Electronic Design®

THE AUTHORITY ON
EMERGING TECHNOLOGIES
FOR DESIGN SOLUTIONS

JULY/AUGUST 2019
electronicdesign.com

POWER PLAY

From semiconductors to EV charging, advances in power tech evolve to meet today's demands

$10.00

IN THIS ISSUE

FEATURES

COLUMNS & DEPARTMENTS

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) is published bi-monthly in Jan/Feb, Mar/Apr, May/June, July/Aug, Sept/Oct and Nov/Dec by Informa Media Inc., 9800 Metcalf Ave., Overland Park, KS 66212-2216. Paid rates for a one-year subscription are as follows: $60 U.S., $90 Canada, $120 International. Periodicals postage paid at Kansas City, MO, and additional mailing offices. Editorial and advertising addresses: ELECTRONIC DESIGN, 605 Third Avenue, New York, NY 10158. Telephone (212) 204-4200. 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 2100, Skokie, IL 60076. POSTMASTER: Send change of address to Electronic Design, PO Box 2100, Skokie, IL 60076. Canadian GST #R126431964.

Altech® Printed Circuit Board TERMINAL BLOCKS with PUSH-IN TECHNOLOGY

- Time Saving *Push-In Connection*
- Easy wire insertion - No tool required !

PUSH-IN PCB TERMINAL BLOCKS

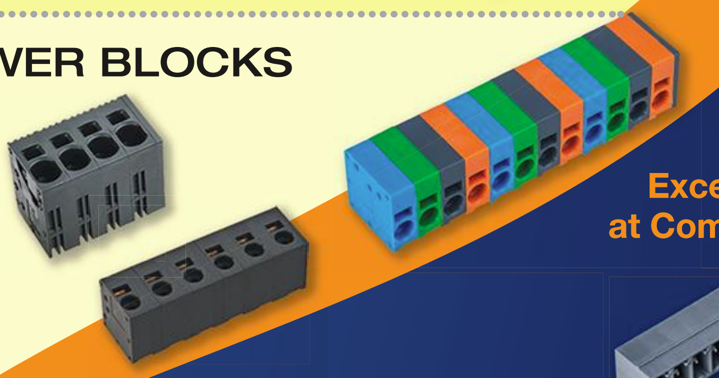
FEATURES

- 2.50 / 2.54mm micro terminals
- Vertical / horizontal types
- Lever version available
- Double row pins for high stability

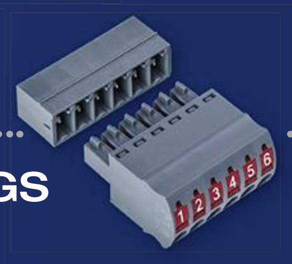
PUSH-IN POWER BLOCKS

FEATURES

- Rated current:
  20A **(AK4200)**
  35A **(AK4210)**
- Double row pins
- Many colors allow easy identification



Excellent Quality
at Competitive Prices



PUSH-IN PLUGS

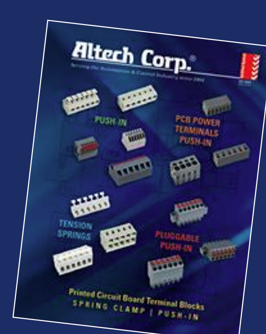
FEATURES

- Ergonomic grip design for quick and easy mounting
- 3.50, 3.81, 5.00 and 5.08mm pin spacing
- Screw flange version for additional stability
- Imprinting and coding available

For a FREE Demo Kit
visit altechcorp.com/PCBPushin

c US

VDE IEC



Altech Corp.®
Your Source for Automation & Control Components

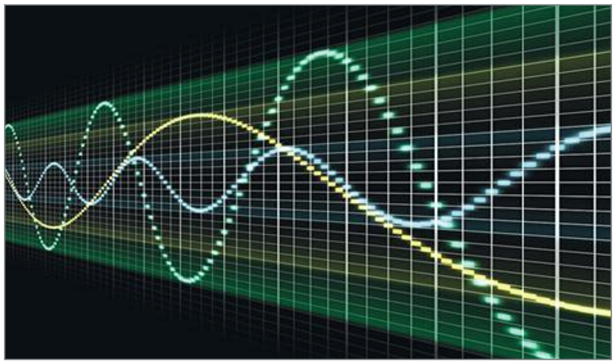
AltechCorp.com/PCBPushin
908.806.9400
for all the details & catalog

ON ELECTRONICDESIGN.COM

Tempest in a Cuisinart Tea Kettle

The tea kettle has a great design and solid assembly, yet it still caught fire. Paul Rako tore it down to find out what the heck went wrong.

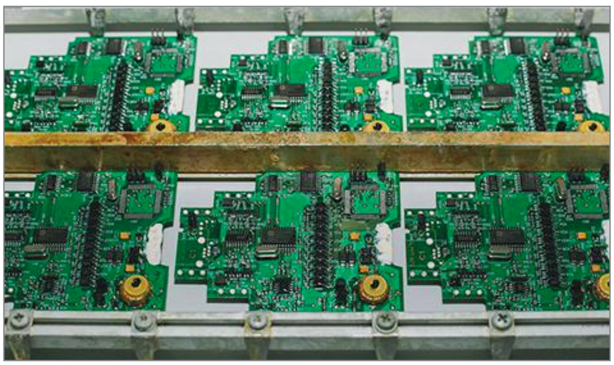
https://www.electronicdesign.com/analog/tempest-cuisinart-tea-kettle



Sending Data Over Sound: How and Why?

Demand grows for better IoT connectivity, with the need for the most basic of smart devices to be able to process real-time audio and perform intelligent DSP. One emerging solution is sending data over sound.

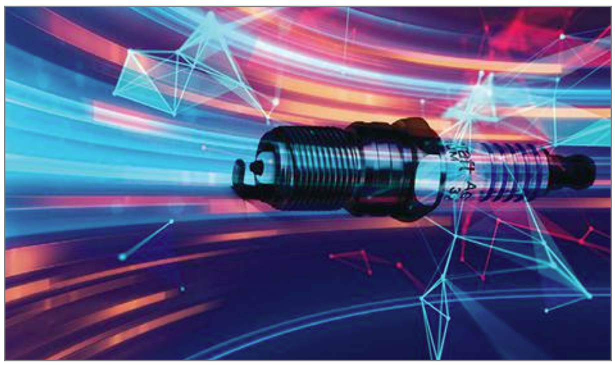
https://www.electronicdesign.com/industrial-automation/sending-data-over-sound-how-and-why



The Economics of ASICs: At What Point Does a Custom SoC Become Viable?

ASICs aren't the preserve of only the richest in Silicon Valley. EnSilica's Ian Lankshear looks at the economics behind developing an ASIC and how to keep costs to a minimum.

https://www.electronicdesign.com/embedded-revolution/economics-asics-what-point-does-custom-soc-become-viable



Forget the Spark Plug: Use a Laser to Ignite ICE Gasoline

A research team replaced the conventional spark-plug ignition of the gasoline engine with a laser-based system, yielding improvements in engine performance and reduced emissions.

https://www.electronicdesign.com/automotive/forget-spark-plug-use-laser-ignite-ice-gasoline

ElectronicDesign.

JULYL/AUGUST 2019

EDITORIAL

GROUP CONTENT DIRECTOR: **KAREN FIELD** karen.field@informa.com
SENIOR CONTENT DIRECTOR: **BILL WONG** bill.wong@informa.com
SENIOR STAFF WRITER: **JAMES MORRA** james.morra@informa.com
TECHNICAL EDITOR: **CHRIS DeMARTINO** chris.demartino@informa.com
CONTRIBUTING EDITOR: **LOUIS E. FRENZEL** lou.frenzel@informa.com
ASSOCIATE EDITOR/COMMUNITY MANAGER: **ROGER ENGELKE** roger.engelke@informa.com

ART DEPARTMENT

GROUP DESIGN DIRECTOR: **ANTHONY VITOLO** tony.vitolo@informa.com
CONTENT DESIGN SPECIALIST: **JOCELYN HARTZOG** jocelyn.hartzog@informa.com
CONTENT & DESIGN PRODUCTION MANAGER: **JULIE JANTZER-WARD** julie.jantzer-ward@informa.com

PRODUCTION

GROUP PRODUCTION MANAGER: **GREG ARAUJO** greg.araujo@informa.com
PRODUCTION MANAGER: **SUSAN POSKIN** susan.poskin@informa.com

AUDIENCE MARKETING

USER MARKETING MANAGER: **DEBBIE BRADY** debbie.brady@informa.com

FREE SUBSCRIPTION / STATUS OF SUBSCRIPTION / ADDRESS CHANGE / MISSING BACK ISSUES
OMEDA **T** | 847.513.6022 **TOLL FREE** | 866.505.7173

SALES & MARKETING

MANAGING DIRECTOR: **TRACY SMITH** **T** | 913.967.1324 **F** | 913.514.6881 tracy.smith@informa.com

REGIONAL SALES REPRESENTATIVES:

AZ, NM, TX: **GREGORY MONTGOMERY** **T** | 480.254.5540 gregory.montgomery@informa.com

AL, AR, SOUTHERN CA, CO, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MI, MN, MO, MS, MT, NC, ND, NE, NV, OH, OK, SC, SD, TN, UT, VA, WI, WV, WY, CENTRAL CANADA: **JAMIE ALLEN** **T** | 415.608.1959 **F** | 913.514.3667 jamie.allen@informa.com

CT, DE, MA, MD, ME, NH, NJ, NY, PA, RI, VT, EASTERN CANADA:
ELIZABETH ELDRIDGE **T** | 917-789-3012 elizabeth.eldridge@informa.com

AK, NORTHERN CA, OR, WA, WESTERN CANADA: **STUART BOWEN** **T** | 425.681.4395 stuart.bowen@informa.com

INTERNATIONAL SALES:

GERMANY, AUSTRIA, SWITZERLAND: **CHRISTIAN HOELSCHER** **T** | 011.49.89.95002778
christian.hoelscher@husonmedia.com

BELGIUM, NETHERLANDS, LUXEMBURG, UNITED KINGDOM, SCANDINAVIA, FRANCE, SPAIN, PORTUGAL:
JAMES RHOADES-BROWN **T** | +011 44 1932 564999 **M** | +011 44 1932 564998 james.rhoadesbrown@husonmedia.co.uk

ITALY: **DIEGO CASIRAGHI** diego@casiraghi-adv.com

PAN-ASIA: **HELEN LAI** **T** | 886 2-2727 7799 helen@twoway-com.com

PAN-ASIA: **CHARLES LIU** **T** | 886 2-2727 7799 liu@twoway-com.com

PLEASE SEND INSERTION ORDERS TO: orders@informa.com

INFORMA REPRINTS: **WRIGHT'S MEDIA** **T** | 877.652.5295

LIST RENTALS/ SMARTREACH CLIENT SERVICES MANAGER: **MARY RALICKI** **T** | 212.204.4284 mary.ralicki@informa.com

DIGITAL

GROUP DIGITAL DIRECTOR: **RYAN MALEC** ryan.malec@informa.com

DESIGN ENGINEERING & SOURCING GROUP

EXECUTIVE DIRECTOR, CONTENT: **KAREN FIELD** karen.field@informa.com
VP OF MARKETING: **JACQUIE NIEMIEC** jacquie.niemiec@informa.com

INFORMA MEDIA INC.

605 THIRD AVENUE
NEW YORK, NY 10158 USA **T** | 212.204.4200

informa

Electronic Design | Machine Design | Microwaves & RF | Source ESB | Hydraulics & Pneumatics |
Global Purchasing | Distribution Resource | Power Electronics | Defense Electronics

the accuracy of *Rubidium*...

PRS10 Rubidium Oscillator (10 MHz)

- Less than 5×10^{-11} aging per month
- Ultra low phase noise (-130 dBc/Hz @ 10 Hz)
- 20 year lamp life
- 1 pps input and output
- RS-232 computer interface

$1495 (U.S. list)

FS725 Benchtop Rubidium Frequency Standard

- 5 MHz and 10 MHz outputs
- 0.005 ppm aging over 20 years
- Built-in distribution amplifier (up to 22 outputs)
- 1 pps input and output
- RS-232 computer interface

$2695 (U.S. list)

SRS rubidium frequency standards have excellent aging characteristics, extremely low phase noise and outstanding reliability.

The PRS10 component rubidium oscillator is designed for easy system integration. It has a 1 pps input for phase-locking to an external reference (like GPS) and provides 72 hour Stratum 1 level holdover.

The FS725 benchtop instrument is ideal for the metrology laboratory as well as the R&D facility – anywhere precision frequency is required. It generates 5 MHz and 10 MHz signals and has a built-in distribution amplifier with up to 22 outputs.

Stanford Research Systems
1290-D Reamwood Ave. Sunnyvale, CA 94089 · email: info@thinkSRS.com
Phone (408) 744-9040 · Fax (408) 744-9049 · **www.thinkSRS.com**

Editorial

WILLIAM WONG | Senior Content Director
bill.wong@informa.com

Have We Learned Machine Learning Yet?

Bill Wong takes a look at the challenges and payoffs of machine learning in embedded applications.

Machine learning (ML) isn't a replacement for conventional programming, but it's one more tool that developers have available when creating solutions. Unfortunately, the scope and variations of ML are complex and changing continually as designers refine existing methodologies and develop new ones. This would be less of an issue if the changes weren't occurring so frequently and if ML had less of an impact. Often, ML is the edge that companies need to stay ahead of the competition.

New models and approaches are being created in the neural-network space where most of the latest ML technology is found, even as hardware designers rush to incorporate ML accelerators in their latest releases. Few processor, GPU, or networking chips lack some sort of ML acceleration these days. These can often provide increased speeds on the order of multiple magnitudes, making many applications possible.

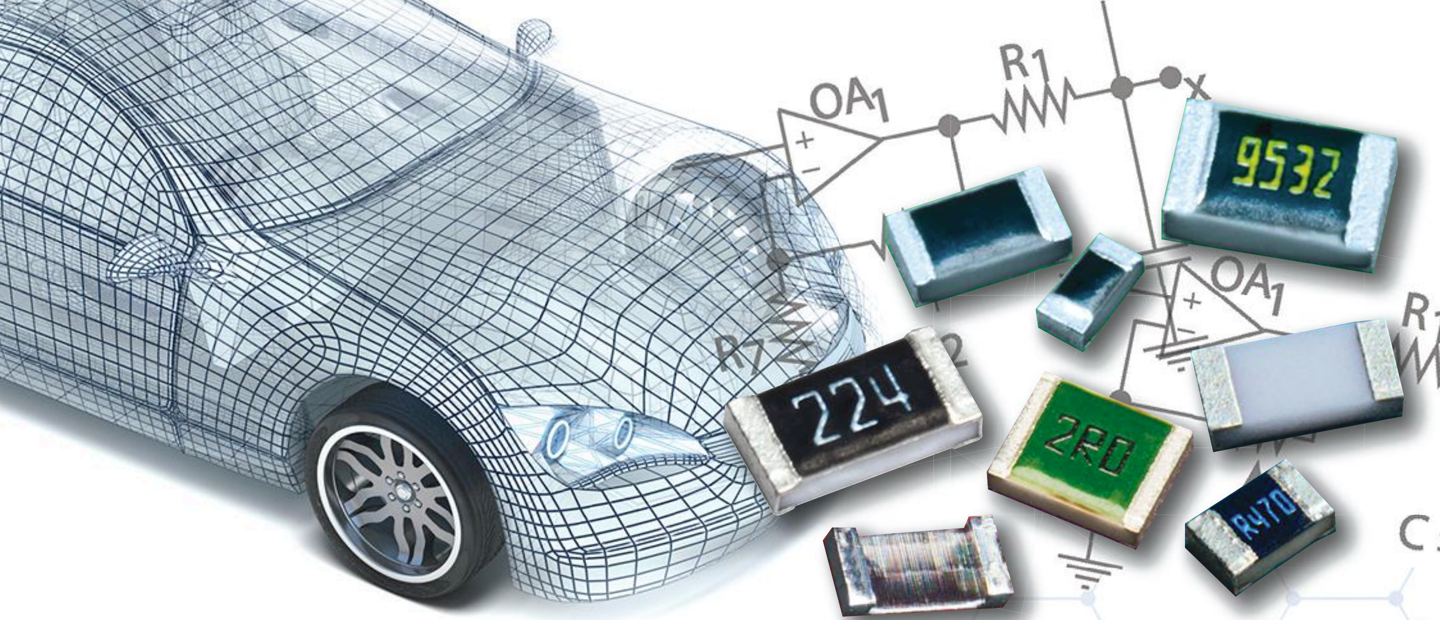
The advantage of the plethora of options is that tools like Xnor's AI2Go and H2O.ai's H20 are allowing developers to choose ML models based on applications and criteria. These can be quickly incorporated into an application. They frequently provide services to train models for a particular application, too.

Likewise, combining ML models can often be similar to combining filters in DSP applications. Sometimes it's as simple as connecting a few blocks together sequentially, but other times the interconnections and interactions can be as complex as the ML models themselves. More applications are utilizing multiple models rather than a single model that addresses just one aspect of the application.

Also, ML is being applied to all aspects of application design, deployment, and maintenance in many cases. ML support is part of most cloud services and management systems these days. ML models are often employed in apps and cloud services given the growth of IoT in embedded applications.

ML tools and models aren't a panacea for embedded developers. They're also not a solution for all applications. In fact, many embedded and IoT applications will not benefit incorporating ML support. By the same token, ML can be a time and money sink with little payoff, especially when the technology is misunderstood or not applicable.

It's a good idea for developers and managers to get more educated about ML and the options available. Therefore, they can determine whether or when it might be applicable to your needs. Remember, neural networks are all about probabilities, so their relevance isn't always a binary choice.



From
Concept to Reality

Bring Your Designs to Life with our NEW Passive Components

Thin Film Resistor for Automotive RN73H
- Improved moisture resistance by special protective coating
- High precision tolerances ±0.05% ~ ± 1%
- High performance TCR ±5 ~ ±100ppm/°C

Surge & Pulse Precision Resistors SG73P/S
- Pulse withstanding; down to ±0.5% tolerance
- Endures the ESD limiting voltage
- Resistance range: 1 ~ 10MΩ

High Voltage Resistor for Automotive HV73V
- Maximum working voltages from 350V ~ 800V
- Resistance range: 10K ~ 51MΩ
- Tolerances: ±0.5% ~ ±5%

Current Sense Metal Plate Resistor TLR
- Power ratings from 0.2W ~ 3W
- Resistance range: 0.5m ~ 20mΩ
- Ultra low TCR (±50ppm/°C) available

Anti-Sulfur Resistors -RT
- Broad family of resistors with excellent anti-sulfur characteristics
- Passes ASTM-809 anti-sulfuration testing
- Excellent heat resistance and environmental resistance

To make your design a reality, check out our NEW products!
www.KOASpeer.com

ISO 9001:2015 CERTIFIED
IATF16949:2016 CERTIFIED

KOA®
KOA SPEER ELECTRONICS, INC.

SOLUTIONS · SERVICE · DELIVERY · Quality 1st

News

NXP'S NEW i.MX 8 UNITES with Microsoft's Azure Sphere

These days, most cloud providers are delivering end-to-end edge computing platforms for the Internet of Things (IoT). Among them is Microsoft's Azure Sphere, which runs on the Azure cloud platform. The Azure Sphere OS is a customized version of Linux designed to work with processors that implement Microsoft's Pluton security hardware subsystem.

NXP and Microsoft announced a partnership around Azure Sphere at this year's NXP Connect conference. NXP will supply the edge-node hardware while Microsoft provides the hooks to the cloud, including the custom, Linux-based Azure Sphere OS *(see figure)*.

The i.MX 8 chip will not be shipping in quantity until next year. It was designed by engineering teams from each company and will be built using fully depleted silicon-on-insulator (FD-SOI) technology to minimize power requirements. There will be single and dual core versions based on the Arm Cortex-A35; a Cortex-M33 core handles real-time chores. Both support Arm's TrustZone security as well. An independent audio/video processing domain is powered by a high-performance HiFi4 DSP core. The chip will be part of NXP's new Edge Verse family.

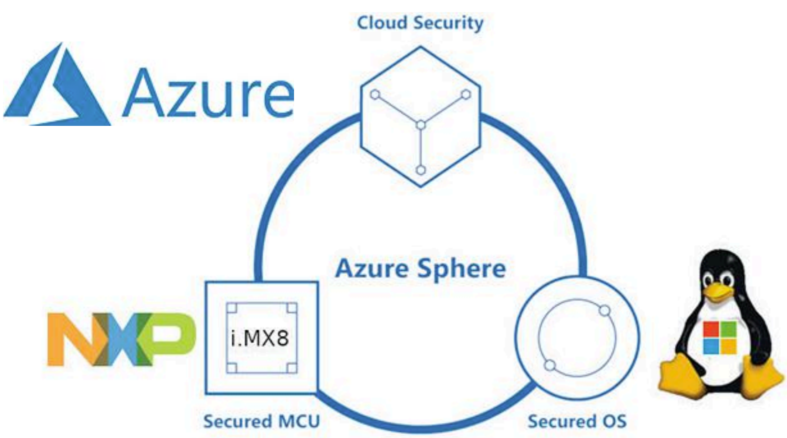
Azure Sphere OS is designed to have a small memory footprint. It runs on the Cortex-A while an RTOS will typically run on the Cortex-M33. Microsoft just picked up Express Logic, so its ThreadX RTOS could be an option.

"NXP's collaboration with Microsoft is yet another step in our ongoing commitment to bring complete security solutions to our customers," said Joe Yu, vice president of the low-power applications processors product line at NXP. "With this Azure Sphere-certified applications processor, customers can build purposeful edge products using the energy efficiency and multifaceted capabilities of the i.MX 8 series, and enjoy the peace of mind that comes from knowing that their products are protected in the field by Azure Sphere security service."

"At a time when the opportunity of innovation is limited only by imagination, security is a persistent challenge. Our collaboration with NXP enables our partners to fully realize the opportunity in front of them by delivering intelligent security that is responsive and always learning," said Galen Hunt, Distinguished Engineer and Managing Director, Microsoft Azure Sphere. "Together with the performance and flexibility of NXP's i.MX application processors, we will help our partners transform their products and the way they service and interact with their customers. This collaboration allows device manufacturers across various industries to achieve more."

Microsoft's Azure Sphere OS boots from the Pluton's root-of-trust. The hardware has its own random-number generator, crypto acceleration, and key management and storage. Each chip has its own private key and Azure's public key installed when the chip is built. It supports remote attestation; the OS and framework also support over-the-air (OTA) updates, with Microsoft delivering Azure Sphere OS updates along with application updates all under company control.

Applications that run on the Azure Sphere OS have a secure and authenticated communications link with the Azure Sphere cloud. The Azure cloud supports applications as well as provides management services for the edge devices. ■



NXP's new i.MX 8 will incorporate Microsoft's Pluton security hardware designed to support the Azure Sphere OS that ties into the Azure cloud.

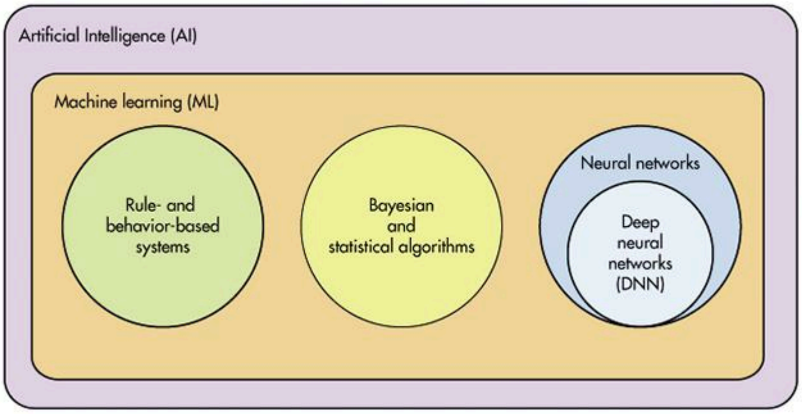

RACING OUR WAY TO THE
MOON
MPD
MEMORY PROTECTION DEVICES
BEST SOURCE FOR LITHIUM COIN CELL RETAINERS AND BATTERY CAGES
WWW.BATTERYHOLDERS.COM

WILLIAM WONG | Senior Technology Editor

Machine Learning: THE MAGIC IS HOW IT WORKS

No overarching artificial intelligence looms on the horizon, but machine-learning tools can make applications do some magical things.

I was talking with a friend recently about artificial intelligence (AI) and machine learning (ML), and they noted that if you replaced AI or ML with the word magic, many of those discussions would be as useful and informative as before. This is due to a number of factors, including misunderstanding about the current state of affairs when it comes to AI, ML, and more specifically, deep neural networks (DNNs)—specifically, what ML models are actually doing and not comprehending how ML models are used together.



1. Neural networks are just a part of the machine-learning portion of artificial-intelligence research.

I hope that those who have been working with ML take kindly to my explanations, because they're targeted at engineers who want to understand and use ML but haven't gotten through the hype that even ML companies are spouting. More than half of you are looking into ML, but only a fraction is actually incorporating it into products. This number is growing rapidly though.

ML is only a part of the AI field and many ML tools and models are available, being used now, and in development *(Fig. 1)*. DNNs are just a part; other neural-network approaches enter into the mix, but more on that later.

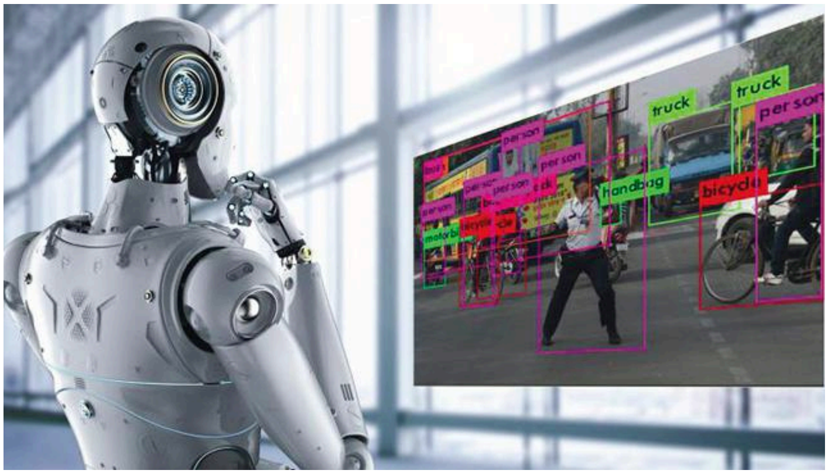
Developers should look at ML models more like fast Fourier transforms (FFTs) or Kalman filters. They're building blocks that perform a particular function well and can be combined with similar tools, modules, or models to solve a problem. The idea of stringing black boxes together is appropriate. The difference between an FFT and a DNN model is in the configuration. The former has a few parameters while DNN model needs to be trained.

Training for some types of neural networks requires thousands of samples, such as photos. This is often done in the cloud, where large amounts of storage and computation power can be applied. Trained models can then be used in the field, since they normally require less storage and computation power as their training counterparts. AI accelerators can be utilized in both instances to improve performance and reduce power requirements.

ROLLING A MACHINE-LEARNING MODEL

Most ML models can be trained to provide different results using a different set of training samples. For example, a collection of cat photos can be used with some models to help identify cats.

Models can perform different functions such as detection, classification and segmentation. These are common chores for image-based tools. Other functions could include path optimization or anomaly detection, or provide recommendations.



2. Different tools or ML models can be used to identify areas of interest that are then isolated and processed to distinguish between objects such as people and cars.

A single model will not typically deliver all of the processing need in most applications, and input and output data may benefit from additional processing. For example, noise reduction may be useful for audio input to a model. The noise reduction may be provided by conventional analog or digital filters or there may be an ML model in the mix. The output could then be used to recognize phonemes, words, etc., as the data is massaged until a voice command is potentially recognized.

Likewise, a model or filter might be used to identify an area of interest in an image. This subset could then be presented to the ML-based identification subsystem and so on *(Fig. 2)*. The level of detail will depend on the application. For example, a video-based door-opening system may need to differentiate between people and animals as well as the direction of movement so that the door only opens when a person is moving toward it.

Models may be custom-built and pretrained, or created and trained by a developer. Much will depend on the requirements and goals of the application. For example, keeping a machine running may mean tracking the operation of the electric motor in the system. A number of factors can be recorded and analyzed from power provided to the motor to noise and vibration information.

Companies such as H2O.ai and XNor are providing prebuilt or customized models and training for those who don't want to start from scratch or use open-source models that may require integration and customization. H2O.ai has packages like Enterprise Steam and Enterprise Puddle that target specific platforms and services. XNor's AI2Go uses a menu-style approach: Developers start by choosing a target platform, like a Raspberry Pi, then an industry, like automotive, and then a use case, such as In-cabin object classification. The final step is to select a model based on latency and memory footprint limitations *(Fig. 3)*.

IT'S NOT ALL ABOUT DNNs

Developers need to keep in mind a number of factors when dealing with neural networks and similar technologies. Probability is involved and results from an ML model are typically defined in percentages. For example, a model trained to recognize cats and dogs may be able to provide a high level of confidence that an image contains a dog or a cat. The level may be lower distinguishing a dog from a cat and so on, to the point that a particular breed of animal is recognized.

THINK INSIDE THE BOX

NEMA RATED ENCLOSURES

Full Line | Same-Day Shipping | Rugged NEMA Enclosures

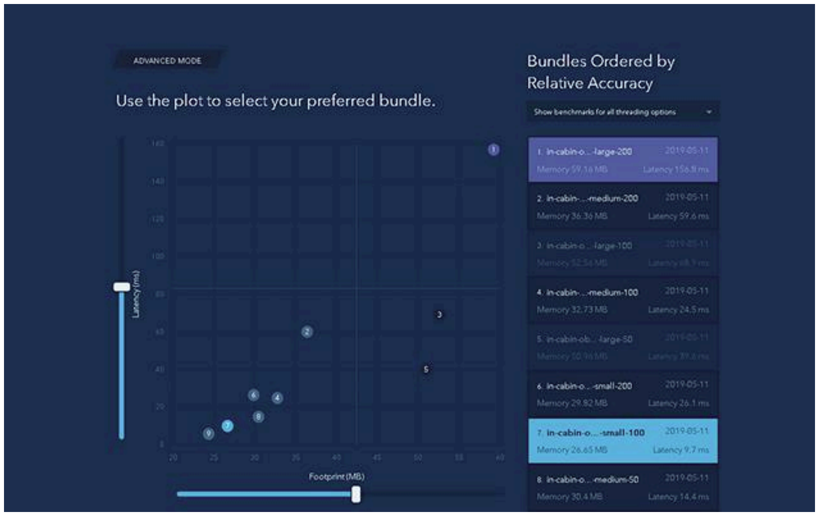
Transtector Systems NEW TEF14 Series Includes:

- Wide range of configurations - 120 VAC, 240 VAC, 12 VDC, PoE-ready and non-powered enclosures
- 45 enclosure configurations and 15 essential accessories
- Weatherproof NEMA design in standard 14" x 12" x 7" sizes
- Rugged construction ideal for indoor and outdoor applications
- Fully gasketed raised lid for secure protection
- Pad lockable stainless steel quick release latches
- UV stabilized polyester reinforced fiberglass construction
- Options include cooling fans, power sources, heaters, insulated, PoE, terminal blocks, DIN Rail, pre-drilled mounting plates, vents and UL listed configuration

The large selection of in-stock TEF14 enclosures is competitively priced to provide high-value performance. Custom modifications are also available. For more information, visit transtector.com or call +1 (208) 635-6400.

TRANSTECTOR
an INFINITe brand

Superior Protection | Patented Technology | Guaranteed Performance



3. Shown is the tail end of the menu selection process for XNor's AI2Go. Developers can narrow the search for the ideal model by specifying the memory footprint and latency time.

The percentages can often improve with additional training, but changes usually aren't linear. It may be easy to hit the 50% mark and 90% might be a good model. However, a lot of training time may be required to hit 99%.

The big question is: "What are the application requirements and what alternatives are there in the decision-making process?" It's one reason why multiple sensors are used when security and safety are important design factors.

DNNs have been popular because of the availability of open-source solutions, including platforms like TensorFlow and Caffe. They have found extensive hardware and software support from the likes of Xilinx, NVIDIA, Intel, and so on, but they're not the only types of neural-network tools available. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and spiking neural networks (SNNs) are some of the other options available.

SNNs are used by BrainChip and Eta Compute. BrainChip's Akida Development Environment (ADE) is designed to support SNN model creation. Eta Compute augments its ultra-low-power Cortex-M3 microcontroller with SNN hardware. SNNs are easier to train than DNNs and their ilk, although there are tradeoffs for all neural-network approaches.

Neurala's Lifelong-DNN (LDNN) is another ML approach that's similar to DNNs with the lower training overhead of SNNs. LDNN is a proprietary system developed over many years. It supports continuous learning using an approximation of lightweight back propagation that allows learning to continue without the need to retain the initial training information. LDNN also requires fewer samples to reach the same level of training as a conventional DNN.

There's a tradeoff in precision and recognition levels compared to a DNN, but such differences are similar to those involving SNNs. It's not possible to make direct comparisons between systems because so many factors are involved, including training time, samples, etc.

LDNN can benefit for AI acceleration provided by general-purpose GPUs (GPGPUs). SNNs are even more lightweight, making them easier to use on microcontrollers. Even so, DNNs can run on microcontrollers and low-end DSPs as long as the models aren't too demanding. Image processing may not be practical, but tracking anomalies on a motor-control system could be feasible.

OVERCOMING ML CHALLENGES

There are numerous challenges when dealing with ML. For example, overfitting is a problem experienced by training-based solutions. This occurs when the models work well with data similar to the training data, but poorly on data that's new. LDNN uses an automatic, threshold-based consolidation system that reduces redundant weight vectors and resets the weights while preserving new, valid outliers.

ML models can address many tasks successfully with high accuracy. However, that doesn't mean all tasks, regardless if they're conventional classification or segmentation problem, can be accommodated. Sometimes changing models can help or developing new ones. This is where data engineers can come in handy, though they tend to be rare and expensive.

Debugging models can also be a challenge. ML module debugging is much different than debugging a conventional program. Debugging models that are working within an application is another issue. Keep in mind that models will often have an accuracy less than 100%; therefore, applications need to be designed to handle these conditions. This is less of an issue for non-critical applications. However, apps like self-driving cars will require redundant, overlapping systems.

AVALANCHE OF ADVANCES

New systems continue to come out of academia and research facilities. For example, "Learning Sensorimotor Control with Neuromorphic Sensors: Toward Hyperdimensional Active Perception" is a paper out of the University of Maryland's engineering department. Anton Mitrokhin and Peter Sutor Jr., Cornelia Fermüller, and Computer Science Professor Yiannis Aloimonos developed a hyperdimensional pipeline for integrating sensor data, ML analysis, and control. It uses its own hyperdimensional memory system.

ML has been progressing like no other programming tool in the past.

> Systems that employ ML aren't magic and their application can use conventional design approaches.

Improvements have been significant even without turning to specialized hardware. Part of this is due to improved software support to optimizations that increase accuracy or performance while reducing hardware requirements. The challenge for developers is determining what hardware to use, what ML tools to use, and how to combine them to address their application.

It's worth making most systems now rather than waiting for the next improvement. Some platforms will be upward-compatible; however, others may not. Going with a hardware-accelerated solution will limit the ML models that can be supported but with significant performance gains, often multiple orders of magnitude.

Systems that employ ML aren't magic and their application can use conventional design approaches. They do require new tools and debugging techniques, so incorporating ML for the first time shouldn't be a task taken lightly. On the other hand, the payback can be significant and ML models may often provide support that's unavailable with conventional programming techniques and frameworks.

As noted, a single ML model may not be what's needed for a particular application. Combining models, filters, and other modules requires an understanding of each, so don't assume it will simply be a matter of choosing an ML model and doing a limited amount of training. That may be adequate in some instances, especially if the application matches an existing model, but don't count on it until you try it out. ed

AHMED EISAWY | Senior Product Marketing Manager, Tektronix
www.tek.com

Probing Methods that Boost Power-Conversion Measurement Accuracy

Accurately evaluating and measuring power-conversion devices to find small performance increases means that highly accurate measurements are of utmost importance.

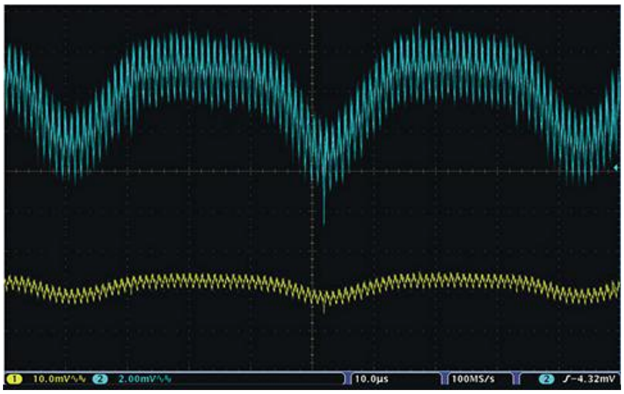
Most oscilloscopes ship with 10X attenuation passive probes because this type of probe is well-suited to making measurements across a broad range of applications. These probes are typically rated from dc to 500 MHz and generally can handle up to a few hundred volts. It's certainly possible to use general-purpose probes for power measurements, but the results will unlikely deliver the level of accuracy needed to drive notable performance improvements in the design.

The primary goal for power-supply designers and test engineers is to find incremental improvements that will increase power-conversion efficiency or reduce losses in the design. This requires the ability to accurately evaluate and measure very small performance increases. The strategy is that enough small improvements will add up to noteworthy improvements.

Using a measurement system with sufficient sensitivity, efficiency losses can be found in nearly every sub-section of a power converter, with key areas of interest often involving switching semiconductors, magnetics, and rectifiers. Performance improvements in the low-digit percent values and fractions of a percent can be meaningful. Accurately evaluating and measuring such small performance increases means that highly accurate measurements are of utmost importance.

SIGNAL SENSITIVITY

General-purpose probes are a must-have for many applications, but there are also times—e.g., many power applications—where something more specialized is needed. Here's why. A common challenge in power-supply design and measurement is to separate noise from ripple voltage. Now suppose we're attempting to probe a 3.3-V power supply using a general-purpose 10X probe. Unfortunately, the 10X probe doesn't provide enough sensitivity to trigger on period noise present in the waveform. These probes are a good choice for many general electronics measurements since they increase the voltage range of the scope and offer relatively high bandwidth.



1. For a 3.3-V supply, a 2X probe (blue trace) proved to be the right choice. The 10X probe is represented by the yellow trace.

However, for measuring low-level signals in the tens of millivolts, a 1X (or 1:1) probe may be a better option, because it doesn't attenuate the signal as much and won't push the signal down into the noise floor of the scope. Of course, the sensitivity benefit is offset by limited bandwidth, usually around 15 MHz. If this bandwidth is insufficient for your measurement, a passive 2X probe may be a better alternative.

In the application represented by the waveforms in *Figure 1*, the 2X probe in fact proved to be the right choice. The yellow trace is a 10X probe adjusted to the lowest vertical setting of 10 mV per division and the blue waveform is a 2X probe. The 2X probe can be adjusted down to its lowest vertical setting of 2 mV per division. Since the output of the power supply produces a signal with 3 mV of ripple, it's clear why a probe with 10X attenuation is not well-suited for this measurement.

DIFFERENTIAL MEASUREMENTS

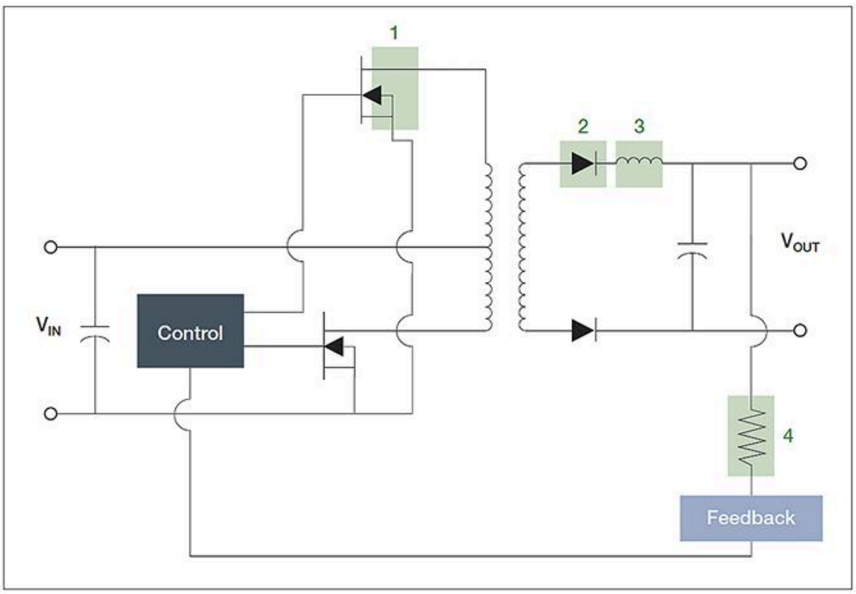
A ripple measurement such as that just discussed is one application where a single-ended (ground-referenced) probe can be used effectively and safely in power-supply design and debug. But many measurements on power conversion need to be made in a floating environment, where reference to earth ground is not available.

Figure 2 identifies several common power-conversion measurements that aren't tied to ground and require differential measurement techniques:

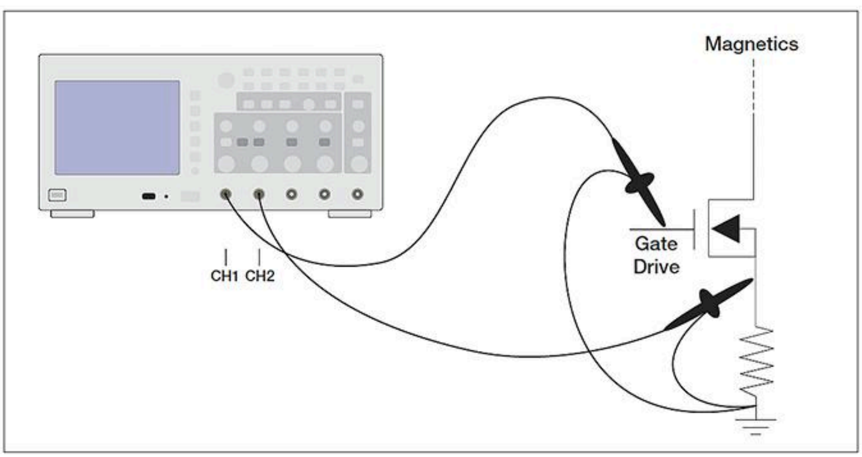
- Drain-to-source voltage (V_{DS}) on a MOSFET
- Diode voltage on a freewheeling diode
- Inductor and transformer voltages
- Voltage drop across ungrounded resistors

There are several ways to perform differential measurements, including:

- Use two single-ended probes and calculate the difference voltage
- Use an oscilloscope with specially designed floating inputs
- Select a differential probe best matched to the measurement



2. Several differential measurement points, as shown here, come into play when testing a push/pull power converter.



3. Quasi-differential measurements can be made using two single-ended probes.

TWO SINGLE-ENDED PROBES

A common technique is to use two single-ended probes with each probe's ground lead tied to earth ground and the tips on either side of the component under test *(Fig. 3)*. The oscilloscope is then set to show the difference between Channel 1 and Channel 2. This is sometimes called "A-B" and it displays the difference voltage between the channels using math in the scope. This technique is occasionally used by engineers when they need to make a differential measurement but don't have the appropriate test equipment available.

There are several problems with this approach. This method will provide good measurement results only when the probes and the oscilloscope channels are very well matched (gain, offset, delay, and frequency response). This method also doesn't provide very good common-mode rejection (nulling out any ac or dc portion of the signal that's common to both inputs). And, if the two signals aren't properly scaled, you run the risk of overdriving the oscilloscope inputs and getting erroneous measurements.

FLOATING INPUTS

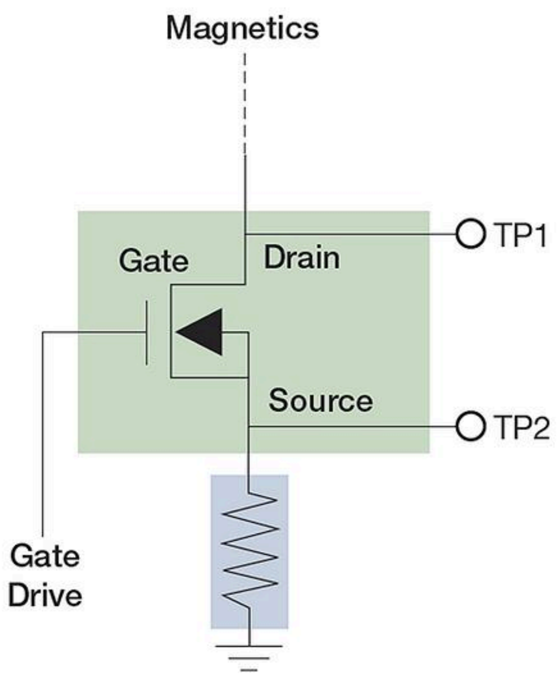
Using a "floating" oscilloscope is another alternative. Each of the input channels of these oscilloscopes is electrically isolated from chassis ground, and then the oscilloscope is powered by its battery. The parasitic capacitance from the oscilloscope chassis to earth ground is also very low.

Together, these isolation characteristics of a floating oscilloscope enable differential measurements with an insulated passive probe. These instruments are convenient, easy to use, and give good results. However, differential voltage probes have lower capacitance and are highly balanced.

MATCHED DIFFERENTIAL PROBE

For the best measurement accuracy, a differential probe with specifications matched to the measurement task is usually the best choice. Differential probes are active devices. They include a purpose-designed differential amplifier in the probe tip that measures only the voltage across the two test points, regardless of the potential between either test point and ground. This greatly simplifies the probing task and eliminates some possible sources of error. And since they measure only the differential voltage, they also can ignore—and null out—common-mode ac swings or dc offset voltages that may be present.

Because measurements in different parts of the device under test (DUT) may have very different requirements, it's important to choose probes carefully. In the example in *Figure 4,* the task at hand is measuring Turn-on Loss, Turn-off Loss, and conduction loss in the MOSFET switching component of a power supply under test. Shown is a simplified diagram of the MOSFET with measurement points, TP1 and TP2.



4. MOSFET Turn-on Loss and Turn-off Loss are measured at measurement points TP1 and TP2.

Testing a universal MOSFET-based power supply designed to be powered from ac line (or "mains") voltage in countries all over the world essentially dictates the use of a differential probe. Let's take a look at the implications for test requirements and equipment.

Input-voltage ratings for this type of device typically range from 80 to 250 V ac or wider. To characterize performance across a variety of input-voltage conditions worldwide, you need to perform not just one measurement, but a series of measurements at several input-voltage levels. This applies to each performance parameter to be tested. Switching characteristics (and therefore losses) can be expected to be different at each of these input-voltage levels, and they may not vary in a linear fashion. This increases not only the total number of measurements to be taken, but also the need for repeatability from measurement to measurement.

With input-supply-voltage levels as high as 250 V ac, the voltage levels between drain and source in the switching MOSFET can be expected to reach 354 V or higher. The probing solution needs to be versatile enough to measure these voltage levels along with much lower levels for some tests.

The power supply under test has a switching rate of 250 kHz. Using the common "5X" rule of thumb for measurement bandwidth, this equates to a measurement bandwidth requirement of 1.25 MHz. However, this is a simplistic view of real-world signal speeds since actual rise times of the switching components can be expected to exceed this by an order of magnitude. The same is possible for spikes, transients, and other noise that may need to be investigated.

If you're measuring signals with rise times in the tens of nanoseconds, the probe should have a rise-time spec in the nanoseconds. For accurate measurements in this example application, the bandwidth of the measurement system should be on the order of 500 MHz or higher.

Resistive and capacitive loading must be considered, too. Important specs for voltage probes also include input impedance, a key to possible interaction with the DUT that can compromise measurement accuracy by loading the circuit under test. Two key probe specifications are input R (resistive impedance) and input C (capacitive impedance.) A high input impedance R value (in ohms) will minimize circuit loading at dc and low frequencies; a low input C value (in picofarads) will minimize circuit loading at higher frequencies, which can affect timing measurements such as rise time.

Another important factor is common-mode rejection ratio (CMRR), which specifies a differential probe's ability to reject any signal that's common to both test points in a differential measurement. In a perfect world, a probe's CMRR would be infinite. However, most differential probes have limitations in this regard. CMRR in dB describes the ratio of signal "reject-

Broadcom Optocouplers

Serving the Industrial Market For More Than 40 years

Optocouplers are the only isolation devices that meet or exceed the IEC 60747-5-5 International Safety Standard for insulation and isolation.

Stringent evaluation tests show Broadcom optocouplers deliver outstanding performance on essential safety and deliver exceptional High Voltage protection for your equipment. Alternative isolation technologies such as magnetic or capacitive isolators do not deliver anywhere near the high voltage insulation protection or noise isolation capabilities that optocouplers deliver.

IEC 60747-5-5 Certified

BROADCOM®
connecting everything®

ed" by the probe to the amount that leaks through and shows up in the measurement. Note that for most differential probes, CMRR declines with frequency.

STEPPING UP TO GaN AND SiC

For some applications, the measurement requirements may exceed the capabilities of even the best conventional voltage probes. In particular, semiconductors based on gallium nitride (GaN) or silicon carbide (SiC) create new challenges for measurement.

For example, switching devices based on GaN offer smaller chip size, low on-state resistance, and the ability to operate at ultra-high frequencies (> 1 MHz) compared to traditional silicon devices. But these desirable performance improvements also create new demands for successful probing and measurement.

The combination of high voltage and higher switching frequency requires a probing solution with extremely good CMRR at high frequency, especially when making measurements such as gate-source voltage on the high-side semiconductor. However, because of their electrical connection to the DUT, conventional differential probes lose CMRR performance as frequency increases.

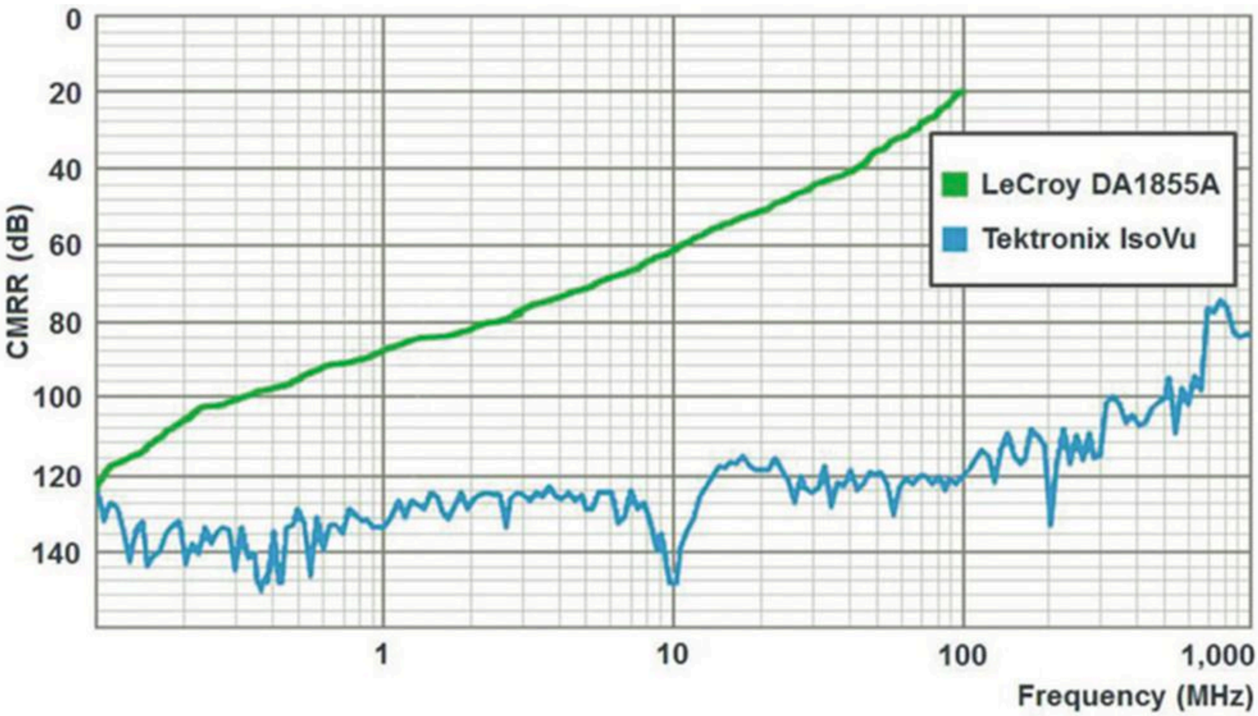
Consequently, the industry is increasingly turning to probes that use electro-optic sensors to convert the input signal to optical. This approach keeps CMRR from derating as frequency increases, opening the door to measurements on GaN or SiC devices that were previously impossible. This is shown in *Figure 5* where an optical differential probe provides near flat CMRR as frequency increases.

POWER INTEGRITY

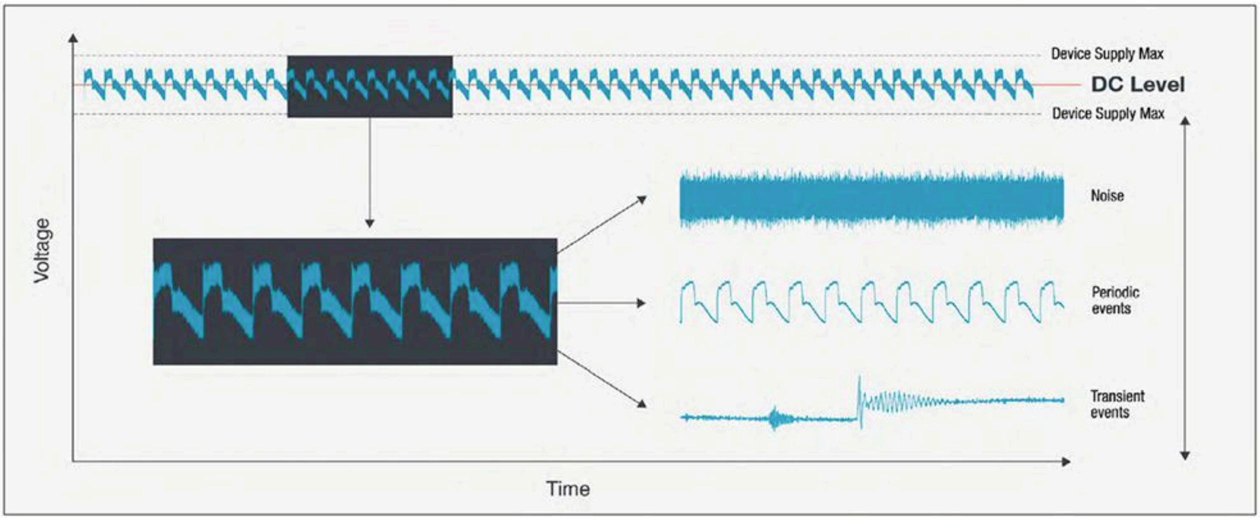
Most of today's electronic designs require many different supply voltages to function properly. In fact, many components within a given circuit require multiple voltages. This is especially true with highly integrated system on chip and microprocessor designs in which multiple technologies interface together.

Performing dc power-rail measurements is becoming increasingly difficult due to several factors, including power-efficiency features like power gating and dynamic voltage and frequency scaling, dynamic loads with fast transients, increased crosstalk, and coupling and switching regulators with faster rise times. Connectivity can also be challenging since it must have low inductance paths to ground and minimum effective capacitance to reduce ringing and provide the most bandwidth.

Many designs have a bulk supply voltage that filters down through various dc-dc converters to the needed supply voltages required by the various ICs and systems. It's often the case that the bulk supply voltage is many times higher than the voltage needed by ICs. Because of this, it's important to choose a probe that offers enough offset to look at all of the rails being tested in a power-delivery network.



5. Isolated optical probes overcome the problem of CMRR derating with frequency.



6. Components of dc power-supply noise including ac noise and coupling.

In addition, it's important to look at each dc line to see if the power supplied is within the tolerance band of a target system or device. This includes the nominal dc value of the line, as well as any ac noise or coupling present, as displayed in *Figure 6*. All of these noise sources impact the quality of power that reaches a device, and it's important to reduce these noise sources to the point that the target device can operate properly.

However, prior to minimizing these noise sources, they need to be accurately measured. Looking at many power-delivery designs, it might seem that measurement system bandwidths of a few tens of MHz are enough. Most switching designs are switching in the hundreds of kHz up to perhaps a few MHz. Larger physical designs and devices that run off higher supply voltages were less sensitive to noise. So, noise content above 20 MHz was rarely a concern.

Now, as design sizes and supply voltages shrink, tolerances follow. Power-distribution networks are being analyzed more as transmission-line environments, examining things like cross-coupling, line impedances, and resonant regions.

While the fundamental switching frequency of power-conversion devices may be relatively slow, the edge speeds and rise times are typically much faster to help reduce switching loss. These edges and other interferers can excite the power-distribution network in a way that generates noise and harmonics at much higher frequencies. Depending on the target device and the function of the circuit, these higher-order harmonics can interfere with operation.

Choosing an oscilloscope and power-rail probe with enough bandwidth to see these events is crucial to diagnose problems related to high-frequency interference. For the latest designs, this bandwidth should be in the 1- to 4-GHz range.

SUMMARY

Selecting the best probe is highly application-dependent. Therefore, it's important to understand the measurement requirements of the application and ensure the probes are well-suited to the job. Differential probes are a clear choice for many power-electronics measurements, especially those that aren't ground-referenced.

For grounded measurements, single-ended probes are a good option, but avoid using 10X probes and over-attenuate small signals. For low-voltage signals, such as ripple, a 1X or 2X probe may be a better choice. For MOSFET design, high-performance differential probes are the best choice, while optically isolated differential probes are the ideal choice for GaN or SiC devices. Specialized power-rail probes are designed to handle the measurement and connectivity challenges faced when looking at dc supplies. ed

AHMED EISAWY is a Senior Product Marketing Manager, business leader for probing solutions at Tektronix. Prior to this role, Ahmed had various roles and responsibilities in the EDA industry in technical and product marketing with a primary focus on analog/mixed-signal design and verification in custom IC design environments. Ahmed holds a B.Sc degree in electrical engineering and a M.Sc degree in analog/mixed-signal design methodologies from Cairo University.

WILLIAM WONG | Senior Technology Editor

FPGA Highways Speed Machine Learning and Communication

Achronix's Speedster7t FPGAs utilize high-speed networks to provide faster communication between machine-learning blocks, storage, peripherals, and FPGA fabric.

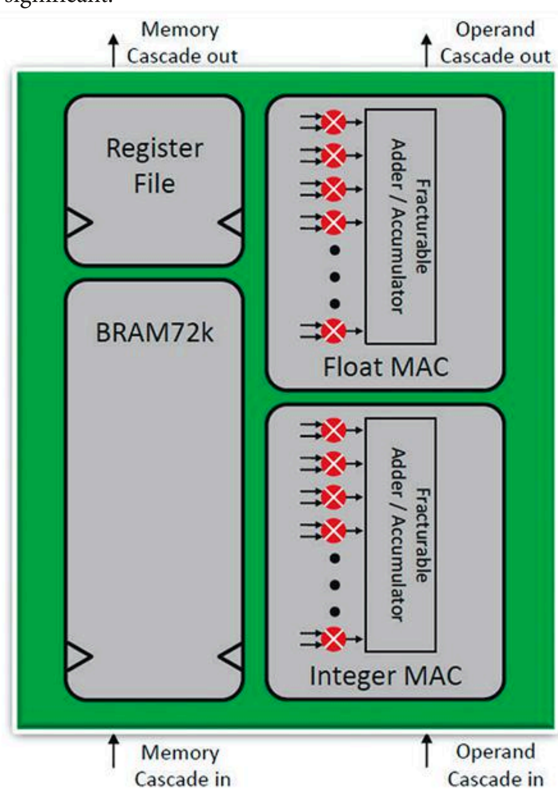
Today's high-speed FPGAs have a passing similarity to their older counterparts. However, large arrays of identical lookup tables (LUTs) and interconnect fabrics have given way to high-speed, networks-on-chip (NOCs) and specialized blocks targeting applications such as machine learning (ML). The solutions are almost more ASIC in nature, gaining the performance and efficiencies of that approach, but retaining the configurability of conventional FPGAs.

Achronix's latest Speedster7t FPGAs incorporate all of these features to provide faster communication between ML blocks, storage, peripherals, and FPGA fabric *(Fig. 1)*. It's built on TSMC's 7-nm FinFET process. Chips and development boards will be available in Q4.

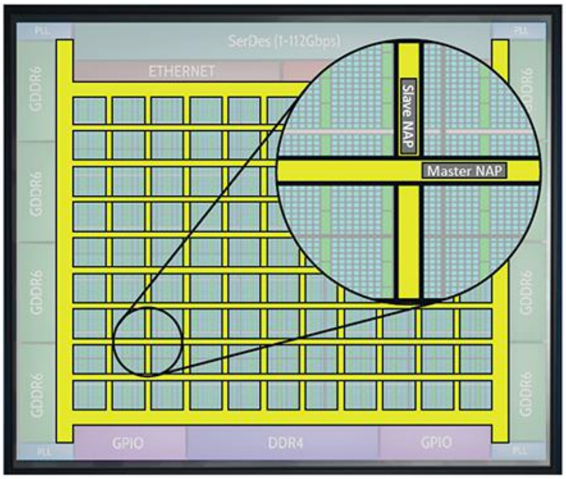
The Speedster7t includes conventional FPGA blocks such as LUTs, DSP engines, and block memory sprinkled throughout the FPGA fabric that connects these components together. The challenge in large FPGAs is timing closure when information from one end of the fabric needs to get to the other side. This normally requires reducing the system clocking rate to meet design needs. DSP engines are one way to reduce connectivity requirements while providing a more efficient implementation, but this is on a small scale.

1. The Speedster7t leverages TSMC's 7-nm FinFET process.

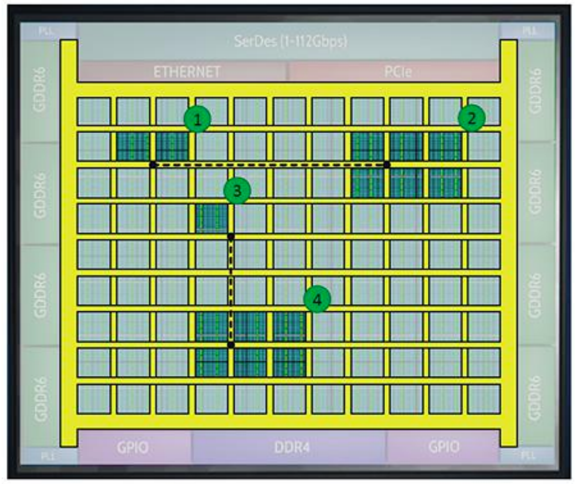
The machine-learning processor (MLP) blocks are significantly larger than a DSP block *(Fig. 2)*. It would take multiple DSP engines to make up the integer or floating-point MACs in the MLP. The MLP also incorporates a register file and 72 kb of block RAM (BRAM). The MLP could be built up from conventional FPGA blocks; however, the resulting area would be two or more times larger and it would run slower. This is true when moving an FPGA design to an ASIC, but the cost is significant.



2. The Speedster7t machine-learning processor (MLP) block targets ML applications.



3. Achronix's network-on-chip (NOC) implements non-overlapping communication highways with network access points (NAPs) near each intersection.



4. The NOC allows for connections between blocks. However, the NOC operation is independent of the surrounding fabric, so no timing closure is required between connected blocks.

The MLP retains the configurability of an FPGA with its surrounding components and fabric while providing an efficient, cascadable ML accelerator. These larger MLP blocks can be combined to handle larger or multiple layers within an ML model. As with most ML accelerators, a single block will not handle an entire model, and even multiple blocks will have to be reused to manage very large ML models.

The MLP blocks are highly configurable themselves. The MACs can handle FP16 and a non-standard FP24, as well as TensorFlow's BFloat16 plus 12- and 16-bit block float formats. The integer MACs support int4, int8, and int16 formats. The floating-point and integer sections can operate in tandem, allowing for custom numeric operations such as accumulating integer results, often used to reduce the size of ML data, as a floating-point value. The MLP BRAM is for general storage, while the register file is typically used to deal with cascading and channelization for larger ML models.

Though one might consider the large MLP blocks to be an enhancement along the lines of a super DSP block, the part of the Speedster7t that deviates from a conventional FPGA is the NOC *(Fig. 3)*. The NOC consists of non-interconnected horizontal (master) and vertical (slave) communication highways. Near each intersection is a network access point (NAP) that allows the nearby FPGA fabric to link to the highways.

The highways are actually a bidirectional set of unidirectional lanes that connect one NAP to another *(Fig. 4)*. Data enters a NAP and flows to the destination that will forward it in the desired direction. The highways overlap but don't connect to each other. To move data diagonally requires moving off via a NAP and onto another one.

The lanes can operate in different modes depending on how data is handled. This includes transaction-oriented operations using an AXI-style interface, a packet-based interface for handling Ethernet-style packets, unpacketized data streams, and a NAP-to-NAP for handling communication from FPGA logic.

Each column or row has a bandwidth of 512 Gb/s. It's a 256-bit-width channel running at 2 Gb/s. This number is key because the chips are designed to handle 400-Gb Ethernet. Total system bandwidth is 20 Tb/s.

The ports around the periphery of the chip include the Ethernet ports that can be configured to handle different speed interfaces up to 400 Gb/s. This also means that the system can distribute the 400 Gb/s quickly and efficiently to different sections of the FPGA along a column or row without having to worry about time closure on the system. That's because the lanes operate independently of the fabric and other blocks like the MLP blocks.

The periphery of the NOC is also optimized for common operations. This lets the PCIe interface move data to and from the GDDR6 memory controllers. GDDR6 was chosen over on-chip HBM storage because of cost. The GDDR6 memory operates more slowly than HBM. It can be significantly larger while being faster than the DDR4 interface, which is also provided. DDR4 is at the bottom of the memory hierarchy, providing very large amounts of slower storage.

The PCIe Gen 5 and Ethernet interfaces are supported by the 112-Gb/s SERDES. The two PCIe ports support x16 and x8 interfaces.

The first chip out of the chute will be the 7t1500 with 692K 6LUT or 1522K 4LUT blocks. It has 2560 MLP blocks that deliver 211 TOPS with int4 data or 86 TOPS with int8 data. It will have thirty-two 112 Gb/s SERDES, eight GDDR6 interfaces, and a single DDR4 interface. The PCIe Gen 5 includes a x16 and x8 interface. It can handle four 4x 100-Gb/s Ethernet ports or two 8x 50-Gb/s ports. The top-end 7t6000 ups the 4LUT fabric to 5.72 million blocks and 134 MLP blocks.

The technology is also supported by Achronix's Speedcore eFPGA IP. This allows the highways and MLP byways to be included in embedded applications. It also simplifies interfacing, since the ASIC can be connected to outer NAPs. ed

JIM YASTIC | Senior Technical Marketing Manager, Macronix America
www.macronix.com

Integrating Secure Non-Volatile Memory in Internet-of-Vehicles Systems

Automotive electronics, particularly in autonomous vehicles, is skyrocketing, and advanced secure memory storage in non-volatile memory will be crucial to next-gen computing architectures for these applications.

Modern society is currently experiencing exponential growth of connected devices on public Internet Protocol (IP) networks. In 2018, the overall number of connected devices worldwide was 17 billion, approximately half of which were Internet of Things (IoT) devices, according to market research firm IoT Analytics. These numbers don't even include smartphones, tablets, laptops, or fixed line phones.[1] Targeted IoT devices range from atmospheric sensors, remote payment systems, IP cameras, smart lighting systems, and home routers to connected vehicles, which is the main focus of this article.

Connected vehicles are, in essence, connected IoT devices or edge computing platforms. In vehicles with autonomous driving-assistance system (ADAS) functionality, they could be also be considered connected artificial-intelligence edge devices. Internet of Vehicles, or IoV, is the term that describes this classification. Vehicles that fall within this definition cover a wide range of types—from ground to air, from consumer to commercial.

TABLE 1: BASIC SET OF APPLICATIONS DEFINITION

Applications Class	Application	Use case
Active road safety	Driving assistance - Co-operative awareness	Emergency vehicle warning
		Slow vehicle indication
		Intersection collision warning
		Motorcycle approaching indication
	Driving assistance - Road Hazard Warning	Emergency electronic brake lights
		Wrong way driving warning
		Stationary vehicle - accident
		Stationary vehicle - vehicle problem
		Traffic condition warning
		Signal violation warning
		Roadwork warning
		Collision risk warning
		Decentralized floating car data - Hazardous location
		Decentralized floating car data - Precipitations
		Decentralized floating car data - Road adhesion
		Decentralized floating car data - Visibility
		Decentralized floating car data - Wind
Cooperative traffic efficiency	Speed management	Regulatory / contextual speed limits notification
		Traffic light optimal speed advisory
	Co-operative navigation	Traffic information and recommended itinerary
		Enhanced route guidance and navigation
		Limited access warning and detour notification
		In-vehicle signage
Co-operative local services	Location based services	Point of Interest notification
		Automatic access control and parking management
		ITS local electronic commerce
		Media downloading
Global internet services	Communities services	Insurance and financial services
		Fleet management
		Loading zone management
	ITS station life cycle management	Vehicle software / data provisioning and update
		Vehicle and RSU data calibration.

Connectivity options—and opportunities for hacking—are numerous.

Non-volatile memory and security features commonly offered in these devices will be the focus here, because they're commonly targeted in attacks, either directly or indirectly. It could be by installing malware by modifying software stored in memory, or installing a sniffer device that modifies software or data in transit between memory and a host. The attack could also be the theft of sensitive data ranging from personal financial data to information that could be used to bring down the defenses of a vehicle causing unauthorized use.

INFUSION OF FEATURES

The amount of advanced electronics in autos is growing exponentially. The main reasons for these increases are the growing inclusion of driver-assistance features that include adaptive cruise control, cameras, object identification, and notification/crash avoidance. The emission and powertrain control systems are more complex to manage hybrid propulsion. Clusters are being converted from mechanical to electronic displays along with the addition of head-up displays (HUDs). Infotainment options and connections to onboard LCD monitors are growing, too. Lighting and other environmental controls are increasing and becoming more automated. Take, for example, a rainfall-detection system that automatically activates the car's windshield wipers when it senses precipitation on the windshield.

And then there are requirements for connectivity to secure servers to download mapping and positioning information as well as entertainment content (video and audio), and to manage software updates for onboard computer systems. Moreover, security systems to authenticate authorized drivers are getting more complex, while recording systems are being added to capture the last seconds before and after an accident to capture driving patterns, aiding law enforcement and insurance companies.

A phenomenal growth in data has accompanied these new features, and with it a need to store them. The data comes in multiple forms: at rest, in use, and in motion. Most data is in use or in transit, and a small percentage that's stored should be done so securely. However, different levels of security are required. Logging or calibration data is typically the least secure. The most secure is in the form of personally identifying information (PII), which are credentials and keys that provide access to services or levels of capability.

OPENING THE DOOR TO THREATS

Many potential threats to the electronics are lurking with the emergence and adoption of what's called vehicle-to-everything, or V2X, technologies:

- V2I: vehicle-to-infrastructure
- V2N: vehicle-to-network
- V2V: vehicle-to-vehicle
- V2P: vehicle-to-pedestrian
- V2D: vehicle-to-device
- V2G: vehicle-to-grid

As highlighted in *Table 1*, from the ETSI TR-102-638 Intelligent Transport Systems technical report[2], the connectivity options and opportunities for hacking are great.

The hacking and controlling of a Jeep by Dr. Charlie Miller and Chris Valasek, through a flaw in the Uconnect system, is a commonly cited example on the dangers of insecure vehicles. This sort of security breach should get the respect it deserves.

However, there are many practical reasons for security in thwarting more likely forms of attack. Take, for example, ensuring that odometers can't be rolled back to change the monetary value of a vehicle (fraud) or hacking into a car's electronics to take control of it (auto theft). Another not-so-obvious example is keeping a competitor from reverse-engineering and copying the design of an electronic control unit with the intent of using it in their vehicle design (product cloning).

In addition, an auto must protect against the vehicle owner or someone else accessing information she/he isn't authorized to access, such as services related to the vehicle experience like mapping, levels of autonomous service, audio, or copyrighted video content. Then there's the emergence of electronic commerce, a good example of which is the adoption of integrated toll modules to replace RFID stickers on windshields. Personal information is increasingly being stored in onboard systems, such as contacts and passwords. Like a computer at home or your smartphone, this data must be kept secure and privacy must be maintained.

TABLE 2: ATTACK SURFACES AND PROTECTION APPROACHES

Attack surface	Vulnerabilites	Protection approaches
Communication channels	• Man in the middle • Snooping, injection • Weak entropy sources • Clear text	• Use of sequenced frames • Strong random-number generators • Use of encryption and authentication
Physical	• Side-channel analysis • Timing, power, EM emissions, acoustics • Fault injection (power and clock glitches) • Data remanence in memory • Brute force: probing, de-capsulation, reverse engineering, uncovering special debug modes	• Fault injection—redundancy and fault-tolerant computing • Hiding and scrambling the bus • Data encryption and de-encryption in a trusted zone • Sensor mesh in top metal layer • Fault detection – memory erase • Other anti-tampering designs
Software	• Fault injection • Version rollback • Interrupts • Buffer overflows • Malware injection • Code in clear text form over non-encrypted channels	• Fault injection • Execution redundancy • Checksums on data transfers • Randomized execution • Encryption/decryption • Creation of trusted execution zones • Designs based on compartmentalization and isolation, and access control • Root of Trust and Chain of Trust implementations
Lifecycle	• Unsecure memory provisioning environments • Code and data programmed and stored in clear text • Provisioning and re-provisioning in the field • Root-key discovery and exploitation	• Secure provisioning environments or services • Root of Trust and Chain of Trust implementations • Encryption and authentication • PKI – public/private keys, certificates, authentication • In the field secure secret key sharing

Various attack surfaces can be met with proven protection approaches.

Finally, an important and intertwining link exists between safety and security. Applying the proper levels of security, in of itself, provides the determinism needed for safe operation. This is often overlooked when thinking about security alone. A robust security framework protects against unauthorized actions taken by individuals, while improving safety by incorporating additional controls in the system design.

GENERAL ATTACK TYPES AND PROTECTION APPROACHES

Three major objectives need to be met to achieve security of information in connected devices:

• *Integrity:* Ensuring that information is authentic and hasn't been compromised. In other words, providing a security feature called non-repudiation that provides proof of authenticity and origin.

• *Confidentiality:* Securing information so that it's private and can't be accessed by or made available to unauthorized users.

• *Availability:* Making sure that information is accessible when it's needed.

The best place to start in understanding device security would be to take a high-level view of it. There are general categories of attacks that are useful in understanding the threat landscape, as well as common approaches to protecting against these threats.

When designing a component for a vehicle, the potential attacker's motivations must be considered. Is it stealing intellectual property, a denial-of-service (DOS) attack, financial theft, or some other type of data pilfering? Next, the would-be attackers need to be understood. Are they government-funded, competitors or just weekend hackers? After analyzing the attacks that seem likely, are they invasive, semi-invasive, or non-invasive? Lastly, what kind of attack methods might be used and how can a secure design thwart the attack or make it too expensive to hack. In short, a thorough security analysis needs to be conducted up front.

Table 2 summarizes some of the major attack surfaces, their vulnerabilities, and protection approaches against likely attacks.

SECURITY FEATURES IN NON-VOLATILE MEMORY

It's always assumed that the details of the cryptographic algorithms are well-known by an attacker, so it's only the secrecy of the key that ultimately provides security. Trying to keep keys secret is one of the toughest challenges in deploying security. Beside the keys, other user-sensitive or private data

AHEAD OF WHAT'S POSSIBLE™

Design Note

Simplifying Design of Analog Input Modules for Process Control

Cathal Casey
Applications Engineer, Analog Devices, Inc.

Introduction

When designing analog input modules for process control applications such as programmable logic controller (PLC) or distributed control system (DCS) modules, the main trade-off considered is usually performance vs. cost. A tradition exists in this application space of using bipolar ±15 V rails to supply active front-end components that are used to attenuate or gain input signals. This impacts the cost of bill of materials (BOM), as well as adding to the complexity of the design with the creation of the isolated bipolar supplies. To save on costs, an alternative approach is to architect with a single 5 V supply. The single 5 V rail greatly reduces the complexity of the design of the isolated supply for the analog front ends. However, it can introduce other pain points and may result in a lower accuracy measurement solution. The AD4111 takes on a lot of the integration needed for voltage and current measurements, and addresses the limitations of a 5 V supply solution.

Integrated Front End

The AD4111 is a 24-bit Σ-Δ ADC that reduces both development time and the cost of the design by implementing an innovative yet simple signal chain. Utilizing ADI's proprietary *i*Passives® technology, it incorporates an analog front end and ADC. This enables the AD4111 to accept ±10 V voltage inputs and 0 mA to 20 mA current inputs while operating off a single 5 V or 3.3 V supply without the need for external components. Voltage inputs are specified to an overrange of ±20 V at which the part will still provide a valid conversion and an absolute maximum specification of ±50 V on a voltage pin. Current inputs are specified to a range of –0.5 mA to +24 mA, which enables accurate current measurements close to 0 mA and provides precise conversions to 24 mA. The voltage inputs of the AD4111 have a guaranteed minimum of 1 MΩ impedance. This enables the removal of ±15 V external buffers, further reducing board space and BOM costs. 5 V designs are limited by the requirement to have a high impedance voltage divider per voltage input, which uses more board space. The design of a discrete solution becomes a trade-off between cost and accuracy of precision resistors. To resolve this, AD4111 incorporates a high impedance, precision voltage divider per input as shown in Figure 3.

Open Wire Detection

Typically, a limitation of single 5 V designs is a lack of open wire detection, which is usually implemented using a high impedance resistor to the 15 V supply rail to pull an open connection to an out-of-range voltage. This is overcome by the AD4111, which boasts the unique capability of open wire

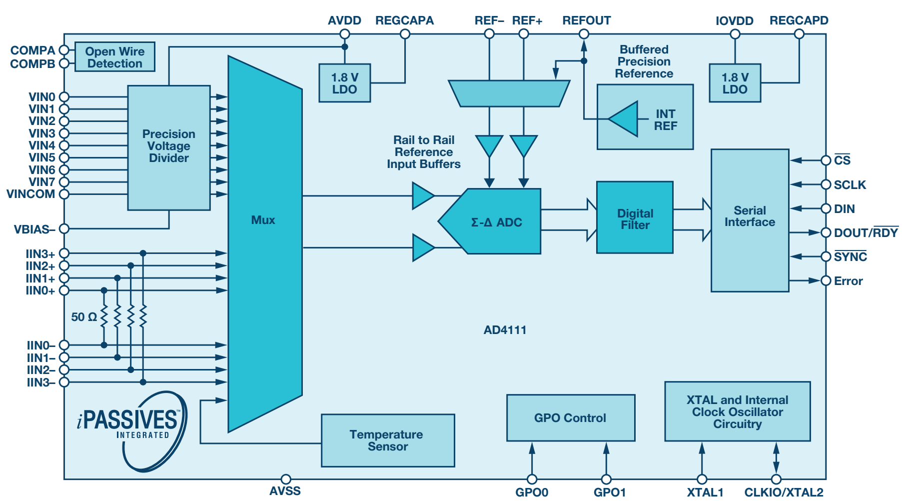


Figure 1. AD4111 functional block diagram.

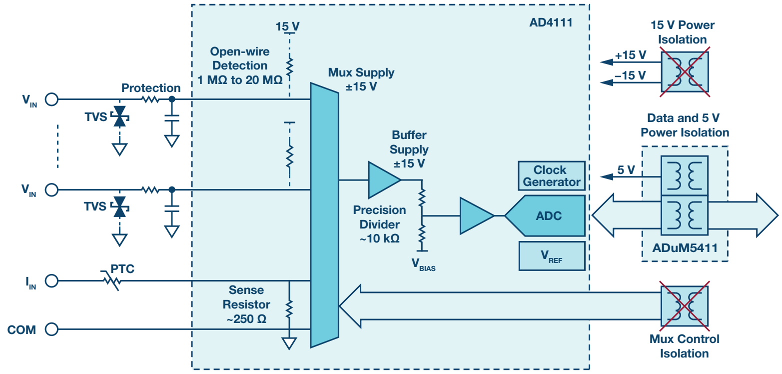


Figure 2. Typical high-end solution.

detection using a 5 V or 3.3 V supply. This separates open wire detection from an out-of-range fault, further simplifying diagnostics. By including this feature internally on the AD4111, the need for a pull-up resistor on the front end is removed, and thus the need for the 15 V supply is also removed, as shown in Figure 2. Eliminating the ±15 V supply reduces the complexity, area, and emissions of the isolation circuitry. For applications where open wire detection is not desired, an alternative generic, the AD4112, is available. This part has all the benefits of the AD4111 but without open wire detection.

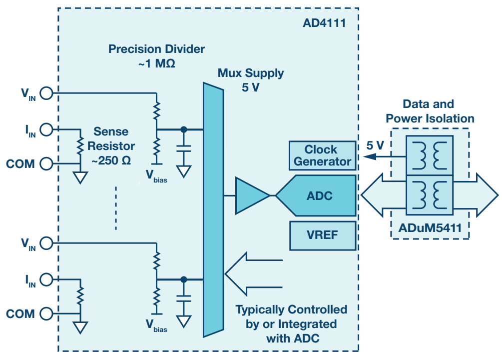


Figure 3. Typical low-end solution.

System-Level Solution

The AD4111's integration of a voltage reference and internal clock help to reduce board size and BOM cost further, while allowing external components to be used if and when a higher precision and lower error conversion over temperature is required. Figure 2 and Figure 3 show typical high-end and low-end solutions, respectively. Highlighted in both Figure 2 and Figure 3 are the proportions of the signal chain that can be entirely replaced by the AD4111. The AD4111's total unadjusted error (TUE) accuracy specifications were targeted to reach system level requirements. For many solutions, the accuracy may be sufficient to remove the need for any additional calibration. In existing high accuracy solutions, modules are often calibrated per channel. The AD4111 has been designed with highly matched inputs so that calibrating one input will provide a similar level of accuracy across all inputs.

EMC Testing

PLC and DCS modules often operate in harsh industrial environments and have to survive electromagnetic interference (EMI) conditions. This adds a greater level of complexity when designing an electromagnetic compatibility (EMC) capable input module since most devices are not rated for EMC, therefore designing the input protection and filtering circuitry becomes complex. This can add significant time to development in terms of design and testing. EMC labs are expensive to rent and failed testing can mean long delays until boards can be redesigned and retested. The AD4111 has been designed into a printed circuit board (PCB) that demonstrates a proven EMC solution. The board is characterized to ensure the circuit performance is not permanently affected by radiated radio frequency (RF) or conducted RF disturbances, and has been shown to have sufficient immunity against electrostatic discharge (ESD), electrical fast transient (EFT), and surge as per the IEC 61000-4-x set of standards. It has also been evaluated for CISPR 11, where the radiated emissions for the board are well below the Class A limits. More information on the AD4111 EMC PCB can be found from AN-1572. This application note details all the necessary information on test procedures used, as well as board design schematics and layout to design an EMC proven input module for the AD4111.

Conclusion

The AD4111 is a system-level ADC with high levels of integration, which includes a comprehensive degree of configurability. Its ability to take ±10 V voltage inputs and 0 mA to 20 mA current inputs, and operate from a single 5 V or 3.3 V supply with open wire detection, along with its many other features, provides a unique solution for analog input module design. It comes in a 6 mm × 6 mm, 40-lead LFCSP and enables modules that once required a full complex PCB to now be replaced by a single device.

For more information on any of the above mentioned products, visit *analog.com/ADC*.

Data sheet download mou.sr/AD411x-DN

are usually stored at the same safety level as cryptographic keys.

In some secure execution environments, a dedicated secure storage is required not only for secret key information, but also for application-specific data. Furthermore, for modern operating-system architectures that support virtualization or multi-tenancy, it's a requirement to support and store the security credentials of multiple users for multiple applications. For complex systems, this secret information will be stored in external secure flash memories due to the storage size requirements.

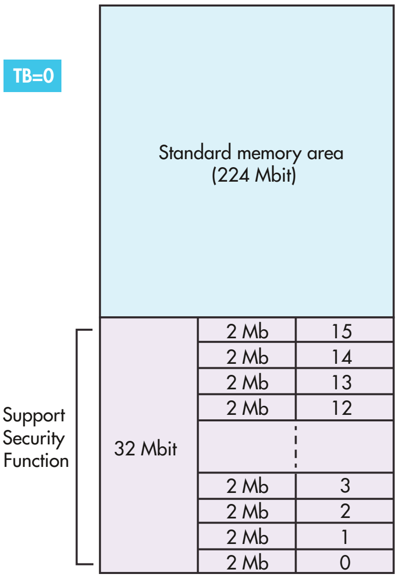
Table 3 (page 36) shows a wide range of standard security features that are common in today's non-volatile memory offerings. They're effective in protecting various threats to its contents.

Why would a product need to implement memory with advanced security features? To start with, an overabundance of existing designs is based on older microprocessors. Thus, if there's a requirement to make the product more secure, it's expensive to do major redesigns of existing products, particularly with complex applications in a hard, real-time environment that have been field-proven. If a requirement has come from an organization's marketing department that it must now be connected to an open network, where will the required keys and credentials be securely stored to support secure access?

Three major advanced security memory products are available today: RPMC, Authentication Flash, and a more full-featured secure flash. RPMC is a memory device whose sole purpose is to provide non-volatile monotonic counters for the host to support sequenced frames in secure communication protocols like TLS/SSL to thwart man-in-the-middle attacks. Several memory providers currently offer this type of flash as a dual-purpose memory. This means there is a region in flash memory reserved for RPMC functionality, with the rest reserved for normal NOR Flash. Authentication Flash devices only perform authentication with the host before a secure operation. A good example of this would be the RPMB feature in eMMC5.1. Full-featured secure flash devices perform authentication and encryption along with a full range of additional security features.

A good example of a secure flash with advanced security features is the Macronix ArmorFlash. It's a NOR flash device and incorporates all of the standard security features listed in *Table 3*, as well as advanced security features listed in *Table 4*. It has a standard SPI interface and leverages the SPI command infrastructure. ArmorFlash introduces special packet operations that provide secure read and write operations to its secure memory region for data storage.

The memory layout shown in the *figure* is one example of how it could be configured. It has a secure memory region that's further partitioned into 16 independent 2-Mbit units for data storage.



Secure flash memory can be configured for two main sectors—standard memory and a secure region that's partitioned into 16 independent units for data storage.

Another standard memory area of 224 Mbits for code storage is accessed directly with regular SPI flash commands after an initial gated authenticated exchange.

Advanced security flash devices have the following security elements:

Non-volatile monotonic counter

The use of non-volatile monotonic counters is the best weapon against replay attacks, even if power outages or glitches occur. This is because monotonic counter values are stored in non-volatile memory. ArmorFlash, for example, has four independent 32-bit monotonic counters, which only respond to authenticated operations. Counter configurations can be programmed to change counter behavior depending on different application scenarios.

TRNG

High-entropy random-number generators (RNGs) are fundamental to almost all secure systems because cryptographic systems depend on secret data that is only known to authorized users. There are two main uses for RNGs. The first use is in producing a random value called a nonce, used in calculating a unique MAC challenge-response authentication value. In this case, the client initially receives a random challenge, usually a message embedded with a nonce value from the host. The device uses this value as one of the inputs into a MAC calculation. The nonce value could also be provided by the memory device.

The second use is in generating encryption keys. In some implementations, a dynamic/session symmetric key is generated from random number. The true RNG (TRNG) in ArmorFlash has spatial and temporal dependence that produces a TRNG with high levels of entropy.

TABLE 3: NON-VOLATILE MEMORY SECURITY FEATURES					
	Security features	NOR flash	NAND flash	e.MMIC	Advanced flash
Hardware	BGA package: The ball grid array under the chip protects against probing.	X	X	x	x
	Protection pin: The block-protect operation using the protection pin can protect the whole chip or selected blocks from erasing or programming.		X		
	Hardware write-protect pin: There are two versions, depends on the flash type: NOR: It protects the register settings that configure the program/erase protection of blocks and sectors. NAND: The memory will not accept the program/erase operation. It is recommended to keep "write-protect" active during power on/off sequence.	X	X		X
Software protection	Temporary block protection: This avoids accidental program/erase to specific blocks.	X	x	x	x
	Solid protection: This permanent block forbids malicious modification to block-protection configurations	X	x	x	x
	Unique ID: This is a value that is unique to each non-volatile memory device	X	x	x	x
	Password-protect block locking: This feature uses an advanced method to protect block-locking configuration from modification	X		x	
	Read protection: Protecting against data corruption	X	X		
	Sanitize: All data is physically erased to prevent data reuse.	X	x	x	x

Standard security features common in non-volatile memory are effective in protecting against threats.

Key storage and management

The storage and management of keys are crucial to authentication and encryption. For example, a specific memory region in ArmorFlash is capable of storing four 256-bit keys. Each key is set up through an independent and lockable process. Several key-management commands are available to set up keys according to their intended use cases. The configuration allows for separate keys for authentication and encryption operations.

TABLE 4: NON-VOLATILE MEMORY ADVANCED SECURITY FEATURES				
Security features	NOR flash	NAND flash	e.MMIC	Advanced flash
One-time programming: Protect configuration setting from others	X	X	X	X
OTP space: A space for OTP data	X	X		X
RPMB: Authenticated access	X		X	
Replay protected monotonic counter (RPMC): Monotonic counter support				X
Secure region: Authenticated access				X
Independent areas for encryption/decryption: Support for multiple users				X
True random-number generator (TRNG)				X
Key generation				X
PUF code: Hardware feature that provides a unique sequence of values based on intrinsic process variations during the manufacturing				X

Advanced security features are found to varying degrees across non-volatile memory types.

Authentication

For its part, ArmorFlash introduces a packet read/write command for various security operations. The packet command is a proprietary link-layer packet protocol that's similar to the RPMB protocol. It has its own sub-command operation codes and data structures. The design for the layer above the link layer involves a shared symmetric key used for the authentication process and a Cipher Block Chaining Message Authentication Code (CBC-MAC), which is used to produce the required MAC values. CBC-MAC is implemented as part of a block cipher algorithm called Counter with Cipher Block Chaining-Message Authentication Code, or CCM. It's a mode of operation of a block cipher algorithm that integrates both authentication of the host and device, and encryption of the payload. It's a NIST standard (NIST Publication 800-38C).

Data encryption

As part of the Advanced Encryption Standard (AES)-CCM algorithm, the data field in the packet contains the encrypted payload based on a symmetric key block cipher algorithm with a minimum block size of 128 bits. In the case of ArmorFlash, AES-256 is used, making for a 256-bit block size.

Unique ID and extra serial number

An additional 8-kbit secure one-time program (OTP) area is added for storing unique identifiers and static data according to system applications' demand. The 8-kbit secured OTP area further decomposes into two rows of 4-kbit configurations. The flash device also could incorporate an extra 8-byte, factory-coded serial number, used in cryptographic calculations to uniquely bond the host and secure flash.

Non-volatile PUF

Although programmed unique ID or serial numbers are okay for identification, a physically unclonable function (PUF) code can be used that's truly unique to each memory device. PUF is becoming common as a unique identifier or digital fingerprint for semiconductor devices. Devices that implement PUF take advantage of random and physical variations of a flash-memory array to produce a non-volatile PUF code. Users can leverage this code both for identification purposes and as an input for key generation.

CONCLUSION

The growth of IoV and connected transportation continues to explode, as does the associated data. Computing architectures in this space are moving to decentralized models to adapt, bringing about the growth of powerful, mobile edge computing platforms that include support for applications such as personal automobiles, autonomous vehicles, mobility-as-a-service (MaaS) systems, and other new forms of transportation like robo-taxis (ground/air).

This all has to be done safely and securely in an untrusted environment. Therefore, non-volatile memory requires a range of security mechanisms and policies to ensure identity, confidentiality, integrity, authenticity, and availability. Advanced secure memory-storage features found in non-volatile memory are crucial to achieving these objectives.

REFERENCES:

1. IoT Analytics, State of the IoT 2018: Number of IoT devices now at 7B – Market accelerating, August 18, 2018, https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/

2. ETSI TR-102-638 Intelligent Transport Systems technical report, Retrieved on 5/7/2019 from: https://www.etsi.org/deliver/etsi_tr/102600_102699/102638/01.01.01_60/tr_102638v010101p.pdf

ASHOK ALAGAPPAN | Lead Consulting Engineer, ANSYS

Power Semiconductors: Past, Present, and Future

Advances in manufacturing processes as well as material and packaging technologies, plus the emergence of new application areas, are creating a resurgence of sorts for wide-bandgap devices.

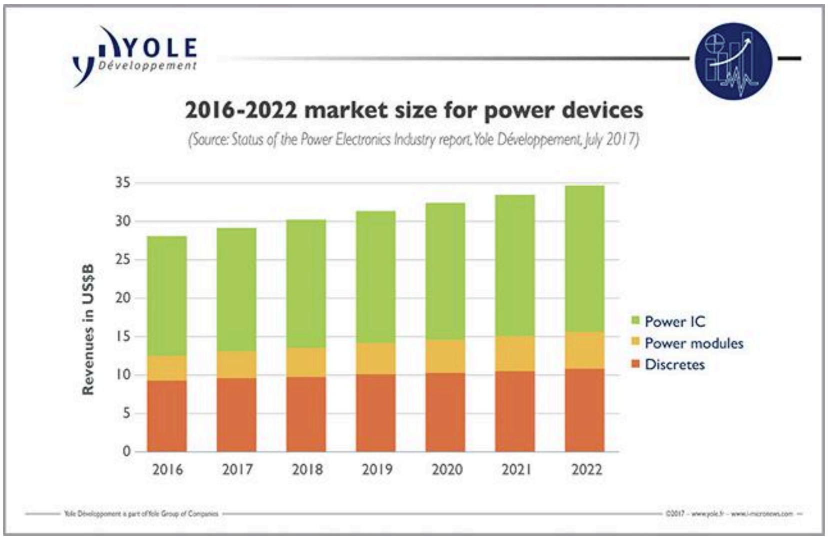
According to industry research firm Yole Développement, electronics manufacturing is growing at a steady rate above U.S. GDP in part due to power electronics *(Fig. 1)*. Power devices have found their way into high-value products across different market segments such as information technology, electric motor drives, grid infrastructure, automotive, and aerospace. While the first three are mature markets, electric aerospace and automotive are developing.

For the power-generation industry, advances in power devices is continuing to drive efficient and reliable power-grid infrastructure. Currently, up to 40% of generated electric power passes through power electronics between generation and end use. As the power consumption of data-center applications grows in relation to the total electrical power consumption, data-center architectures rely on high-efficiency designs. Innovations in power electronics drive more electric aerospace applications such as no hydraulic networks with more or full electric engines.

In electric vehicles, power devices are used in energy storage, electric drive and power conversion, interior, thermal, powertrain, and chassis. Electric-vehicle adoption presents a high-volume opportunity for power electronics. It's expected to grow to a $3B market by 2022.

The two major technologies that drive the growth are SiC (silicon carbide) and GaN (gallium nitride) thanks to superior material properties. SiC is typically used in high power applications (10 kV and beyond) and have the benefits of lower switching losses and lower cost, but with less reliability. GaN is usually employed in high-speed applications (higher switching frequencies). Other advantages are high operating temperatures, high breakdown voltage, and being more robust, reliable, and rad-hard. However, it has the disadvantage of material mismatch (GaN on Si or SiC) and higher cost. The *table* provides a summary of key material properties.

SiC and GaN technologies have found their space in the applications as shown in *Figure 2*.



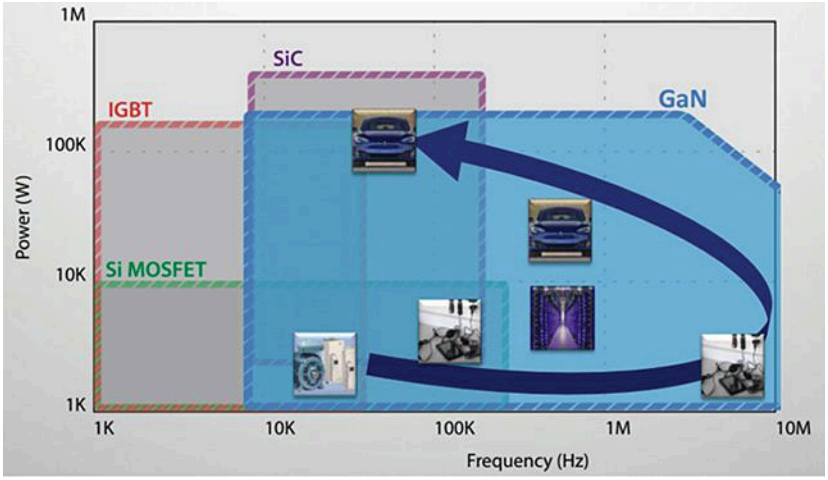
1. The market for power devices continues to expand and grow.

Key Material Properties of Wide-Bandgap Semiconductors

Material Properties	Si	SiC	GaN
Bandgap (eV)	1.1	3.2	3.4
Thermal Conductivity ($W/cm^2 K$)	1.5	5	1.3
Coefficient of Thermal Expansion (CTE)	2.6	4.2	5.6

MANUFACTURING CHALLENGES

Due to limited market volume until recently, there hasn't been much



2. Wide-bandgap semiconductors will find homes in an array of applications. *(Source: GaN Systems, Styles)*

advancement in manufacturing in this space. The industry was able to meet the demand mostly on 2- and 4-in. wafer diameter processes until the end of the last decade. Industry adopted 6-in. wafers a couple of years back and the limited volume of production using 8-in. wafer process is a sign that demand is picking up.

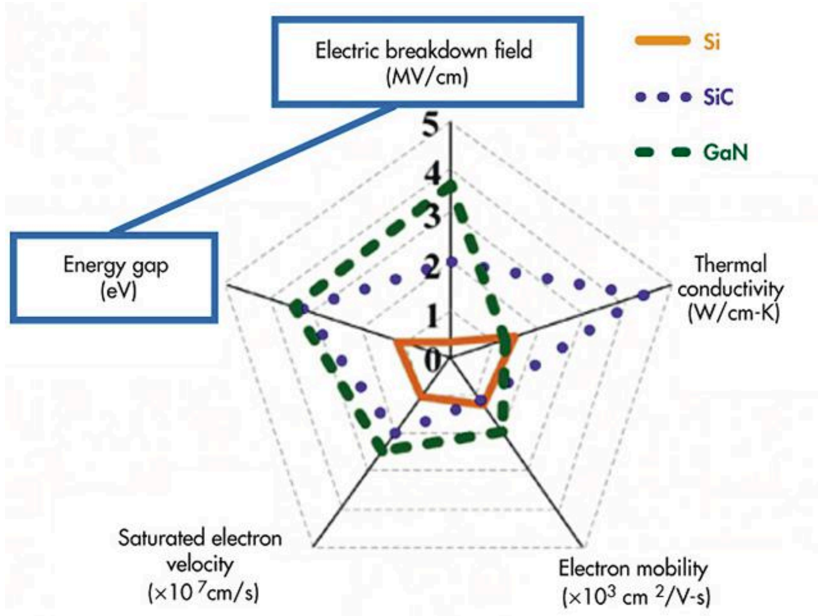
However, it pales in comparison to the digital logic industry, which has migrated to a 12-in. wafer process a long time ago. Obviously, there's supply where there's demand. Market adoption of power devices driven by evolving applications and maturity in manufacturing (in terms of cost and defect reduction) are going to be the main growth drivers for years to come.

Let's look at one of the major aspects that involves the so-called extrinsic defects originating mainly from the manufacturing process itself. One of the important processing steps when it comes to the origination of a defect is epitaxial deposition. Bulk drift layers are formed using this step. It's a long processing step that needs to have high purity, but it has a high propensity to generate more defects.

Another subsequent yet critical step is the so-called ion implantation, where the appropriate dopants are added to the bulk layer. The process variation needs to be tightly controlled across the wafer and between the wafer and lots to achieve uniform thickness. Variation outside the tolerance significantly alters the device characteristics.

Device performance metrics such as breakdown voltage, which is indicative of blocking voltage across the device, are determined by the quality of the process in the form of thickness and doping uniformity. Higher defect density in gate oxides can be a limiting factor for failure mechanisms such as time-dependent dielectric breakdown (TDDB).

Advances in manufacturing techniques will play a key role in not only enabling high-volume production capabilities, but delivering a quality product to the market as well.



3. Wide-bandgap semiconductors offer a number of advantages over silicon. *(Source: PowerAmerica, Veliadis)*

ADVANTAGES OF WIDE-BANDGAP DEVICES

The most telling graphic that sums up the advantages of the wide-bandgap device compared to its silicon counterpart is the famous web graph shown in *Figure 3.* As shown, the large bandgap and the critical electric field enable high voltages to be blocked with thin layers, resulting in lower resistance and associated conduction losses. Thin layers not only provide low on-resistance, but allow for smaller form factors and reduced capacitance, leading to higher frequency operation.

Large bandgap also produces low intrinsic carrier concentration, resulting in low leakage and high temperature operation. Large thermal conductivity is a key benefit for high power operation with limited cooling requirements.

DEVICE TYPES

There are numerous device structures; each type is unique and is used to address specific needs. For SiC technology, the criteria all device types try to achieve is lower on-state resistance ($R_{DS(ON)}$) and high breakdown voltage (V_{BR}). Lateral devices have been commonplace, but they have the limitation of large areas required for blocking voltage. SiC devices with blocking voltages up to +900 V are typically of vertical configuration.

Multiple device types are fabricated in SiC technology. Unipolar devices are those where conduction is by one carrier type such as an electron. Some of the unipolar devices are MOSFET, JFET, and junction barrier Schottky diode. These have higher conduction losses but faster switching. Some of the bipolar devices are BJT, thyristor, ISBT, and PIN diode. They have lower conduction losses due to both electrons and holes as charge carriers, but they utilize slower switching with higher losses.

Due to its superior properties, the high electron mobility transistor (HEMT) is a device structure commonly used with GaN technology. It's purely of the lateral type and the device operates in both depletion mode and enhancement mode. Depletion mode is normally on and can be turned off with negative voltage. Enhancement mode is normally off and turned on by the application of a positive voltage at the gate. It has lower conduction losses ($R_{DS(ON)}$) with faster switching speed as shown in *Figure 2*. There are no losses associated with this device type due to ringing or EMI effects thanks to the absence of body-diode operation.

RELIABILITY

Wide-bandgap devices are susceptible to different failure mechanisms, some of which are common to silicon devices. Due to the high blocking voltage for these devices, the following breakdown phenomena are possible: gate-oxide breakdown, SiC breakdown, edge-termination breakdown, threshold voltage drift over time, or reduction in current flow due to increase in resistance.

There's a need for high electric-field reliability due to high blocking voltage. Accelerated lifetime tests such as HTOL, HTGB, HTRB (high-temperature operating life, gate bias, and reverse bias, respectively) are performed to screen parts that are marginal prior to shipping. However, for certain device types, they don't completely capture or characterize the intrinsic wearout and aging failure mechanisms like dielectric breakdown (TDDB) or bias temperature instability (BTI).

Dielectric breakdown occurs when there's a formation of interface and bulk traps due to high electric field, eventually resulting in a short. Depending on the device types and materials used, other failures like avalanche breakdown could occur at test voltages before the occurrence of intrinsic reliability mechanisms.

Threshold voltage stability (P or NBTI, depending on the polarity of applied bias) is another important intrinsic reliability mechanism that causes the gradual shift in device threshold voltage (V_{TH}) over time, affecting the on-state and blocking characteristics. GaN devices are susceptible to dielectric breakdown in the gate dielectric during forward bias and near the field plate and gate corners during off state.

Depending on the application, there's a third-quadrant operation of power devices that can't be ignored. Reliability of the body diode is typically stressed in the HTOL tests, but the failure mechanisms corresponding to such operation aren't completely characterized. Product qualification, depending on the market it serves, requires performing standard tests to demonstrate zero failures with industry-standard models and parameters (for example, 0.7-eV activation energy). They're not test-to-fail type tests based on a mission profile, which brings out the true performance characteristics and reliability of the devices.

Wide-bandgap devices are susceptible to terrestrial neutrons. When a device gets impacted with a neutron particle, it has the propensity to collide with lattice atoms, creating an ionization path in the trajectory. This results in large current transients leading to gate burn out. This is a parasitic bipolar turn-on effect or gate rupture that's a gate-oxide failure.

Some reliability issues crop up with packaging. Silicon packaging technology isn't enough for wide-bandgap devices. To minimize thermal resistance and thermo-mechanical stresses generated due to high power dissipation, double-sided cooling is often required. Wire-bond techniques typically employed for low-voltage logic devices are to be replaced with planar interconnects using topside copper tabs soldered to die. Encapsulation material ought to have required dielectric strength to meet the high operating voltage and temperature of the die. Finally, the die attachment and interface material must have a high melting point with good thermal conductivity.

Standard qualification doesn't suggest the part can meet a desired lifetime nor that it has lower defectivity. Demanding an often-different mission profile than what's been tested could produce field acceleration and failure modes that haven't been screened by standard tests. The reliability challenges must be addressed with the right software analysis tools and techniques prior to product or system design to prevent expensive field failures.

EMERGING TECHNOLOGY

Ongoing research in the field of emerging technology is focused on bringing ultra-wide-band devices to

(Continued on page 46)

MARCO WOLF | Principal Field Applications Engineer, TE Connectivity

Thermal Modeling: How to Meet EV High-Power Charging Requirements

Powertrain electrification, automation, and new business models are shaping the next generation of mobility. These trends will profoundly impact the electrical power and electronic architectures of future vehicles.

The next generation of vehicles will generate, process, and communicate much more data than current vehicles. Wireless networking via mobile technologies (e.g., 5G, V2X) enables communication with other vehicles or with the surrounding infrastructure, as well as makes software updates over-the-air (OTA) possible.

At the same time, high current power will be transmitted within electrified cars. Today's electric cars already have in excess 120 kW of engine power. The high-power levels required for this performance produce strong electromagnetic fields that require protection of nearby signal lines and electronic components against interference and malfunction (high data rates of up to 20 Gbits vs. high power).

Put simply, the physical layer will play a key role as the backbone of future vehicle functionality and ensuring its reliability. That means low-voltage data connectivity networks and high-voltage (HV) drive systems must work ultra-reliably and safely in parallel.

THE ELECTROMOBILITY FRAMEWORK

Powertrain electrification serves to reduce vehicles' consumption of fossil fuels despite a worldwide increase in the demand for mobility. It's the only way to meet the ever-more stringent limits on greenhouse-gas (CO_2) emissions both in the medium- and long-term.

Long distance in electric vehicles (EVs) will be enabled by fast charging with dc and a future charging power of 350 kW, which is classified as high-power dc charging (HPC DC).

THE IMPORTANCE OF HPC

Until recently, greater attention was dedicated to the driving side of electromobility instead of the issue of charging. This can be attributed to a lack of maturity in the business models of the two industry segments that are involved—carmakers (OEMs) and energy producers.

The typical EV use case varies globally. While European EV drivers expect their car to be capable

of occasional long-distance trips, Asian EV drivers tend to use their cars for short distances in mega cities. HPC DC would enable EVs to be used in all cases. However, simply expanding the inner-city ac charging station network would not be sufficient because lower power charging times would result in prohibitively long waiting times and queues.

Increasing battery capacities to lengthen driving ranges can only be exploited in a helpful way if "bigger" batteries don't lead to even longer charging times. With a charging power of 350 kW, it would be possible to gain up to 300 km of additional range within up to minutes maximum. This would turn EV "refueling stops" into acceptably short breaks (comparable with combustion-engine-driven cars), and the dc charging station would quickly be available for the next vehicle. However, 350 kW of charging power at currents of up to 500 A are the peak load for the complete current path from the charging station to the vehicle battery.

The high current, flowing along this path, causes high heat losses since the electric resistance of all components (connectors, cable) inevitably generates heat. This heat loss needs to be factored into the design and dimensioning of all electrically conductive components to avoid overloading or overheating, or a controlled derating of the charging current should the battery begin to overheat during charging. While derating protects the battery, it also prolongs the charging time. This divergence of objectives needs to be solved in an optimal way.

Thermal management can do this by predicting the exact state of all components in every segment of the structure at any time.

THE CHALLENGE OF HPC

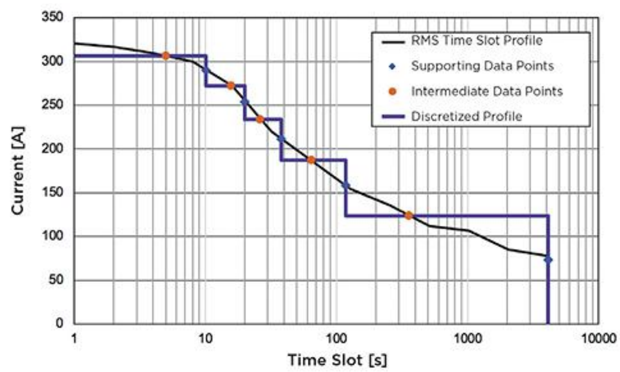
HPC DC represents a peak load state for the electrical system in an EV. The high-charging current of HPC DC causes a significant rise in temperature in all components. This is further exacerbated when the vehicle isn't moving, because there's no convection available for cooling. Therefore, to facilitate HPC DC, the complete electrical system from the charging point to the vehicle battery needs to be designed and dimensioned electrically and thermally.

A major contributor to this challenge is that the higher the current, the larger the required cable cross-section to carry the power at the same level of voltage without overheating. Within the vehicle, this is primarily a matter of weight and available space. For example, it makes a considerable difference, in terms of cost, weight, and bulk, as to whether a 50-mm2 cross-section or a 95-mm2 cross-section conductor between the inlet and the battery will suffice.

An attractive option is to increase the voltage, so the same amount of power can be transmitted at a lower current level. While dimensioning electrical components potentially accounts for unwanted additional mass in the vehicle, it also approaches weight limits in the case of fixed charging cables (mode 4 cables). If HPC DC is to be a realistic proposition, the over-dimensioning of the cable and all other electrical components must be avoided.

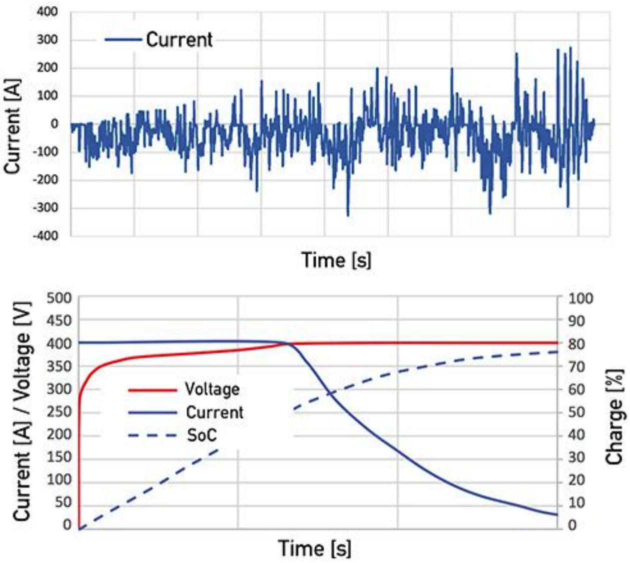
TODAY'S ELECTRICAL COMPONENT DESIGN

To date, design of electrical components along the high-current path has been based on assumptions that aren't really suited for either the dynamic load profiles of driving or the requirements of HPC DC. Existing standards are based on static load points originally used for the design of relays and (switch) fuses, which are determined by statistical methods reflecting the frequency at which they occur and their importance. This leads to the root-mean-square (RMS) values representing static conditions *(Fig. 1)*.



1. This is the quantification method behind the current profile in *Fig. 2.*

Electrical interconnection components are designed in accordance with this type of load profile—which does not reflect reality—and a safety margin of 20% is added. The actual load profile in an EV, though, differs dramatically from previous vehicle applications and their RMS values *(Fig. 2)*.



2. Driving profile compared to an HPC load profile.

Figure 2 explains why thermal design is so essential for charging. While driving results in a very dynamic current profile, consisting of load changes between high peaks and lows, the constant high load during HPC DC isn't reflected at all in the load profile derived from driving. To facilitate a peak load of 350 kW charging power requires a different approach to designing the electrical components.

While the electric energy stored in a battery is typically retrieved over a time span of several hours during driving, three to four times this amount of energy flows into the battery during HPC DC within minutes. Accordingly, the complete high-voltage/high-current path has to be analyzed at a system level to understand its behavior during charging *(Fig. 3)*. Root mean squares aren't very helpful for this as was detailed earlier.

It's essential to know where a constant load can cause overheating that might lead to a critical system state. This thermal angle must be analyzed more closely. The methods currently employed don't deliver the needed answers.

As a result, current systems are statically over-dimensioned due to safety considerations. With 350 kW of charging power, this approach isn't sustainable due to the implications on weight, installation space, and usability.

To address this challenge, TE Connectivity is developing a new design approach within its ZVEI activities (German Electrical and Electronic Manufacturers' Association). The thrust of the approach is to dynamically determine the temperature increase caused by components, as well as the heat dissipation in the system, via established principles of simulation (as used for electrical systems). This methodology makes it possible to examine component design earlier to predict its performance during operation.

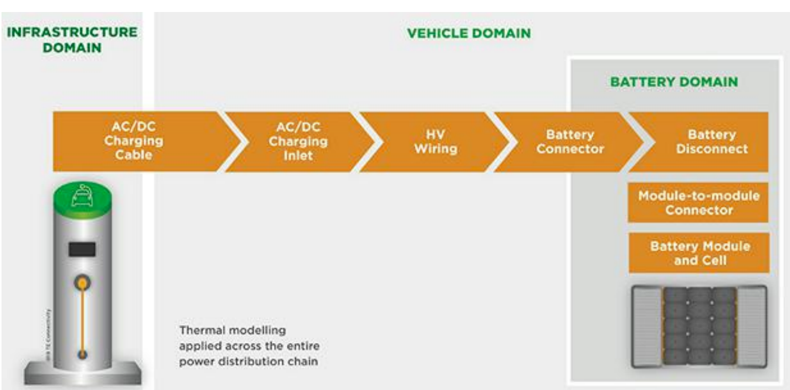
Model-based thermal simulation provides a verifiable basis for future load profiles that facilitate a proof of safety, reliability, and availability of all interconnection components along the high-voltage/high-current path.

THE IMPORTANCE OF THERMAL SIMULATION

The physics of transmitting electric energy causes power dissipation in the shape of heat losses along the wired energy flow. The root cause is the electrical resistance (measured in ohms) of all metallic conductors.

This resistance is known for every element of the high-voltage path. However, the ohmic resistance changes with the temperature increase during operation. The amount of power dissipation that occurs at a specific component can be calculated for a certain current, voltage, and temperature—albeit only for a stationary state when all paths of heat dissipation are balanced.

Existing methods to dynamically calculate the complete high-voltage path on a system level aren't very practical. To apply a well-known method such as finite element analysis (FEA), it would be necessary to make multiple calculations in rapid sequence for each operating point. It's clear a continuing thermal calculation in real time (in the vehicle) requires a different methodology that needs much less computing power.



3. Shown are differently stressed components along the current path of an EV.

One part of the challenge is that heat dissipation along a high-voltage path leads to a comparatively slow system. Depending on an individual component's mass and the adjacent available heat sinks, the individual component will react differently to changing load profiles.

Lightweight components with a limited chance for heat dissipation can become a bottleneck for thermal management. If the generated heat can't be sufficiently dissipated, the component will temporarily become an adiabatic element (i.e., a condition with no heat exchange with the environment) without any chance to externally influence its heating-up process. Thermal bottlenecks of that type need to be understood so that no unnecessary limits or stress are based on the system.

Furthermore, heat dissipation occurs on several paths. In addition to conductive heat dissipation within the material, there's also the share of heat radiation and heat dissipation via cooling air or coolant flows (convection). For each component along the high-voltage path, the mix of these three elements will be different.

When electrical components heat up, they also undergo an aging process that changes the electrical (and or mechanical) properties of the component over time. The stronger the heat entry, the faster this aging process and the smaller the residual performance level of the component.

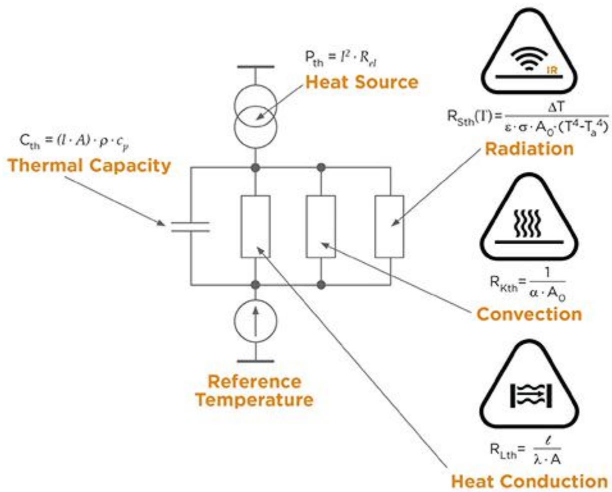
Electrical		Thermal	
Current	I	P	Heat Flow
Voltage	U	T	Temperature
Resistance	R	R_{th}	Thermal Resistance
Capacity	C	C_{th}	Heat Capacity

4. The correlation between electrical and thermal values forms the basis for equivalent circuit diagrams.

THE CHALLENGE OF HPC

It's therefore necessary to find a different tool that can provide a timely definition of a safe and economically feasible current-path design for HPC DC—and provide proof of its safety. Using a proven systemic thermal simulation makes it easy to automatically test an almost unlimited number of possible load profiles in advance. This will reveal potential thermal bottlenecks in the system that can be addressed via design changes.

Employing this methodology can lessen subsequent troubleshooting effort. The reduced investigative effort can be considerable because the thermal system is so complex. The exact root cause may not be in the originally diagnosed component, but rather in an adjacent component along the heat path.



5. As seen in this equivalent circuit diagram for thermal simulation, resistors represent the three ways of heat dissipation.

SYSTEMIC SIMULATION METHOD

This advanced systemic simulation methodology, which calculates heat losses along the high-voltage path under dynamically changing load conditions, is based on Kirchhoff's circuit laws. His "point rule" and "loop rule" state that the sum of all currents at a "point," and the sum of all voltages in a "loop," must be zero.

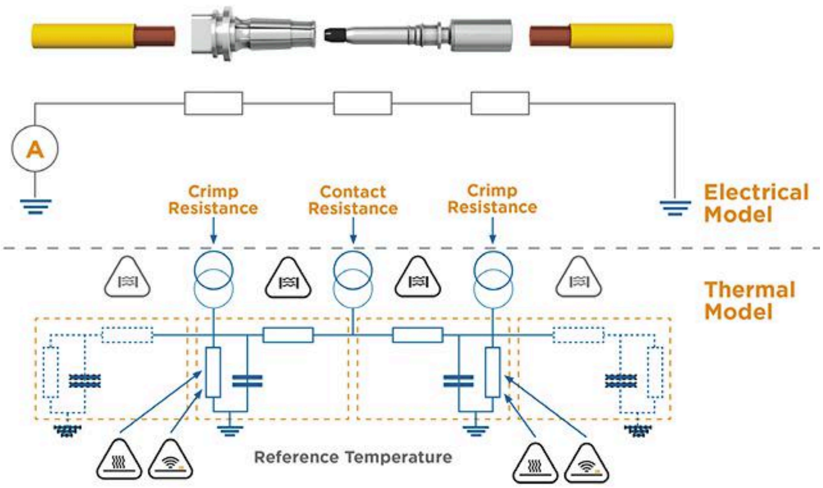
At the same time, the rule states that energy is always conserved. That means the current that's transformed into heat (heat loss) due to the electrical resistance isn't lost. Instead, this heat energy is exactly equal to the difference between the electrical energy flowing into the circuit and the energy that's available at the target system. Equivalent circuit diagrams exploit the immediate and linear relationship between electrical and thermal behavior *(Fig. 4)*.

Consequently, equivalent circuit diagrams *(Fig. 5)* serve to simulate the linked electrical and thermal behavior. In the same way that a voltage sends a current through a resistor, a temperature difference causes heat transport. The different physical forms of transport (conduction, convection, radiation) are each represented by a resistor. Stored algebraic equations in the component model continually calculate the heat generation depending on the applied current and voltage, as well as the ambient temperature.

Based on this heat generation, the different possibilities for heat dissipation are represented by resistors (thermal barriers) and thermal masses/capacities in the equivalent circuit diagram. The diagram represents the heat transport resolved over space via conduction within the material, via radiation, and via convection.

Using this fairly simple method, it's possible to simulate individual components (e.g., a contact), whole products (e.g., a connector, such as in *Fig. 6*) or a high-voltage path, as heat generation and heat dissipation are predictable through loop-formation.

Once cable models are made available from the cable manufacturers, the intermediate sections also can be calculated. In



6. A connector is broken down into loops: There's equivalence between the electrical points of contact in a connector and the thermal simulation.

addition, it's possible to integrate components from different manufacturers (as per the on-board net)—all it requires is to enter the manufacturer-specific electrical parameters. Within the model, these parameters are applied to algebraic equations that follow Kirchhoff's circuit laws. In essence, the model describes the heat generation and heat exchange with the environment.

The simulation can determine, for example, the location of heat sources and heat sinks; when a temperature level becomes critical and begins to shorten a component's service life; how a component integrates into a larger cluster; and where adiabatic states can be found and what impact they're making.

During the original model development, iterations between simulation and testing (raw data from lab testing) served to refine the algebraic part of the model until the accuracy of the prediction matched the test results.

With the resulting simulation methodology, dynamic load profiles can be tested for each component on the high-voltage path with a minimum of computing power.

SAFETY GAIN

The computing power needed for thermal simulation based on equivalent circuit diagrams is so low that it's feasible to run this procedure as a continuous routine task on a typical automotive electronic control unit (ECU). Actual load profiles of real-world driving can thus be calculated in real time.

The simulation delivers data that helps improve functional safety. Simulation and sensor data mutually complement one another as heterogenous diagnostic routines. For automated vehicles, which require multiple redundancy for safety reasons, this can contribute to the safety concept.

DESIGNING HV COMPONENTS FOR THE VEHICLE

Systemic thermal simulation strongly advances the load-oriented design of high-voltage components for the vehicle toward real operating conditions. As manufacturers know, it's highly valuable to be able to predict the performance of a component during its development phase. The systemic and dynamic thermal simulation precisely reveals the expected effects resulting from wear and aging during operation. Thus, a complex system like the high-voltage path can be simulated and its behavior can be predicted. In addition, simulation can cover a breadth of testing that would never be achievable in the testing lab. ed

(Continued from page 41)

market. Materials such as diamond and aluminum nitride (AIN) have much higher breakdown field strength compared to existing material types Si, SiC, GaAs, or GaN, but they have limited practical applications.

Ultra-wide-bandgap materials in the range of 4.5 to 5.0 eV aim to push the breakdown strength much higher without compromising other critical properties such as electron mobility and thermal conductivity. More device structure optimization is required to obtain the desired performance metrics to satisfactory levels.

CHALLENGES AND FUTURE OPPORTUNITIES

As the market requirements for evolving market segments such as automotive, electric aerospace, and data-center applications gets more stringent than traditional markets, wide-bandgap device capabilities have to increase to demonstrate a high junction temperature (175 to 200°C) range of operation for higher-quality-grade products, a target of 0 PPM, demonstrate significantly higher field hours, and a more in-depth understanding of failure modes related to die, package, and system levels.

As emerging applications start to drive growth and define the requirements, opportunities will present itself for advances in manufacturing and device design to enable higher volume with better quality, reliability, and reduced cost. ed

ASHOK ALAGAPPAN is a Lead Consulting Engineer with ANSYS. He has 15 years of experience in semiconductor design and manufacturing, managing products through a range of activities from design to qualification. He has worked with customers in a variety of high-reliability applications from aerospace to telecommunications to provide expert analysis and recommendations for defining and improving reliability of electronics products and IC components. He has developed tools to predict the lifetime of IC components due to intrinsic wearout failure mechanisms that are featured in Sherlock Automated Design Analysis software.

INDEX

For more information on products or services visit our website www.electronicdesign.com. The advertisers index is prepared as an extra service. *Electronic Design* does not assume any liability for omissions or errors.

WILLIAM WONG | Senior Technology Editor
bill.wong@informa.com

Flying Above the Shoulders of GIANTS

NXP's HoverGames lets developers build on the open-source PX4 flight-control system with a real drone.

Thanks to open-source platforms, developers can stand on the shoulders of software giants to build bigger and better things. Linux is probably the biggest example, but a host of other open-source platforms are out there as well. One of those is the PX4 autopilot, which is the heart of many drones.

PX4 runs on a number of platforms, including Gumstix's AeroCore 2 for NVIDIA Jetson that supports a Jetson TX2. One of the latest platforms is NXP's RDDRONE-FMUK66 *(Fig. 1)*, which is designed to support NXP's new HoverGames competition. This flight management unit (FMU) is based on NXP's Kinetis K66 Arm Cortex-M4 microcontroller. The FMU can control servos and supports the two-wire 100Base-T1 automotive Ethernet standard.

The FMU is paired with NXP's drone *(Fig. 2)*. They're all part of a kit designed for the HoverGames. The challenge for this year's HoverGames is to develop a drone system to assist firefighters and other first responders. It's somewhat open-ended compared to more specific goals of other robotic competitions, but the hardware platform lends itself to more advanced solutions.

In particular, all of the kit's parts have been integrated, which is a big task. Developers could spend months just getting hardware and software to the point where a drone is flying. This includes the use of the PX4 autopilot software that's managed by DroneCore.org, which is also part of the Linux Foundation. In addition, DroneCode.org hosts projects like MAVLink, a communication protocol used by PX4, and a ground-station control system.

A host like an i.MX 8 or NVIDIA Jetson TX2 running the Robot Operating System (ROS) can control the PX4 system via MAVLink using MAVROS. The FMU has headroom for applications, but it lacks the horsepower to analyze video streams that one of these other hosts can easily handle.

HoverGames competitors will build on many open-source platforms and protocols. They will be able to concentrate on applications such as thermal imaging to identify people and animals, or track the progress of fires to provide firefighters and first responders with information they would normally lack without drone support.

The FMU will likely crop up in other projects, since the hardware and software are applicable to other robotic applications as well. I'm looking forward to seeing the winners of this year's competition. ed

1. The heart of the kit is a flight management unit (FMU) based on NXP's Kinetis microcontroller running the PX4 open-source software.

2. NXP delivers a drone kit for those competing in its HoverGames competition.