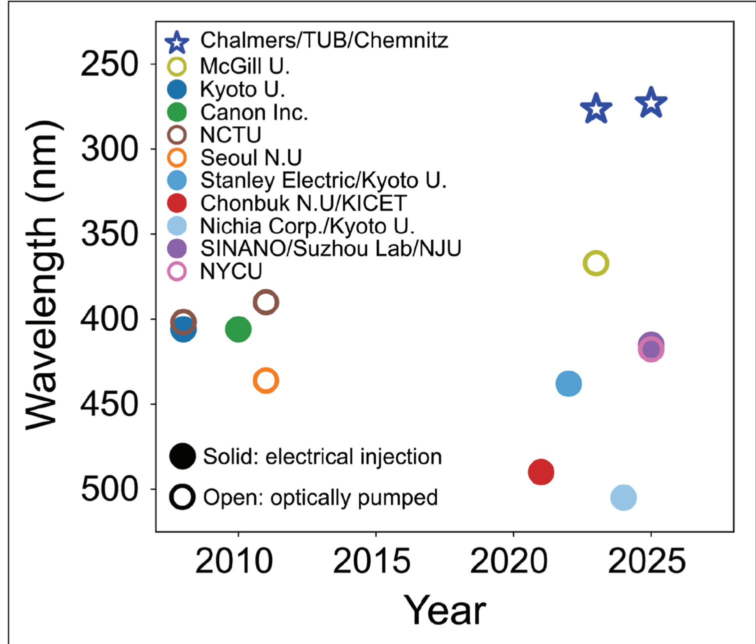
Helping to drive the latest progress is our team from Chalmers University of Technology, Sweden. Our important milestone towards shorter wavelengths has been the demonstration of the first deep-UV PCSELs under pulsed optical pumping at room temperature. Thanks to this, PCSELs now cover a larger region of this domain.

Growth of our AlGaN epitaxial stack on a sapphire substrate has been undertaken by our collaborators at the Technical University of Berlin (see Figure 2). They have provided us with material with an estimated defect density of 5 x 10^8 cm^{-2} – much higher than that for GaAs-based devices, which have defect densities typically ranging from 10^4 cm^{-2} to 10^6 cm^{-2}.

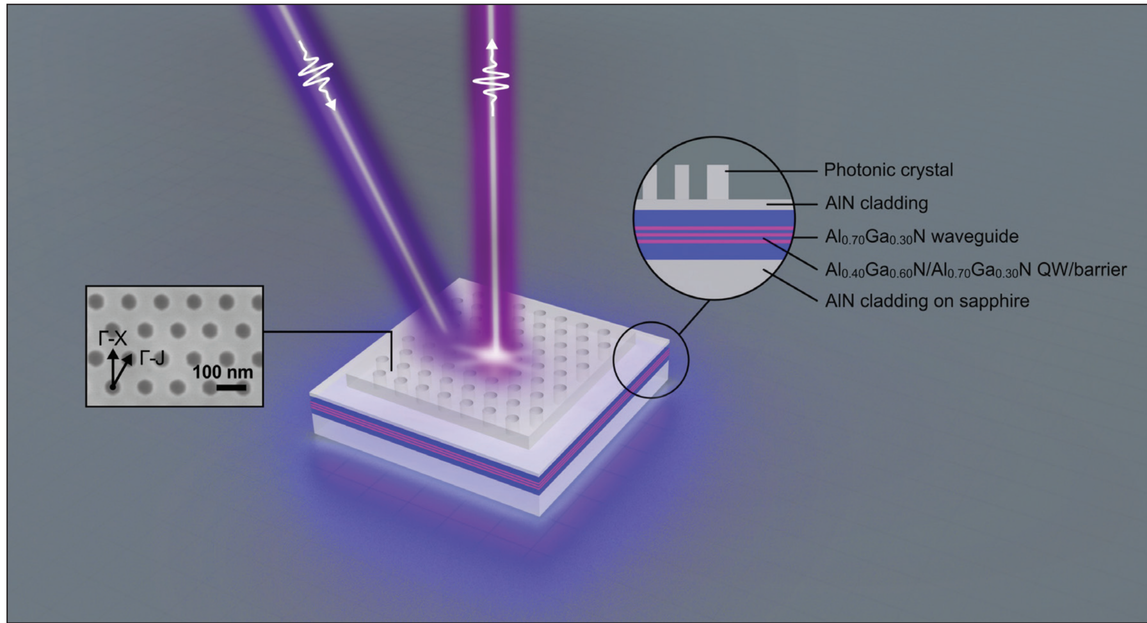
Device fabrication began by spinning a high-resolution negative electron-beam resist, HSQ, on the epiwafers, prior to electron-beam exposure and development. When exposed or annealed, HSQ turns into SiO_x, creating a hard mask for etching. But that's not all – this oxide helps avoid imperfections introduced by pattern-transfer methods, such as lift-off. That's highly valued when producing devices for the deep-UV, because feature size scales with wavelength. After exposure, we directly transfer 140-µm-wide photonic crystals with a lattice constant of 140 nm into the top AlN cladding, using Cl_2/Ar dry etching. A buffered oxide etchant removes the remaining resist.

Most producers of PCSELs encapsulate the photonic crystal in a subsequent regrowth step. We take a different approach, bringing both benefits and trade-offs. Our regrowth-free approach simplifies the fabrication process, which is expensive and time consuming, but poses the risk of introducing plasma-induced damage to the quantum wells, due to deeper photonic crystal etching.

One of the key requirements for the PCSEL is that the optical mode in the waveguide has sufficient



➤ Figure 1. Emission wavelengths of III-nitride PCSELs that have been demonstrated by various groups over the years.



➤ Figure 2. Illustration of the pump beam and the surface emission in a deep-UV PCSEL, with insets showing a topview SEM of the hexagonal photonic-crystal pattern and the AlGaN epitaxial stack.

confinement from the photonic crystal to be modulated by the periodicity. This is accomplished by ensuring that the photonic crystal is sufficiently close to the quantum wells. If we fail on that front with an etch that's too shallow, the lasing threshold is then too high, due to insufficient confinement. But go too far the other way, with an etch that's too aggressive or too close to the quantum wells, and then there's the threat of degraded quantum wells, increasing threshold, and in the worst case, preventing the device from lasing.

By going regrowth-free, we are adopting an approach that's applied to existing edge-emitting laser structures that are proven to lase under current injection. In terms of epitaxial structure and fabrication, our regrowth-free PCSELs are not that different to edge-emitting lasers, with a waveguide between two claddings and a photonic crystal. What this means is that it should be possible to easily integrate a 2D photonic crystal and realise electrically-driven PCSELs in the deep-UV.

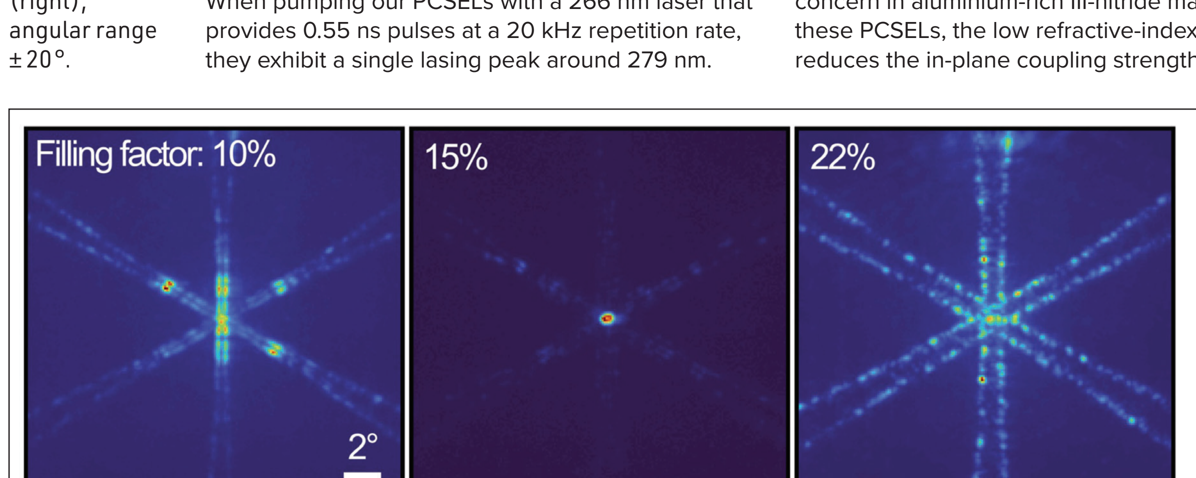
When pumping our PCSELs with a 266 nm laser that provides 0.55 ns pulses at a 20 kHz repetition rate, they exhibit a single lasing peak around 279 nm.

Initial demonstrations produce lasing thresholds of around 17 MW cm^{-2} at room temperature, so significantly higher than optically pumped deep-UV edge-emitting lasers, which have values generally around 1 MW cm^{-2} at similar wavelengths. But that gap is closing, with our most recent devices having thresholds around 5 MW cm^{-2}, thanks to optimised photonic-crystal fabrication. Note that the pump beam is absorbed only by the quantum wells.

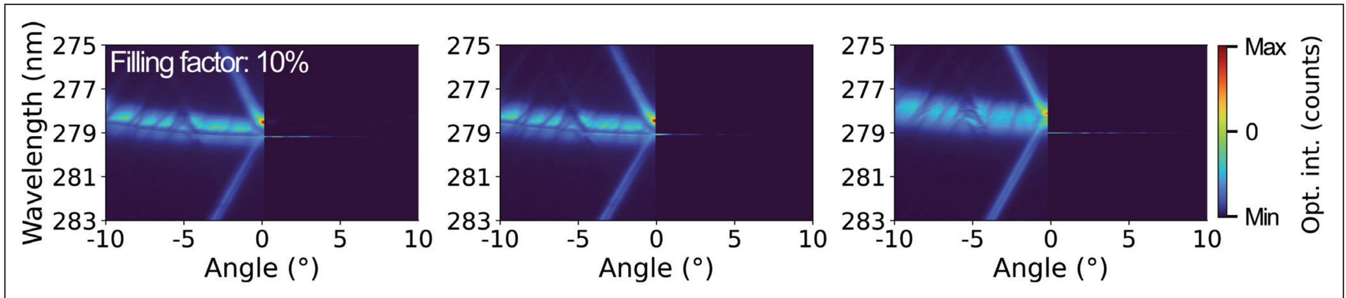
Strengthening 2D coupling

Unlike a distributed feedback laser, a PCSEL does not couple light in just one direction. Instead, it is coupled in 2D, essentially leading to a 2D standing wave. However, it's not easy to realise sufficient 2D coupling in a square lattice with circular holes, because the wavevector must change direction by 90°. If the 2D coupling is weak, oscillation tends to occur mostly in 1D, causing the 2D photonic crystal to act as a 1D grating. In this case, a highly divergent beam results.

This undesirable emission profile is a particular concern in aluminium-rich III-nitride materials. For these PCSELs, the low refractive-index contrast reduces the in-plane coupling strength. Another



➤ Figure 3. Measured PCSEL far-field emission patterns for filling factors of 10 percent (left), 15 percent (center) and 22 percent (right); angular range ±20°.



➤ Figure 4. Measured PCSEL band structures for filling factors of 10 percent (left), 15 percent (center), and 22 percent (right). The left half of each figure shows the below threshold band structure, the right half shows the above threshold band structure.

issue is that the short wavelength, and therefore the small feature size, makes the device more sensitive to surface roughness and fabrication imperfections. Taken together, these drawbacks make it particularly challenging to realise robust 2D operation.

Moving to a hexagonal lattice is a solution. As 2D coupling occurs at 60° and 120°, it is easier to obtain sufficient 2D coupling strength with a hexagonal lattice. We have pursued this design, using a hexagonal lattice with circular holes. This architecture has helped increase symmetry and decrease the threshold.

If a PCSEL with circular holes were to have an infinite photonic crystal, there would be no surface emission, since the fields are perfectly symmetric and the modes are symmetry-protected. But real photonic crystals have a finite size. Thanks to this, the electric field is perturbed by the edge of the photonic crystal and symmetry is broken, enabling vertical outcoupling.

However, as the device size increases, wall-plug efficiency falls. This occurs because the rate of increase in the device area exceeds the rate of increase in the area that light can outcouple. One option for overcoming this problem, and enhancing outcoupling, is to introduce asymmetric hole shapes that intentionally break the symmetry and enhance vertical outcoupling. More recently, some groups have adopted a double-lattice square lattice, an architecture that increases modal discrimination while maintaining sufficient 2D coupling strength.

This has enabled high-power single-mode operation in the infrared and blue. However, transferring this approach to the deep-UV is challenging, because a single-lattice square has a smaller lattice constant than a hexagonal lattice for the same operation wavelength – for example, 138 nm for a hexagonal lattice, and 119.5 nm for a square lattice – and it's far from easy to fit a double hole in the unit cell.

Ensuring low divergence

For PCSELs, an important design parameter is the size of the air holes. This is often defined as the filling factor: the ratio of the air hole area to the unit cell area.

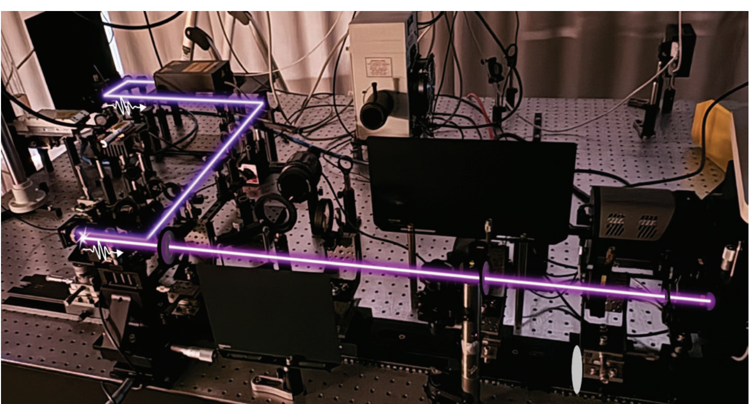
For small filling factors, such as 10 percent or so and associated with small hole sizes, the 2D coupling strength is weak and 1D oscillations dominate emission. This leads to highly divergent far-field emission patterns, extending up to 20° (see Figure 3).

Increasing the filling factor to 15 percent boosts the 2D coupling strength, and delivers benefits associated with the 2D periodicity of the photonic crystal. This fill factor leads to a donut-shaped emission pattern with a low divergence angle.

Going to even larger filling factors of around 20 percent or more ensures high 2D coupling. However, this also results in a peculiar divergent emission pattern that is an exact copy of the cross-section of the band structure at the lasing wavelength. As filling factors increase, they cause the device to lase at a different bandedge mode. In our band structure measurements, we have observed these changes in the lasing band edge mode (see Figure 4). This has led us to conclude that selecting the optimal filling factor is extremely important in deep-UV PCSELs.

Next: current injection

Current injection is always the ultimate goal for any semiconductor laser. We are working towards this, having shown that the PCSEL concept can be extended into the deep-UV using aluminium-rich III-nitrides, despite their low refractive index contrast



➤ Figure 5. Angleresolved photoluminescence setup used for deep-UV PCSEL characterisation, with illustrations of the 266 nm pump beam and the PCSEL emission.

and the small feature sizes required at these wavelengths. As technology matures, we expect deep-UV PCSELs to operate under current injection, just like their infrared and visible counterparts. However, success will require addressing several questions, including how to spread the current efficiently across a large photonic crystal area.

With our PCSELs, while the epitaxial layers are similar to those of an edge-emitting laser, the current spreading requirements are very different. In edge-emitters, the current only has to be spread over a lateral distance of the order of the ridge width; but in PCSELs, this distance can range from a few tens of micrometres to millimetres. For blue and infrared PCSELs, one option is to use conductive substrates to aid lateral spreading. Unfortunately, that's not possible with deep-UV PCSELs, due to the lack of conductive substrates. One solution is substrate removal, but that does not guarantee current uniformity, which could impact lasing performance, for example, by promoting higher-order modes or preventing the formation of a 2D standing field.

Solving these issues will be challenging, but the path is clear. There's good reason to believe that thanks to progress in materials and photonic crystal design, it will not be long before we have the first electrically driven deep-UV PCSEL.

FURTHER READING

- D. Apaydın *et al*. "Deep UV Photonic Crystal Surface Emitting Lasers" Laser Photonics Rev. e00271 (2025)
- Emoto *et al*. "Wide-bandgap GaN-based watt-class photonic-crystal lasers" Commun Mater. **3** 72 (2022)
- K. Yoshida *et al*. "High brightness scalable continuous wave single mode photonic crystal laser." Nature **618** 727–732 (2023)
- N. Taguchi *et al*. "Green-wavelength GaN-based photonic-crystal surface-emitting lasers." Appl. Phys. Express **17** 012002 (2024)
- H. Amano *et al*. "The 2020 UV emitter roadmap." J. Phys. D: Appl. Phys. **53** 503001 (2020)
- Z. Zhang *et al*. "Continuous-wave lasing of AlGaN-based ultraviolet laser diode at 274.8 nm by current injection." Appl. Phys. Express **15** 041007 (2022)
- R. Miyake *et al*. "Room-temperature continuous-wave lasing at 318 nm on a relaxed AlGaN template grown on a sapphire substrate." Appl. Phys. Lett. **128** 023502 (2026)

Combatting compositional pulling in red InGaN microLEDs

Growing quantum wells at different temperatures improves indium uniformity in the active region of red-emitting InGaN LEDs

THROUGH the setting of new standards for spectral purity and minimal wavelength shift, a Chinese collaboration is claiming to have improved the performance of InGaN red-emitting microLEDs.

This breakthrough – revealed by a partnership between Xiamen University, the Future Display Institute of Xiamen, and Southern University of Science and Technology – will aid the development of displays based on III-N microLEDs, which are held back by a strong and stable source of red emission.

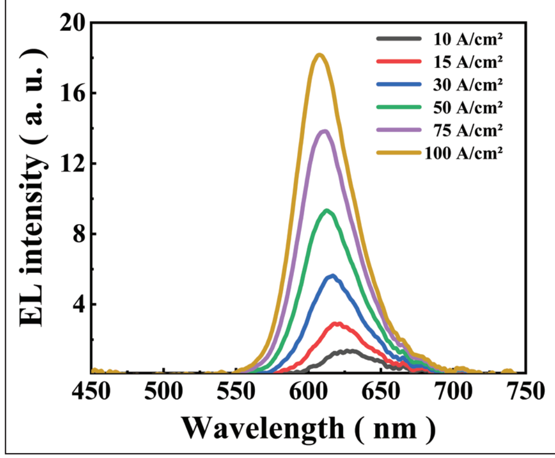
Another potential material systems for producing red microLEDs is AlGaInP, the quaternary sitting at the heart of LEDs deployed in many applications. However, as the dimensions of this form of red LED shrink, its external quantum efficiency (EQE) plummets. For sizes below 5 µm, EQE is less than 1 percent, primarily due to Shockley-Read-Hall non-radiative recombination stemming from sidewall damage.

A more promising alternative is the InGaN microLED. Scaling the dimensions of this device, which also offer superior thermal robustness, is more successful if proper sidewall passivation is applied. What's more, as blue and green InGaN LEDs are well-established, there's the potential to make full-colour displays from a single material system, eliminating material incompatibilities and simplifying display fabrication.

But there are issues with red InGaN microLEDs. They require an indium content that exceeds 35 percent, and due to an 11 percent lattice mismatch between InN and GaN, there's significant compressive strain in the quantum wells. This strain degrades crystal quality and introduces strong piezoelectric fields that pull apart electrons and holes and hamper radiative recombination. Also at play is phase separation and compositional inhomogeneity, broadening emission from the LED and degrading its performance.

While the team from China are by no means the first to try and address the poor EQE of red LEDs, they have broken new ground for realising spectrally pure, single-peak red emission by combatting compositional pulling – it's caused by compressive strain in the quantum wells, and leads to vertical inhomogeneity of the indium composition in the multi-quantum well region. When compositional pulling is not addressed, spinodal decomposition may result, creating a yellow or green shoulder in the emission profile of a red LED. This impacts spectral purity and colour stability, metrics that matter in display technologies.

To minimise compositional pulling, the team adjusts the growth temperature during deposition of the active region, with the first two quantum wells grown at a temperature 5°C higher than the three remaining quantum wells. Note that team spokesman Hongjian Li, an affiliate of the Future Display Institute of Xiamen and Xiamen University, told *Compound Semiconductor* that a similar result may be realised by varying the flow of trimethylindium into the MOCVD chamber.



➤ Thanks to improved uniformity from well to well, InGaN microLEDs produce a single red emission peak. Another strength is the relatively low shift in emission wavelength with varying drive current.

To quantify the consequences of compositional pulling effects, Li and co-workers employed cross-sectional cathodoluminescence to scrutinise a control sample, with all quantum wells grown at the same temperature. This technique revealed a distinct blue-shift of 15-25 nm between upper and lower quantum wells.

Further evidence for variations in indium content came from energy-dispersive X-ray spectroscopy, which determined a difference in indium content between the first and fifth quantum wells of the control of 12 percent. For the structure produced with different quantum well growth temperatures, indium content varies by just 4 percent.

The superior uniformity enabled the team's 15 µm by 15 µm LEDs to produce a single red emission peak over a wide range of drive current densities, ranging from 1 A cm^{-2} to 100 A cm^{-2}. Driven at 5.2 A cm^{-2}, the researchers device delivers an EQE of 7.9 percent.

Li says that the next goal for the team is to increase the EQE of its InGaN red microLEDs and deploy them in full-colour displays with an ultra-high pixel density.

REFERENCE

➤ C. Liu *et al*. Appl. Phys. Lett. **128** 191103 (2026)

Giving ß-Ga_2O_3 an insulating buffer

To combat an unwanted silicon peak, buffer growth provides an attractive alternative to etching the substrate in hydrofluoric acid

RESEARCHERS from the University of California, Santa Barbara (UCSB), are claiming to have developed a superior process for addressing an unintentional silicon peak in ß-Ga_2O_3 heterostructures.

This peak, occurring at the interface between ß-Ga_2O_3 substrates and homoepitaxial layers, degrades device performance in a number of ways. One downside of the silicon peak is that it creates a low-mobility electron channel further from the gate terminal that adds to delays associated with resistance and capacitance, and ultimately hampers fast switching and high-frequency performance in lateral devices. In addition, the peak leads to a higher pinch-off voltage and an increased subthreshold slope; and it can increase leakage and lead to premature breakdown.

Back in 2023, a collaboration between UCSB and the University of Utah reported that it's possible to remove the silicon peak by dipping the substrate in 49 percent hydrofluoric acid prior to growth. However, this process is time-sensitive, with the peak returning in around 10 minutes under ambient exposure.

Now, Sriram Krishnamoorthy and co-workers are arguing that they have developed a superior process, involving growth, by MOCVD, of an insulating buffer that compensates the interfacial charge.

According to the team, the benefits of the insulating layer extend beyond charge compensation to include aiding the performance of vertical and lateral FETs. It is said that the addition of the buffer offers an advantage in vertical devices, as it allows the inclusion of wider fins in vertical enhancement-mode devices, by creating a potential barrier in the fins.

The insulating buffer could act as a current-blocking layer, using acceptors such as nitrogen, magnesium or iron. A major merit of nitrogen is its deep acceptor level, 2.9 eV below the conduction band, that reduces the likelihood of inadvertent activation. The acceptor level for iron is not as deep; and while that's not an issue for magnesium, its weakness is that it lingers in the reactor, and is unintentionally incorporated in epilayers in subsequent growth runs.

To investigate the impact of a nitrogen-doped buffer, Krishnamoorthy and co-workers have grown a number of samples in an Agnitron Agilis 1000 vertical cold-wall reactor. Prior to growth, iron-doped (010) ß-Ga_2O_3 substrates were cleaned in acetone, methanol and deionised water. Omitting the etch in hydrofluoric acid allowed the team to see if the nitrogen-doped layer compensated for the silicon peak.

The researchers grew the nitrogen-doped layer at 910°C, as higher temperatures reduce the level of hydrogen, which is predicted to cause compensation of nitrogen and act as a shallow donor. What's more, experiments show that the incorporation of hydrogen roughens the surface of ß-Ga_2O_3 epilayers.

After growing the nitrogen-doped layer, the team shut off ammonia flow and added a 150 nm-thick silicon-doped layer with an intended doping of 5×10^{17} cm^{-3}.

When growing the nitrogen-doped buffer layer, it's crucial to hit a sweet spot. Too much nitrogen degrades the crystal quality of the buffer layer, and may increase unintentional impurities in the silicon-doped channel, if there's surface segregation or nitrogen diffusion into active channel layers. But if nitrogen flow is insufficient, there will be incomplete compensation of unintentional silicon impurities.

To establish optimal flow, the team produced samples with a range of ammonia/nitrogen flows.

Capacitance-voltage measurements on these samples determined that an ammonia flow rate of 1.8×10^{-6} mol min^{-1} provided partial compensation, while flows of 2.7×10^{-6} mol min^{-1} and 4.0×10^{-6} mol min^{-1} both showed no carrier spike at the interface, indicating that the concentration of nitrogen in the buffer layer exceeds that of silicon at the interface.

To verifying full compensation, the team also produced buffer leakage structures. Measurements determined that as ammonia flow increased, so did resistance.

To quantify the concentration of nitrogen in these structures, the team produced a test structure, based on alternating layers of unintentionally doped Ga_2O_3 and nitrogen-doped Ga_2O_3 at various flow rates. According to secondary ion mass spectrometry, nitrogen levels in layers with flows of 2.7×10^{-6} mol min^{-1} and 4.0×10^{-6} mol min^{-1} are 1.7×10^{19} atoms cm^{-3}. As nitrogen is not thought to saturate at this level, the constant value is attributed to growth variation.

Based on these results, the team has determined an optimum flow of 2.7×10^{-6} mol min^{-1}. They have also investigated the optimum buffer thickness, which is 50 nm, according to plots of capacitance versus voltage.

REFERENCE

➤ R. Kahler *et al*. Appl. Phys. Lett. **128** 20213 (2026)

Scrutinising 50 mm AlN substrates

AlN substrates feature local variations in defect density and buried aluminium platelets

THE RECENT availability of AlN substrates with diameters of up to 100 mm could be a game-changer for a number of compound semiconductor devices. This foundation provides a native platform for UVC LEDs and for power electronics based on AlN. The latter has the potential to outperform rivals based on SiC and GaN, thanks to a much wider bandgap and a higher critical electric field. However, the AlN substrate will only provide a bedrock for success on these fronts if it is of sufficient material quality, and can ensure the growth of excellent epilayers.

To investigate if these crucial conditions are met, a US team, led by researchers at the US Naval Research Laboratory (NRL) and including engineers from Crystal IS and Agnitron Technology, has been scrutinising AlN substrates with a diameter of 50 mm, and AlN epilayers grown on this foundation.

Offering a summary of the findings, team spokesman Alan Jacobs from NRL told *Compound Semiconductor*: “The wafer-scale changes in photoluminescence point toward varying point-defect populations, while the X-ray topography and high-resolution diffraction mapping show there can be off-cut variation or local dislocation density variations, potentially impacting growth, impurity incorporation, and device characteristics.”

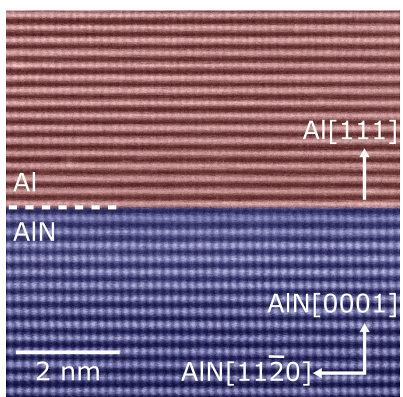
The team’s comprehensive study involved Crystal IS AlN substrates with a surface orientation of ± 3°, an absorptivity of less than 30 cm^{-1} at 265 nm, and a surface bow of less than ± 10 µm.

Atomic force microscope of as-received AlN substrates determined a root-mean-square roughness of about 0.1 nm, and the presence of clear atomic steps, with edges aligned parallel to the flat of the wafer. The team also found a number of depressions, with a depth of around 1 nm, diameters ranging from 100 nm to 400 nm, and curved step edges at the circumference.

After chemical cleaning of the substrates, they were loaded into an MOCVD reactor and annealed at 1250°C in ammonia to clean and stabilise the surface, prior to the addition of an unintentionally doped, 1 µm-thick layer of AlN at 1250°C.

Compared to the AlN substrate, its epilayer has a slightly higher macro-roughness, but is free from small depressions. However, there are surface perturbations, with co-located peaks and valleys, having a height deviation of 0.5 nm and a spatial extent of 1.5-2 µm. These features may result from overgrowth of the initially observed depressions.

When inspecting AlN material by optical microscopy, the researchers observed bright features, described as spatially inhomogeneous in the radial direction. To uncover their origin, cross-sections were formed with a focused ion beam and imaged with various forms of microscopy. Electron microscopy revealed that these bright features are associated with platelet regions parallel to the wafer surface. The platelets are mainly aluminium, according to energy-dispersive X-ray spectroscopy. This finding is supported by high-angle annular dark-field scanning transmission electron microscopy, which shows that the atomic spacing within the platelet is essentially equal to that for bulk aluminium.



➤ The platelets have an atomic spacing that is essentially equal to that for bulk aluminium, according to high-angle annular dark-field scanning transmission electron microscopy.

“We were 100 percent surprised by the metallic inclusions,” admits Jacobs, who initially worried about their impact on the performance of power devices, but is now far less concerned, due to their low volume fraction, small size and isolated nature.

“Perhaps there will be some noticeable RF or dynamic switching losses associated with these defects, however, that remains to be seen, and given the sparse nature – and distribution through the full thickness – they may not be near enough devices to have an impact.”

X-ray topography has been employed to determine the dislocation density, which is significantly lower than that of GaN. It’s around 10^2 cm^{-2} in the centre of the wafer, and as high as mid-10^5 cm^{-2} near the edge.

The team plans to continue to characterise AlN substrates and epilayers, with investigations hopefully extending to 4-inch wafers.

“We are also focusing our efforts toward devices built on these substrates, including AlGaN/AlGaN HEMTs for high power and high temperature operation, vertical device topologies for high voltage and high-power scaling, and potential novel devices enabled by the unique characteristics of AlN,” said Jacobs.

REFERENCE
➤ A. G. Jacobs *et al*. Appl. Phys. Lett. **128** 202102 (2026)

CORPORATE PARTNERS DIRECTORY

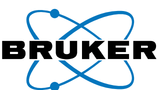
Tel: +1 952 937 7505
Email: sales@agnitron.com
www.agnitron.com

Tel: +49 2407 9030 0
Fax: +49 2407 9030 40
Email: info@aixtron.com
www.aixtron.com

Tel: +1 (408)-429-2774
Email: apc-e@luminus.com
www.luminus.com/products/APC-SiC-MOSFETs

Tel: +001 610-647-8744
info@appliedenergysystems.com
www.appliedenergysystems.com

Tel: 1 610-647-8744
purification@appliedenergysystems.com
www.armpurification.com



Tel: +44 (0)191 332 4700
Fax: +44 (0)191 332 4800
sales.semi@bruker.com
www.bruker.com

Tel No: +49 (89) 96 24 00-0
Fax No: +49 (89) 96 24 00-122
Email: sales@csclean.com
www.csclean.com

www.mbe-komponenten.de

Tel: +886 4 2534-7800
Fax: +886 4 2534-8869
Email: sales@efab.com.tw
www.efab.com.tw

Tel: + 41 81 403 8000
Fax: + 41 81 403 8001
www.evatecnet.com

Driven by Innovation

www.freiberginstruments.com

PHOTON IS OUR BUSINESS

www.hamamatsu.com

Tel: +1 732 494 8660
joanne.lowy@horiba.com
www.horiba.com/scientific

Tel: +49 8728 911 093
Fax: +49 8728 911 156
Email: sales@35reclaim.com
www.35reclaim.com

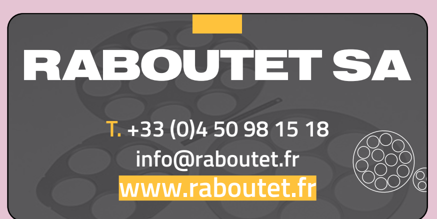
k-Space Associates, Inc.
Email: requestinfo@k-space.com
www.k-space.com

www.malvernpanalytical.com

https://masimosemiconductor.com

Tel: + 1 203 949 8697
Email: info@nelhydrogen.com
www.nelhydrogen.com

https://innovation.us.nipponsanso.com



www.riber.com

Tel: +1 925 245 5817
Fax: +1 925 449 4096
Email: lwarren@ferrotec.com
www.temescal.net

Tel: +82 10 9671 9675
Email: oesales@hites.co.kr
www.www.hites.co.kr

Tel: +31 24 350 0809
info@trymax-semiconductor.com
www.trymax-semiconductor.com

All Industrial All The Time
Tel: +001 805-368-6653
www.thericogroup.com

Tel: +49 (0) 3641 79980
Fax: +49 (0) 3641 7998222
www.vistec-semi.com

Showcase to the Compound Semiconductor industry for 12 months

COMPOUND SEMICONDUCTOR MAGAZINE is the primary information resource for professionals working with compound semiconductor materials and devices.

Each issue of the magazine is distributed in print and digitally to over 60,000 professionals worldwide.

With the largest and most diverse global audience of qualified buyers served by any compound semiconductor focused magazine, secure a corporate partnership and buyers guide entry and receive 12 months of promotion and exposure to the compound semiconductor industry across the magazine portfolio including:

- 12 months of promotion in nine issues of the print and digital versions featuring your company logo and company details
- 12 months of promotion and exposure on the Compound Semiconductor website compoundsemiconductor.net featuring a Buyers Guide entry and hyperlinked company logo
- Sponsorship of one e-news alerts sent out to over 60,000 industry professionals with text and a standard web banner.

Ranjodh Avern
T: +44 (0)2476 718970
E: ranjodh.avern@angelbc.com

Discover the Hidden Treasure in Your Lift-Off Coating Process!

Up to 40% Deposition Efficiency

With gold at an all-time high, optimizing the collection of gold during thin film deposition is more crucial than ever for cost-effectiveness and operational efficiency in semiconductor manufacturing.

Ferrotec's Coating Platform with patented HULA system provides an alternative lift-off
compliant wafer motion to traditional box coaters. By providing planetary motion to the wafers during deposition, HULA eliminates the need for masks and enhances the collection efficiency and uniformity of the deposited material.

Visit: www.temescal.net/auratus-gold
Contact: thinfilmsolutions@ferrotec.com

DIG UP THE SAVINGS!

	Conventional Box Coater	Ferrotec UEFC-5700
Performance		
Wafers Per Batch	20	42
Batches Per Day	15	20
Wafers Per Day	300	840
Operational Cost of Unreclaimed Gold*		
Cost of gold perf wafer	$19.28	$9.60
Cost of 100,000 wafers	1,928,000	$960,000

*Assuming $110/gram and reclaim cost of 80%

© 2013 Ferrotec (USA) Corporation. All rights reserved.