www.eenewseurope.com

October 2020

EE News europe

electronics europe News

Quantum lift

european business press


Delivering Excellence
in 48 hours to most areas in Europe
FREE SHIPPING
ON ORDERS OVER € 50 OR $60 USD*
FIND CONTACT AND ORDERING INFORMATION FOR YOUR REGION AT
DIGIKEY.COM/EUROPE
Digi-Key ELECTRONICS
9.2 MILLION+ PRODUCTS ONLINE | 1,100+ INDUSTRY-LEADING SUPPLIERS | 100% FRANCHISED DISTRIBUTOR
*A shipping charge of € 18.00 will be billed on all orders of less than € 50.00. A shipping charge of $22.00 USD will be billed on all orders of less than $60.00 USD. All orders are shipped via UPS, Federal Express, or DHL for delivery within 1-3 days (dependent on final destination). No handling fees. All prices are in euro or United States dollar. Digi-Key is a franchised distributor for all supplier partners. New product added daily. Digi-Key and Digi-Key Electronics are registered trademarks of Digi-Key Electronics in the U.S. and other countries. © 2020 Digi-Key Electronics, 701 Brooks Ave. South, Thief River Falls, MN 56701, USA
ECIA MEMBER
Supporting The Authorized Channel

CONTENTS

Shockwaves of a deal

The prospect of the largest deal in the semiconductor industry has put the IP business under scrutiny. Nvidia's $40bn offer to buy ARM from Japanese tech conglomerate Softbank faces tremendous obstacles, not least as a pawn in the US-China trade war. The prospect of key IP resting with a large US competitor raises issues for China and for multiple global semiconductors.
These issues were not a surprise. ARM was in play for months pr ior to the deal, but efforts to form a consortium failed to materialise. The Nvidia deal will take 18 months to secure, if it happens at all. Qualcomm's $44bn bid back in 2018 to buy NXP failed to win approval in China and failed. The geopolitical forces at work are even more pronounced today.

The ARM ecosystem is a key part of the embedded, IoT and automotive markets, and gaining traction in the data centre and high performance computing. What happens with the deal, and afterwards, matters to the entire industry.

Nick Flaherty

Cover
A breakthrough in quantum communication technology by a group led by the University of Bristol is securing networks. Image credit ÖAW / Klaus Pichler / QET

World first with multiple Covid-19 test machine

By Nick Flaherty

Bosch has reduced the test time for its Vivalytic machine to 39 minutes with five Covid-19 tests simultaneously in a world first

Bosch has developed a new rapid test for its Vivalytic analysis device to detect Covid-19 with five simultaneous tests per cartridge. This helps tackle the major drawback of throughput and cartridge production for DNA test machines and is a world first, says the company.

The 39 minute test is currently the fastest polymerase chain reaction (PCR) test worldwide and is aimed at mobile test centres at motorway service stations or in airports to obtain a reliable result while at the testing site. The new machine is CE-approved and available in Europe now.

From early October, it will be possible to simultaneously evaluate five samples in one test cartridge and at a comparable speed, a world first according to the company. This will increase the throughput with fully automated processing of more than 160 samples a day. Optimized software to be rolled out in the next few weeks will further reduce the time to result for positive Covid-19 samples.

“One of the keys to fighting the coronavirus pandemic is to rapidly identify sources of infection. That’s why we focused on following up on our first coronavirus test with an even faster one,” says Dr. Volkmar Denner, chairman of the board of management of Robert Bosch. “This will now enable us to put people’s minds at ease even more quickly.”

The development of the new Bosch PCR singleplex test is part of a research and development project funded by the German Federal Ministry of Education and Research (BMBF).

“I believe it’s important that people have clarity about their state of health as quickly as possible. In this respect, insights from science and research can bring people huge benefits. Over the next few months, we will be confronted with the particular challenge of having to test more people,” said the German minister minister for education and research Anja Karliczek. “The improved testing procedure developed by Bosch with the BMBF’s support has the potential to be a tremendous help with this complex job. The rapid improvement of our technological capabilities shows what innovative achievements German companies can deliver in times of crisis.”
The test has a sensitivity of 98 percent and a specificity of 100 percent. To develop it, the Bosch subsidiary Bosch Healthcare Solutions joined forces with the German biotechnology company R-Biopharm – a leading provider of highly sensitive manual PCR tests. This replaces Randox Labs in northern Ireland that was the lab partner for the first version launched at the end of March after just six weeks’ development. This was a multiplex test that simultaneously checks samples for the SARS-CoV-2 virus and nine other respiratory diseases in two and a half hours.

This is comparable to the multiplex one hour test developed by DNAnudge in the UK thas has been shown a sensitivity of 94.4% accurate compared against standard NHS lab-based tests and a specificity of 100 percent to avoid false positives.

“The decentralisation of mass testing is now being recognised as the key to addressing this unprecedented public health crisis, and these results very clearly support the use of CovidNudge as a highly accurate, rapid and near-patient testing solution that can be delivered on-the-spot and at scale, in both clinical and non-clinical settings,” said Prof Chris Toumazou, CEO and co-founder of DnaNudge.

The new, accelerated test from Bosch is exclusively for Covid-19 to reduce the test time.

“With our different coronavirus tests and variable analysis strategies, we open up a range of testing scenarios with a Vivalytic device – from screening all the way to supporting differential diagnosis for diseases with similar symptoms,” said Marc Meier, president of Bosch Healthcare Solutions.

By the end of the year, Bosch wants to reach capacity for one million tests and says it is working closely with its suppliers to maximize capacity and further increase supply of both machines and cartridges using its existing manufacturing capabilities at Bosch Healthcare Solutions at Waiblingen in Germany.

Bosch
Bosch Healthcare Solutions

ARM sale to Nvidia agreed at $40 billion

By Peter Clarke

SoftBank Group Corp. has announced it plans to sell ARM Ltd. to GPU and AI technology vendor Nvidia Corp. for up to $40 billion.

Although many have spoken out against the deal on the grounds it could weaken ARM's independent IP licensing model, Nvidia has provided some assurances to try and dispel doubts.

Nvidia said it will continue ARM's open licensing model and customer neutrality while adding Nvidia technology into ARM's portfolio. In addition, Nvidia said it would retain the ARM brand and that ARM's intellectual property would continue to be registered in the UK.

Nvidia said ARM would continue to be headquartered in Cambridge, UK and that Nvidia would expand ARM's R&D presence there. The expansion would include the creation of an AI research and education center and the building of an ARM/Nvidia AI supercomputer for research purposes.

The terms have been agreed by the boards of SoftBank, Nvidia and ARM, but completion of the deal is dependent on regulatory approval from the UK, China, the European Union and the United Stares. The size and strategic significance of the deal means that Nvidia is forecasting it could take up to 18 months to complete.

The deal is being funded with shares and cash and as a result SoftBank will acquire a stake in Nvidia, which is expected to be under 10 percent. It also includes a $5 billion bonus for SoftBank if ARM achieves certain financial performance targets that would take the value up to $40 billion.

Jensen Huang, founder and CEO of NVIDIA, said: "AI is the most powerful technology force of our time and has launched a new wave of computing. In the years ahead, trillions of computers running AI will create a new internet-of-things that is thousands of times larger than today's internet-of-people. Our combination will create a company fabulously positioned for the age of AI."

Nvidia said that SoftBank and ARM remain committed to satisfying the undertakings made by SoftBank when it acquired Arm in 2016 and which are scheduled to complete in September 2021. In its statement Nvidia did not speak about what would happen after 2021 or any further time-bound undertakings.

The deal is constructed with $2 billion payable at signing and an eventual total of $12 billion to be paid in cash plus $21.5 billion in Nvidia common stock. Nvidia will also provide $1.5 billion in equity to ARM employees.

The deal does not include ARM's IoT services group, which was split away from the main company a few months ago.

Defending the Nvidia-ARM deal

By Peter Clarke

Jensen Huang, CEO of Nvidia, and Simon Segars, CEO of ARM, have been explaining the proposed combination of the two companies.

The $40bn deal, one of the largest in the semiconductor industry, is set to produce a major force in fabless semiconductors targeting artificial intelligence but has been attacked by observers for removing the independence of IP provider ARM and potentially impairing its business. By Nvidia's own estimation the deal could take as much as 18 months to clear regulatory hurdles that could be thrown in its path.

The conference calls painted a picture of a combined Nvidia-ARM that would be a 30-year champion for the era of artificial intelligence. And that ARM would continue to license technology out using an open business model. Indeed, Huang frequently said that one purpose of the deal was to give Nvidia's GPU and AI technology access to customers via the IP licensing business model.

First line of defence

Most analysts bowled the fairly straight ball at Huang that if the deal went through, Nvidia would be competing with its customers; that this could be bad for Nvidia's chip sales and for ARM's IP licensing.

Huang's first line of defence was that as it stands ARM and Nvidia barely compete with each other and so IP licensing would continue in an open and neutral manner. In 2019 Nvidia shipped 100 million chips mainly into the data center while ARM technology shipped in 22 billion chips across all embedded fields, he said. "We're [Nvidia is] in very few mobile communications. We're in very few embedded applications." He added: "We love the business model. It will stay open and fair and we'll offer even more IP," said Huang.

So, Huang's position appears to be different delivery mechanisms for different application sectors.

But Nvidia's adoption of the IP licensing model while also selling chips does not fundamentally address the objection of conflict of interest. It just adds Nvidia GPUs and AI cores to the conflict. One analyst picked up on that and used his question to probe further.

Huang's second line of defence was to say that Nvidia wel-

comes competition throughout the data center platform from chip to software. He said Nvidia sells chips, boards and software in any combination. Some customers take the chips and make their own boards or write their own software. Others take the software and make their own chips. Some take the platform.

In effect, Huang said Nvidia already competes with its customers everywhere from chip to board to software. Adding ARM will merely add competition at the IP level and is nothing to be perturbed about. "We will stay neutral. We are committed to the open business model for boards, chips and throughout the software stack," Huang said.

Three reasons

Huang said there are three reasons why Nvidia wants to buy ARM as soon as possible:

1) to license Nvidia's IP through the ARM ecosystem
2) To create a first-class data center computing platform based on the ARM architecture
3) Invent the future of cloud-to-edge computing

Segars appeared keener on the second two of the three points. "I truly believe AI is the defining technology of the future. We are just at the starting point. [The deal will provide] so much more weight and resource. We will also have more to go and sell, which is nice."

Segars also admitted he had spent some time speaking with customers about the deal but did not characterise their attitude as either opposition or acceptance. "Independence is part of our strength. He said the combination with Nvidia will bring more investment but there is the intent to maintain the independent business model.

A journalist pointed out that UK government had extracted time-based legally-binding conditions on ARM's current owner SoftBank Group Corp. when it acquired ARM for $31 billion back in 2016 and that these conditions were due to expire in 2021. Had Nvidia made similar undertakings?

"We're open-minded to talk to the UK government," said Huang before adding that the UK government should be very keen on the deal because of all the things Nvidia is promising to do in Cambridge. These include expanding ARM R&D and building an AI research center to house an Nvidia-ARM based supercomputer there.

"How that's documented, we're more than delighted to have that discussion. We've just started the conversation with the UK government. We have plenty of time to do it," said Huang.

It would appear Huang and Nvidia are not as keen on making legally-binding commitments as they are on making verbal promises, but that should not be a surprise.

Getting down to the nitty gritty an analyst asked: "Would a combined Nvidia-ARM continue to develop ARM's Mali graphics processor architecture and continue to use the RISC-V processor architecture?" Huang was quick with his answer; "Yes and yes."

He pointed out the Nvidia graphics and Mali graphics are aimed at different markets and they both serve them well. There was room for both to go on developing. On RISC-V he said: "We are enthusiastic users of both ARM and RISC-V. They are very different things." He said the ARM architecture and cores have the advantage of a rich ecosystem helping get products to market, legacy software and lots of developers working on applications and software – but not all use cases need that. "We use RISC-V internally all over our large chips and we will continue to do that," said Huang.

Several analysts and journalists alighted on the issue of regulatory approval.

Huang's first point was that for the purposes of export controls the country of origin of the technology is the relevant factor, not the legal location of the owner. He said ARM's intellectual property would continue to be mainly registered in the UK and so the transfer of ownership from a Japanese company to a US company would make no difference.

Callers pressed Huang as to what China's regulatory response to the proposed take over is likely to be. "We've been through the regulatory process in China with Mellanox," said Huang. He added that the China regulator is pro-competition and pro-choice for the consumer. "They will love it," he said. This response does not explain why Chinese regulators did not approve Qualcomm's proposed acquisition of NXP Semiconductor, which it appears may have had geopolitical overtones. But Huang has his own experience and an undeniable enthusiasm for the deal.

If enthusiasm were enough the deal would be done in a lot less than 18 months.

Why the Tesla Battery Day matters for Europe

By Nick Flaherty

Tesla has re-engineered its battery technology and processes from scratch to be ready to scale production to TWh at sites including Berlin.

The financial markets were distinctly underwhelmed by the Tesla Battery day. After all, it was a tutorial on how to build a battery for an electric vehicle, and that was the point.

Current battery cell manufacturing is struggling to scale – which is why there are huge investments around the world on building more battery gigafactories. Even Tesla's own 150GWh Gigafactory in Nevada can't scale, says Elon Musk, Tesla's CEO (above).

To reach the 120TWh energy requirements in a future of electric vehicles requires a fundamental re-evaluation of the production process, he says. "The goal is TWh, tera is the new giga ," he said. "We need 100x growth in battery production to transition the global fleet of vehicles to electric."

This has implications for Europe, which has also been recognised by politicians in the EU.

"For a company that's growing rapidly its important to reduce the supply chain, having the parts move very quickly though the factory and then ship to the customer is vital. That's why its important to have a factory in each continent," he said, highlighting the new factory being built in Berlin, Germany, alongside plants in Fremont, Nevada and Shanghai.

"Berlin is making rapid progress, and the Model Y made in Berlin will be more efficient [than the ones currently built in Fremont, California]," he said. That global manufacturing footprint means that Berlin will also add battery cell manufacturing. "We will be manufacturing cells in Berlin," he said.

The design of the battery cell design has been tweaked and changed to support new high volume manufacturing processes that are seven times more efficient than today.

This re-engineering drives up efficiencies that are the equivalent of building seven new gigafactories, and highlights Tesla's

strategy. While it continues to work with the large battery cell supplier, it is optimising its technologies and processes to provide the additional volume itself.

"This is a 75% reduction in investment per GWh," said Musk. "We are able to get in a smaller formfactor than Giga Nevada many times the cell output – 1TWh in less space than 150GWh. This will be production of 100GWh/yr by 2022 a3TWh/yr by 2030 for internal cell production supplemental to what we buy from our suppliers, Panasonic, CATL and LG Chem. We will continue to use them as suppliers."

But it is not easy to achieve this scale, and Tesla can only do this if it controls the entire production chain, from sourcing the materials to building the cells and the end vehicle.

Tesla uses a German subsidiary to provide the manufacturing equipment. Tesla Grohmann Automation , headquartered in Prüm near the border with Belgium, has also worked with BMW and Daimler but is now a dedicated engineering unit building automation systems for Tesla globally.

"The difficulty of designing the machine that makes the machine is immense compared to the vehicle itself," said Musk. "It's at least 10 to 100 times harder to build the factory than the prototype, as manufacturing of new technology is the hardest thing by far. Creating the prototype is perhaps 10 percent of the challenge."

"The vertical integration with the machine design teams allows us to design the machine to be one machine and remove unnecessary steps," he said.

One example of this is the 'gigapress' that will be used in Berlin to press out the front and back of the Model Y in two sections using Tesla's own custom aluminium alloy. This will use the new batteries as part of the structure of the vehicle in a fundamental change to the way cars are designed.

"We had to develop our own aluminium alloy that did not require coating or heat treatments and that interfaces to the structural battery. This is quite profound," he said.

All of this will be incorporated into the Berlin Gigafactory.

All the improvements announced were relatively small, from 5 to 20 percent, and come from going back and re-engineering materials, design and processes from scratch. This is a huge move for the company, and put together, these changes drive down the cost of battery cells, but more importantly, are designed to scale. This is why the changes are profound for the industry and a significant challenge for other battery cell makers.

www.tesla.com

Tesla Battery Day will mark dramatic industry shift

By Nick Flaherty

Panasonic's role as a supplier will come under scrutiny at Tesla's battery day later this week

The postponed Battery Day by electric car maker Tesla this week will have a profound effect on the industry. While many rumours circulate about the content of the day, one thing is certain – it will mark a dramatic shift in the way the battery industry operates. This has consequences for the global economy, as the European Union pointed out last week.

No longer can battery makers develop new battery systems themselves – car makers have realised that this is a key capability. Tesla commissioned leading researcher Dr Jeff Dahn at Dalhoustie University in Canada five years ago to work on new battery technologies, from simpler electrolytes for 'anode free' metal batteries to new test methodologies, and we are likely to see the results later this week.

Tesla's $200m purchase of ultracapacitor maker Maxwell Technologies similarly gives it higher efficiency dry electrode technology for longer life time battery cells. This is potentially behind the Million Mile battery, boosting the lifetime of the battery pack from 120,000 miles to day to over 1m.

The role of Panasonic will come under renewed scrutiny after the Battery Day. Setting up the first battery Gigafactory in Nevada for Tesla's high volume Model 3 production with its own technology was a bold, and expensive, move. Panasonic is only just into profit on the deal, five years on. And it hasn't given the company the dominance as a supplier it expected. The automotive division in Panasonic is still listed as 'recovering', and other Tesla suppliers such as CATL and LG Chem are looking hungrily at the opportunities.

CATL in particular has said it has a million mile battery, although the technical details are limited.

Panasonic's deal is also for the one model. While most of the cars are made in Nevada, it also ships batteries to Shanghai for Tesla's Chinese plant, while at the same time CATL ships batteries over from China to Fremont for Tesla's other models.

However Tesla CEO Elon Musk has said the company would expand its business with all three suppliers as the sticking point is the availability of low cost cells. He is also at pains to point out that Panasonic and Tesla continue to have a strong partnership.

All this highlights the move back to a more integrated technology stack for electric car makers, where the brands control the battery technology and form factor and the manufacturers are added value licensees. This is happening with GM's Ultium batteries, the Mercedes deal with Faradis, VW with NorthVolt and many others, including Panasonic and its joint venture with Toyota and SK Innovation with Ford.

These are all using different battery chemistries, limiting the economies of scale. One of the drives is improving the lifetime of NMC (nickel, manganese, cobalt) lithium ion cells while using less cobalt to make them more sustainable. SK Innovation in Korea for example sees this as a way to catapult it into the top three battery suppliers, but there are many other battery chemistries being used.

Similarly there are also new form factors, such as larger battery cells with sahpes to assist high volume manufacturing, which is also expected to be part of the Tesla Battery Day announcements.

As a result, this approach limits the ability to build factories large enough to second source and to drive innovation that benefits the whole industry. The Battery Day will shift the responsibility for that innovation back onto the car makers. For Tesla, that maybe a smart move. For others, not so much.

www.tesla.com; www.panasonic.com

Tesla moves to cobalt-free silicon battery cell with a new form factor

By Nick Flaherty

Tesla is combining a silicon anode, cobalt-free cathode, tabless cell design, high speed manufacturing and a new form factor to slash the cost of vehicle battery packs

Electric car maker Tesla has developed a cobalt-free, silicon lithium ion battery cell that is says will dramatically change the way cars are powered.

The cell is in a new, larger form factor measuring 46mm in diameter and 80 mm long. This compares to the previous 1865 and 2170 cells that are 18mm and 21mm in diameter.

"We have a plan to halve the cost per kWh with engineering and industrialisation," said Drew Baglino, senior vice president of powertrain and energy engineering, speaking at Tesla's Battery Day (above left).

This part of a plan to scale battery cell production to 20TWh per year. The current battery Gigafactory in Nevada built with Panasonic will produce 150GWh a year. "We would need 135 gigafactories like that," said Baglino.

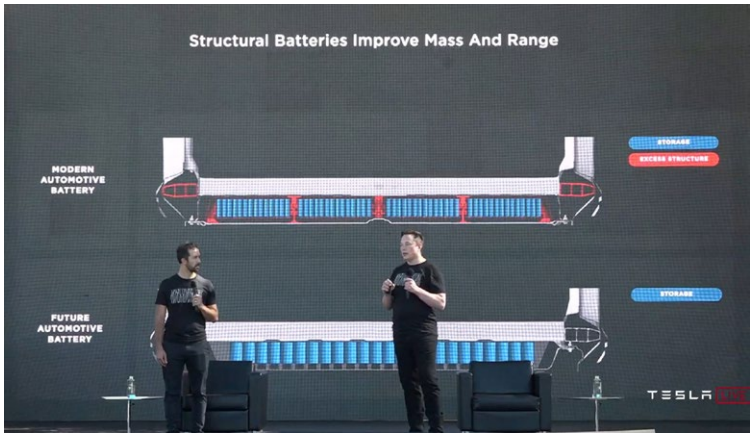
The new cell is key to the scaling up of production. It uses a tables construction with a dry electrode process acquired from ultracapacitor maker Maxwell Technologies, with a simple silicon anode and cobalt-free high nickel cathode. The electrolyte wasn't mentioned though.

All of this is built on a high speed continuous production line similar to a bottling plant. The pilot plant in Fremont California has made tens of thousands of the new cells, but the yield is a problem, says Elon Mush, CEO of Tesla.

"The dry coating they had was proof of concept and we have revised the machine four times since the acquisition," said Musk. "There is still a lot of work to do to go to pilot to volume production, its insanely difficult to scale up, but we have made

tens of thousands of cells. The yield is not good but there is a clear path to success," he added. "We will probably be on machine revision 6 or 7 for volume production with a new rev every three or 4 months.2

The new line produces 20GWh of batteries, 7x the capacity of existing lines, through the high speed and the higher energy density of silicon.



The high speed production comes from the tabless cell construction. Rather than using tabs top and bottom, the substrate of the 'jelly roll' that holds the anode, electrolyte and cathode has copper edges that are laser patterned. These fold over to produce the connections top and bottom. This avoids having to stop and start the line to insert the tabs

Using silicon provides higher energy density but suffers from cracking. "Silicon stores 9x more lithium than graphite but expands 4x when fully charged," said Baglino. "Current approaches use engineered silicon materials and don't scale – what we are proposing is a step change in capability and cost. We use the base silicon, which costs $1.2/kWh, and stabilise the surface with an elastic ion-conducting polymer coating, then use a highly elastic binder."

Tesla is also mining its own lithium in Nevada using a saline process, and will build a cathode plant nearby. This will build cathodes with no cobalt with a variety of other materials such as iron and manganese that it already uses. Cutting out cobalt avoids sustainability and ethical production issues.

"This uses metal powder directly, eliminates billions in battery grade nickel production, with simpler mining and simpler recycling," said Baglino. "We want to make sure we are not constrained by nickel supply but we need a three tier approach with iron [for stationary storage], Nickel Manganese [for mid range vehicles], then high nickel for long range for the cybertruck and the semi," said Musk.

The answer to how Tesla addresses the swelling of the silicon may well come from the packaging of the cell which will be used as part of the structure of the vehicle, rather than in a separate battery pack.

"This has a dual use as energy and structure – this is quite profound," said Musk ."This allows higher packing as there is no intermediate structure, so there's more space, the pack itself is structural. Instead of a flame retardant filler in the pack, the filler is structural adhesive that is also flame retardant. This allows shear transfer between top and bottom sheets and this gives stiffness. We use the steel shell case of the battery to transfer the shear," he said.

In this case the steel of the cell casing could be significantly thicker to be used as structure and to prevent the casing bursting when it swells.

It is this combination of new materials and new processes from the mining to the final testing that is the way to reduce the overall cost of the cell and the battery pack, say Baglino and Musk. This halves the pack cost per kWh (by 56 percent), although Tesla doesn't give a cost. Current cell costs are already under $150/kWh, so the overall pack cost for Tesla would be well under $100/kWh. This allows a lower cost electric vehicle platform.

"It will take us a year to 18 months to start to realise the advantages and probably about 3 years overall but this bodes well for the future," said Musk. "We are confident we can design and manufacture a compelling $25,000 vehicle three years from now that's also fully autonomous with an EV powertrain that costs less than a combustion engine."

www.tesla.com

European chip firms concerned over US export controls

By Peter Clarke

The European Semiconductor Industry Association (ESIA) has expressed concern about tightening US export control measures and has asked for increased dialogue.

The US has instituted unilateral export controls requiring all semiconductors designed or made with US technology to be licensed for export to certain entities. Even though much IC design work is done outside the US, design tools and manufacturing tools from the US are more or less a necessity to complete chips.

ESIA has issued a statement saying that it is concerned that, following the US announcement of 17 August 2020, the export control measures will have "have significant impact and bring disruption to the global semiconductor industry."

ESIA is not the first organization to protest at the US administration's effort to take control of global chip exports. The SEMI and SIA organizations did so back in April (see Semiconductor industry pushes back against US export controls). ESIA includes both European headquartered chip companies such as Infineon, NXP, STMicroelectronics and X-Fab but also overseas-headquartered chip companies with a significent European presence such as Intel, Globalfoundries, Renesas and Texas Instruments.

"Export controls should be transparent and accountable to serve as tools of global non-proliferation and be multilateral to be effective for that purpose. ESIA would welcome a discussion on export control measures between the US administration and international partners such as the European Union," said ESIA in its statement.

www.eusemiconductors.eu

US looks to add SMIC to trade embargo

By Peter Clarke

The US government is considering including Chinese foundry Semiconductor Manufacturing International Corp. (SMIC) in its ban on the supply US technology, according to a Reuters report.

This extension of the trade blacklist would come after administration of President Trump has effectively cut Chinese communications equipment company Huawei off from its sources of ICs by including the company on its "entity list.".

Reuters cited a spokeswoman at the Department of Defense as its source. Such a move would force US and other suppliers to seek a special license before supplying goods and services to the company.

Although SMIC cannot yet supply leading-edge ICs, with its highest performance chips being at about the 14nm FinFET node, it is China's best hope of moving towards self-sufficiency in leading-edge semiconductors in the long term.

Many US companies are key providers of semiconductor manufacturing equipment and no leading-edge wafer fab can operate without these. US company Applied Materials is the world's largest semiconductor manufacturing equipment supplier. Also at the leading-edge ASML Holding NV (Bilthoven, The Netherlands) is a monopoly supplier of the extreme ultra violet (EUV) lithography equipment required to make chips at 5nm and below.

ASML and most other companies in the west are operating in line with US requirements on supplying China.

Applying the licensing requirement to SMIC would bring China's move towards self-sufficiency and global significance in semiconductors under tight US control.

Although such an embargo would not cause the immediate closure of SMIC, a lack of on-going support and spares for equipment would have an impact on productivity. In the medium to longer term it could cause SMIC to change its methods of manufacturing and what it is able to make.

The news caused SMIC's share price to fall by 20 percent when it opened Monday September 7. However, due to the strong run on the shares after they opened on the Shanghai stock market the price has only fallen back to where it was on June 1.

Trade war risks for wafer supply

By Nick Flaherty

The US-China trade war risks the global supply chain, particularly for the supply of wafers, says the chief executives of Global Foundries and X-Fab.

"We need a little more balance in the supply chain," said Tom Caulfield, CEO of Global Foundries. "Wafer supply is dominated by Taiwan." GF has 300mm fabs in the US and Germany, as well as a 200mm fab in Singapore.

That view is reflected by European foundry X-Fab, which also has a fab in the US at Lubbock, Texas.

"The evolution of the global politics and the whole trade war is of course not good for the semiconductor industry and for industry overall as it puts in barriers and that's not good," said Rudi de Winter CEO of German foundry X-fab (above). "At the same time it drives changes and that maybe opens up opportunities but overall is more negative than positive

"There are in certain areas definitely concentrations that could pose risks," he said in agreement. "We source wafers all regions in the world – we have a wide diversity of technologies from more mature on 6in and 8in wafers, but we don't have particular problems but we do see the whole evolution of the trade war as an obstacle to doing business when people should be focussing on growing business and removing barriers."

STMicroelectronics has also been securing its supply chain for its specialist wafers, with deals to buy wafer makers Norstel for silicon carbide (SiC) and ExaGaN for gallium nitride (GaN). It works with TSMC in Taiwan for mainstream silicon process technologies.

Like Global Foundries, de Winter is also looking at local financial support, whether that is called co-investment or subsidy, particularly with the US Creating Helpful Incentives to Produce Semiconductors (CHIPS) for America Act signed in June.

"In all the locations where we are active we are using local measures to support our activities and if you look at the CHIPS Act to stimulate US business that is well aligned with our activities in Lubbock [Texas]. The CHIPS Act is a good thing for semiconductor industry and particularly for the US semiconductor companies, so we look forward to see how that can help our growth in Lubbock."

European locations need the same kind of support, he says.

"I hope it will be complemented as there is the same awareness in Europe that things need to happen but not in the same proportion as the US. There is support and awareness but it could be further stepped up, and the European community is not always helping, in particular the rules on subsidy are very strict and that is hindering strategic area like semiconductors that has a huge leveraging effect on the overall industry. They are essential. A lot of our industrial equipment depends on semiconductors and if these trade barriers are going to make things more difficult we need to make sure we continue to have access to state of the art technologies."

"We need an even playing field around the world, and we should look at what the other regions of the world like China are doing. I am not in favour in general of subsidies but we need an even playing field so we need to find the right balance," he said.

Globalfoundries ready for US government co-investment

By Peter Clarke

Tom Caulfield, CEO of Globalfoundries Inc., has expressed support for US government plans to stimulate domestic chip manufacturing with capital.

Speaking ahead of his company's virtual Global Technology Conference, Caulfield spoke of a global semiconductor supply chain imbalance and Taiwanese dominance of global wafer production. "We need a little more balance in the supply chain," he said while observing that US-headquartered companies serve 47 percent of the total chip market but only 12 percent of the world's chips are manufactured in the US.

In recent months there have been political moves to use US tax payers' money to support domestic chip manufacturing.

When asked if Globalfoundries could benefit from US government subsidies and use them to expand manufacturing or even build an additional wafer fab, Caulfield said: "I don't like the word subsidy. It sounds like a government hand-out. I prefer co-investment. It used to be the case that we let open-market competition decide who the winners and losers are. But countries are not doing that. They are treating semiconductors as strategic and the US is realizing we have to do the same. That's what the CHIPS Act is about."

Caulfield continued: "We are growing our global footprint. We are accelerating growth and we will decide how best to do that." He said Globalfoundries has already invested billions of dollars in US manufacturing and asked who is better placed than Globalfoundries to invest further funds under US government policy.

When asked if Globalfoundries was still on course to stage an initial public offering of shares in the company, despite problems such as the Covid-19 pandemic, Caulfield said: "We have conviction. There are certain metrics you have to meet. Can you meet the top-line [revenue] you want? We're on-track for a late 2022 IPO."

www.globalfoundries.com

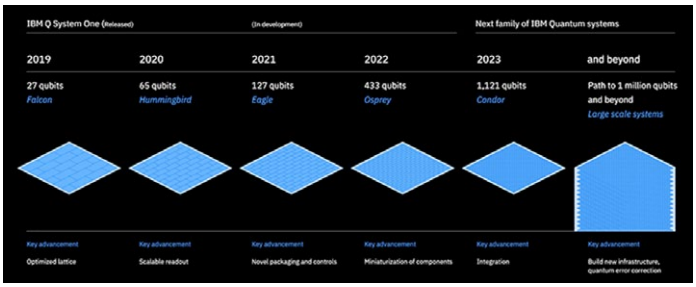
IBM aims for million qubit quantum computer

By Nick Flaherty

IBM has launched its roadmap for quantum computing, aiming to deliver a system with 1000 qubits by 2023. This will form the basis of systems with millions of qubits.

The 1000 qubit system in 2023 will have addressed the scalability challenges for much larger systems, says IBM.

"We think [this] will take us from the noisy, small-scale devices of today to the million-plus qubit devices of the future," said Jay Gambetta, IBM Fellow and Vice President of IBM Quantum. "Our hardware roadmap sits at the heart of a larger mission, to design a full-stack quantum computer deployed via the cloud that anyone around the world can program," said Gambetta.



The IBM system is based around the electronic quantum states of artificial atoms known as superconducting transmon qubits, which are connected and manipulated by sequences of microwave pulses in order to run these circuits.

The company has been working on superconducting qubits since the mid-2000s, increasing coherence times and decreasing errors to enable multi-qubit devices in the early 2010s. It now has over 24 stable systems on the IBM Cloud for clients and the general public to experiment on. These include the 5-qubit Canary processors and 27-qubit Falcon processors.

The 1000 qubit device, IBM Quantum Condor, will be launched by the end of 2023 and to house larger systems IBM is developing a dilution refrigerator larger than any currently available commercially.

"Simultaneous to our efforts to improve our smaller devices, we are also incorporating the many lessons learned into an aggressive roadmap for scaling to larger systems. This month we quietly released our 65-qubit IBM Quantum Hummingbird processor to our IBM Q Network members," said Gambetta.

This device features 8:1 readout multiplexing, combining readout signals from eight qubits into one, reducing the total amount of wiring and components required for readout and improving the ability to scale. IBM says it has also significantly reduced the signal processing latency time in the associated control system in preparation for upcoming feedback and feed-forward system capabilities. This will allow control of the qubits based on classical conditions while the quantum circuit runs.

"Next year, we'll debut our 127-qubit IBM Quantum Eagle processor," said Gametta. Eagle includes through-silicon vias (TSVs) and multi-level wiring to effectively fan-out a large density of classical control signals while protecting the qubits in a separated layer in order to maintain high coherence times. This will also include concurrent real-time classical compute capabilities that will allow for execution of a broader family of quantum circuits and codes.

All of this is in preparation for the 433-qubit Osprey system in 2022, to be followed by 1,121-qubit Condor processor in 2023. This will require larger cooling systems. The 3.3 x 2m (10 x 6ft) "super-fridge," internally codenamed "Goldeneye," is designed with a million-qubit system in mind.

"We think of Condor as an inflection point, a milestone that marks our ability to implement error correction and scale up our devices, while simultaneously complex enough to explore potential quantum advantages—problems that we can solve more efficiently on a quantum computer than on the world's best supercomputers," said Gambetta.

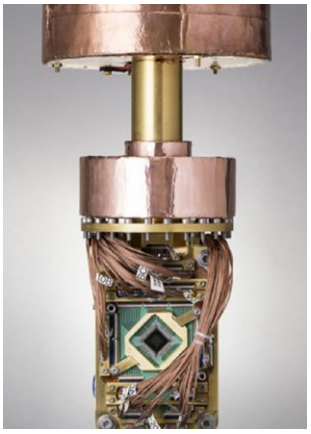
www.ibm.com

Study: How the automotive industry will benefit from quantum computing

By Christoph Hammerschmidt

Quantum computers are far from being technically mature, but practical applications are already emerging. In particular, the automotive industry could benefit greatly from the use of these machines. The consulting firm McKinsey has investigated where the potential lies.

After companies such as IBM with its Q System One or D-Wave Technologies made headlines in recent years with supposedly usable quantum computers, various companies in the automotive value chain have taken a closer look at this technology - the promises made by manufacturers were too seductive. According to their pledges, quantum computers are ideal for solving certain problems that the best scientists have long been brooding over, such as route optimisation, fuel cell optimisation and the durability of materials.



According to the McKinsey study, some of these early users have already achieved a certain success. Volkswagen, for example, has teamed up with D-Wave to develop a traffic management system that optimises the routes of buses in urban traffic. The automotive supplier Bosch has invested $21 million in the start-up company Zapata Computing (Cambridge, Massachusetts).

However, the reluctance still far outweighs the commitment to this innovative computing technology, write the authors of the McKinsey study. The novelty of the technology and the still very narrow market have so far discouraged many companies from intensively engaging in quantum computing. It will take another five to ten years before this technology has become established in the long term. By then, quantum computing will have overcome several hurdles: Quantum Supremacy must be achieved; the practical benefit must be proven beyond doubt; application software must be available to solve concrete problems; and above all, a Quantum Turing Machine must be available. The latter means that a universally applicable quantum architecture with quantum memory and conventional main memory (RAM) must be available. Such a machine, as described by the experts at McKinsey, will be able to work with the number of qubits required by the users and execute arbitrary algorithms. Such a machine will be available in one to two decades, the study says.

The automotive industry will be one of the primary value pools for QC, with a high impact noticeable by about 2025. Most of the early value added will come from solving complex optimization problems, including processing vast amounts of data to accelerate learning in autonomous-vehicle-navigation algorithms. In later years, QC has the potential to have a positive effect on many areas in the automotive industry, such as vehicle routing and route optimization, material and process research, and the security of connected driving. Moreover, QC can also provide a boost to automotive players transitioning into the electric-vehicle (EV) era by notably accelerating R&D of novel technologies. For instance, companies can speed the transition of their more traditional technology spectrum towards more relevant technologies such as cooling of EV batteries. Likewise, the simulation of material process research for batteries and fuel cells could be a field where QC could be deployed with a chance of success.

Near-term opportunities for QC – for the time frame through 2025 – are expected to surface in product development and R&D. Relevant use cases will primarily relate to solving simple optimization problems or involve parallel data processing for simple quantum artificial-intelligence/machine- learning (AI/ML) algorithms. These QC applications will be executed as part of a hybrid solution, where bits of a larger problem, processed by a High-Performance Computer (HPC), are outsourced to a quantum computer and results are fed back into the HPC flow. Possible optimization use cases include the combinatorial optimization of multichannel logistics, highly local traffic-flow optimization, and improvements in vehicle routing. Quantum AI/ML might involve the time-efficient training of autonomous-driving algorithms due to an increase in the parallel processing of large amounts of data.

In the midterm – that is, the timeframe from 2025 through 2030 – the authors expect the QC activities in the automotive industry will focus on things like simulations (heat and mass transfer, fluid dynamics as well as material properties at the atomic level – relevant for the development of battery and fuel cell materials). In addition, more complex city traffic simulations could become possible as well as solving large-scale multimodal fleet routing problems. Plus, the capability of more advanced quantum computers to process large amounts of data will help engineers and developers to implement solutions for enhanced pattern recognition.

In the long run, that is from 2030 onward, quantum- computing applications will build on at-scale access to universal quantum computers, the experts estimate. Prime factorization algorithms to break common encryption keys will therefore be universally available. The focus will likely move toward digital security and risk mitigation as players try to prevent the quantum hacking of communications in autonomous vehicles, on-board electronics, and the Industrial IoT. Cloud-hosted navigation systems of shared-mobility fleets will improve their coverage algorithms through regular training enabled by quantum computing. Other promising fields of application include investigating and optimizing crash behaviour, cabin soundproofing or training for AI-based autonomous driving algorithms.

Such applications require a steep maturing process for the QC industry. The McKinsey experts admit that today it is unclear how the QC hardware industry will be able to reach this degree of maturity, but they see a number of possible ways. For instance, existing QC approaches will evolve over time. QC will establish itself as a cloud service, which will relieve users of the need to acquire and run their own hardware. Similarly, the market researchers expect that the QC software will evolve – with the difference that in contrast to hardware where almost all competencies are concentrated in the US, European players will also become relevant.

While many uncertainties persist, the analysts express optimism that most of the problems ahead will be solved within the time frame specified. For 2030, they expect an economic impact of these technologies in the automotive industry alone of some $2 billion to $3 billion by 2030.

UK buys Rigetti quantum computer in £10m deal

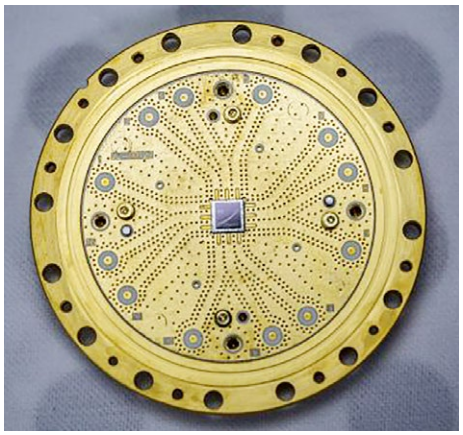
By Nick Flaherty

The UK is to buy a quantum computer from Rigetti Computing as part of a £10m (€11.2m) project to provide a cloud quantum computing service housed at Harwell, Oxfordshire.

The Rigetti quantum computer system will be hosted at the newly announced National Quantum Computer Centre in Abingdon, Oxfordshire. Rigetti has also developed a cloud-based platform allowing computer programmers to write quantum algorithms and will work alongside Oxford Instruments, Standard Chartered and Bristol and London-based quantum software start-up Phasecraft, as well as the University of Edinburgh.

A Rigetti superconducting quantum computer is already commercially available in the Amazon Web Service (AWS) Bracket cloud, alongside other US-based systems using different approaches from D-Wave and IonQ.

The UK government expects quantum computing to provide £4 billion of economic opportunities globally by 2024, while in the coming decades productivity gains resulting from quantum computing are expected to surpass over £341 billion globally.



"Our ambition is to be the world's first quantum-ready economy, which could provide UK businesses and industries with billions of pounds worth of opportunities. Therefore, I am delighted that companies across the country will have access to our first commercial quantum computer, to be based in Abingdon," said UK Science Minister Amanda Solloway.

"This a key part of our plan to build back better using the latest technology, attract the brightest and best talent to the UK and encourage world-leading companies to invest here," she said.

"We are excited to deliver the UK's first quantum computer and help accelerate the development of practical algorithms and applications," said Chad Rigetti, CEO of Rigetti Computing.

There are currently only a small number of quantum computing platforms being developed around the world – presenting an opportunity for the UK to be at the forefront of this technology. The activities announced today will help promote quantum computing across the UK economy, providing businesses with the best opportunity to take advantage of these new technologies in the years to come.

The £93m Centre, first announced in 2018, will bring together academia, businesses and the government to address key challenges to quantum computing, such as scaling-up the technology and making it commercially viable and exploring how to create economic value.

Working closely with industry and the research community, the Centre will also provide businesses and research institutions with access to quantum computers as they are developed around the world and grow the UK's thriving quantum computing industry.

"Quantum computers are extraordinary new tools with the potential to allow us to tackle previously insurmountable challenges, promising benefits for all of society through applications in areas such as drug discovery and traffic optimisation," said Professor Dame Ottoline Leyser, chief executive of UK Research and Innovation.

"The National Quantum Computing Centre will tackle the key bottlenecks in quantum computing by bringing together experts from across the UK's outstanding research and innovation system from academia and industry to unlock the potential of this exciting new technology," she added.

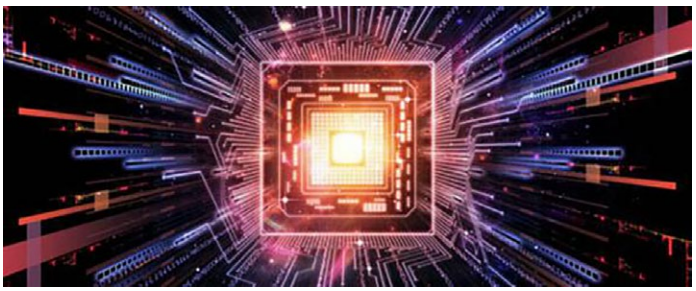
"The next steps initiating centre recruitment and commissioning technology work packages are very welcome tangible steps as the centre moves from initialisation and conceptual design to facility construction and operational delivery," said Dr Michael Cuthbert, director of the National Quantum Computing Centre Director.

Breakthrough for quantum key distribution networks

By Nick Flaherty

Researchers at the Quantum Engineering Technology (QET) Labs in Bristol have built a multiplexed eight user quantum key distribution system with just eight receivers, a fraction of the current requirement for QKD networks

An international team of researchers has developed the first distributed QKD network for sharing quantum keys in a breakthrough design.



So far, quantum key distribution has been point to point, even over satellites, but this limits the use in a network. The multiplexed photonic quantum key distribution (QKD) system, published in Science Advances , supports eight users and can be easily scaled up.

"This represents a massive breakthrough and makes the quantum internet a much more realistic proposition," said Dr Siddarth Joshi, who headed the project at the Quantum Engineering Technology (QET) Labs at the University of Bristol, UK. "Until now, building a quantum network has entailed huge cost, time, and resource, as well as often compromising on its security which defeats the whole purpose."

"Our solution is scalable, relatively cheap and, most important of all, impregnable. That means it's an exciting game changer and paves the way for much more rapid development and widespread rollout of this technology," he said.

Photonic QKD systems use entangled photons to ensure an encryption key is not intercepted.

"Until now efforts to expand the network have involved vast infrastructure and a system which requires the creation of another transmitter and receiver for every additional user. Sharing messages in this way, known as trusted nodes, is just not good enough because it uses so much extra hardware which could leak and would no longer be totally secure."

The team includes researchers from the UK's University of Leeds, Croatia's Ruder Boskovic Institute (RBI) in Zagreb, Austria's Institute for Quantum Optics and Quantum Information (IQOQI), in Vienna, and China's National University of Defence Technology (NUDT) in Changsha.

The team used multiplexing to develop an eight user system with eight transceivers, rather than the 56 that would previously be needed for each user to have a point to point link. The receivers were connected to optical fibres via different locations across Bristol and the ability to transmit messages via quantum communication was tested using the city's existing optical fibre network.

"Besides being completely secure, the beauty of this new technique is its streamline agility, which requires minimal hardware because it integrates with existing technology," said Joshi.

The network was created within months for less than £300,000, enabling secure networks for a fraction of the cost today. The system also features traffic management, delivering better network control which allows, for instance, certain users to be prioritised with a faster connection.

"With these economies of scale, the prospect of a quantum internet for universal usage is much less far-fetched. We have proved the concept and by further refining our multiplexing methods to optimise and share resources in the network, we could be looking at serving not just hundreds or thousands, but potentially millions of users in the not too distant future," said Joshi.

"The ramifications of the COVID-19 pandemic have not only shown importance and potential of the internet, and our growing dependence on it, but also how its absolute security is paramount. Multiplexing entanglement could hold the vital key to making this security a much-needed reality."

The research received funding from the Quantum Communications Hubs of the Engineering and Physical Science Research Council (EPSRC), Ministry of Science and Education (MSE) of Croatia, and the Austrian Research Promotion Agency (FFG).

bristol.ac.uk/qet-labs/

Sigfox sells German IoT network in European restructure

By Nick Flaherty

Sigfox alliance with Cube Infrastructure Managers sees restructuring to create Europe's largest IoT network

French network operator Sigfox has sold its German low power, long range network for the Internet of Things (IoT) to a fund manager to raise cash.

The deal with fund manager Cube Infrastructure Managers (Cube) is pitched as strategic alliance, as it is also buying a majority stake in Heliot Europe, the owner and operator of the Sigfox networks in Austria, Switzerland and Liechtenstein, to create a Europe-wide IoT network.

The value of the German sale was not disclosed but it will allow Sigfox to finance its continued innovation efforts in data value extraction and improvements in cloud algorithms to reduce energy consumption and allow the implementation of even more cost-effective devices and sensors says the company

Sigfox has networks in 72 countries and regions, some owned directly such as Germany but most owned by partners such as Heliot, called Sigfox Operators. These operators are the owners of the networks, which they operate as exclusive connectivity providers of Sigfox IoT services, offering worldwide connectivity to customers.

Cube sees the deal as bringing together the networks of all four countries to create the largest IoT network in Europe. The aim is to accelerate the growth of this essential and exclusive IoT infrastructure in the region in joint venture with the operating management of the networks, which retains a minority stake in Heliot Europe.

"The acquisition of Sigfox Germany is a key milestone in our development and we will pursue commercial development initiated by Sigfox Germany with tier one customers in the country and beyond, in retail, automotive and logistics," said Thomas Scheibel, CEO at Heliot. "This combined network from

the Adriatic to the North Sea reflects our clients expectations and plays well to our operational strengths. We are looking to further accelerate and foster the adoption of IoT at the heart of European industry. This is made possible with the successful and entrepreneurial team already fully integrated into the Sigfox ecosystem as well as the powerful financial backing and experience of Cube Infrastructure Managers. This is also an unprecedented opportunity to reinforce our strategic and long-term alliance with Sigfox," he said.

Rather than competing with cellular technology, Sigfox has pitched its proprietary low power, long range network as complementary as it uses far less energy and radio spectrum and cost far less. It provides the cloud services to the network operators as well as licensing the technology to operators, device makers and chip makers such as ST Microelectronics and ON Semiconductor.

"From the very beginning, Sigfox has been aiming to offer customers an ever-better service, through innovation, densification, a mature ecosystem, and a high return on investment. Selling Sigfox Germany to dedicated and long-term partners such as Heliot and Cube will allow us to keep this promise," said Ludovic Le Moan, CEO and cofounder of Sigfox. "Our goal at Sigfox, with our operators and our ecosystem partners, is to minimize data extraction costs. Sharing infrastructure is a first step toward lower TCO, and higher reliability and interoperability."

Cube started out backing fibre optic infrastructure companies.

"We are known as pioneers among infrastructure funds for investing early in optic fibre infrastructure companies in Europe, long before fibre became the "fourth utility" and have over the last decade accompanied the growth of several European infrastructure operators," said Henri Piganeau, Managing Partner at Cube Infrastructure Managers.

"Today, we see the development of massive IoT as a critical enabler and driver of economic, social and environmental progress in the context of the digital transformation of our economies. With this investment in the Heliot platform and strategic alliance with Sigfox, we are now aiming at consolidating a new communication infrastructure, harnessing its potential to enhance the competitiveness of our industries and territories, and contributing to the efficient use of public services and scarce resources consistent with our ESG objectives," he said.

www.sigfox.com
www.cubeinfrastructure.com

UK project to build quantum key distribution satellite

By Nick Flaherty

A £23m (€25m) project led by Arqit with BT, Fraunhofer UK and Toshiba Research, aims to develop a QKD payload

The UK's innovation agency has awarded a £23m project to build a quantum key distribution (QKD) payload for a satellite. The first small satellite for low Earth orbit (LEO) is being built by the European Space Agency (ESA)

The "Quantum Payload Factory" project commissioned by InnovateUK is led by a London-based service provider startup called Arqit and aims to establish the world's first commercial QKD satellite constellation. This is a considerable challenge given China's lead with a QKD satellite in space and European satellite operator SES also working on the technology (see links below).

The partners developing the payload include BT, Toshiba Research labs, Fraunhofer UK Research as well as NU Quantum, London-based quantum fibre specialist Orca Computing and STFC Laboratories. The photonic quantum chips for the payload will be developed by AegiQ in Sheffield: Startup raises £1.4m for quantum photonic chip

Quantum keys can encrypt data for transmission over conventional fibre links across any distance, but QKD itself is limited over fibre to around 150km. Beyond this, 'trusted nodes' are required, but at major risk of creating security vulnerabilities.

A number of fibre QKD networks are being built, including in the UK, but satellites provide the means for distributing keys across very large distances between end users spread across countries or continents.

Satellite components in QKD networks have already been demonstrated by Fraunhofer in Germany.

InnovateUK sees this as an opportunity for the UK quantum technology industry to leapfrog other countries by creating a capability to manufacture the next generation of space QKD payloads here in the UK.

The technologies will be used in the second generation of Arqit satellites and opens up the potential for use in systems such as the OneWeb LEO constellation which was rescued by the UK goverment earlier in 2020 to be a technology platform.

www.gov.uk/government/organisations/innovate-uk

Nuvia raises $240m to build ARM data centre chip

By Nick Flaherty

Nuvia raises $240m (€206m) in series B funding to get its data centre system-on-chip to silicon

A US startup less than 18 months old has raised nearly $300m (€258m) to develop and produce a new generation of chips for the data centre.

Nuvia was publically launched in February 2019 by three engineers from Apple to build a new high performance processor core around the ARM architecture, and use that for a system-on-chip for the data centre.

Nuvia aims to more than double the performance of previous chips based around ARM's Cortex-A79 core with its Phoenix core within the same power budget. This is a clear competitive driver for ARM's latest datacentre roadmap, which split into the N2 core for scaling and the V1 vector core for higher performance at comparable performance levels.

Nuvia's Orion chip would compete with Amazon's in-house Graviton3 chip, where ARM would expect to see a design-in, and Facebook's internal server chips, but may have an opportunity with Microsoft's Azure data centres and the Google Cloud. But it would take advantage of the move to ARM-based servers with the software eco-system support.

Nuvia's latest funding round was led by Mithril Capital, which has largely invested in healthcare technology in recent years, as well as Sehat Sutardja and Weili Dai, founders of Marvell Technology Group). Funds and accounts managed by BlackRock, Fidelity Management & Research and Temasek, with additional participation from Atlantic Bridge, Redline Capital, Capricorn Investment Group, Dell Technologies Capital, Mayfield, Nepenthe and WRVI Capital.

The funding builds on a $53M Series A round, raised in November 2019.

"The opportunity in front of Nuvia has never been brighter, with an industry that's looking for a new way to get the performance needed to power the next generation of cloud and enterprise computing," said Gerard Williams, CEO of Nuvia. A former Senior Director at Apple and Chief CPU Architect for nearly a decade, he spent spent over 10 years at ARM, as an ARM Fellow, and serving on the ARM Architectural Review and Technical Advisory Boards. While at ARM, he served as a technical advisor for the ARM architecture and CPU development to many key ARM partners, and started his career at Intel and Texas Instruments. "We're very fortunate to have an incredible group of investors behind us as we close Series B and take the next steps in our vision to redefine performance, energy efficiency, scalability, compute density and total cost of ownership within the data centre."

The cash won't go very far. Although Nuvia hasn't disclosed many technical details, the Orion server chip needs to be built on a 5nm process to be competitive in the market, and that can cost up to $600m, so expect signficant Series C funding to get the chip to market in volume. A 7nm design is half the wafer cost but would reduce the performance advantages. UK data centre chip designer Graphcore for example is working with TSMC on 5nm and 3nm technology.

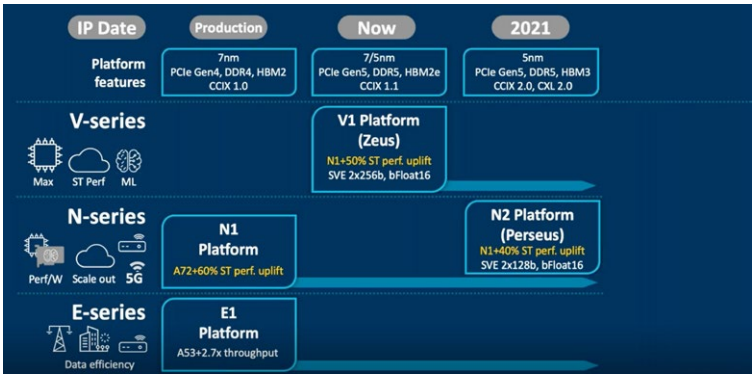
www.nuvia.com

ARM splits its Neoverse datacentre server chip designs

By Nick Flaherty

ARM has split its Neoverse datacentre processors into separate lines for high performance vector processing and for scaling in existing racks and for edge AI.

The company launched its first Neoverse core, the N1, last year, which in a 7nm process saw a 60 percent performance boost over the previous Cortex-A72. The N1 has been used for multiple datacentre server chip designs, including Amazon's Graviton 2. Altran and NXP have also used for core for the next generation Layerscape 5G network processors.



Now it is splitting the line with the successor N2, codenamed Perseus, and the V1, codenamed Zeus, the company's first scalar vector processor for high performance server designs. This has even more significance with the proposed acquisition of ARM by Nvidia with a focus on the data centre alongside GPU cores.

"The emergence of ARM in the data centre is powered by many factors, all built on performance," said Chris Bergey, general manager of the ARM infrastructure business. "We started with a power performance message and this didn't resonate – we had a performance gap to close."

"The N2 Perseus we expect to offer 40 percent higher performance over N1 with the same area and power efficiency for scale out cloud, smartNIC and edge devices," he said.

The V series will be optimised for higher performance than N2, with larger buffers, cache, windows and queues that allow a single thread to execute faster. It will also support two 256bit vector instructions for high performance computing (HPC), cloud computing and AI. SiPearl and the European ExaScale project are planning to use the cores in high performance system-on-chip designs for supercomputers.

This will support scalar vector extensions (SVE) with full control over voltage and frequency transitions when moving between linear and vector instructions. "This has run at full frequency with SVE code and transition seamlessly between memory register widths for helper code," said Bergey.

The split gives developers a choice, says Berger. With the N2, a 42U rack in a data centre with a 25kW power supply can run significantly more thread throughput and the cloud provider can host more customers per rack and the customer gets better performance per core, he says. This could have up to 192 cores in a chip with a thermal profile (TDP) of 350W in teh data centre, or smaller core counts for edge and network applications.

"There's a powerful story for 5G and edge computing where there are even more constraints on power. The defining characteristic is an equal focus on performance, power and area, whether that's a 250W cloud SoC or 20W 5G basestation SoC," he said.

V1 would have less cores but optimised for 40 percent higher performance than N2. "What we see is customers have a TDP to optimise around, going from air cooling to liquid cooling have different cost profiles so really it's about the balance of TDP and computing performance, which is the thing with V1 vs N2. That kind of customisation is really interesting to the ecosystem," said Bergey. "About 36 cores of V1 would be the high end of cooling."

The IP will be launched later this year with more technical details on the architectures, and designs are progressing on TSMC 7nm and 5nm processes.

www.arm.com

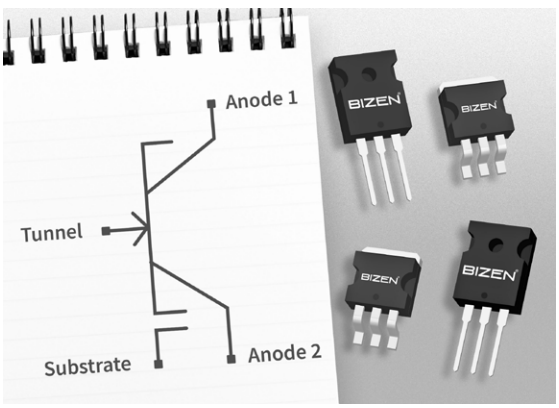
First Bizen quantum tunnelling transistors launched

By Nick Flaherty

A UK startup is to ship its first 1200V power devices using a new silicon architecture called Bizen that fits into TO247 or TO263 packages

The first devices to use the Bizen process technology include three parts rated at 1200V/75A, 900V/75A and 650V/32A, available in the industry-standard TO247 or TO263 power MOSFET packages. These are made using standard silicon substrates on conventional, larger-geometry silicon processing lines. The initial pilot line for testing the Bizen technology was set up at Semefab in Scotland.

"To get this level of performance from traditional silicon-based MOSFETs, the device size must be much bigger. 1200V/75A in a TO247 housing can be achieved using wide bandgap materials like silicon carbide, but this approach has other well-known issues," said David Summerland, CEO and founder of Nottingham-based Search For The Next (SFN) which developed the Bizen technology. This is being used for the parts, called Quantum Junction Transistors (QJT), from subsidiary Wafertrain.



"SiC, for example, takes much longer to process and has a significant manufacturing carbon footprint. Also, regardless of the roadmaps, SiC does not scale like silicon, and the economic argument that SiC can match silicon does not take into account the advances made possible by Bizen. By

contrast, the data we have obtained from physical wafer tests proves that by using Bizen on silicon substrates, our QJTs deliver the same performance as SiC or GaN. Yet the production equipment required to make a QJT is exactly the same as for a standard silicon MOSFET, and the Bizen process adds no extra manufacturing complexity."

Bizen applies quantum tunelling to a traditional bipolar wafer process. The result is a very rugged and reliable device with the heritage and pedigree of traditional bipolar silicon technology. Bizen also reduces leadtimes from 15 weeks for a CMOS device (although not a SiC MOSFET) to less than two weeks, and halves the number of process layers. The QJTs use these same eight layer bipolar process.

Wafer tests also show that the Bizen process exhibits an effective current gain of over 1 million, says Summerland. This will enable direct connection between the 1200V/75A QJT power transistor and a low voltage, low current CPU output port such as a PWM for a monolithic design.

"The QJT is the first power device on the Bizen family roadmap. This will shortly lead to the PJT (Processor Junction Transistor), an integrated Bizen device with its own processor which can also be produced on a manufacturing cycle time of eight days, heralding a new era of intelligent power devices," he said.

SFN has also released other comparative performance metrics for a 1200V/100A part – also in TO247 – which is on its short-term roadmap. The losses at rated current will be a quarter (<300mV) of those exhibited by the SiC device, and its input capacitance will also be four to five times less (< 1pF).

In July Semefab also started making 1200V Schottky diodes on an SiC process and is planning 1700V devices.

www.wafertrain.com/blog

Israeli startup raises $6.5m for IoT security

By Nick Flaherty

Sternum raises $6.5m in Series A funding to provide embedded cybersecurity and real time visibility for IoT devices

Sternum in Israel has raised $6.5 million in Series A funding for its embedded cybersecurity technology in the Internet of Things (IoT).

Sternum will use the funding to expand its R&D team and commercialisation of its scalable, embedded software that is injected into a device's binary code and can be integrated into a wide range of IoT devices – low-end, high-end, old, new, existing devices, RTOS, and Linux-based – with no configurations or changes to existing code required.

The Tel Aviv-based company recently proved the effectiveness of its solutions by successfully blocking the exploitation of multiple critical zero-day vulnerabilities known as Ripple20. It has already teamed with module maker Telit and a leading global medical device manufacturer, to provide first-of-its-kind endpoint security for safety critical IoT devices.

"I'm excited to say that we are the only company offering both an embedded, agentless, and holistic security solution for all IoT devices, as well as the ability to monitor them through an intelligent cloud platform, no matter the operating system or resources available, thus creating the first on-device technology acceptable by all IoT devices," said Natali Tshuva, CEO and co-founder of Sternum.

0 billion IoT devices are predicted to be in use by 2025, and until now, there has not been a scalable, endpoint cybersecurity solution that can be installed on individual devices. Instead, manufacturers have heavily relied on vulnerability patching or network security. Manufacturers are also not capturing valuable data points from devices in operation. Device data can provide key insights for optimizing performance, supporting future developments, and profiling and mitigating third-party behaviour and risks.

The round, led by Square Peg, was joined by existing investor and global business leader Merle Hinrich, European venture capital firm btov, and private investors including Boston-based veteran entrepreneur Eyal Shavit and Founder & CEO of CyberArk, Udi Mokady. This brings the total raised to $10m.

"This round, closed during a global pandemic, is a vote of confidence in our dual vision for IoT cybersecurity and on-device intelligence, our technology, and most importantly, our team. With these new funds, we will continue actively recruiting researchers and engineers to grow our industry-leading R&D team in our mission to secure all IoT devices and obtain insights that will vastly improve our connected world," said Tshuva.

"We are impressed with Sternum's innovative products and diverse team, whose technologies will power our connected future with uncompromising security protection and rich, data-driven insights," said Philippe Schwartz, Partner at Square Peg. "We look forward to supporting the Sternum team as they continue to help their customers protect and secure their business-critical IoT devices."

"Authentication, security and real-time monitoring across all IoT devices and applications is a critical problem for manufacturers and consumers," said Merle Hinrich, Founder and Chairman of Hinrich Foundation. "Sternum's solutions are unique, scalable, and proven. We are delighted to be invested in and supportive of Sternum's team and technology, which will assist in bringing trust and dependability to industries implementing IoT."

Following this investment, Shavit, Schwartz and Hinrich will join Sternum's Board of Directors.

www.sternumiot.com

Driving 5nm chip design from Europe

A UK chip design house is aiming to become one of the world's major custom chip supply chain services providers. Graham Curren, chief executive of Sondrel, talks to Nick Flaherty about its 7nm and 5nm designs, recruitment and the plans for the company.

There is a company in Europe regularly doing leading edge, large chip designs on 7 and 5nm process technologies. Sondrel in Reading, UK, is the largest chip design house in Europe and aiming to be one of the largest supply chain companies for the custom chip business. It already works with the major foundries, but is now offering support all the way from concept to finished, packaged chips.

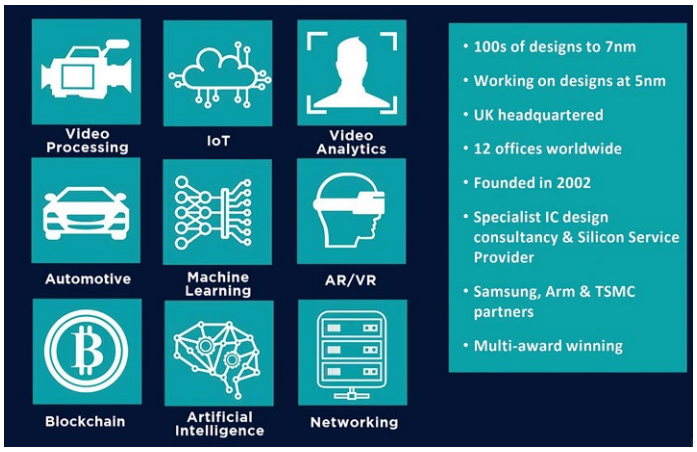
"We do design and ASIC supply, focussed on digital designs in newer technologies, principally FinFET. In terms of the history we are 19 years old with design centres in China, India, Morocco and the UK and with a sales and technical support office in the US in Santa Clara," saod Grham Curren, founder and CEO of Sondrel

"There's a lot of design services companies in Europe but we are bigger with over 200 staff and have the capability to do designs. Everything is below 28nm and the vast majority below 16nm and there isn't really anyone else in Europe with that capability and not many in the world – GUC, Socionext, Faraday Technology and Alchip," he said.

The self-funded company started in 2001 and has 18 percent backing from design tool vendor Mentor Graphics, which is now part of Siemens.

"Our first customer was in Germany and providing supply chain services was in the business plan but I wanted to do it with a strong offering," said Curren. "When you are self funded, you are susceptible to having to do what is presented today rather than set your agenda, so it takes longer to get where you want to be. A few years ago we felt we had the supplier relationships to provide a real ASIC supply service. Rather than supplying companies like GUC, we would be becoming a competitor."

"Today we are still in the early stages, we are not looking at shipping chips yet, we are designing. With a big design win,



execution, from concept to architectural design, including architectural studies, all the way through, its an 18 month to 2 year design cycle to prototype and then another year to production. These are big chips, so it takes a lot of time and effort," he said.

Independence is a key advantage for the company.

"We are able to ship to TSMC, Samsung, GF as we are on their approved lists. It's a balancing act – each one has its own strengths, its own markets, and we have to be careful not to go behind someone's back," he said. "For example, if we are working with TSMC on a project and suddenly switched to Samsung someone would be upset. We work in an open and trustworthy manner which allows us to be independent. That's particularly important in EDA and IP. There is a lot of anti-competitive practice in this industry, whether its bundling, discounting or doing favours for people, and there is a line between trust and partnerships and anti-competitiveness."

The challenge is scale, as most customers deal with the foundries direct, he says.

"If you look at TSMC for example, 95 to 98% of their business is done direct. GUC is arguably their biggest supplier at $400m, that's 1 percent of their business," he said. "So like all distributors we have to add value – and we have to make margin, so either the foundry sells to us more cheaply or we add value. That's the design part and the supply chain management."

"There is definitely a minimum level at which it is sensible to do your own chip and not many companies pass that threshold. From a business point of view it doesn't often make sense to maintain a chip design team."

This is part of understanding what the core business is of a company, and some areas of Europe are better at this than others.

"What is the core value of the company? Is it the chip design, software, system understanding, the customer relationship, the brand? A lot of companies get sucked into doing the chip themselves without realising the difference," he said.

"Israel is better at seeing where they add value, the UK has always been very much 'do it yourself', in Germany it varies by industry. As an ASIC supplier we have the expertise and experience to make use the chip works first time and hits the performance requirements."

The company saw a boost in May 2017 with the acquisition of the ImgWorks design team from Imagination Technologies, a neighbour in the UK.

"When Imagination decided to get rid of parts of the com-

pany we discovered they were selling the ASIC part, so we bid for it, the management liked our proposal, they had some contracts that needed finishing, so the team joined us. Its not easy to find a bunch of god engineers and it fitted exactly with what we were doing. They were already working on FinFET designs, multiprocessor, CPU, GPU, complex interfaces so it fitted with our core business. They were 80 people, we had 120 at the time," he said.

"There's always been the two hardest things for growth have been cashflow and the ability to hire engineers, and that more so. It continues to be challenge.

"There are clear differences between Europe and Asia for hiring," he says. "Everyone in Europe is over 40 – in Asia under 40 so when you put them together it works well. You have to address both of those. In Europe we are increasingly hiring graduates. We had successful graduate recruitment in China, some in Morocco and we want to do more there and this year is the first year we have tried to do that in the UK.

"We have recruited five graduates in the UK, with assessments even before and after lockdown. The university programmes tend not to be related to the real world so we are looking at other ways for our 2021 graduate programme," he said.

In the film, The Imitation Game, potential recruits were found by setting code to be cracked. Hidden in Sondrel's recruitment brochure about the programme in 2021 is a code, which, if cracked correctly, fast tracks the applicant to the shortlist.

"Last year, we had over 150 applicants so cracking the code to be shortlisted can really make a difference to their chances. And we've not made it easy to find the code because in real life working at Sondrel, they need observation skills as well as analytical skills. A CV won't show us that, hence the cunning code challenge.

"One of the key things we look for in candidates is their ability to generate ideas and challenge assumptions. Exceptional engineers look at every problem with fresh eyes, understand what the hard limits are, and generate numerous ideas on how to push the boundaries where the limits are not so fixed."

These graduates will work on leading edge designs for companies around the world.

"We don't do much large digital chip designs for European customers, the mask costs are huge at 7nm and 5nm which puts people off – 12nm is a lot more manageable," said Curren. "We have a very broad range of designs from 22nm to 5nm and we have just completed a test device at 5nm, at 6nm and a lot of stuff at 7nm.

"There's quite a few startups around AI and ADAS automotive, a lot at 12nm and 16nm and fully depleted silicon (FDS). For FDS the drivers are mask related. The power is important but they are not stretching the power and using back bias to get the lower leakage current."

This means the business is global. "We see lots of prospective customers in China, lots of companies in the US, I guess the market hasn't changed that much in Europe, its mostly Israel, with systems companies in aerospace and automotive," he said.

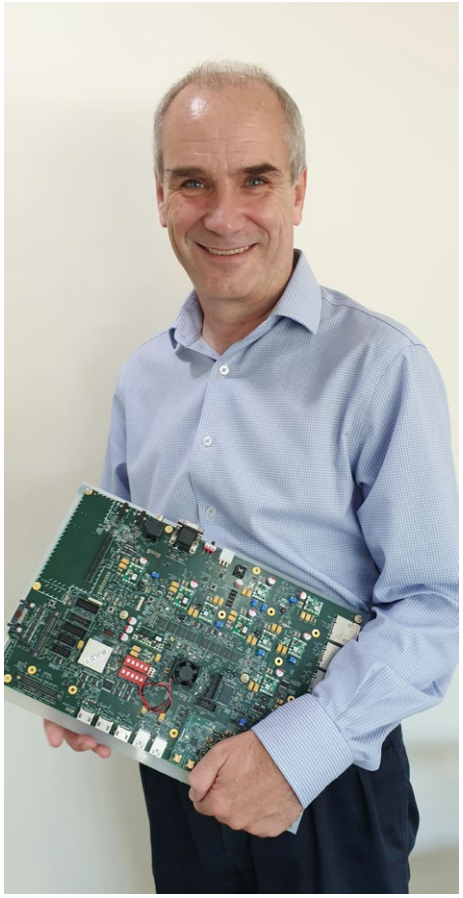
"The real challenges are around the complexity. We think of ourselves as a system integration company – you are not doing that much gate level designs, it's about the integration of CPU, GPU and interfaces so the problem is different."

Memory design

"We taped out an MRAM design with ARM, Samsung and Cadence about a year ago – there has been a lot of take up on that for IoT designs so there is a lot of interest, but MRAM designs tend to be smaller and we are more focussed on larger designs, typically with 30 to 40 people on the team," he said. "The biggest design we have done was 600mm2 in 7nm for a networking chip but that was very regular. We taped out a 400mm2 design that was very irregular in 16nm with many CPUs and GPUs, that was harder to implement with the network on chip (NOC) [interconnect].

Future plans

"We have our sights very much set on supplying full supply chain for these complex digital designs and we think there are few people with the technical depth to do that. Nothing is holding us back today," said Curren. "We are still doing services, both are valuable, some people want one and some want the others. Companies such as GUC and Socionext typically have 5 to 10 times our revenue per head so the way forward is to increase the revenue per head, and the key to that is the technical competence."



The US China trade-war is just part of the cost of doing business, as are the changes in Europe as the UK leaves the EU.

"China is part of our business and we do need to manage those risks, but it is also a good potential, as it's a huge market – there's more new companies that we are talking to about projects, so its not an area to turn your back on," he said. "We are seeing issues with European funding, for example as a UK company whether we will be able claim French tax credits, so there are barriers to European business as well."

Advancing 5G test using software-defined radio techniques

By Gunes Karabulut Kurt and Selahattin Gokceli at the Wireless Communication Research Laboratory, Istanbul Technical University

The wireless communications research laboratory of Istanbul Technical University conducts research projects on many different signal processing topics, including fading channels, multicarrier communications systems, modulation techniques such as orthogonal frequency division multiplexing (OFDM), and advanced wireless communications security issues. Projects may cover theoretical aspects of wireless communication systems, practical test and measurements, or a mixture of the two.

In most cases, a research project takes place using a common, reliable platform to share expertise and foster collaboration easily. Over time, the research laboratory has opted to use a variety of open-source tools and mathematical modelling languages to simulate and prototype software-defined radio (SDR) projects.

One recent project focused on investigating a communication waveform technique used for 5G. With the considerable increase in mobile device adoption, particularly as 5G deployment gains traction, the communication waveform used is increasingly important. The waveform technique needs to provide acceptable levels of quality, even when the synchronization requirements are less strict. This requirement is particularly the case where communications density is high and limited resources need sharing across a high volume of users.

A popular waveform technique that suits many different types of application is OFDM. Unfortunately, OFDM has a reasonably tight synchronization requirement, has high spectral side lobes which make it exhibit inefficient spectral properties, and has a high cyclic prefix overhead. These limitations make OFDM not suitable for use in 5G, so some new waveform techniques, one of which is universal filtered multicarrier (UFMC), has been identified as a likely alternative. Compared to OFDM, UFMC has a simpler synchronization architecture and is easier to predict in dense network scenarios. As such, UFMC is now the leading candidate for use in 5G.

The UFMC v. OFDM project

The UFMC project would compare and contrast UFMC against OFDM. It would require several critical technical test environments, including I/Q imbalance, dealing with synchronization issues, and peak-to-average power ratio (PAPR). Rather than selecting the standard set of open-source tools, the research team highlighted the need for a more flexible and reliable platform on which to conduct the tests. The decision resulted in selecting a platform from NI. Of particular note were the depth and breadth of the NI ecosystem and the comprehensive training that was available. By selecting NI's platform, the research team saw the ability to achieve faster results and increase levels of productivity.

The new UFMC test platform hardware consisted of two NI USRP-2921 SDR devices, an NI PXIe-5644R transceiver module, and an NI PXI-6683H timing module. See Figure 1. Test software included LabVIEW, the Modulation Toolkit, and the Spectral Measurements Toolkit. The USRP-2921 devices are SDR nodes, one configured as a receiver and the other a transmitter, that can operate in the 2.4 GHz to 2.5 GHz and the 4.9 GHz to 5.9 GHz frequency range with a bandwidth up to 20 MHz. Synchronization, absolutely crucial for investigating multicarrier waveforms, was achieved with an NI PXI-6683H timing and synchronization module. The timing module uses a TCXO to deliver a 10 MHz clock signal with an accuracy of 1 ppm. A PXIe 5644R receiver channel measured the spectrum of the modulated transmitter signal.

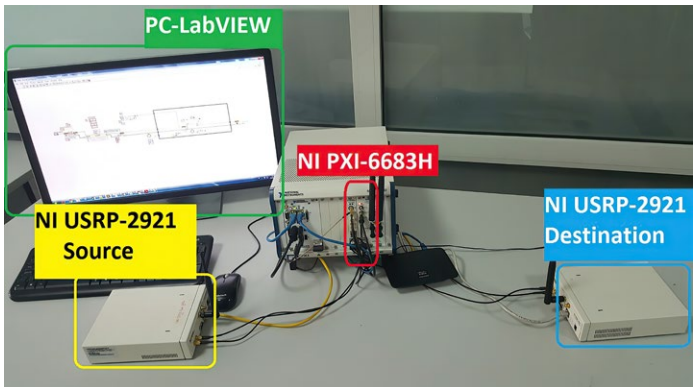


Figure 1 - The complete testbed setup (source Wireless Communication Research Laboratory, Istanbul Technical University)

Using LabVIEW, two different transceiver applications formed the testbed for the OFDM and UFMC performance measurements. A source virtual instrument (VI) and a destination VI realized the transmitter and receiver functions.

UFMC v. OFDM results summary

Overall, the comparison of the UFMC and OFDM testbed results indicated that UFMC is a clear winner.

In just one example, the error vector magnitude (EVM) was 18.50 % compared to 34.96 % for OFDM when measured with a 1 dB transmit power. Sidelobe suppression of a 5G waveform, as mentioned earlier, is a crucial aspect, and the research team created a simulation environment using LabVIEW. Based on the results, the research team proposed a slightly modified form of UFMC to facilitate even better sidelobe suppression.

This Precoded-UFMC has spectral containment properties make it suitable for a variety of different 5G applications than UFMC alone.

Another set of tests focused on the potential to reduce the peak-to-average power ratio. The research team has concluded there might be an opportunity to lower PAPR. In any multicarrier system, where there are lots of subcarriers, the amplitude of each independently modulated subcarrier varies dynamically. If all carriers achieve their maximum value all at the same time, the output power envelope will significantly peak, resulting in a massive peak compared to the average for the whole system. The occurrence of such a rise can have an impact on the RF power amplifier performance, so anything that can reduce the overall PAPR is valuable. The method of PAPR reduction proposed by the research team indicated a PAPR reduction of 1 dB, as illustrated in Figure 2.

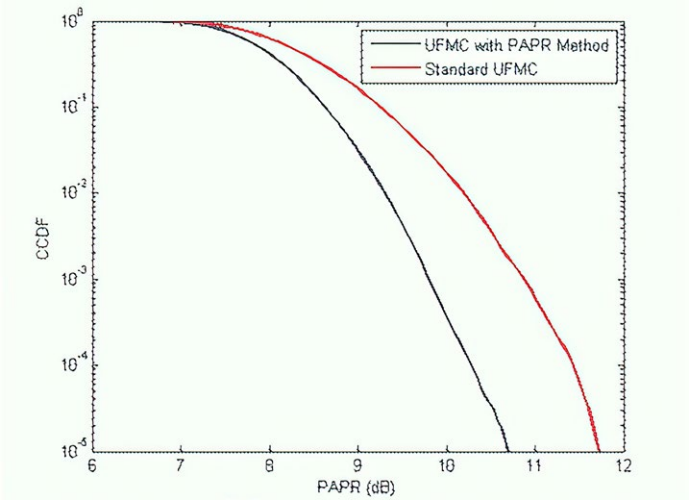


Figure 2 - Performance of the proposed PAPR reduction method for UFMC (source Wireless Communication Research Laboratory, Istanbul Technical University)

The research team believe that the results of their UFMC and OFDM comparison project will provide valuable insight into the challenges involved in 5G deployment from the aspect of modulation techniques. The results indicated, for the first time, the use of SDR to implement the UFMC waveform and to investigate the real-time challenges of 5G in deployment. The NI software-defined radio equipment significantly aided the wireless research team to achieve the project goals quicker and with fewer complexities.

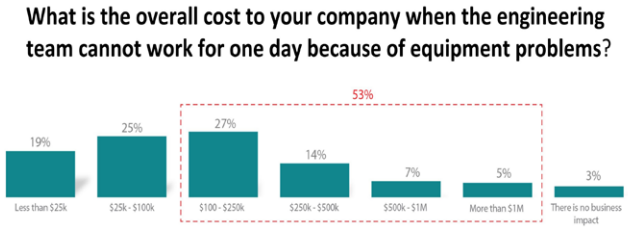
The Business Impact of Test Equipment Issues

By Ted Burns, Global Director of KeysightCare at Keysight

Testing is a crucial element in the development cycle of any product, solution or service. No matter what sector a business operates in, it will be subject to specific frameworks, standards and regulations governing testing to assure that its products operate effectively and safely.

But the demands placed on R&D engineers have proliferated in recent years, as the landscape becomes more complex with increased time pressures to bring new products and services to market. There are growing incidences of workflow issues related to use of test equipment; whether because equipment has been misconfigured, test engineers lack the necessary training or knowledge, or environmental interoperability or equipment failures cannot be debugged and resolved quickly enough. This materially adds to the intensity of development cycles.

The survey, conducted by Dimensional Research, consisted of 305 R&D engineers from a range of global organizations, including those in the technology and telecoms sectors. The results were clear – more than 90% of companies have experienced a revenue loss because of preventable delays related to test and measurement equipment. For more than half – 53% of companies – that revenue loss amounts to more than $100,000 of waste per day waiting to resolve technical support issues. Loss of time results in loss of money. So how did this situation arise, and what can be done about it?



Test cases increase by orders of magnitude

To remain competitive, companies must eliminate obstacles that cause delays. Design

complexity, frequent design changes and unfamiliar new technologies with rapidly developing standards can force engineers to troubleshoot technical issues on the fly. Product or service development schedules do not account for time required to troubleshoot and debug such issues. To maintain aggressive schedules, they must do so without the right resources, training or expertise.

In the past, engineers would develop a deep familiarity with test equipment design and test standards. Today, they simply do not have the time to research how their test equipment works, how to optimize a configuration, or how to troubleshoot a complex setup to achieve the results specified in specs and industry standards. And with constant pressure to shorten new product development life cycles, market-leading companies cannot afford to wait days to access technical support experts for issue resolution.

As a practical example – in the past, R&D engineers would routinely download technical manuals to learn the inner workings of a product. They had an intimate understanding of the operation, test configuration and use cases.

Today, test matrices have spiralled vastly increasing test cases both in number and complexity. It is no longer feasible to master those cases while meeting customer expectations and time-to-market goals. R&D engineers rely on automation and sophisticated software to achieve speed. But if a test case fails, troubleshooting can be difficult and time-consuming.

Escalating business issues

Keysight's survey found 91% of respondents said they experience business-impacting issues related to calibration, technical problem resolution or equipment performance. In addition, 46% reported their business suffers within the first day with test equipment not functioning to expectations. Further, 72% said calibration issues affect their business within three days, yet and average of 65% said that the time it takes to calibrate and repair equipment is not fast enough, or predictable, lacking the needed Service Level Agreements (SLAs) to help mitigate schedule slips.

All in all, 97% of respondents had experienced problems with test equipment which resulted in project delays. This included 63% of respondents who said they had experienced test equipment failure, which then required repair. 56% said they had experienced problems due to test equipment being improperly set up – problems with configuration, cabling and so on. 50% had experienced equipment being out of calibration, and 46% had experienced problems with improper tool use – employees making mistakes, lack of training, programming issues and so on. 29% had experienced delays waiting for new test equipment to be set up.

Direct impact on the business bottom line

All these problems can have critical business impacts. If the R&D engineering team cannot work for a single day because of problems with test and measurement equipment, the cost can be severe. 53% of survey respondents said that the cost of this was $100,000 a day or more – and 5% said it was more than $1 million.

In addition, 91% of respondents explained that such issues had a tangible and material business impacts, from reducing product yield (cited by 53%), to products being rejected by buyers (47%), to increased product returns (45%) and even product recall (28%). Such occurrences are costly. And those costs can linger for multiple quarters as firms struggle to regain their customer's confidence as a quality supplier.

A material difference

In this complex environment, it is not surprising that 60% of reported problems relate to incorrect test equipment setup or improper use. These factors alone suggest that as product designs and test solutions increase in complexity, access to technical support expert resources and information grows more important. When problems arise, companies need fast, reliable guidance to troubleshoot and resolve issues. Having the right technical support can reduce time to resolution when troubleshooting and diagnosing advanced measurement techniques.

Note how often test professionals require assistance: 95% said they need help every month, and 59% said they experience six or more technical support issues each month. Ninety-four percent of electronic test professionals said they need committed, fast-response technical support.

What is even more apparent is that expectations and demands upon technical support have dramatically changed. Nearly half the professionals surveyed said existing technical

support models do not meet their expectations and business needs. They said technical support models need to be highly knowledgeable, immediately accessible with faster, committed response times 68% thought they could save multiple days each year using a priority support model.

Modern testing and development require a modern technical support approach, tailored to meet the needs of agile, connected design and test. Each generation of technology increases this complexity. Test and measurement companies must respond and resolve technical support issues faster to help their customers stay ahead of test standards, test methodology and equipment maintenance. Overall, priority access to test experts with committed response times and proactive notifications, calibration services and repair facilities with committed turnaround times and a state-of-the-art digital experience, can make a material difference, whether engineers are designing or manufacturing a product. The right support can enable a team to deliver a quality product on schedule and help avoid the design or production issues that result in costly delays.

Measurements for optical devices Current Drivers

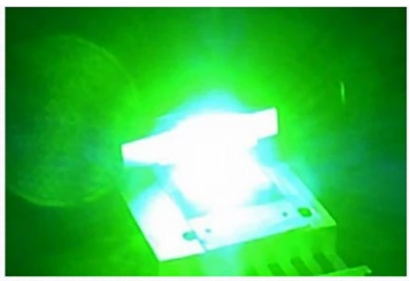
By Andrea Vinci, Technical Marketing Manager at Tektronix/Keithley in Europe

LED based devices are literally everywhere, not just in lighting applications. The supplying circuit for LED devices is a challenge designers, squeezing into small sizes high efficiency AC-DC converters. To control the energy flowing towards the device the typical design exercise is to control a modulated signal for the forward current.

Today the market is full of ultra-packaged drivers with different capabilities according to the application. In same case the same driver can be used broadly for LEDs but also for other infrared light sources like lasers. The key parameters for selecting a driver package capable of sustaining the average optical power are the amplitude of the voltage, the current pulse duration and the current pulse repetition rate

Complex projects need to be more selective, for instance also able to withstand high operating temperatures, or to support protection modes and temperature controller input so as not to overheat the device it is driving.

If the driver is continuous or quasi continuous, the current range and the supported frequency range must be in line with your design requirements too. If in pulse mode, also aspects like pulse duration, duty cycle and peak current often require to be finely adjustable.



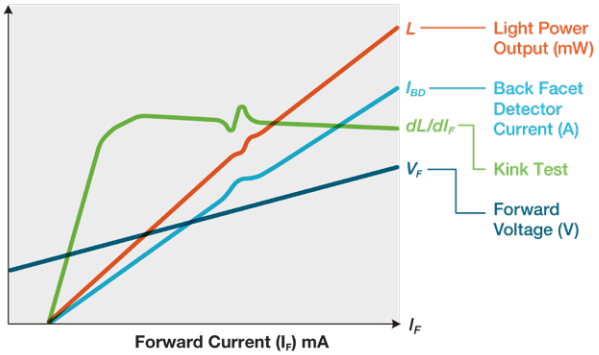
Among lasers, VCSELs are widely spread in the datacom market for high speed optical communication.

These same devices, with different specs and characteristics are being used within compact package arrays also in the consumer market, the industrial market and the mil/gov market for applications related to “sensing”, spanning from smartphones face recognition to Lidar in self-driving cars.

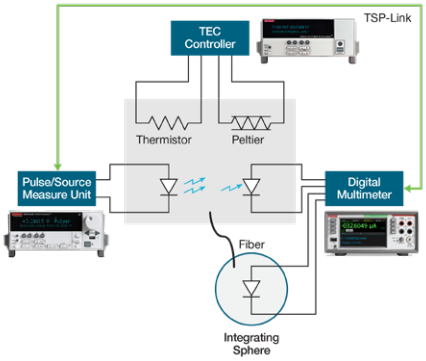
A solid-state 3D laser scanning system to detect objects around the vehicle and map their distances have different test requirements depending on the contest and the vehicle. Same thing for requirements related to the pulsed light source driver to associate.

These devices require to be tested in combination with a specific driver solution, which has to be able to supply precise and controlled current pulse trains. This is basically necessary to ensure that the optical power level to generate does not cause failures to both the optical device and all other embedded components in the system.

The L-I-V (light-current-voltage) characterization of optical devices requires a test with a current source to the optical device while measuring light, power and voltage. Depending on the level of current sourced, the right instruments can measure and log accurately the forward voltage, the reverse voltage and the leakage current of the device. This also allows a trace of the output power characteristic for ensuring efficiency ratings or saturation/thermal run off is within specs.



This requires Source Measure Units (SMUs) that can act both as current sources and current/voltage meters at the same time, with high accuracy and over large span ranges.



Another typical element of an optical devices measurement setup is a fast and accurate oscilloscope, with a low noise front end and a good dynamic range. These optical test and characterization systems stations typically embedded an additional high-speed photoreceiver; this photoreceiver works in a spectral range that can span from DC to the several MHz. The photo receiver provides a trace of the fast pulse edges by converting them into a voltage signal that the oscilloscope can sample and validate, while also verifying rise and fall times are within specs.

Here, a msixed signal oscilloscope with a 12bit ADC such as

the Tektronix 6 series fits well as the low noise front end provides the necessary signal integrity conditions.

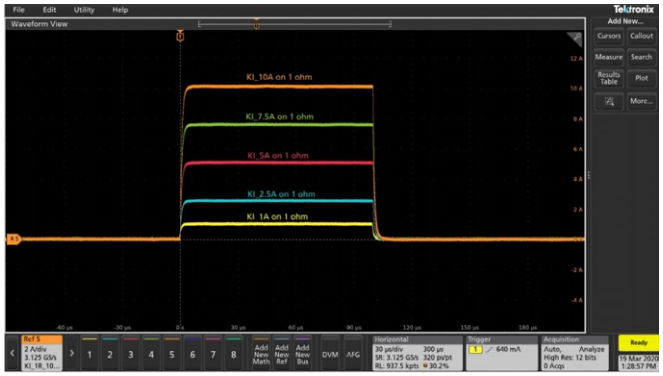
The world of solid-state lasers is really wide and complex and sometimes power laser diodes can be quite expensive devices. Generally, the users of these distinguish between low power and high power but also between standard and “fast”. So how fast is “fast”?

In some applications with 10s or 100s of Watt of optical power (illumination, pumped solid state lasers), a “fast pulse range” means the ability to test from pulses from 20 us to 1ms, and pulses are considered fast when the rise time less than 10us. These pulses have to reach several Amps of current, sometime 100A or even more. In these cases, load deviation during a pulse in the few Amps range can cause some headache with the voltage and present current profile errors.

In some other applications (Time of flight sensing drivers) the current peaks at less than 1A but the minimum pulse width duration can be as low as few nanoseconds.

To add further complexity, the most challenging test situations are typically when the device unmder test (DUT) is on a tiny packaged module mounted on a remote prober/test station. This test does not use the embedded driver on the PCB but a remote current pulse generator probably connected through relatively long wires and pins. An important factor to consider is that the shape of pulses transferred from the driver pulser to the optical DUT strongly depend on the type of connection and voltage/current levels in use.

The photodetector output connected to the high speed oscilloscope demonstrates this. Leaving the current pulser as it is and just changing the path impedance displays different pulse traces on the scope despite the device and the wire bonding stays the same!



The SMUs and the scope and all the rest already use significant space on the test rack aside an already large device tester/sorter. Having the pulser functionalities embedded in one of the mandatory units for integration. The tests also have to be reliable and repeatable when generating pulses in sequence for several DUTs. Customers told us that when they look at their pulses on the oscilloscope, they appreciate changes from device to device. This is because the driver connection path impedance may change and this is affecting the measurement accuracy and repeatability.

Because of the above mentioned problem, customer have to manually tune pulses shape looking at the feedback given by the oscilloscope. Rise time ringing oscillation and distorted flat plateau of the pulse require adjustment of the impedance mismatch from the parasitics.

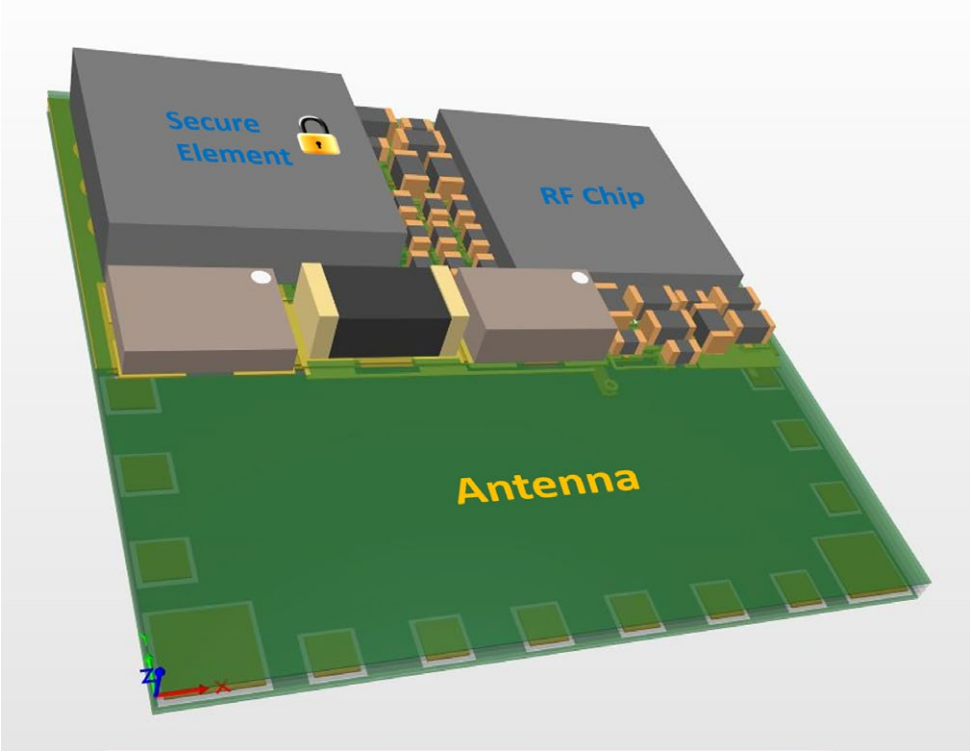
To address this challenge, the 2601B-PULSE pulser SMU combines the capabilities of a 2601B SMU with a programmable current pulse source able to generate 10A precisely shaped current pulsed from DC down to 10us, with steep precise rise time.

Securing systems and data in IOT nodes

By Nick Wood, Sales & Marketing Director, Insight SiP

The explosion of the Internet of Things (IOT) has led to a proliferation of connected devices. These vary from small sensor devices, medical wearables, and connecting existing electronic industrial and consumer equipment. Many of these are connected ultimately to the internet and use one or more wireless channels.

This opens up many interesting possibilities to manage systems and assets more efficiently, but at the same time, such connections become a potential channel for malign attacks. Here we look at the core issues in securing such systems.



Core IoT security issues

The first point to note is that there will never be 100 percent security. Therefore, the key is to analyse what risks one is trying to address, the severity of the possible consequences – and the costs of prevention.

Malicious code

Attacks can take many forms; here we will concentrate on a few types. The first is malicious code being placed on a device. If a hacker has physical access to a device, it is very difficult to prevent them from replacing the software on it. However, what can be done is to prevent them having easy access to security keys that would allow the hacker to further penetrate the larger system. There are several options now for either processors with secure zones, or for the strictest security, including a secure element to hold trusted information.

Physical access is not an easy attack channel. A greater vulnerability comes through the possibility of Over the Air Updates. If a malicious update can be injected, a device can be compromised remotely, with possible serious consequences, particularly in the case of medical devices.

Digital signing of software can be a way to mitigate this threat and provide an end-to-end secure method of ensuring the right update arrives at the end device. This in turn requires a robust key management system, and the ability to store such keys securely on the end device. A secure boot mechanism is then also required to ensure the software is checked prior to being executed.

Malicious monitoring

A different type of risk is that of malicious monitoring data sent from IOT devices. The consequences here could be someone gaining access to sensitive health data from a medical device, but also data from a connected home could reveal if an occupant's absence.

Although wireless links may seem insecure, in general, the point to point links are encrypted with reasonable security. What is harder to be sure of is the security of the end to end travel, via an internet where one has little control. The only way to isolate oneself from this dependency is to provide end to end encryption. This may sound simple, but as with the software signing above, it requires a robust key management and storage process. A hacker can buy a device and analyse it, so any keys or encryption processes need to be secured against spying.

Spoofing

Another more subtle risk is spoofing. Imagine some kind of distributed data collection system. A hacker could create "rogue" devices and inject false data into the system. In such a way, they could create imaginary faults and in extremis cause operators to take harmful actions based on false information. The rogue device could just be a genuine article rigged to generate certain data.

To prevent this kind of attack, a robust authentication process would be required. Data could be signed via individual device keys, so that any rogue devices would not present correctly authenticated data.

Security cost and risk balance

Of course, all of the above does not come cost free. Processes to distribute and manage security keys are complex. A final decision on an IoT security strategy will always have to balance cost and risk when considering the measures to be taken.

Security cannot however be a "last minute add-on". It needs to be designed in from the start, using components designed to provide the security features required. The more tightly they are integrated in, the harder it will be for a malicious actor to enter the system.

Insight SiP are experts in RF circuit miniaturization, System-in-Package and Antenna-in-Package.

How system level noise in digital interfaces can lead to spurious errors in serial Flash memory

By Paul Hill, Senior Technical Marketing Director and Gordon MacNee, Field Applications Manager EMEA at Dialog Semiconductor

System designers often rely on proven, standards-based interfaces that are effective right "out of the box". For example, standard interfaces, such as SPI and I2C, provide a relatively simple and effective way of interconnecting devices from different suppliers. Using this practice, developers can create complex systems using 'standard' technologies.

While this level of integration and configuration is undoubtedly an enabling element that provides considerable advancements in product design, it may also present engineers with unexpected challenges. When a configuration doesn't 'just work', it can lead to confusion, especially if the source of a fault is misconstrued.

Finding faults can be difficult, time-consuming, and costly, even under ideal conditions. Often a fault will be observed by its effects, which will provide enough evidence to allow engineers to identify the cause.

If the cause of a fault is a low-level feature that hasn't been initialized correctly, then finding it could become even more perplexing. Understanding how the initial state of the hardware platform could impact an entire design requires a much higher appreciation of the overall system.

System noise, in all its forms

Any distortion to a signal can be interpreted as noise, which is often more apparent in a communications environment; the signal received is not the signal sent. This direct correlation is relatively straightforward to find, but in some cases the cause and effect are not so easily identified. The challenge is compounded when the fault becomes intermittent.

Today's microcontrollers are designed to deliver reliable operation with minimal configuration. In the case of a serial interface this may include defaulting to high drive currents on the I/O pins in an attempt to counter the effects of long PCB tracks or high capacitive loads. In some instances, this can result in over-driving an interface which, in turn, can lead to derivative effects that are interpreted as errors or faults.

As an example, serial Flash memory devices offer numerous advanced features that ensure reliable operation and allow the device to be interrogated. This can include noise filters, advanced adaptive programming, and erase algorithms that manage cell margins. Some manufacturers also include error-correcting code (ECC) in the storage elements, saving additional meta data with each write operation to allow single or multi-bit errors to be detected and corrected. But this ECC fix will not help when noise taints the basic message transaction on the communication interface bus.

Noise on the SPI interface can be misinterpreted as additional clock pulses. As SPI is a clock driven interface, this would have ramifications such as commands being ignored, data being misinterpreted, wrong commands being used, etc. However, noise also carries energy and in some cases this energy can itself introduce errors in the operation of a device.

Charge pumps and overshoot

In most cases, some overshoot or undershoot in a signal can be endured by digital interfaces without consequence. However, the energy under the curve is still present and in some circuits this can be troublesome.

A working example is the charge pump circuitry in serial Flash memory. If the SPI bus signals contain significant noise, there is a chance that the energy in that signal can propagate through to the charge pump and disrupt its operation.

The charge pump in Flash memory is a critical function, as it provides the power needed to change the bias of a memory cell and, effectively, store a logical 1 or 0. The write/erase process is a crucial time in the operation of Flash memory, any disruption to the charge pump during this time can cause write or erase errors and, while these errors may be detected, there is a chance that they will not be evident.

An error of this sort can easily be interpreted as a fault in the Flash memory device. The fact that Flash memory has a finite number of read and write cycles guaranteed by the manufacturer is well understood by embedded designers, but often the importance of providing a clean interface devoid of too much overshoot or undershoot is not appreciated.

As an example, consider the image in Figure 1. It shows healthy cell margins for six Flash devices. Two distinct patterns emerge between cells programmed with data representing logical 1 (2V to 5V) and 0 (>6v). By comparison, the image in Figure 2 shows memory cell margin for three Flash devices that have suffered data corruption caused by overshoot and undershoot on the control lines.

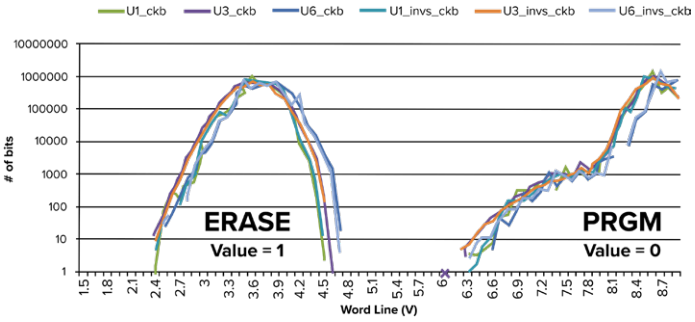


Figure 1: This image shows good cell margin separation data for Flash memory that has been programmed and erased.

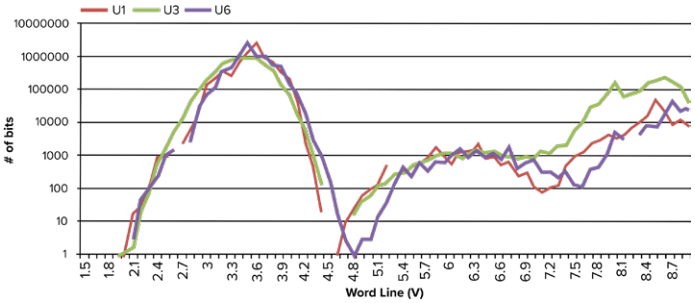


Figure 2: This image shows poor cell margin separation data for Flash memory where there has been significant noise present on the SPI lines

Multiple factors can contribute to the noise level, such as frequency of operation, amplitude of the signal, MCU drive levels, and the energy contained in the noise spikes. PCB design and cross talk between signals may also be contributing factors.

The data in Figure 2 shows the effects of excessive overshoot and undershoot on the serial interface. Figure 3 below shows a representation of what this overshoot looks like in a real application.

The result of this noise was erroneous device operation, which showed up as errors in the serial flash memory stored values. Initially, the true impact of the errors was missed, as polling the STATUS register on a less frequent basis reported fewer errors, leading the designer to make incorrect assumptions on the root cause of the failure.

Although this fault appeared as a memory failure, the root cause was not with the Flash devices. This was discovered by engineers by probing the SPI signals and identifying the system noise present. While the noise could be partly attributed to an impedance mismatch present on the PCB track between the MCU and Flash memory, it wasn't the full story.

The source of the noise was actually the MCU interface, which defaults to a high drive level at power-up. The excessive drive was enough to cause overshoot and undershoot on the SPI lines, which in some cases can be misinterpreted as signal transitions, leading to read or write errors. However, in this instance it was found that the overshoot held sufficient energy to disrupt the Flash charge pump, which was in turn causing the errors.

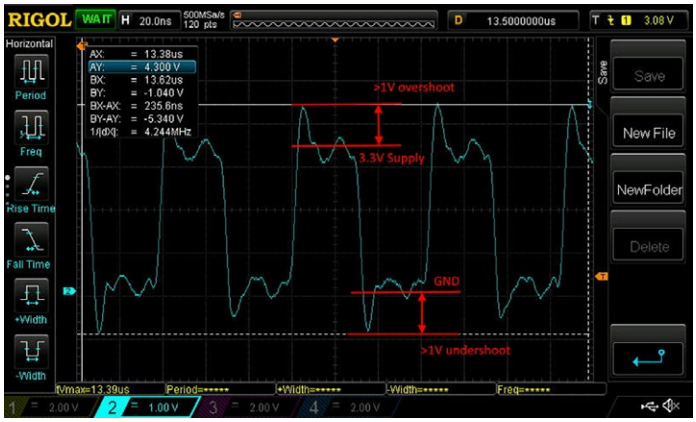


Figure 3: This trace image clearly shows the overshot and undershoot present on the SPI lines was resulting in a peak-to-peak voltage of 5.65V, which exceeds the absolute maximum value documented in the Flash memory specification.

In this design, the microcontroller being used provided a configurable drive current for its I/O, which defaults to HIGH at start-up. As the application code did not modify this level during initialisation it remained high in normal operation. The impact of this may not be apparent for other devices on the SPI bus, as digital interfaces are typically designed to be robust. The sensitive nature of Flash memory, the need to operate at much higher frequencies, and in particular, the operation of the charge pump, made the memory susceptible to overshoot/undershoot. This led to erroneous operation that was initially misinterpreted as a fault in the Flash memory device.

Reducing the drive current, through firmware, reduced the overshoot and undershoot to effectively zero (Figure 4), and in turn resulted in error-free operation of the Flash memory.

Figure 4: With no overshoot apparent, the serial Flash memory's charge pump was able to operate correctly and provide reliable functionality

The nature of the fault showed that the Flash device was making every effort to compensate for the effect of the error, which was the excessive system noise on the SPI interface.

The cause was actually a design feature of the MCU used, which defaulted to an operating mode that, in most situations, would be entirely acceptable. The combination of a high drive output and imperfect PCB inductance created a condition that resulted in intermittent failure. Reducing the drive output on the MCU, through a simple firmware change, solved the issue.

Conclusion

System noise can easily be dismissed when there is no apparent impact. Intermittent errors are particularly difficult to locate under optimal conditions, but when the errors are misinterpreted it makes the challenge even harder.

Overshoot is possibly the least apparent form of system noise, but as explained here, its impact can be significant. Flash memory is a reliable technology but one that depends on a carefully designed interface. Excessive noise on the serial interface has the potential to propagate through to the charge pump circuitry, detracting from the operation of the programming and erase circuitry.

In this instance, replacing the Flash memory and assuming the issue was solved could have resulted in products going to market that were likely to fail at some point.

Instead, by examining the default settings, engineers were able to improve programming and erase consistency by a significant factor, with effective endurance jumping from an unacceptable ~20K cycles before errors were detected to over 2.5M cycles with no errors and no requirement for supplementary error detection and correction routines.

What may appear to be a hardware fault could easily be fixed through software. What looks to be a failure in a one component could be traced to an incorrect configuration in another component. Even though default settings are meant to help, they should be verified.

Optimization of these settings can lead to significant improvements in system performance and reliability.

HBM to 4.0 Gbit/s and beyond

Frank Ferro, Senior Director of Product Marketing, IP Cores, Rambus

HBM (High Bandwidth Memory) has been available since 2013, developed to address a need for greater memory bandwidth than DDR could offer. It presented a new memory choice that could deliver tremendous bandwidth, with great power efficiency, in a very compact footprint.

Originally developed for graphics cards, HBM has since been adopted for other bandwidth-hungry applications, such as AI accelerator cards used for model training, high-performance computing (HPC) and 5G base stations.

These applications themselves are driving the evolution of HBM. With the memory bandwidth available via HBM, 5G base stations can carry out more processing at the edge of the network, instead of merely passing data through to cloud compute resources. This capability enables new services and applications, thereby heightening the demand for greater bandwidth in an ongoing cycle.

Similarly, refining AI models to increase accuracy, and hence usability, requires exponentially more training data to be processed. In applications such as autonomous vehicles, the increased utility of Advanced Driver Assistance Systems (ADAS) will lead to more miles driven using the systems, and more data for refining models, which needs to be processed as fast as possible to improve the driving capabilities of autonomous vehicles.

While there may be no such thing as too much bandwidth, the application of HBM is, due to its implementation complexity, primarily tilted toward high-value applications.

A wide interface and stacked memory

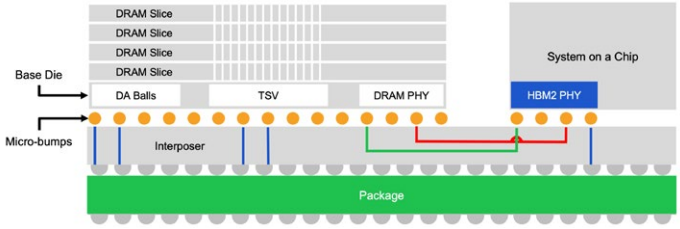
Even today, the latest version of HBM2E runs at a "relatively low" data rate of 3.6 Gbit/s per pin but achieves very high bandwidth through the use of an extremely wide interface.

Specifically, the GPU, AI Accelerator, or network ASIC Connects to its associated HBM DRAM through an interface of 1,024 data lines. With command and address (control plane) lines, the number of connections grows to about 1,700.

This number of connections is far more than can be implemented on a standard PCB. Therefore, a silicon interposer is used as an intermediary to directly connect the processor and memory. Like within a logic chip, finely spaced data traces can be etched in the silicon interposer to achieve the desired number of connections needed for the HBM interface. Since the processor and memory sit next to each other and atop the interposer, this is referred to as a 2.5D structure.

The HBM memory itself consists of a 3D stack of SDRAM die. Up to 8 in HBM and HBM2, and 12 in HBM2E. This enables a HBM system to achieve a very high capacity in a compact area.

The combination of these two innovations, the use of an interposer to implement an ultra-wide interface, and 3D memory device stacking, is referred to as a 2.5D/3D architecture.



This design not only delivers the desired high bandwidth, it also minimises the path length between memory and processor. The downside is that the assembly of these structures is a more complex endeavor than with conventional memory architectures.

The evolution of HBM

The roots of HBM can be seen in die-stacked memory released by Toshiba and Hynix Semiconductor (now SK Hynix) for NAND flash in 2007. Die stacking was subsequently used in DDR3 by Elpida Memory (acquired by Micron) for the first 8GB DDR in 2009.

Adopted as a JEDEC standard in 2013, HBM operates at up to 1 Gbit/s delivering 128 GByte/s per HBM memory stack.

The next generation, HBM2, adopted by JEDEC in 2016, pushed the data rate to 2 Gbit/s and 256 GByte/s bandwidth per HBM2 device. State-of-the-art accelerators like Nvidia's A100 Tensor processor employ six HBM2 memory devices to deliver over a TBytes/s of aggregate bandwidth.

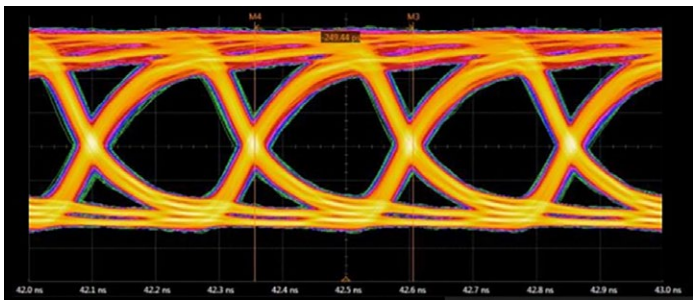
In late 2018, HBM2E was announced by JEDEC. It extended performance of the architecture to 3.2 Gbit/s and 410 GByte/s per device. HBM2E memory devices can support stacks of up to 12 DRAM.

Rambus, in partnership with TSMC, Hynix and Alchip, has

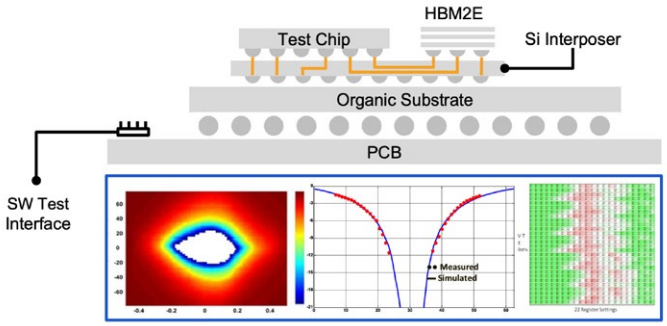
raised the HBM2E data rate to 3.6 Gbit/s, and demonstrated performance in silicon to 4 Gbit/s. At 3.6 Gbit/s 'per pin', a HBM2E memory device can achieve 461 GByte/s, and a next-generation AI accelerator with six such devices can achieve a whopping 2.765 TBytes/s of bandwidth.

Design challenges

The biggest challenges are the physical design between the interposer and the DRAM. To get the HBM2E system to work correctly requires investing a lot of time optimising that channel. The thousands of data lines running through the interposer are single-ended signals. So, they are very susceptible to crosstalk and insertion loss. This channel is also very resistive because it is going through silicon. The task becomes one of balancing all the parameters that affect signal integrity, such as managing the crosstalk, resistive channel, insertion loss, and reflections. Significant simulation work is required to make sure that those signals are going to be nice and clean.



Rambus understands these challenges well. It has long focused on signal integrity and has worked closely with numerous customers not only on integrating its interface (PHY and controller) designs but also assisting with interposer channel and package design. These elements are seen as a significant design risk as it is often the case that chip engineering teams do not have deep experience of HBM design. As part of Rambus' standard interface license, a reference design is provided which includes the interposer. Support is then provided to help the customer iterate through that design to meet actual system requirements.



The industry needs more speed – as well as addressable memory if it is to keep up with the exponential growth in workloads such as AI/ML training.

A little over a year ago, the largest training model had 10 billion parameters, now the largest is over 100 billion. For the past decade, the pace of increase has been 10X per year so AI/ML training models with a trillion parameters can be expected in 2021 or 2022. The training models of today can take days or weeks to run. More bandwidth and capacity are key to accelerating that process.

Taking HBM2E to 3.6 Gbit/s puts aggregate system bandwidth within reach of 3 TB/s, a huge step up over the best-in-class systems of today. The industry is working on designs that continue to push the limits of performance with innovative PHY, controller, and interposer designs.

Whatever the challenges, it is clear that new generations of HBM will emerge to deliver the bandwidth at the network edge and in the data center, to accelerate the next wave of technology.

Persistent Memory for Artificial Intelligence and Machine Learning Applications

Data centers should take advantage of persistent memory to eliminate bottlenecks and accelerate performance in Artificial Intelligence and Machine Learning technologies.

By Arthur Sainio, Director of Product Marketing, SMART Modular Technologies

In today's enterprise datacenters, limited memory capacity and the input/output (I/O) performance of mass storage are the two biggest causes of bottlenecks. These two pain points have historically been perceived as different computing concepts: memory is a temporary store of code and data to support a running application, while discs and other persistent storage hold data on a long-term basis. When an application needs to access data from disc (which happens frequently with large data sets that cannot be held in memory), the slow access imposes a significant penalty on the application's performance. The introduction of persistent memory has marked a turning point in the traditional data center memory and storage hierarchy through the possibility of a new unified hyper-converged architecture that dramatically accelerate enterprise storage server performance.

The Growth of AI and ML Applications

The explosion of data has resulted in huge growth in Artificial Intelligence (AI) and Machine Learning (ML) applications, but traditional systems are not designed to address the challenge of accessing these large data sets. The key hurdle for AI and ML applications entering the IT mainstream is reducing the overall time to discovery and insight based on data intensive ETL (Extract, Transform, Load); and checkpoint workloads. AI and ML create highly demanding I/O and computational performance for GPU accelerated ETL. Varying I/O and computational per-

formance is driven by bandwidth and latency. The high-performance data analytics needed by AI and ML applications require systems with the highest bandwidth and lowest latency.

According to the International Data Corporation (IDC) Worldwide Artificial Intelligence Spending Guide, spending on AI and ML systems will reach $97.9 billion in 2023, more than two and a half times the $37.5 billion that will be spent in 2019. In turn, the data processing required needs to keep up with this expansion will be exponential in growth. Conventional memory solutions today lack the vital component to answer this push: non-volatility, even as parallel architectures are being designed to answer future data needs. However, while these architectures are being refined, power losses could cost data centers millions of dollars. Hence the immediate need for non-volatile memory.

Moving Non-Volatile Memory Closer to the CPU

Checkpointing is a process where the state of the net being trained is stored to ensure that the result of the learned data is not lost. Checkpointing is a particular challenge for AI and ML applications because it wastes processing capacity and burns a lot of power, without directly offering a benefit to the application itself. Processing in other nodes may also be halted when writing data to a central store. The operation is also write-intensive, compounding the problem in some situations as conventional storage such as hard drives are inefficient when data is written to them.

As checkpointing to a central memory can significantly reduce the speed to insight in AI and ML applications, engineers are moving non-volatile memory closer to the CPU to minimise the impact of this essential process. This produces a better balance between data and compute, enabling the system to deliver the overall production needs.

NVDIMMs in AI and ML Applications

Persistent memory, in the form of NVDIMM (a Non-Volatile Dual-Inline Memory Module), is being used to increase the performance of write-latency sensitive applications, effectively providing a persistent storage model with DRAM performance. Data centers have a unique opportunity to take advantage of NVDIMMs to achieve the low latency and increased performance requirements of AI and ML applications without major technology disruptions.

When NVDIMMs are plugged into a server, they are mapped by the BIOS as a subsection of persistent memory within main memory. The application is then free to use this persistent memory for high-speed checkpointing. The alternative is the traditional approach in which the checkpointing data is transferred through the I/O stack, over NVMe and then saved to an SSD. This system incurs the latency penalty of the I/O stack and the NAND Flash.

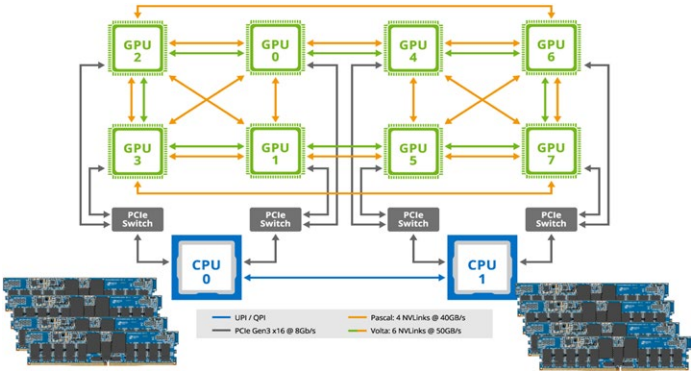


Figure 1. Four 32GB NVDIMMs are used for each CPU providing a total of fast byte-addressable persistent memory.

NVDIMMs are an ideal solution for high-performance AI and ML servers. Data intensive ETL and checkpointing workloads can use the persistent memory region of main memory, allowing them to operate at DRAM latencies (<100ns) and DRAM bandwidth (25.6GB/s).

While NVDIMMs are used to accelerate checkpointing for AI applications, they can also be used for ML to increase performance and protect data being collected by algorithms. GPU configured storage servers run algorithms which are part of simulation and ML. NVDIMMs are used to protect the GPU servers from losing simulation data. Typical algorithm data set sizes vary from Kilobytes (kB) to Terabytes (TB), and lost data would cause a need to restart work. When four servers are configured with NVDIMMs, dataset sizes up to 1TB can use persistent memory, as opposed to traditional storage, to dramatically improve performance without risk of losing data.

The most common method used to process AI, ML and simulation datasets (which all have similar characteristics) is for the datasets to come through the network via InfiniBand or Ethernet into the AI/ML server then cached into the SSD to eliminate the risk data loss. Portions of the datasets are then moved to DRAM by the GPU where the calculations can be performed. An example of this process would be performing calculations on a dataset to determine if the data represents a picture of a dog or cat. Once the calculation is completed the response is sent back out to the network. If there is a system crash during this process all calculations are lost. By switching to NVDIMMs, this process can be dramatically streamlined. There is no need to cache the incoming datasets into the SSDs. The datasets can be moved directly to DRAM where the GPU can immediately start its calculations. The response to determine if a specific dataset represents a picture of a dog or cat can occur magnitudes faster. At the same time, there is no risk of losing the datasets or the calculations because the NVDIMMs are persistent.

NVDIMMs are not only well suited for AI and ML applications, they can also be used in financial applications commonly referred to as FinTech. FinTech applications demand high performance (reducing latency and increasing transaction

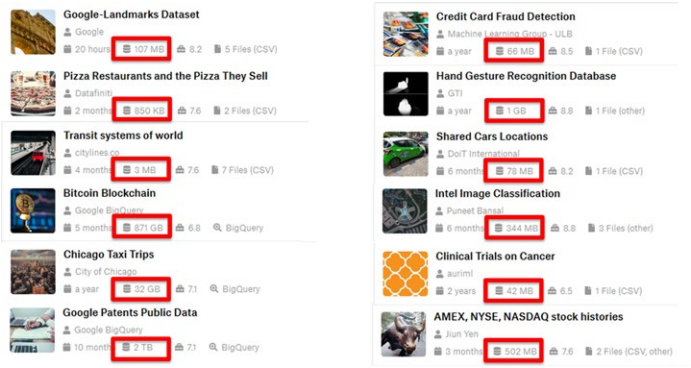


Figure 2. Examples of machine learning datasets ranging from 850KB to 2TB.

rates) because time is money. Processed transactions need to be logged synchronously before the next transaction can be started. This synchronous function, while critical for auditing, also creates a significant bottleneck for many systems, slowing the transaction velocity. By utilising NVDIMMs the current process of logging data to SATA or NVMe SSDs can be eliminated. Instead of sending logging data through the I/O to the Flash SSD, the logging data can be put directly into high-speed DRAM made persistent with the use of NVDIMMs. The NVDIMMs enable the system to begin the next transaction with the confidence that the previous transaction is logged to a secure location with no risk of data being lost.

While NVDIMMs have been around for more than a decade, the benefits of using this type of persistent memory for AI and ML applications is still being investigated by various sectors from banking and retail to discrete manufacturing, process manufacturing, healthcare and professional services. The support ecosystem for NVDIMMs including the operating systems, hardware enablement and JEDEC standardization was the result of many companies working together to adopt persistent memory. NVDIMMs are intersecting with the growth of AI and ML to provide an ideal way to increase system performance.

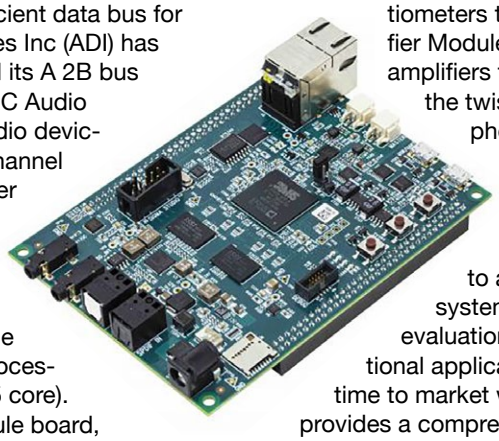
Multi-channel 10bit digitizer with up to 4 GS/s sampling

Flexible 10bit PCIe digitizer with software-configurable channel count and sampling rate up to 4Gsamples/s Teledyne SP Devices in Sweden has developed a modular data acquisition board with configurable channel count and sampling rate. The PCI Express digitizer complements the previously released ADQ8-8C by offering a higher sampling rate and software-selectable two- or four-channel mode of operation. The high channel density, flexible mode of operation, and open FPGA architecture make it suitable for large-scale physics installations and Original Equipment Manufacturer (OEM) product integration. The programmable analog front-end (AFE) supports multi-purpose operation and can therefore be used with a wide variety of detectors and in applications such as particle physics, scientific instruments, time-of-flight applications, and more. “ADQ8-4X extends our 10-bit product portfolio by offering a flexible analog front-end and faster sampling at two different rates. This helps our customers optimize channel count, sampling rate, and cost,” said Jan-Erik Eklund, Digitizer Product Manager at Teledyne SP Devices. This allows large multi-channel systems to use a combination of different digitizer models with simultaneous acquisitions on a large number of channels distributed over many chassis with a timing alignment of better than 200 picoseconds and extended over time. Additional capabilities of the ADQ8-4X include an AFE with programmable channel count, sampling rate, DC-offset, and input voltage range, 10bit resolution with 2 GS/s sampling rate in 4-channel mode and 4 GS/s in 2-channel mode as well as 1 GHz analog input bandwidth. It supports open Xilinx FPGA with resources available for customized real-time digital signal processing along with 1 Gbyte of onboard acquisition memory and hardware trigger and highly accurate multi-channel synchronization capabilities. There is extensive software suite including easy-to-use evaluation/integration software Digitizer Studio

Flexible

www

Complete audio system opens up A2B bus



A2B inventor Analog Devices (ADI) has introduced an end-to-end prototyping system that drives the efficient data bus for applications beyond the car. Analog Devices Inc (ADI) has launched a full audio system based around its A 2B bus technology. The system features the SHARC Audio Module (SAM) for the creation of digital audio devices, including audio FX processors, multi-channel audio systems, MIDI synthesisers, and other DSP-based audio systems. In addition to the main SHARC audio module board, ADI offers daughter boards to provide added functionality to the main board and expand the audio system. SAM includes the dual-SHARC+ core ADSP-SC589 audio processor SoC (with an integrated Arm Cortex-A5 core). In addition to the main SHARC audio module board, the company offers daughter boards to provide added functionality to the main board and expand the audio system. The Audio Project Fin board mates directly to the main board, providing MIDI input/output as well as pushbuttons and potentiometers to modify audio effects. The A 2B Amplifier Module features two high-efficiency Class-D amplifiers to output digital audio data received over the twisted-wire pair A 2B bus from PDM microphones and/or serial TDM sources on the main board (or another connected A 2B node). This complete audio system delivers high-fidelity multichannel digital audio with low and deterministic latency to a fully synchronised distributed audio system. The system is aimed at fast prototyping, evaluation projects, demonstrations, and educational applications. It enables users to realise a shorter time to market with a “ready to go” prototype system that provides a comprehensive hardware and software solution.

Analog Devices (ADI)

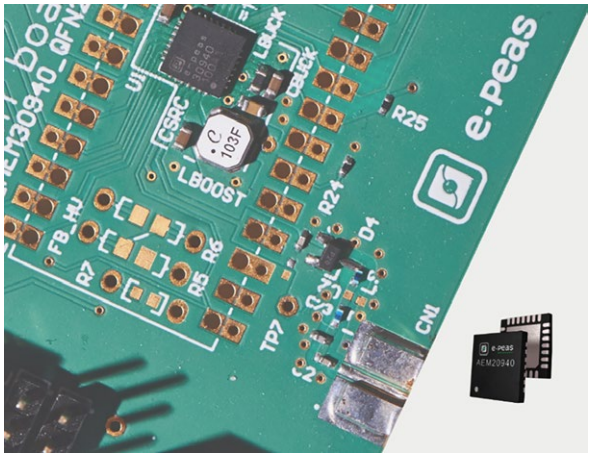
www.analog.com

e-peas signs global supply chain deal with Mouser

Mouser will stock the complete portfolio of power management ICs (PMICs) from e-peas, as well as related evaluation boards, ready for immediate shipment.

Belgian energy harvesting chip designer e-peas has strengthened its sales channel through a global distribution agreement with Mouser Electronics.

The PMIC technology is suitable for a broad array of potential use cases, including industrial control/monitoring, building automation, smart homes, agriculture, environment monitoring, healthcare and smart metering and seeing widespread deployment in locations all over the world.



The Ambient Energy Manager (AEM) PMIC series offers devices specifically optimized for solar, thermal, vibration and RF oriented energy harvesting. This allows IoT hardware can be powered indefinitely, without the need to periodically replace batteries.

The AEM series includes the AEM10941 for extracting DC power from photovoltaic cells, covering 380mV to 3µW. The AEM20940 is dedicated to thermal energy harvesting, and can manage DC power from 60mV, taken off thermo-electric generator (TEG) outputs. The AEM30940 is used in situations where vibrational-based energy harvesting is applicable, this PMIC (combined with an adapted rectifier) can deal with the DC power provided by piezo generators, micro turbine generators or RF antennas. The AEM40940 extracts AC power from ambient RF waves, leveraging the available input power from 20dBm up to 10dBm.

Housed in compact 5mm x 5mm QFN packages, the PMICs extract power from their respective harvesters as soon as the 380mV and 3µW threshold is met. The cold-start circuitry allows operation to be initiated with empty storage elements from a far lower input power than for competing devices. They can then supply this power to IoT hardware via 2 independent regulated voltages lines (thanks to the two LDOs integrated), delivering currents of up to 80mA.

The PMICs can simultaneously store the extracted energy in rechargeable elements (such as battery cells or supercapacitors). Each device has a built-in ultra-low power boost converter that operates with input voltages of between 50mV and 5V. Only a small number of passive components are required to accompany these PMICs, saving board space and keeping bill of materials costs down.

Mouser

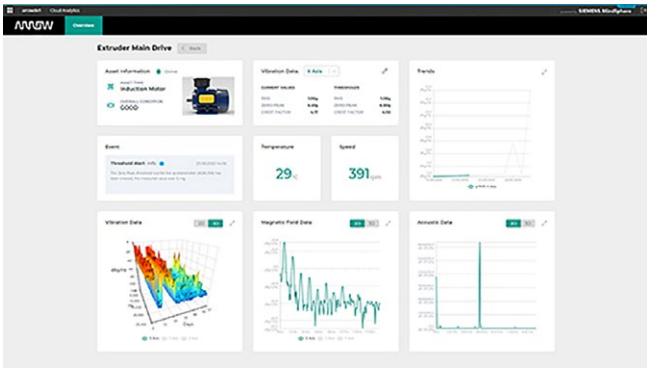
www.mouser.com/manufacturer/e-peas

End-to end cloud analytics for the IIoT

Arrow Electronics has worked with codestryke, Shiratech and Siemens to develop an end-to-end cloud analytics system for the Industrial Internet of Things (IIoT)

Highlighting the changing role of global distributors, Arrow Electronics has teamed up with three other companies to develop an end-to-end cloud analytics solution for applications in the industrial internet of things (IIoT).

Arrow worked with codestryke, Shiratech and Siemens to cover everything from sensors and connectivity hardware to analysis and reporting software, all running on the Siemens MindSphere industrial IoT-as-a-service solution.



The networking and monitoring of plants and factories is a key trend in th emove to Industry 4.0, and condition monitoring of machines and devices is essential for efficient predictive maintenance, having a significant impact on overall equipment effectiveness.

Arrow's system gathers data from sensors and transmits it to the cloud via an edge gateway, where it can be visualized and analysed through a customised dashboard. Continuous monitoring and collection of all relevant operating data means patterns can be observed and emerging faults identified, enabling maintenance to be carried out on a more accurately scheduled basis. In addition to helping users avoid the disruption and costs associated with unplanned downtime by optimizing maintenance intervals, these cloud analytics can also analyse running costs and help increase overall service life of equipment.

Arrow worked with Shiratech to develop iCOMOX, a hardware and software solution for condition-based monitoring of industrial plants, machines and other assets connected to the cloud with MindSphere. This uses the data from connected products, plants and systems to optimize operations, create better quality products and deploy new business models.

codestryke develops customised IoT solutions and offers consultancy to help organisations plan and roll out IIoT deployments successfully and at scale. Its expertise covers app development, edge and cloud analytics and the operation of cloud infrastructure. codestryke also has an edge gateway offer (including a runtime), PLC programming and connectivity solutions.

Arrow is involved in the project planning and commissioning to after-sales support. Customer requirements are identified and specified during a workshop with Arrow's support team and installation and assembly are coordinated and implemented jointly.

Arrow Electronics

fiveyearsout.com

www.arrow.com

eeNews
EUROPE

PUBLISHER
André Rousselot
+32 27400053
andre.rousselot@electronicseurope.net

EDITOR-IN-CHIEF
Nick Flaherty
+44 7710 236368
nick.flaherty@electronicseurope.net

EDITORS
Christoph Hammerschmidt
+49 8944450209
chammerschmidt@electronicseurope.net
Peter Clarke
+44 776 786 55 93
peter.clarke@electronicseurope.net
Nick Flaherty
+44 7710 236368
nick.flaherty@electronicseurope.net
Jean-Pierre Joosting
+44 7800548133
jean-pierre.joosting@electronicseurope.net
Alistair Winning
+44 7834 770342
ally.winning@electronicseurope.net

CIRCULATION & FINANCE
Luc Desimpel
luc.desimpel@electronicseurope.net

ADVERTISING PRODUCTION & REPRINTS
Lydia Gijsegom
lydia.gijsegom@electronicseurope.net

ART MANAGER
Jean-Paul Speliers

ACCOUNTING
Ricardo Pinto Ferreira

REGIONAL ADVERTISING REPRESENTATIVES
Contact information at:
http://www.eenewseurope.com/content/sales-contacts

eeNews Europe
is published 10 times in 2020
by **European Business Press SA**
Chaussée de Louvain 533,
1380 Lasne, Belgium
Tel: +32-2-740 00 50
Fax: +32-2-740 00 59
email: info@electronics europe.net
www.eenewseurope.com
VAT Registration: BE 461.357.437.
Company Number: 0461357437
RPM: Nivelles.
Volume 22, Issue 8 eeNews P 304128
It is free to qualified engineers and managers involved in engineering decisions – see:
http://www.eenewseurope.com/subscribe

The shape of manufacturing post-COVID

Chris Purcell, Product Strategist at Episerver

There's no doubt that the coronavirus pandemic has had a devastating impact on the UK economy. In April 2020, GDP fell by just over 20% — the largest fall since monthly record began in 1997.

Despite widespread weakness across many UK sectors during August, including wholesale and retail trade, transportation and food and beverage, this decline in GDP was primarily due to the fall of 24.3% from the manufacturing sector.

The strict lockdown measures put in place at the end of March in a bid to prevent the spread of the virus have had a significantly negative effect on manufacturing; factories have been forced to close and supply chains disrupted, while consumer demand for essential items has surged.

The Covid-19 experience has been the catalyst for change across many sectors. We see four key predictions that are likely to shape the manufacturing and distribution industries in the next five years.

Increased convenience

The pandemic forced many business-to-business (B2B) companies to come up with new and creative solutions to ensure customers could receive their orders.

A great example is Distributor Corporation of New England, which introduced pickup lockers that allow customers to easily and safely pick up products outside the warehouse. Others have offered click & collect methods with text and email delivery tracking.

These strategies were not designed to ensure customer service and continuity for a few days or weeks. Consumers have experienced these quicker, more convenient ways of collecting orders for months now and it's shaping their expectations for the future.

To stay ahead, manufacturers and distributors must ensure their pivot to convenience — for example clear visibility into order status, inventory availability, and fast fulfilment — continues long after the pandemic.

A shift in buying preferences

Global lockdown measures meant millions were forced to stay at home and both consumers and B2B buyers were quickly forced to make the shift to purchasing goods online — where they expected the Amazon-like experience.

While ecommerce sales have not increased across every industry, reliance on seamless digital purchasing experiences are here to stay.

At the height of the pandemic, companies were seeing an increase in online activity even from their most technology-adverse customers. It's clear that customers are growing more comfortable with digital tools and those across the supply chain need to pivot as the adoption of digital continues to rise.

Investment in ecommerce

The shift in buying preferences to ecommerce means manufacturers and distributors will need to innovate to meet new customer expectations. B2B companies that rely on selling products in particular – whether that's car parts for the automotive industry or materials for building merchants — are finding out the hard way that they are not fully equipped with the right tools to serve customers in a digital-first world.

For many, it's tempting to pause any technology investments to ensure efficiency during this uncertain time, but thinking strategically now, and investing in the right way, could well save businesses in the long run.

When it comes to the manufacturing and distribution industries, an ecommerce strategy is no longer a matter of "when" — it's a matter of "how". It's clear COVID-19 has and will continue to accelerate digital transformation, leaving many organisations with no choice but to embrace and invest in digital technologies. In fact, our latest research discovered that nearly a third (32 percent) of manufacturers believe that over 41 percent of their revenue in 2025 will be derived from ecommerce.

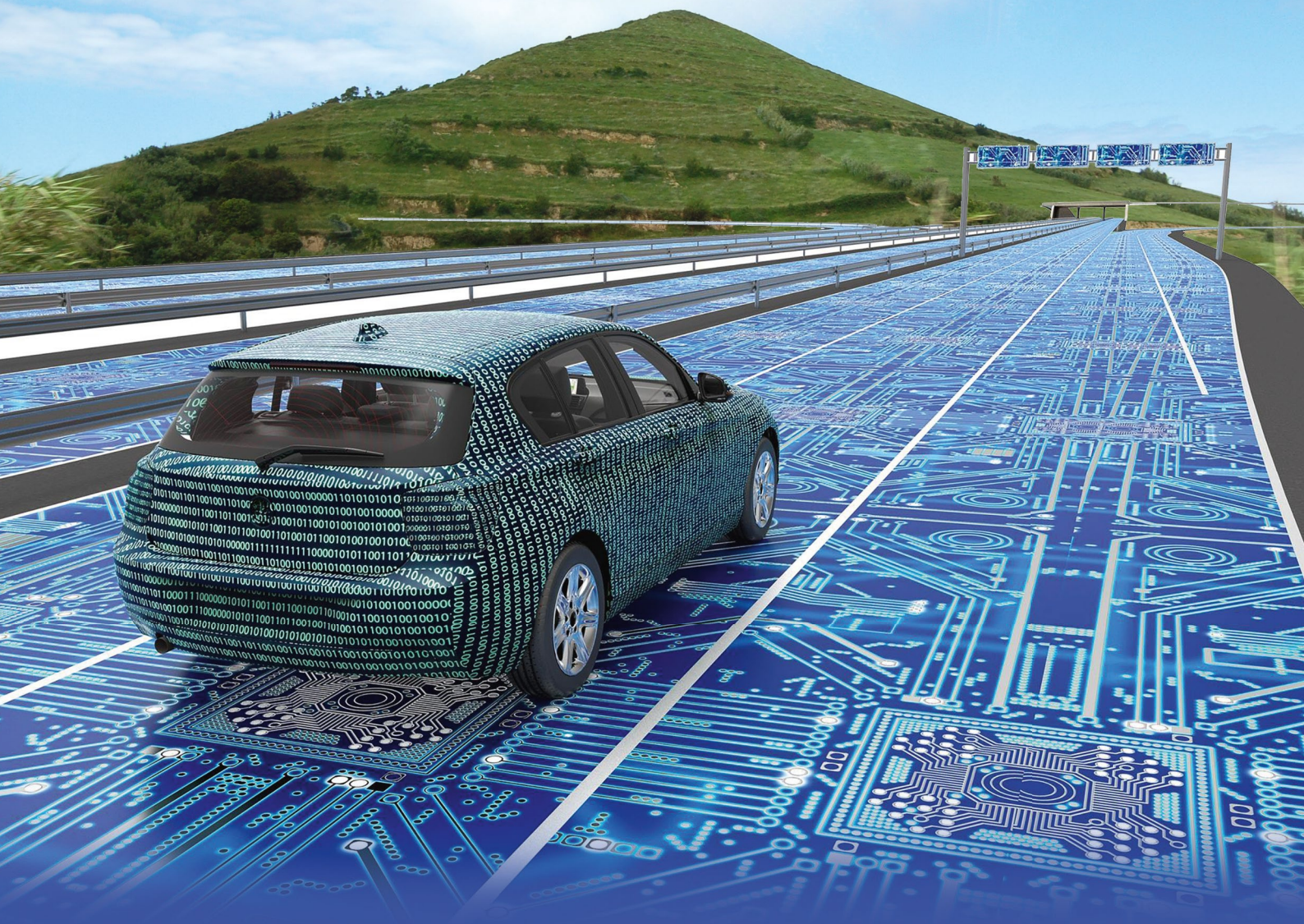
Creating digital-first businesses

Like many industries, manufacturing and distribution are going through drastic changes. As the global pandemic continues to have an impact on the wider economy, manufacturers and distributors need to be agile and poised to pivot at a moment's notice.

eeNews

AUTOMOTIVE

www.eenewsautomotive.com



european business press

eeNews Embedded

eeNews EUROPE

MWee RF - Microwave

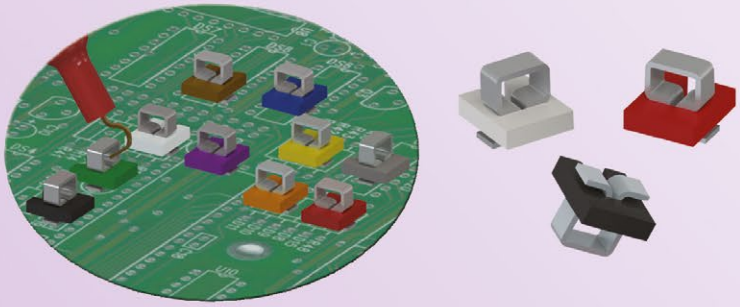
eeNews ANALOG

eeNews LED LIGHTING & DISPLAYS

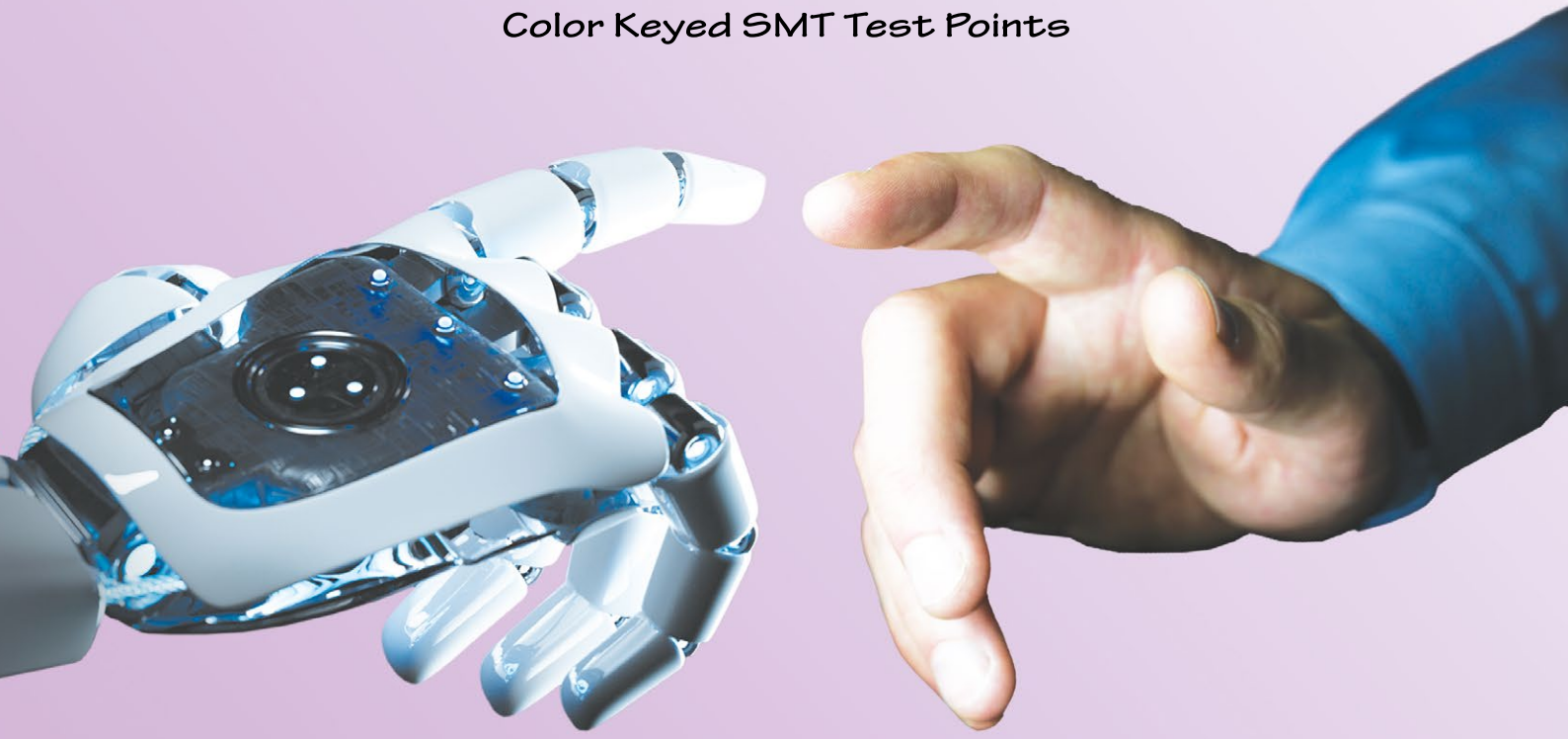
eeNews POWER MANAGEMENT

eeNews TEST & MEASUREMENT

THINK™
HIGH PERFORMANCE



Color Keyed SMT Test Points



Designers & Manufacturers of Electronic Components and Hardware

• Color Keyed for visibility and quick identification • Ideal for production testing and service troubleshooting • Wire loop provides convenient access point for J-hooks and test probes • Manufactured from Phosphor Bronze with Silver Plating and High Temperature Nylon Insulator • Silver Plating allows for superior solderability and low electrical resistance • High Temperature Nylon Insulator designed for SMT reflow soldering • Supplied on Tape and Reel

IT'S WHAT'S ON THE INSIDE THAT COUNTS

KEYSTONE®
ELECTRONICS CORP.

View our Dynamic Catalog M70 at www.keyelco.com

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