Electronic Design®

THE AUTHORITY ON EMERGING TECHNOLOGIES FOR DESIGN SOLUTIONS

NOVEMBER/DECEMBER 2020
electronicdesign.com

A DEEP DIVE INTO SmartNICs

p10

$10.00

11.5 MILLION+
PRODUCTS ONLINE

1,500+
INDUSTRY-LEADING
SUPPLIERS

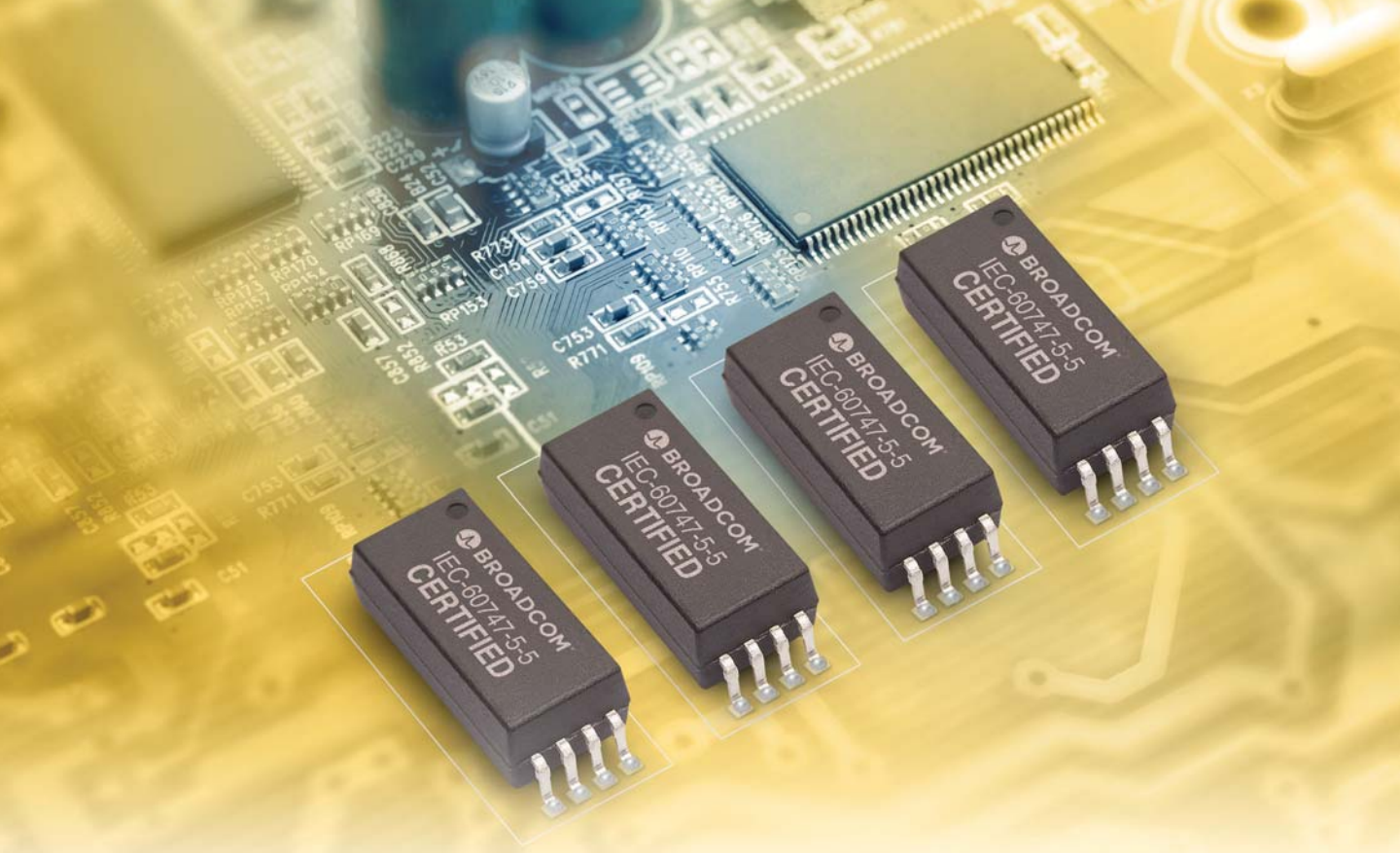
2.6 MILLION+
PRODUCTS IN STOCK

NEW TECHNOLOGIES
ADDED EVERY DAY

SAME-DAY SHIPPING
on orders entered by 8:00 pm CENTRAL
NEXT-DAY DELIVERY

1.800.344.4539
DIGIKEY.COM

Digi-Key ELECTRONICS



BROADCOM OPTOCOUPLERS

Safety Certified Protection...Worldwide!

Only Optocouplers Meet or Exceed International Safety Standards for Insulation and Isolation.

Stringent evaluation tests show Broadcom optocouplers deliver outstanding safety performance and exceptional high voltage protection. Alternative galvanic isolation technologies such as capacitive and magnetic isolators do not perform anywhere near the levels of insulation and isolation provided by Broadcom optocouplers.

GO with OPTO! 50 YEARS
BROADCOM® OPTOCOUPLERS

IEC IEC 60747-5-5 Certified

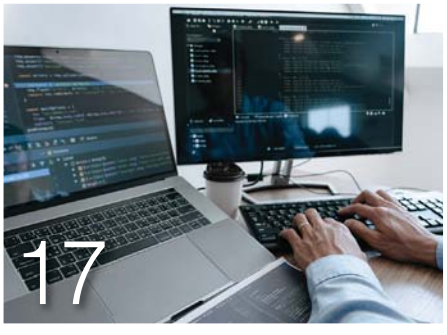
To learn more visit
broadcom.com/optocouplers

BROADCOM®
connecting everything®

IN THIS ISSUE

FEATURES

ID 78618930 © Sashkinw | Dreamstime.com



COLUMNS & DEPARTMENTS

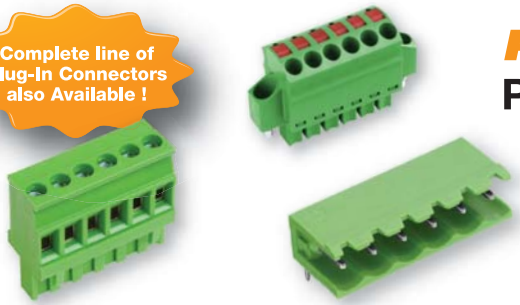
EDITORIAL MISSION:
To provide the most current, accurate, and in-depth technical coverage of the key emerging technologies that engineers need to design tomorrow's products today.

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, 1233 Janesville Ave., Fort Atkinson, WI 53538. Paid rates for a one-year subscription are as follows: $60 U.S., $90 Canada, $120 International. Periodicals postage paid at Fort Atkinson, WI and additional mailing offices. Printed in U.S.A. Title registered in U.S. Patent Office.

Altech®

Top Components for designing Electronics

Complete line of Plug-In Connectors also Available !



Push-in TECHNOLOGY

Printed Circuit Board Terminal Blocks

- 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

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

Customization & Labeling Available

State of the Art

TEST & MEASUREMENT

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

Largest Selection

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

Altech Corp.®

Your Source for Automation & Control Components

info@altechcorp.com • 908.806.9400

www.altechcorp.com

ON ELECTRONICDESIGN.COM

Transpilers, Emulators, and Containers, Oh My

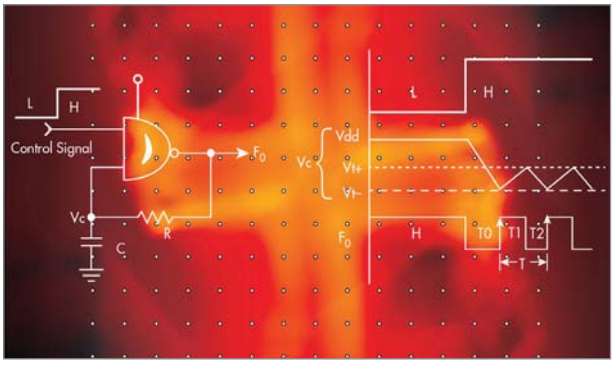
Catching up on several interesting developments that slipped by due to volume, here's a look at the TeaVM translator, Nvidia's Jetson emulator, and Ubuntu's MicroK8s container management project.

https://www.electronicdesign.com/altembedded/article/21146296/electronic-design-transpilers-emulators-and-containers-oh-my

How Simulation Provokes Better EV Battery Designs

Electric-vehicle manufacturers are leveraging simulation to not only improve battery design, but also to scale production and continue meeting industry demands for sustainable transportation solutions.

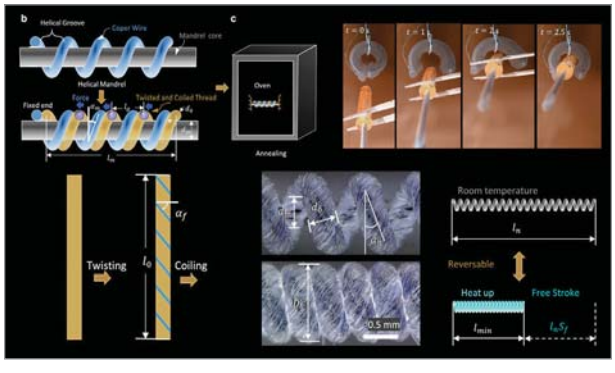
https://www.electronicdesign.com/markets/automotive/article/21145287/how-simulation-provokes-better-ev-battery-designs



Gated Oscillator Starts Without Delay

This Idea for Design shows how proper component selection can help reduce startup time for gated oscillators.

https://www.electronicdesign.com/resources/ideas-for-design/article/21140690/gated-oscillator-starts-without-delay



Coiled Conductive-Thread Actuators Eerily Emulate Muscle Motion

Twisted and coiled conductive threads, when heated by current flow, can contract and provide substantial linear and bending forces for grippers in soft robotics.

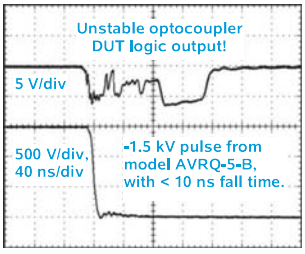
https://www.electronicdesign.com/industrial-automation/article/21147106/electronic-design-coiled-conductivethread-actuators-eerily-emulate-muscle-motion

FOLLOW US ON FACEBOOK AND TWITTER

(ElectronicDesign) (@ElectronicDesgn)

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

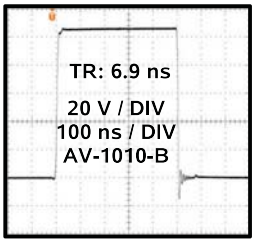
AVTECH

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

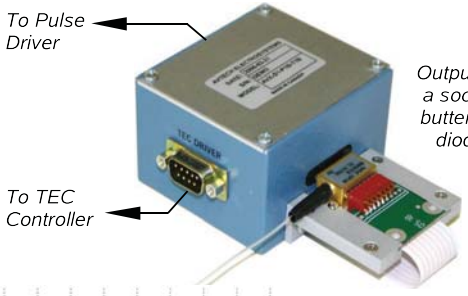
Variable baseline and burst mode options and output currents to 8 Amps with accessory transformers.

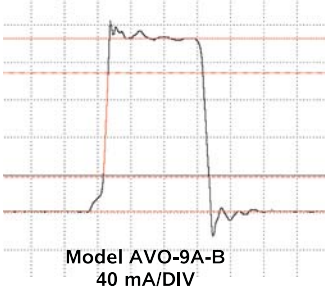
AVTECH

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.

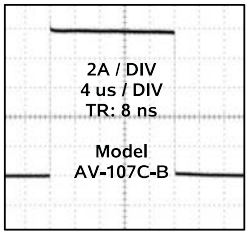
AVTECH

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

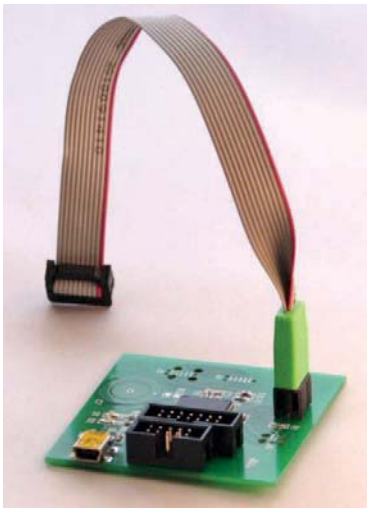
Nanosecond Electronics
Since 1975

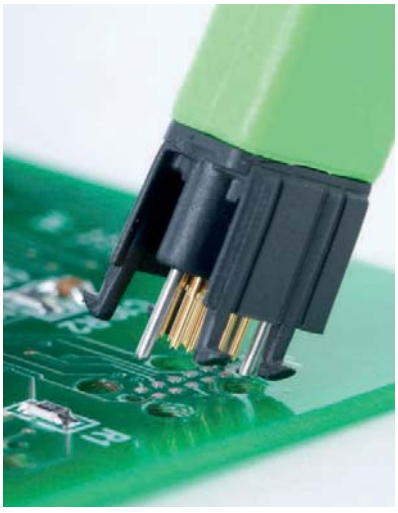
Pricing, manuals, datasheets and test results at:
http://www.avtechpulse.com/current

ElectronicDesign

NOVEMBER/DECEMBER 2020

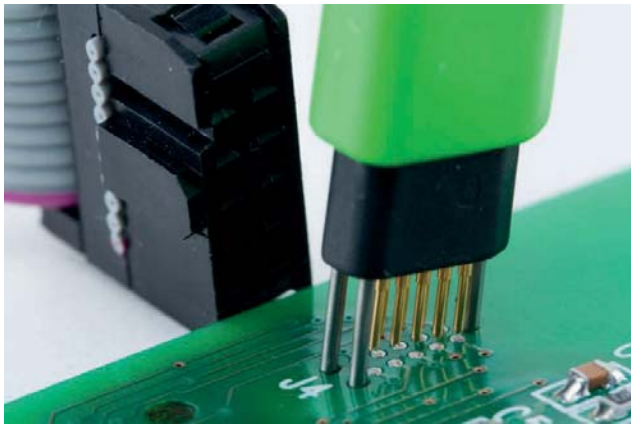
Tag-Connect

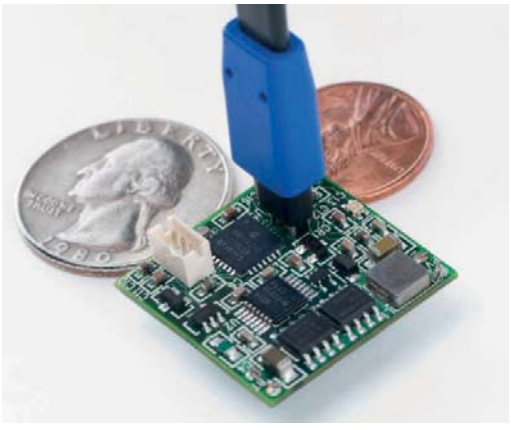




JTAG Connector Plugs Directly into PCB!!

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

FPGAs Emerge Everywhere

FPGAs aren't just for special embedded applications—now they're in everything from NICs to SSDs.

I remember when FPGAs were specialized devices requiring custom programming that could only be generated by wizards. They were often hidden inside radar systems or other high-performance platforms where the cost of the devices and programming was warranted. These days, that's all changed.

Low-cost, flash-based FPGAs are in all sorts of embedded devices, and high-performance FPGAs are turning network interface cards (NICs) and solid-state drives (SSDs) into programmable computing platforms. It was once the realm of ASICs, where high volume made these practical for applications like encrypted disk drives. This assumes standard software interfaces with sufficient adoption to make the solutions practical.

Nowadays, FPGAs are mainstream. FPGA boards are found in the cloud alongside GPGPU boards and machine-learning/artificial-intelligence (ML/AI) accelerators. FPGAs can provide such support, but custom chips are often better for optimized ML/AI support. FPGAs offer a more flexible approach, allowing more than just ML/AI acceleration to be incorporated into hardware. Standardizing the APIs and driver interfaces has changed the way people view FPGAs, GPGPUs, and ML/AI support in the cloud.

The same is happening at the peripheral level. SmartNICs with bundled FPGAs accelerate and offload network processing chores from the host. This helps minimize system bandwidth requirements as well. It can also help keep data moving at wire speeds, where a host might otherwise become overwhelmed.

Xilinx's SmartSSD computational storage device (CSD) is an example of how FPGAs are playing a role in storage *(see figure)*. It's not the first FPGA/SSD device on the market, but the push for use with standards such as the PCI Express-based (PCIe) NVMe make it a very interesting platform.

Like SmartNICs, SmartSSDs can utilize the FPGA to implement a variety of features that might otherwise need to be handled by a host processor. For example, a SmartSSD may be programmed to handle data compression and encryption. It could also turn the basic SSD into a content-addressable memory or even an ML/AI engine. Applications like real-time multimedia transcoding fit in both SmartNIC and SmartSSD arenas.

Xilinx's SmartSSD computational storage device (CSD) was shown at this year's Flash Memory Summit. It looks like a standard U.2 SSD, but it actually appears as two logical devices: an SSD and a compute system.

Moving computational chores closer to the peripherals makes sense, especially as data centers become disaggregated. It also makes sense in embedded applications, where functionality can be distributed to facilitate development and broaden modularity.

Keep an eye out for other peripherals, from cameras to motor control, to come with much more intelligence than in the past. ed

The higher standard of high voltage DC/DC power supplies.

UMR Collection & PMT Series

Cost effective, analog controls and monitors, and the best performance...

...for what you need now

SPS Series

Fully digital, extremely customizable, with serial communications...

...for making the future now

Outputs up to 6kV @ 30W

Available as biasing or capacitor charging modules

Contact DTI today to discuss your high voltage design.

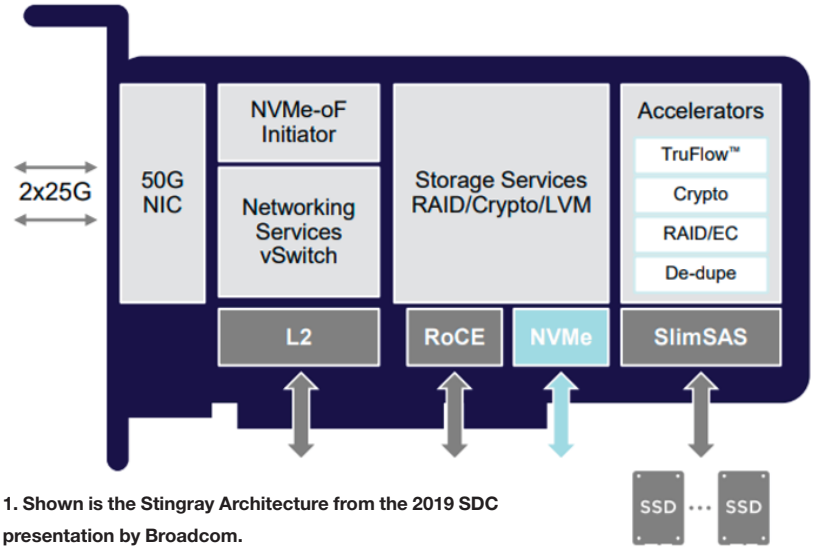
DEAN TECHNOLOGY

+1.972.248.7691 II www.deantechnology.com

SCOTT SCHWEITZER | Technology Evangelist, Xilinx

SmartNIC Architectures: A Shift to Accelerators and Why FPGAs are Poised to Dominate

SmartNICs incorporate various additional computational resources beyond a generic NIC. But like snowflakes, these architectures differ, so we'll dive deep into several approaches from the biggest and most popular vendors.



1. Shown is the Stingray Architecture from the 2019 SDC presentation by Broadcom.

Generic network interface cards (NICs) are built around a single application-specific integrated circuit (ASIC) engineered to be an Ethernet controller. Good examples of these are the ConnectX line by Mellanox, the Broadcom NetXtreme, or the Xilinx XtremeScale. Often, these chips are further optimized around a second design objective; for example, the ConnectX line also supports Infiniband, while the XtremeScale is focused on kernel bypass. These controllers are excellent at what they do, they represent the finest in the industry, but they are not SmartNICs.

For the purpose of this article, we're defining a SmartNIC as a NIC that allows for additional software that can be loaded into the NIC at some point after it's purchased to add new features or support other functions. Much like when you buy a smartphone and later install apps from that vendor's app store.

The capability to load future code into a NIC, making it a SmartNIC, requires additional computational power and onboard memory not found in generic NICs. Most SmartNIC approaches start with a foundational Ethernet controller either on the chip in silicon, as firmware, or as a separate chip on the adapter. Then one of three approaches below are used to make an otherwise generic NIC smart by increasing its computational power through the addition of a:

- Collection of many Arm cores; some call it clustering, others use the term grid or tiles.
- Flow processing cores (FPCs), which are custom-designed network processors, often P4.
- Field-programmable gate array (FPGA), programable logic.

Many of these SmartNICs will often use one or more Arm cores for control-plane management within the NIC. Some even allow for loading a modified Linux kernel into one or more of these cores. These Arm cores typically handle loading code into the other processing elements, gathering statistics and logs, and watching over the health and configuration of the SmartNIC. They don't touch any of the network packets, and they often run "out-of-band," meaning that they can't be accessed via "normal" network interfaces or PCIe commands.

Also, the cores should only accept duly signed firmware bundles over previously secured interfaces. Beyond architectural sketches, we won't call out these control-plane Arm cores below, as they're required plumbing. And on their own, they don't generally add value to the feature set offered by that SmartNIC.

PRODUCT COMPARISON

To understand how SmartNICs differ from generic NICs, let's take a dive into the leading SmartNIC products

from four of the biggest NIC companies on the planet, and two upstarts to see what improvements they offer. The six companies selected are Broadcom, Intel, Nvidia (previously Mellanox), Netronome, Pensando, and Xilinx. We'll also throw in some insights on a stealth project currently called Fungible.

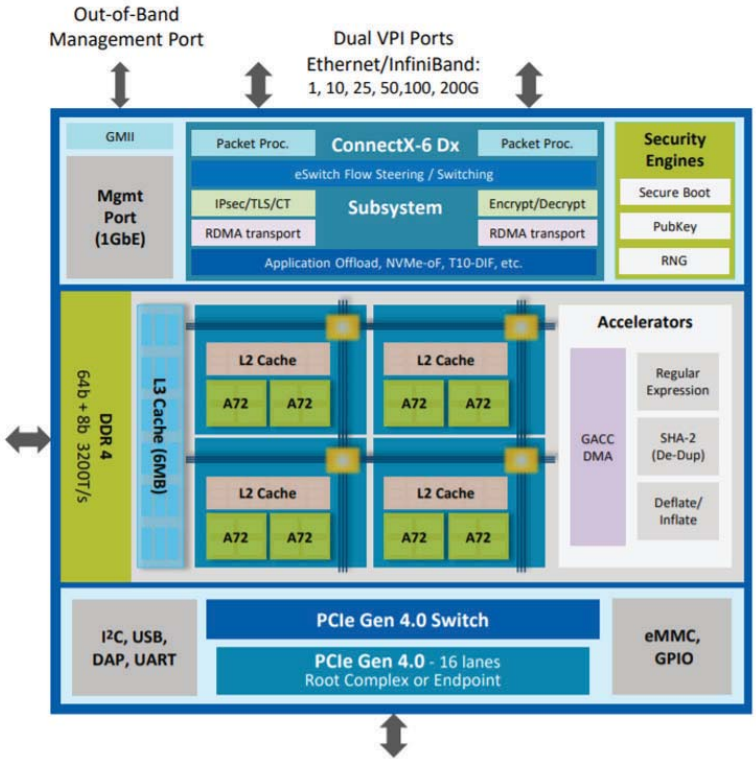
Broadcom

Broadcom is the undisputed leader in the commodity Ethernet NIC controller market. So, when it came time to assemble its Stingray SmartNIC and enter the fray, the company went with a single-chip approach *(Fig. 1)*. Single-chip SmartNIC solutions are always less costly to produce at the board level than many chip boards by other competitors.

Broadcom designed the NetXtreme-S BCM58800 chip at the heart of Stingray by starting with the logic from its NetXtreme E-series controller as a base. Then eight Arm v8 A72 cores clocked at 3 GHz were laid down in a cluster configuration. At 3 GHz, these may be the fastest SmartNIC Arm cores. Also, Stingray can be configured with up to 16 GB of DDR4 memory. Next, some logic was mixed in to offload encryption at up to 90 Gb/s as well as offloads for storage processing like erasure coding and RAID.

Finally, Broadcom added its somewhat mysterious TruFlow technology. This is a configurable flow accelerator for moving common network flow processes into hardware; we believe the company is using a P4 processor. This frees the Arm cores to focus on more computationally intensive tasks at both the flow and packet level. From what's published, TruFlow seams to offload tasks like Open vSwitch (OvS) in hardware. The company also claims that TruFlow implements many of the classic software-defined networking (SDN) concepts like classification, matching, and action in hardware. So, Stingray has two programable components, TruFlow and a cluster of four 3-GHz, dual-core Arm v8 A72 complexes.

Broadcom is preparing to move Stingray to the 7-nm process later this summer, which will enable scaling from eight to 12 cores. Understanding the complexity of the product it's providing, the company also offers a Stingray developer's kit for both SmartNIC application development and storage controller development. This is less common than you might think, and it's a necessary component for a complete SmartNIC product offering.



2. Mellanox's Bluefield 2 Architecture as taken from the product datasheet.

Nvidia

Nvidia defined graphical processing units (GPUs), which has become the darling accelerator of choice in high-performance computing (HPC). Earlier this year, Nvidia finally closed on the purchase of Mellanox for $7 billion. In its quest to own the HPC market, the company has picked up the premier Infiniband interconnect vendor, so that it can offer HPC customers a complete solution. This is very similar to what Cray has done in the past.

Nvidia also recently acquired Cumulus Networks, which is the leader in open-source Ethernet switch OSes. Software has always been a Mellanox weak point, and clearly, Nvidia realized this early. With regard to SmartNICs, Nvidia also picked up an interesting Easter egg with the Mellanox acquisition.

Mellanox is one of the oldest entrants in the SmartNIC space, but it was through acquisition. Its current Bluefield 2 solution *(Fig. 2)* was built from the purchase of Tilera via EZchip in 2015. Tilera had one of the first highly parallel SmartNIC implementations using intellectual property, which evolved from a much earlier MIT research project.

Essentially, Tilera arranged processing cores as tiles on their chip with each core having a high-speed bus to the four cores around it. Its flagship product back in 2013 supported up to 72 MIPS cores, memory controllers, encryption modules, a PCIe block, and the mPipe, a collection of channels out to several MACs with SFP+ connectors. Mellanox advanced this forward by replacing the cores with Arm and swapping out the mPipe for ConnectX logic. The current

core count is eight Arm v8 A72 cores, just like Broadcom, but only clocked at 2.4 GHz. These are arranged as a cluster of four dual-core Arms. Bluefield is currently using the Avago 16-nm process, but like Broadcom, it too should be going to 7 nm this summer and moving from eight to 12 cores.

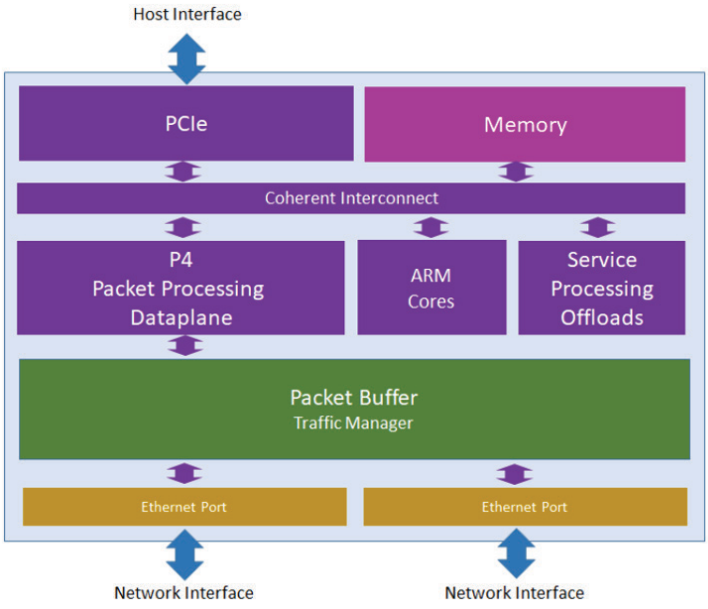
While this approach is very similar to the Broadcom Stingray, it lacks the parallel P4 processor at the heart of Broadcom's architecture. P4 is where everyone is going, just not everyone has stated it, or gotten there, yet. Broadcom, Xilinx, and Pensando are all focused publicly on P4. This again is where Cumulus Networks comes in—it has experience in P4 programming, and it wouldn't be a surprise for Nvidia to craft a P4 packet processing engine to front end a future Bluefield offering.

Pensando

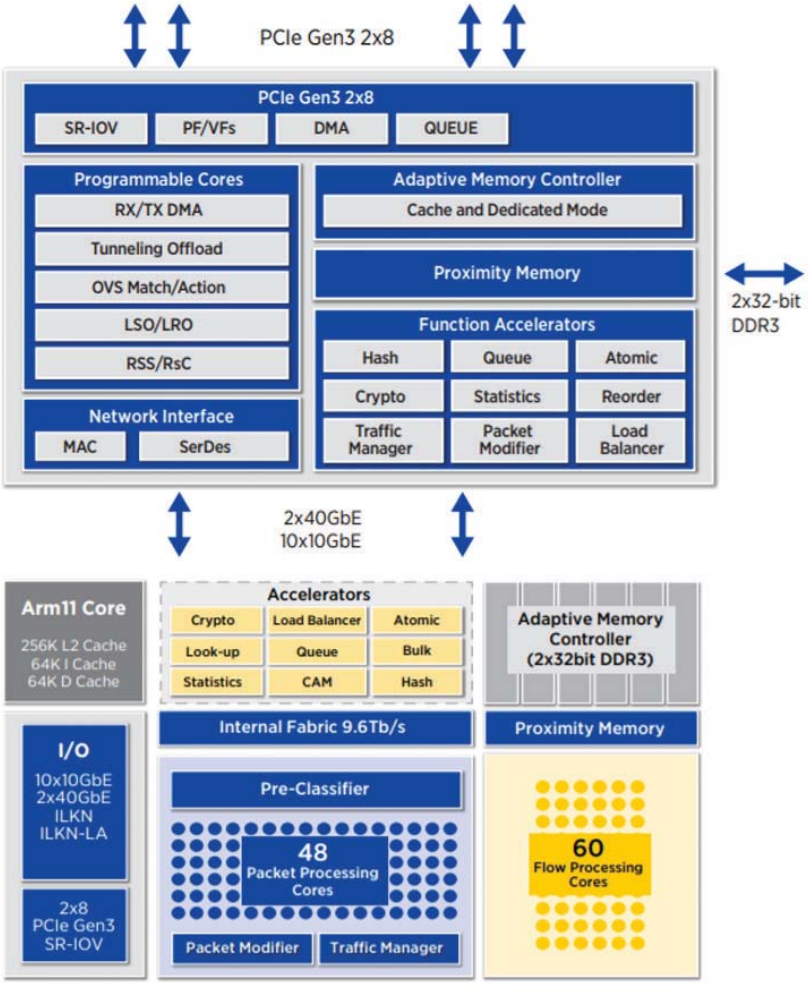
The newest SmartNIC startup is Pensando, founded by the "MPLS" team of Mario Mazzola (Vice Chairman of the Board), Prem Jain (CEO), Luca Cafiero (Board member & technical advisor), and Soni Jiandani (Chief Business Officer), a team of engineers who spearheaded the development of several of Cisco Systems's key technologies, and founded four startups that were acquired by Cisco. John Chambers—the former CEO of Cisco—is the Chairman of the Board.

Given the reputations of the founding group and Chairman of the Board, and prior experience, the general consensus is that Pensando is proving its technology works with some anchor customers, and then it will be packaged up for a Cisco acquisition. Cisco has some generic NIC technology and several internal SmartNIC projects, but the word on the street is that these may not see the light of day, so Pensando is clearly looking to fill this void.

Pensando originally had two products, but recently dropped back to a single one, the DSC-25 Distributed Services Card, which even sounds like a Cisco product name. It appears to be a single P4 processor chip, with an Arm



3. The Pensando DPU Architecture (from P4.org blog post 02/24/2020).



4. Netronome's NFP4000 Flow Processor architecture.

for some auxiliary processing, both of which the datasheet claims support up to 4 GB of onboard memory *(Fig. 3)*.

The processor, called Capri, is a P4 programmable unit with multiple parallel stages; the exact degree of parallel processing is unknown, though, as is the packet performance, latency, and jitter which haven't yet been published. Pensando keeps the P4 applications tight so that they remain inside Capri's cache as a cache miss, resulting in a memory fetch for an instruction that kills performance across all metrics. Other additional computational units, called Service Processing Offloads, handle encryption, storage processes, and other tasks. Pensando claims that Capri can provide wire-rate performance.

Netronome

Netronome the is grey-bearded startup in this space, having started in 2003 and taken five rounds of funding thus far for a total of $73 million. The company has been actively promoting P4 since 2015, when it demonstrated its first SmartNICs utilizing the technology. Since then, Netronome has made some significant strides, but lately the rumors are that it's stumbled and may be exiting the market.

Figure 4 shows Netronome's current NFP4000 Flow Processor architecture. Instead of a single P4 processing engine, the company leverages two classes of programable cores: 48 packet processing cores, and 60 flow processing cores. Additional silicon is dedicated to classification, modification, and management. All of these cores can be programmed in P4.

The result is that Netronome claims the cores can sustain a single 100-Gb/s link at wire-rate, 148 million packets per second, with millions of exact matches and wild-card flows. Also, they can support 100K+ tunnel connections requiring encapsulation. An extremely long list of target applications that the technology can support includes intrusion detection systems (IDS), intrusion prevention systems (IPS), next-generation firewall (NGFW), routers, load balancers, packet brokers, SDN, NFV, and a whole host of other applications.

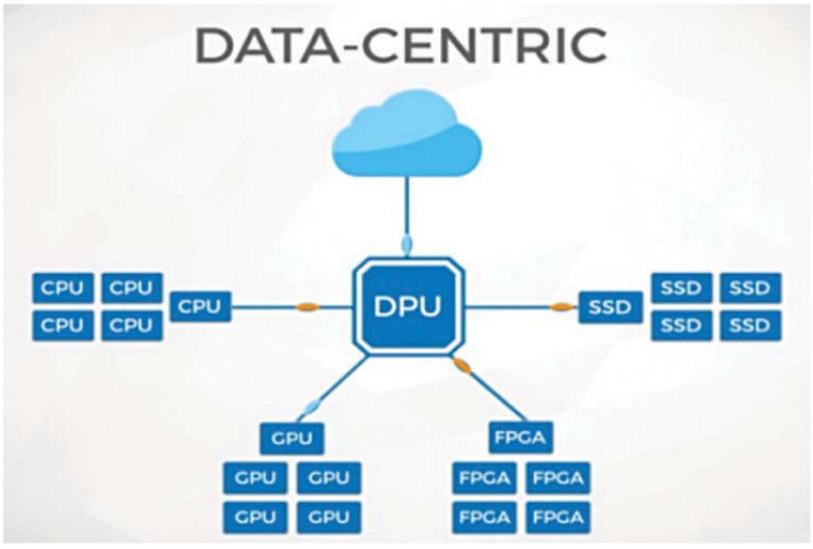
Fungible

One other potentially large snake is still lurking in the P4 grass that has yet to raise its head, namely Fungible. It's on the verge of announcing a product after having taken three rounds of funding totaling nearly $300M, with $200M of that coming from Softbank Vision in the C-round last year. Currently, the company has 180 people, no products, no revenue, and no visible customers.

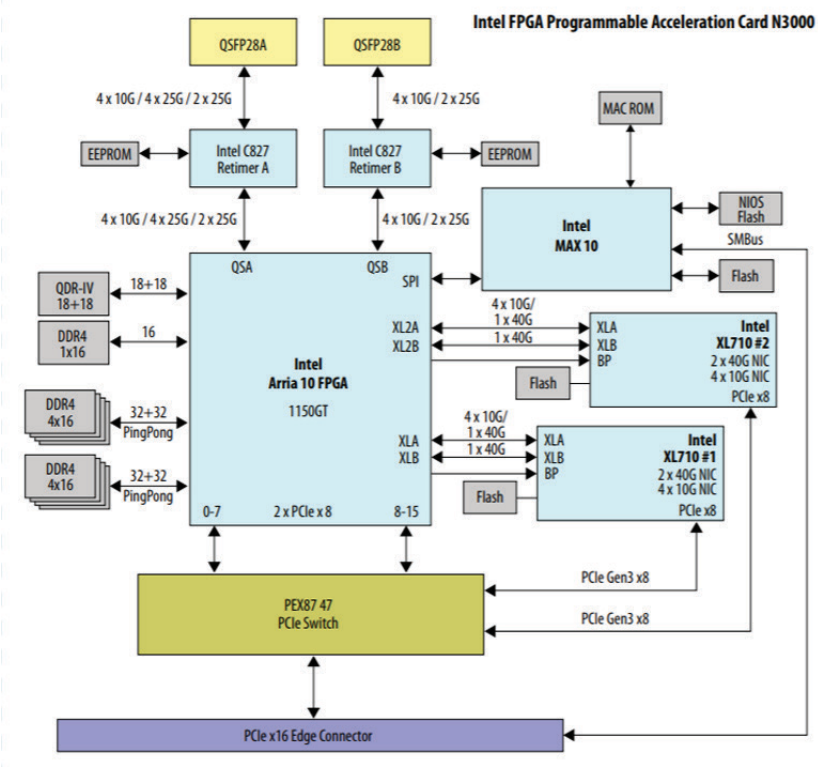
Fungible claims to be producing a data processing unit (DPU), but

the actual architecture and makeup is unclear. The company provided the diagram shown in *Figure 5*, which is about as cloudy as an architecture diagram can become before becoming pointless. Various articles have mentioned that its product will be released this summer, so we wanted to make sure this company was on your radar.

One of Fungible's founders, and the Chief Architect, is formerly a 10-year veteran of Chelsio Communications, which specialized in storage-centric Ethernet NICs. The VP of Software and Firmware Engineering is also a former Chelsio employee, but for 13 years. So, while Chelsio was ASIC-driven in the past, the cost of producing a leading-edge SmartNIC Ethernet controller ASIC from scratch these days would easily consume over $50M of its precious capital.



5. Fungible high-level architecture (retrieved from BlocksAndFiles.com blog post, 2/11/20).



6. Illustrated is Intel's N3000 SmartNIC architecture (from the product datasheet).

The speculation is that Fungible will take the path of least resistance to reach revenue and utilize an FPGA platform for its initial product with the secret-sauce ASIC design loaded into this platform. It could then shake out the design while beginning to chase customers and revenue. Loading a design into an FPGA would enable the company to easily fix design flaws and support rapid customer-driven improvements. Today, FPGAs are enormous, and we've started seeing full-blown processor architectures like RISC-V being loaded into these platforms.

Intel

When it comes to a chip company, you don't get any bigger than Intel. For more than a decade, Intel has maintained an unbroken line of high-performance 10-GbE controllers. Its XL710 platform has shipped millions of units and is a staple in many data-center servers.

For the new N3000 SmartNIC, Intel has crafted a board using five of its chips *(Fig. 6)*. This is an expensive approach, as most vendors strive for single-chip designs. Intel fused together a pair of its XL710 Ethernet controllers and an Arria 10 FPGA using a 48-lane PEX8747 PCIe Generation three-switch chip. Eight lanes go to each of the XL710s, 16 lanes to the Arria, and 16 lanes down to the PCIe socket. The fifth chip is a MAX 10 FPGA baseboard management controller (BMC) that manages the FPGA, much like Arm cores are used on other SmartNICs to handle control-plane management.

The board has dual QSFP28 ports that go straight to the FPGA. Then eight lanes of 10G leave the FPGA to each XL710. This is a classic bump in the wire approach, and it enables the FPGA to work on packets prior to them being passed to the XL710s.

Building a NIC using an off-the-shelf Ethernet controller and an FPGA is nothing new. Back in 2012, Solarflare Communications placed an FPGA between the two QSFP ports on a NIC and its Ethernet controller to create the Application Onload Engine (AOE) platform. This was a precursor to the N3000 design selected by Intel, but it enabled Solarflare to deliver impressive tick-to-trade results for financial customers, at the time 350 ns. Today, some eight years later, the record is 24.2 ns.

Intel's approach allows the FPGA to do packet processing prior to the XL710. The company's FPGA has 1,150K programable logic elements and two banks of 4 GB of DDR4 memory, each giving it ample room to handle SmartNIC tasks like:

- Virtual Broadband Network Gateway (vBNG)
- Hierarchical Quality of Service (HQoS)
- Packet classification, policing, scheduling, and shaping
- Virtualized Evolved Packet Core (vEPC)
- 5G Next-Generation Core Network (NGCN)
- Internet Protocol Security (IPSec)
- Segment routing for IPv6 (SRv6)
- Vector Packet Processing (VPP)
- Virtual Radio Access Network (vRAN)

While Intel has targeted the above workloads for the N3000 platform, it's unclear at this point if the company delivered all of the necessary software to offload each of these applications on this SmartNIC. SmartNIC consumers will learn that the devil is in the software—all of these companies are excellent at hardware, but software delivery is a whole other discipline.

Xilinx

The other standout FPGA entrant in the SmartNIC space is from Xilinx, the first company to commercialize the FPGA in the mid-1980s. Today, Xilinx stands atop the FPGA hill while Intel is running a distant second. The company acquired Solarflare Communications in the fall of 2019, and Solarflare has been building both ASIC- and FPGA-based NICs for electronic trading since 2012. The Solarflare engineering team out of Cambridge in the U.K. was behind Xilinx's Alveo U25 SmartNIC *(Fig. 7)*. Hence, it leverages the company's nearly 10 years of experience in this market.

The Alveo U25 attaches the dual SFP28 ports directly to a Zynq series chip. The Zynq is an actual system-on-chip (SoC) as it includes not only an FPGA, but also a quad-core Arm A53 for packet processing. The Zynq then connects directly to the host server via eight lanes of PCIe Gen 3 or through the SerDes to the X2 Ethernet controller chip, which is also connected via eight lanes of PCIe Gen 3 to the host. This

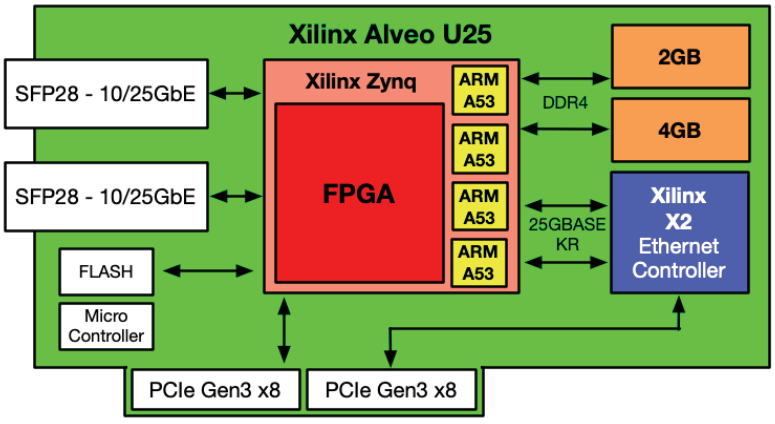
approach enables the Zynq to process packets before they're passed to the X2 chip, or to bypass the X2 entirely.

The Alveo U25 also includes 6 GB of DDR4 memory accessible to the Zynq's FPGA and Arm cores by programs running on this chip. The FPGA has 520K logic elements available, but the provided quad-core Arm more than makes up for the reduced available gate count.

Xilinx is bringing the Alveo U25 to market to initially address customers that have requested Open virtual Switch (OvS) offload capability. The company announced that in the near future, it will be adding offloads for IPsec, machine learning (ML), deep packet inspection (DPI), video transcoding, and analytics. Exhibiting its SmartNIC versatility, two years ago, in the run-up to acquiring Solarflare, the two companies were partners, publicly demonstrating the XtremeScale X2 controller logic running as a soft-NIC inside a much larger FPGA at the OCP Summit.

Like Intel, Xilinx has several computational chip product lines: Kintex, Virtex, Zynq, and Versal. Kintex and Virtex are pure FPGAs, and some models in this line have nearly 3,000K logic cells, almost triple what Intel has used in its N3000. Xilinx has also worked with silicon interposers and layered up to 16 GB of high bandwidth memory (HBM) on Virtex chips. This technology is also appearing in other chips across all four chip lines.

Zynq, which is Xilinx's SoC chip line, includes FPGA programable logic along with quad-core Arms, real-time Arm cores, DDR controllers, and connectivity logic for both Ethernet and PCI Express. Versal goes way beyond SoC to an adaptive compute acceleration platform (ACAP) built on 7-nm chip technology. The ACAP extends the Zynq architecture by adding hundreds of artificial-intelligence (AI) cores, digital-signal-processing (DSP) engines, and much more. AI cores are somewhat new, but they're essentially single-precision compute engines. Eventually, the company will couple its publicly demonstrated SoftNIC with Versal.



7. Xilinx's Alveo U25 architecture diagram. (Source: Xilinx)

CURRENT SMARTNIC CLIMATE

As we've seen from Netronome and even Solarflare Communications, SmartNICs have been a long time in coming. Potentially big customers like Google and Amazon have removed themselves from the market by designing and building their own in-house solutions. Meanwhile Facebook and Microsoft have provided high-level architectures that industry vendors have since rushed in to fill.

While all this was being sorted out, SoCs and more importantly FPGAs have matured to a point where they can now become the foundational technology for SmartNICs. A decade ago, we were awash in GPU products as the first significant wave in hardware-acceleration technology hit. Now with FPGAs scaling above three million logic units, we're also tightly coupling them with other composable processing blocks for networking, memory, storage, and compute. Compute, in this case, means on-chip clusters of cores via SoC blocks or even ACAPs.

With such advances, we're starting to recognize the formation of the second wave of hardware acceleration. Where GPUs needed new programming APIs and tools to support those platforms, the same will be true of FPGAs. The difference, though, is that this technology has been evolving for over 35 years, and its time is coming. So, as the SmartNIC market is finally emerging, it will converge with the next wave of FPGA-based hardware accelerators. That will form a superposition of sorts in the acceleration market, perhaps facilitating a sea change and altering our perspective of computing going forward.

SmartNICs are pushing computation and, therefore, acceleration to the edge of the network, freeing up server CPUs to work on more solutions focused on complex business-critical processing. Studies have shown that networking in highly virtualized environments can consume upwards of 30% of the host's CPU cycles' processing tasks like OvS transactions.

Imagine if storage functions, encryption, DPI, and sophisticated routing could be done in a SmartNIC? This could potentially deliver back to the host CPU complex a significant portion of the CPU cycles usually spent processing these workloads.

To stay ahead of the big guys, new companies like Pensando and Fungible will continue to breathe innovative features and capabilities into the SmartNIC market. At the same time, technology leaders like Xilinx, Intel, Broadcom, and Nvidia will improve foundational computing cores and specialized P4 processing engines. Exciting times are ahead. ED

Smoke Alarm System 2.0

Christoph Kämmerer, Analog Devices, Inc.

Underwriters Laboratories (UL), the author of the U.S. and Canada smoke detection standards, released a revised specification (8th edition) that took effect in May 2020. These new standards are significantly more technically challenging to meet than the previous regulations.

New Smoke Detector Tests

One major change to the standards is the introduction of the hamburger nuisance test. In this test, hamburgers are placed in an oven set at a high enough level to eventually burn, and the detectors must not issue an alarm before a certain amount of smoke has been generated. This requirement to not alarm will bound the maximum sensitivity of a detection system. This test is designed to reduce the number of false alarms generated due to cooking events, as residents disconnecting alarms due to high false alarm rates is one of the leading cases of death in fire-related events.

Another addition to the standard is the flaming polyurethane (PU) test, also known as the burning couch cushion test. Due to the optical scattering cross-sections and physics of different smoke types, sensor response from flaming PU is lower than the response from other smokes at a similar obscuration. As such, the ability to detect flaming PU smoke at specified levels will bound the minimum sensitivity of most, if not all, optical detection systems. In practice, the sensor response to the flaming polyurethane smoke can be difficult to separate from the hamburger nuisance test. In the past, setting the pass/fail criteria for a detector was straightforward, at least to minimally meet agency requirements. For the upcoming requirements, the manufacturing and calibration margins are much tighter and may require an increase in algorithmic complexity. The flaming polyurethane and hamburger nuisance tests occur on different time scales and it is straightforward to create a simple algorithm that looks at the slope or rate of change of the smoke to distinguish between the two fires to pass the UL requirements. However, there is a question of how effective this algorithm is in real-world applications.

In fire room tests, the pass criteria are specified in either time passed since the test was initiated or at a defined obscuration level. A typical obscuration sensor is shown in Figure 1, with a light on one end and the photodetector on the other end. For UL tests, the beam is a sodium vapor lamp 4" diameter and 5' long. Particles in the path of the beam absorb or scatter light out of the beam path, reducing the amount of light that reaches the detector. For different types of smoke, the relationships between an optical scattering system and an obscuration are different. In the case of the hamburger nuisance test and flaming polyurethane test, a 3× difference in obscuration can be nearly impossible to differentiate in an optical scattering system.

A Typical Smoke Detector

A typical smoke detector is made up of a detector, a microcontroller with an algorithm, and additional components such as loudspeakers, LED indicators, and CO sensors. Photoelectric smoke alarms often use a discrete LED (typically near-infrared, 850 nm or 880 nm) and a discrete photodiode with a typical 135° angle between them and a separation of several cm.

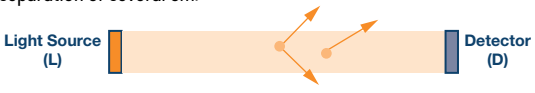


Figure 1. Reference measurement.

LED light of a specific wavelength is scattered by particles onto a photodiode. As shown in Figure 2, the distance between the LED and the photodiode is usually a few centimeters.

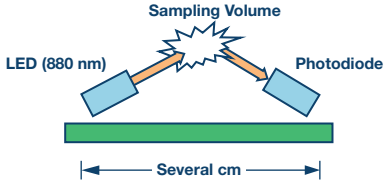


Figure 2. Forward scatter system with an infrared LED.

However, both the discrete design of the smoke alarm and the measurement method result in a few disadvantages. The main one is that monochromatic LEDs lead to a higher false alarm rate because they make it harder for different particles to be distinguished from one another. In addition, a discrete implementation is large and associated with a higher power consumption. Laborious calibrations are also necessary. The technology for the optical components has advanced to the point that the LEDs and the photodiodes, along with the optical front end, can be integrated into a small housing.

Addressing the New Challenge

ADI has created a technology to help address these issues: the ADPD188BI. It directly integrates two LEDs (blue and infrared), a photodiode, and an analog front end. Digital output over I^2C or SPI enables a connection to a microcontroller. A block diagram of the ADPD188BI is shown in Figure 3. As can be seen in the figure, the complete signal chain is realized in a single 5 mm × 3.8 mm chip.

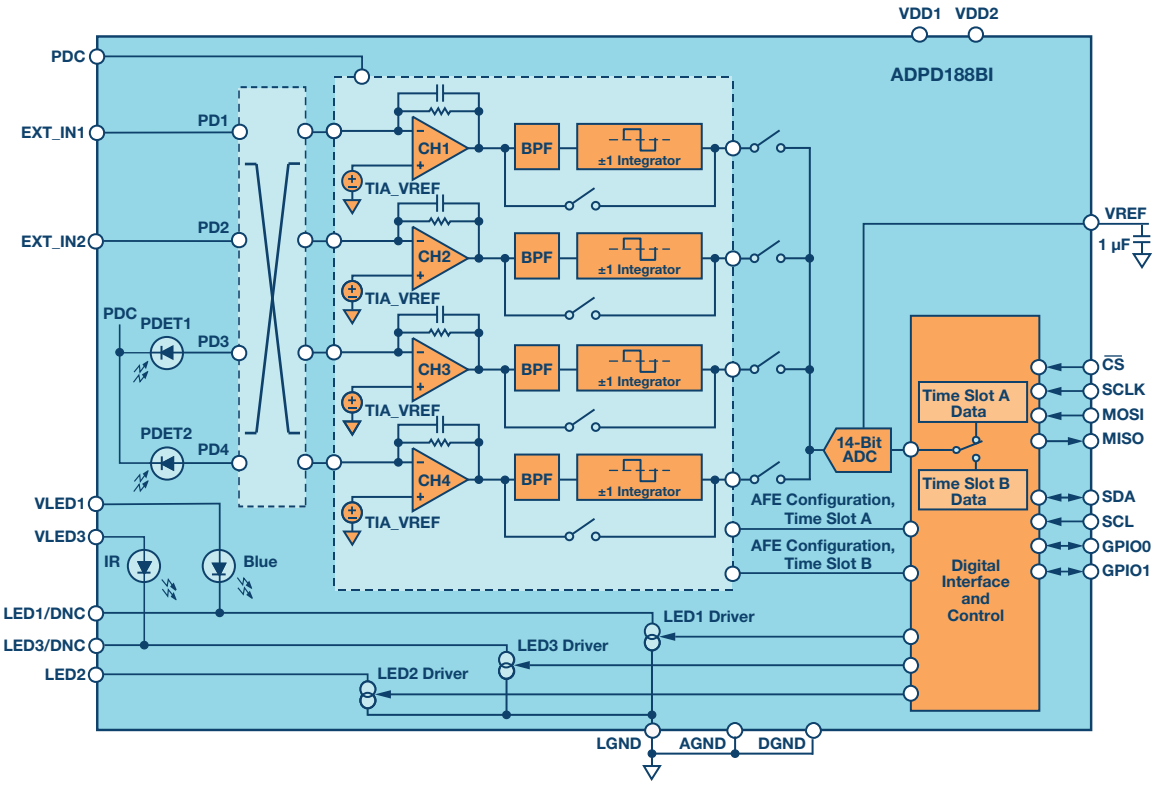


Figure 3. Block diagram of the ADPD188BI.

The ADPD188BI works by emitting a short LED pulse of a few microseconds. Some of this light is scattered by the smoke particles back onto the photodiodes (see the cross-section of the ADPD188BI in Figure 4). The analog front end (AFE) includes the transimpedance amplifier, band-pass filter, integrator, ADC, LED drivers, and digital control (see the middle of Figure 3). There are many options inside the AFE to enable optimization for different applications and use cases. The AFE also provides the ability to reject ambient light, such as from lamps or solar radiation at levels up to 80 dB.

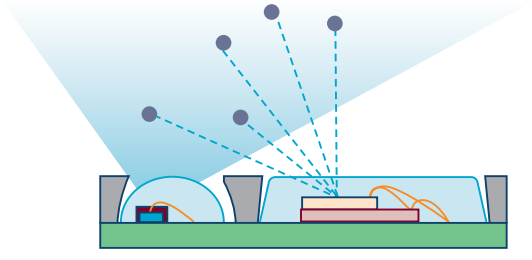
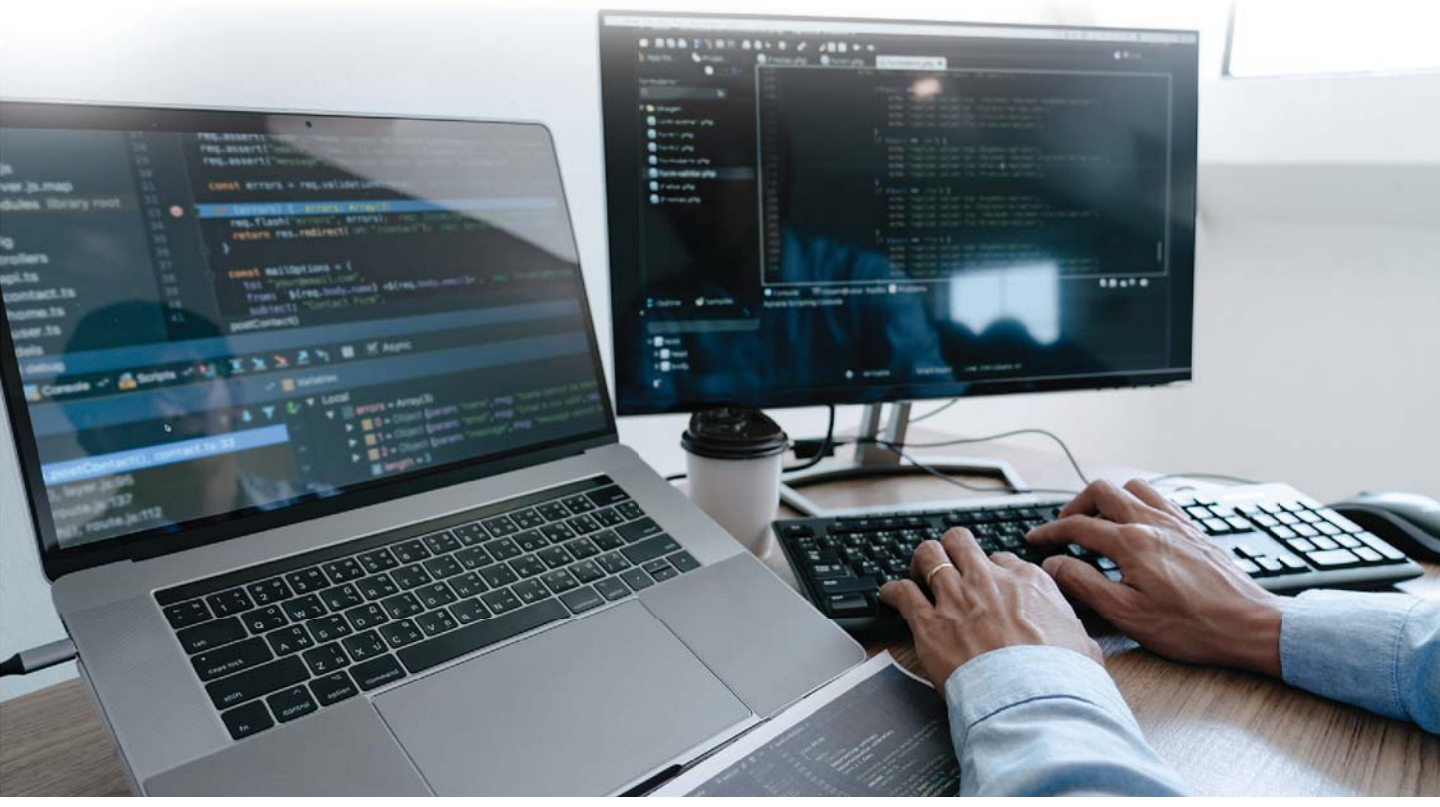


Figure 4. Cross-section of the ADPD188BI.

This principle provides many significant advantages. The short distance between the LED and PD results in much more efficient use of light, which reduces the power dissipation of the system for a required sensitivity and results in a longer battery lifetime.

Two different LED colors are included in the device. The amount of light scattered by the particles is a function of wavelength. This provides limited use for separating the hamburger and flaming polyurethane smokes, but can be used to distinguish between relatively small smoke particles (100 nm to 300 nm diameter) and much larger smoldering plastic or steam particles (10 µm diameter). The highly configurable AFE inside the ADPD188BI provides for a very high dynamic range that is software configurable and can be adjusted on the fly. The SNR of the system can also be easily adjusted to optimize for power or performance on the fly. One example would be to dynamically increase the sample rate or SNR when smoke has been detected to more accurately differentiate between a nuisance source or a real fire. The integration of the system also enables Analog Devices to calibrate the loop response (LED driver → LED → PD → AFE) of the parts and burn calibration coefficients into the AFE to limit the part-to-part variation to better than ±10%, which reduces or eliminates the need for expensive and time-consuming sensitivity calibrations in smoke tunnels. The ADPD188BI provides the features and capabilities to increase the performance of smoke detectors to more accurately separate nuisance sources from fire events. As well as performance, there are further advantages: the integration of LEDs means separate LED sourcing and stocking is eliminated, and the small form factor allows for integrated smoke detection across intelligent building components.

ANDERS HOLMBERG | General Manager, Embedded Development Tools, IAR Systems



Tracing and Debugging to Increase Efficiency

Debugging embedded software can be a time-consuming activity, from chasing down a specific bug to standard project activity. This article will cover tactics that might not eliminate all the hassles of debugging, but at least minimize the magical part.

It's no surprise to anyone that newly written software is rarely 100% bug-free. However, steps can be taken in the beginning to help reduce the number of issues that may appear in your code. Or, to put it another way, to make it so you'll have less debugging to do. The clear starting point is laying down some basic rules for code hygiene:

- Use a coding standard like MISRA and CERT C. Striving for MISRA compliance helps avoid quite a few of the pitfalls inherent in C and C++. CERT C can add a security perspective to the list of things to avoid. The first corollary is to pay close attention to your compiler warnings. The second corollary is to use an automated static checker to check your compliance.
- Use your own, or someone else's, hardware abstraction layer. Avoid inlining code that directly manipulates hardware in your code. For example, if you need to start a timer, call a HAL function to set up and start the timer instead of directly manipulating the timer registers. You'll realize numerous advantages by heeding this advice, one of which is almost completely avoiding copy/paste errors and typo mistakes when dealing with several different timer invocations in different places of the code base. Moreover, a compiler can often do a better job at optimizing, and might even inline the code, so you

gain in both performance and code size. A corollary to this advice is to keep the individual HAL functions as small as possible—avoid creating "Swiss Army Knives" (large functions with many responsibilities). Not only are small, single-purpose functions easier to understand and maintain, they're also often easier for a compiler to optimize really well, which may seem a bit counterintuitive.

- Give a little extra thought to how you use memory. For example, do you really need dynamic memory management? Is the stack really a good place to store complex data structures? Standards for functional safety and high-integrity software often advise strongly against dynamic memory management and storing complex or large data structures on the stack, and these are for good reasons.
- If your toolchain supports worst-case stack depth analysis, the investment to read up on and use that functionality will pay off quickly.

TO printf OR NOT TO printf SHOULD NOT BE THE QUESTION

One of first things to realize (or remember) is that if you're developing and debugging embedded software, you're very likely to do so in an environment where executing code on the target is done through a debugger. For example, if you're working in an IDE, the easiest way to execute your program is by firing up the debugger.

Let's get down to the nitty gritty and look at the power of breakpoints. But first, let's throw some shade at the venerable printf as a debugging tool.

The most important reason to not use printf is that adding printf statements in your code can dramatically affect how your code is compiled. Not only is the printf a function call, but the arguments to the call will have to be accounted for. This means that stack and register usage will look completely different and many compiler optimizations will not be performed, especially if the statement is located in a tight loop.

Such a scenario can have unpredictable consequences if your code is complex or relies on C/C++ behavior that's implementation-defined or even undefined by the C/C++ standards. What might happen is that your code behaves perfectly well when adding the printf to the code, but breaks when you remove the printout, or vice versa? By the way, this is a very good reason to strive for MISRA compliance.

Another good reason is that printf is a weak tool as it can only display data. A third reason is that to change the behavior of the printout or add more printing statements, you need to rebuild the

application and download it to the target again. Finally, at some point you will have to go through the code base and remove all of the statements you added, even if they're all guarded with #ifdefs.

THE POWER OF BREAKPOINTS

So, let's take a break from the preaching and look at the different types of breakpoints available. A breakpoint, in its simplest form, is a stop sign at a particular source statement that ensures execution breaks unconditionally when reaching the right spot. A good debugger will let you examine the content of variables, registers, and the call stack as well as memory in general. Such a code breakpoint is very useful in and of itself, but it can also be associated with an expression whose truth value determines if execution stops or not.

This lets you focus on the interesting cases instead of examining the interesting variables every time execution passes through the breakpoint location. For example, if you want a closer look at what's going on in a specific range of value in a loop index variable, you can set up the expression to stop only when the index is in that range rather than stopping each time you hit that code. Of course, you can also construct more complex stop expressions based on any variables that are in scope.

Sometimes you really need to see the value of one or more expressions. This can be done easily by using a log breakpoint—a breakpoint whose only purpose is to print a message in the debug log window without stopping execution. It's essentially a debugger-supplied printf that can be combined with a Boolean expression to determine whether or not the message should be generated.

A very powerful type of breakpoint is the data breakpoint. This triggers a specific variable or a memory location is accessed. It can be extremely helpful if you're trying to figure out why data values in a specific location aren't what's expected.

Why would this be needed? There can be several reasons, but one of the sources for such issues is pointers. If you use pointers, there's a fair chance that at some point you will get some pointer arithmetic wrong. While reading from or writing to the wrong address might not make the program fail, it can produce very strange results. These kinds of issues can be very tricky to debug, as the actual bug and the place where you experience the effect often aren't related in any way.

Combining data breakpoints (or any type of breakpoint, for that matter) with the call stack window can be very revealing *(Fig. 1 on page 20)*. The call stack window will show you where you came from. It also gives you the opportunity to move up and down the call chain and examine parameter values.

Some of these types of breakpoints may not always be available, depending on the exact device running the program, and/or the specific debug probe.

Some targets support live reading of memory, so that the debugger can continuously display variable values and other information during execution with a standard debug probe.

A PATH TO ENLIGHTENMENT

If you can stand a few extra buzzwords and adjectives, let's talk about a debugging tool that's truly amazing. Trace is a way to record the execution and other types of data flow on your device, like interrupt information and other hardware events *(Fig. 2)*. For example, viewing combined event data in a timeline can be very revealing about how a system behaves: Are your interrupts firing when they should, and how does it correlate with other activity?

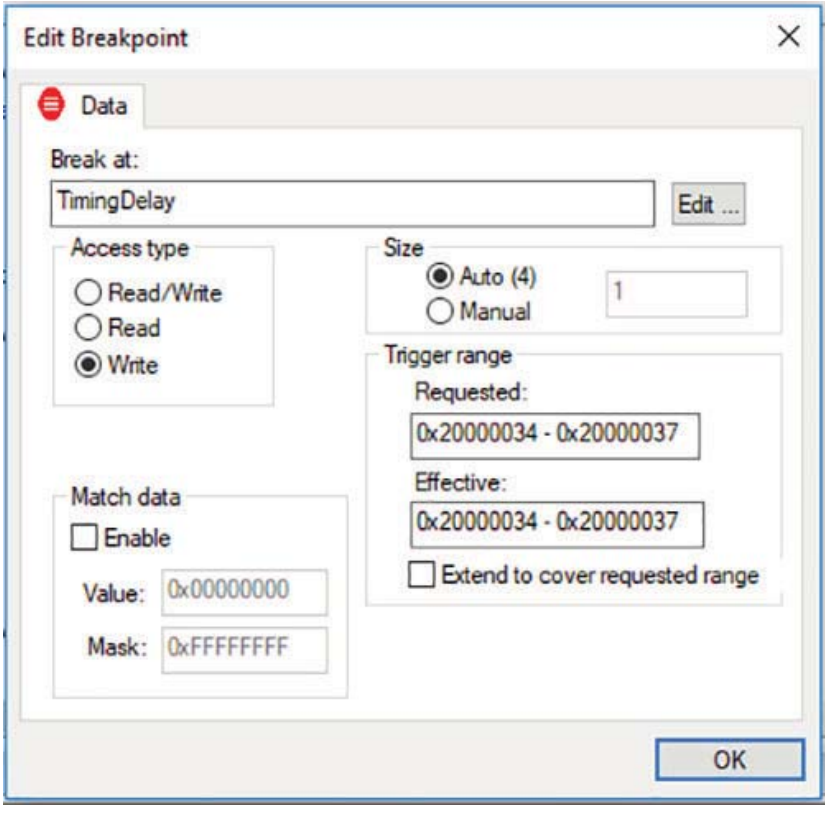
What makes trace more complex than regular debugging is that there are many different types of trace technologies, and different ways to access the trace data. In addition, you may need a trace-enabled probe. To utilize the power of trace in the best way means thinking about what you need to do to use it at the beginning of your project:

- One thing to consider is the choice of device. Does it have trace functionality, and if so, what kind? Is the device available in versions with and without trace? If so, you may build development versions of your board with trace and go to production without it to keep cost down.
- Trace can also be an enabler for profiling and code coverage data, so thinking upfront about your needs in that area can be beneficial.

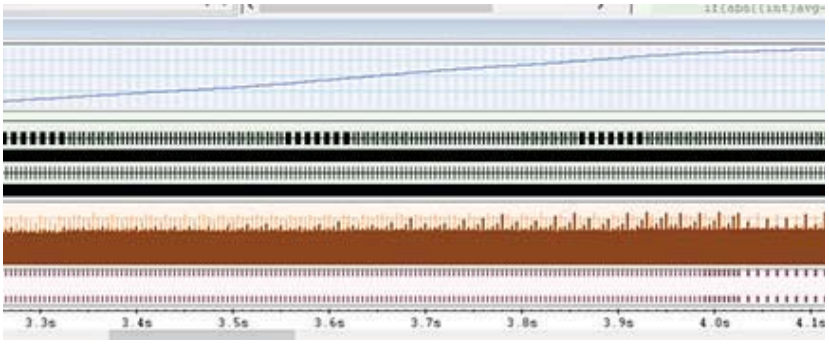
High-quality trace tools are designed to take away the pain of trace complexity and use all available trace information, but you still need to understand your needs on the hardware side. However, investing some time and resources upfront in trace as a debug and code-quality tool will pay off when you hit that first tricky issue.

THE PATH TO INCREASED EFFICIENCY

Some of the topics in this article might seem borderline trivial, but the best solutions to tricky problems often fall into this category. Finding the root cause of a software problem can take days or even weeks, or it may be a quick and easy process. One way to reduce the difficulty is to spend a moment thinking about how to best use your knowledge of the code base in combination with the features of your debugger and trace tools, rather than always reaching for a printf statement. Over time, this way of working will boost productivity and efficiency, not to mention peace of mind.



1. Edit Breakpoint. Combining data breakpoints (or any type of breakpoint, for that matter) with the call stack window can be very revealing.



2. Trace is a way to record the execution and other types of data flow on your device, like interrupt information and other hardware events.

WILLIAM G. WONG | Senior Content Director

Linux and Security for Today's Embedded Medical Devices

Discover how embedded developers can ease development of advanced Linux-based medical devices for reliability and safety.

Electronics and software going into medical devices has become increasingly more sophisticated. Platforms utilizing embedded Linux are also common these days. Likewise, safety and security remain paramount for medical devices.

Electronic Design's Bill Wong talked with Scot Morrison, Platform Business Unit G.M. of Mentor's Embedded Systems Division, a Siemens Business, about medical device development using Linux, managing security and safety, to ensure product performance success.

Scot, how can Linux open-source software be used for medical device safety?

Linux has been deployed safely in a wide variety of medical devices, but to use Linux in a medical device that has a safety requirement, embedded developers need to follow the process defined by the certification standard for compliance and certification.

So, can Linux be pre-certified for use in a medical device?

Not really. Certain real-time operating systems (RTOSs) such as the Nucleus RTOS from Mentor, can be acquired pre-certified, as can other embedded software components from a number of vendors. To achieve this kind of pre-certification, the vendor must be able to show that the complete software development process—requirements, design, development, testing and verification, and all of the steps of development—has been performed to medical industry standards such as ISO 13485 and/or IEC 62304. Linux and other open-source components aren't developed to these standards, so they're not pre-certified.

There have been efforts to show conformance of Linux to the over-arching concepts of functional safety, like mapping to IEC 61508, from which many industry standards are derived, including IEC 62304. While this approach isn't successful, Project ELISA, a current process, is showing promise by improving the processes for open-source software development, and in mapping the higher-quality output to these standards. However, this promise is likely years away from being completely realized.

What are embedded developers relying on now to ensure the safety of their Linux-based medical devices?

Instead of pre-certification, Linux is generally handled using a concept from IEC 62304 called Software of Unknown Providence, or SOUP, for today's medical devices. Under these guidelines, Linux is considered as part of the risk assessment of the overall device, and potential failures of Linux as used in the device must be considered, and mitigated, if they might cause harm to a patient. This risk assessment must meet the requirements of the FDA's pre- and post-submission guidance.

So, on the front end, it requires considerations on the use of Linux in the design, implementation, testing, and verification of the device. Then, the use of Linux, and all open-source software, must consider the possibility that issues will be found after product release. Certifiers are taking a very close look at this aspect of open source, especially as far as security issues are concerned.

Both safety and security are necessary when we're looking at medical devices. We hear that you can't have safety without security, but why is that?

Security is something that can be looked at as standalone. Even in medical devices, not all aspects of security are tied to safety. For example, when we talk about protecting someone's personal information, this is an aspect of security that doesn't overlap with safety. But when we talk about safety, things that could go wrong will impact the patient's health. If the device isn't secure, it makes it possible for bad actors to make these negative impacts happen either accidentally—or purposefully.

Does the use of Linux and other open-source software help protect these devices?

Linux is the most heavily used operating system for devices with a large, worldwide developer base. This global developer community focuses on ensuring that Linux works as expected in all conditions, and that it's as secure as possible.

As the most studied operating system in the world, the vast majority of developers are conscientiously working on improving Linux and other open-source

packages. But there's also a small number of actors who are looking at ways to break into Linux for their own purposes. The security of applications using open source is a constant tug-of-war between these opposing forces. Without security, you can't have safety.

How is it possible that Linux can have so many security flaws that we're always finding more?

Linux and other major open-source packages like OpenSSL or SQLite are large packages that can have unpredictable interactions with other software running in the same system. This is combined with the fact that many flaws are hard to find in code reviews, normal testing, or by static analysis. And they're undetectable unless software is combined with task switching and interprocess communication. Best practices will not identify every possible flaw or exploit, and much of the open-source software that we rely on was not originally developed with today's best practices in place.

However, the most important pieces of open-source software used in devices worldwide are much more stable and secure now, compared to five years ago. This is mainly due to the hard work and diligence of engineers all over the world in identifying avenues of exploitation, fixing those when they find them, and with the worldwide community looking for similar issues in their own projects. The work will never be complete, but it's becoming harder and harder to find exploitable flaws in this important infrastructure software.

What happens when a security issue is found in Linux?

Security issues in Linux, including other important software like OpenSSL, are found by engineers either by happenstance, like a bug that they uncover during a project, or through concerted efforts to find exploits, like "white hat" hacking. Or, an exploit will be found during a post-mortem analysis of an attack, but that's uncommon.

In either case, the exploit discoverer will notify the community of the offending open-source component. Then, the discoverer or Linux community member will notify the Common Vulnerability and Exposures (CVE) group, run by MITRE, an organization closely related to the U.S. National Vulnerability Database (NVD) that's managed by the National Institute of Standards and Technology (NIST).

Once a vulnerability is understood and a fix is available, the CVE is publicized by inclusion in these lists. If the exploit is sufficiently serious, the issue is discussed by the security community worldwide. This is the point where devices are potentially most vulnerable. Since most vulnerabilities are found by the "good guys," the bad actors will find out about them as will the rest of the world. These bad actors can then deploy exploitations that take advantage of the newly found vulnerability.

That said, this publicity is very important, since it alerts the worldwide community of both the issue and the fix. Thus, an organization can determine if a particular exploit might affect their devices, and if it is, they can mitigate the issue before it may be attacked. Of course, not everybody will be able to update their devices, which will leave them open to attacks. But since there are no real secrets in the world, this openness prevents more issues than it causes.

Back to safety. How safe is Linux?

An operating system like Linux doesn't directly do anything to make a device safer. The operating system doesn't prevent a failure from occurring, nor does it make the system recover when a failure occurs. When you put Linux on a system with no other application and turn it on, Linux boots; however, it just sits there at a login prompt. It's not doing anything until applications that leverage Linux are running, and it's those applications that contribute to the

overall safety of the device. While an operating system isn't a safety mechanism, it enables them and is considered to be a safety element.

With today's advanced and affordable microprocessors, how does a multiprocessor system affect safety?

Today's microprocessors are powerful and complex, designed to support heterogeneous multiprocessing. They comprise powerful general-purpose cores to run an OS like Linux, and more specialized cores to handle other functions. Designing for safety is an integrated systems issue, not just hardware or software.

To take full advantage of the board BOM costs and higher integration of components in an advanced multiprocessor for a safety-sensitive design, applications must be kept separate—what's known as mixed-safety criticality.

Simultaneously, the safety-critical portion of the system runs on a separate cluster dedicated to real-time processing. It has features like tightly coupled data and instruction memory with extremely low fetch cycles, and highly deterministic performance, or lockstep mode for error detection.

Advanced multiprocessing systems contain hardware-enforced isolation that keep the application world and the safety-critical world separate. However, the software designer must use middleware such as the Mentor Hypervisor or Mentor Multicore Framework to take advantage of those hardware features. These software packages enable important system-level functions like secure Inter-processor Communication (IPC) between the processor clusters to be possible.

Scot, thank you. Where can our readers learn more about Linux and Mentor's embedded software?

Our website www.mentor.com/embedded-software/ provides a broad range of white papers and on-demand webinars on topics such as Linux, mixed-criticality, safety, and security to enhance embedded development.

SCOT MORRISON is the general manager of the platform business unit, Mentor Graphics Embedded Systems Division, overseeing the Linux, Nucleus, and AUTOSAR product lines, middleware, and professional services. Prior to joining Mentor Graphics in 2012, Morrison served as GM and SVP of Products at Wind River Systems Inc., where he previously served as VP of Engineering. He joined ISI in 1986, where he spent 14 years in various management positions, last serving as a VP and GM of the design automation solutions business unit in 1999, responsible for operating systems and associated middleware and tools.

BOB ZOLLO | Solution Architect for Battery Testing, Electronic Industrial Solutions Group, Keysight Technologies

Preventing Li-Ion Thermal Runaway During Battery-Cell Test

Installing water and gas safety systems are common preventative measures in case of fire, but the best course of action is to take steps that will remove thermal runaway from ever becoming an issue.

Much research and development today is spent trying to make lithium-ion cells safer. Some chemistries are inherently safer than others. For example, lithium-iron-phosphate (LFP) cells use chemically stable LFP, which doesn't exhibit the energetic thermal runaway experienced by metal-oxide lithium-ion cells.

However, one must always be careful when handling and processing cells. When testing cells, the goal of the test may be to evaluate cell performance under stressful conditions, such as extremes of temperature or high charge/discharge rates. During such stressful tests, care must be taken to monitor for signs of the start of thermal runaway and then proceed with appropriate action.

WHAT IS THERMAL RUNAWAY?

First, what is thermal runaway? Thermal runaway in a cell occurs when the cell internal temperature gets high enough to ignite the electrolyte, which is an organic liquid. Once the electrolyte is ignited, the oxide material in the cathode will break down and release oxygen. Now, in the damaged cell you have fuel (liquid organic electrolyte) and oxygen (from the oxides in the cathode)—ingredients for a fire that can generate its own oxygen, making it extremely difficult to extinguish *(see figure on page 26)*.

For thermal runaway to start, enough heat must be generated in the cell to ignite the electrolyte. Normally, this happens because of damaged separator material. The thin poly-plastic material can be damaged by improper manufacturing or being punctured by metallic dendrite growth.

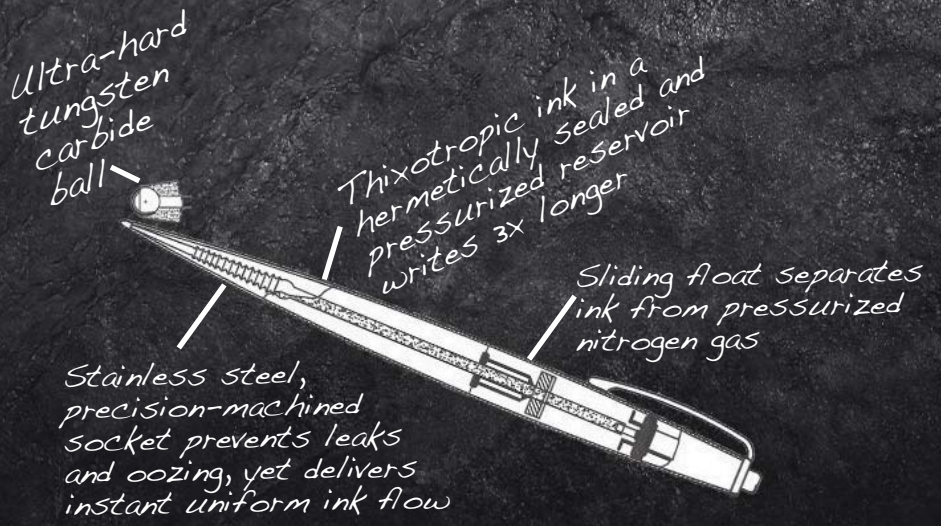
If the separator is punctured by a dendrite, the dendrite shorts the electrodes together. Current can freely flow as the anode and cathode come in contact through a short circuit where the damaged separator no longer keeps them apart. As current flows, heat is generated in the internal short and thermal runaway begins. More heat means more separator damage as the separator melts. More damage means more internal shorts, and more shorts mean more heat. Eventually, this heat can grow to the point where it ignites the electrolyte. The heat also causes the breakdown of the oxides in the cathode, which releases oxygen to feed the fire of the burning electrolyte.

ENDING THERMAL RUNAWAY

Safety systems can be installed in laboratories or in manufacturing, basically wherever there's the possibility for thermal runaway to start. Safety systems include gas/oxygen deprivation and water. Among the more common gas systems are carbon dioxide, argon, or nitrogen.

Lessons in Overdesign

NASA spent millions inventing a pen that would work in zero gravity.



Cosmonauts used a pencil.

Don't overdesign your battery.
Learn more about powering remote wireless devices with simply better batteries that are smaller, lighter, cheaper, and last longer. Visit us at www.tadiranbat.com

TADIRAN BATTERIES

* Tadiran LiSOCL2 batteries feature the lowest annual self-discharge rate of any competitive battery, less than 1% per year, enabling these batteries to operate over 40 years depending on device operating usage. However, this is not an expressed or implied warranty, as each application differs in terms of annual energy consumption and/or operating environment.

PROVEN

40

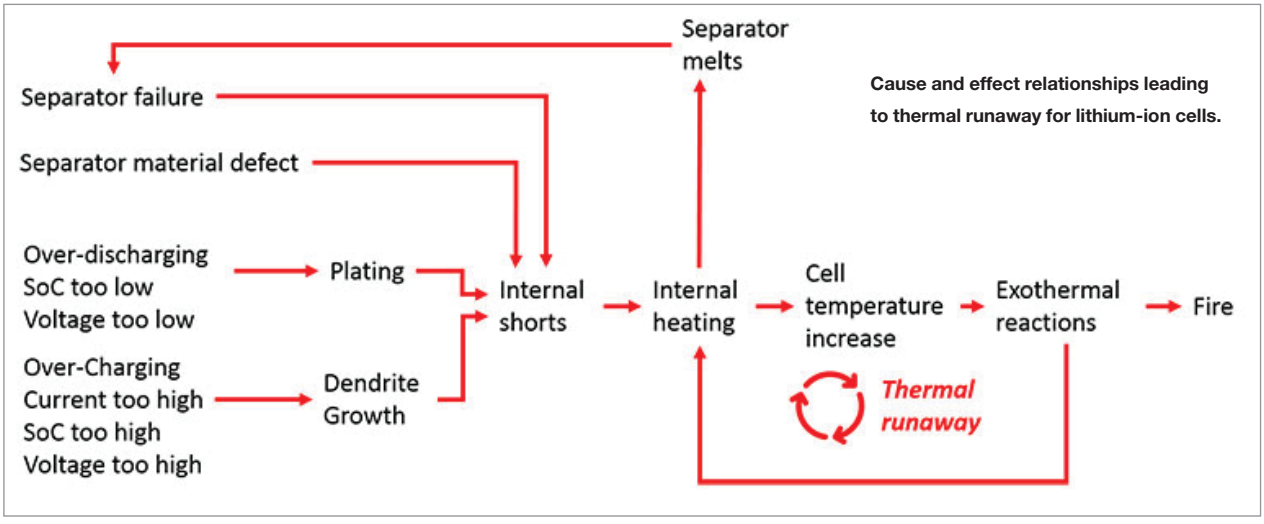
YEAR

BATTERY OPERATING

LIFE*

Tadiran Batteries
2001 Marcus Ave.
Suite 125E
Lake Success,
NY 11042
1-800-537-1368
516-621-4980

www.tadiranbat.com



Cause and effect relationships leading to thermal runaway for lithium-ion cells.

With these systems, the idea is to flood the area with gas that doesn't contain oxygen and, therefore, makes it difficult for anything to burn. The problem is that since the damaged cell will generate its own oxygen, these systems will not put out the cell fire. The gas systems may prevent the spread of fire to nearby flammable materials (wires, shelves, plastic storage containers, etc.).

Water, on the other hand, works differently. Instead of trying to deny oxygen to the fire, water is applied to the cell to drop the cell temperature. The goal is to remove the heat to ensure electrolyte doesn't burn as fuel and further oxygen isn't released. Of course, a side benefit of water is that it will also stop nearby materials from burning. So, water is commonly used, even though it creates quite a mess as it's sprayed onto a burning test setup.

PREVENTING THERMAL RUNAWAY

The best way to manage thermal runaway is to prevent it from starting.

First, cells should be manufactured with high-quality materials and with carefully controlled assembly processes to avoid separator defects. Ensuring proper alignment of electrodes and separators will prevent shorts around the edges of the stack (in pouch cells) or jellyroll (in prismatic and cylindrical cells).

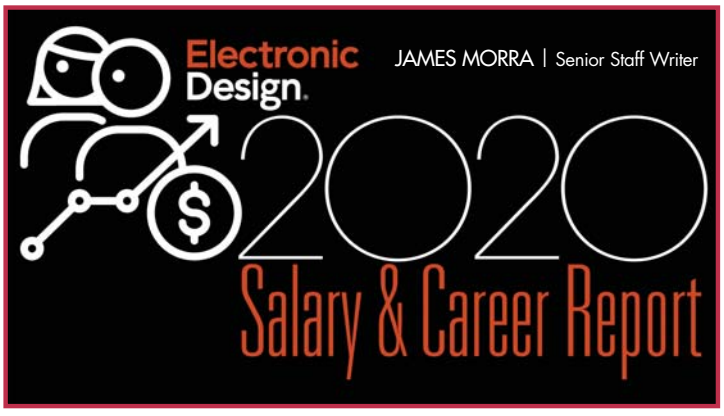
During charging or discharging, it's critical to ensure that cells aren't overcharged or over-discharged, which can cause plating and internal shorts. By continuously monitoring cell open-circuit voltage, this will ensure cells remain within their safe operating range. By continuously monitoring the amp-hours of energy put into the cell, it will ensure too much charge isn't pushed into the cell, which will cause overcharging.

These electrical limits are easily managed by properly designed cell test equipment that can immediately stop charging or discharging the cell if it detects electrical parameters that are incorrect. Such parameters include voltage too high, voltage too low, too much energy (amp-hours) pumped into the cell, or broken leads, which can cause measurement errors that obscure accurate detection of overcharging or over-discharging. For additional information about prevention of overcharging of cells, check out the article "Prevent Overcharging of Li-Ion Cells" on *www.electronicdesign.com*.

Also during charging or discharging, cell temperature should be monitored. It's to be expected that cell temperatures will rise during high-rate charging and discharging. When these high-rate charge/discharge steps are applied during elevated environmental temperature, it may be difficult to detect that the cell is getting too hot until it's too late. Setting a proper thermal threshold would allow an alarm signal to be tripped and appropriate action taken before the onset of thermal runaway.

Once the thermal limit is tripped, that appropriate action should include immediate termination of all electrical activity; i.e., immediately stop charging or discharging the cells. Even after applied current is halted, the temperature of the cells should continue being monitored. If temperature continues to rise, the cells should be moved immediately to a water tank to cool down the cells. Alternatively, a sprinkler could be activated to douse the cells to lower their temperature.

These actions will keep the cell heat low enough to prevent separator melting, electrolyte ignition, and oxygen release from the cathode's oxide coating. ed



Here's an overview of the state of salaries and overall compensation in engineering, as revealed by responses to our 2020 Salary & Career Report survey.

Engineering Salaries

As the world struggles to contain a rapidly-spreading virus that has ravaged the global economy, reordered everyday life for large parts of the population, and inflicted a heavy death toll, electrical and electronics engineers have had to tolerate more uncertainty than they may be used to.

Despite the global economic collapse caused by the coronavirus, most engineers are not seeing major changes in their compensation. According to a survey by *Electronic Design* and Endeavor's Design Engineering Group, most of the highly-skilled workers in engineering are as encouraged as ever about their jobs. The results also revealed that, unlike large swathes of the workforce, engineers are mostly feeling confident about their future prospects for compensation.

More than 1,100 engineers responded to the short-form survey, volunteering to share details about their overall compensation with *Electronic Design, Microwaves & RF*, and *Evaluation Engineering.*

There are challenges, however. Most of the respondents are struggling with the stresses of working remotely while also grappling with tightening deadlines, continuous education, and other issues that have lasted for years. Many are also working for employers that are delaying raises, reducing wages or canceling bonuses as the global economy sours. But for the most part, engineers are prospering.

The results revealed a median base salary of $100,000 to $124,999 for the engineering profession. The base salary is supplemented with a median bonus in the range of $1,000 to $1,999, according to the survey results. At the top of the pay scale, engineers in management and executive roles are padding their pay packages with thousands of dollars from share-matching programs and stocks.

Buoyed by the soaring demand for engineering talent, more than 65% feel that the potential for salary advancement is as favorable as any point in the last five years. Around 68% said that they are adequately compensated for what they do, and 70% said that they never consider leaving the profession to find another line of work.

While many of the respondents feel as though they should be making more money, around 92% said that they would recommend it as a career to younger generations. Even though many feel that they are under more pressure from management than ever before, the intellectual challenges of the job and other factors are lifting their morale.

There are many factors that impact engineering pay, including education, experience, location, title, age, company culture and profit as well as economic conditions and the status of the labor market.

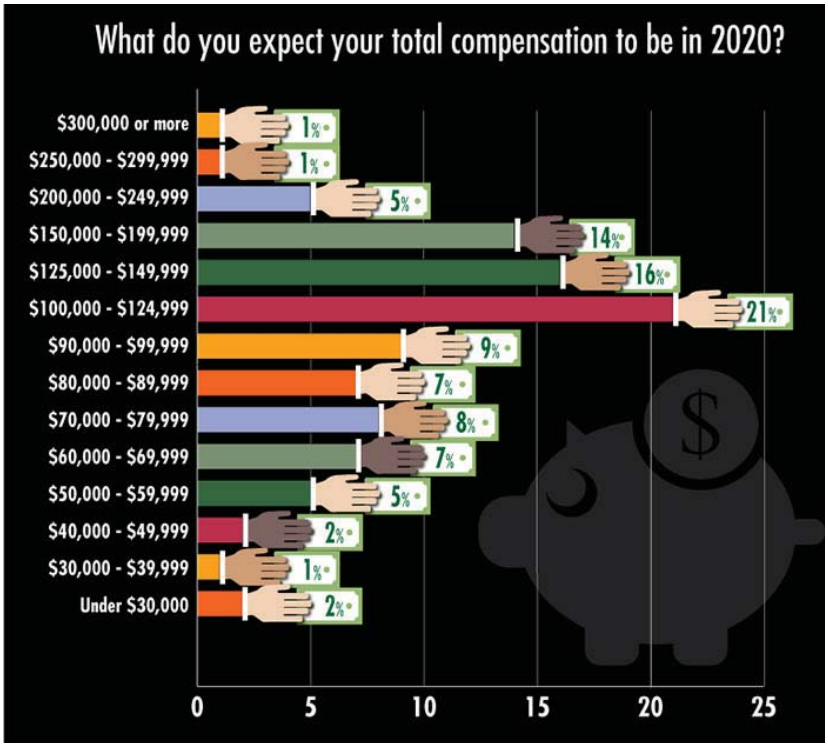
The virus, which has a death toll of more than 225,000 as of this writing in the United States, also remains unpredictable. Total cases of the virus could continue to increase around the world, prolonging the pain in the global economy, leading to permanent job cuts, and creating uncertainty. Or alternatively, the world could start to bring the virus under control, easing the economic woe. Around 30% of the respondents said that they have been stressed out about the state of the economy in 2020.

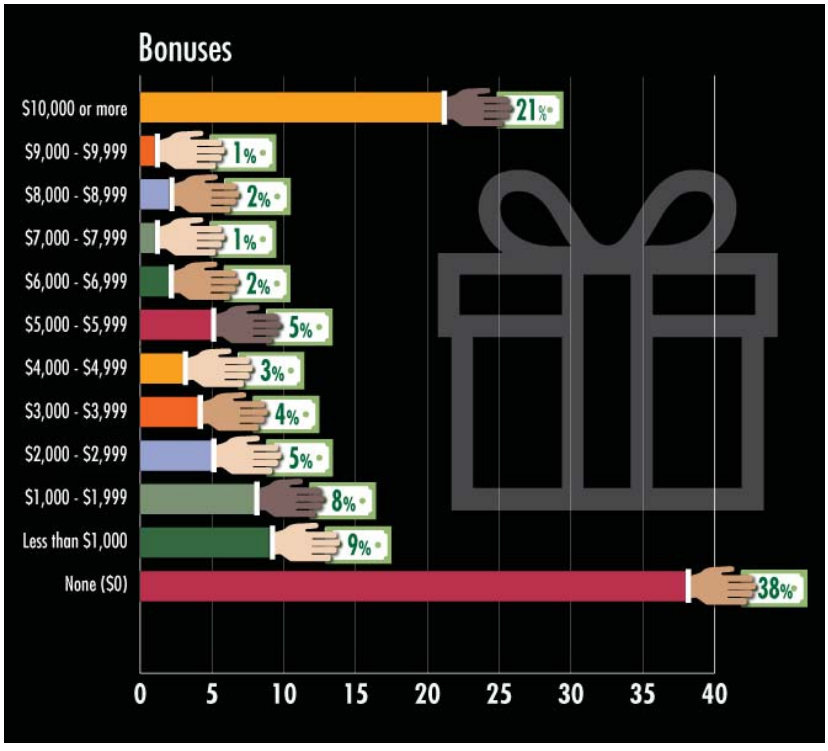
But many highly-skilled workers, including electrical and electronics engineers, are in a privileged position. While the economic fallout from the deadly virus has forced millions out of work in the US alone, paychecks have continued to come through for many of the engineers replying to the poll. While many respondents are struggling with burnout and family responsibilities, most are able to work remotely without many difficulties and are covered by company healthcare plans.

"Most of us can work from home without having results impacted too much," said one respondent. "The current situation shouldn't affect long-term perspectives on engineering as a career."

Still, the virus is hurting short-term perspectives on the profession. Over 26% of respondents said that their overall compensation will fall due to the economic fallout from Covid-19, the potentially lethal malady caused by the novel coronavirus. Around 68% said there would be no unforeseen changes to their overall rate of pay. Only 6% said that they expect an increase in their earnings.

Among the respondents seeing pay cuts, 50% said that they are seeing their base salary reduced, while 57% work at companies that have canceled bonuses. Another 60% are facing salary freezes. Out of all respondents, 29% said that their company has slashed jobs as a result of the continuing fallout. Only 18% said that their company has been forced to furlough employees due to the virus.





"For us, raises and promotions are going forward, although the money pool for raises has been cut so pay increases will not be as significant as they would've been without the economic pressures of Covid-19," warned one respondent. Another lamented that the fallout from the virus not only cut into orders from customers, but also closed production plants and worsened component shortages.

Still, over a quarter of the respondents revealed that their pay would increase by 1% to 3% this year, while another 15% said that their overall compensation package this year would be up 4% or more.

Roughly 36% of respondents said that their overall pay would remain the same, a major jump from previous years. In addition, slightly more than 5% of respondents said that wages would contract by 1% to 3% in 2020. According to the survey, about 15% revealed that their compensation would fall more than 3% compared to last year due to delayed bonuses, pay cuts, or other factors.

Many employers plan to pay out bonuses in spite of the economic slowdown. More than 32% of respondents are on pace to land more than $5,000 in bonuses in 2020. Another 20% said that they would earn bonuses of $1,000 to $4,999, while 10% would get up to $1,000. On the other hand, 38% are not bringing in any bonus pay in 2020.

More than half pointed to personal performance as a key factor in calculating their bonus, while another 61% said that their company's finances factored into the final tally. For 34%, a company's profit-sharing policies calculate to their bonus. Around 12% said that patent grants contribute to their bonuses. Moreover, 12% said they tend to land bonuses after meeting major engineering deadlines.

For many design engineers, other perks, including travel allowances, have been paused. Some said that their company has paused 401(k) matching programs to reduce fixed costs during the crisis.

Even though many employers have paused hiring plans, many engineers are on the lookout for new jobs. Changing jobs can result in a raise that may take longer to get through an internal promotion. Around 10% are actively seeking a new position, while another 33% responded that, while they're not looking to change jobs, they would follow up if contacted personally with a promising job offer. About 25% said they would follow up if they heard about an interesting opportunity.

But as the virus continues to spread uncontrollably in many parts of the world, other respondents said that they feel secure in their current position and generally enjoy their jobs. Around 32% said that they have no plans to change jobs in the foreseeable future.

Many feel that moving to a management role or changing jobs are the only ways to guarantee wage growth in engineering. But at the same time, some respondents said that, given all the challenges of engineering, these salary gains are not as generous as they should be. "It always seemed to me that you have to leave real engineering and move to management to make more," said one respondent.

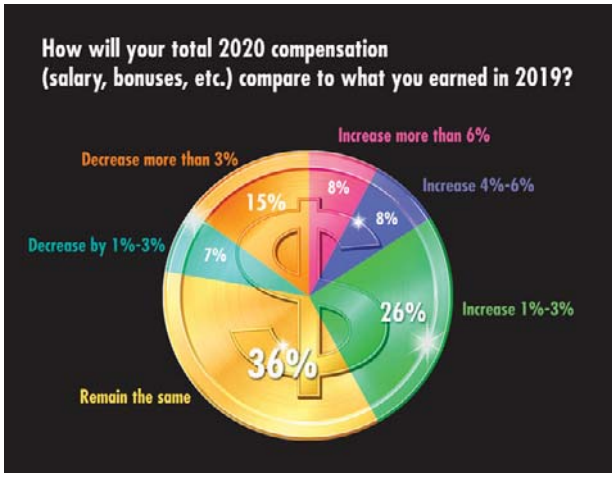
"Engineering salaries start higher and cap lower than many other professions," remarked one of the electronics engineers replying to the survey. "As a career, engineering is in a saturation period, where only one sector of the profession—software—is showered with salary advancement and higher than average pay, even in entry-level jobs," another worker said.

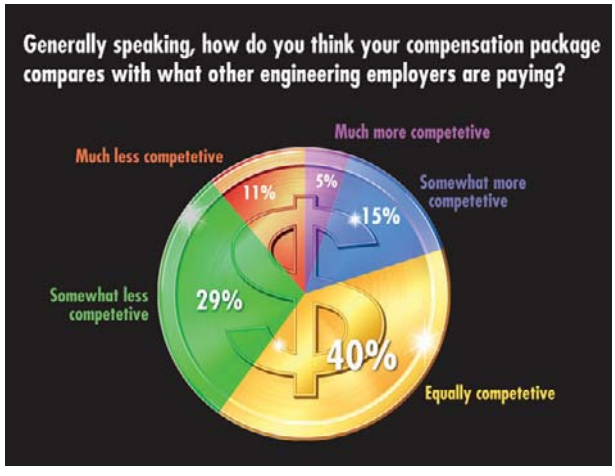
Last year, many companies continued to raise salaries and roll out better bonuses to keep employees happy and lure out engineers for hard-to-fill positions. Many firms also shelled out for continuing education, as the tight market for engineering jobs—at least in the US—underscored the need to nurture high-end skills internally. But according to the survey results, many of these perks have been scaled back or halted as the virus continues to hammer the global economy.

Other factors could hurt wages in the short term. Employers may not be in a position to promote workers in a recession, and engineers debating a job change think twice in a struggling economy.

Overall, engineers are also incurring indirect costs. Only 20% said that their company reimburses employees working remotely who require upgrades in internet speed or a change in their service provider. On the other hand, some respondents said that mandates on remote work have helped save on commuting costs, including car services, tolls, and parking. There are pros and cons.

More than 68% of respondents reported feeling content with their 2020 compensation package. But among engineers that feel undervalued, more than 50% said that they deserve a wage hike of 10% to 25%. Roughly 20% of respondents who feel undercompensated said that their compensation package is up to 10% out of line with what they deserve, the survey showed.





Engineering talent continues to be highly coveted. More than half of the respondents said that their company is struggling to locate candidates for open positions, and that could protect them against pay cuts. Only 30% said that their company offers signing bonuses to lure out candidates, and 16% have canceled signing bonuses within the last couple of years.

For many employers struggling with the skills shortage, the virus is complicating the calculus. Only 28% of the study's respondents said that their company plans to create more engineering jobs in 2020, a steep decline from previous years. More than 60% replied that they would maintain current staffing levels, closing open positions or delaying new hires to ride out the economic slowdown.

Only around 12% of respondents said that their company plans to reduce headcount by buying out employees who agree to leave their jobs or resorting to involuntary job cuts to reduce fixed costs. But given the challenges of scouting and hiring highly skilled engineers, it's clear that many companies are trying to hold onto the workers they already have. Around 69% of the respondents feel that their company is as focused on employee retention in 2020 as it was last year.

Lab Bench
WILLIAM WONG | Senior Content Director
bwong@endeavorb2b.com

RISC-V Lands on Mini-ITX Motherboard

SiFive put one of its RISC-V designs on a Mini-ITX motherboard, which features the company's FU740 SoC that's built around four U74 cores and one S7 core in a SMP configuration.

One of the popular compact motherboard form factors is the Mini-ITX. Hundreds of cases support the form factor, and most are likely to hold x86-based motherboards. There are numerous Arm motherboards as well, but, unlike the x86 motherboards, they typically require a customized operating system due to less standardization in this space.

Linux tends to be the norm for Arm-based solutions, and there's often an array of Linux distributions to choose from. Low-power solutions are typically the reason why designers opt for an Arm-based platform. However, it's often the functionality of the system-on-chip (SoC) at the core of the platform that makes the difference.

One platform that's presenting a challenge to Arm is RISC-V. Like most Arm solutions, RISC-V platforms are licensed from a standard platform such as an Arm Cortex-A76. The big difference is that RISC-V is a really just an instruction-set specification that's open, so anyone can create an implementation. And a number of companies are going in that direction, including SiFive.

Like Nvidia's Arm, SiFive is in the business of selling CPU IP. Much of this winds up in custom SoC solutions, but that makes it hard for developers to check out the platform or deploy it without having their own chip made, which isn't an inexpensive exercise. The plethora of x86 and Arm Cortex-M and Cortex-A platforms highlights how important real hardware is to the growth of an echo system.

SiFive's HiFive Unmatched Mini-ITX motherboard features an FU740 SoC with four SiFive U74 cores and one SiFive S7 core.

SiFive has actually been involved in the hardware business, starting with the 32-bit RV32IMAC Freedom E310. The RV32IMAC designation is an abbreviation for the standard RISC-V features, including 32-bit support (RV32), integer support (I), hardware integer multiplication and division (M), atomic real-time instructions (A), and support for the 32-bit and compact (C) 16-bit instruction set. The chip has 16 32-bit registers and no hardware stack. As with many RISC systems, it uses a jump-and-link (JAL) instruction to save a return address in a register.

The E310 would be on par with an Arm Cortex-M solution, which has a memory protection unit (MPU) but not a memory management unit (MMU) that's found in Arm Cortex-A platforms that can run operating systems like Linux. RISC-V is an extensible instruction set that supports MMUs, virtualization, and 64-bit platforms among other features. SiFive and other RISC-V IP vendors have been delivering 64-bit, MMU RISC-V solutions for a while, and they have found their way into numerous SoC solutions, but, again, typically for a captive audience.

RISC-V MINI-ITX MOTHERBOARD

That long-winded introduction is to highlight the significance of SiFive's latest announcement for the embedded space. I expect to eventually see more standard form-factor solutions like this in the future—the rise of RISC-V and this board has occurred a few years from RISC-V really taking off, which is impressive.

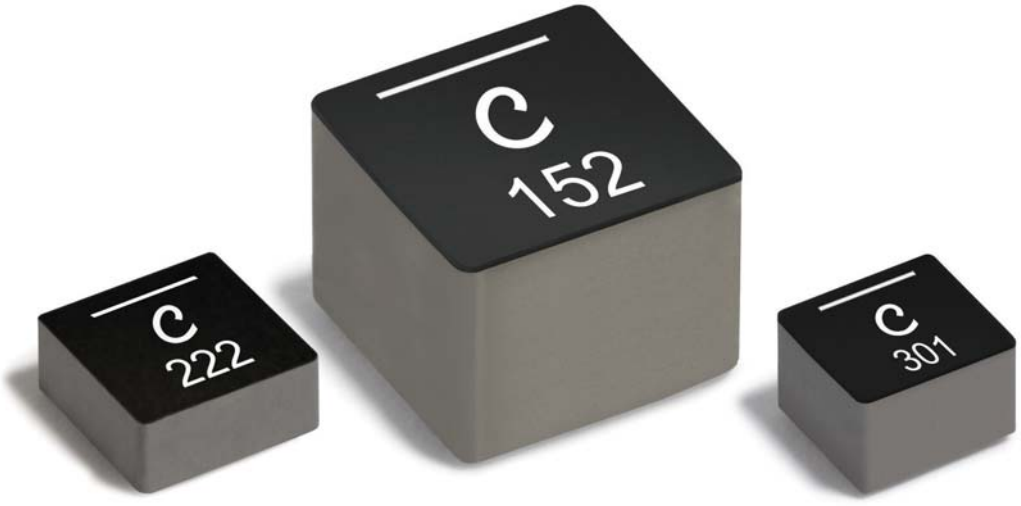
SiFive's HiFive Unmatched Mini-ITX motherboard holds the company's FU740 SoC *(see figure)*. The FU740 is built around four SiFive U74 cores and one SiFive S7 core in a SMP configuration like Arm's big.LITTLE approach, which is now Arm's DynamiQ architecture. The main shared memory is 8 GB of DDR4. It also incorporates a 32-MB QSPI flash and microSD card for booting the system.

As with most Mini-ITX boards, there's a single PCIe slot with a x8 PCIe Gen 3 interface. Furthermore, two M.2 slots are in the mix: one is a shorter version suitable for interface cards, such as wireless support, while the longer version can handle everything up to a PCIe Gen 3 x4 M.2 device, such as an NVMe SSD. Four USB 3.2 Gen 1 Type A ports are available on the rear panel along with a micro-USB Type B and a 1-Gb Ethernet port.

The board uses a standard ATX power-supply connector, facilitating its use with standard Mini-ITX cases. Board pricing starts at $665, including a microSD card with a custom version of Linux. ed

Advertisers Index

The (lowest) Loss Leaders



XGL Family power inductors feature the industry's lowest DC resistance and extremely low AC losses for a wide range of DC-DC converters

Coilcraft's XGL Family of molded power inductors are available in a wide range of inductance values (from 100 nH to 47.0 µH) and current ratings up to 43 Amps. With up to 45% lower DCR than previous-generation products, they are the most efficient power inductors available today!

Their ultra-low DCR and higher Irms also allow XGL Family inductors to operate much cooler than other components.

All XGL Family inductors are qualified to AEC-Q200 Grade 1 standards (with a maximum part temperature of 165°C) and have no thermal aging issues, making them ideal for automotive and other harsh environment applications.

Download datasheets and request free samples at **www.coilcraft.com/XGL.**

Coilcraft®

WWW.COILCRAFT.COM

ANALOG DEVICES

AHEAD OF WHAT'S POSSIBLE™

1500 V_{DC} Isolated Transceivers for Solar Energy and HV Batteries

ADM3058E

- Protocol options for CAN FD or RS-485.
- 5.7 kV reinforced *i*Coupler® digital isolation and IEC ESD for signal integrity in harsh environments.
- CISPR 32 Class B emissions on a 2-layer PCB reduces design complexity and time to market.

Buy Now at mou.sr/ADM3058Ex

MOUSER ELECTRONICS