Electronic Design®

THE AUTHORITY ON EMERGING TECHNOLOGIES FOR DESIGN SOLUTIONS

MARCH/APRIL 2020
electronicdesign.com

5G TEST Takes to the Air

p24

$10.00

THINK™
HIGH PERFORMANCE

Densi-Pak® Water & Dust Resistant Battery Holders

• IP65 rated to withstand water and dust • Cover gasket seals out water and dust • Equipped with a built-in, sealed, on-off switch for additional security • Holders include coil spring contacts • Sealed, 6" long, pre-tinned wire leads • For "AA" batteries in two, three, and four cell configurations (Wired in series)

IT'S WHAT'S ON THE INSIDE THAT COUNTS

KEYSTONE®
ELECTRONICS CORP.

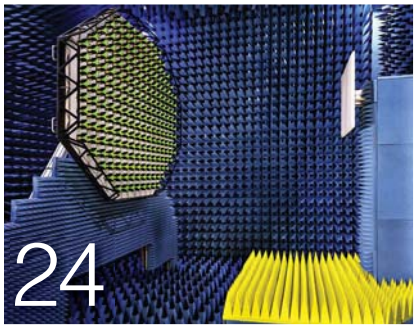
View our Dynamic Catalog M70 at www.keyelco.com

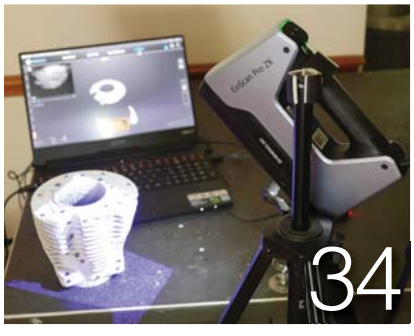
(516) 328-7500 (800) 221-5510

Request a copy of our Product Design Guide M70

IN THIS ISSUE

FEATURES





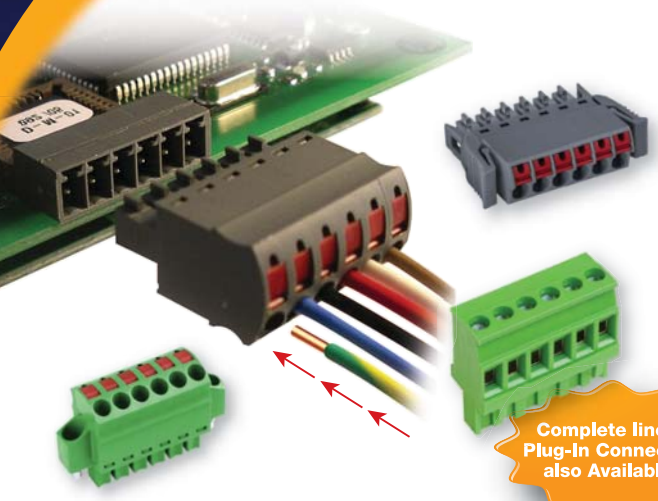
COLUMNS & DEPARTMENTS

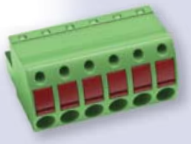
EDITORIAL MISSION:
To provide the most current, accurate, and in-depth technical coverage of the key emerging technologies that engineers need to design tomorrow's products today.

ELECTRONIC DESIGN (ISSN 0013-4872) is published bi-monthly in Jan/Feb, Mar/Apr, May/June, July/Aug, Sept/Oct and Nov/Dec by Endeavor Business Media, 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.

Altech®

Your Source for Electronic Components





Printed Circuit Board Terminal Blocks with ...

Push-in TECHNOLOGY

- Unique ergonomic 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

Complete line of Plug-In Connectors also Available !

The LARGEST Selection of

DIN ENCLOSURES

- 8 different series with over 240 parts
- DIN Rail and Panel Mount
- Fixed and Pluggable Terminals
- Integral PC Board Guides
- Many terminal to PCB connection options
- Excellent Quality at Competitive Prices

Customization & Labeling Available

High Quality & Versatile

ENCLOSURES

- 12 different series with over 700 enclosures
- UL approvals and IP protection ratings up to IP67 (NEMA 4X)
- Non-metallic and aluminum
- Light and heavy duty environments
- Corrosion proof, impact and chemical resistant, non-flammable and halogen free materials

State of the Art

TEST & MEASUREMENT

- Plugs
- Sockets
- Alligator Clips
- Test Probes
- Clamp-Type Test Probes
- Oscilloscope Probes
- Measuring Leads
- Sets

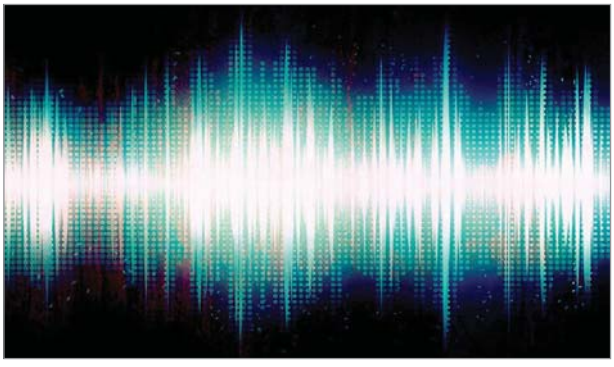
Altech Corp.®

Your Source for Automation & Control Components

info@altechcorp.com • 908.806.9400

www.altechcorp.com

ON ELECTRONICDESIGN.COM



Small Devices Deliver Big Audio with Boosted Amplifiers

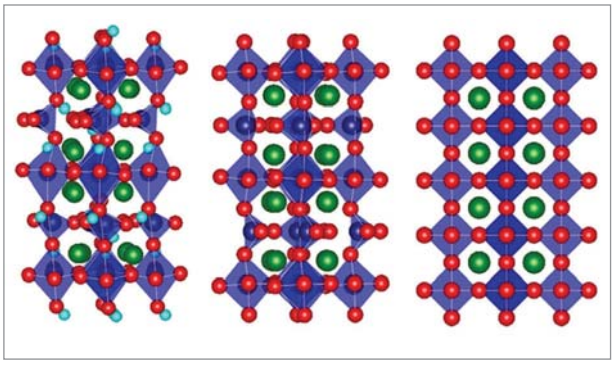
Users are consuming audio from very small devices, but lithium batteries and low-voltage power supplies often can't deliver loud-enough audio. Boosted amps are becoming an increasingly popular way to solve that problem.

https://www.electronicdesign.com/power-management/whitepaper/21125176/small-devices-deliver-big-audio-with-boosted-amplifiers

DragonSpeed Safety on Display at Sebring

DragonSpeed prepares for its first IndyCar race of the season at Sebring International Raceway. In particular, the new Advanced Frontal Protection cockpit protection system was put to the test.

https://www.electronicdesign.com/markets/automotive/article/21125701/dragonspeed-safety-on-display-at-sebring



At Last, Voltage-Tunable and Adjustable Thermal Conductivity

By adding oxygen or hydrogen to a thin-film oxide, researchers could reversibly change its thermal conductivity to higher and lower values than nominal, perhaps eventually leading to voltage-controlled, tunable thermal paths.

https://www.electronicdesign.com/technologies/analog/article/21126357/at-last-voltagetunable-and-adjustable-thermal-conductivity

Common PCB Soldering Problems to Avoid

Soldering can make or break a PCB, both figuratively and literally. This article offers basic tips and nine specific problems/mistakes to look out for when doing a soldering job.

https://www.electronicdesign.com/industrial-automation/article/21126406/common-pcb-soldering-problems-to-avoid

FOLLOW US ON FACEBOOK AND TWITTER

(ElectronicDesign) (@ElectronicDesgn)

Rethink everything you know about power supplies.

high voltage EVOLVED

- Patented Technology
- Remote Monitoring and Control
- CONFIGURE or CUSTOMIZE Features & Performance with Software

Long design times are a thing of the past.

Introducing the SPS Series of DIGITAL High Voltage Power Supplies

Outputs Up to 6kV & 30W

Contact DTI today to learn what the SPS Series can do for your design.

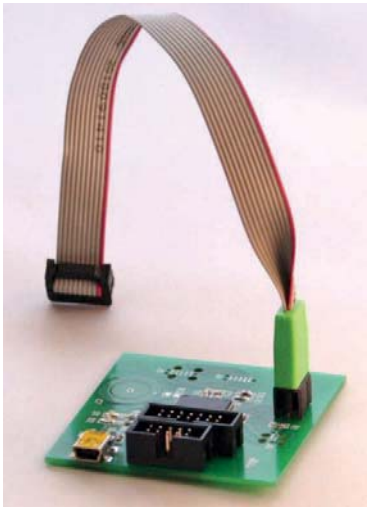
DEAN TECHNOLOGY

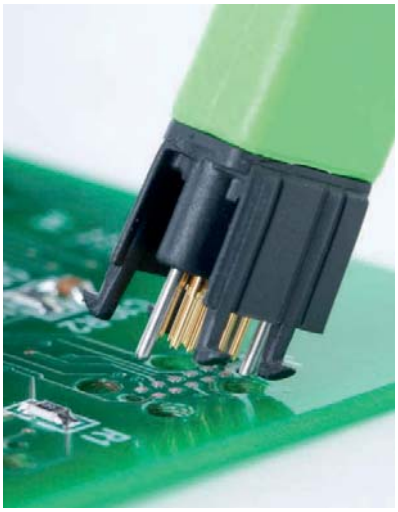
1-800-548-0344 II www.deantechnology.com

ElectronicDesign

MARCH/APRIL 2020

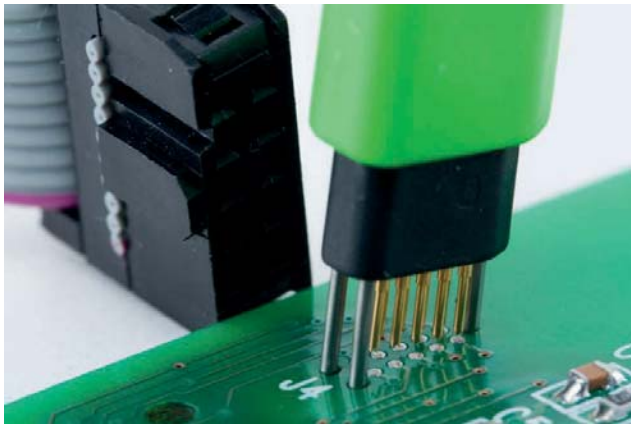
Tag-Connect

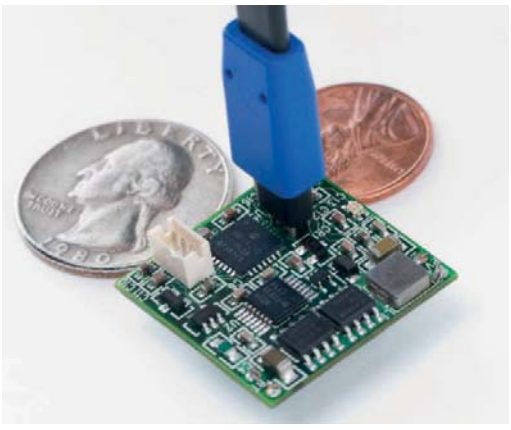




JTAG Connector Plugs Directly into PCB!!

No Header! No Brainer!





Our patented range of Plug-of-Nails™ spring-pin cables plug directly into a tiny footprint of pads and locating holes in your PCB, eliminating the need for a mating header. Spring-Pins tested to over 150,000 cycles!

Save Cost & Space on Every PCB!!

Solutions for: PIC . dsPIC . ARM Cortex . Atmel . JTAG . Altera . Xilinx . TI DSP . MSP430 . SPI / IIC . FTDI UART & More

Tag-Connector footprints as small as 0.02 sq. inch (0.13 sq cm)

www.Tag-Connect.com

Editorial

WILLIAM WONG | Senior Content Director
bwong@endeavorb2b.com

COVID-19 and the Electronics Industry

Editor Bill Wong comments on how COVID-19 is affecting the electronics industry, from supply chains to materials to trade shows.

Normally, the editorial for the print issue of *Electronic Design* is about the latest technology trend or articles in the issue, but the coronavirus, COVID-19, has turned everything on its head. I was at Embedded World this year, which is one of the few trade shows that hasn't been cancelled or moved. This was partially due to the fact that it occurred relatively early and steps like the U.S. travel ban to Europe were not in place. Almost any travel needs to take these kinds of actions into account in addition to considering the virus itself.

COVID-19 is having a significant impact on virtually every front. On the plus side, people will have more time to read *Electronic Design's* print issues and online articles, and quite a few will probably take advantage of our webinars and bootcamps because they're online.

Unfortunately, there's a much larger downside in our industry let alone the medical impact. Supply chains are being disrupted. Asia is the major supplier of electronics and China is where the virus problem is at its worst. A significant amount of uncertainty exists because of the changing scope as well as the reactions by governments and companies.

A number of the larger companies pulled out of Embedded World as they implemented a general travel ban. While it impacted the show, such a ban also affects their sales and support efforts in general. I was able to meet with many of the small- to medium-size companies that stayed, and some of the reasons they gave was that the show was one of the few where they can get exposure.

Many companies may not survive this crisis, especially if it lasts for a long time. Just-in-time delivery is feeling the effects as well as advertising and support. Dependencies in the supply change and availability of critical components can easily cause an otherwise profitable company to go bankrupt. Counting on tight margins is fine if everything is good, but few are likely to have contingency plans for what we're seeing with the impact of COVID-19.

Our industry tends to be readily supported by remote interaction support, such as video conferencing and other collaboration tools, but certain still actions still need be done in person. The question is how quickly everyone can pivot to utilize these alternative tools and get procedures in place to continue to work with customers and clients.

Another question to ask is what will happen after the pandemic runs its course, which it will eventually.

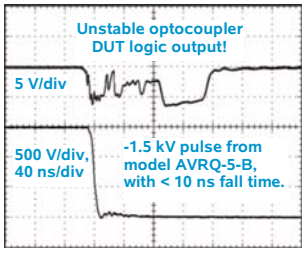
Are alternate sources of materials going to change how a company operates? Will the customer still be there in a few months, or has the impact of the virus caused them to fail or no longer need your product? Likewise, how will functionality like video conferencing, collaboration tools, and web-based tools affect the way you operate in the future? Many are inadequate as they stand, but there may be improvements or business processes may be altered to accommodate those limitations, making them a viable solution when face-to-face interaction becomes safe.

This industry was tough enough when dealing with new technologies ranging from securing the Internet of Things to adding artificial intelligence to the software mix. Throw in a potentially fatal virus and a host of new remote collaboration tools and the resulting, hopefully, workable solutions may change how we all work forever. We're all being forced to reconsider how we work with each other, what tools we will use, and what venues marketing and advertising will take. There are few winners and many losers at this point.

The picture may become clearer in a few months, particularly if things like a vaccine for the virus come about. Here's hoping that the effects of COVID-19 go away quickly and that our worries only concern whether or not that self-driving car down the road is safe to be around.

Transient Immunity Testers from AVTECH

The Avtech AVRQ series of high-voltage, high-speed pulse generators is ideal for testing the common-mode transient immunity (CMTI) of next-generation optocouplers, isolated gate drivers, and other semiconductors.



- Kilovolt amplitudes (±1, ±1.5, -2 kV)
- Transition times down to 10 ns, dV/dt rates up to 120 kV/us
- Switchable daughterboards to handle a variety of DUT package styles
- GPIB, RS-232 ports standard
- Ethernet / VXI optional

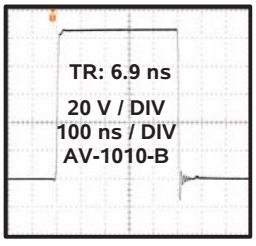
AVTECH

Nanosecond Electronics
Since 1975

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

30, 50 and 100 Volt Lab Pulsers

Avtech offers an extensive series of user-friendly 30, 50 & 100 Volt general-purpose lab pulsers. We can provide an alternative for the discontinued Agilent 8114A or HP214!



Model AV-1015-B: 50 Volts, 10 MHz
Model AV-1010-B: 100 Volts, 1 MHz, 25 ns to 10 ms, 10 ns rise time
Model AV-1011B1-B: 100 Volts, 2 ns rise time
Model AV-1011B3-B: 30 Volts, 0.5 ns rise time

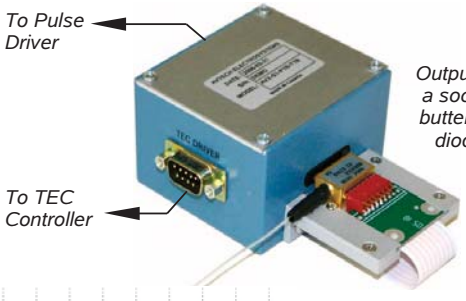
Variable baseline and burst mode options and output currents to 8 Amps with accessory transformers.

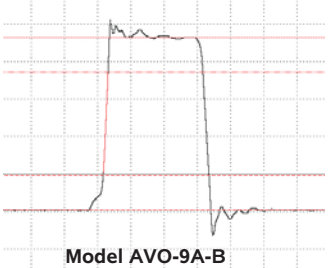
AVTECH

Nanosecond Electronics
Since 1975

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

Nanosecond Laser Diode Drivers With Butterfly Diode Sockets





Model AVO-9A-B
40 mA/DIV
1 ns/DIV

Each of the 19 models in the Avtech AVO-9 series of pulsed laser diode drivers includes a replaceable output module with an ultra-high-speed socket suitable for use with sub-nanosecond rise time pulses. Models with maximum currents of 0.1A to 10A are available with pulse widths from 400 ps to 1 us. GPIB, RS-232, and Ethernet control available.

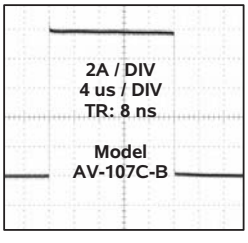
AVTECH

Nanosecond Electronics
Since 1975

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

2 to 200 Amp Current Pulsers

Avtech has pioneered the design of user-friendly 2 to 200 Amp constant current pulsers. We offer over 35 models in five series for laser diode, air bag, squib, fuse and other applications. The basic specifications are as follows:



Series	I, V	PW	TR
AV-107	2 - 20 A, 60 V	0.2 - 200 us	10 - 30 ns
AV-106	5 - 100 A, 100 V	0.5 us - 1 ms	50 ns - 1 us
AV-108	12.5 - 200 A, 100V	2 us - 1 ms	5 - 15 us
AV-109	10 - 100 A, 5 V	10 us - 1 s	10 us
AV-156	2 - 30 A, 30 V	1 us -100 ms	0.2 - 50 us

Avtech has a long history of producing one-of-a-kind custom units.

AVTECH

Nanosecond Electronics
Since 1975

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

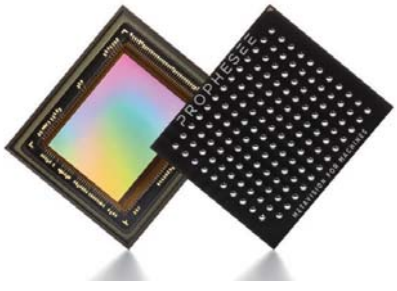
News

EVENT-BASED IMAGE SENSOR Views Videos in a New Way

Image processing is a demanding task and one reason is that it normally involves processing all of the pixels in an image. Multiple frames must be processed when dealing with video streams. Determining what changes from one frame to another is useful in detecting objects and other alterations, but it requires lots of horsepower and bandwidth to do that. This approach is needed because the image capture devices deliver a frame at a time.

But what if that wasn't the only way to get image information?

Prophesee's Metavision image sensor *(Fig. 1)* takes a different approach—it's an event-based sensor. Aim it at a solid color image and it generates almost no data because the pixels aren't changing. Move a hand in front of the background and the sensor will start sending a stream of events that indicate what pixels change. This tends to be a fraction of the overall number of pixels, which is typical of most scenes.



1. Prophesee's Metavision can deliver 66 million events/s with a dynamic range over 120 dB. A comparable imaging system would need to operate at 10,000 frames/s to keep up.

2. A conventional system would record full frames, but most of the pixels in the image don't change significantly. That's why data compression works well with video. The gray areas highlight what the Metavision chip would report as changes.

The sensor actually works on a programmable threshold for each pixel, allowing the system to have a wide dynamic range of over 120 dB. This is significantly higher than most image sensors. As a result, the sensor can detect changes even with a wide dynamic range, which would cause problems with a conventional sensor.

For example, sun glare or low light can a problem because the typical image sensor has a more limited range for each pixel. The sensor can be set up for a low-light situation, but then it would deliver a maximum white value if a bright light is in the scene. A typical color sensor may have RGB values, though only 8 or 16 bits per pixel.

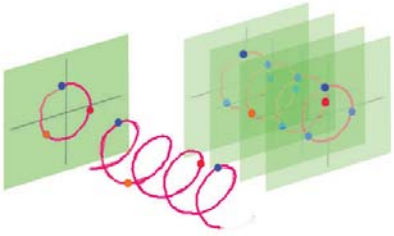
Figure 2 highlights the area where data would be generated by the Metavision sensor. In this case, the video is of a person swinging a golf club. The highlighted area is where pixels are changing.

There's a downside to the sensor as it doesn't report the color of each pixel—only the changes. An application may combine a conventional image sensor with the Metavision sensor if this type of information is needed, but the Metavision sensor changes the way your application analyzes video. It may only be interested in following the changes in an image.

For instance, an application may be tracking a hand gesture. It doesn't matter whether a person is wearing a blue glove or not. The application simply wants to recognize the gesture and the Metavision sensor can provide that information more economically.

FRAME-BASED VS. EVENT-BASED

Figure 3 attempts to highlight the difference between using a frame-based imaging system and event-based system. Imagine that there's a rotating disk with a dot on the periphery. The spiral is a mapping of the blue dot's position over time. Orange and red dots are highlighting the blue dot's position for particular points in time. The frames to the right would be how a conventional imaging system would report the data, although the actual dot is all we're concerned with. The circles and other dots are to provide a perspective.



3. Prophesee was demonstrating a rotating disk with a dot on the periphery. A conventional frame-based system would capture a full image, but the dot would be moving more quickly than the frames could be captured. The spiral highlights the data that the Metavision system would deliver.

The frame-based system would essentially show the dot jumping from one point to another. A sufficiently high frame rate would reveal the rotating nature of the system. Essentially, we have a Nyquist sampling issue.

Of course, a sufficiently fast frame-based solution will provide enough information for the video to be analyzed. However, this also means that a lot of data must be processed. If a machine-learning algorithm is being applied, then even more processing power is necessary.

The event-based system would deliver a significantly lower amount of information even though it could easily track the rotation. In fact, a microcontroller could easily handle the amount of data from this type of imaging system with a more complex scene and scenario. Of course, it could be overwhelmed by massive changes with a multitude of changing light conditions, reflections, etc., but that would be unusual in most cases.

Likewise, many applications have a more controlled environment, or the areas within the scene may be partitioned or managed in some fashion. The threshold that the Metavision sensor has for each pixel can be adjusted as well, allowing areas to essentially be ignored.

Prophesee is now delivering its sensor. "This is a major milestone for Prophesee and underscores the progress in commercializing our pioneering Event-Based Vision sensing technology. After several years of testing and prototyping, we can now offer product developers an off-the-shelf means to take advantage of the benefits of our machine-vision inventions that move the industry out of the traditional frame-based paradigm for image capture," said Luca Verre, co-founder and CEO of Prophesee.

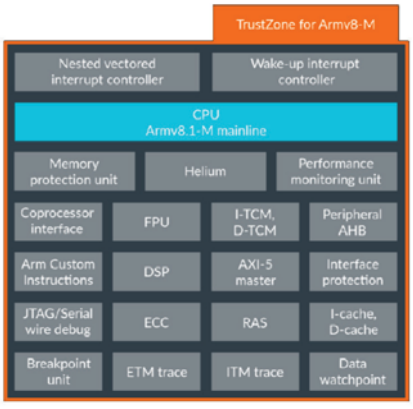
Frame-based video processing remains a useful paradigm. However, event-based video processing opens up a whole new area, potentially providing very-low-end platforms with the ability to handle image-processing chores that they couldn't handle if a frame-based input stream was used.

Development hardware and software available from Prophesee gives developers the ability to implement the Metavision sensor right away. The current sensor has a 640-×-480 resolution with a 15-µm pixel size in a 0.75-in. format. It has a 0.04-lux low-light cutoff and under 1-mHz background noise activity. The sensor comes in a 13- × 15-mm PBGA package. ■

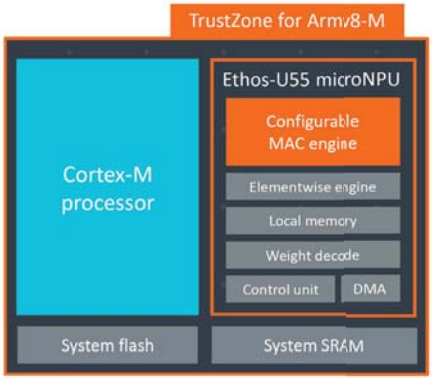
CORTEX-M Gets Heavy-Duty Machine Learning

MACHINE LEARNING (ML) has taken the developer community by storm, but implementing many algorithms with any efficiency have required FPGAs, multicore CPUs, or high-performance GPGPUs. Though developers have used ML on microcontrollers, the ML models are more limited. ML performance on these smaller conventional platforms is also a bit slower—useful but not capable of handling more ambitious projects.

This is about to change with Arm's announcement of the Cortex-M55 and the Ethos-U55. The Cortex-M55 can be used alone as it has its own ML augmentation. The Ethos-U55 can be added if ML applications are especially demanding. It could be paired with other Cortex-M platforms, too, but it will make more sense to be matched with the Cortex-M55, allowing ML support to be distributed.



1. The Cortex-M55 has a conventional Cortex-M that is augmented by the Helium ARMv8.1-M machine learning hardware acceleration.



2. The Ethos-U55 microNPU can have up to 256 MACs. It supports on-the-fly weight decompression to make more efficient use of the available memory bandwidth.

The Cortex-M55 *(Fig. 1)* is built around a conventional Cortex-M that includes the Helium ARMv8.1-M machine-learning hardware-acceleration support. Helium adds 128-bit vector extensions including gather/scatter support and complex math support such an 8-bit vector dot product. This enables the Cortex-M55 to deliver 15X the performance of a basic Cortex-M platform.

The Ethos-U55 *(Fig. 2)* is what Arm calls a microNPU (neural processing unit). It can provide 32X the performance of a base Cortex-M. Together, the Ethos-U55 and Cortex-M55 and deliver up to 480X the ML performance of a non-accelerated microcontroller.

The microNPU is designed to handle the ML heavy lifting while the Cortex-M manages the system and takes on lightweight ML models. This distributed approach allows many models to be employed in the system. It also lets developers balance performance and power. It's possible to employ just the Cortex-M55 for lighter-weight applications using less power while Ethos-U55 is powered down. Of course, the Ethos-U55 is able to efficiently handle video streams that would cripple lesser systems.

The Ethos-U55's scalable design can contain 32, 64, 128, or 256 MACs. DMA support for moving model weights supports on-the-fly decompression. There's also a built-in weight decoder. The accelerator communicates with the micro via system SRAM. The Ethos-U55 can be configured in a number of ways that run applications in parallel to the micro.

Both platforms support popular ML frameworks. For example, the Cortex-M55 can run TensorFlow Lite models. Both will handle multiple models at the same time. A unified development tool chain allows models to target either platform.

Expect to see chips based on the Cortex-M55 and the Cortex-M55 with an Ethos-U55 in about a year. A Cornerstone-300 reference design is available to get chip developers started.

"Enabling AI everywhere requires device makers and developers to deliver machine learning locally on billions, and ultimately trillions, of devices," said Dipti Vachani, senior vice president and general manager, Automotive and IoT Line of Business, Arm. "With these additions to our AI platform, no device is left behind as on-device ML on the tiniest devices will be the new normal, unleashing the potential of AI securely across a vast range of life-changing applications."

The Cortex-M55/Ethos-U55 sets the bar for microcontroller-based artificial intelligence (AI). ML applications that once targeted higher-end platforms will be practical on this micro-based solution that will consume less space and power. ■

MEETING RF DEMAND FOR CONNECTIVITY

ANYTIME, ANYWHERE!

Available for Same-Day Shipping!

TSX Series

4.3-10 Premium RF Surge Protection
Assuring Small Cell, RF Network Reliability

- 40 kA lightning surge rating
- Low let through energy (<300 NJ)
- Ultra low PIM typical -120 dBM
- Frequency ranges from 698 MHz to 2.7 GHz
- Fully weatherized
- 4.3-10 connector female-female and female-male

When small cell network reliability is a requirement, the only choice is PolyPhaser! Learn more at polyphaser.com or call +1 (208) 635-6400.

PolyPhaser™
an INFINITE brand

IMS

Connecting Minds. Exchanging Ideas.

LOS ANGELES 2020

CONNECTIVITY MATTERS

HOW DOES CONNECTIVITY MATTER TO YOU?

10,000

Attendees from 48 Countries of which 37% are First-Time Attendees!

Over **625** Exhibiting Companies

including **10%** First-Time Exhibitors!

20%

of Attendees are Senior Management

33%

of Attendees are Design Engineers or Engineering Management

11%

of Attendees are in Research and Development

5G

Dedicated tracks on 5G, Autonomous vehicles and Aerospace

Discover New Solutions and Gain Valuable Insights

Countless Networking Opportunities

Can't Miss Event for the RF & Microwave Industry!

REGISTER TODAY!

IEEE

www.ims-ieee.org

MTT-S

21-26 June 2020

IMS
Connecting Minds. Exchanging Ideas.

New!

LOS ANGELES 2020
CONNECTIVITY MATTERS

Technical Lectures at IMS and RFIC 2020!



"FUNDAMENTALS OF PHASED ARRAYS"
Marinos Vouvakis - *University of Massachusetts Amherst*

"QUANTUM COMPUTING: AN RF CONTROL PERSPECTIVE"
Evan Jeffrey - *Google Inc.*

"N-PATH MIXERS AND FILTERS: CONCEPT, THEORY AND APPLICATIONS"
Alyosha Molnar - *Cornell University*

"TRENDS IN AUTOMOTIVE RADARS: WAVEFORM, SYSTEM IMPLEMENTATION, AND IC TECHNOLOGIES"
Cicero Vaucher - *NXP Semiconductors*

"SILICON-BASED MILLIMETER-WAVE PHASED ARRAY DESIGN"
Bodhisatwa Sadhu - *IBM T. J. Watson Research Center*

"UNDERSTANDING OSCILLATOR PHASE NOISE AND LOCKING"
Ali Hajimiri - *Caltech*

"INTUITIVE MICROWAVE FILTER DESIGN WITH EM SIMULATION"
Daniel Swanson - *DGS Associates, LLC*

"TAXONOMY OF RF RECEIVERS: FROM BASICS TO LATEST TECHNIQUES"
Peter Kinget - *Columbia University*

IMS2020 has introduced Technical Lectures to create a unique opportunity to get a condensed, curated technical look at current RF and microwave design topics taught by the top researchers.

REGISTER BY 26 MAY 2020 AND SAVE 30%!

FOR THE LATEST ON IMS AND MICROWAVE WEEK VISIT WWW.IMS-IEEE.ORG

IEEE 21-26 JUNE 2020 | LOS, ANGELES, CA

MTT-S

RFSoC Board Aligns with SOSA Reference Architecture

Pentek's latest RFSoC board puts it on the vanguard of SOSA Technical Standard adoption.

With the launch of its Quartz model 5550 RFSoC board, Pentek stakes out a leadership position in the defense/aerospace sector's adoption of the Sensor Open System Architecture (SOSA) Technical Standard and the reference architecture it defines. To that end, the board implements connector technology that enables a major goal of the SOSA reference architecture—backplane-only I/O. It incorporates the ANSI/VITA 67.3D VPX backplane interconnect standard for both coaxial RF and optical I/O. In addition, the Model 5550 includes a 40-GbE interface and a shelf-management subsystem that are also required by the SOSA reference architecture.

The 3U OpenVPX board, equipped with PCI Express Gen 3 capabilities, comprises an eight-channel analog-to-digital converter (ADC) and digital-to-analog converter (DAC) and is based on the Xilinx Zynq UltraScale+ RFSoC FPGA. It's aimed squarely at applications in communications, electro-optics, electronic warfare, and radar and signals intelligence.

Pentek's modular approach to hardware and software enables quick adaptation to new and changing customer requirements. The Model 5550 uses the Model 6001 QuartzXM eXpress module containing the RFSoC FPGA and all needed support circuitry implemented on a carrier module designed specifically to align with the technical standard for the SOSA reference architecture. This allows for easy upgrades to third-generation RFSoC modules when available.

The Model 5550 is pre-loaded with a suite of Pentek IP modules to provide data capture and processing solutions for many common applications. Modules include direct-memory-access (DMA) engines, DDR4 memory controller, test signal and metadata generators, data packing, and flow control. The board also comes pre-installed with IP for triggered waveform and radar chirp generation, triggered radar-range gate selection, wideband real-time transient capture, flexible multimode data acquisition, and extended decimation. For many applications, the Model 5550 can be used out-of-the-box with these built-in functions, requiring no FPGA development.

The front end accepts analog IF or RF inputs on eight coax connectors located within a VITA 67.3D backplane connector. After balun coupling to the RFSoC, the analog signals are routed to eight 4-GS/s, 12-bit ADCs. Each converter has built-in digital downconverters with programmable 1x, 2x, 4x, and 8x decimation plus independent tuning. The ADC digital outputs are delivered into the RFSoC programmable logic and processor system for signal processing, data capture, or routing to other resources. A stage of IP-based decimation provides another 16x stage of data reduction, ideal for applications that need to stream data from all eight ADCs.

Eight 4-GS/s, 14-bit DACs deliver balun-coupled analog outputs to a second VITA 67.3D

coaxial backplane connector. Four additional 67.3D coaxial backplane connections are provided for clocks and timing signals.

The Model 5550 also uses the VITA-67.3D backplane connector for eight 28-Gb/s duplex optical lanes to the backplane. With two built-in 100 GigE UDP interfaces or a user-installed serial protocol in the RFSoC, the VITA-67.3D backplane interface enables gigabit communications independent of the PCIe interface.

NAVIGATING IP

For cases in which IP development proves necessary, Pentek's Navigator Design Suite streamlines the process. The suite includes the Navigator FDK (FPGA Design Kit) for custom IP and Navigator BSP (Board Support Package) for creating host software applications. The Navigator FDK includes the board's entire FPGA design as a block diagram that can be edited in Xilinx's Vivado tool suite. All source code and complete documentation is included. Developers can integrate their IP along with the factory-installed functions or use the Navigator kit to replace the IP with their own. The Navigator FDK Library is AXI-4 compliant, providing a well-defined interface for developing custom IP or integrating IP from other sources.

The Navigator BSP supports Xilinx's PetaLinux on the Arm processors. Users can work efficiently using high-level API functions, or they can gain full access to the underlying libraries including source code. Pentek provides numerous examples to assist in the development of new applications.

Serving as a ready-to-use Quartz development platform, the Model 8257 is a low-cost 3U VPX chassis well-suited for developing applications on Pentek's Model 5550 Quartz RFSoC board. Providing power and cooling to match the 5550 in a small desktop footprint, the chassis allows access to all required interfaces and the Model 5901 rear transition module. The 8257 can be configured with optional real-panel dual MPO optical connectors to support the 5550's dual 100-GbE interfaces and coaxial RF connectors.

Designed for air-cooled, conduction-cooled, and rugged operating environments, the 5550 board starts at $38,745. Options for optical interface, GPS support, and memory are available. Deliveries begin in 3Q 2020. The Navigator BSP and Navigator FDK are priced at $2,500 and $3,500, respectively. Both include free lifetime support. ed

IAN SAGAN | Field Applications Engineer, Marvell

Determining the Best Option for NVMe-over-Fabrics

The NVMe-over-Fabrics data-storage architecture could be a game-changer for system administrators in terms of making more efficient use of available memory resources.

Non-volatile memory express (NVMe) presents a streamlined means for delivering low-latency operation, and it's being adopted by all major server and storage vendors *(see figure)*. The architecture is replacing Small Computer System Interface (SCSI) in modern solid-state-disk (SSD) resources.

As a result of its adoption, NVMe-enabled SSDs are now eliminating the bottlenecks that were inherent in traditional storage implementations—leading to substantial improvements in the access speeds for a range of applications from mobile platforms to enterprise data centers. This article answers critical questions concerning NVMe-enabled SSDs deployment and use at scale.

WHY IS NVMe SEEING SUCH WIDESPREAD PROLIFERATION?

NVMe is designed from the ground up to communicate at high speed with flash storage, and only requires 30 commands that are specific to dealing with SSDs. In addition, this architecture supports multiple deep command queues in order to take advantage of the parallel-processing capabilities of the latest multicore processors. With up to 64K commands per queue and support of up to 64K queues, NVMe represents a major advance over traditional SCSI, SAS, and SATA technologies, which were originally developed for spinning hard-disk drives (HDDs).

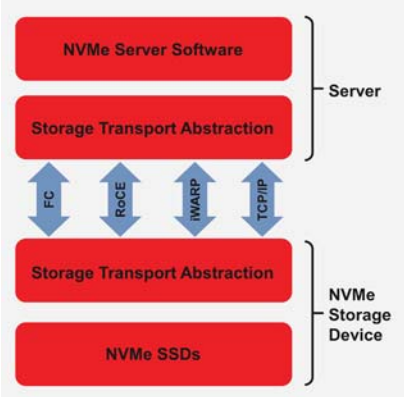
Global sales in NVMe-based SSD drives are now overtaking those relating to SAS and SATA SSD storage.[1] This is due to the dramatically improved performance offered by NVMe, for both current and next-generation SSD technologies (such as 3D XPoint and non-volatile DIMMs, or NVDIMMs).

WHY DO YOU NEED TO CONSIDER NVMe-OVER-FABRICS?

When conceived, NVMe's primary objective was to allow central processing units (CPUs) to access NVMe-based SSDs within the server using the PCIe bus. However, as storage administrators are aware, local server storage provides major administration headaches—particularly in terms of having to over-provision expensive SSD storage resources (so that there's adequate headroom factored in to cope with any excess demand).

Different servers require different amounts of high-performance NVMe SSD storage depending on their application workloads. These applications can migrate to different physical servers but still require the same amount of SSD storage. To stop the practice of overpopulating every server with expensive SSD storage, it's more economical and efficient to create a pool of shared NVMe SSD storage, which can be dynamically allocated depending on workloads.

With local storage, it's paramount that all data is backed up, in case the server fails. Furthermore, there are serious security implications, and replication between sites can become difficult to manage. Through shared storage, administrators can avoid these issues. In other words, CIOs can utilize high-performance flash storage to its full degree across servers—with all of the elevated availability and security capabilities of modern storage arrays, plus similar performance and latency advantages associated with local NVMe SSD storage.



This is an overview of the key NVMe fabric connectivity options currently available.

HOW FAR DOWN THE ROAD IS NVMe-OVER-FABRICS?

To help describe this, let's compare shared storage arrays to automobile engines. Specifically, traditional Fibre Channel (FC)/iSCSI shared storage arrays can be equated to conventional combustion engines. They have been used for many years, are reliable, and

will provide a good method of transportation for a prolonged period of time.

Hybrid cars are becoming more commonplace and provide some of the benefits of electric and gas. In a similar manner, newer NVMe arrays use a mix of NVMe internally inside the array, but they connect to the host via SCSI commands over either FC or Ethernet transport protocols.

Though most people agree that electric cars will be the future, currently they're not mainstream due to their cost versus traditional alternatives and the infrastructure isn't broadly available yet to support electric charging. Native NVMe arrays can be thought of in the same way as all-electric vehicles. The infrastructure needed to make them a reality is NVMe-over-Fabrics. In time, this will become the dominant communication standard for connecting shared storage arrays to servers. Still, it will take time before NVMe-over-Fabrics gains widespread proliferation and all of the related teething problems are resolved.

WHAT NVMe-OVER-FABRICS OPTION SHOULD YOU CHOOSE?

The biggest dilemma for storage administrators is deciding on the right technology to invest in. As with any new technology when it first emerges, there are multiple ways to implement the overall solution. In this respect, NVMe-over-Fabrics is no different. NVMe commands can be passed across FC, RDMA-enabled Ethernet, or standard Ethernet using TCP/IP. Let's look at the key differences in these approaches:

NVMe-over-FC (FC-NVMe)

FC-NVMe is a great choice for those who already have FC storage-attached-network (SAN) infrastructure in place. The NVMe protocol can be encapsulated in FC frames using 16GFC or 32GFC host bus adapters and switches. Support for FC-NVMe on Linux servers can be gained by upgrading to the latest FC firmware and drivers. Therefore, investing in modern 16- or 32-Gb FC host bus adapters and SAN infrastructure provides future-proofing for FC-NVMe arrays when they're released. It's also worth noting that both SCSI (FCP) and NVMe (FC-NVMe) can coexist on the same fabric. Thus, legacy FC-SCSI-based arrays can run concurrently with new NVMe native arrays.

NVMe-over-Ethernet Fabrics using RDMA (NVMe/RDMA)

For this RDMA-capable Ethernet, adapters are mandated. There are two different types of RDMA implementation: RDMA-over-converged-Ethernet (RoCE) and Internet wide-area RDMA protocol (iWARP). Unfortunately, these protocols aren't interoperable with one another. Their benefits include:

- *NVMe-over-RoCE (NVMe/RoCE):* When using Ethernet only networks, NVMe-over-RoCE is a good choice for shared storage or hyper-converged infrastructure (HCI) connectivity. Consequently, many storage array vendors have announced plans to support NVMe-over-RoCE connectivity. RoCE provides the lowest Ethernet latency and works well when the storage network involved is small scale, with no more than two hops. As the name implies, RoCE requires a converged or lossless Ethernet network to function. This approach enables extra network capabilities, including data-center bridging, priority flow control, plus more complex setup and network-management mechanisms. If low latency is the overriding goal, then NVMe-over-RoCE is likely to be the best option, despite the heightened network complexity.
- *NVMe-over-iWARP (NVMe/iWARP):* The iWARP RDMA protocol runs on standard TCP/IP networks, and is therefore much more straightforward to deploy. Though its latency isn't quite as good as that of RoCE, its greater ease of use and much lower administration overheads are very appealing. At this stage, storage array vendors aren't designing arrays to support iWARP, so iWARP is for the moment best suited for software-defined or HCI solutions like Microsoft Azure Stack HCI/Storage Spaces Direct (S2D).

NVMe-over-TCP (NVMe/TCP)

NVMe-over-TCP is the new kid on the block. Ratified in November of 2018, it runs on existing Ethernet infrastructure with no changes necessary (taking advantage of the incredible pervasiveness of TCP/IP). The performance that NVMe-over-TCP delivers may not be quite as fast as NVMe-over-RDMA or FC-NVMe, but it can be easily implemented on standard network interface cards (NICs) and network switches. The key benefits of NVMe SSD storage can still be derived without requiring significant hardware investment. Some NICs like the Marvell FastLinQ 10/25/50/100GbE have the potential to offload and accelerate NVMe/TCP by leveraging built-in capabilities of offloading the TCP/IP stack in the NIC.

SUMMARY

Whichever NVMe-over-Fabrics route you decide to undertake, Marvell can offer assistance in the implementation process, and with its product, particularly the QLogic 16- and 32-Gb FC host bus adapters that support FC-NVMe. In addition, the company's FastLinQ 41000 and 45000 series of 10/25/40/50/100Gb Ethernet NICs and converged network adapters support both NVMe-over-RoCE and NVMe-over-iWARP functionality, as well as NVMe-over-TCP (thanks to the Universal RDMA feature incorporated). ED

REFERENCE:

1. According to figures from Research and Markets *https://www.prnewswire.com/news-releases/global-data-center-flash-storage-market-to-2024---nvme-flash-storage-to-replace-sassata-flash-solutions--emergence-of-qlc-nand-flash-drives-300854343.html*

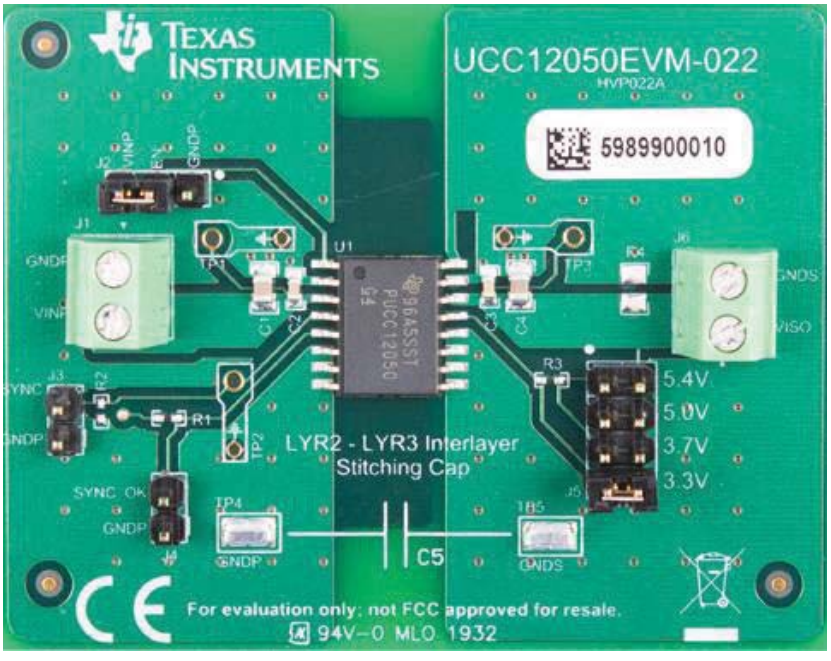
LEE GOLDBERG | Contributing Editor/Power, *Electronic Design*

POWERBITES

A Round-Up of News-Making Power Products

TI power supplies, MagnaChip's BCD process, prototyping tools for ECUs from STMicro, and MIT's potentially groundbreaking electrode for Li-ions raised eyebrows.

How did TI manage to shoehorn an isolation transformer into its new dc-dc supply chips? Will MIT's process for creating pure lithium anodes enable quantum leaps in Li-ion battery capacity and longevity? Can STMicro's new automotive development platform really automate most of the details involved with creating custom ECUs? What new applications will MagnaChip's automotive-certified 0.13-micron process enable by making it possible to fabricate bipolar, CMOS, and DMOS on a single chip? The answers for these and other pressing questions can be found in this week's installment of PowerBites.



1. By integrating an isolation transformer into its package, TI's UCC12050 requires up to 80% less PCB space than a conventional solution. *(Courtesy of Texas Instruments)*

TI'S ISOLATED POWER SUPPLIES FEATURE COMPACT FOOTPRINT, INDUSTRY-LEADING EMI

Texas Instruments has introduced the first in a series of high-efficiency isolated dc-dc converters that use an integrated isolation transformer to achieve what they believe to be the industry's lowest electromagnetic-interference (EMI) levels. TI's UCC12050 can deliver up to 500 mW of isolated 5.0- or 3.3-V output power at 60% efficiency *(Fig. 1)*. Built for industrial and medical requirements, the UCC12050's 5-kVrms reinforced isolation and 1.2-kVrms working voltage protects industrial transport, grid infrastructure, and medical equipment against high-voltage spikes without any additional components.

To achieve high performance and reduce component count, the converter's transformer and its companion EMI shield are integrated into the device's package along with the chip. By integrating a complete isolated dc-dc solution into a 16-pin small-outline integrated circuit (SOIC) package measuring 10.3 × 10.3 × 2.65 mm, a converter based on the UCC12050 requires up to 80% less PCB space than a conventional solution. TI also offers the UCC12040, a lower-cost device intended for less-demanding applications, which provides all of '2050's features and basic isolation (3 kVrms).

The UCC12050 and UCC12040 are available in volume production from TI and authorized distributors. Pricing starts at US$3.90 and US$3.15, respectively, in 1,000-unit quantities. Engineers can evaluate this product with the UCC12050EVM-022 evaluation module, which goes for US$99.

MAGNACHIP'S 0.13-µM BCD PROCESS TARGETS AUTOMOTIVE POWER APPS

Korea-based MagnaChip Semiconductor Corp. announced today that it added an enhanced-performance, 0.13-µm bipolar-CMOS-DMOS (BCD) process to its portfolio of foundry processes intended to help automotive power-semiconductor designers build more competitive products. The BCD process technology combines three different process technologies onto a single chip: CMOS for digital functions, bipolar for analog signal control, and DMOS high-power functions, such as motor drivers and actuators. It supports an IP for an electrically erasable programmable ROM (EEPROM), which can be reprogrammed at least 1000 times.

The process has been certified as Grade-1 under the AEC-Q100 automotive reliability standard, making it suitable for manufacturing many types of automotive power devices, including motor-driver ICs, battery-management systems (BMSs) and dc-dc converters.

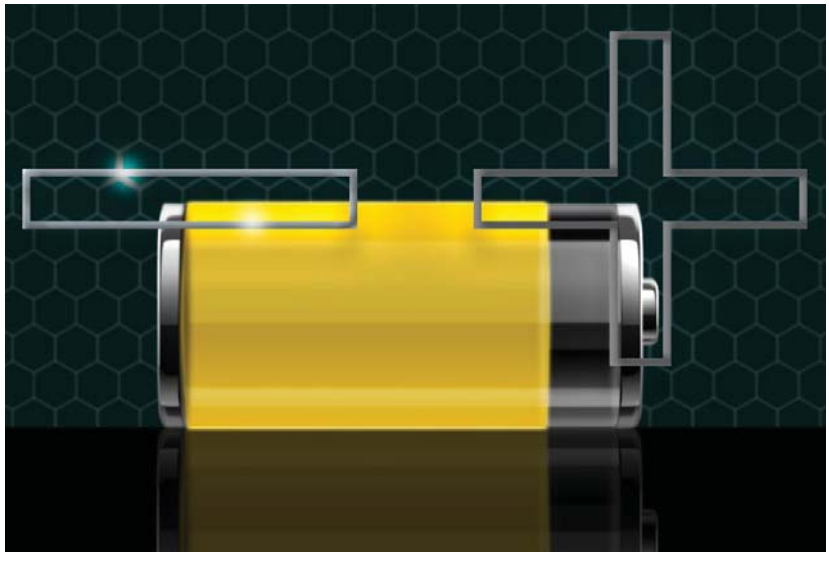
For more information, visit *www.magnachip.com*.

STMICRO's NEW DEVELOPMENT TOOLS SIMPLIFY PROTOTYPING OF AUTOMOTIVE ECUs

The latest version of its AutoDevKit ecosystem developed by STMicroelectronics includes tools that assist the development of electronic control units (ECUs). ECUs manage the numerous the growing number of motors, actuators, and other components found in modern electric and conventional vehicles. Available for free, the AutoDevKit library is a software environment that lets users select the microcontrollers and functional boards from ST's wide automotive portfolio. Once configured, the AutoDev platform guides users to connect the boards, generate code, and compile and download firmware, without forgetting prototype testing and debugging.

Also included within the AutoDevKit ecosystem are AEKD System Solution Demonstrator components that provide direct access to pre-assembled system-demonstrator boards. Furthermore, board kits and non-electronic hardware assemblies such as car-like components and loads can help users closely emulate the desired solution.

Further information and free downloads can be found at *http://www.st.com/autodevkit*.



2. MIT's process for fabricating all-lithium anodes is part of a concept for developing safe all-solid-state batteries, which don't require the liquid or polymer gel typically used as the electrolyte material between the battery's two electrodes. *(Courtesy of MIT)*

MIT'S NEW ELECTRODE MAY LEAD TO MORE POWERFUL, LONGER-LIFETIME LITHIUM BATTERIES

New research by engineers at MIT and elsewhere could lead to batteries that pack more power per pound and last longer, based on the long-sought goal of using pure lithium metal as the battery's anode. The design is part of a concept for developing safe all-solid-state batteries, which don't require the liquid or polymer gel often currently used as the electrolyte material between the battery's two electrodes.

An electrolyte allows lithium ions to travel back and forth during the charging and discharging cycles of the battery. An all-solid version could be safer than liquid electrolytes, which have high volatility and have been the source of explosions in lithium batteries.

The concept for the new electrode comes from the laboratory of Ju Li, the Battelle Energy Alliance Professor of Nuclear Science and Engineering and professor of materials science and engineering. It's described in the journal *Nature*, in a paper co-authored by Yuming Chen and Ziqiang Wang at MIT, along with 11 others at MIT and in Hong Kong, Florida, and Texas.

"There has been a lot of work on solid-state batteries, with lithium metal

electrodes and solid electrolytes," says Li, but these efforts have faced a number of issues.

One of the biggest problems is that when the battery is charged up, atoms accumulate inside the lithium metal, causing it to expand. The metal then shrinks again during discharge, as the battery is used. These repeated changes in the metal's dimensions, somewhat like the process of inhaling and exhaling, make it difficult for the solids to maintain constant contact. Subsequently, it tends to cause the solid electrolyte to fracture or detach.

To overcome this, Li and his team adopted an unusual design that utilizes two additional classes of solids—"mixed ionic-electronic conductors" (MIEC) and "electron and Li-ion insulators" (ELI), which are absolutely chemically stable in contact with lithium metal. They then developed a three-dimensional nanoarchitecture in the form of a honeycomb-like array of hexagonal MIEC tubes, partially infused with the solid lithium metal to form one electrode of the battery, but with extra space left inside each tube *(Fig. 2)*.

When the lithium expands in the charging process, it flows into the empty space in the interior of the tubes, moving like a liquid, even though it retains its solid crystalline structure. This flow, entirely confined inside the honeycomb structure, relieves the pressure from the expansion caused by charging, but without changing the electrode's outer dimensions or the boundary between the electrode and electrolyte.

Additional details can be found in the complete story, available at: *https://news.mit.edu/2020/solid-batteries-lithium-metal-electrode-0203.*

> "There has been a lot of work on solid-state batteries, with lithium metal electrodes and solid electrolytes," says Li, but these efforts have faced a number of issues.

RENESAS' RAD-HARD DUAL-OUTPUT BUCK CONVERTER + LDO TARGETS CIVILIAN, MIL-SPACE APPS

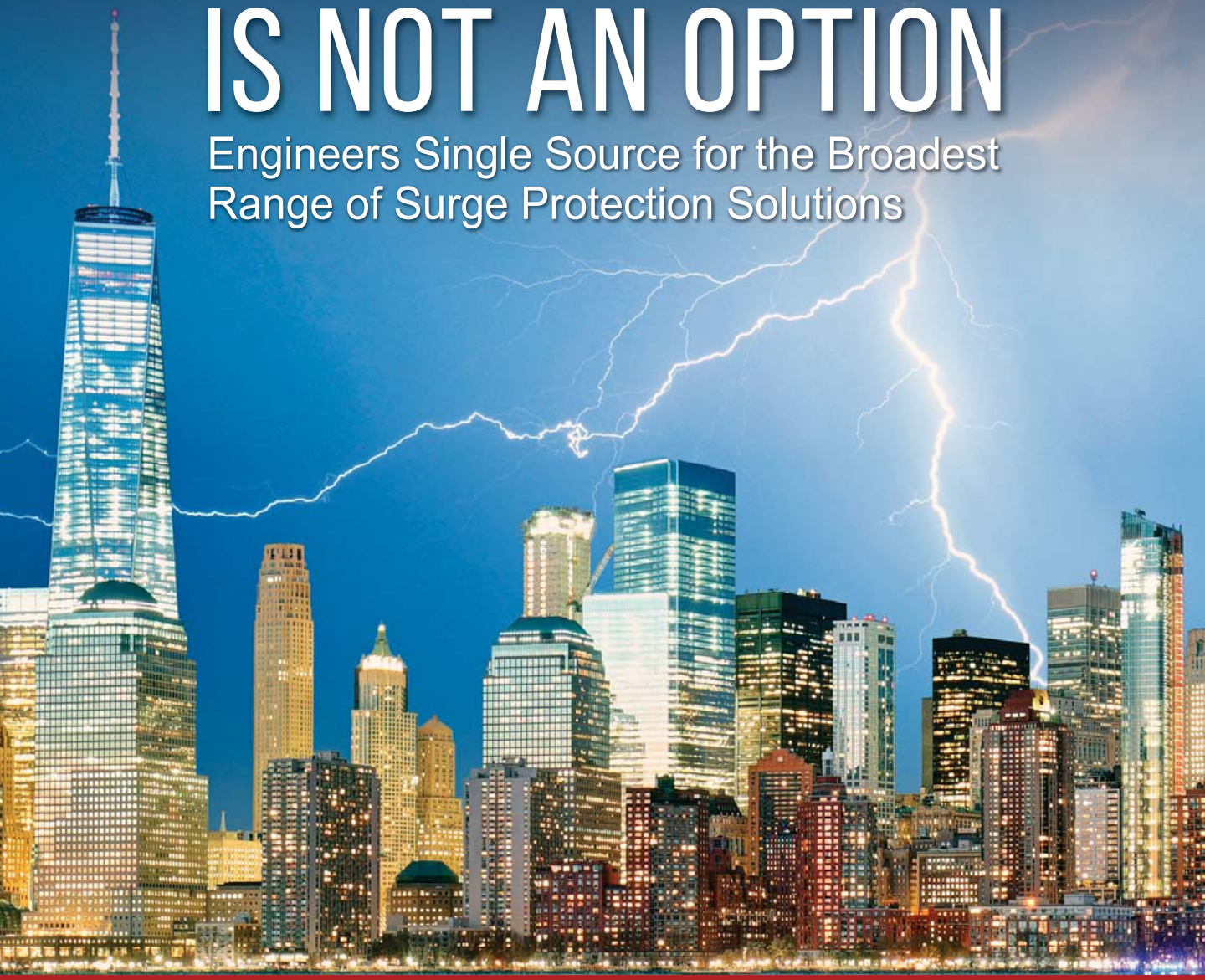
The ISL70005SEH is a rad-hard single-chip synchronous buck and low-dropout (LDO) regulator intended for spaceflight payload applications, such as low-power FPGAs, DDR memory, and other digital loads. By integrating a 95% efficient synch buck regulator and an LDO, Renesas' rad-hard point-of-load power solution reduces size, weight, and power (SWaP). It also simplifies the assembly and component count of the electronics used in medium-Earth-orbit (MEO) and geosynchronous-Earth-orbit (GEO) long-duration missions. Key features include:

- Synchronous buck V_{IN} range of 3 to 5.5 V
- LDO V_{IN} range of 600 mV + V_{DO} to V_{CC}-1.5 V
- 1% reference-voltage accuracy
- Separate V_{IN}, enable, soft-start, and power good indicator
- LDO stable with 150 µF; 3X less output capacitance than competitive solutions
- Full military temperature range of −55 to +150°C
- Wafer acceptance tested to 100 krad(Si) over high dose rate
- Tested for ELDRS up to 75 krad(Si) over low dose rate

The ISL70005SEH radiation-hardened dual-output PoL regulator is available now in a 28-lead ceramic dual flatpack package or in die form. An evaluation board also is available to evaluate device features and performance. For more information, visit: *www.renesas.com/products/ISL70005SEH.*

WHEN DOWNTIME IS NOT AN OPTION

Engineers Single Source for the Broadest Range of Surge Protection Solutions



AVAILABLE FOR SAME-DAY SHIPPING

- AC | DC | Data | EMP Surge Protection
- Weatherized NEMA Enclosures
- Wide Range of Product Configurations
- Dedicated Customer & Technical Service
- Ecommerce Website Offering Same-Day Shipping

Call +1 (208) 635-6400 or go to transtector.com to keep your power online.

TRANSTECTOR™

an INFINITE brand

CLINTON LINVILLE | Application Engineer, Rohde & Schwarz

Delivering 5G Devices to Market Will Bank on OTA Testing

As 5G begins to leverage high-frequency mmWave bands, engineers that used to work at RF frequencies will once again need to sharpen their skill sets and adopt new design and testing techniques.

Shown is Rohde & Schwarz's ATS800R rack-mount CATR system that performs OTA testing.

The biggest challenge in testing 5G devices, as opposed to 4G/LTE, is the reliance on active antenna arrays and high RF frequencies. These frequencies are far higher than what we have ever used before in a commercial communication system. Frequency range 1 (FR1), which will transmit most of the traditional cellular communications traffic, has been designated for 450 to 7125 MHz. Frequency range 2 (FR2) will employ millimeter-wave (mmWave) frequencies (24,250 to 52,600 MHz) to deliver short-range, high-data-rate transmission.

Higher frequency bands combined with extended bandwidths in the mmWave range will place much higher demands on the components for 5G communications devices and systems, including the filters, mixers, amplifiers, analog beamforming chipsets, and antennas.

For FR1, which is sub-6-GHz, testing can still be performed using cables. But at FR2, it will be necessary to consider the entire assembly as one entity, including the antenna, phase shifters, amplifiers, attenuators, and more. Testing needs to be performed at a system level, so that connectors and cables don't interfere with the system characterization.

NEW TECHNOLOGIES, MORE COMPLEXITY

FR2 products are becoming very complex. They're comprised of phased-array antennas; each is built of many antenna elements, every one with its own phase shifter and amplifier, working collectively to steer a signal in a desired direction.

In this case, if testing is performed using a connector that bypasses the antenna array, it's impossible to gauge how the overall system performs: If the system lacks proper beamsteering or gain, or if the components behind it are too noisy and impact the modulation of the signal, it will not be discovered during the testing phase. Therefore, it's critical to characterize the entire system—the radio and the phased-array antenna together.

In multiple-input, multiple-output (MIMO) technology, multiple antennas are used at both the transmitting and receiving points, creating a circuit that minimizes errors and optimizes speed. Traditionally, testing MIMO signals has focused on testing the transmitter/receiver system and the quality of the

Matchmaker

0402DC Series Designer's Kit (C472-2):
20 samples each of 112 inductance values, including 0.1 nH increments from 2.8 nH to 10 nH

Looking for the perfect high-Q inductor for impedance matching in RF/microwave antenna circuits? This kit has it!

Coilcraft 0402DC Series wirewound chip inductors offer the industry's highest Q factors in an 0402 (1005) size for super low loss in high frequency circuits. And with 112 values from 0.8 to 120 nH, including **0.1 nH increments from 2.8 nH to 10 nH**, you'll have exactly what you need for all your RF and Microwave applications.

The 0402DC also features wirewound construction for extremely high self resonance – up to 28.8 GHz – and offers DCR as low as 25 mΩ, significantly lower than other inductors this size.

Equip your lab with the ultimate impedance matching resource. Our C472-2 Designer's Kit has 20 samples of all 112 values! Purchase one online at **www.coilcraft.com/0402DC.**

Coilcraft®
WWW.COILCRAFT.COM

coilcraftdirect.com
No min order. Next day delivery.

channel. Facilitating commercially feasible testing in 5G devices could be a challenge.

However, when designing a massive-MIMO active antenna system, development engineers face new challenges that include phase-shifter tolerances, thermal effects of the power amplifiers (PAs), and frequency drifts between modules that affect the desired beam patterns.

In an active antenna system, the transceiver front-ends are integrated together with the antenna array, which means that traditional RF output ports are no longer accessible. In addition, a fiber interface replaces the traditional RF input ports for digital I/Q data. Consequently, over-the-air (OTA) testing becomes the default use case for massive-MIMO systems and for modeling the spatial properties of the propagation channel. Due to the different sizes of massive-MIMO systems, testing in far-field conditions requires a variety of shielding environments.

It's still a bit early to consider what a MIMO test solution for 5G will look like. Nonetheless, we anticipate the need to consider multiple angles of arrival of signal, application of the proper fading channels, etc. Currently, it's difficult to see what this looks like for FR2.

The 3rd Generation Partnership Project (3GPP), which unites seven telecommunications standard development organizations (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, and TTC), is actively developing the reports and specifications that define 5G technologies. The project covers cellular telecommunications technologies, including radio access, core network, and service capabilities, which provide a complete system description for mobile telecommunications.

The 3GPP technologies from these groups are constantly evolving through generations of commercial cellular/mobile systems. With its LTE, LTE-Advanced, LTE Advanced Pro, and 5G work, 3GPP has become the focal point for the vast majority of mobile systems beyond 3G. A number of Rohde & Schwarz representatives regularly attend 3GPP's meetings globally, including those of the standards committees, and the company has contributed input and feedback on testing devices and systems to meet the specifications the organization lays out. When the standards are finalized, they will drive the future certification process.

The major focus for all 3GPP Releases is to make the system backward- and forward-compatible where possible, to ensure that the operation of user equipment is uninterrupted. For 5G, many operators are starting with dual connectivity between LTE and 5G NR equipment—using a non-standalone specification detailed in Release 15 that was completed earlier this year. Care has been taken to build forward compatibility into non-standalone NR equipment, to ensure that it will be fit for use on standalone 5G NR systems.

THE EVOLUTION OF OTA TESTING

In terms of testing, the standards are changing from connectorized measurements to an indirect far-field measurement using a compact antenna test range (CATR), as it provides the most flexibility and a practical solution moving forward. For example, if you were to measure an FR2 phased array in a true far-field chamber, you would be looking at a range length of several meters, which is impractical.

Today's systems can execute the antenna characterization or antenna check using test solutions such as a compact antenna range that emulates far-field conditions within a smaller area/footprint. Moving forward, CATR will be the test system of choice to test phased arrays at high frequencies for user equipment.

One such compact solution is the ATS800R from Rohde & Schwarz, a vertical rack-mount CATR system *(see figure on page 24)*. The device can be placed on a table to characterize the phased arrays and facilitate quick mea-

Some Basic Functional Measurements

- *Effective isotropic radiated power:* What's the power level from my reference point at my measurement antenna?
- *Error vector magnitude (EVM):* Looks at modulation quality. How well we can transmit and receive a modulated 5G signal and then receive it?
- *Adjacent channel leakage ratio (ACLR):* How much energy is transferring from one channel to a channel next to it?
- *Spurious:* How much or how high are the levels coming from a particular device?
- *Antenna pattern*
- *Antenna gain*
- *Antenna directivity*

surements—such as EVM, ACLR, EIRP, or beamsteering measurement—within a 20-cm quiet zone. Although not automated, an adjacent device under test (DUT) can be quickly, fully characterized, which is beneficial for R&D functions as well as for a production environment.

OTA testing will be critical for ensuring that 5G devices will perform in the real world. The 5G device is placed in a test chamber and tested in simulated conditions to see how it responds. In addition to verifying the performance, such testing will certify that products meet specified standards, from both a modulated perspective and an antenna perspective. For example, a vector signal generator can provide a modulated 5G signal, while a spectrum analyzer can help analyze those signals with special measurement profiles, to verify unimpeded transmission between the signal source and the device.

White box testing was sufficient for sub-6-GHz devices for 2G, 3G, and 4G, as the far-field conditions were not as vast as we'll see with the FR2 band. The latter frequency range is designed for applications that would not allow for a direct line of sight from the DUT to measurement probe. White box testing enables the measurement antenna to look directly at the center of rotation, and the tester knows the antenna's precise location. Positioning is very important in white box testing, since it uses the far-field condition directly.

While a direct far-field system may allow for testing of larger devices, it's not practical at higher frequencies. FR2 will require black box testing, either because the antenna position is unknown or there's a need to measure an entire system or product. This testing method uses indirect far-field (CATR), which doesn't require exact positioning because an indirect wave measures the device. The measurement system is pointed at the reflector and filters out the spherical wave component to reflect the planar wave back toward the DUT.

Modern highly integrated chipsets, front-ends, and antenna systems require new techniques for OTA measurements in a multitude of development steps. Integrating the antenna or antenna array into the chipset poses challenges in beamforming verification and chipset or amplifier testing. With the use of state-of-the-art test and measurement equipment in lab environments, shielded boxes, or large anechoic chambers, the challenge of OTA testing largely comes down to understanding key challenges in antenna measurement and chamber setup.

It's important to keep in mind that conformance testing will change considerably for FR2 frequencies, leaning toward OTA testing over conducted

tests. Engineers will need to know how to calibrate the systems and understand what they are measuring. This will require a bit of homework on their part, perhaps working closely with antenna manufacturers.

5G PRODUCTION TESTING

In addition, 5G production testing, which is still evolving, should be recognized as much more of a challenge for R&D due to the high volumes involved. Production speed ultimately will be impacted by the types of measurements required as well as the length of time it takes to measure and record the measurement.

However, OTA systems for FR2 can be used without a front door in an R&D environment, which saves time placing and removing devices in the chamber and is further conducive to a production environment when a reflector is mounted above the manufacturing line.

5G production testing, which is still evolving, should be recognized as much more of a challenge for R&D due to the high volumes involved.

Such a setup enables a CATR system to easily measure bore sight, off-peak, null definition, beamsteering accuracy, and more.

Production testing for FR2, though, currently remains in the R&D stage for the most part, as manufacturers focus on developing the right design. When 5G becomes mainstream, test-and-measurement designs will need to quickly evolve and align with standards and specifications.

CLOSING THE KNOWLEDGE GAP: THE NEW DIGITAL DIVIDE

Millimeter-wave engineering is considerably different than RF engineering. At mmWave, components act differently. At low frequencies, engineers don't need to account for the phase properties of a wavelength since wavelengths are large in comparison to the component size. But at mmWave, in which frequencies are high and wavelengths are small, the wave properties of the signals must be considered.

As 5G begins to leverage high-frequency mmWave bands, engineers that used to work at RF frequencies will once again need to sharpen their skill sets and adopt new design and testing techniques. To learn more about the technologies shaping our future wireless world, please visit the 5G Learning Center at *www.mobilewirelesstesting.com* or the Rohde & Schwarz website at *www.rohde-schwarz.com/5G*.

TDK

Engineered to make life easier!

Tired of spending hours resolving radiated EMI issues? Use a power supply that actually meets Class B, even in a Class II construction (no earth ground).

CUS-M Series

High Performance Medical and Industrial AC/DC power supplies with low EMI

Features:

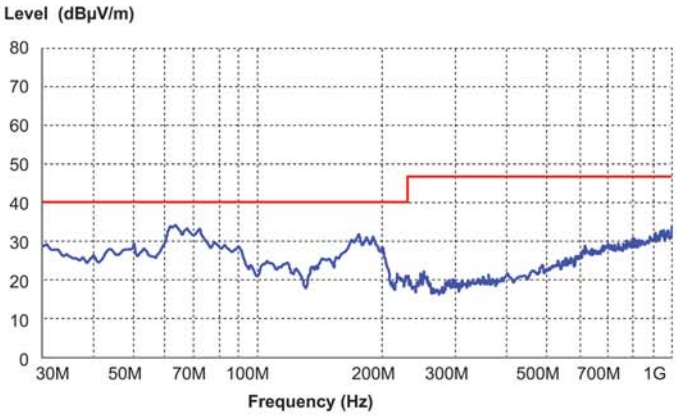
- Industrial and medical safety certifications
- Industry standard footprints
- Efficiencies up to 94%
- 4kVAC isolation and low leakage currents
- Suitable for B and BF rated equipment
- Class B conducted and radiated EMI

30-1500 Watt AC/DC Power Supplies

Model	Watts	Package	Size	Construction
CUS30M	30W	O, E, P*	2x3"	Class I / II
CUS60M	60W	O, E, P*	2x3"	Class I / II
CUS100ME	100W	O, E, B	2x4"	Class I / II
CUS150M	150W	O, E, B, F	2x4"	Class I / II
CUS200M	200W	O, E	3x5"	Class I
CUS1500M	1500W	E	5x2.5x10.3"	Class I

* E = Enclosed, O = open frame, P = pcb mount, F internal fan, B conduction cooling

EN55032 Class B, QP Limit Class II Construction



In stock for immediate purchase through our Distribution Partners:

ALLIED ELECTRONICS & AUTOMATION · arrow · AVNET · Digi-Key · MOUSER ELECTRONICS · Newark · SAGER POWER SYSTEMS

For more information on how TDK-Lambda can help you power your unique applications, please visit our website at **www.us.tdk-lambda.com** or call **1-800-LAMBDA-4**

Industrial Automation

BRYCE KEENEY | Product Manager, Teguar
www.teguar.com

Android Brings Advantages to OEM Industrial Rugged Computers

The performance and adaptability of industrial computers used on OEM equipment gets a boost from the Android OS previously used on smartphones.

Businessweek calls the smartphone revolution an Android revolution. So much so that the article recounts how the operating system was on Steve Jobs' hit list. Worried about the impact on the Apple iPhone, in 2010 he declared, "I'm going to destroy Android."

As we well know, Android is still standing, and now when many people think of this operating system, smartphones pop to mind. Though many regard smartphones as handheld computers, Android is a case where technology developed for smartphones has gone on to become operating systems for many forms of computers, including the industrial-panel PCs that OEMs install in their equipment. Considering Android's origins in low-cost cell phones, this operating system until recently was thought of as suitable for business or in the industrial space.

However, the coming years will lead to changing minds. A growing number of organizations need flexibility in their industrial operations. More and more plants and DCs are turning to panel PCs for monitoring and controlling production and material-handling operations that have Android operating systems installed on them. These computers will be finding their way into the equipment and systems being produced by manufacturers for their production and supply-chain users.

WORKING IN THE INDUSTRIAL ENVIRONMENT

Given the complexity of the industrial manufacturing process and the supply chain, the increasing compliance standards, the tight profit margins in this business, and the risk to a company's reputation, if a breakdown in safety occurs, this isn't a place for outmoded technology.

Many in management claim they like to be on the cutting edge. Yet, Android is an example of a technology that's not finding its way into the business space as fast as the cutting edge would imply.

Despite the promise of this and other technology, no matter how revolutionary the development that appears over the horizon is, these advances generally seep in rather than charge into general use on the processing plant or DC floor. Often, management views change as risky, waiting to find out if these advances work elsewhere before they bring new systems, equipment, or processes into their operations. For many managing these operations, their jobs depend on the type of approaches and equipment they bring onto the floor.

Anyone who owns a cell phone—even iOS users—know that the Android operating system has been around for more than a decade. The release of the Android 10 OS was just this past September. It's a long history for a software platform and shows consistency and continual improvement that benefits both the cell-phone market and the industrial business market.

YOU WILL BE USING ANDROID—EVENTUALLY

The time of Android as a business tool is coming. This system has been operating in a growing number of rugged mobile devices and touchscreen panels in processing, production, distribution, and other business essential applications, particularly for two reasons.

Android is a more robust and powerful operating system than legacy compact Windows technology. However, if management isn't sold yet on this operating system, then consider Microsoft is sunsetting mobile OS platforms such as Windows Embedded CE 6.0, Windows Embedded Handheld 6.5, and Windows Embedded Compact 7. The deathwatch of these long-used platforms is ticking onward.

A recent survey of logistics professionals found that 56% of respondents planned to increase Android use over the next three years. The holdouts are no doubt looking to wring a few extra years out of their legacy systems before device manufacturers turn off the support spigot.

So, with the door being shut on these legacy Windows OS platforms, operations will eventually come to find that they will be benefitting from Android's benefits. These include low cost, its built-in Bluetooth, near-field communications for transferring data between two Android devices, and an effective support for voice sensors.

ADVANTAGES OF THE ANDROID OPERATING SYSTEM

For production and material-handling operations with unique requirements, app development can be an enormous benefit. The evolution of Android computing is accompanied by an army of Android software developers that take advantage of a simplified software-development environment.

Android is a great system to develop your apps. For machinery or equipment that needs only a specific program to run on them, an Android panel PC would be ideal to operate those processes.

Customizable: According to Tom Warren, writing for *THE VERGE* on a closed-source OS like Windows, the code can only be modified by Microsoft, along with a few selected customers like big companies. With the code underlying open-source operating systems like Android, that code is freely available for anyone to view and modify under the terms of open-source licenses. This gives the end user more complete control over the software running on their equipment.

Android developers can easily customize applications to accommodate unique I/O devices and create GUIs to manage any industrial application. The mobility and device-agnostic aspects of this software is the basis for many new industrial-automation infrastructure implementations.

Energy efficient: As it's designed for battery-powered devices, Android uses fewer system resources than most other operating systems, so it requires a less powerful processor. This design saves energy and allows Android to run on smaller devices. If your application is simple, and you only needed to run one or a few programs. You would spend less, getting a panel PC with less computing power that gets the job done just the same.

Cost-effective: Android panel PCs are a lower-cost alternative to traditional Windows panel PCs for several reasons. Android computers typically include Arm processors, which are more cost-effective than Intel processors. Android industrial computers aren't subject to the operating-system license fees required with Windows-based computers.

Android industrial computers don't require additional storage capacity and memory requirements that are necessary for Windows-based computers. The significant cost savings associated with these differences is driving the industrial factory automation industry toward Android-based computing.

Since units operating on Android are also generally smaller with less-demanding computing requirements,

Design Note

Analog Devices' Latest PLL/VCO Device, the ADF4371, Delivers Groundbreaking Low Phase Noise and Spur Performance

Ian Collins, Applications Engineer

Introduction

Increasing demands for frequency bandwidth, throughput, and dynamic range in communications systems, together with the need for higher antenna frequencies used in millimeter wave 5G, have all placed further demands on the quality of the local oscillator (LO) or clock used in communications or mixed-signal systems, respectively.

The newly released integrated phase-locked loop (PLL) and voltage controlled oscillator (VCO), the ADF4371, and the similar ADF4372, showcase Analog Devices' efforts to address the needs of these demanding applications.

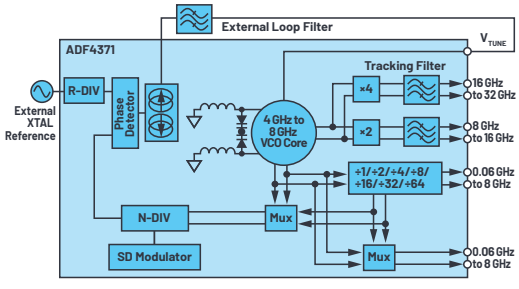


Figure 1. ADF4371 block diagram.

Frequency Coverage

To maximize frequency coverage, the ADF4371/ADF4372 VCO covers an octave range from 4 GHz to 8 GHz, and, by using frequency dividers at the output, dividing by 1/2/4/8/16/32/64 allows full frequency coverage at the main RF8 output of between 62.5 MHz to 8000 MHz. A second identical output is provided to allow a user to drive a converter clock. The open-loop VCO phase noise is –109 dBc/Hz at 100 kHz offset for the 8 GHz output frequency.

Until recently, generation of high frequencies required the use of external frequency multipliers, typically fabricated on GaAs processes, and they often demanded additional filtering, along with amplification, to overcome the effect of the filtering.

To achieve higher frequencies, the ADF4371/ADF4372 contains an integrated frequency doubler, which provides 8 GHz to 16 GHz output at the differential RF16 pins. The ADF4371 also features a frequency quadrupler that generates from 16 GHz to 32 GHz at the RF32 differential output. To minimize the generation of unwanted multiplier products, the ADF4371/ADF4372 contain tracking filters that optimize the power level of the desired frequency while suppressing the unwanted multiplier products. On the doubled output, the VCO feedthrough is –45 dBc. On the quad output, the suppression is approximately –35 dBc.

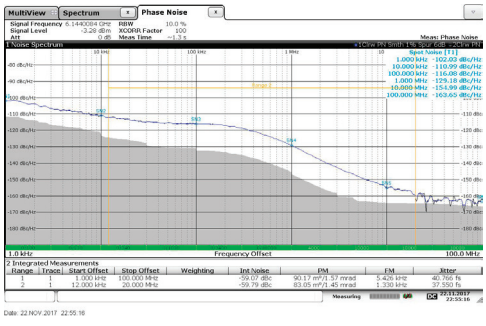


Figure 2. RMS jitter at 6.144 GHz.

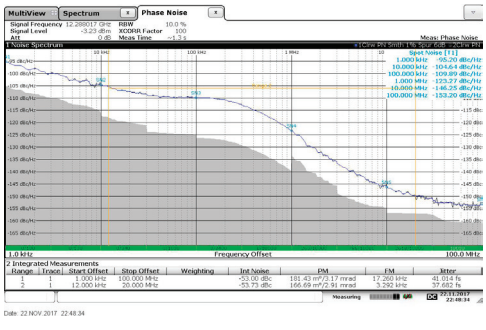


Figure 3. RMS jitter at 12.288 GHz.

Leading PLL Performance for Converter Clocks

Improvements to the PLL circuitry mean the ADF4371/ADF4372 devices have a PLL figure of merit (FOM) as low as −234 dBc/Hz that, when together with a correspondingly low 1/f noise of −127 dBc/Hz (normalized to 1 GHz output frequency at 10 kHz offset), allows users to generate clocks with an rms jitter number as low as 40 fs (1 kHz to 100 MHz integration limit), making them very suitable for use in the most demanding converter clock applications. A simple low-pass filter with small resistors is recommended in order to minimize the resistor noise, which may appear in the loop. A high frequency (250 MHz or 125 MHz using the reference frequency doubler) ultralow noise reference source is essential for achieving such low noise. The phase frequency detector (PFD) can operate up to a maximum of 250 MHz in integer-N applications. The doubled VCO differential output at RF16 can be used to interface directly to some ADI converters without the need for external balun circuitry, which would increase cost and performance. No degradation in performance from 6.144 GHz to 12.288 GHz is expected.

Communications and Instrumentation LOs

For wireless and instrumentation applications, the ADF4371/ADF4372 contain 39-bit resolution sigma-delta modulators, which enable frequency generation with submillihertz resolution with 0 Hz error. In this case, the ADF4371 PFD operates up to a maximum of 160 MHz PFD frequency. In these applications, the ADF4371/ADF4372 deliver <48 fs rms jitter. The ADF4371 also has industry-leading PLL spurious performance, with PFD spurious as low as −100 dBc and in-band, unfiltered integer boundary spurious as low as −55 dBc. This performance level greatly simplifies frequency planning and reduces time to market. Many fractional-N PLL and VCO devices have unpredictable fractional-N spurious mechanisms, which can lead to additional unplanned characterization and frequency planning, which add complexity and cost.

Small Size

The ADF4371/ADF4372 PLL/VCO devices are available in 7 mm × 7 mm, 48-lead land grid array (LGA) packages. Minimal additional decoupling is required, meaning exceptional performance exists in a small footprint solution.

To achieve the best performance, the use of high quality low dropout (LDO) regulators such as the ADM7150 or LT3045 are recommended. The VCO can be supplied with either 3.3 V or 5 V, and the remaining circuitry is powered from a 3.3 V rail. The ADF4371 can be simulated in ADIsimPLL™ to assist the user in designing the appropriate external component circuitry required to implement a full PLL system.

Conclusion

Industry-leading frequency coverage, performance, and small form factor combine on the ADF4371 to address the high demands of new communication and instrumentation systems.

About the Author

Ian Collins graduated from University College Cork with a degree in electrical and electronic engineering, and he has worked in the RF and Microwave Group at Analog Devices since 2000. He is currently an applications manager in the Microwave Frequency Generation Group, which focuses mainly on phase-locked loop (PLL) and voltage controlled oscillator (VCO) products. When not spending time at work or with his young family, Ian enjoys photography and the theater (both on- and off-stage), reading, and listening to music. He can be reached at *ian.collins@analog.com*.

operations can reduce costs by deploying just the right-sized device needed. Android Panel PC sizes can range from 7 in. up to 21.5 in., while industrial tablet PCs and handheld devices can range in screen sizes from 4 to 12 in.

CAN THE PC PANEL LIVE UP TO INDUSTRIAL ENVIRONMENT DEMANDS?

Just like all of the other equipment and systems that make an industrial environment run, here are some critical musts when selecting a PC Panel to run Android for these operations:

Touchscreen Design

The guts of any PC that manufacturers work with, of course, is essential to performance. But to the operation of these systems, the screen is all that matters.

Seeing what's on that screen is essential. The information must be viewed off the screen quickly, with any miscomprehension leading to severe consequences. Consider units that offer high resolution, readable under bright lights (sunlight or artificial), TFT screens that are CCFL backlit with high brightness, and a wide viewing angle.

Many software applications are designed to run at specific LCD resolutions and aspect ratios. A web-based application may require a panel PC screen resolution of 1920 × 1080 pixels, as the layout of the on-screen menus and tabs are scrunched when displayed on LCDs with lower resolutions.

Bigger isn't always better. Often there are physical space constraints in production areas. In this case, the application would require a 15.6-, 18.5-, or 21.5-in. LCD size, as these are the only common LCD sizes that support 1920 × 1080 pixels. It's crucial to consider the software application requirements when selecting the LCD size of your panel PC.

In some environments, these screens must be sensitive to gloved hands. Computer panels with resistive touchscreens are still the most prevalent in low-temp computer applications, as they're pressure-sensitive, which means they can be used with gloves, as well as bare hands, or a stylus.

Water Resistant

Given that panel PCs house sensitive electronics, they must withstand blasts of water and chemicals from clean-in-place operations in food, pharmaceuticals, and similar processes. PCs with stainless-steel front bezels protect the components and will not degrade or rust from exposure to liquids. An IP66-rated front panel will stand up to the water jets.

Keeping it Clean

Not only do panel PCs need to stand up to the equipment in the area being cleaned, they also must be suitable for frequent, easy cleaning to avoid contamination. Fanless, rugged panel PCs meet that requirement with a minimum of openings where germs can hide, as well as housings that can withstand industrial cleaning agents.

Taking Computing Through to the Next Generation

In the tech world, change is a fact of life. Inevitably, Android will someday be replaced by another form of technology. Though it's one thing to take advantage of the latest advance, it's another for the device the system runs on to make it all the way through the current technology cycle.

TIME FOR A CHANGE

Industrial-grade computers are the tools for manufacturing and moving product. These are the human-machine interfaces into whatever system controls the operation. As production and material handling head into the future, the industrial-rugged PCs specified to monitor and control these functions are gradually and surely migrating to the Android operating system.

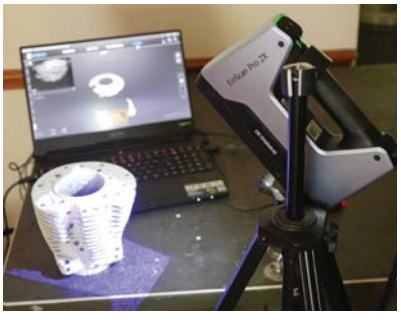
Along with the advantages the Android OS gives your equipment users, you have the ability to provide practically whatever they want. Some of the rugged industrial computers are geared toward the needs of industries such as food and beverage, and industrial automation. Your rugged-industrial PC source has an array of options to find the form and function of your equipment and systems. ed

PAUL RAKO | Contributing Editor

What's All This 3D Scanning Stuff, Anyhow?

3D scanning does work, but you have to be patient and know the tricks to get any kind of decent results.

I recently bought a 3D scanner to complement the 3D printer I purchased to work on a motorcycle engine case redesign. The Einscan Pro 2X model I bought needs a Nvidia GTX video card to support its software *(Fig. 1)*. The Nvidia Quaddro cards I have in my CAD workstations will not work. I wanted a dedicated laptop for this, so I purchased a GigaByte gamer-class laptop with a Nvidia GTX 1070 card—only one revision down from the fastest GTX 1080 video card, but about 1,000 dollars cheaper.



1. The Einscan Pro 2X scanner costs about $5,000, and can achieve point resolutions of 0.15 mm, whereas the Pro 2x Plus model is faster with a larger field of view, but a 0.24-mm point distance. I opted for the finer detail of the cheaper 2X model.

The Newegg description was "GIGABYTE Aero 15X v8-BK4K4P Thin Bezel 15.6" 4K/UHD GTX 1070 8 GB VRAM i7-8750H 16 GB Memory 512 GB PCIe SSD Windows 10 Pro 64-bit Gaming Laptop." It cost $1,949.00 in late 2018. I got the scanner on eBay for $5,224.00 from a vendor that I now see only sells junk jewelry. I should have seen if my SolidWorks vendor offered this model and tried to get them to match the eBay price. You want support with things this difficult to use.

GOOD SCANNERS COST A LOT

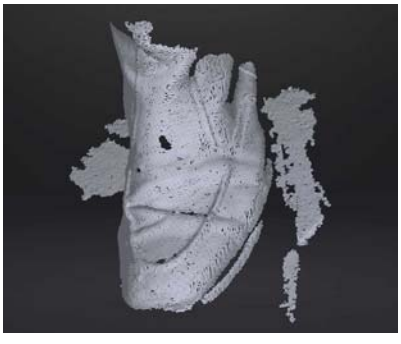
While $7,000 is a lot of money, there are far more expensive scanners that are even more accurate. The Einscan uses patterns of light to detect the objects, similar to the Artec scanners. The Artec scanners achieve 0.1-mm accuracy without the need for calibration plates or stickers, but they cost between $10k and $25k—too far out of my price range. At the other end of the spectrum are 100-dollar scanners that use the Microsoft Kinect game module. Newark/Element 14's Ben Heck did a nice video trying to get a decent output out of one of those models *(go to https://youtu.be/TLFonu92Wn4)*.

I wanted a handheld scanner that could do something as large as a motorcycle exhaust system, so the many cheap 3D scanners that used a turntable for small items were not of interest to me. The Einscan product line occupied a niche with good resolution and a price, while high, was not heart-stopping.

GETTING STARTED SCANNING

I unpacked my new laptop and suffered with the despicable Windows 10. I have Win7 on all my machines, but I have to admit Win10 was not as infuriating as I feared. I got the Einscan software installed and plugged the scanner into a USB port. Lots of misery since the scanner software would only recognize the scanner if it first had power and I did a reboot on the laptop.

I tried and tried to scan a Harley cylinder I had handy. Of course, I did not read any instructions—that's the American Way. I was very disappointed and, not being able to get anything, I just happened to wave the scanner past my shoe *(Fig. 2)*. Miraculously, a pretty decent point cloud appeared.



2. I could not get any points scanned of a black Harley cylinder. Then I accidentally moved the scanner past my shoe that was speckled with white paint.

This taught me two things. First, the scanner likes nice "monotonic" shapes like shoes and wizard figurines. These were work shoes that I had done painting and plastering in. The shoes were specked white and that gave the scanner enough reflections to sense the shape. That was the second lesson, my trying to scan a black cast-iron cylinder was a fool's errand.

I then dragged over a DeWalt drill and scanned that in *(Fig. 3)*. To do a fast scan without using little circular paper markers plastered on the part, you must

3. Realizing the scanner needed lighter-colored objects, I scanned in a hand drill. Putting the part on black felt would have eliminated the spurious parts of the scan. Inexperience accounts for the noise and shabbiness of the results.

go very slowly and methodically. The software is actually recognizing the shape of the item and using it to register the relationship between the new points it's adding.

THE SCANNING SOFTWARE

The Einscan software has one of those whizzy dumbed-down interfaces that remind me of a Speak-and-Spell toy *(Fig. 4)*. It's as silly as that "ribbon" menu Microsoft added to Word. You can't switch between high-definition and rapid-scan modes easily; you have to start the project by deciding which you want.

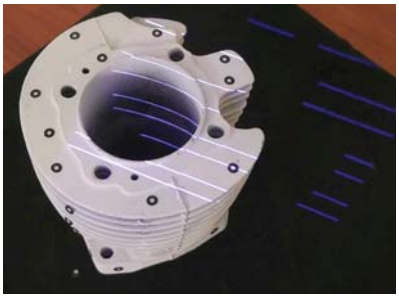
4. The Einscan software does not follow Windows standards for menu items. To do a calibration with the included image plate, you have to know to click your mouse on the tiny little dot under "Aug.13 – 11:24." The question mark is a goofy contextual help that tries to make up for the barren oversimplified interface. I now know pressing shift and the left mouse lets me erase points. That there is an edit mode is completely hidden from the user.

Once in the rapid-scan mode, the interface lets you choose whether to use markers or part features or both. It also lets you select a texture scan, in case you have the optional camera attachment, which paints a bitmap on the point cloud. You can either select the resolution or choose a mode that lets you specify the point-cloud resolution after the scan. "Classic" mode is 15fps (frames per second), and "Lightning" operation mode is 30fps. You can tell programmers are involved in the UI since they create artificial variable names that only they know the meaning of, rather than labeling the buttons "15fps" and "30fps."

The rapid-scan mode paints what looks like 2D bar-codes over the object *(Fig. 5)*. If the object has many distinguishing features, the scanner can keep its place. On areas with uniform appearance, like the right side of that cylinder, the software gets "lost." Your only hope is to mess with distance to try and get some unique features in.

5. The rapid scan mode offers less resolution, but it doesn't require you put little circular markers on the part. It projects what looks like a 2D bar-code on the part. You can judge the proper distance, since the pattern goes into focus when you're at the optimum distance from the part.

Though the HD mode provides greater resolution, it requires the use of markers so that the software can maintain registration. Instead of the bar-code like pattern, it uses several rows of light beams *(Fig. 6)*. It's intoxicating to watch the laptop screen and see the point cloud appear as you sweep the lights over the part.



6. HD mode has up to 0.2-mm resolution, but you have to plaster little paper markers all over the part. They must be close enough so that the software can recognize several of them in the same field of view, even as you twist down to do the sides. This is very difficult to do.

This is where I learned you can add markers in the middle of a scan to help the software maintain registration. It takes a lot of markers. You must have enough of them to "turn the corners," so to speak. Getting a scan of the top of the cylinder was easy. Where the software kept getting lost was when I would angle the scanner more level to scan the sides of the cylinder.

AN EXASPERATING SCANNING EXPERIENCE

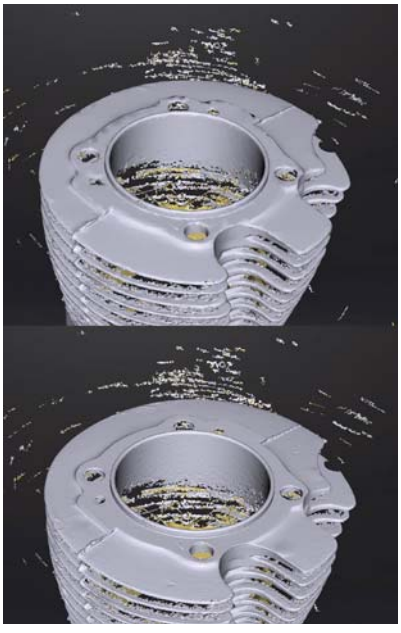
The Einscan user interface can be exasperating *(Fig. 7)*. That little green bar on the left edge is telling you the scanner is at a good distance. If it runs up, it turns blue and you get a warning "Too far." If the bars bounce down to the bottom, they turn red and you get a warning "Too close." The nice thing is that you can see the light pattern on the part go into focus when you're at the right distance, so that helps you keep the scan going.

When doing a rapid scan using the part's features, this screen shows the most infuriating aspect of the software. That little green splotch of point cloud underneath the "Preview" light box is what the scanner is detecting. That's what it has to pattern-match to when you start scanning.

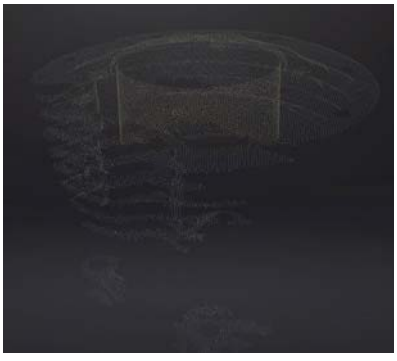
Often, when you hit the "Play" triangle, either on this screen or on the scanner button itself, the scanner just

7. When the Einscan scanning software actually gets to the point of taking data, this is what you see. The green bar on the far left shows the distance you have to the part is OK. The picture at upper left shows what the scanner is seeing, and you set the brightness so that there's no red warning areas washing out the part. The green points speckled under the "Preview" box is the point cloud. If you press the scan button again, it leaves preview mode, and promptly gets lost since there aren't enough features captured at the instant of the first frame. Then you can't get into preview mode again, unless you reopen the project.

8. This is the dreaded "Track Lost" warning. The scanner has successfully used feature detection to maintain registration as I waved it over the top of the cylinder. However, it loses its place when I move down to do the sides, and it can't be recovered. It was better to flip the cylinder upside down and start from that position. Impressively, you can flip the part back over, and if you have enough features taken already, the software will re-find its place in space. Then you can add to the scan. All of this tedium significantly increases the noise and errant points recorded.



9. You can speed things up and have smaller files with a low-resolution mesh (top). A high-resolution mesh does keep the head gasket area flat (bottom). The "Lightning + Refine" mode lets you save the point cloud with different resolution, as opposed to making the mesh with different simplifications.



10. If you specify a 2-mm point-cloud resolution, the scan becomes very sparse. What I worry about is that there are also no detailed features for the software to register to. Thus, the program might "Lose Track," which is a constant bedevilment.

seems to quit, and you get a "No Data" warning. Then the only way to get back into preview mode is to use the folder icon to reopen the whole project. You learn to get a really thick and rich point cloud in the preview mode, so that the software has something to "latch onto" as far as feature recognition.

Once you have that initial point cloud, the software is pretty amazing in that it can find that shape again if you go back to the exact same place. This is why it's criminal that Einscan did not put a 1/4-20" tripod mount nut on the bottom, so you can fix it in place. It might be snazzy and glamorous to have a handheld scanner, but not after 10 or 20 times restarting the same project because you can't get the scanner past its preview mode. If the scanner was fixed, it would let you move the part incrementally to add to the scan.

LOSING TRACK

When you move too fast or the part has no unique features, you get the dreaded "Track Lost" warning *(Fig. 8)*. Worse yet, the software plays this obnoxious calliope music that makes you want to toss the scanner at the laptop screen. I would rather have a white-noise hiss or a tone or anything else, but I guess I will have to figure out the volume setting on Win10, right after I find where they hid the Control Panel. I still can't stop the keyboard backlight from shifting colors like some infantile toy.

It's often the "big concept" usage that goes unspoken in documentation. All I needed was one paragraph that explained the Einscan software has a project file folder you create, and it stuffs a project file into it like .hd_prj or .rap_prj. When you hit the "stop" square icon, it generates a point cloud, letting you select the resolution; that is, the density of the point cloud if you're using that Lightning + Refine mode *(Fig. 9)*. Then you can save the scan as a point cloud; they have tiny little text at the bottom of the dialog box that gives the option of .asc or .p3 file. The program won't let me select the .p3 format for some unknown secret reason.

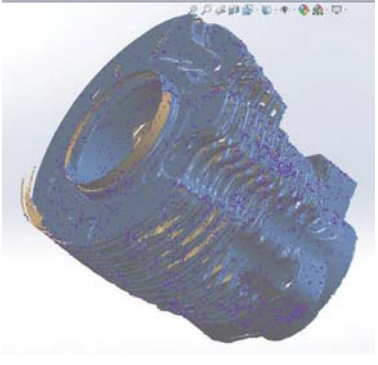
You can also decide to just take a low-resolution point cloud to begin with *(Fig. 10)*. What confused me was that bright green arrow in the "Mesh Model" box on the lower right of the screen. See, the programmers at Einscan have decided to have fun calculating a mesh of triangles from the point cloud, when they should just let me do that in SolidWorks. Once you do hit that "Mesh Model" check button, you get more choices.

Then, when you hit the little floppy-disk save icon, it lets you save as an .stl file, along with a handful of other formats. It shows an .asc format—I'm not sure if that would overwrite a point-cloud .asc file with a mesh .asc file. I don't have the patience to find out.

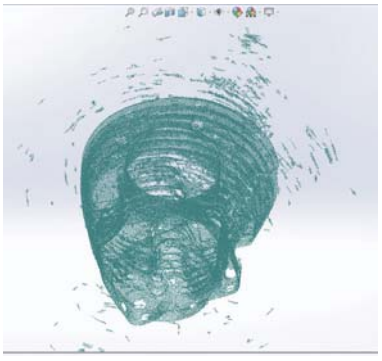
GETTING INTO SOLIDWORKS

Getting an .stl mesh to get into SolidWorks had its own set of headaches. Because my file was so big, SolidWorks can only open the file as a graphical object *(Fig. 11)*—pretty much useless for doing anything with. Knowing just how amazing SolidWorks is, I knew to look on the web, and learned about the ScanTo3D add-in that's included in my Pro version, but not turned on by default. Using the tools>Add-ins menu pick, I turned it on, and activated it at startup.

With ScanTo3D working, I then realized I wanted the .asc point cloud from the Einscan software *(Fig. 12)*. After a while to figure out where the "next" arrow is in a SolidWorks wizard, I went through the steps, although SolidWorks seemed to crash a few times. That lost about a half-hour of "wizard time" for each occurrence, so I started bailing from the wizard with the green check icon, doing a save, and then just running past the beginning wizard stuff again. This was critical after SolidWorks lets you erase all of the noise and errant points.



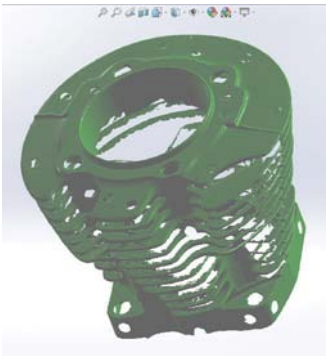
11. You can load the mesh created by Einscan into SolidWorks, but it's so large it only comes in as a useless graphical object. I did not know how to enter edit mode in Einscan, so there are a lot of trash meshes in this.



12. I think it's better to let SolidWorks do the meshing. Here's what the .asc point-cloud files look like when first imported to SolidWorks with the ScanTo3D plugin. A wizard helps you scrub out the extraneous points, simplify, and regularize the cloud before SolidWorks does a very sophisticated meshing operation.

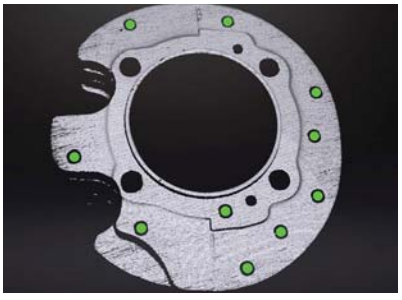
A save then protects all of the time and tedium of lassoing those little dogies.

After SolidWorks goes through the wizard, you can then make surfaces from the points with yet another wizard. Instead, I stopped with the mesh, and that resulted in a workable structure *(Fig. 13)*. You might laugh at calling that hole-ridden monstrosity workable. The thing is, you're just using this as a guide to build real solid features, like 3D tracing. This mesh is pretty lame, but I can see getting



13. After running through the ScanTo3D wizard, SolidWorks can come up with a haggard but usable mesh representation of the motorcycle cylinder.

14. Once you have the mesh built or imported into SolidWorks, you use the section command to slice it in places. You put in a plane and sketch at these places, where you can then trace over the features in your mesh to make a real 3D solid model. You can subsequently suppress the mesh in the SolidWorks feature tree.



15. Despite a rough start, there's amazing potential for 3D scanning. This scan of the simpler top area of the motorcycle cylinder shows parting lines and even the texture of the cast iron in places. With some more research and improved technique, I can reverse-engineer some parts that are very difficult to measure with hand tools.

better with the scanner and taking a full morning to do a scan that will result in a more monotonic closed form.

One thing I want to try is hanging the part from a ceiling fan with thin wires. Then mount the scanner on a tripod, and once it gets a decent preview, kick in the scan while slowly turning the part. Since the part is just hanging from three wires, once I get most of the features into the software, I can use the scanner in handheld mode to get the underside of parts and inside the cylinder. I also will experiment with tempura paints to get better reflection of the part, though the white ceiling paint I used this time worked pretty well.

USING THE SECTION COMMAND

Once you get even a lame mesh in Solidworks, you can then use the Section command to slice it at various points *(Fig. 14)*. Then you create a plane at the section line and do a sketch where you trace around the feature—in my case, the cylinder fin. Then you can extrude that as a fin or use normal modeling techniques. I do wonder about my alternative method, which is to buy a cylinder on eBay and saw it up at every fin. Then just lay the fins on a 100-dollar flatbed scanner and use the scans to trace over and make .dxf geometry in TurboCAD. I can then import the .dxf files into SolidWorks as a set of sketches. It sure would be cheaper than the 7,000 dollars for this setup.

The thing is, new stuff is always frustrating. I remember spending hours trying to optimize JPEG settings in GIMP. Then I got into video and it was days of playing with codecs and frame rates and all those things. I am sure this 3D scanning has a steep learning curve, but it's like a video game that frustrates you continuously, only to occasionally give you some fantastic achievement that keeps you stuffing quarters into the game.

This cylinder digitization is a nice start, but my real need was scanning Harley cylinder heads so that I can figure out the arbitrary angle that the exhaust and intake ports come out at. I can see it will take a while, but the path to that goal is doable, and it might even be fun, especially if I bother to read the instructions.

I do realize that scanning a deeply finned motorcycle cylinder was a very difficult objective. Yet when I scan the simpler parts of the cylinder *(Fig. 15)*, I can see the awesome results that I might expect with more patience and understanding. While I prefer the SolidWorks meshing algorithm, I note that the Einscan meshing code will automatically remove those green markers and pave them over with mesh. Those little joys will keep me playing with this technology for a long time.

ElectronicDesign MachineDesign

are proud sponsors of

2020 DragonSpeed IndyCar Race Team!

22 MAY
Behind the Scenes Tour

ENTER OUR DRAWING
for tickets to the Indy 500*

*Offer valid for residents of the United States.

24 MAY
Indy 500 Race Day

What you'll win:

- 2 Indy 500 Tickets
- 3 Nights Lodging
- 2 Three-Day Passes
- $1,000 travel voucher

Join Electronic Design Sr. Content Director, Bill Wong for a meet and greet to start your behind-the-scenes tour at Indy on May 22.

Visit this webpage to enter the drawing and join us in the pits:

http://bit.ly/2vXegDA

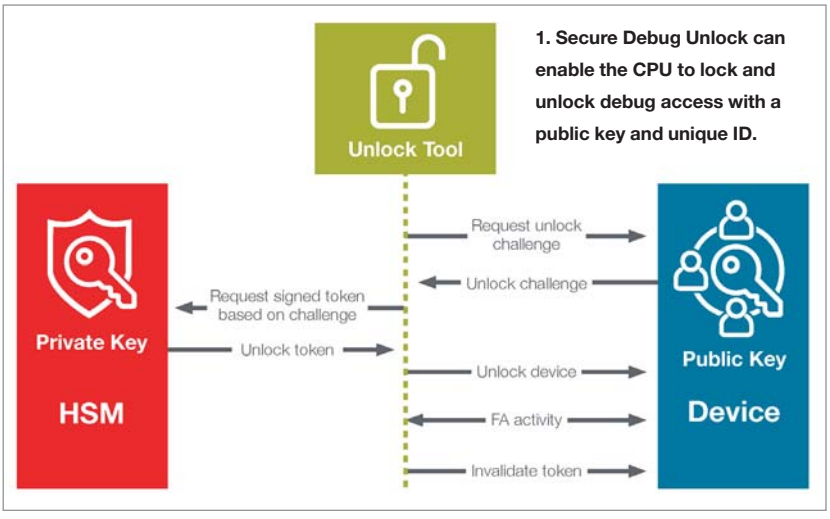
JOSH NOREM | Staff Systems Engineer, IoT Products, Silicon Labs
www.silabs.com

Securing Low-Cost Embedded IoT Devices

Many simple techniques and best practices can be applied to low-cost or older embedded devices to make end products and systems more secure, often without requiring advanced or costly security hardware.

Securing embedded devices for both IoT and non-IoT applications is an increasingly important concern. In response, device manufactures are building embedded processors and microcontrollers with a wealth of integrated security features.

Unfortunately, security comes at a cost both in the price of these new devices and in the effort required to redesign and requalify hardware to make use of new ICs. The good news is that much can be done on low-cost or older hardware to enhance system security. Let's explore the techniques we can use to improve security on these systems.



1. Secure Debug Unlock can enable the CPU to lock and unlock debug access with a public key and unique ID.

GENERAL STEPS TO IMPROVING SECURITY

Many simple techniques are available to greatly improve the security of low-cost systems. The first and easiest is to ensure that hardware and firmware details aren't readily available. It's common to see exploits presented in academic papers that begin with finding the firmware or schematics for a product on a website or FTP server. Schematics, source code, and binaries should all be access-controlled. In addition, firmware updates should be encrypted to prevent code from being easily extracted from an update. It's also possible to take more extreme measures such as requesting custom-marked ICs to obfuscate what hardware is being used.

Limiting information in this way is simply obfuscation and not true security. It does increase the difficulty of an attack and improves the system's security. However, limiting access to key information makes it less likely for researchers and other "white-hat" actors to analyze your product and alert you to security vulnerabilities.

One way to mitigate this is to engage with third parties who conduct code reviews and penetration testing, and provide this information under NDA. In addition, consider making information available under NDA for researchers or academic groups who wish to analyze and attack your product.

Another common technique is to secure the debug interface to the product. While low-cost ICs may not have hardware features such as Secure Debug Unlock *(Fig. 1)*, they often enable the CPU to lock and unlock debug access. This capability can be used to implement a reasonable debug unlock feature.

In the simplest form of Secure Debug Unlock, each device would be programmed with the same public key (unlock_keypub) and a unique ID. To unlock a device, an unlock token must be generated by signing its uniqueID with the private unlock key (unlock_keypriv), and that certificate is sent to the device. The device firmware checks the signature to ensure that the uniqueID is being sent by the private key holder, and then checks that the uniqueID matches the uniqueID of that part. If both conditions are met, the debug interface can be unlocked.

While there are more advanced implementations, this simple example represents a significant increase in security over not locking the interface, using a simple password-based unlock, or using a device that allows anyone to unlock the device after erasing flash.

In addition, some devices offer the ability to permanently lock the debug interface. In systems where debug access isn't needed for service or failure analysis, this is often a very secure way of disabling the debug port.

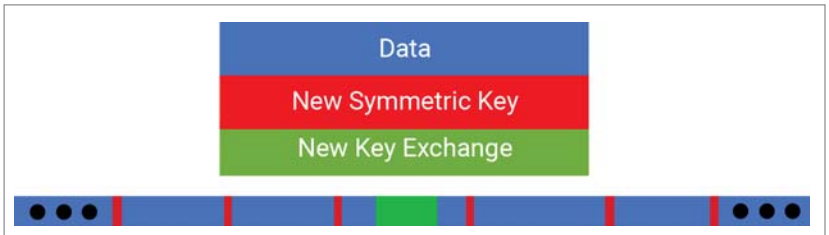
In certain systems, debug terminals are left running. Like debug ports, these should be disabled to prevent attackers from using them to compromise the system. The same techniques used in locking debug access can be applied for this purpose.

PROTECTING SYMMETRIC KEYS FROM SIDE-CHANNEL ATTACKS

The consideration is how to better protect symmetric keys from extraction via side-channel attacks. Although these attacks take many forms, they all rely on observing some aspect of the device (power consumption, EMI, timing) while decryption is being performed. Then those observations are used to extract the key via statistical analysis.

Newer embedded devices can contain hardware countermeasures to prevent a usable signal from being observed. While it's tempting to try to reproduce the effect of these countermeasures in software, the reality is that the statistical analysis performed by attackers is often too effective for software to be able to sufficiently obfuscate the signal.

In the absence of dedicated hardware countermeasures, the goal is to thwart side-channel attacks by preventing use of the key. If the key isn't used, the attacker can't capture the number of observations (traces) required to perform their analysis. In some cases, this isn't possible, and we must live with the fact that the key is vulnerable to this style of attack. In other cases, we can effectively limit the use of the key and prevent the attacker from collecting the needed data.



2. A communications protocol can frequently rotate the symmetric key by having the transmitter generate a new key at fixed intervals.

A prime example of this is a secure bootloader, which will decrypt an image as part of the installation process. The key should only be accessed when installing a valid update. In most systems, updates are performed a few dozen times a year. This means we can limit key use to a few times a year, which is far less than an attacker requires to mount a successful attack. First, the bootloader must prevent the loading of the same image over and over. This can be easily accomplished by including a version number in the image and only decrypting it if the version is newer than the currently installed image. This prevents attackers from provoking use of the key by repeatedly loading the same image or by flipping between versions (i.e., N, N-1, N, N-1….).

Next, we need to prevent the bootloader from loading a corrupt or altered image. This is easily accomplished by signing the image and any metadata such as the version. It prevents the attacker from simply editing the version of the image and fooling the bootloader into thinking it's getting hundreds of valid updates.

Initially, the bootloader must prevent the repeated loading of the same image. The final attack vector is a bit more difficult to handle. With these mechanisms above, the bootloader will only decrypt the image if it's newer than the current version and correctly signed. However, attackers can still load the image many times if they abort the installation process by either removing power or pulling the reset pin.

When a reset occurs during image installation, most bootloaders will identify that a bootload was in progress and automatically retry the installation. This ensures that a failed installation doesn't "brick" the part. It also causes decryption of the image to occur with every retry.

By continually resetting the part midway in the installation, an attacker can cause the bootloader to repeatedly decrypt the image. To avoid this scenario, the bootloader may include a "failed update" counter that's incremented every time the installation begins. This counter will increase with each retry until it reaches a value (such as 20 retries) indicating that an attack is probable, and then it will brick the device (erase the decryption key and erase the application) to prevent any further attacks on the key. In most systems, this set of bootloader rules will greatly increase the difficulty of executing a side-channel attack.

Another example is a symmetric key used to protect communication with an external entity. Normally, a public key exchange is used to exchange symmetric keys that encrypt the actual communication between the two parties. The symmetric keys are often utilized for an extended period, and their use can be observed long enough to allow side-channel extraction. Rotation of the symmetric key at regular intervals reduces the number of times it's used and the probability that it can be extracted through a side-channel attack.

Since generating and distributing a new key consumes power and time, there's a cost to this method. However, when the traffic being protected is sensitive, the increase in security is generally worth the performance cost.

Figure 2 shows a communications protocol that frequently rotates the symmetric key (perhaps every 1 ms) by having the transmitter generate a new random symmetric key at fixed intervals and sending it as part of the encrypted data. Once sent, the link switches to the new key. If attackers gain access to any symmetric key, they will be able to read all future messages since they can decode the next key with the broken one. To mitigate this vulnerability, an asymmetric key exchange can be performed at longer intervals (for example, 1 second) to ensure that if a key is broken, the attacker can decode only a second of data.

INCREASING THE DIFFICULTY OF EXPLOITING THE APPLICATION

Exploits of the application code are a concern for low-cost devices that don't have hardware support for isolating software to contain exploits. Code exploits come in two varieties. The first style of attack attempts to manipulate memory by using an existing interface whose inputs aren't correctly validated in an unexpected way.

The most well-known attack of this type is a buffer overflow *(Fig. 3)*, where an attacker sends more data than expected to an interface, which results in overwriting application code or data with their own values. This attack can be used to activate application code in an unintended way, such as getting a command intended to retrieve a variable to instead retrieve configuration information. In the worst case, code can be injected into memory, causing the CPU to jump to that code and allowing the attacker to execute arbitrary code and take control of the hardware.

The second style of attack attempts to cause code to malfunction by indirectly injecting a fault such as a power supply or clock glitch. The goal of this attack is to modify program flow. This is typically accomplished by injecting a fault or disturbance at the precise moment in time that the CPU is making a security decision (or responding to one) by corrupting a branch instruction or the instructions leading up to a branch instruction. For example, there may be test code that sends secret information from a UART for test purposes that's not supposed to run in normal operation. However, an attacker may be able to get that code to run by injecting a fault that causes an if-statement to make the wrong decision.

The best way to mitigate this type of attack is to reduce the attack surface, e.g., pare down the number of interfaces that can be attacked and reduce the complexity of the overall application. For example, it's tempting to create a single image containing the application and test code needed for board test. This image can be programmed, a board test can be run, and the product shipped. However, it results in the presence of test code and a board test interface on devices in the field.

Instead, we can create separate test and application images. Program the board test image, run the tests, and apply the application image. This will simplify the end application and eliminate a possible point of attack. Similarly, production images should not contain unneeded code. For example, if a feature is partially developed and then eliminated for cost or time-to-market reasons, that code should be removed from the production image rather than left in place. Similarly, rather than include optional features that can be enabled in the field, create two separate images and enable the optional features via a firmware update.

Firmware updates are also a good way to mitigate code exploits. Regular firmware updates have two beneficial effects. First, if an attacker develops a permanent exploit that installs itself into flash, pushing an update will either overwrite that exploit with a new image or force the attacker to not install the new image, which will likely be detected when the expected changes in behavior aren't observed.

Firmware updates can also include intentional (or unintentional) changes in code, which disrupt the function of

3. A buffer overflow occurs when an attacker sends more data than expected to an interface.

exploits of the previous version. For example, if the structure of data in RAM or code in flash is changed, then an exploit relying on overwriting a specific address will not function since that address no longer contains the variable that the exploit was trying to overwrite. Similarly, if a communication protocol is changed on an update and a compromised device chooses to suppress (not install) the update, it will be immediately obvious that the device did not update properly.

Hardware functionality can also be reduced to limit the tools available to an attacker. For example, many ICs allow RAM to be disabled to save power. It's possible for an application running on a device with 4K of RAM but using only 2K to disable the other 2K and prevent an attacker from using it. Similarly, unused pages of flash can be locked to prevent attackers from installing an exploit into them.

```
// Do work to see if
// we should unlock

//Unlock
If(status != LOCKED)
{
  DBG->LOCK.UNLOCK = 1;
}

// Continue with
execution

return;
```

```
//If(status != LOCKED)
LDR SP+0x1234, r0
LDR #0xACED  , r1
CMP r0, r1
BE  over

//DEBUG->LOCK.UNLOCK = 1;
MOV #0x1     , r0
STR r0       , #ADDR

Over:
```

4. A glitching attack opens a system vulnerability by forcing a hardware fault.

Other easy options available on some hardware include implementing stack limit/overflow detection (or stack canary), setting the "no execute" bit on data memory (if available) and address space layout randomization (ASLR). ASLR is an application-processor class feature, but the stack limit and "no execute" capabilities are found on microcontrollers as well. A variety of coding techniques can make code more robust to glitching attacks *(Fig. 4)* and side-channel observations.

In addition to these techniques, it's recommended to perform penetration testing on firmware images to identify and fix vulnerabilities. This can be carried out through third parties or internally using open-source tooling. A combination of internal and third-party testing is advisable. The most common form of analysis for interface exploits is fuzzing; numerous online resources explain this technique.

SECURING FLASH AGAINST TAMPERING

While a true secure boot requires an immutable root of trust (typical in ROM), many low-cost MCUs can make a reasonably secure root of trust in flash. This relies on the ability of hardware to lock pages of flash to prevent their erasure or programming. If this feature is available, a non-updatable bootloader can be placed into a flash page and that page is locked.

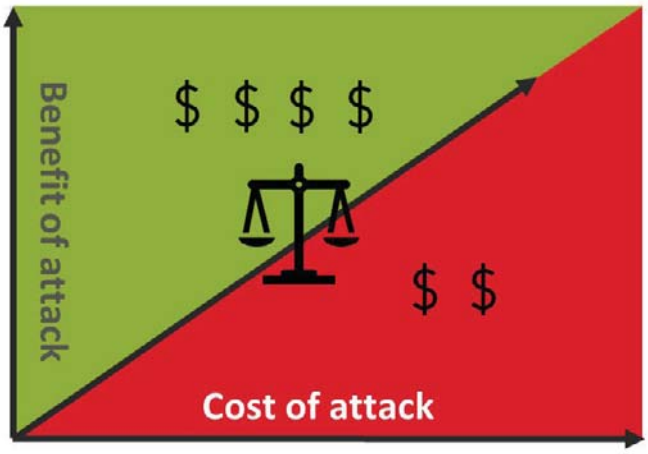
In addition to providing bootload services, the immutable bootloader can also provide secure-boot services to check the signature of the application before allowing it to execute. In this case, the bootloader contains a public key used to validate images to be installed and to validate the flash contents on boot. If an attacker attempts to edit the contents of the application, the bootloader will detect that the signature no longer matches and prevent the execution of the tampered code.

PROTECTING CONFIDENTIAL DATA

Protecting confidential data on a low-cost IC is difficult. While some techniques make it more difficult for attackers to extract confidential data, ultimately hackers will find a way to extract the information if it's valuable enough.

The primary defense of a low-cost system is to simply not have valuable data present in the first place. Persistent (flash) and transient (RAM) data should be evaluated to see if it's required. If not, the data should be removed from the system. If required, it should be stored for as short a time as possible. If the system has other components that can store sensitive data, such as a smartphone or cloud application, move the data to those platforms because they typically have better capabilities to physically protect stored data.

In addition, implementation of features that require storage of confidential data should be carefully evaluated.



5. Focus on adding security features that make the cost of an attack higher than the potential benefits of the exploit.

If a desired feature can't be adequately secured on a low-cost device, it should not be implemented, or the system should be updated to a more secure IC.

External Secure Element ICs do a good job of storing confidential data, but if the application is compromised, the attacker can do everything the application can do. For example, while storing a key in a Secure Element will prevent a compromised application from getting the key value, it will not prevent the application from using that key to sign or decrypt messages. That ability can be just as damaging as gaining access to the key value itself. Anything the application can do using an external Secure Element can also be performed by a compromised version of that application.

Remember that confidential data is accessible beyond the device. Protecting a symmetric key on a device is ineffective if the key is being emailed around the engineering office or stored in an unsecured source control repository.

Cryptography users should consider how confidential data (i.e., cryptographic keys, firmware images, or source code) is stored and accessed in their development office. This is critical to ensuring that confidential information is well protected. Developers should use hardened storage mechanisms such as a hardware security module (HSM) or trusted platform module (TPM), institute policies and procedures for accessing secure information, implement access controls, and implement correct logging and auditing procedures.

FUNDAMENTAL LIMITATIONS OF LOW-COST SYSTEMS

It's important to understand the reasons why IC vendors include hardware-security mechanisms in their chips and why many OEMs pay for them. Low-cost systems have fundamental limitations and will never be as secure as a system with the proper dedicated hardware:

- Systems without hardware side-channel attack countermeasures will leak key information.
- Systems without process isolation hardware will be vulnerable to application code exploits.
- Systems without an on-board secure boot will be vulnerable to firmware tampering.
- Systems without a Secure Element (secure key storage) will be vulnerable to key extraction.

Since these vulnerabilities can't be eliminated in low-cost systems, focus on manipulating the attacker's cost/benefit calculation *(Fig. 5)*. Attackers will develop an exploit if the benefit they gain (money, enjoyment, notoriety, geopolitical aims) exceeds the cost (money,

effort) required to generate it. Instead of making key extraction impossible, reduce the value of the key and increase the cost of extraction to the point where that cost is higher than the value of the key.

If the cost of the attack exceeds the value of that exploit, the designer must consider moving to a more expensive security-enabled device.

CONCLUSIONS

While low-cost or legacy devices have limitations, you can take many steps to make end products and systems more secure. Designers should carefully consider the security requirements of both new systems and upgrades to existing systems. In some cases, these requirements will require advanced hardware features. Applying simple, common-sense techniques and best practices can also enable legacy or low-cost devices to thrive and succeed in today's IoT market.

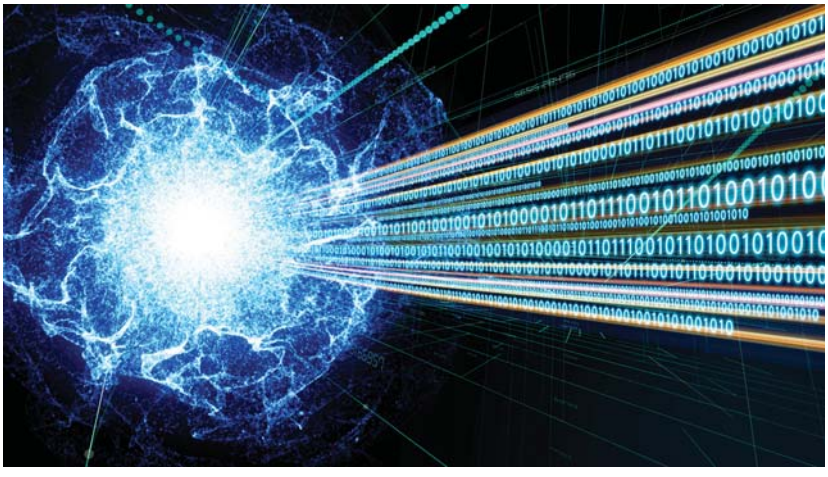
AD INDEX

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.

JEFF LEWIS | Senior VP of Business Development, Spin Memory
www.spinmemory.com

The Challenge at the EDGE OF MEMORY

The recent drive to bring intelligence to the edge, rather than the cloud, has created a conundrum—hardware constraints are beginning to cripple innovation. But new memories are being developed to replace the industry standards of SRAM and DRAM to solve this dilemma.



If data is the new oil, then artificial intelligence (AI) is what will process data into a truly invaluable asset. It's this belief that's causing the demand for AI applications to explode right now. According to PwC and MMC Ventures, funding for AI startups is rapidly increasing, reaching over $9 billion last year with tech startups that have some type of AI component receiving up to 50% more funding compared to others.

This intense investment has led to rapid innovation and advances for AI technology. But the traditional AI use model of "sweep it up and send it to the cloud" is breaking down as latency or energy consumption can make transmission impractical. Another major challenge is that consumers are increasingly uncomfortable having their private data in the cloud potentially exposed to the world.

For those reasons, AI applications are being pushed out of their normal data-center environments, allowing their intelligence to reside at the edge of the network. As a result, mobile and IoT devices are becoming "smarter," and a whole variety of sensors—especially security cameras—are taking up residence at the edge. However, this is where hardware constraints are beginning to cripple innovation.

Increasing the amount of intelligence living at the edge requires much more computational power and memory compared to traditional processors. Studies have repeatedly shown that AI model inference accuracy strictly relies on the amount of hardware resources available. Since customers require ever-higher accuracy—for example, voice detection has evolved to multifaceted speech and vocal pattern recognition—this particular problem only continues to intensify as the complexity increases with these AI models.

One significant concern is simply the need for electrical power. Arm has predicted that there will be 1 trillion connected devices by the 2030s. If each smart device consumes 1 W (security cameras consume more), then all of these devices combined will consume 1 terawatt (TW) of power. This isn't simply an "add a bigger battery" problem, either. For context, the total generating capacity of the U.S. in 2018 was only slightly higher at 1.2 TW. These ubiquitous devices, individually insignificant, will create an aggregate power catastrophe.

Of course, the goal is to never let the power problem get to that point. AI developers are simplifying their models, and hardware power efficiency continually improves via Moore's Law and clever circuit designs. However, one of the major challenges remains the legacy memory technology, SRAM and DRAM (static and dynamic RAM, respectively). These memories are hitting a wall on size and power efficiencies and now often dominate system power consumption and cost.

THE EDGE-COMPUTING CONUNDRUM

The core promise of AI is also its biggest challenge for the edge: the model needs to be constantly adapting and

improving. Not only do AI models require a colossal amount of data and time to learn, but they're never truly "done." If it was that simple, self-driving cars wouldn't still have so much difficulty simply getting out of a parking lot.

Even when AI models have been transitioned to the edge, they still need to be capable of continuing to learn and develop without the hardware amenities available in a data-center environment. These models need to continue tuning themselves while minimizing processing power and energy consumption and simultaneously optimizing and maximizing the available local storage. Some applications will minimize this requirement by improving the AI model in a cloud environment, and then frequently updating the edge device with the latest model version.

More interesting, though, are hybrid approaches, such as Google's Federated Learning Model, which enable the model to be optimized using local data. This requires robust edge compute power to support incredibly frequent neural-network model updates from the cloud. However, since the learning is never "complete," the model must consistently devote substantial processing power and memory to continue improving.

This is precisely the problem, as the AI models struggle to accomplish these goals at the edge.

MEMORY: IT'S WHERE THE POWER GOES

As the apocryphal story goes, when the infamous bank robber Willie Sutton was asked why he chose banks as his targets, he replied "because that's where the money is." For many edge AI devices, most of the power is consumed in the memory system. AI processing—especially training—is very memory-hungry and utilizing off-chip memory has become a necessity to keep up with performance improvements. Google has found that in a mobile system over 60% of the total system power budget is used to transfer data back and forth between on- and off-chip memories. This is more than the processing, sensing, and all other functions combined.

The obvious answer is therefore to eliminate these data transfers by putting all of the memory on-chip.

However the current on-chip memory of choice, SRAM, is simply too large and power-hungry. If transferring data off-chip is the biggest power hog, close behind it is the power consumed by the SRAM on-chip memory. And due to SRAM's large size, one quickly runs out of area on the chip to add enough memory for AI applications.

To make AI at the edge truly successful, memory must be able to address performance demands on-chip and perform perception tasks locally, with high accuracy and energy efficiency.

NEW MEMORY FOR THE EDGE

All of these factors have made the AI landscape a fertile ground for experimentation and innovation with new memories that have unique or improving characteristics. Hardware is becoming the key performance bottleneck, and solutions to the bottlenecks become differentiators. That's the reason why leading internet players, such as Google, Facebook, Amazon, and Apple, are rushing to become silicon designers in search of a hardware competitive edge. Hardware has emerged as the new AI battlefield.

Necessity begets invention, and the necessity for faster AI chips that use less power has opened opportunities for potentially denser, more efficient memory technologies. One such promising technology is magnetic RAM (MRAM), a memory that's bound to cross paths with AI as it rapidly moves toward higher density, energy efficiency, endurance, and yields. The semiconductor industry is beginning to invest heavily in MRAM, as the technology's potential slowly becomes reality. Initial research has shown it offers a number of benefits that are ideal for intelligent edge applications.

The ubiquitous on-chip working memory today is SRAM, but it has flaws. It's the largest memory type, meaning it's the most expensive per bit, and every bit "leaks" (wastes power) whenever the memory is powered on. MRAM is the only promising new memory that has the speed and endurance to replace SRAM.

Since MRAM uses a very small memory bitcell, it can be three to four times denser than SRAM, allowing for more memory to reside on-chip and thus eliminating or reducing the need to shuttle data off-chip. MRAM, is also non-volatile, meaning that data is retained even when the power is shut off. This virtually eliminates memory leakage, which is critical for applications where the AI chip remains idle for extended periods of time.

MRAM isn't the only memory getting attention. The demand for AI applications and intelligence at the edge is leading a memory revolution within the semiconductor industry for a wide variety of applications.

Other new high-density non-volatile memories, such as 3D XPoint (Intel's Optane), phase-change memory (PCM), and resistive memory (ReRAM) also bring new possibilities and unique advantages for storage applications. While neither as fast or high-endurance as MRAM, and therefore not replacements for SRAM, these non-volatile technologies are extremely dense and provide unique speed and power advantages over flash memory.

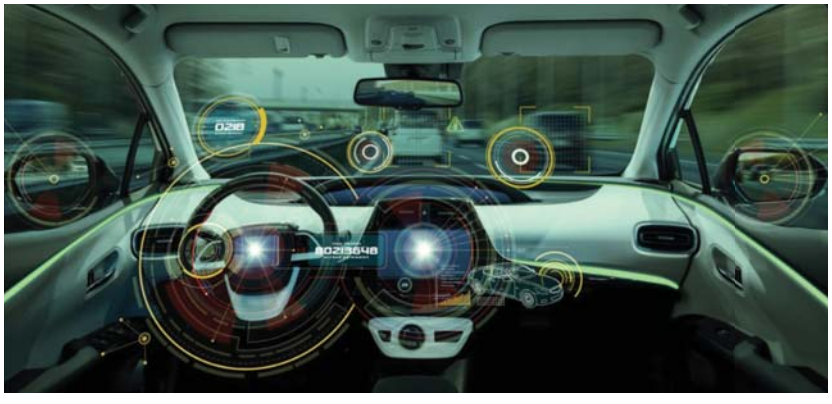
In addition, significant neuromorphic research is investigating using these new memory bitcells directly as the synapses and/or neurons of a neural net. Most research is focusing on ReRAM, although other technologies such as MRAM are also being explored.

For the first time in decades, silicon startups are shaping our future with new, innovative memory technologies. MRAM and the other memories are the catalysts that will change the possibilities of modern technology and applications. ed

WILLIAM WONG | Senior Content Director
bwong@endeavorb2b.com

Ada and RISC-V Secure Nvidia's Future

Nvidia is using RISC-V for its security processor and programming it using Ada/SPARK.



Nvidia tends to hype the machine-learning (ML) capabilities of its system-on-chip (SoC) solutions like its DRIVE AGX Orin that targets automotive applications. Often lost in the mix is the security processor that is part of the package but these are critical to safe and secure applications.

Nvidia's security support has used a custom processor, but it's moving to RISC-V for future implementations. The company isn't alone in its adoption of RISC-V. Western Digital is taking advantage of SiFive's RISC-V designs across the board for its storage solutions. Nvidia isn't changing the SoC's core processors at this point. These are still Arm Cortex cores, but the security processor is essentially isolated from the rest of the system and it runs its own firmware.

That firmware will not be written in C though. It is being done in SPARK, a provable subset of the Ada programming language. Ada 2012 added contracts to that language and SPARK takes advantage of this feature. It allows programmers to specify details like the characteristics of procedure inputs and outputs. The compiler can then enforce these rules for calls to the procedure as well as how the results will be used.

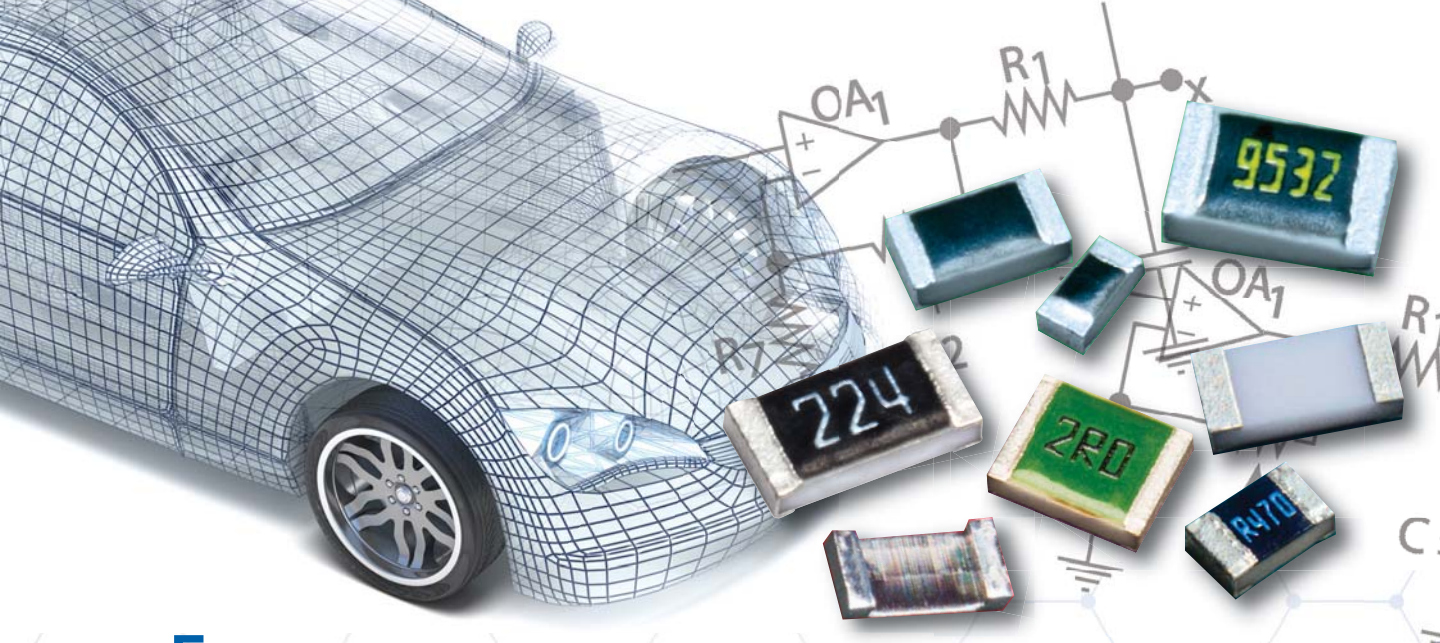
The contract support enables the compiler to prove that a procedure does what it is desired, and the code that calls and uses the results will operate as specified by the contracts and implicit contracts within the language definition. An example of the implicit language checking is the range checking done by Ada for arrays and strings. One of the biggest problems with C and C++ code has been buffer overruns, which can't happen with Ada.

The advantage of including the contract information and allowing the compiler to do the checking is that it can also remove many of the runtime checks often associated with Ada, because they are unnecessary. For example, the implicit array access range checks can be removed from the runtime if it's known that an index to the array can never exceed the size of the array.

Ada's use in avionics is well-known, but it's also ideal for any embedded application. It can be extremely useful in safety applications like medical and automotive. Though the amount of code in a self-driving car will typically exceed that in even an advanced fighter jet, it will need the same level of scrutiny to provide safe and reliable operation. It makes a lot of sense to have the compiler checking that the code is doing what's desired rather than having humans do that same chore by examining the code.

"Self-driving cars are extremely complex and require sophisticated software that needs the most rigorous standards out there," said Daniel Rohrer, vice president of Software Security at Nvidia. "Taking measures like incorporating Ada and SPARK languages into Nvidia platforms can improve the robustness and assurances of our automotive security."

C and C++ remain the primary languages for embedded programming. However, there are advantages to using Ada and SPARK, including cost savings when looking at the total cost of ownership (TOC). Open-source versions of Ada and SPARK tools are available as well as online training.

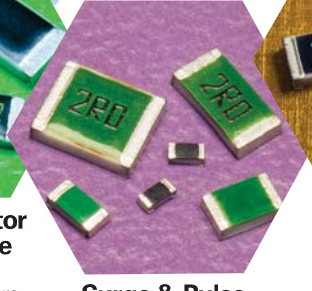


From Concept to Reality

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

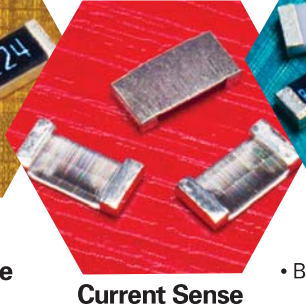


Surge & Pulse Precision Resistors SG73P/S

- 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
IATF16949:2016

KOA SPEER ELECTRONICS, INC.

ANALOG DEVICES

AHEAD OF WHAT'S POSSIBLE™

ADI A²B®

A²B® Transceivers

Delivering high fidelity audio, reducing cable and system complexity

- Low latency (<50 µs) for echo cancellation
- No network stack or memory required
- Eliminates power supply circuitry on slave nodes

Learn More at mou.sr/AD242xW

MOUSER ELECTRONICS