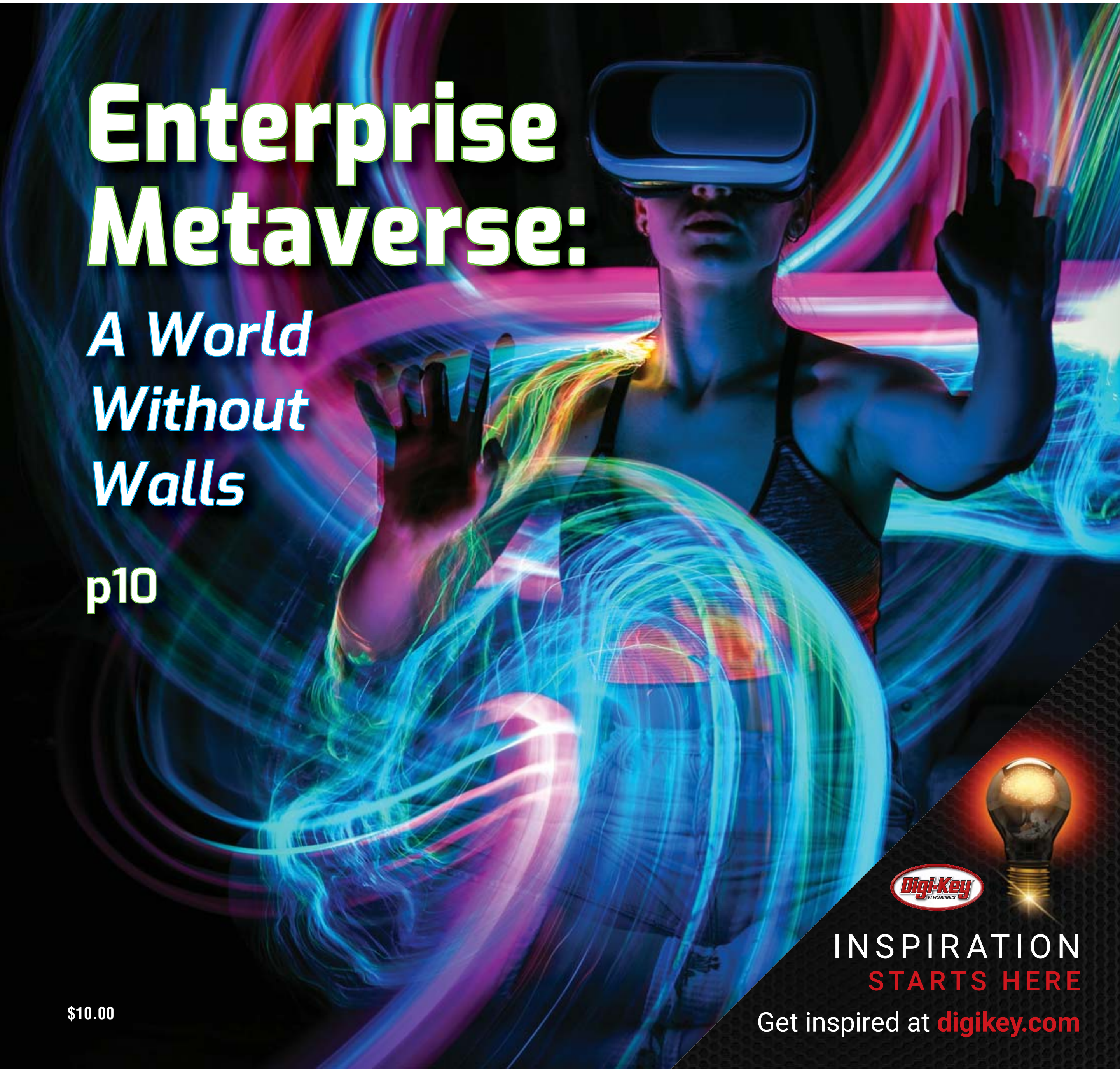
Electronic Design®

70 YEARS

THE AUTHORITY ON EMERGING TECHNOLOGIES FOR DESIGN SOLUTIONS

SEPTEMBER/OCTOBER 2022
electronicdesign.com



Enterprise Metaverse:

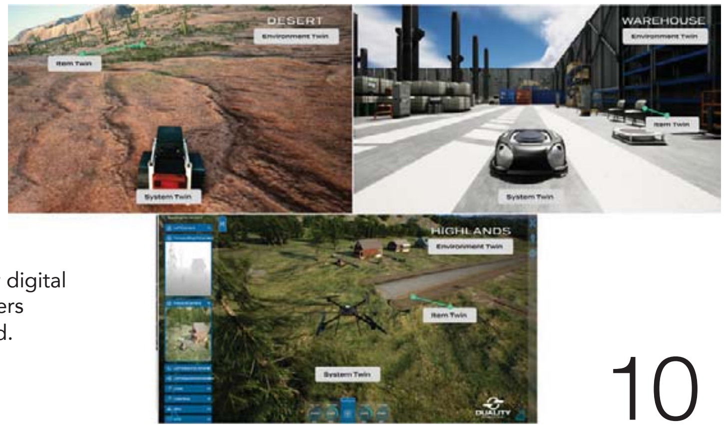
A World Without Walls

p10

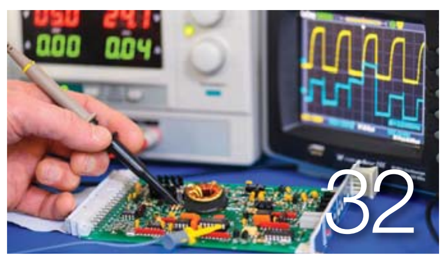
$10.00

IN THIS ISSUE

FEATURES



10



COLUMNS & DEPARTMENTS

FOLLOW US ON FACEBOOK AND TWITTER

(ElectronicDesign)

(@ElectronicDesgn)

EDITORIAL MISSION:
To provide the most current, accurate, and in-depth technical coverage of the key emerging technologies that engineers need to design tomorrow's products today.

ELECTRONIC DESIGN ISSN 0013-4872 (Print) / 1944-9550 (Digital) is published bi-monthly in Jan/Feb, Mar/Apr, May/June, July/Aug, Sept/Oct and Nov/Dec by Endeavor Business Media, 1233 Janesville Ave., Fort Atkinson, WI 53538. Paid rates for a one-year subscription are as follows: $60 U.S., $90 Canada, $120 International. Periodicals postage paid at Fort Atkinson, WI and additional mailing offices. Printed in U.S.A. Title registered in U.S. Patent Office. For subscriber services or to order single copies, write to Electronic Design, PO Box 3257, Northbrook, IL 60065-3257. POSTMASTER: Send change of address to Electronic Design, PO Box 3257, Northbrook, IL 60065-3257. Canadian GST #R126431964.


Altech®
Broad Selection of Quality
Printed Circuit Board
Terminal Blocks
FIXED PCB TERMINALS
SCREW TERMINAL BLOCKS
Wire Guard Type
Lift Type
Excenter Type
SPRING TERMINAL BLOCKS
Push-In Type
Tension Spring Type
Lift Type
PLUGGABLE MULTI-CONNECTOR
SCREW TERMINAL BLOCKS
Wire Guard Type
Lift Type
Excenter Type
SPRING TERMINAL BLOCKS
Push-In Type
Tension Spring Type
PIN STRIPS
State of the Art
Test & Measurement
• Plugs • Sockets
• Alligator Clips
• Test Probes
• Test Probes
• Oscilloscope Probes
• Measuring Leads
• Sets
Altech Corp.®
Your Source for Automation & Control Components
info@altechcorp.com | 908.806.9400 | AltechCorp.com

ElectronicDesign 70 YEARS

SEPTEMBER/OCTOBER 2022

EDITORIAL
Senior Content Director: **Bill Wong**
bwong@endeavorb2b.com

Senior Editor: **David Maliniak**
dmaliniak@endeavorb2b.com

Managing Editor: **Roger Engelke**
rengelke@endeavorb2b.com

Editor-at-Large: **Alix Paultre**
apaultre@endeavorb2b.com

Senior Editor: **James Morra**
jmorra@endeavorb2b.com

CONTRIBUTING EDITORS
Louis E. Frenzel
Lee Goldberg
Bill Schweber
Cabe Atwell
Murray Slovick
Steve Taranovich

ART DEPARTMENT
Group Design Director: **Anthony Vitolo**
tvitolo@endeavorb2b.com

Art Director: **Jocelyn Hartzog**
jhartzog@endeavorb2b.com

PRODUCTION
Production Manager: **Brenda Wiley**
bwiley@endeavorb2b.com

Ad Services Manager: **Deanna O'Byrne**
dobyrne@endeavorb2b.com

AUDIENCE MARKETING
User Marketing Manager: **Debbie Brady**
dmbrady@endeavorb2b.com

Article Reprints: *reprints@endeavorb2b.com*

SUBSCRIPTION SERVICES
OMEDA
847-559-7598 or 877-382-9187
electronicdesign@omeda.com

LIST RENTAL
List Rentals/Smartreach Client Services Manager: **Mary Ralicki**
mralicki@endeavorb2b.com

SALES & MARKETING
Gregory Montgomery
gmontgomery@endeavorb2b.com
AZ, NM, TX

Jamie Allen
jallen@endeavorb2b.com
AL, AR, SOUTHERN CA, CO, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MI, MN, MO, MS, MT, NC, ND, NE, OH, OK, SC, SD, TN, UT, VA, WI, WV, WY, CENTRAL CANADA

Elizabeth Eldridge
eeldridge@endeavorb2b.com
CT, DE, MA, MD, ME, NH, NJ, NY, PA, RI, VT, EASTERN CANADA

Stuart Bowen
sbowen@endeavorb2b.com
AK, NORTHERN CA, NV, OR, WA, WESTERN CANADA

AUSTRIA, BELGIUM, FRANCE, GERMANY, LUXEMBURG, NETHERLANDS, PORTUGAL, SCANDINAVIA, SPAIN, SWITZERLAND, UNITED KINGDOM

Diego Casiraghi
diego@casiraghi-adv.com
ITALY

Helen Lai
helen@twoway-com.com
PAN-ASIA

Charles Liu
liu@twoway-com.com
PAN-ASIA

DIGITAL
VP Digital Innovation Data: **Ryan Malec**
rmalec@endeavorb2b.com

ENDEAVOR BUSINESS MEDIA

ENDEAVOR BUSINESS MEDIA, LLC
331 54th Ave N., Nashville, TN 37209 USA
www.endeavorbusinessmedia.com

CEO: **Chris Ferrell**

President: **June Griffin**

CFO: **Mark Zadell**

COO: **Patrick Rains**

Chief Administrative and Legal Officer: **Tracy Kane**

EVP Special Projects: **Kristine Russell**

EVP Marketing Solutions & Data: **Jacquie Niemiec**

DESIGN & ENGINEERING GROUP
EVP, Design & Engineering Group: **Tracy Smith**
tsmith@endeavorb2b.com

Group Content Director: **Michelle Kopier**
mkopier@endeavorb2b.com

Electronic Design, Machine Design, Microwaves & RF, Power & Motion, SourceESB, Source Today, 3DX

Editorial

WILLIAM WONG | Senior Content Director

Your Metaverse or Mine?

Where will you be collaborating in the future? It will likely depend on your depth of knowledge of these new worlds.

238853596 © Pop Nukoonrat | Dreamstime.com

REMEMBER CYBERSPACE? It's now called the metaverse. It can be used to look through a new house or building—virtually. Or to remotely design and debug an assembly line. The possibilities are endless.

The metaverse isn't the ubiquitous and all-enveloping, realistic simulation of *The Matrix* or a *Star Trek* holodeck, but hardware and software have progressed to the point where augmented reality (AR) and virtual reality (VR) are practical tools for collaboration.

Few of you use these tools for production work at this point. Nonetheless, the number is growing, and engineers and software developers should be aware of what's possible so that they can determine when to take advantage of some amazing functionality.

A lot of parts have been coming together, from high-frame-rate, high-resolution GPUs to tiny displays that can fit into a pair of glasses without feeling like you're wearing an iron mask. Collaboration software has been expanding to take advantage of the hardware. We're almost to the point where software will become the determining factor of success.

The advent of video conferencing due to COVID-19, with almost everyone involved, is a good indication of potential adoption and challenges for AR/VR. There are a dozen different video-conferencing systems in general use, with Zoom and Microsoft Teams leading the way.

Each of those systems lives within its own walled garden, with most providing a browser-based interface in addition to dedicated applications. All that's needed is a sufficiently powerful PC. Microphones, speakers, cameras, and even touchscreens are optional. All are required for the full effect.

Entering a metaverse is more of a challenge, because having the right hardware is mandatory. Plus, the level of sophistication of the hardware and matching software can significantly affect the experience. We haven't reached the point where resolution, frame-rate differences, etc., are insignificant as they are with something like a car, where minimum standards are required.

Moving from one metaverse to another is a bit of a hassle these days since it means at least switching software. The bigger challenge will be differences between the control and experiences within the virtual reality of the chosen environment. Unlike a car's steering wheel, the way to navigate through a metaverse is often unique to the application or the virtual environment.

Like machine learning, digital twins, and self-driving vehicles, knowing what they are, how they can be applied, and when to take advantage of them requires a good understanding of the technology. ed

United States Postal Service

Statement of Ownership, Management, and Circulation **(Requester Publications Only)**

1. Publication Title: ELECTRONIC DESIGN
2. Publication Number: 172-080
3. Filing Date: 9/9/2022
4. Issue of Frequency: Bi-Monthly - Jan/Feb, Mar/Apr, May/Jun, July/Aug, Sept/Oct, Nov/Dec
5. Number of Issues Published Annually: 6
6. Annual Subscription Price: Free to Qualified
7. Complete Mailing Address of Known Office of Publication (Not Printer): Endeavor Business Media, LLC, 1233 Janesville Ave, Fort Atkinson, WI 53538 — Contact Person: Debbie M Brady, Telephone: 941-208-4402
8. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not Printer): Endeavor Business Media, LLC,30 Burton Hills Blvd., Ste. 185., Nashville, TN 37215
9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor - Publisher: Tracy Smith, Endeavor Business Media, 30 Burton Hills Blvd., Ste. 185., Nashville, TN 37215; Editor: William Wong, Endeavor Business Media, 30 Burton Hills Blvd., Ste. 185., Nashville, TN 37215; Managing Editor: ,
10. Owner - Full name and complete mailing address: Endeavor Media Holdings I, LLC, 905 Tower Place, Nashville, TN 37205; Endeavor Media Holdings II, LLC, 905 Tower Place, Nashville, TN 37205;Resolute Capital Partners Fund IV, LP, 20 Burton Hills Blvd, Suite 430, Nashville, TN 37215;RCP Endeavor, Inc, 20 Burton Hills Blvd, Suite 430, ,Nashville, TN 37215;Northcreek Mezzanine Fund II, LP,312 Walnut Street, Suite 2310,Cincinnati, OH 45202;Invergarry Holdings, LP,44235 Hillsboro Pike,Nashville, TN 37215;Everside Endeavor International Blocker, LLC,155 East 44th St, Suite 2101 - 10 Grand Central,New York, NY 10017
11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities: None
12. Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: N/A
13. Publication Title: ELECTRONIC DESIGN

14. Issue Date for Circulation Data: July/August 2022

15. Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)	36,888	36,818
b. Legitimate Paid and/or Requested Distribution (By Mail and Outside the Mail)		
(1) Outside County Paid/Requested Mail Subscriptions stated on PS Form 3541. (Include direct written request from recipient, telemarketing and Internet requests from recipient, paid subscriptions including nominal rate subscriptions, employer requests, advertiser's proof copies, and exchange copies.)	31,146	32,045
(2) In-County Paid/Requested Mail Subscriptions stated on PS Form 3541. (Include direct written request from recipient, telemarketing and Internet requests from recipient, paid subscriptions including nominal rate subscriptions, employer requests, advertiser's proof copies, and exchange copies.)	0	0
(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid or Requested Distribution Outside USPS®	1,471	1,275
(4) Requested Copies Distributed by Other Mail Classes Through the USPS (e.g. First-Class Mail®)	0	0
c. Total Paid and/or Requested Distribution (Sum of 15b (1), (2), (3), and (4))	32,617	33,320
d. Nonrequested Distribution (By Mail and Outside the Mail)		
(1) Outside County Nonrequested Copies Stated on PS Form 3541 (include Sample copies, Requests Over 3 years old, Requests induced by a Premium, Bulk Sales and Requests including Association Requests, Names obtained from Business Directories, Lists, and other sources)	3,504	2,744
(2) In-County Nonrequested Copies Stated on PS Form 3541 (include Sample copies, Requests Over 3 years old, Requests induced by a Premium, Bulk Sales and Requests including Association Requests, Names obtained from Business Directories, Lists, and other sources)	0	0
(3) Nonrequested Copies Distributed Through the USPS by Other Classes of Mail (e.g. First-Class Mail, Nonrequestor Copies mailed in excess of 10% Limit mailed at Standard Mail® or Package Services Rates)	0	0
(4) Nonrequested Copies Distributed Outside the Mail (Include Pickup Stands, Trade Shows, Showrooms and Other Sources)	133	100
e. Total Nonrequested Distribution (Sum of 15d (1), (2), (3), and (4))	3,638	2,844
f. Total Distribution (Sum of 15c and 15e)	36,254	36,164
g. Copies not Distributed	634	654
h. Total (Sum of 15f and g)	36,888	36,818
i. Percent Paid and/or Requested Circulation (15c divided by 15f times 100)	89.97%	92.14%
16 Electronic Copy Circulation		
a. Requested and Paid Electronic Copies	-	-
b. Total Requested and Paid Print Copies (Line 15c)+ Requested/Paid Electronic Copies (Line 16a)	32,617	33,320
c. Total Requested Copy Distribution Distribution(Line 15f) + Requested/Paid Electronic Copies (Line 16a)	36,254	36,164
d. Percent Paid an/dor Requested Circulaltion (Both Print & Electronic Copies) (16b diveded by 16c x 100)	89.97%	92.14%

[x] I certify that 50% of all my distribution copies (electronic and print) are legitimate requests or paid copies:

17. Publication of Statement of Ownership for a Requester Publication is required and will be printed in the: September/October 2022 issue of this publication.

18 Debbie M Brady, Manager User Marketing — Date 9/9/22

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

PS Form 3526-R, July 2014

APURVA SHAH | Founder and CEO, Duality Robotics

What's the Difference Between the Enterprise and Consumer Metaverses?

The enterprise metaverse is a Fortnite but for digital twins. Presented is a perspective for developers on the challenges—and opportunities—ahead.

Credit: 238554561 @ Irinayeryomina | Dreamstime

The "enterprise metaverse"—the metaverse for commercial, industrial infrastructure, and manufacturing applications—will be far removed from the social, gaming, and entertainment ecosystem that many of us regard today as the "consumer metaverse," where entities like Fortnite and Roblox have staked an early claim.

We intuitively understand the applications for a consumer metaverse, where the core mechanics are focused on game play and socialization. But what are the applications for an enterprise metaverse, and what are they good for? Today we see three main focal points for enterprise metaverse applications emerging from our engagement with customers:

1. Synthetic Data Generation (Open-Loop Simulation)

The idea here is to create an environmental context where virtual sensors are placed either in the environment itself or on systems that inhabit it. The goal is to collect data—typically, perception data—for the purpose of training machine-learning (ML) models, or other forms of autonomous software.

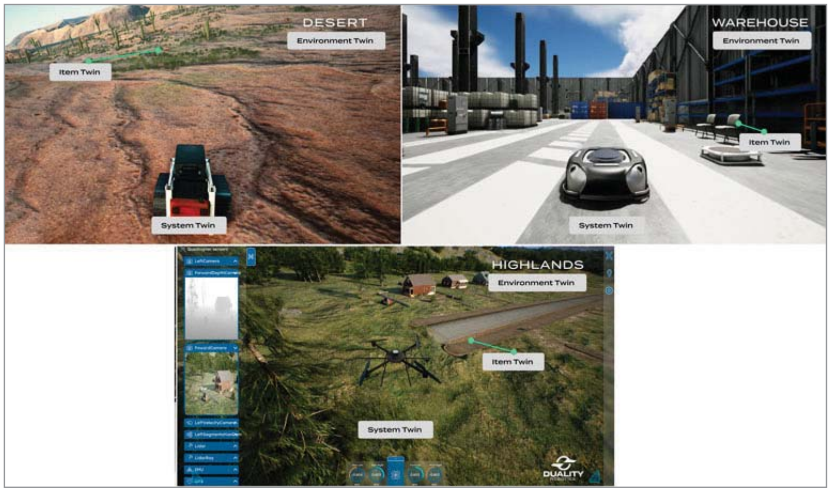
Is it worth double-clicking on what we mean by an autonomous system? It's a mechanical system that knows how to respond dynamically to its context. A robot or other mechanical system that's not autonomous will do exactly what it's programmed to do. What makes it autonomous or semi-autonomous is when it can make sense of its dynamic environment and then act accordingly—and independently.

But human experience has taught us that it can take a long time to learn how to operate successfully within a new context! The same is true for machines. The way to accelerate that learning is by exposing autonomous systems to more diverse information. It's not enough simply to generate more data. It's important that the data itself represents the diversity of situations that you want your autonomous system to respond to.

Collecting this data in the real world is often very expensive, time-consuming, sometimes dangerous—and in some cases, virtually impossible. It's not possible to order up edge cases on demand, yet they're exactly what autonomous systems need to be able to recognize and handle to operate safely and robustly.

Thankfully, when driving around town in our vehicles in the real world, traffic accidents seldom happen. Yet that's exactly what you'd want a self-driving car to experience and learn from. It just doesn't happen very often.

So how do you go and collect this data very intentionally? The enterprise



Falcon, Duality's Digital Twin Simulator, utilizes digital-twin counterparts based off of real-life appearance, physics, and behaviors. *Photos courtesy of Duality Robotics*

metaverse holds this promise: The ability to generate very accurate, scalable, and intentional data for ML models.

2. Validation of Autonomous Systems and Processes Automation (Closed-Loop Simulation)

We can also think of the enterprise metaverse in terms of validation and optimization. This differs from Open-Loop Simulation/Synthetic Data Generation in that you don't merely have a virtual context in which a system is operating; you also have actuation of those systems. These systems can run their external software brain based on input from virtual sensors and then determine how they should respond. And you can test this in a real-time closed-loop simulation.

Thus, you can run many thousands of tests before you ever deploy a system like this in the real world. Whereas with a self-driving car, a change applied to even just one line of code could make the software—and vehicle—behave unpredictably or even crash (think of how common software bugs are…). We obviously can't afford for this to happen when we introduce autonomous systems into real-world operating environments.

How do we ensure that these systems are robust and validated? It's not realistically possible every time you change a line of code to go drive a thousand miles for validation. The same applies in a construction site or a warehouse. Every time you change a line of code in your mobile robot's logic, if that robot is deployed at 40 customer sites, you can't realistically test that robot at all 40 sites before you release the software. So, how do you actually validate these systems effectively?

Until recently, we've been thinking in terms of single systems. But now imagine that in a warehouse, or an airport, we have a broad collection of systems that are autonomous, semi-autonomous, or human operated, all working together in close orchestration. At this point, it becomes increasingly challenging to think about how to go about optimizing a workflow like that from an operations perspective.

The enterprise metaverse holds this promise. For the first time, we're allowed a fully instrumented environment where a complex set of systems can interact and orchestrate, and subsequently analyze all of the output data in a closed loop to figure out what and where to optimize. Then, we can run A/B tests or myriad configurations in parallel to find the optimal workflow.

3. B2B2C Interaction

In some cases, we want to expose these digital twins to end customers or consumers. A simple example: a shaving blade. How do you really show off the features of that blade? Until recently, manufacturers have invested as much as 4X more money in the packaging than the shaving blade itself, all to influence the purchase decision. But how might you influence consumers' purchasing decisions when over 90% of those purchases are made online?

One way is to allow them to interact directly with a digital twin of your product, or your service, so that they can more fully understand and appreciate the value proposition. And this is where the lines between the enterprise metaverse and the consumer metaverse begin to blur.

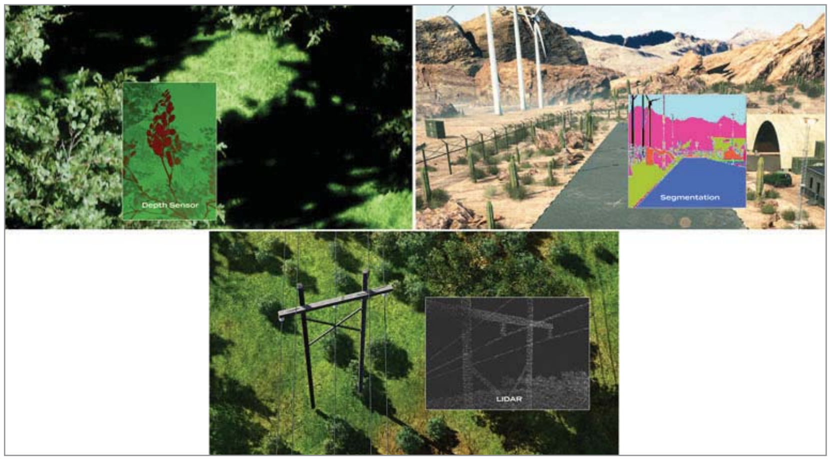
Digital Twin Challenges

The three dynamics discussed are what we've seen most frequently with customers, and all of them can drive significant business value. Some shift the bottom line, some the top line, but these are firmly grounded use cases addressing significant pain points for companies today. In other words, the metaverse future is already here.

It's a paramount concern to ensure that these autonomous "smart" systems aren't suddenly going to become "dumb" when you simply change a couple of lines of code. It's also important to allow your customers to interact directly with these digital twins so as to get a much better understanding of them.

Today there are online product configurators for almost every new automobile, a simple example of a digital twin. But it can cost half a million to a million dollars and six months working with a consulting agency to develop one of these configurators. How can we practically do this for an $8 to $12 product?

These are among the many metaverse challenges that enterprises face today. They're not solutions in search of problems, but rather acute problems that enterprises currently grapple with, and frankly don't have great solutions to overcome them.



To facilitate machine-learning training, hyper-realistic synthetic data is necessary. System twins can be modeled with a suite of sensors, including depth, segmentation, and LiDAR.

Enterprise vs. Consumer Metaverse

With the applications now established, what are the key differences between the consumer metaverse and the enterprise metaverse? There are three main differences to consider.

Who is in the metaverse? In the consumer metaverse you'll find avatars—avatars that often represent the users who are in that metaverse with their friends, performing activities, playing games, etc.

The enterprise metaverse, on the other hand, consists of digital twins. These are digital counterparts that accurately capture the appearance, physics, and behavior of physical systems, items, and spaces making up the actors and context of enterprise workflows.

If you have a digital storefront, the store is an *environment twin*. The products on the shelves are *item twins*. A reshelving robot or a frictionless checkout counter would be *system twin*s. The number one difference from the consumer metaverse is that the enterprise metaverse isn't predominantly a metaverse of humanoid avatars. It's a metaverse of digital twins.

Permeability of the data between the real and the virtual: In a consumer metaverse, what matters most is that the metaverse itself is engaging, whether it follows real physics or takes creative liberties (lighting, shading, physics, etc.). Fortnite, Roblox, and Minecraft *embrace* a very high degree of creative expression and stylization. They're not meant to resemble the real world, and this is by design.

An enterprise metaverse, on the other hand, *must be precise*, mainly because it must generate data intended to be permeable with the physical world. For a simple example, envision the digital twin of a dump truck. Whether it's driving through the terrain or dumping its load, the speed and physics must behave the same way in both the metaverse and the real world. Otherwise, the data produced by the metaverse just isn't meaningful for machine learning or validation.

Now envision a consumer shopping interaction. If a digital twin of your product behaves differently than it would in the real world—perhaps it's a garment that responds to lighting in a different way as it moves—then the experience is largely useless. The consumer can't make an informed purchasing decision based on that interaction because the metaverse product isn't representative of the real-world product.

Data generated in the enterprise metaverse must therefore be permeable, which is a two-way street. The way these digital twins are created and tuned to work correctly is typically based on data that's harvested from the real world. The enterprise metaverse is created and tuned using real-world data, and the behaviors in *that* world also produce data that's meaningful in the real world.

What is the metaverse for? The enterprise metaverse isn't an interaction-centric metaverse. It's a data-centric metaverse. The very reason for operating the metaverse is to produce data of some kind. This could be learning data for ML, or validation data for an autonomous system.

Even when consumers are interacting with the enterprise metaverse, they're creating data to evaluate how that product will work for them in the real world. The value proposition here isn't entertainment, it's functional.

Therefore, the goal of the enterprise metaverse is really data generation. In that sense, two things become very important: The entire enterprise metaverse must be heavily instrumented for efficient, scalable data generation. And, because the data doesn't live on its own, it must have a clean way to integrate and flow into and coexist with other enterprise data and analytics pipelines.

Easy, Right?

Some major challenges emerge when attempting to make all of this work elegantly at scale.

Achieving data permeability

On the face of it, it sounds simple: In the enterprise metaverse, things should behave just like in the real world. But think of the complexity and diversity and variety of what can happen in the real world—it's essentially unbounded. One has to be very focused and disciplined about what data permeability really means in a given context.

In what domain does the metaverse need to operate? Even after that's defined, it's essential to ensure that the digital twins behave accurately, and virtual sensors produce data at the desired speed and fidelity. What's typically required here is *lots* of deep domain knowledge—knowledge of simulation, physical systems, and mechanical systems—to truly make all of this work. A diverse range of skills, talents, and expertise in the team, vendors, and partners is critical.

Customer learning

Customers are very comfortable today working with data and analytics, and harnessing them to impact how key decisions are made. It's how any successful business operates today.

AI and ML are just starting to achieve that level of mainstream understanding and adoption. Oftentimes, even when we don't think we're using ML, it's there under the hood and it's solving practical challenges in seamless ways within current workflows.

When it comes to digital twins, companies today are just beginning that journey. The tools, pipeline, and plumbing doesn't really exist. The early adopters (brave souls!) will forge the way forward for everyone else. There's a significant learning curve for companies just starting to adopt digital twins in their workflows, and it's best to acknowledge and plan for that.

Standardization and decentralization

The third challenge is perhaps the trickiest to solve, with major implications for metaverse technology and commerce in the future.

Today, when we talk about the consumer metaverse, you don't think of a lonely place to occupy by yourself. You'll be with family, friends, colleagues, sometimes strangers. The metaverse is enriched by these interactions with others.

From the enterprise view, if companies want to bring together assets from different teams, other companies—perhaps from their own supply chain—can be incredibly difficult because they understandably control their assets very carefully. Bringing physical assets together become massive logistical challenges—legal friction, logistical friction, protections that are there for a reason.

But different companies, different teams—sometimes different teams within the same company—can learn a lot by bringing their assets together in a shared context. The enterprise metaverse has the promise to fulfill this potential. It represents a brand new way for companies to think, work, and operate if the friction to collaboration and developing shared contexts can be removed.

This, however, requires that the metaverse itself doesn't wind up being a walled garden for each company, or a set of walled gardens from different vendors.

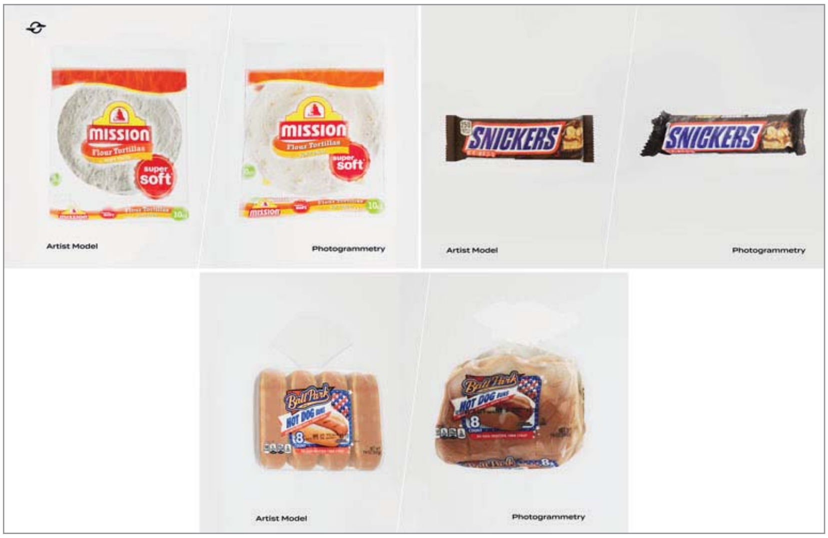
Digital twins in the enterprise metaverse must be expressed in a standardized and decentralized way so that they can be exchanged with others.

Unfortunately, on the consumer side, the trend is very much aligned toward walled-garden metaverses where it's not possible to exchange assets between them. Consumer metaverses *want* that consumer lock-in.

On the enterprise metaverse side, there's significant value for participants to not be locked in. Our hope is that the economic value the participants can potentially gain from having a common metaverse construct—a standardized digital-twin format—will result in a metaverse without walls. The vector we want to push for is standardization in terms of formats and interoperability, and decentralization.

The enterprise metaverse—the Fortnite of digital twins, if you will—presents significant challenges, but also a significant opportunity. There's real value to customers participating in standardized, interoperable, and decentralized platforms.

There will be plenty of opportunity to generate revenue in the enterprise metaverse, but it won't and shouldn't be achieved at the expense of locking in customers into walled gardens. This only hurts the ecosystem and reduces the overall value that enterprises can gain from the metaverse. ed



Techniques like photogrammetry allow for digital-twin acquisition that's based on real-world data.

KORY SCHROEDER | Director of Marketing and Product Engineering, Stackpole Electronics

Dreamstime | 537587

RoHS and Its Impact on the Resistor Industry

RoHS stands for Restriction of Hazardous Substances. Also known as Directive 2002/95/EC, it originated in the European Union and restricts the use of specific hazardous materials found in electrical/electronic products (known as EEE). All applicable products in the EU market after July 2006 must pass RoHS compliance.

The substances banned under RoHS include lead (Pb), mercury (Hg), cadmium (Cd), hexavalent chromium (Cr6), polybrominated biphenyls (PBB), polybrominated diphenyl ethers (PBDE), and four different phthalates (DEHP, BBP, BBP, DIBP). The restricted materials are hazardous to the environment, pollute landfills, and are dangerous in terms of occupational exposure during manufacturing and recycling. Exceptions are the maximum permitted concentrations in non-exempt products be 0.1% or 1000 ppm (except for cadmium, which is limited to 0.01% or 100 ppm) by weight.

The restrictions are on each homogeneous material in the product. That means the limits don't apply to the weight of the finished product, or even to a component, but to any single material that could (theoretically) be separated mechanically—for example, the sheath on a cable or the tinning on a component lead.

Fully lead-free requirements for thick-film chip resistors will inevitably increase in the next few years. For manufacturers looking to make the switch, there are options for general-purpose types, but finding stock on any of them will be difficult.

RoHS Exemption 7c-1

The RoHS Directive allows for exemptions from its restrictions, under certain conditions defined in article 5(1), adapting the Annexes to scientific and technical progress. Seven exemption groups have been approved for the use of lead in certain applications under EU RoHS Annex III for a few more years. Exemption 7c-1 has the largest impact on the electric and electronic components industry, which exempts lead in glass or ceramics as stated below:

"Electrical and electronic components containing lead in a glass or ceramic other than dielectric ceramic in capacitors, e.g., piezo electronic devices, or in a glass or ceramic matrix compound."

It's important to understand exemptions and their effective duration. Industry regularly applies for the renewal of exemptions or for additional applications to be exempted from the Directive's requirements.

Each request must be evaluated, and when appropriate, an exemption is granted. Historically, lead-based exemptions have been extended, but it's never a guarantee and staying vigilant on the RoHS site is critical.

RoHS Today

The EU is expected to take the industry recommendations and extend this exemption for lead containing glass for three more years from the original expiration date in 2021. In the past, lead has been an essential element in the dielectric glass used on nearly all thick-film chip resistors, which comprise the bulk of the resistors used globally in all markets.

Lead in the dielectric glass is critical to stabilize the laser calibration trim. It's worth noting that the lead in this glass isn't free lead; it's chemically bound to the thick-film dielectric material itself. Therefore, the potential for lead contamination due to the lead content in this material is highly questionable. However, the worldwide electronics industry has committed to removing lead from all components and materials used in manufacturing regardless of the true potential risk in each individual usage case.

Lead-Free Thick-Film Chip Resistor Evolution

Some manufacturers have had fully RoHS-compliant chip resistors for many years. Some early designs simply removed the lead-containing dielectric layer completely. Still, these designs were never successful as they had substantial instability with respect to voltage, parasitic capacitance and inductance, and pulse performance.

Later designs utilized more expensive dielectrics or may have required multiple dielectric layers to be applied prior to calibration trim. Cost for these early fully compliant series was typically 5X to 10X the cost of the equivalent general-purpose chip resistor.

Resistor Options for Completely Lead-Free Design Requirements

Most reputable manufacturers of thick-film chip resistors have completely lead-free options for general-purpose thick-film chip resistors. The current demand for this type of part is very small, though, which leads to several issues that must be addressed before a complete market changeover to totally lead-free thick-film chips.

First, the performance of these lead-free options is simply not as uniform across all manufacturers as is the case currently with thick-film chips that have the 7C-1 exemption. Design engineers and end users can be fairly confident that if they're using a reputable manufacturer of these resistors, the performance will be consistent and predictable.

For fully lead-free chip resistors, performance differences still exist between materials from different thick-film suppliers, let alone differing methods on implementation of these new materials. As volumes increase, the performance differences will be less noticeable as material suppliers and resistor manufacturers make the expected continuous improvements to materials, designs, and processes.

Another major obstacle for fully lead-free thick-film chip resistors is cost. With low volume, the new lead-free dielectric materials will be far more costly until economies of scale can be reached for current dielectrics. It will take a leap of faith by the electronics industry to make the move to these higher-cost components in the short term, with the expectation that the cost of these series will eventually come down.

This short-term cost impact will be felt in every market. However, it will obviously be magnified in markets such as cell phones, laptops, and other consumer handheld electronics, as well as the automotive industry, as these markets consume a massive amount of chip resistors.

For other specialty thick-film resistors, the completely lead-free options are much more limited. Many manufacturers don't have thick-film, fully RoHS-compliant options currently available for anti-surge/pulse withstanding, anti-sulfur, high-voltage, current-sensing, precision, or high-frequency requirements.

Although general-purpose thick-film resistors are the bulk of the global thick-film chip-resistor market need, many designs will utilize one or more of these specialty resistors. It's critical to have fully lead-free versions of these specialty thick-film resistors before the market moves away from options that require the 7C-1 exemption.

It should be noted that for certain applications, there are non-thick-film resistor options available to the design engineer. General-purpose thin-film chip resistors, with tolerance and TCR similar to thick-film chip resistors, are an option at minimal increased cost. Precision requirements may use thin-film nichrome or tantalum-nitride chip resistors. Thin-film resistors don't require a dielectric to stabilize the calibration trim area and therefore are, and have been, fully RoHS-compliant for some time.

For current-sensing applications, foil on ceramic carrier, thin-film/metal-film, and all metal sense resistors offer electrically superior performance options, but at increased component cost.

Summary

Fully lead-free requirements for thick-film chip resistors will inevitably increase in the next few years. For electronics manufacturers looking to make the switch currently, there are options for general-purpose types. However, finding stock on any of them will be difficult.

Furthermore, for specialty thick-film requirements, very few options are available from any manufacturer, and finding stock will be next to impossible. Some of these specialty applications may utilize other technologies, such as thin-film/metal-film, metal element, and wirewound, as a bridge until fully RoHS-compliant thick-film options are developed and stock becomes available. ed

JAMES MORRA | Senior Editor

Credit: Ayar Labs

Is It Finally Time for Silicon Photonics to Shine?

A new silicon-photonics process developed by GlobalFoundries has the backing of Ayar Labs, Broadcom, Cisco, Marvell, and NVIDIA.

GlobalFoundries believes there is a bright future for chips that harvest the potential of photons, the building blocks of light, instead of electrons to propel data faster at a fraction of the power and cost.

To get there, the U.S.-based foundry giant is banking on its second-generation silicon-photonics platform, called GF Fotonix. It has landed design wins with leaders in server networking chips such as Broadcom, Cisco, Marvell, and NVIDIA, as well as startups Ayar Labs, Lightmatter, PsiQuantum, and Ranovus to make chips that move data at the speed of light.

The contract chipmaker is doubling down on silicon photonics after falling behind more generally in the chip sector when it stepped out of the race with Intel, Samsung, and TSMC to make the most advanced processors.

GlobalFoundries has pivoted, with a new focus on feature-rich chips for everything from smartphones to cars that are based on mature technology nodes. Business is booming. But it believes silicon photonics is its ticket back to the leading edge.

The executive leading this effort is Anthony Yu, vice president in its Wired and Computing business unit. While it could take years for silicon photonics to make a serious dent in the data center, he said, GF is hoping to get a foot in the door with companies betting on the technology to power up artificial intelligence and even quantum computers.

GlobalFoundries is hoping to hit it big with GF Fotonix following its IPO last year. It's pointing to early wins with companies that plan to use the platform as a sign that silicon photonics is ready for mass deployment.

Yu said GF Fotonix will open the door to the next generation of silicon photonics called co-packaged optics, which promise power and cost savings when used in switch chips that call the shots in data centers. Even chip-to-chip interconnects will have to use silicon photonics to limit the share of the processor's power budget wasted on I/O. Yu noted some customers want to use GF Fotonix to create chiplets that fit the bill.

"We wanted to announce that silicon photonics has arrived, and that may sound strange given all the virtues of silicon photonics," said Yu. "But it has been in the lab for a long time, and people have doubted it."

A Long Time Coming

Tech giants have used the power of photons to send data between geographically distant data centers over fiber optics and even undersea cables for decades. But they are also increasingly using light to move data within colossal data centers, hauling it between tens of thousands of server racks. To do so, they are utilizing optical networking modules that can be plugged directly into switches to convert light into electricity and vice versa.

Shorter distances between servers and various chips inside them are still spanned with electrical interconnects that move data via copper wires. "The data center is the foothold for silicon photonics right now," said Yu.

Today, moving data at the speed of light means using specialty materials. These include indium phosphide (InP), the gold standard in lasers and other technologies that can propel photons over optical fibers, and silicon germanium (SiGe), which is widely used in the high-speed mixed-signal electronics that keep the light under control.

But bringing all of these building blocks into inexpensive slabs of silicon that can be mass-produced is a monumental challenge that presents a unique set of problems compared to cramming more transistors on a CPU, according to Yu.

Thousands of different components must be integrated in a silicon-photonics IC, including the modulator that translates data into photons, the waveguide that shines light around the IC, and the detector that converts photons to electrons to be processed. Packaging is also important for moving data into and out of the package, as lasers and fiber optics have to be attached separately, said Yu.

GlobalFoundries rolled out its first-generation silicon-photonics platform in 2016, tapping into the trove of silicon-photonics technology acquired in its takeover of IBM's semiconductor manufacturing arm.

Yu said the company had the foresight to create a separate silicon-photonics business the same year, in a bet that the world would have to use the power of light to move enormous amounts of data within and between the data centers popping up across the globe. While the industry standard for bandwidth was 40 GB/s at the time, technology giants today are gearing up for data rates of 400 GB/s and 800 GB/s.

"Everyone could see that bandwidth was about to explode, particularly as data centers became more hyperscale," said Yu. "We thought the current technology to supply that—indium phosphide—would not be able to scale up."

Data movement is also power-hungry. As data centers move to higher bandwidths, they are expending more power to move data through copper cables, and as a result, modern data centers run exceptionally hot. Yu said silicon photonics promises to solve this problem, as it burns through only a third of the power in terms of picojoules per bit.

"We were able to take silicon and get it to behave like indium phosphide through use of differentiated features and unique materials, so that we could apply scale to this problem of supplying enough bandwidth."

GlobalFoundries said GF Fotonix delivers cost-, space-, and power savings that will help shore up its position as a manufacturing leader in the market for pluggable optics, which is expected to surge to $4 billion in 2026.

Lighting it Up

In an industry keen to crown winners and losers based on process technology, GlobalFoundries said GF Fotonix will help its customers solve some of the biggest challenges facing the data center in a whole new way.

"I would assert that silicon photonics is probably as leading edge—or maybe even more so—than single-digit process nodes. The ability to control light, manage the light budget on the chip, and even operate at the frequencies of photons requires solving challenges that are totally separate from what companies at single nodes are contending with," said Yu. "Leading edge is not only defined by process nodes in the single digits."

"We're using materials and processing techniques that no other process out there uses," he said, adding CEO Tom Caulfield has been willing to invest "very significant" amounts on R&D related to silicon photonics.

GlobalFoundries is not alone in mounting a major offensive. Intel has been investing for years to advance the

GlobalFoundries said chips based on GF Fotonix will be made at Fab 8, its most advanced manufacturing site that's located in upstate New York. *Credit: GlobalFoundries* state-of-the-art in silicon photonics to solve bottlenecks in system bandwidth, power, and heat dissipation.

The company said GF Fotonix stands out as a monolithic platform that unites all of the technology's building blocks for the first time, including 300-mm photonics features and 300-GHz RF-CMOS on a silicon wafer.

"The principal feature we have that no one else can touch is that we are a completely monolithic process, meaning that we can basically put together a photonics system-on-a-chip (SoC)," Yu told *Electronic Design.*

All major components, including the passive and active photonics, radio frequency (RF), and CMOS, were previously manufactured onto separate chips that then had to be bundled together in a package, noted Yu.

Moving to a monolithic architecture has many pros. The company said GF Fotonix can deliver data rates of up to half a terabit per second (Tb/s) over a single fiber, the fastest data rate of any foundry offering. That will allow for optical chiplets running at 1.6 to 3.2 Tb/s, offering faster data transmission plus better power efficiency and signal integrity. The platform also supports 2.5D packaging technology to glue chiplets together.

Yu said GF Fotonix is based on 45-nm CMOS technology, a step up from 90 nm in its first-generation node released in 2016. Customers can get access to its Process Design Kit (PDK) to start designing chips for GF Fotonix.

GlobalFoundries also pointed out the high level of integration in GF Fotonix, which opens the door for customers to pack more functions into the same silicon-photonics IC and reduce bill-of-materials (BOM) costs.

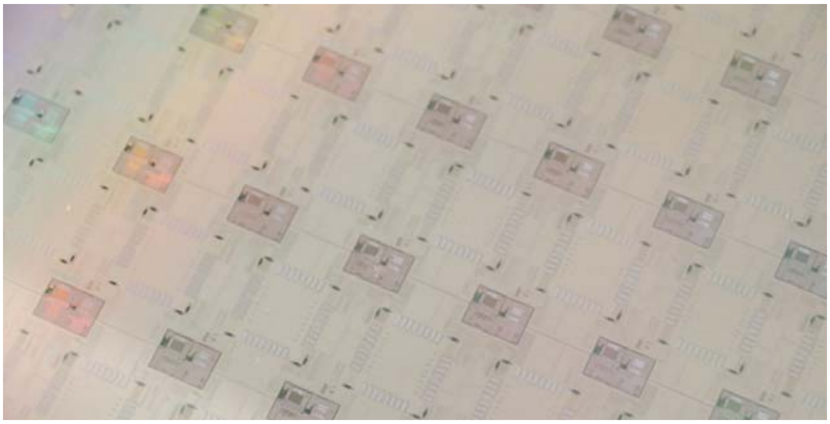
Flipping the Switch

According to GlobalFoundries, it collaborated with a long list of "customer-partners" that plan to use GF Fotonix to make sure the platform meets their needs while addressing challenges in packaging, assembly, and test.

Yu said it worked closely with them to build out the platform through "the prism of high-volume manufacturability" to make sure it will scale to mass production. "We have a long list of partners that are allowing us to drive this forward, and so we portend a surge of silicon-photonics-based ICs."

GlobalFoundries said it has partnered with Cisco on a custom silicon-photonics solution for networking gear and interconnects in data centers. The companies are also co-developing an "interdependent PDK" for GF Fotonix.

"Our heavy investment and leadership in silicon photonics, combined with GF's feature-rich manufacturing technology, allows us to deliver best-in-class prod-



AI chip startup Lightmatter has also signed up to use the new GF Photonix platform.
Credit: Lightmatter

ucts," said Bill Gartner, senior vice president of Optical Systems and Optics at Cisco, in a statement.

GlobalFoundries is also trying to help companies take another step on the road to co-packaged optics, giving them the ability to shift optics out of the pluggable module and into the same package as the switch ASIC.

Reducing the distance between the optics and the switch confers several key advantages, including reducing power dissipation. Heat is also limited, opening the door for higher port density and, as a result, bandwidth.

"The innovation of semiconductor engineers has been able to keep co-packaged optics at bay because they have found ways to increase bandwidth and reduce power using different design and packaging technologies," said Yu. But they have basically reached a wall."

Other executives have cautioned that co-packaged optics is unlikely to be competitive in the foreseeable future as the semiconductor industry enters the era of 112-Gb/s SerDes and starts looking ahead to 224-Gb/s SerDes.

There are no guarantees that the world can continue to push the envelope in electrical I/O. As a result, Yu said customers are hedging their bets, with plans to roll out prototypes of switches and processors with in-package optics by the end of the year.

GF Fotonix supports a wide range of chip-packaging technologies, too. GlobalFoundries said that it can chisel "cavities" into the silicon die, giving customers an open slot to bond lasers directly to the die, resulting in cost, power, and space savings. The company also is able to carve

(Continued on page 39)

WILLIAM G. WONG | Senior Content Director

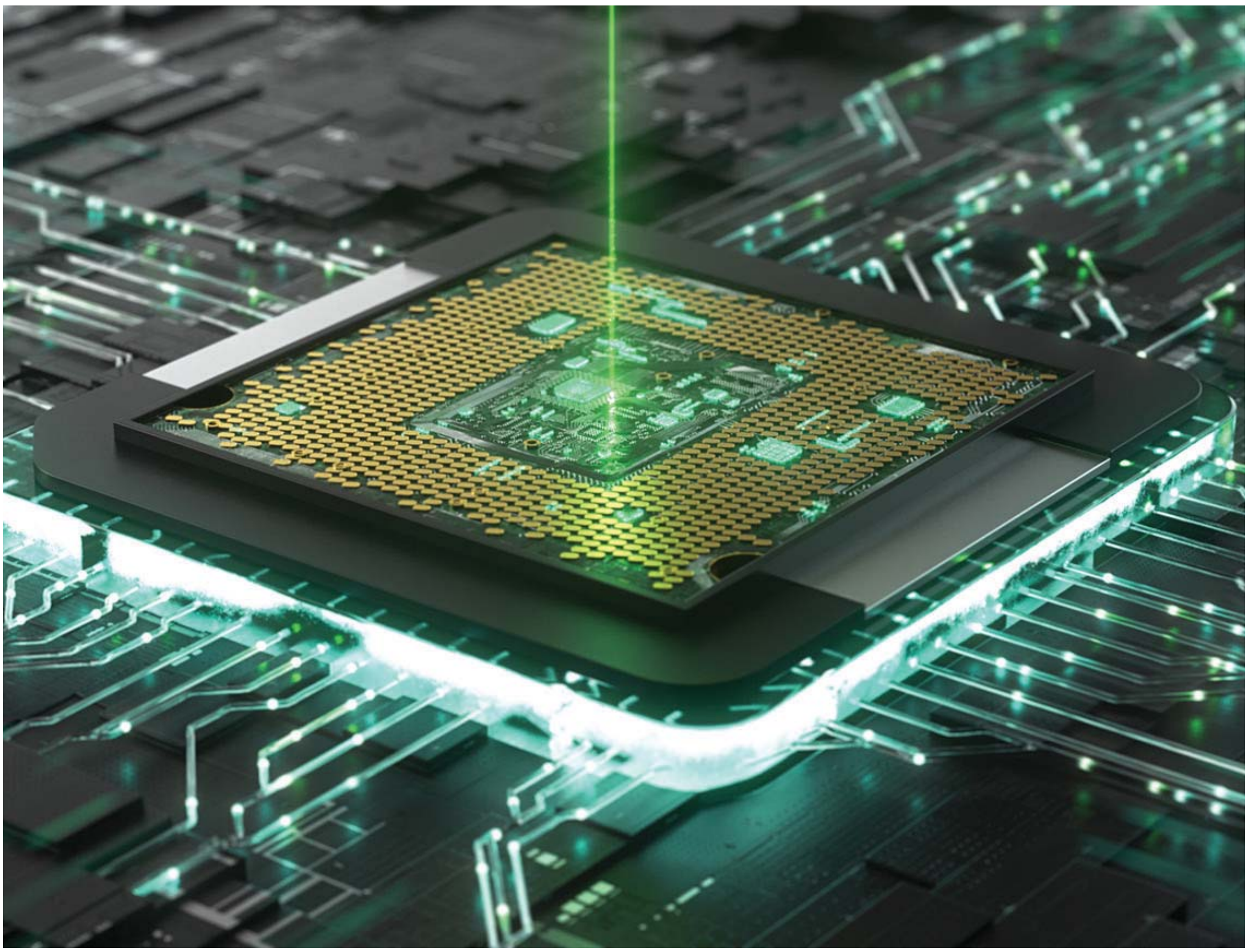
Circuit Editing Evolves to Meet Semiconductor Needs

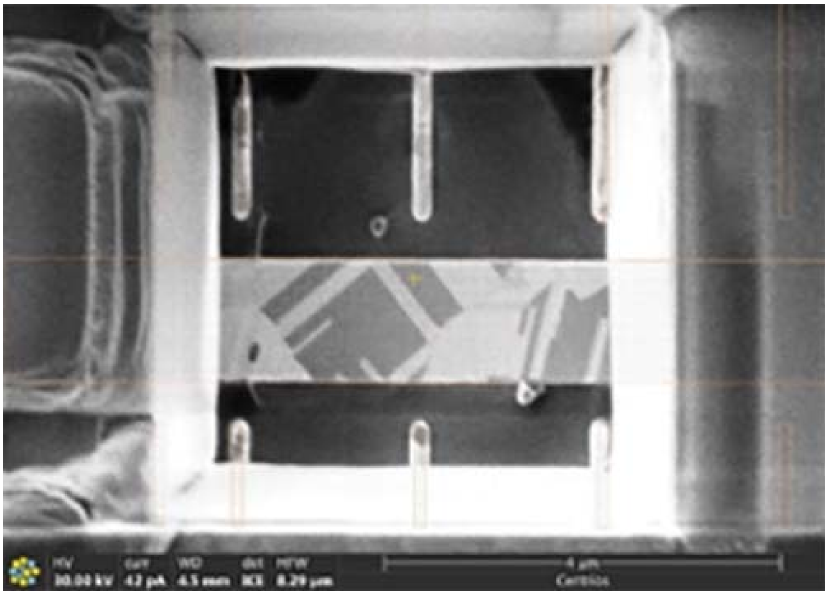
Editor Bill Wong talks with Thermo Fisher's David Akerson about circuit edit and the strategic role it plays in advancing semiconductor product development and manufacturing.

In this Q&A, Editor Bill Wong talks with Thermo Fisher's David Akerson, Senior Global Market Development Manager, about circuit edit and the strategic role it plays in advancing semiconductor product development and manufacturing.

What is circuit editing?

Circuit editing is a well-established technique that enables the direct repair of integrated-circuit (IC) defects. A precisely tuned and placed ion beam with nanoscale resolution, in conjunction with process gases, selectively uncovers internal circuitry, allowing a skilled focused-ion-beam (FIB) operator to make functional changes to the device or the copper wiring pattern. The FIB operator then reseals the chip surface to produce a functional device with revised circuit logic functions.





Removing layers of material to access the region of interest.

Why is FIB circuit editing important for semiconductor device development?

For IC developers, FIB circuit editing enables debugging and validating fixes, exploring design optimization changes, duplicating and scaling pre-production parts, and prototyping new devices without committing to the expense and lengthy delay of a mask spin. For many, circuit editing provides real tangible benefits that are quantified in cost savings, productivity, and time-to-market.

In terms of cost savings, the ability to rapidly prototype provides the ability to eliminate unnecessary wafer costs and schedule delays. With a full set of masks costing millions of dollars, the financial savings provided by circuit editing can be substantial.

There's lots of pressure on the semiconductor industry right now. How does FIB circuit editing accelerate the time-to-market for a chip?

From a time-to-market perspective, the ability to quickly prototype enables internal and external development teams, or partners, to continue product and platform development without waiting six to eight weeks for wafer-processing time to get new samples.

Internally, the ability to rapidly prototype and deliver functional devices supports the efforts of the software and firmware development; testing, validation, and qualification; manufacturing process; and sales teams. This allows them to move projects forward and avoid schedule delays.

Externally, functional prototypes can support ecosystem enablement for key technologies and end-customer's design-in and platform validation efforts. This helps fellow travelers hit their milestones and intercept market opportunities.

Logic architectures have evolved and increasingly become more complex. What are some common design and manufacturing challenges?

Each generation of architecture and fabrication process node advancement brings new challenges for semiconductor designers and circuit edit, as design and integration complexity increases.

For example, the introduction of FinFETs in 2011 on 22-nm processor nodes with new architectural structures, multi-patterning, and unique thermal and reliability behaviors brought a fundamental change to circuit editing. Where point-to-point wiring changes were scalable on planar FETs, the thinner, dielectric layer in FinFETs was more susceptible to transistor performance alterations.

For circuit editing, this necessitated the need for circuit-edit solutions with better aspect ratios for smaller cuts, higher resolutions, improved CAD navigation capabilities, and lower landing energies to keep the beam from penetrating and damaging sensitive circuits.

With new architectures, semiconductors are becoming more three-dimensional with higher aspect ratios, and more complex with smaller, more fragile features. The combination creates two new challenges—the need to to go deeper into the device and to carefully navigate around more delicate circuits. This dictates the need to "operate" in a more sensitive environment with a higher degree of surgical precision.

How are advances in circuit-editing technology helping semiconductor manufacturers overcome some of the current challenges associated with FIB circuit editing?

As semiconductors advance, circuit-edit solutions are advancing as well to meet the market's needs. For example, as semiconductor manufacturers move to gate-all-around (GAA) FET architectures and 3D semiconductor packaging, new challenges have been encountered that require advances in circuit-edit tools.

To expand on this, GAAs feature a more complicated architecture that separates the transistor blocks, the interconnects, and power distribution. The GAA also includes wiring layers above the transistors, power rails below, and higher transistor densities.

With transistor layers, interconnects, and power distribution stacked on top of each other, a 3D high-aspect ratio structure is created, making it more challenging to access the active area at depth for circuit editing. With the higher number of transistors in a given area, this makes it easier to damage them.

As a result, GAA circuit editing requires a higher degree of precision and higher level of control over the rate of the process.

Thus, there's a need for circuit-edit tools with lower landing energies and beam currents than what existed up until today.

Where many of today's circuit-edit solutions operate at 1.2 pA, state-of-the-art circuit-edit solutions, such as the new Thermo Scientific Centrios HX, can operate at 0.33 pA. This lower landing energy and lower beam current reduces damage to sensitive circuits and allows the operator to remove very thin layers with surgical precision.

Aside from semiconductor device development, are there any other important applications where FIB circuit editing can be used?

When people hear the term "FIB circuit edit," they tend to focus on the circuit-edit capabilities and don't consider the FIB's capabilities to go beyond. For example, with an advanced circuit-edit solution, the user has the sharpest scalpel in their toolbox with the ability to simultaneously capture high-resolution images and remove very small amounts of material in a controlled manner.

These characteristics make a circuit-edit solution attractive for other applications, such as deposition of semiconductors containing beam-sensitive transistors or materials, or for preparation of a device for failure analysis investigation.

What do you predict is next for the industry regarding circuit editing?

Going forward, many of the challenges encountered by IC designers will be similar in nature to those that occurred with previous inflection points. For instance, when semiconductors moved from planar to FinFETs, this created a need for circuit-edit tool improvements.

From a circuit-edit instrument perspective, as semiconductors evolve again, the device complexity places a greater premium on accurate ion-beam positioning, and the ability to make precise, delicate cuts to access and service regions of interest. As a leader in providing circuit-edit instruments, these are areas we will continue to focus on to meet the semiconductor market's needs.

> **For circuit-edit practitioners, it's a very exciting time and we expect the circuit-edit function will continue to grow in strategic value.**

For circuit-edit practitioners, it's a very exciting time and we expect the circuit-edit function will continue to grow in strategic value. As devices and manufacturing processes become more complex with more steps, the probability of functional defects becomes higher. This makes circuit edit and the ability to deliver a functional device very valuable to an organization. ed

Fast Pulse Test Systems from Avtech

Avtech offers over 500 standard models of high-speed pulsers, drivers, and amplifiers ideal for both R&D and automated factory-floor testing. Some of our standard models include:

AVR-E3-B: 500 ps rise time, 100 Volt pulser

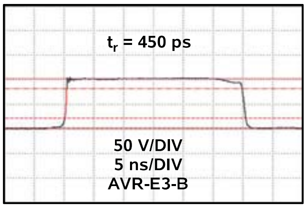
AVRQ-5-B: Optocoupler CMTI tests, >120 kV/us

AVOZ-D6-B: 1000V / 200A, 1-10 us pulser

AV-1010-B: General purpose 100V, 1 MHz pulser

AVO-9A-B: 200 ps t_r, 200 mA laser diode driver

AV-156F-B: 10 Amp current pulser for airbag initiator tests



Nanosecond Electronics
Since 1975

Pricing, manuals, datasheets and test results at:
www.avtechpulse.com

AVTECH ELECTROSYSTEMS LTD.
US & Canada 888-670-8729, Worldwide +1-613-686-6675
info@avtechpulse.com

High Output - Low Risetime Pulsers from Avtech

Nanosecond and faster pulsers covering a wide range of amplitudes!

GPIB / RS-232 / Ethernet standard on "-B" models.

Standard and customized models available.

Ampl	t_{RISE}	Max. PRF	Model
5 V	80 ps	1 MHz	AVP-AV-1S-C
10 V	120 ps	1 MHz	AVP-AV-1-B
20 V	120 ps	1 MHz	AVP-AV-HV2-B
20 V	200 ps	10 MHz	AVMR-2D-B
40 V	150 ps	1 MHz	AVP-AV-HV3-B
50 V	500 ps	1 MHz	AVR-E5-B
100 V	500 ps	100 kHz	AVR-E3-B
100 V	300 ps	20 kHz	AVI-V-HV2A-B
200 V	1 ns	50 kHz	AVIR-1-B
200 V	2 ns	20 kHz	AVIR-4D-B
400 V	2.5 ns	2 kHz	AVL-5-B-TR

Nanosecond Electronics
Since 1975

Pricing, manuals, datasheets and test results at:
www.avtechpulse.com

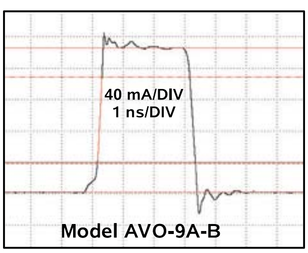
AVTECH ELECTROSYSTEMS LTD.
US & Canada 888-670-8729, Worldwide +1-613-686-6675
info@avtechpulse.com

Pulsed Laser Diode Drivers from Avtech

Each of the 18 models in the Avtech **AVO-9 Series** of pulsed laser diode drivers includes a replaceable output module with an ultra-high-speed socket, suitable for use with sub-nanosecond pulses.

Models with maximum pulse currents of 0.2A to >10A are available, with pulse widths from 400 ps to 1 us.

GPIB, RS-232, and Ethernet control available.



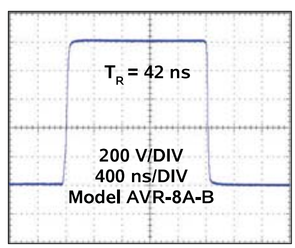
Nanosecond Electronics
Since 1975

Pricing, manuals, datasheets and test results at:
www.avtechpulse.com/laser

AVTECH ELECTROSYSTEMS LTD.
US & Canada 888-670-8729, Worldwide +1-613-686-6675
info@avtechpulse.com

100 to 1000 Volt Lab Pulsers

Avtech offers a full line of 100, 200, 500, 700 and 1000 Volt user-friendly pulsers capable of driving impedances of 50 Ω and higher. The **AVR Series** is suitable for semiconductor and laser diode characterization, time-of-flight applications, attenuator testing, and other applications requiring 10, 20, or 50 ns rise times, pulse widths from 100 ns to 100 us, and PRFs up to 100 kHz. GPIB, RS-232, and Ethernet ports are standard.



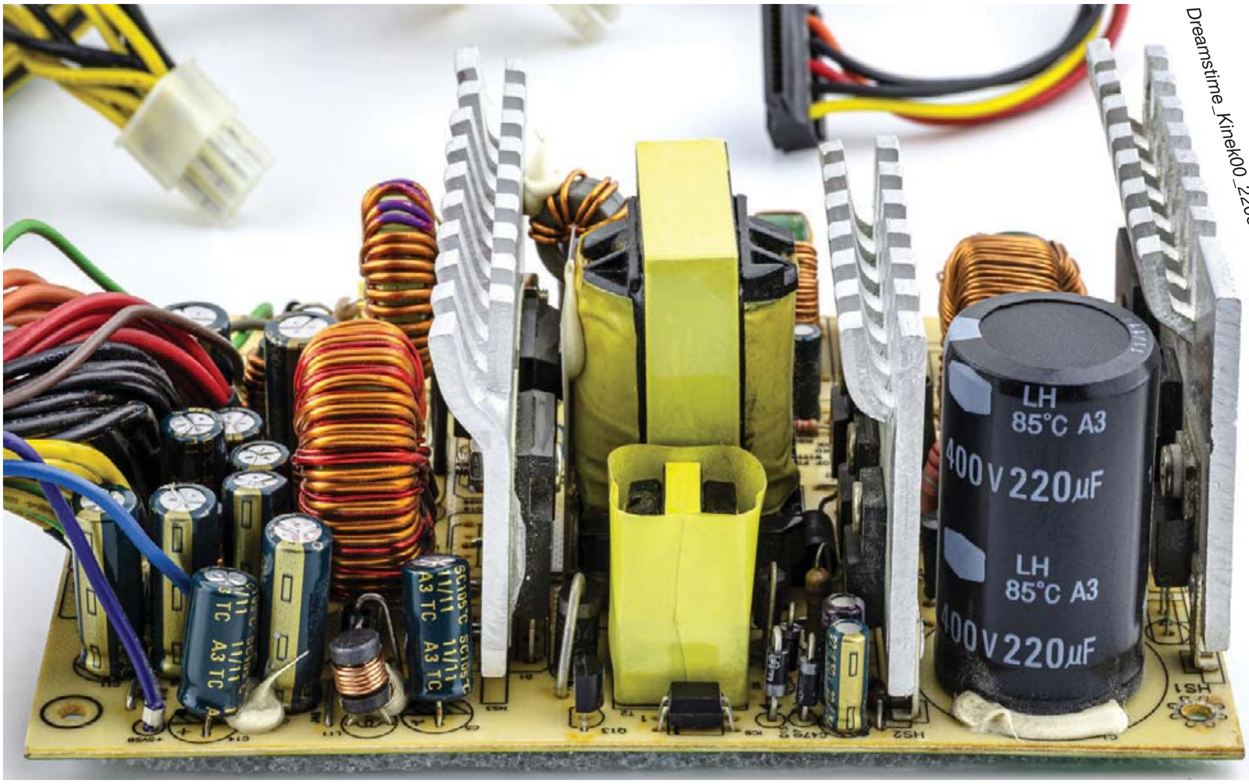
See additional test results at:
http://www.avtechpulse.com/medium

Nanosecond Electronics
Since 1975

Pricing, manuals, datasheets and test results at:
www.avtechpulse.com/medium

AVTECH ELECTROSYSTEMS LTD.
US & Canada 888-670-8729, Worldwide +1-613-686-6675
info@avtechpulse.com

LEE BOURNS | Director of Marketing, Circuit Protection Division at Bourns



Dreamstime_Kinek00_220589736

Essentials for Effective Protection Against **Overvoltage Events**

While there's no one-size-fits-all circuit protection solution, robust overvoltage protection is a necessity in virtually any application that connects to a power line. This article explores how to pinpoint the right solution based on app requirements.

Potentially damaging electrical overvoltage threats are an everyday occurrence in today's electrical and electronic world. Their intensity ranges from very light electrostatic-discharge (ESD) events to very intense lightning-strike-induced surges on data lines and power lines. These events have the potential to lock up microprocessors, damage sensors, cripple computer communications ports, cause severe damage to equipment, and even threaten harm to users through electrical shock or cause a fire.

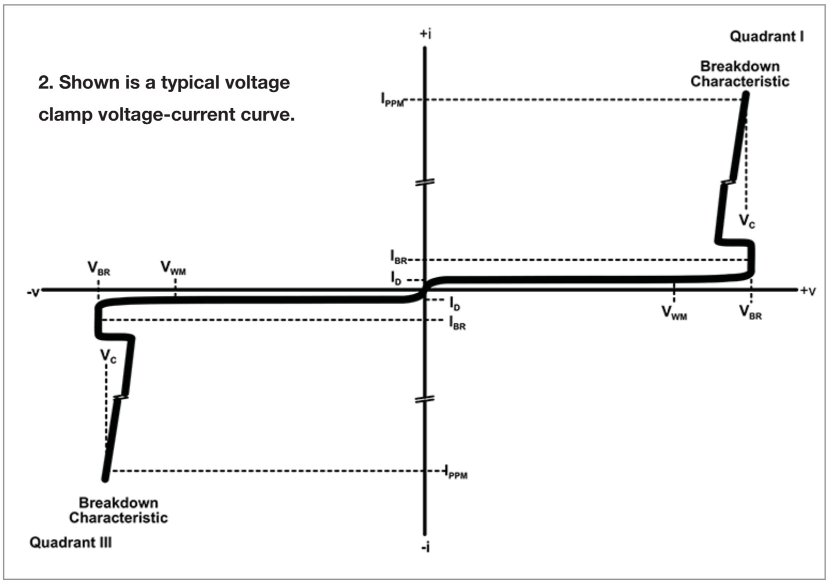
To address this wide range of threats, an equally wide range of circuit protection technologies is available. Currently available components span from small, fast PCB-mounted components to large, rugged wall-mounted devices. Some of these devices are binary in nature—they switch on or off. Others are more proportional or linear in their response to events.

Specifying circuit protection components is a bit like buying insurance. You need adequate coverage to prevent damage, but no one wants to overpay for insurance they will likely never need.

Because of the broad selection of circuit protection technologies and components available, it's helpful first to categorize the action that an overvoltage protection component takes to mitigate damage. Overvoltage protection devices are placed across the line being protected.

The response to an overvoltage event can be classified into one of the "3Ds":

1. *Divert excess energy to ground:* Often referred to as "voltage switches," these devices switch their impedance to a very low level once their terminal voltage reaches a threshold value chosen by the designer, sending the excess current to ground.
2. *Dissipate excess energy:* Regularly known as "voltage clamps," these devices lower their impedance across the protected line to attempt to limit or regulate the voltage to a level chosen by the designer.
3. *Disconnect the load from the line:* This unique technology attempts to open like a fuse and limit or block current flow when the line voltage exceeds a value chosen by the designer.



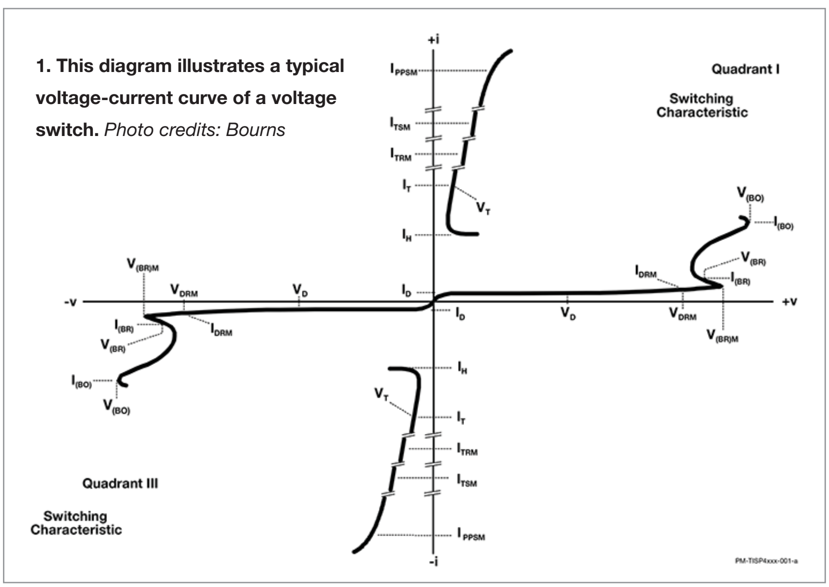
2. Shown is a typical voltage clamp voltage-current curve.

Voltage Switches

Voltage switches have very high resistance when the voltage across them is below their threshold voltage *(Fig. 1)*. Once their voltage threshold is exceeded, they transition to very low resistance, diverting the fault current to ground.

Voltage switches are binary devices—they're either "on" or "off." In the "off" state, they have voltage across them, but very little leakage current. In the "on" state, they have large currents, but low terminal voltages. This way they dissipate very little power, which means they can be much smaller devices than similarly rated voltage clamps.

However, not all circuits can tolerate being shorted out by a voltage switch. If installed on an ac power line, the voltage switch could cause a fuse to blow or breaker to trip every time it was triggered. Also, to reset the voltage switch to the high resistance state, the current through the device must be interrupted—or at least fall below the specified "holding current."



1. This diagram illustrates a typical voltage-current curve of a voltage switch. *Photo credits: Bourns*

Due to these limitations caused by this "latching on" action, voltage switches are usually found only on data lines or lines with limited power capability, such as those used in telecom or HVAC signaling.

Gas discharge tubes (GDTs) are voltage switches made from ceramic and metal materials. These are formed as an enclosed gap between two metal electrodes in a controlled atmosphere. Such ceramic devices are very rugged and relatively inexpensive. They have very low capacitance, making them ideal for high-speed data lines—especially those exposed to areas where lightning-induced surges can be very strong.

Thyristor protection devices are unidirectional or bidirectional 4-layer (PNPN) silicon structures. They switch faster than GDT devices but exhibit higher capacitance and lower surge ratings. Thyristors are often used to protect data ports, sensitive sensors, and other delicate interfaces.

For ESD protection, many suppliers offer voltage-switching silicon punch-through diodes and air-gap devices.

Voltage Clamps

Voltage clamps also exhibit very high resistance at voltages below their "threshold" voltage *(Fig. 2)*. Once the voltage across the clamp rises above this threshold, the current begins to increase rapidly—as their resistance decreases.

Metal-oxide varistors (MOVs) are bidirectional voltage clamps fabricated from sintered ceramics based on zinc oxide. They come in two basic structures: leaded and surface-mount devices.

Leaded devices are typically formed into discs with leads attached to each metallized side and covered in an epoxy powder coat. Surface-mount devices, made in thin layers of ceramic "tape," are commonly referred to as multi-layer varistor (MLV) devices. MLV devices are typically used in low-voltage signal applications, while leaded MOVs are often used in ac power protection.

Moreover, suppliers are developing new hybrid protection devices that combine an MOV and GDT.

MOV devices are known to have inherent weaknesses. Constant exposure to ac line voltages and seemingly insignificant line noise can increase the MOV's leakage current over a long period of time. With increasing leakage current, the MOV begins to dissipate continuous power, which warms the device and further increases the leakage current. This can lead to a thermal runaway condition, resulting in MOV failure.

The GDT in a hybrid protection device effectively removes the MOV from the line voltage. When a surge event occurs, the GDT reconnects the MOV to the line in less than a microsecond. Once the surge event is over, the GDT disconnects the MOV from the line.

The GDT in a hybrid protection device effectively removes the MOV from the line voltage. When a surge event occurs, the GDT reconnects the MOV to the line in less than a microsecond. Once the surge event is over, the GDT disconnects the MOV from the line. This means the MOV is connected to the ac line only when it's needed.

Another type of hybrid protection device is constructed by forming the MOV from two specially shaped thin discs, whereby the GDT function is performed by a gap between the two discs. This type of hybrid device is able to better manage the surge current, handling twice the surge current as an MOV of the same diameter. Because it's not constructed with an external GDT, it features the same thickness as the MOV, providing a significant space-saving advantage.

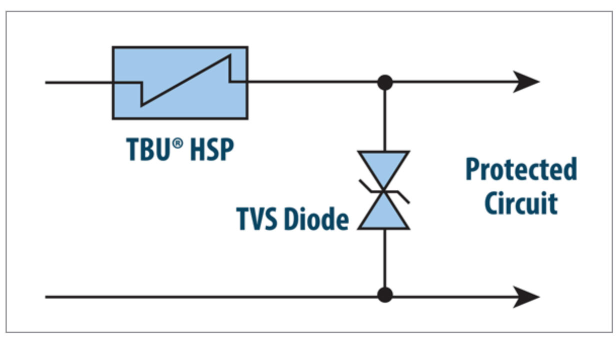
Both hybrid protection solutions also drastically lower the capacitance when compared to an MOV alone. Thus, they're ideal for use on lines where high-speed data is present.

Silicon-based voltage clamps can be either unidirectional or bidirectional. They're available in multiple sizes to handle surges from tiny ESD pulses to near-direct lightning strikes of 10 kA or more. Such clamps include transient voltage suppressors (TVS), power TVS (PTVS), silicon avalanche diodes (SAD), and Zener diodes.

The breakdown voltages of silicon devices can be more precisely controlled than with ceramic devices. Their breakdown voltage also is more stable over time, making them a premium choice for very sensitive applications.

In their active state, voltage clamps conduct current while maintaining a high voltage across the line. This leads to very high levels of power dissipation. To handle these high-power levels, voltage clamps tend to be larger per their increased surge and voltage ratings.

3. By integrating MOV and GDT technologies into a single space-saving device, hybrid overvoltage surge protectors such as Bourns' GMOV and IsoMOV are able to boost performance, reliability, and lifetime operation compared to a traditional MOV.



4. Shown is the circuit arrangement between the TVS diode and the Bourns TBU HSP device.

Part Number	Technology	Rated Voltage	Capacitance	Holding Current	Surge Rating (8/20 µs)	Package
SMBJ15CA	TVS Diode	15 V	900 pF	N/A	123 A	SMB
TISP4030H1BJ	Thyristor	15 V	62 pF	50 mA	123 A	SMB

Overvoltage protection component solutions.

In their active state, voltage clamps conduct current while maintaining a high voltage across the line.

Electronic Current Limiters

Combining two technologies, electronic current limiters (ECL) are a form of blocking device that uses a small TVS diode to sense the overvoltage event and trip an ECL such as a Bourns Transient Blocking Unit (TBU) High-Speed Protector (HSP) *(Figs. 3 and 4)*.

ECL devices are FET-based two-terminal integrated circuits that sense their current and react in about a microsecond to currents that exceed their rated trip current. Once tripped, the internal FET devices are held in the off state using a small quiescent current (~1 mA) until the fault is removed. Once the fault is cleared, the ECL will return to its normal low-resistance state.

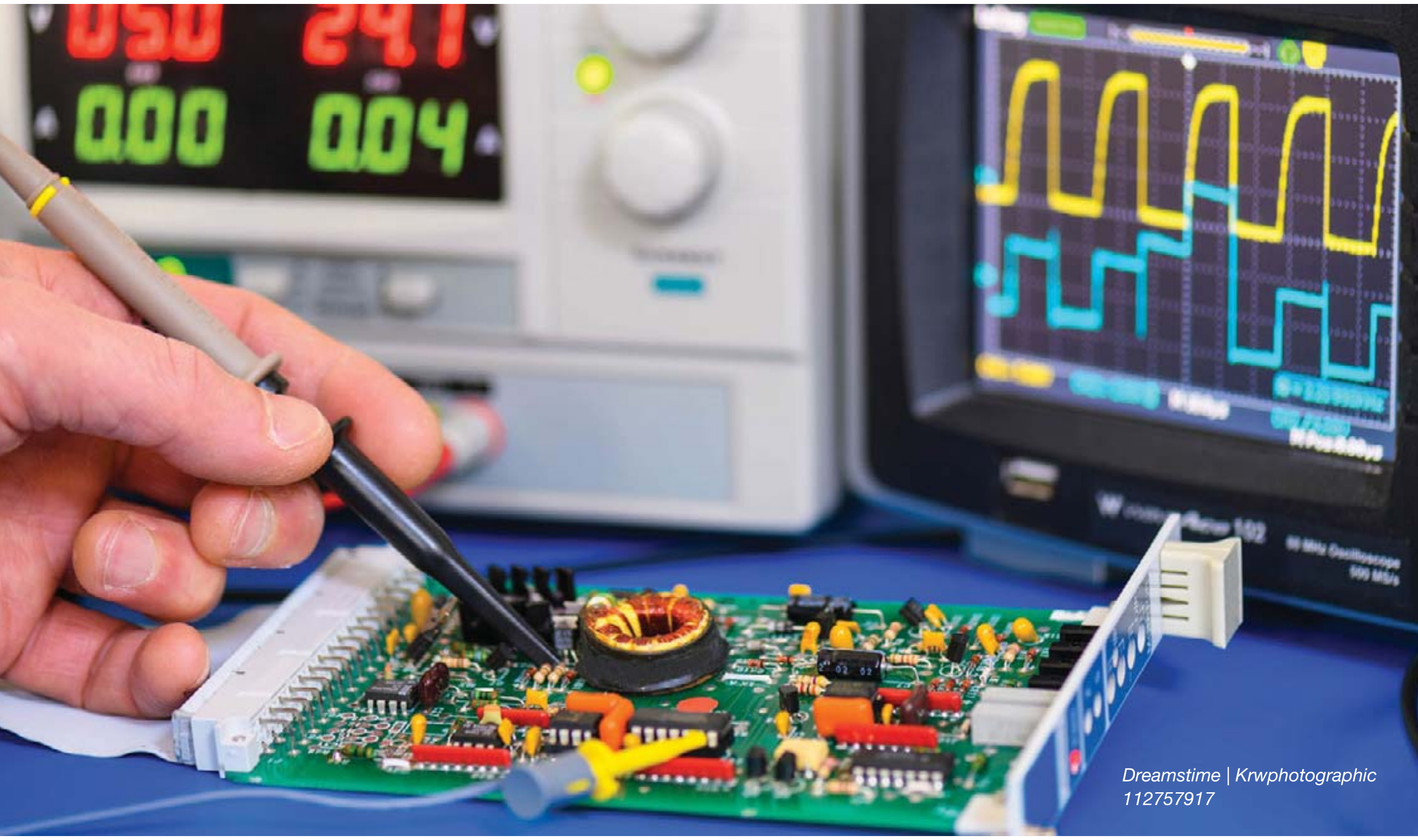
In this arrangement, the TVS device begins to conduct when the threshold voltage is exceeded. That causes current to flow in the TVS device, which in turn trips the device. This action disconnects the protected circuit from the overvoltage source. Because the ECL limits the current through the TVS device, the maximum voltage let through to the protected circuit is fixed no matter the size of the incoming surge event.

Continued Design Protection

Keeping our electrified world working in the face of overvoltage threats from lightning, induced surges, utility power voltage-regulation problems, or even just ESD is the mission for overvoltage protection devices *(see table above)*.

There's a lot to be considered when designing effective protection circuitry. Nonetheless, careful selection and appropriate testing helps to mitigate the risk of overvoltage threats, thereby increasing the reliability and uptime in today's modern electronic equipment designs. ed

JOHN STEVENS | Director of Business Development, Navitas Semiconductor



Dreamstime | Krwphotographic 112757917

Current Sensing: Past, Present, and Future

While measuring voltage is often a simple task, measuring current is usually not so straightforward. This article demonstrates a new, highly integrated, "lossless," and localized approach to current sensing that deals with many of the challenges.

For any electrical control system, input signals of both voltage and current are critical for the system to operate according to the intended design and in a safe manner. In power supplies and motor drives, for example, current is measured for control (like peak-current mode or average-current-mode control in power converters), protection (like short-circuit overcurrent protection), and data-logging purposes.

Shunt Resistors

One of the simplest methods for measuring current is to add a shunt series resistor into the current path of interest, and then measure the voltage across it as a representation of that current—knowing the relationship of voltage = current × resistance, ($V = I \times R$), and assuming a stable, or linear, resistance.

This popular method is simple and accurate. It can be used for dc or ac current measurements and is flexible for a variety of applications. While not intrinsically isolated, some vendors offer isolated amplifiers or isolated ADCs for this approach.

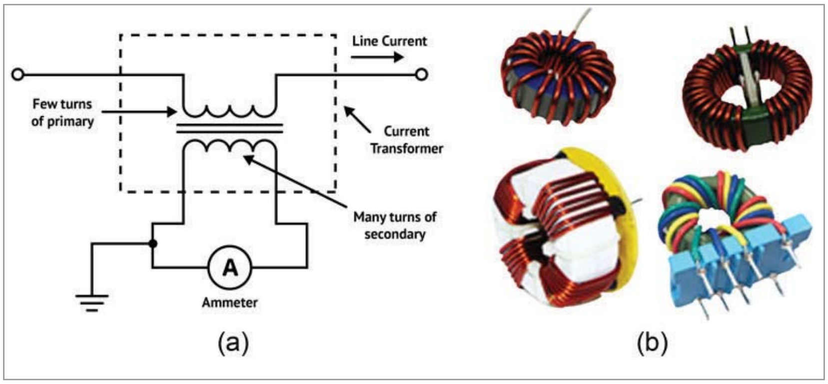
Vendors like Allegro, Broadcom, onsemi, Skyworks, Texas Instruments, and many others offer solutions in this area.

However, this current-sense method does introduce an additional power-loss element, especially at high current, which can be a severe limiting factor in high-power, high-efficiency systems. It also adds parasitic source inductance in the power path and gate loop of low-side switches, where it's most unwanted, causing delayed turn-off and voltage spikes.

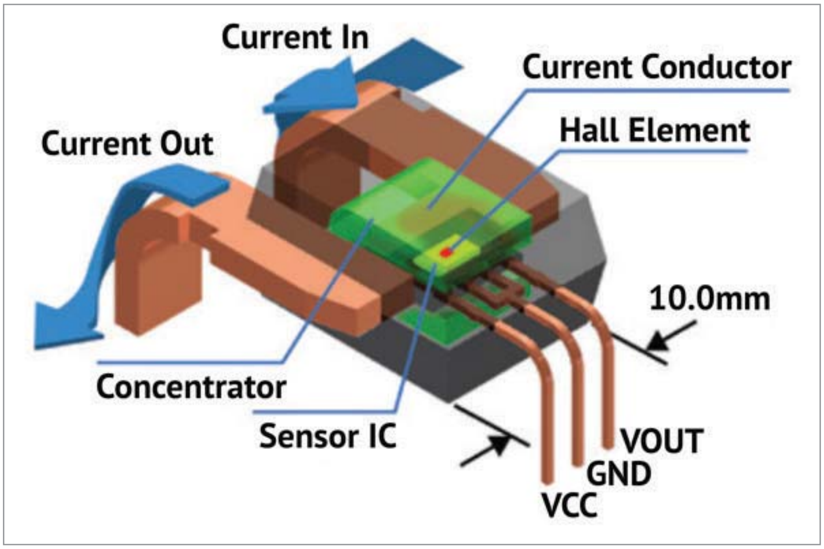
These resistors aren't standard off-the-shelf components either. To carry the full rated load current, and meet typical accuracy of 1% resistance over temperature, very low on-resistance ($R_{DS(ON)}$) FETs and highly optimized shunt resistors must be used. To avoid heavy losses, a variety of low- $R_{DS(ON)}$ FETs with large thermal package options are available from current-sense resistor providers such as Bourns, Ohmite, Susumu, Vishay, and others.

Pushing for low power loss and low voltage drop across these resistors then requires excellent op amps to gain and condition the signal properly for use. Further, to meet efficiency and thermal requirements with this added series resistance, designers also must use lower $R_{DS(ON)}$ and more expensive components to reduce the other power loss in the power path to compensate (for example, using even lower $R_{DS(ON)}$ power MOSFETs).

What started out as a simple measurement now requires multiple additional components that need to be carefully selected, designed, and powered properly, while also consuming PCB space and cost.



1. Shown is current transformer operation and various transformer examples. *Photo credits: Navitas Semiconductor*



2. These are the components of a Hall-effect current sensor.

Current Transformers

Another common method of current measurement that simplifies the system when isolation is needed, or with high-power, high-efficiency systems, is using a current transformer (CT). Gain can be accomplished with the transformer turns ratio, isolation is built in, and bidirectional current can be measured. In addition, no bias supply is required for the CT, regardless of whether it's used in high-side (floating reference) or low-side (ground-referenced) configuration.

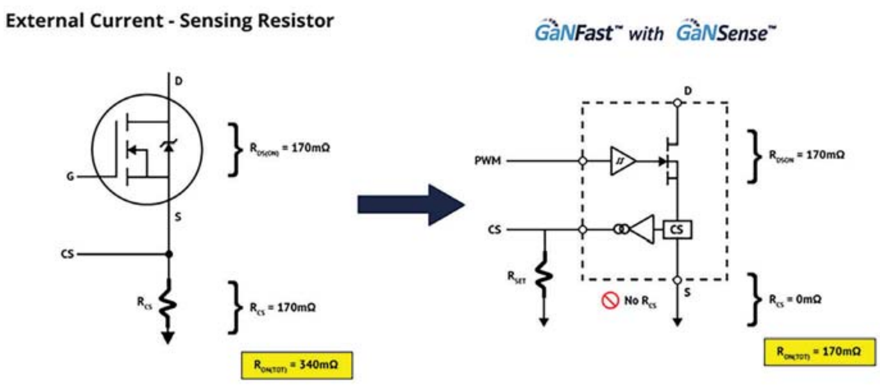
The downsides, though, are that dc current can't be measured; duty-cycle limitations may prevent transformer saturation; and these CTs are usually large components—most typically toroids like those in *Figure 1*, which may limit their adoption in high-density systems. LEM, Renco, Wurth, and many other magnetics vendors offer dedicated products focused on this approach.

Hall-Effect Sensors

For applications at still higher power, or when dc current information is required, designers often look at sensing methods that minimize the lossy pass elements being added to the system (like shunt resistors or CT primary windings), and instead rely on measuring field effects of the system.

Considered "non-contact," "lossless" sensors, methods like Hall-effect current sensors *(Fig. 2)* operate under the principle that for a given conductor (like a copper trace on a PCB) with current flowing through it, a proportional magnetic field is created around the current-carrying conductor. By measuring such a magnetic field, information on the value of the current that produced it can be obtained. The sensing element often has the PCB copper flow through the sensing-element package, and some others place the sensor above the copper trace and sense through proximity only.

Hundreds and thousands of amps can be measured with low loss because resistance and inductance aren't being added to the system. However, the method becomes inaccurate at low current, can be made inaccurate by external fields or mechanical orientation, and often requires a zero-current offset and a concentrator to magnify the signal for a narrow range of current of interest. Furthermore, a dedicated bias power supply is required for such a measurement IC. Allegro, Koshin, Melexis, and many other vendors offer products focused on this approach.



3. This is a simplified block diagram of GaNSense technology with lossless current sensing and cycle-by-cycle overcurrent protection as seen in NV6134.

Anisotropic Magnetoresistive (AMR) Sensing

The AMR IC is capable of sensing both ac and dc signals and offering a high-bandwidth solution up to 1.5 MHz, with an isolated output. It operates by having the current flow into the device through a low-resistance "U-bend" in the leadframe, where it generates either a forward or reverse magnetic field that's sensed by two differential current sensors.

This approach is attractive in that it's fairly compact, resistant to external fields and noise, and has low offset error while also responding very quickly (<300 ns). Thus, it can be used in high-frequency control applications such as power supplies and motor drives using wide-bandgap semiconductors such as gallium nitride (GaN) and silicon carbide (SiC).

The Future: Introducing "GaNSense" Technology

With the continued push for higher efficiency, GaN-based power converters have become more popular in the market. In addition, with high levels of integration used in monolithic GaN power ICs, power converters could be pushed to higher frequency with a minimal number of external components, allowing for much smaller systems.

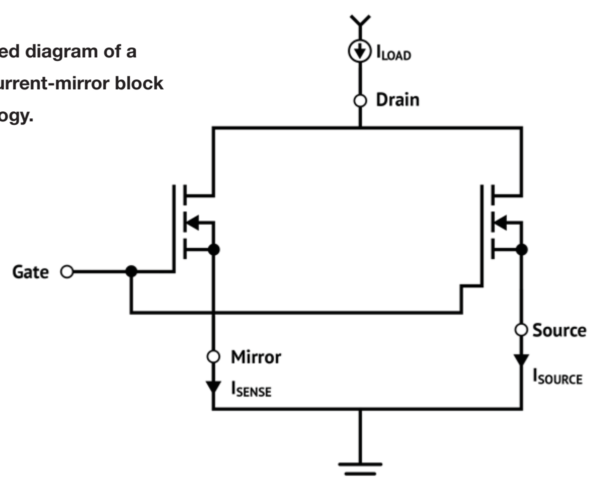
Groundbreaking products such as the latest fast mobile-phone chargers have been able to achieve power densities even 3X higher than best-in-class previous solutions. But the efforts to extract all of the capability out of GaN continued, and in late 2021, GaNSense technology offering further integration and expanded features was released.

In its simplest form, GaNSense is lossless current-sensing in-circuit, removing the shunt current-sense resistor and its associated headaches *(Fig. 3)*. It improves the overall efficiency of the system with lower total series resistance, while also increasing robustness with fast, internal, 30- to 100-ns short-circuit protection.

This approach is "localized" in that it's sensed and acted on locally to the current and power elements of interest. Localized control and response time allows this method to be fast and effective, while also minimizing corruption of the control signals from system noise, long traces, etc., compared to other methods. Therefore, it becomes possible to use this method not only for overcurrent failsafe and short-circuit protection, but also cycle-by-cycle current limit and current-mode control and regulation.

Such a self-governing power block may indeed represent the future of power-stage sensing and protection, enabling higher reliability, pushing the limits of performance, and freeing the main system controller to focus on

4. Shown is a simplified diagram of a current-sense (CS) current-mirror block in GaNSense technology.

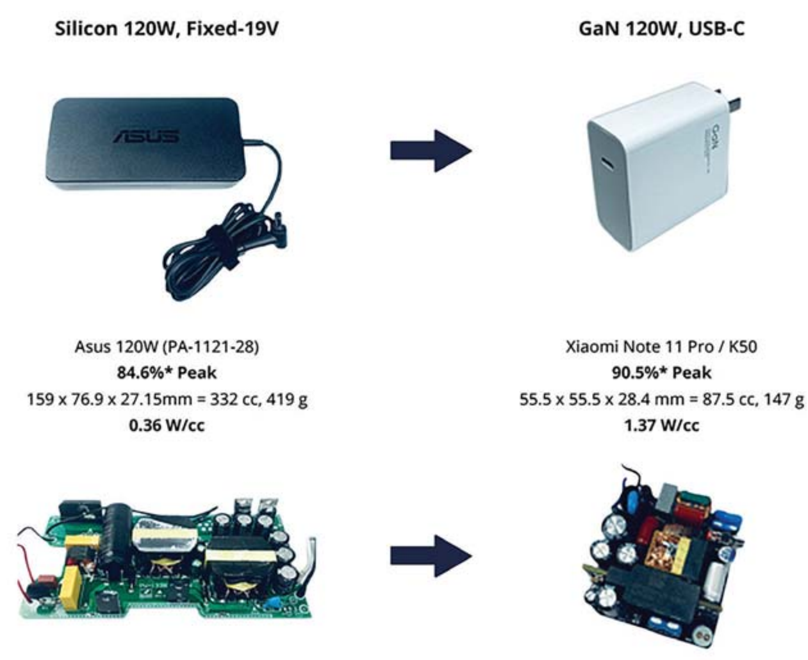


more complicated control algorithms and responsibilities. This "lossless" current sensing in the current-sense (CS) block is implemented through a popular parallel current-mirror design technique, and further optimized to be used with high-speed GaN FETs.

As seen in *Figure 4*, within the CS block, the main power FET device is connected to the common drain and gate connections. For simplicity's sake, we will show the source connected commonly as well. By using well-matched devices and a high on-resistance sense-FET (maybe >1000-1500X higher $R_{DS(ON)}$ than the main power FET), a small portion of the load current branches off to the sense FET and can be measured accurately through a variety of techniques. $R_{DS(ON)}$ and temperate affects are cancelled out naturally, too.

Because of the tight matching, the current is based on the ratio of the resistances of the devices only. As the sense FET has much higher $R_{DS(ON)}$ than the main power FET, the loss from this approach is negligible, especially compared to alternatives like shunt current-sense resistors in the main power path. The current limit is still adjustable and programmable, with the Current Sense (CS) pin and programming resistor.

All of these benefits translate to system improvements. For example, as seen in *Figure 5*, designers leveraged GaNSense technology to achieve an ac/dc 120-W mobile-phone charger that was 70% smaller, 65% lighter, and almost 6% more efficient than the previous production solution.



5. Here, GaNSense technology is leveraged in a real power supply.

Summary

In motor drives and power supplies, voltage and current will continue to be signals that need to be monitored and measured for accurate and reliable real-time control, protection, and data logging. With a variety of applications and requirements comes multiple current-sensing approaches to meet the design goals in the most cost-effective and size-efficient manner.

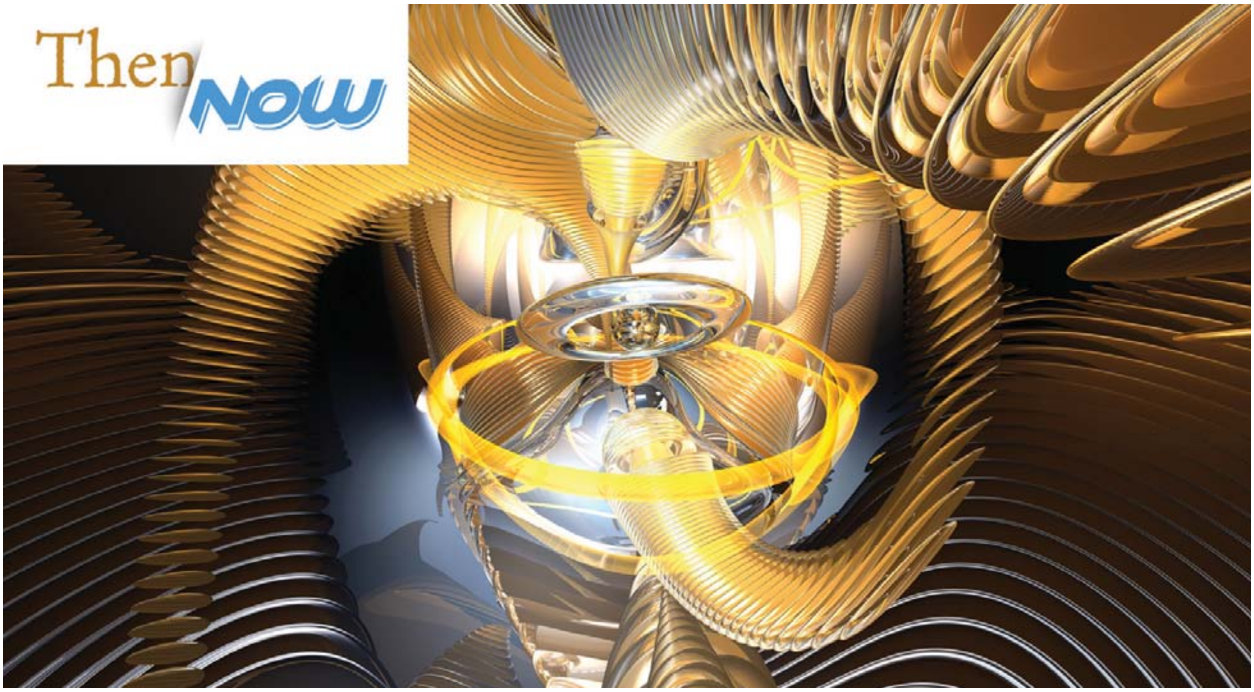
A technology and approach has been discussed here with the advent of near-ideal, lossless "GaNSense" current sensing fully integrated with the power device that's fast, accurate, and can serve the critical functions of both control and protection with effectively zero size and cost impact. While not an exhaustive study, some of the most common methods used in industry today have been discussed and are summarized in the *table below*.

Common Current-Sensing Methods and Their Performance

	Bandwidth	Response time	Accuracy	Requires bias supply	Current Range	Power Loss	Relative Size	Sensing cost
Shunt resistor + op amp	kHz-10 MHz	300 ns-1 ms	1-5%	Yes	mA-A	mW-W	Medium	$$
Current transformer	60 Hz-1 MHz	300 ns-1 ms	3-5%	No	mA-kA	mW	Medium-Large	$$
Hall-effect	kHz-100 kHz	10 us-1 ms	5-10%	Yes	A-kA	mW	Large	$$$
AMR	Hz-1.5 MHz	<300 ns	0.6%-3%	Yes	mA-A	<40 mW	Medium	$$$
GaNSense lossless current sensing	Hz-8 MHz	30-100 ns	2%	No	mA-A	<10 mW	0	0

WILLIAM G. WONG | Senior Content Director

Dreamstime | Janybaby 878202



Then and Now: Graphics Processing

Here's a graphical look back and forward at display processing technology, from CRTs to ray tracing.

A lot has happened over the past 70 years in the electronic and display space. *Electronic Design* has been covering the technologies and continues to present the latest that are significantly more powerful than those of the past. In fact, graphics processing has exceeded what conventional CPUs can do.

Displays and printed output have been a way to get information that's been fiddled with by a computer into a form that we poor humans can understand. Printers are still around and hard copy remains a necessity, but video displays dominate the human machine interface (HMI) these days. It shows up in smartphones, HDTVs, and nav systems in cars. Underlying these systems are display processors and typically more robust graphics processing units (GPUs).

The history of graphics processors is an ongoing effort by Jon Peddie in our *Graphics Chip Chronicles* series *(go to www.electronicdesign.com)*. It's up to Volume 7 and covers dozens of chips. I'm snarfing a significant chunk of this article from this series that provides more in-depth coverage of each chip.

Rise of the CRTs

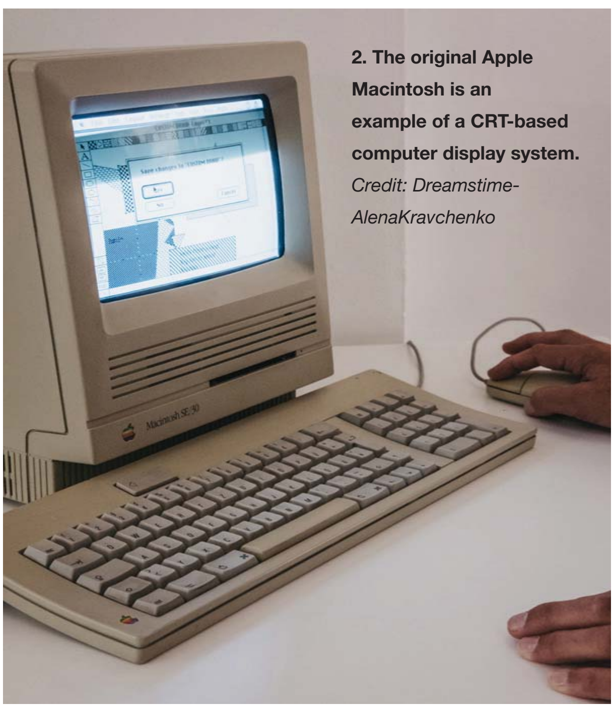
We'll skip past the individual status lights found on early computers to the cathode ray tubes (CRTs), which became ubiquitous with the rise of the television. CRTs were driven by custom controllers that usually processed signals based on analog standards like those from the National Television System Committee (NTSC).

The CRT was versatile in that it supported vector and raster images *(Fig. 1)*. Actually, the raster image is nothing more than a fixed format vector image. The difference is that the encoding for a raster image maps to the regular movement of the electron beam, while a vector display moves the beam based on positions provided by the display controller. An analog oscilloscope is an example of a vector display designed to show analog signals versus a time frame.

1. A cathode ray tube can display vector and raster images depending on the drive electronics. *Credit: Cebas, Dreamstime*

Raster displays are simply a 2D matrix of pixels mapped to a physical display. The value used to encode a pixel varies depending on the device and associated controller. A single bit works for a single-level, monochrome display while the latest displays use as many as 48 bits to describe a single pixel (16 bits for red, green, and blue). Multiple encoding schemes have RGB (red, green, blue), YUV, HIS (hue, saturation, intensity), and so on.

CRTs were the basis for analog television and early digital displays, and they've now been replaced by flat-panel LCD and LED technology. CRTs also were used in computer display terminals, often referred to as video display terminals (VDTs) *(Fig. 2).*



2. The original Apple Macintosh is an example of a CRT-based computer display system. *Credit: Dreamstime-AlenaKravchenko*

Enter the Display Controllers

One of the first display controllers listed in the *Graphics Chip Chronicles* is the NEC µPD7220 Graphics Display Controller. According to Jon Peddie, "In 1982, NEC changed the landscape of the emerging computer graphics market just as the PC was being introduced, which would be the major change to the heretofore specialized and expensive computer graphics industry."

The µPD7220 was a 5-V dc part that consumed 5 W and was housed in a 40-pin ceramic package. It had support for a lightpen and could handle arcs, lines, circles, and special characters. The processor featured a sophisticated instruction set along with graphics figure drawing support and DMA transfer capabilities. It could drive up to 4 Mb of bit-mapped graphics memory, which was a significant amount for the time.

A slew of chips followed with increasing functionality and performance. One of the most dominant chips was the IBM VGA used with the IBM PC, which brought the personal computer (PC) to the forefront. Its big brother, the IBM XGA, was a superset with more resolution and performance.

A significant feature of the VGA was the use of color lookup tables (cLUTs) along with a digital-to-analog converter (DAC) in a single chip. A ROM was used to store three character maps. The VGA supported up to 256 kB of video memory and had a 16-color and 256-color palette. Each color had an 18-bit color resolution, 6 bits each for red, green, and blue. The VGA offered a refresh rate up to 70 Hz. The 15-pin VGA interface also became a ubiquitous connector on PCs and laptops.

3D Displays

Three-dimensional displays have been around for a while, from movie theaters using 3D glasses to HDTVs. These days virtual and augmented reality are likely homes for 3D content. There are 3D displays that don't require glasses for viewing, but they've yet to achieve much success at this point.

3Dfx was one of the first companies to deliver 3D support in graphics processors with its Voodoo Graphics chipset introduced in 1996. Originally delivered as an add-on card, 3Dfx chips eventually found their way into display adapter cards for graphics adapter companies like STB.

3D presentation on 2D displays is another matter that we'll touch on later.

GPUs Take Over

NVIDIA's GeForce 256 was the first fully integrated GPU—it was a single-chip processor that incorporated all of the lighting, transformation, rendering, and triangle setup and clipping. It delivered 10 Mpolygons/s. The 256-bit pipeline was built from four 64-bit pixel pipelines.

At this juncture, gaming was a big factor in GPU design. Other GPUs targeted CAD. These days, there are still GPUs optimized for one of many applications. In general, though, they're all powerful enough to handle any chore, just not necessarily as fast or as efficient as a GPU optimized for a particular application.

The industry has seen significant consolidation in the GPU space with AMD, Intel, and NVIDIA now dominating the space. AMD and NVIDIA sell chips that third parties can incorporate into their products. Some wind up in embedded systems while others target consumers. They all sell their own brands as well.

GPGPUs Add Computational Chores

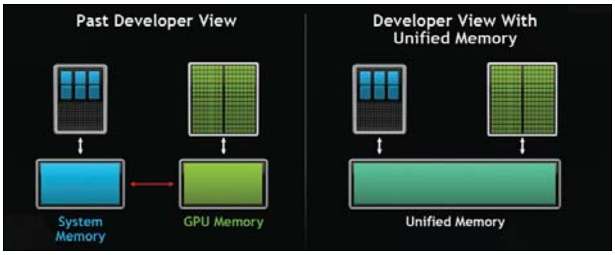
Initially, GPUs were an output-only device. They were programmable to a degree and typically connected to a display such as a CRT, projector, or flat panel. Unfortunately, GPUs started out as closed/proprietary systems that were only accessible via drivers provided by the chip vendors. Nonetheless, engineers and programmers eyed the parallel programming potential that often exceeded CPU performance.

Eventually general-purpose GPU (GPGPU) computing became available. Standards like OpenCL and NVIDIA's proprietary CUDA make these platforms suitable for general processing chores. Programming was different from a CPU, but languages like C were augmented to work with the new, exposed hardware.

Part of the challenge was mapping programs to the hardware that started out supporting gaming and graphical displays. Mapping 3D environments onto a 2D presentation were part of the driving factors for developing the single-instruction, multiple-data (SIMD) and single-instruction, multiple-thread (SIMT) architectures. The surge of this type of programming is relatively recent, starting about 2006.

Since then, the scope and chip architectures have changed radically, especially with the incorporation of machine-learning (ML) and artificial-intelligence (AI) acceleration. The GPGPUs were ideal for handling AI/ML models compared to CPUs, although the GPGPUs are being bested by AI/ML accelerators optimized for specific tasks. Likewise, CPUs and GPGPUs now implement AI/ML acceleration and AI/ML-optimized opcodes along with data handling, making them even better for these kinds of applications.

Originally, GPGPUs were a compute island with their own memory and programming elements *(Fig. 3)*. The original split-memory approach was one reason why GPGPU parallel-programming object code was delivered in "kernels" that were executed on the GPU. Data and kernels were copied to the GPU's memory and the resulting data was copied back to the CPU's memory.



3. GPGPUs have moved from a split-memory architecture (left) to a unified memory architecture (right). *Credit: NVIDIA*

The next step was having the GPU move data around, finally unifying the CPU and GPU space, which put the GPU on the same level as a CPU. A CPU is normally required within a system with one or more GPUs. However, in many instances, the CPU simply handles the initial boot sequence and acts as a traffic cop to manage the overall system. GPUs often deal directly with peripherals and storage.

One reason for unifying the environment is the dominance of PCI Express (PCIe) and the standards built on top of it, such as Non-Volatile Memory Express (NVMe) storage and RDMA over Converted Ethernet (RoCE).

2D, 3D, and Ray Tracing

3D displays are one thing, but these days 2D displays still dominate this space. Mapping 3D environments to 2D displays is one of the major chores performed by GPUs, driven greatly by 3D gaming.

Initially, 3D to 2D mapping was done without taking ray tracing into account because of the overhead required. Graphic artists handle textures and lighting explicitly to mimic how a scene would be rendered in the real world.

Explicit transformations done by artists and designers provide the fastest but least realistic way of turning a 3D gaming world into a 2D presentation. Ray casting is a methodology that's fast and provides more realistic rendering based on a 3D model, while ray tracing matches a real-world scenario only limited by the resolution of the system. Check out Jon Peddie's "What's the Difference Between Ray Tracing, Ray Casting, and Ray Charles?" article on *www.electronicdesign.com* if you get a chance.

Ray-tracing hardware, now available from a number of sources, is found in very-high-end GPGPUs. This includes embedded GPGPUs like those intended for mobile devices (e.g., smartphones). These provide amazingly realistic and beautiful images.

Imagination Technologies identified five levels of ray-tracing support from software through low-level hardware support to full hardware ray-tracing support with a scene hierarchy generator.

What makes the ray-tracing support interesting these days is that it's being done in real-time. Ray-tracing support initially used in movie computer generated imagery (CGI) special effects often required days and a supercomputer of that time to render part of a video clip. This is now being accomplished on a single GPGPU card at 4K resolution in real-time and will eventually be done on a mobile device.

Not all graphical display applications require this level of performance or sophistication, but computer image and video content continue to dazzle users. Expect the future to make things even more interesting. ed

To view more images associated with this article, go to the online version at www.electronicdesign.com.

(Continued from page 21)

out "grooves" in the die to support passive attachment of fiber optics—up to 16 fibers in the case of co-packaged optics—increasing bandwidth density.

AI at Speed of Light

As the decline of Moore's Law takes a toll on the technology industry, other companies are trying to push silicon photonics even deeper into data centers.

Today, data dashes through optical fiber in data centers before it slows to a crawl at copper interconnects. These bottlenecks occur at copper pins and wires on circuit boards. As a result, major semiconductor firms and startups are setting their sights on using silicon photonics to transfer data over shorter distances, such as between CPUs, GPUs, and other computer chips in a server or even on a circuit board (PCB).

NVIDIA indicated that it's designing high-bandwidth, low-latency, power-efficient optical interconnects based on GF Fotonix into some of its "leading-edge" data-center systems to handle increasingly heavy AI workloads.

Yu said one advantage of GF Fotonix is that the monolithic architecture reduces the rate that errors occur in data transmission, lowering latency by a factor of 10, which in turn translates to higher throughput for AI workloads.

To make things easier for itself, NVIDIA has partnered with Ayar Labs, a startup designing optical interconnects that can be bundled into various processors and accelerators. The interconnects come in the form of a chiplet that can be packaged into everything from CPUs to GPUs to supply up to 1,000X more bandwidth than electrical I/O—using one-tenth of the power.

CEO Charles Wuischpard said it partnered closely with GlobalFoundries to integrate its unique requirements into GF Fotonix. Ayar Labs was also the first company to build a prototype on top of the platform.

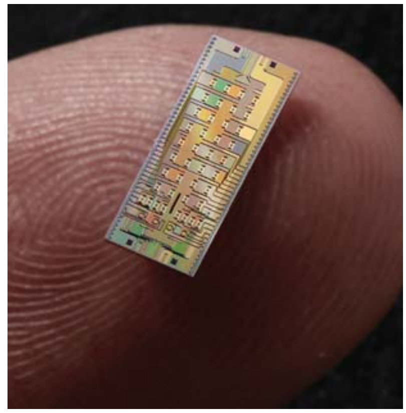
GF Fotonix sets the stage for the company, which has also partnered with HPE to design a future generation of its Slingshot interconnect for high-performance computing, to supply thousands of units of its chiplet this year.

If You Build It, They Will Come

Yu said the collaboration with its current roster of customers helped it to create a silicon-photonics platform for everyone else to use, even in areas such as telecom, aerospace, defense, and automotive.

Lightmatter, a startup using silicon photonics to accelerate AI workloads in data centers, as well as improve energy efficiency, also plans to use the GF Fotonix platform for its first accelerator chip due out in 2022. Said Lightmatter CEO Nicholas Harris, "Together we're changing the way the world thinks about photonics."

PsiQuantum is building out a quantum computer called Q1 with help from GlobalFoundries. Photons are used to solve problems many millions of times faster, and even carry out computations that are impossible today. But prospects for Q1 and other systems to change the world remains years out.



Xanadu, a startup relying on silicon photonics for quantum computing, has also backed GF Fotonix. *Credit: Xanadu*

To lend a helping hand to current and future customers, GlobalFoundries is building out a more vibrant ecosystem of software tools, support, and services around the GF Fotonix platform. Ansys, Cadence Design Systems, and Synopsys are offering suites of electronic design tools that support photonics-based chips and chiplets.

"I don't want to say it's one-size-fits-all," said Yu. "It's not that simple. But with one foundry platform with differentiated features and using unique materials, you can open up photonics to a variety of markets."

GlobalFoundries plans to complete qualification of GF Fotonix to support a production ramp-up by 2024. ed

Ad Index

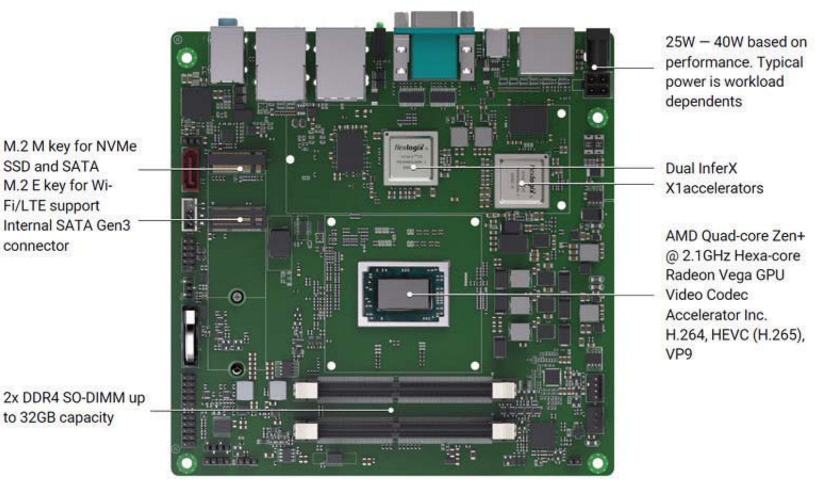
WILLIAM G. WONG | Senior Content Director

Mini-ITX Motherboard Packs Two AI Accelerator Chips

Flex Logix's InferX Hawk motherboard is designed for applications such as machine-learning-based video systems that handle tasks like object detection.

I'm looking forward to checking out the InferX Hawk from Flex Logix *(see figure)*. The Mini-ITX motherboard has a quad-core, AMD Zen+ host processor with a Radeon Vega GPU and video codec accelerator. However, the machine-learning (ML) heavy lifting is done by a pair of Flex Logix InferX X1 artificial-intelligence (AI) accelerators that can be used individually or in tandem. It brings ML/AI acceleration to the edge in a complete package.

The InferX X1 chip utilizes multiple single-dimensional Tensor processors that can be configured on the fly. It supports a high-precision Winograd acceleration option that increases system utilization and overall efficiency. The reconfiguration feature allows for support of two- and three-dimensional tensors.



The InferX Hawk sports a quad-core, AMD Zen+ processor with a Radeon Vega GPU to support twin InferX X1 machine-learning accelerator chips. *Credit: Flex Logix*

The InferX X1 chip utilizes multiple single-dimensional Tensor processors that can be configured on the fly.

The chip also incorporates embedded FPGA (eFPGA) logic, opening the door to further customization. With the on-chip memory and reconfigurable fabric, intermediate results can be kept locally instead of bringing off-chip memory into play. The system is able to handle standard conv2d models as well as depth-wise conv2d models. Each chip can manage multiple models depending on their size, or the two chips may be combined to deal with one large model.

Almost any ML model works with the InferX X1 chips, but video processing is an area where it excels. This includes applications such as object recognition within a video stream. Video streams are normally provided via USB or gigabit Ethernet (GbE). Taking this into consideration, the AMD processor was paired with the two InferX X1 chips to balance the overall system throughput and take advantage of AMD's video-processing hardware capabilities.

The motherboard supports up to 32 GB of DDR4 SO-DIMM memory. There's an M.2 M key and an M.2 E key socket—the former targets NVMe SSDs, while the latter supports smaller peripherals such as a wireless adapter. The backplane features a pair of DisplayPort sockets, two GbE sockets, two serial ports, plus five USB ports for USB 2.0, 3.1, and 3.2 (Type-C) connections. The system requires from 25 to 40 W of power depending on the workload.

The system is supported by Flex Logix's EasyVision, which includes pretrained and configured models for video applications such as face-mask and license-plate identification. In addition, it can check for PCB defects and perform battery inspection. The EasyVision models work with the InferX Model Development Kit (MDK) and the Run Time Kit (RTK). ed