AsianScientist
Issue 07
January 2020

SUPERCOMPUTING ASIA

+

NOT JUST INDUSTRIOUS, BUT INTELLIGENT

KEEPING SMART CITY ALGORITHMS ACCOUNTABLE

STUDENTS TAKE THE STAGE

CITYSIM

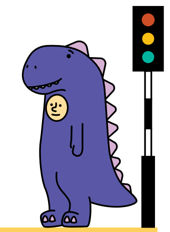
HOW SIMULATING CITIES MAKES THEM SMART

MCI (P) 037/04/2019

CONTENTS

Issue 07
January 2020

FEATURES



EDITOR'S NOTE

Though few members of the general public will ever get the chance to see a supercomputer up close, supercomputing nonetheless touches almost every aspect of our everyday lives. Every time you sit in a car or an airplane, you are sitting in a machine that was likely designed with the help of a supercomputer (*Not Just Industrious, But Intelligent*, p. 10). Industrial uses of supercomputing also extend to smart factories, semiconductor design and even the operations of data centers—the backbone of the internet that few of us could last a day without.

But supercomputers are coming even closer: to a streetlight near you. In our cover story (*CitySim*, p. 16), we talk to Charles Catlett, founding director of the Urban Center for Computation and Data at the University of Chicago. He shares with us how the Array of Things is being used to monitor the health of the City of Chicago, and how high performance computing at the edge is being used to keep the data collected safe.

Last but not least, check out *The Cost of Cooling* (p. 38), a handy summary of the different types of cooling available and the costs and benefits of each.

Rebecca Tan, Ph.D.
Editor-in-chief
Supercomputing Asia

/asianscientist @asianscientist
Asian Scientist Magazine asianscientist

www.asianscientist.com

SUPERCOMPUTING ASIA

EDITORIAL ADVISORY COMMITTEE
Prof. Tan Tin Wee
Prof. Satoshi Matsuoka
Prof. John Gustafson
Yves Poppe

CEO & PUBLISHER
Dr. Juliana Chan

EDITOR-IN-CHIEF
Dr. Rebecca Tan

DIGITAL EDITOR
Dr. Jeremy Chan

CONTRIBUTORS
Eugene Low
Kamila Navarro
Sara Watson

ART DIRECTOR
Lam Oi Keat

SENIOR DESIGNER
Shelly Liew

SALES & MARKETING
Samantha Yeap
Grace Phua

PUBLISHED BY
Wildtype Media Group Pte Ltd

DISTRIBUTED BY
Pansing Distribution Pte Ltd

HEAD OFFICE
Wildtype Media Group Pte Ltd
71 Ayer Rajah Crescent #07-03
Singapore 139951
Tel: 65-6909-2279
Website: www.asianscientist.com/subscribe

WILDTYPE MEDIA

Combining savvy communication with technical rigor, Wildtype Media Group is Asia's leading STEM-focused media company, spanning digital, print, custom publishing and events.

Brands under Wildtype Media Group include the flagship *Asian Scientist Magazine* and *Supercomputing Asia*, award-winning titles available in print and online. Through its Asian Scientist Intelligence consulting division, Wildtype Media Group provides a full suite of marketing and communication services to industry and academic clients hoping to engage Asia.

www.wildtype.media

To order reprints or e-prints, or request permission to republish *Supercomputing Asia* content, please contact hello@wildtype.media

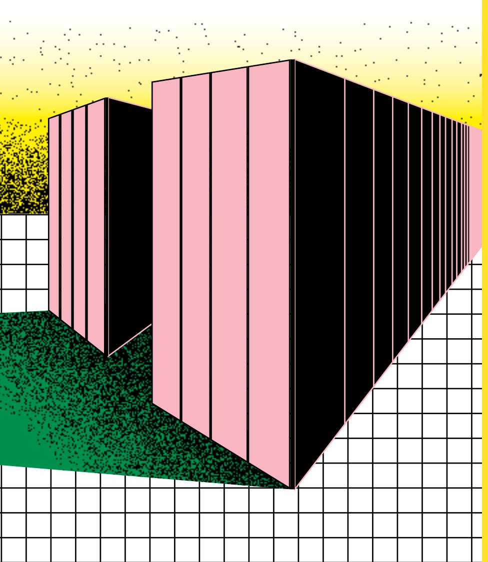
Digital Dispatch

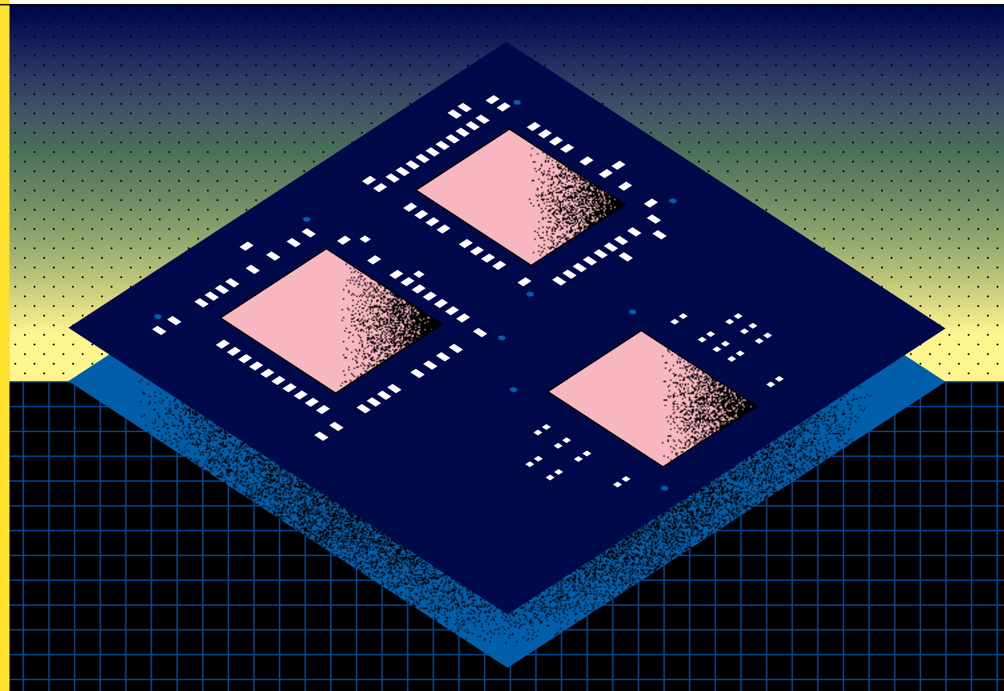
CHINA'S SEVENTH SUPERCOMPUTING CENTER PLANNED FOR 2020

The Chinese government plans to build their seventh supercomputing center in Henan province, according to *Science and Technology Daily*, the official newspaper of the Ministry of Science and Technology. The facility at Zhengzhou University will have a computing capacity of 100 petaFLOPS and is scheduled to come online in the first half of 2020.

China currently has six supercomputing centers in Tianjin, Jinan, Changsha, Shenzhen, Guangzhou and Wuxi, respectively. The country's most powerful supercomputer, the 93 petaFLOPS Sunway TaihuLight, is housed at the National Supercomputing Center in Wuxi, and is being used as a cloud computing open platform for China's small and medium enterprises.

While not intended to be an exascale facility, the new center in Henan is likely to be used to prototype software for China's planned exascale systems, reports *Data Centre Dynamics*. In particular, the Henan center has been earmarked for developing applications in artificial intelligence, equipment manufacturing, precision medicine and agriculture.





CHIPS GO 3D FOR EXASCALE COMPUTING

The European Exascale Processor & Memory Node Design (ExaNoDe) project has unveiled their solution to the impending end of Moore's Law: stacking integrated circuit components in three dimensions (3D).

To reach the next milestone of exascale computing, the computing density and power efficiency of existing hardware has to be dramatically increased. However, integrated circuits are approaching the limits of miniaturization, necessitating radically new approaches to chip design.

"Affordability and power consumption are the main hurdles for an exascale-class compute node. In the ExaNoDe project, we have built a complete prototype that integrates multiple core technologies, which will enable European technology to satisfy the requirements of exascale high performance computing," said CEA-Leti engineer and project coordinator, Mr. Denis Dutoit. CEA-Leti is one of the thirteen partners for ExaNoDe.

The 3D integrated circuit combines multiple system-on-chip chiplets that can be stacked in a modular fashion. This approach reduces the cost of customization and cuts energy consumption by reducing inter-chip communication distances. In addition to the hardware prototype, the ExaNoDe team has also developed a high performance software stack and applications in materials science and engineering.

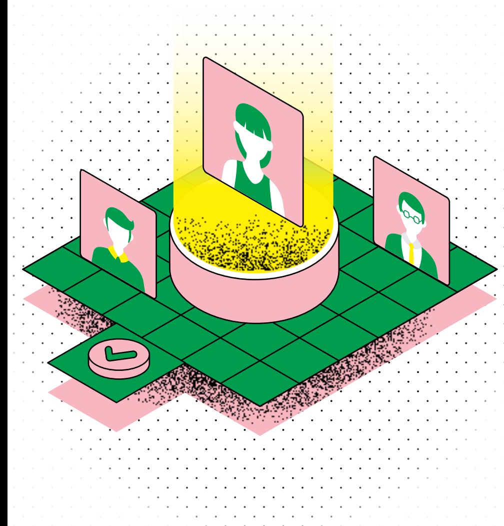
AI SINGAPORE BUILDS IN-HOUSE HPC CAPABILITIES

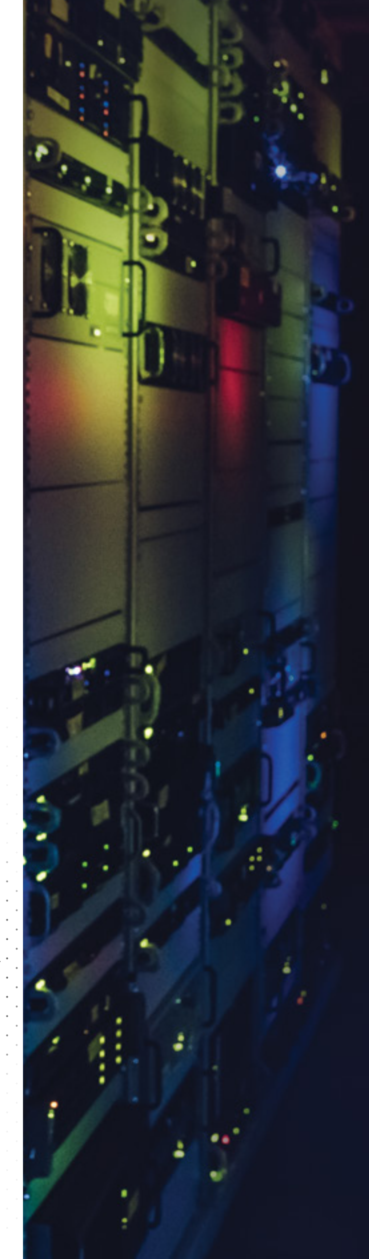
To meet a growing demand for computing capacity, Singapore's national artificial intelligence (AI) program, AI Singapore, has set up its own on-premise supercomputer.

Under AI Singapore's 100 Experiments initiative, companies are encouraged to develop customized AI solutions to industry-specific problems through talent matching, infrastructure support and up to S$250,000 in matching funds. 100 Experiments AI workloads were previously run on the cloud, but as the program grew in popularity, it became more practical for AI Singapore to invest in its own in-house HPC capabilities.

"Within a week, the researchers can easily incur S$5,000 to S$10,000 in cloud resources—that's enough to buy a server in two weeks," said Mr. Laurence Liew, director for AI industry innovation at AI Singapore.

The new infrastructure will also help with data privacy requirements that do not allow large corporations and government agencies to run projects on the cloud. On top of infrastructure support, AI Singapore conducts workshops for students on related topics.





FRONTIER SUPERCOMPUTER TO EXCEED 1.5 EXAFLOPS

Although it will not be the first exascale supercomputer in the world or even in the US, the Frontier supercomputer at Argonne National Laboratory is likely to be the most powerful in the world when it is completed in 2021.

Exascale computing, made possible by supercomputers that can perform over 10^{18} calculations per second, is a symbolic goal being pursued by the US, China, Japan and the European Union, among others. Currently the owner of the two fastest supercomputers in the world, the US plans to launch its first exascale supercomputer—named Aurora—in 2021, followed next by Frontier, which is expected to be even faster than Aurora with at least 1.5 exaFLOPS of peak performance.

"Frontier will accelerate innovation in AI by giving American researchers world-class data and computing resources to ensure the next great inventions are made in the United States," said Mr. Rick Perry, the US Secretary of Energy.

To be co-designed by AMD and Cray, Frontier will build on the AI abilities of the world-leading Summit system, providing new capabilities for deep learning, machine learning and data analytics.

Digital Dispatch

The winners of the 2019 Gordon Bell Prize on stage at SC19 to receive their award from Professor Jack Dongarra, SC19 awards chair (left); Professor Cherri Pancake, president of the Association for Computing Machinery (second from right); and Professor Arndt Bode, chair of the 2019 Gordon Bell Prize Award Committee (right).

Photo credit: SC Photography

ASEAN TO HARNESS HPC

The Association of Southeast Asian Nations (ASEAN) has agreed to develop high performance computing (HPC) in the region through collaborations between its 11 member states. The ASEAN HPC Workplan was endorsed during the 18th ASEAN Ministerial Meeting on Science, Technology and Innovation and the 77th ASEAN Committee on Science, Technology and Innovation (COSTI), held in Singapore on October 6–11, 2019.

Aimed at strengthening the region's digital connectivity and readiness for the Fourth Industrial Revolution, the ASEAN HPC Workplan aims to increase access to HPC resources and expertise. The Workplan also covers public-private partnerships to build HPC capabilities in Southeast Asia's workforce.

The initiative will be co-led by Singapore and Thailand through the Agency for Science, Technology and Research (A*STAR) Computational Resource Centre, and the National Electronics and Computer Technology Centre, respectively.

"Through our collective efforts in growing innovative capabilities and building a strong talent base, we will position ASEAN well to capture opportunities in emerging trends such as digitalization, and improve the lives of ASEAN citizens," said Dr. Raj Thampuran, Singapore's COSTI chairman and senior advisor to A*STAR.

ETH ZURICH TEAM WINS 2019 GORDON BELL PRIZE

Computing could get much cooler thanks to the winners of the 2019 Gordon Bell Prize, a highly-coveted award recognizing the best uses of HPC. A six-member team from the Swiss Federal Institute of Technology in Zurich (ETH Zurich) won the prize for their framework, which allowed them to simulate the movement of electricity through a transistor.

A single microchip can contain hundreds of millions of transistors, each of them generating heat as they operate. Cooling is estimated to account for approximately 40 percent of a data center's electricity bills, and heat generation is a bottleneck limiting further chip miniaturization.

To understand how heat is generated and dissipated in transistors, the ETH Zurich team simulated how electricity is transported through a two-dimensional slice of a transistor consisting of 10,000 atoms. Their DaCe OMEN code ran 14 times faster than an earlier framework used for a system ten times smaller.

This ability to simulate where heat is generated and how it is removed could inform the development of new semiconductors with improved heat dissipating properties.

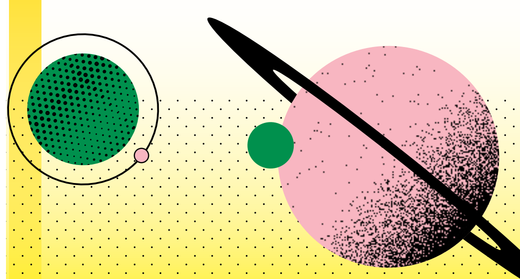
GOH ENG LIM WINS NASA'S EXCEPTIONAL TECHNOLOGY ACHIEVEMENT MEDAL

For his role in sending supercomputers to space, Dr. Goh Eng Lim has been awarded The Exceptional Technology Achievement Medal by the US National Aeronautics and Space Administration (NASA).

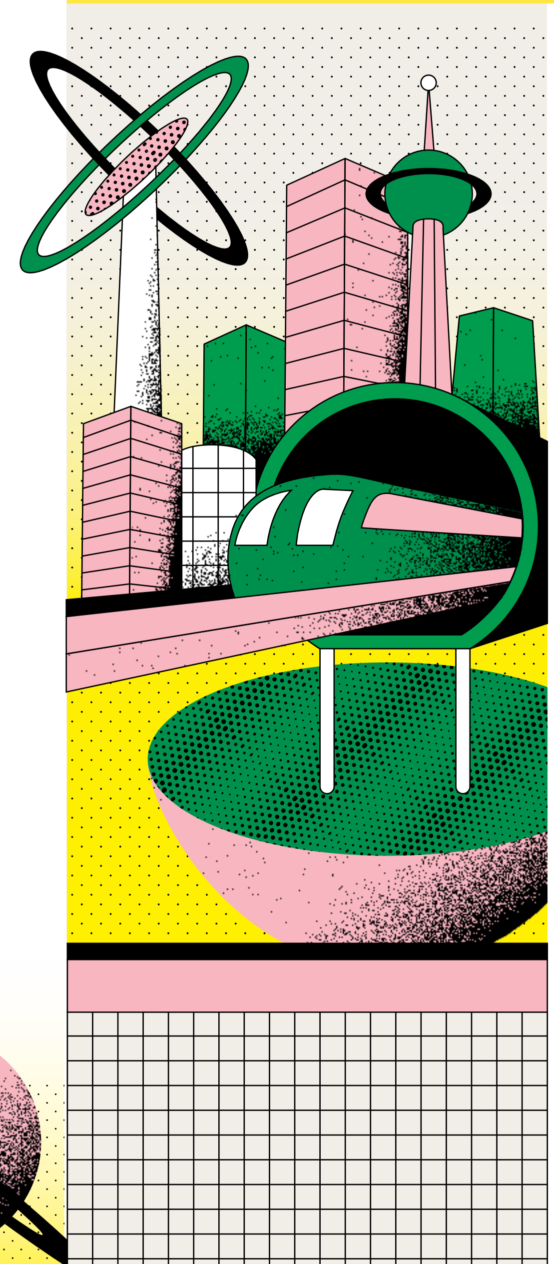
The further we travel away from Earth, the more time it takes to transmit data to and from earth-bound computers, making onboard computing capabilities necessary. However, space is an unforgiving place for supercomputers, with harsh environmental conditions such as cosmic radiation.

Goh, the vice president and chief technology officer of HPC and AI at Hewlett Packard Enterprise (HPE), led a mission to develop a supercomputer onboard the International Space Station. The resulting Apollo Spaceborne Computer showed that off-the-shelf hardware was capable of executing over one trillion calculations per second for a year without requiring a reset. Now that supercomputing in space has been shown to be practical, Goh hopes that supercomputers will next make their appearance on the Moon and eventually, Mars, he said in an interview with *The Register*.

Apart from the Exceptional Technology Achievement Medal, Goh's contributions to HPC have been recognized with prizes such as the Singapore Visionary Award, which he won at the SupercomputingAsia conference in 2018.



WHAT'S UP!



A 2020 VIEW OF TOMORROW'S CITIES

SupercomputingAsia (SCA), Asia's premier gathering of the region's leading HPC experts, returns to Singapore in 2020. The annual event, which brings together thought leaders from both academia and industry, will be held on February 24–27 at the Suntec Singapore Convention and Exhibition Centre.

Organized by the National Supercomputing Centre (NSCC) Singapore, this year's SCA will shine a spotlight on the role of HPC in enabling smart cities. The event will be graced by guest-of-honor Dr. Vivian Balakrishnan, Singapore's minister for foreign affairs and minister-in-charge of the Smart Nation initiative. Mr. Charles Catlett (p. 16), founding director of the Urban Center for Computation and Data, will deliver the keynote speech.

Last year's event saw more than 700 delegates from around the world. This year, apart from co-located events such as the Asia Pacific Research Platform (APRP) Conference, SCA will feature a new track focused on supercomputing in India. Save the date and see you there!

For more information,
visit https://www.sc-asia.org

WHAT

SupercomputingAsia 2020

WHEN

February 24–27, 2020

WHERE

Singapore

LOOK OUT FOR ISC HIGH PERFORMANCE 2020

The International Conference for High Performance Computing, Networking, Storage and Analysis (ISC High Performance) will be held in Frankfurt, Germany, from June 21–25, 2020.

Since 1988, the SC Conference series has showcased the latest developments in the world of HPC via its highly competitive technical program. Chaired this year by Professor David Keyes of the King Abdullah University of Science and Technology, Saudi Arabia, and deputy chair Professor Martin Schulz of Technische Universität München (TUM) Munich, the program includes talks, panels, research papers, poster presentations and tutorials, all designed to keep the community up to date on the bleeding edge of HPC.

Over 400 speakers and 160 exhibitors will gather to discuss the hardware and software demands of researchers and businesses in the fields of HPC, artificial intelligence, machine learning and data analytics.

For more information,
visit https://www.isc-hpc.com

WHAT

ISC High Performance 2020

WHEN

June 21–25, 2020

WHERE

Frankfurt, Germany

NOT JUST INDUSTRIOUS, BUT INTELLIGENT

5 ways supercomputers support industry

High performance computing is transforming the face of manufacturing while augmenting the way factories and facilities operate.

By **Jeremy Chan**

Looking back on the past two centuries of industrialization, much has changed in terms of the economic value of a human being. In the 1800s, human labor used to matter most on the assembly line or manufacturing floor. Then along came machinery, and what workers could achieve with their hands became less important—and less valuable—than what they could achieve with their heads. Now, with digital technologies becoming an integral part of factories in the Fourth Industrial Revolution, even tasks that once solely belonged to the domain of grey matter are being taken over by intelligent systems.

A key addition to the 'nervous system' of modern industry is supercomputers, capable of performing thousands of trillions of calculations per second. The massive processing power of these devices promises to transform the way products are designed, engineered and maintained. Asia, home to more than half of the world's supercomputers, is already benefitting from the use of supercomputers in a wide range of industries.

CRASH TEST SIMULATIONS

While no automobile manufacturer wants to see their product wrecked in a crash, they have a responsibility to ensure that their vehicles offer passengers an adequate degree of protection should a traffic collision occur. Crash testing is therefore routinely carried out by automobile manufacturers to ensure safety standards are met before a new model is introduced.

However, the very nature of crash testing means that dummy vehicles are built only to be destroyed, making the process expensive and time consuming. To reduce the number of dummy vehicles required, automobile manufacturers can tap on supercomputers for crash test simulations.

For instance, the K supercomputer in Japan is being used by Japanese carmakers to run crumple simulations of automobile chassis floors during collisions. This typically involves creating a digital representation of a car comprising more than a million components, then analyzing how these components behave under the extreme physical stresses of a crash. A wide range of parameters—such as chassis material and structure, as well as the speed of the car—can be modeled and monitored *in silico*, and adjustments can be made to the car's design to optimize safety before actual crash testing is carried out.

Illustrations by Lam Oi Keat/*Supercomputing Asia*

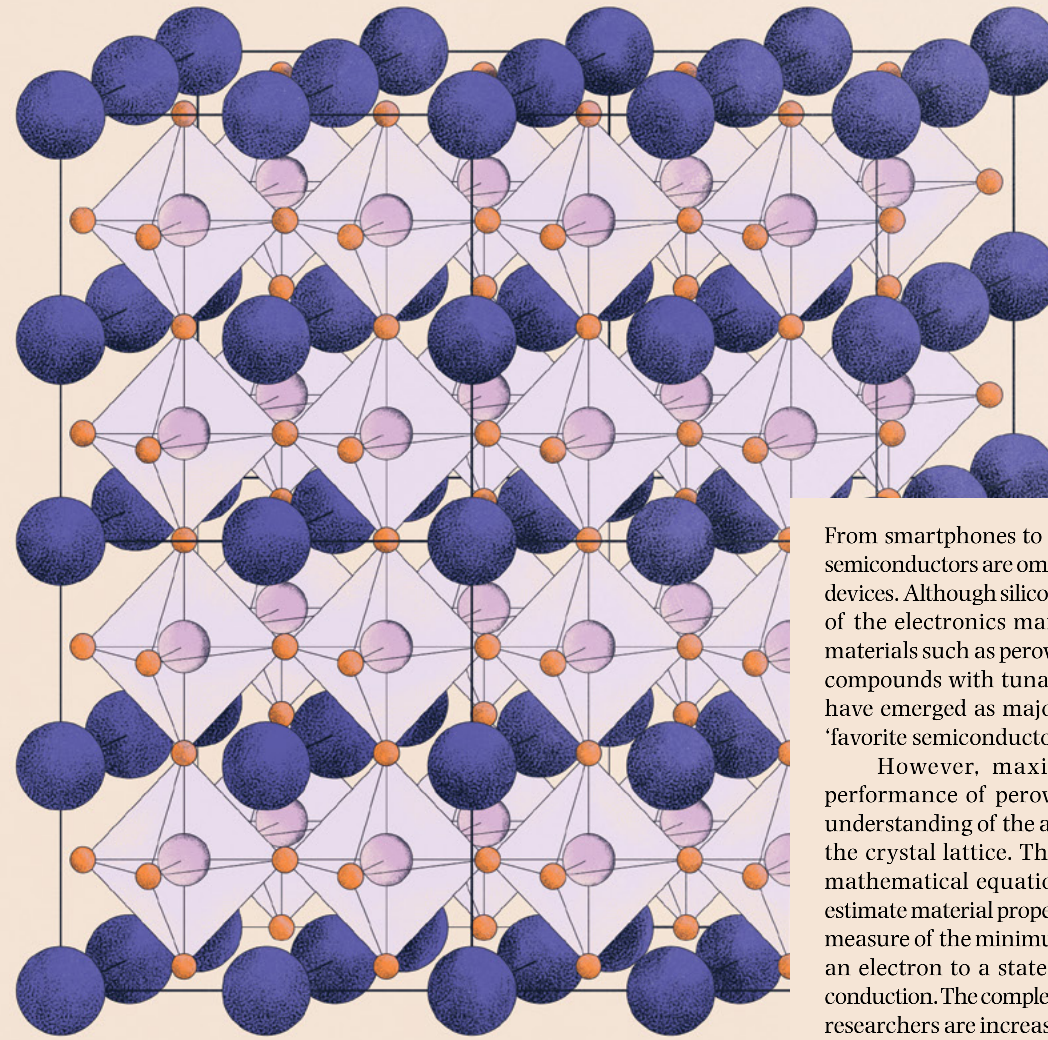
AERODYNAMICS ASSESSMENTS

Since the Wright brothers built the world's first successful airplane in 1903, the aviation industry has come a long way with larger aircraft capable of carrying heavier loads and traveling longer distances. Despite these improvements, aircraft manufacturers are not resting on their laurels and are seeking better airplane designs to reduce drag, which in turn translates to better fuel economy and lower operating costs.

Boeing and Airbus have dominated the aircraft manufacturing industry for decades, but a new contender has emerged in recent years—the Commercial Aircraft Corporation of China, Ltd., or Comac for short. Comac's first model, the C919, is a narrow-body twinjet airliner that is 38.9 meters long, with a wingspan of 33.6 meters and a carrying capacity of 190 passengers.

The design of the C919 was informed by the Tianhe-2, ranked as the world's fastest supercomputer from 2013–2016. Crucially, the Tianhe-2 was used for computational fluid dynamics simulations of the external flow field around the C919. According to personnel involved in the analyses, the highly precise calculations, which would have required two years on conventional computing systems, were completed in just six days on the supercomputer.

The C919 first took the skies in a test flight on May 5, 2017, but is still undergoing refinement at the time of writing. It is expected to make its commercial debut with China Eastern Airlines in 2021.

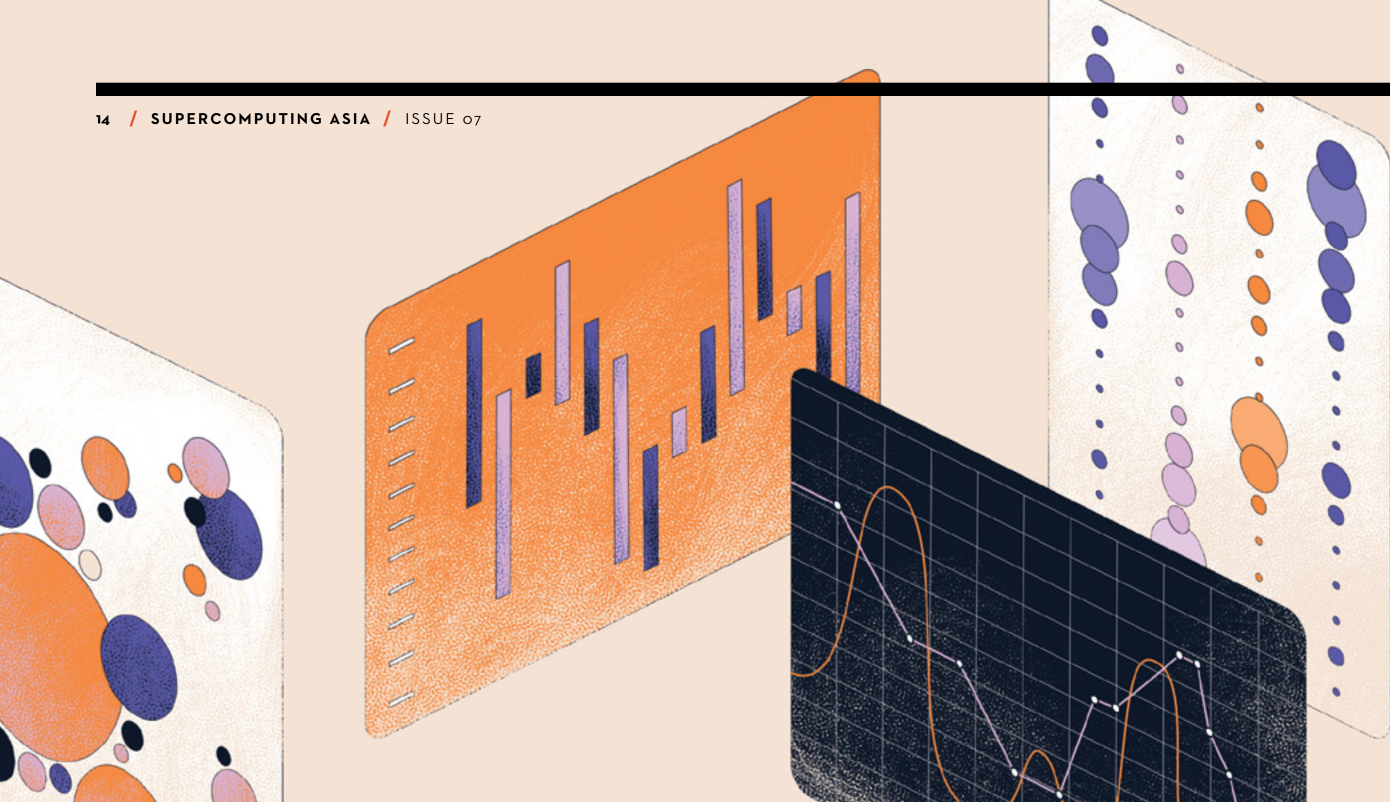


SEMICONDUCTOR SOLUTIONS

From smartphones to LED lamps and solar panels, semiconductors are omnipresent in today's electronic devices. Although silicon has thus far been the darling of the electronics manufacturing industry, novel materials such as perovskites—a class of crystalline compounds with tunable conducting properties—have emerged as major contenders for the title of 'favorite semiconductor.'

However, maximizing the function and performance of perovskites requires a thorough understanding of the arrangement of atoms within the crystal lattice. This requires solving complex mathematical equations for millions of atoms to estimate material properties such as the band gap—a measure of the minimum energy required to elevate an electron to a state where it can participate in conduction. The complexity of the problem means that researchers are increasingly relying on the powerful number-crunching capabilities of supercomputers.

At the Korea Institute of Science and Technology Information (KISTI) in South Korea, the Nurion supercomputer—ranked the 14th most powerful supercomputer in the world as of November 2019—has been used to model the nanostructure of a metal halide perovskite by performing calculations on over a million atoms. Researchers at KISTI reported good agreement between simulated and experimental band gap values for the studied material. Importantly, findings from the simulation have resulted in ideas to overcome a bottleneck in the design of light-emitting diodes and may contribute towards the development and large-scale manufacture of next-generation optoelectronic devices.



SMART FACTORIES

With the advent of the Internet of Things (IoT), factories can now rely on a network of connected sensors to monitor production processes in real time. Market intelligence firm IDC forecasts that the manufacturing sector will account for almost US$200 billion in IoT spending in 2019, with major use cases in manufacturing operations and production asset management.

Concurrently, the amount of data generated by IoT in 'smart factories' is expected to grow tremendously. Without accurate and timely analyses, this data remains idle and cannot be used to enhance production workflows or inform predictive maintenance on machine parts. This is where high-performance computing (HPC) becomes essential for data mining.

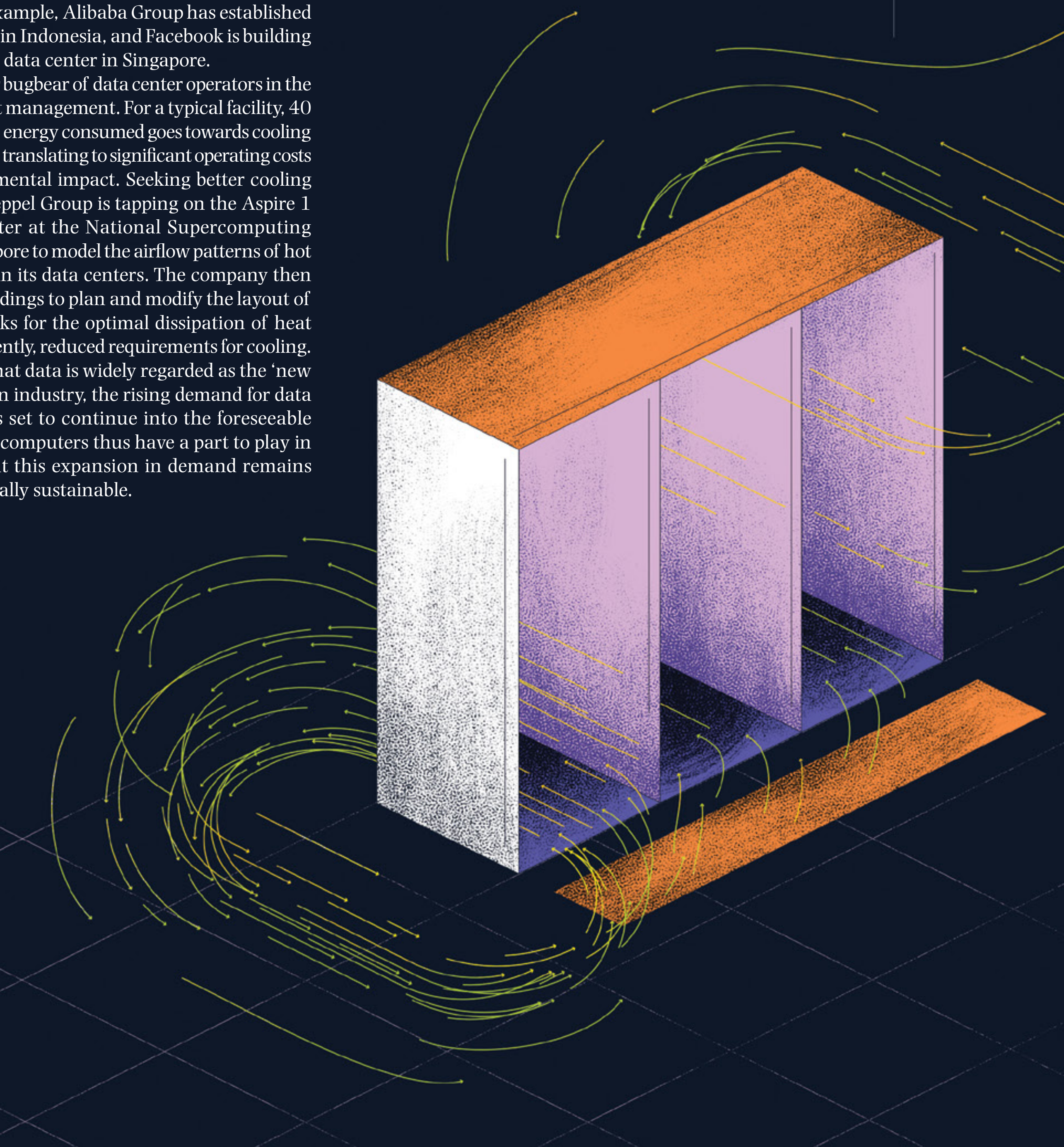
In Asia, companies like Huawei and Fujitsu are actively supporting smart factories with hardware and software for computationally intensive big data analytics. For example, Huawei's Smart Factory Solution includes scalable HPC resources in the teraFLOPS range of processing power. HPC-enabled methods can also be used to track inventory and streamline supply chains, ultimately contributing to greater efficiency and lower manufacturing costs.

DATA CENTER COOLING

Despite existing in the digital realm, data has a footprint in the physical world in the form of servers and networking equipment, which are typically housed in data centers. Southeast Asia has become a hotbed for data center launches in recent years, with many global IT players setting up operations in the region. For example, Alibaba Group has established a data center in Indonesia, and Facebook is building its first Asian data center in Singapore.

A major bugbear of data center operators in the tropics is heat management. For a typical facility, 40 percent of the energy consumed goes towards cooling the hardware, translating to significant operating costs and environmental impact. Seeking better cooling strategies, Keppel Group is tapping on the Aspire 1 supercomputer at the National Supercomputing Centre Singapore to model the airflow patterns of hot and cool air in its data centers. The company then uses these findings to plan and modify the layout of its server racks for the optimal dissipation of heat and consequently, reduced requirements for cooling.

Given that data is widely regarded as the 'new oil' of modern industry, the rising demand for data centers looks set to continue into the foreseeable future. Supercomputers thus have a part to play in ensuring that this expansion in demand remains environmentally sustainable.



CitySim

Cities are quite possibly the most complex things mankind has ever made. Each one teems with millions of people, embedded in interlocking energy, transport and building systems that are evolving by the minute. The dynamism of life in the city is what draws people to them; but it is also what makes cities messy, chaotic and unpredictable—a nightmare for city planners and municipal authorities trying to ensure the reliability of essential services.

How simulating cities makes them smart

Simulations and edge computing could help to tame some of the complexities of cities and make them more livable places for all, says Charles Catlett, director of the Urban Center for Computation and Data at the University of Chicago.

By **Rebecca Tan**

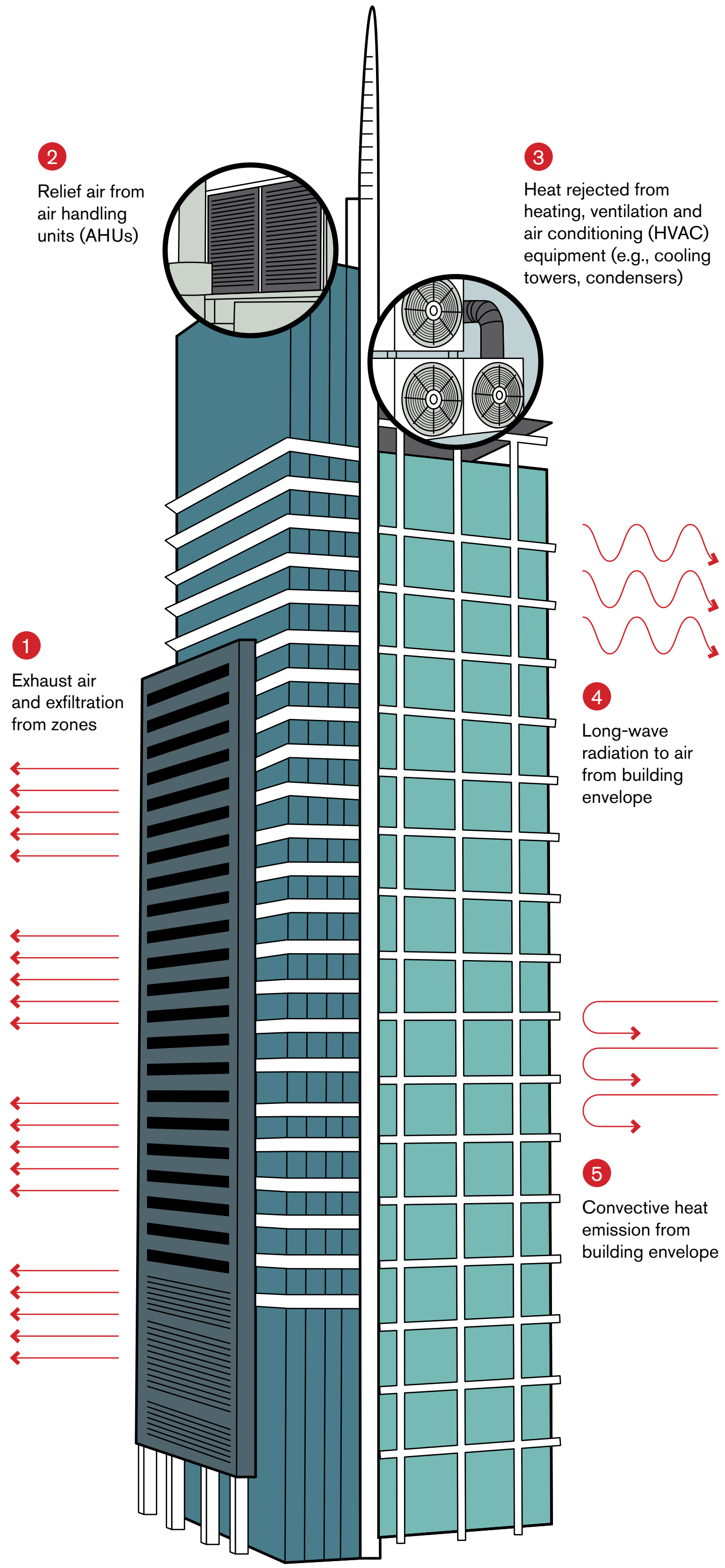
Planning and managing the world's ever-growing cities has never been more important, directly impacting the lives of 60 percent of the population by 2030. Cities generate 80 percent of global gross domestic product, but they are also responsible for 70 percent of global greenhouse gas emissions, making the efficient running of cities an issue of planetary significance. These problems are particularly pressing for Asia, which is expected to have at least 30 megacities by as soon as 2025.

While data can never tell the full story about any living organism, much less one as vast and multifaceted as an entire city, it is now possible to test increasingly realistic scenarios using simulation. This is where high performance computing (HPC), which has been used to simulate everything from aircraft wings to the birth of the entire universe, can help make cities smarter, said Mr. Charles Catlett, founding director of the Urban Center for Computation and Data (UrbanCCD) at the University of Chicago and senior computer scientist at the US Department of Energy's (DOE) Argonne National Laboratory.

"When you're looking at something as complicated as a city, it's not like you can come up with an optimal district design because it's not a closed system or a fixed design like a jet nozzle or something that you might do with traditional fluid dynamics," Catlett said. "Rather, what we are looking for is a sense of the possible and the probable," Catlett told *Supercomputing Asia*.

TWO MODELS ARE BETTER THAN ONE

One very real possibility that the City of Chicago is bracing itself for is extreme heat events. In 1995, a heatwave claimed the lives of over 700 people over a five-day period. The heatwave was unusually severe as the unrelentingly high temperatures were amplified by urban heat islands caused by the high concentration of buildings in Chicago's downtown. Already due for the next big heatwave since 1995-like conditions are thought to occur every 23 years, the city also needs to factor in the impact of climate change, which might make such devastating heatwaves occur as frequently as three times a year by 2100.



Heat emissions from buildings

> **"What we are looking for is a sense of the possible and the probable."**

Understanding how a heatwave might impact the city, however, is much more complicated than simply simulating the weather. While undoubtedly an important parameter and a challenge to simulate in and of itself, the weather is merely one factor contributing to the potential fallout of the next big heatwave. As in the case of 1995, how the buildings trap and radiate heat and the presence or absence of green spaces are crucial determinants of local temperatures; these in turn are influenced over time by socioeconomic factors such as building demand and zoning policies. Furthermore, the common response to heatwaves is to increase the use of air conditioning, which paradoxically drives up the amount of heat generated by buildings and places a strain on the energy supply.

"If I were the City of Chicago, I would want to know which of the buildings in my city are not going to be able to keep up with the demand for cooling during a heatwave. I'm particularly interested not so much in office buildings, but perhaps hospitals, retirement homes or public housing, where those most likely to suffer from a heatwave, like the elderly, live," Catlett explained. "Then I want to ask which of the buildings are able to keep up with the demand for cooling, but do so with inordinately high energy consumption."

"With that information, you can take it one step further and identify the buildings that are likely to have the most difficulty, then do a HPC ensemble that allows you to evaluate multiple energy retrofits for these buildings, and work with the owners to implement the optimal retrofits for their buildings."

What this looks like in practice involves coupling several computational models together. To be able to address the questions that Catlett and his team are asking, a weather simulation forecasting the temperatures during a hypothetical heatwave would have to interact with a separate simulation modeling the energy performance and demand for each building.

"What we did as a proof of concept through the DOE Exascale Computing Project was couple DOE's building energy model, called EnergyPlus, with an equally widely used weather model, Weather Research and Forecasting (WRF). We ran WRF with a spatial resolution of 100m^2 and 50 vertical layers," Catlett said.

"EnergyPlus typically runs a single building on a PC, taking anywhere from several hours to several days to compute an entire year of hourly weather. Instead, we put one building on each computational core, using 20,000 cores for as many buildings, and used WRF hourly output to compute energy performance and demand for each building during a multi-day heat event. We then created an ensemble workflow to evaluate multiple possible retrofit strategies for a subset of these buildings, looking to optimize the retrofit investment relative to cost-benefit."

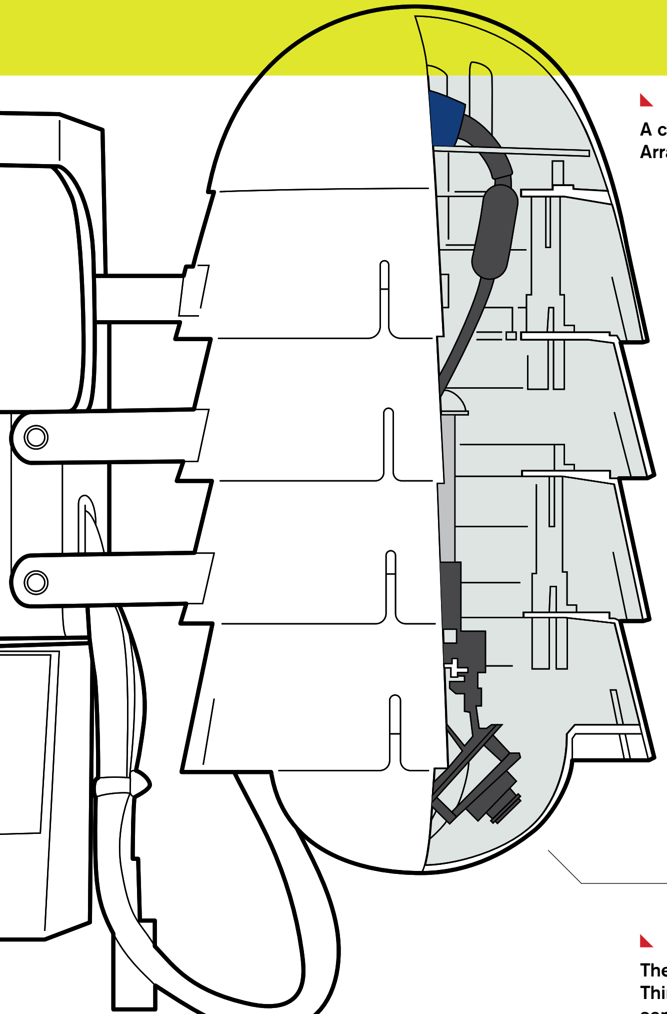
For example, this approach might reveal that a retrofitting improvement that costs $10 million would lead to a 20 percent improvement in building performance through the heatwave while a $12 million retrofit improves performance by 35 percent. "By running these ensembles with HPC, you can ask hundreds or thousands of different questions," Catlett said. "That way, you could at least guide the building owner in making those investments."

TAKING THE CITY'S PULSE

As useful as they might be, simulations still ultimately need to be verified with real-world data. But what kind of data would help you understand the health of a city? And importantly, how can that data be captured, analyzed, moved and stored in a safe and cost-effective way?

"What measurements would assist the scientific community working with cities on diagnosing things like the impact of air quality on school performance, noise on the health of elderly residents, or the flow of pedestrians and safety issues in the downtown area? These are some of the questions that have driven the measurement strategies of the Array of Things," Catlett said.

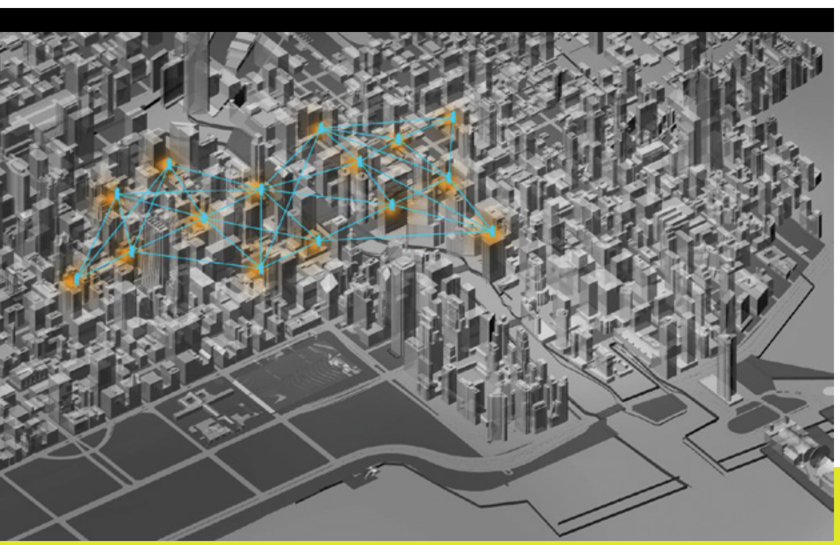
A play on words combining the Internet of Things and array telescopes, the Array of Things is an UrbanCCD project that seeks to use a network of sensors as a city-wide health tracker, analogous to a Fitbit for cities. Designed to be modular to accommodate future sensors, each node can measure environmental factors like temperature, pressure, humidity, light and even ambient sound intensity; common air pollutants like particulate matter, ozone and carbon monoxide; and pedestrian and vehicular traffic. Machine learning is then applied to analyze images locally within each 'node.' In 2020, with planned upgrades to the machine learning hardware in the nodes, more sophisticated machine



A cutaway view of an Array of Things node.

The Array of Things is an urban sensing network of programmable nodes that collect real-time data about cities.

Credit: Center for Urban Computation and Data.



learning could enable the sensors to detect flooding, detailed flow of traffic at an intersection or among people in a park, and other factors of interest to city planners.

All data captured by the sensors is open to the public and available for free on a web-based portal and through a near-real-time API, allowing everyone from scientists to the residents themselves to explore or even develop apps based on the data. This open source approach extends beyond the data itself to the underlying software and hardware platform, called Waggle, developed by Argonne National Laboratory, with all software published as open source to encourage participation and transparency.

"We've done a major test in Chicago with about 120 nodes, some of which have been up for two years. We've also recently received a new grant from the US National Science Foundation to take our Waggle platform and apply it not just to cities, but to ecological and environmental projects as well," Catlett said.

FROM REACTIVE TO PROACTIVE

Apart from improving simulations, data collected from the Array of Things and other sources can also directly help planners in the day-to-day running of a city. For example, the City of Chicago used data analytics to decide how to prioritize food safety inspections. "The city has 35 inspectors who have to inspect 15,000 stores and restaurants ideally every year, which works out to 500 inspections per year, per inspector," Catlett said. "Chicago's data scientists worked with several companies to build a machine learning model, and we were approached to help make this model more efficient and effective."

The model ranks each food establishment as high, medium or low in terms of their risk of food safety violations. This ranking is based on over 30 leading indicators, including the results of previous inspections, the size of the restaurant or the neighborhood the restaurant is in. Instead of selecting establishments to inspect more or less by random, the inspectors now order the inspections based on risk, which has resulted in significant improvement in how quickly safety problems are uncovered and addressed.

"Another indicator that might be counterintuitive until you talk to people familiar with it, is whether the restaurant has a liquor license or not," Catlett shared. "If it has a liquor license, that lowers the risk because they know that if they have a food safety violation, the city will threaten to take their liquor licence, which is a very valuable thing."

In a 60-day double blind test, the team found that the risk-based inspections were finding food safety violations about a week sooner than they would have otherwise. The model has now become part of the city's operational inspection process, generating a list of places to inspect every morning.

"This approach applies not only to food safety but any kind of inspection, such as elevator inspections or fire safety inspections," Catlett said. "When our team got involved, we introduced a framework called AutoMOMML, developed by Argonne computer scientists, that allows people who are not data scientists to evaluate their variables and see which ones are co-dependent so that their models can be simplified, and tests the performance of multiple machine learning techniques. With this framework for testing models, we estimate that we can save the data analytics team several months of work on any given problem."

"What we're trying to do is figure out how we can use data science to get out in front of problems, which was a goal articulated by Chicago's previous mayor, Rahm Emanuel. The goal is to move from reactive policies and measures to spotting trends and taking proactive measures," he said.

DEMOCRATIZING DATA

For Catlett, the real potential of HPC in smart cities is making the city a more livable place for all. Citing the example of the 'rat patrol,' which responds to non-emergency calls reported through the 311 line, Catlett said that the frequency of calls is typically a poor indication of where the problems are.

"For example, in a city like Chicago, there are potholes every spring after the roads thaw out. If you just went by 311 calls, you would think that the more wealthy parts of the city had worse roads, as they are the ones calling about potholes," he said. "In reality, the poorer parts of the city actually have worse roads; it's just that they don't call 311 about them."

When the city adopted a model for predicting rat infestations rather than depending solely on calls, they came across the worst rat infestation in recent history in a part of the city that doesn't normally call 311. "That was a real turn of the corner with big data because it showed that the new processes could be more equitable than the traditional processes and the city is now able to go to places where the problems are. To me, that was an important contribution that big data has made to cities," Catlett said.

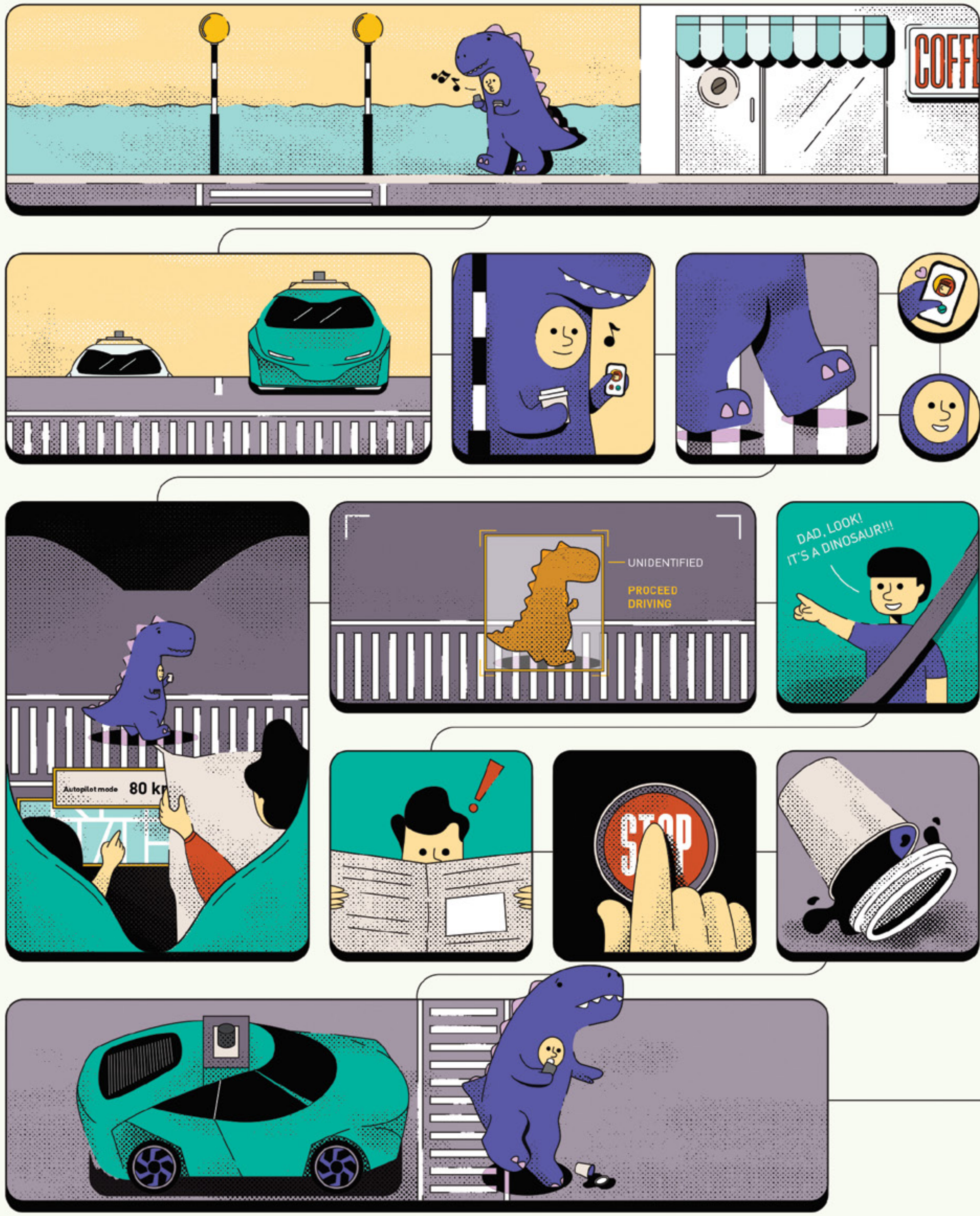
Nonetheless, Catlett acknowledges that data can be dangerous if used improperly. This is why he says every research project should begin with two questions: what is it that you're trying to accomplish, and is it an ethical thing to do in the first place? "Beyond accomplishing your goal, you also need to think of ways that the capability could potentially be misused," he said.

> **"By doing all our artificial intelligence at the edge, we can have this very strong, auditable and verifiable privacy policy."**

Though most of the measurements from Array of Things are about the physical environment or aggregate flow of people and vehicles, the policies and mechanisms that would protect privacy with respect to images was one of the key issues raised during public consultations, underlining how privacy is a key concern for citizens where smart cities technologies are being deployed. To ensure the safety of images and sound data collected by the sensor nodes, the team adopted HPC on the devices themselves, analyzing the images to extract information such as the number of people or vehicles and then deleting the images rather than transmitting them to be analyzed by centralized servers and saved.

"This is HPC at the edge. By doing all our artificial intelligence at the edge, we can have this very strong, auditable and verifiable privacy policy," Catlett explained. "We also wanted to make decisions at the edge with a latency that was short enough for smart city research where infrastructure might be controlled in real time. Here, if you send information back to a server and wait for an answer it will take too long. This was particularly important to transportation researchers who wanted to experiment with intelligent intersections and vehicle-to-infrastructure communication."

Whether out at the edge or back in a supercomputing center doing more simulations, HPC is going to be critical in almost every aspect of making cities smarter. "It's an exciting time to be in HPC and we're just getting started," he concluded.



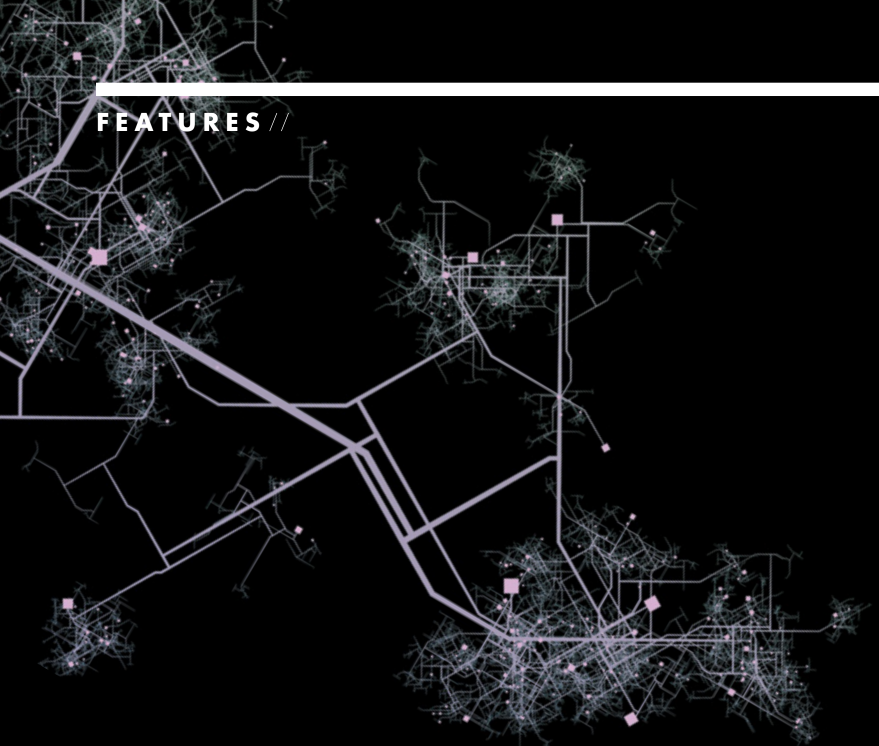
Illustrations by Lam Oi Keat/*Supercomputing Asia*

KEEPING SMART CITY ALGORITHMS ACCOUNTABLE

Why humans are still needed in the loop

Although artificial intelligence promises to give city planners unprecedented insight into urban life, humans must remain involved in decision making.

By **Sara Watson**



Despite the hype, smart cities are nothing new. Take it from Rob Kitchin, a professor of human geography at the National University of Ireland, Maynooth, who has seen the discourse shift from 'wired' cities in the 1980s to 'cyber' cities in the 1990s and even 'sentient' cities in the early 2000s.

In its latest manifestation as 'smart,' much of the focus has centered on the data and sensors that support fine-grained digital feedback. But there's far more to smart cities than data, however voluminous. What differentiates the smart cities movement from previous attempts to make urban life better will be the use of artificial intelligence (AI) to make sense of all the data pouring out of smart cities.

"Cities have been using automated systems for a long time, for example, intelligent transport systems that can configure traffic lights based on the volume of traffic," Kitchin told *Supercomputing Asia*. "With AI, however, these systems are starting to become autonomous or semi-autonomous, as opposed to having humans in the loop."

While AI enables city planners to process vast amounts of data and uncover hidden patterns, the use of AI also raises the attendant questions of fairness, accountability and transparency. These questions of smart city governance are situated in broader debates about the ethics, governance and accountability of AI, but will have to be tackled head on—perhaps more urgently than for other applications of AI—if city planners are to maintain the trust and confidence of their constituents.

DATA AND ITS LIMITS

The aspirational vision for AI in smart cities is a god's-eye-view dashboard of city management and planning. Tech firms are leaping to solve the city-wide data management problem with offerings like Alibaba's City Brain and Huawei's Intelligent Operations Center and +AI Digital Platform to create an operating system layer for the city.

AI promises to rationalize disparate datasets and manage real time data flows. But the reality for many cities today is that AI remains limited to one-off inquiries and planning exercises; it is not yet a part of day-to-day operations running real-time data streams.

One reason for this is that AI is primarily a tool for automating decision making, a prospect planners are primed to be wary of. As recently as the 1950s, modernist planners believed in a comprehensive, predictive and objective science of planning. They imagined that even the most messy human spaces could be understood with the right data and models to analyze the city. This assumption resulted in top-down plans that were limited by what they measured, and which failed to understand the lived experience of citizens on the ground.

Led by urban activists like Jane Jacobs, who famously succeeded at preventing a highway from being built through the middle of Greenwich Village in Manhattan in the 1960s, planning has since taken a strong turn towards participatory modes to include constituents and stakeholders. Today, planners recognize that space is a highly politicized subject, with a complexity that often cannot be reduced to numbers.

So it is perhaps not surprising to find planners and urban scientists hesitant to cede control to the black box once again. Leaving city-wide dashboards and masterplans to AI systems risks ushering in a new wave of technocratic directives that could mistake data for ground truth.

DESIGN-DRIVEN DATA

Although AI has advanced our ability to analyze data, that data and the decisions made with it are by no means perfect or objective. "If decisions about what is best for society are ceded to algorithms, whose notion of civic paternalism or stewardship is embedded in those systems?" Kitchin asked.

These are some of the most pressing questions in assessing the impacts of AI. And it is only more pressing in municipal applications where constituents' lives and wellbeing are at stake. As Kitchin points out, those deploying AI technologies have a responsibility to evaluate if their values align with those of the algorithms or models. No computational process is neutral; fitting data to a curve is a value-laden choice about what we are optimizing towards. Are we optimizing for cost savings? Profit? Sustainability? Fairness and inclusion?

Rather than designing cities using data, what could be more important is ensuring that our use of data is underpinned by design thinking principles, said Associate Professor Bige Tuncer of the Singapore University of Technology and Design (SUTD). "It is not so much about data-driven design as design-driven data; you need to understand what you want to get out of your computation," she explained.

"We are going in the direction of responsive cities, where instead of the technology being the most dominant aspect, we look at how the collection, analysis, visualization and interpretation of the data can support the processes that designers and planners have to undertake to make cities more livable."

For example, in one of her research projects on designing for 'liveliness' in public spaces, Tuncer used workshops to incorporate the feedback of stakeholders such as residents, turning their input into weights of a computational model. Tuncer described this method as an example of human-in-the-loop planning in practice, where weights and values in the model are expressly informed by local expertise.

WATCHING THE WATCHERS

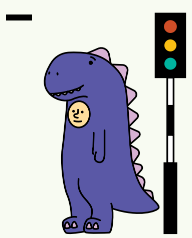
On the technical side of things, AI governance research has centered on methods to mitigate the potential challenges and harms of AI decision making. Computer scientists are formalizing tests that can be run on algorithms to audit their decision making outcomes for fairness, accountability and transparency. These may become part of the auditing system for due diligence either in deploying home-grown algorithms or in procuring industry solutions.

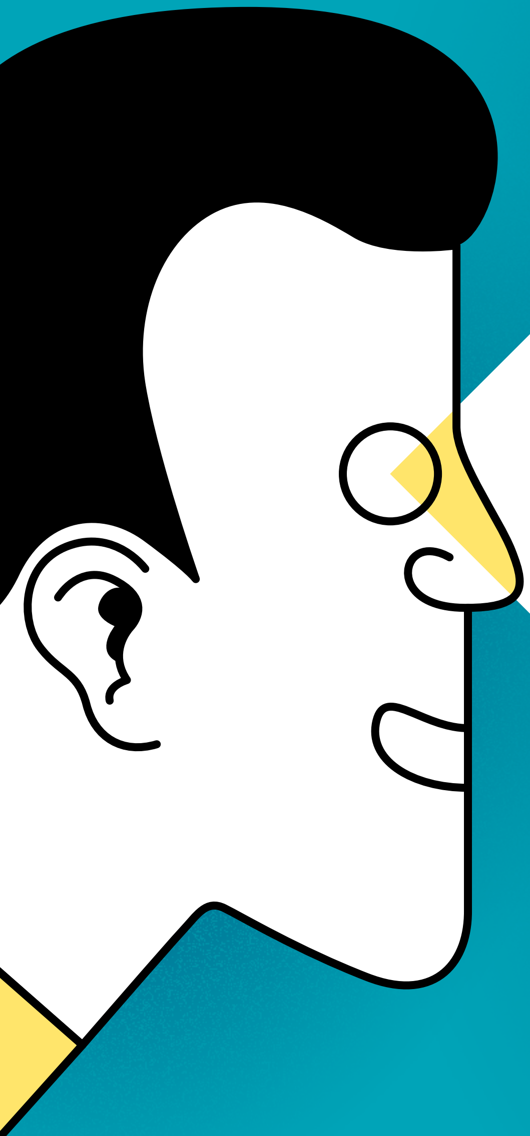
In planning physical spaces, authorities typically incorporate public consultation as part of their planning review process. Smart city procurement processes, however, rarely include public review—putting sensors on a lamppost is seen as a technical operational choice rather than a political one. Kitchin observes that seems to be especially true in city governance structures where authority is consolidated and centralized, as is the case in leading smart cities like Singapore and Barcelona.

While it is early days for smart city-specific AI auditing, Kitchin expects that auditing processes will become commonplace in the near future. AI accountability reviews would slot in naturally with existing data compliance audits to meet requirements for regulations such as the European Union's General Data Protection Regulation and Singapore's Personal Data Protection Act. Such oversight will require city planners to express the values and goals for which these systems ought to be optimizing.

If we hope to avoid another pendulum swing in planning practices back towards unexplainable black boxes, smart cities will have to build in mechanisms for constituent participation, accountability and governance of the algorithmic systems that make cities smart. The truly smart city is one that serves its constituents and accounts for them in both their planning process and AI models.

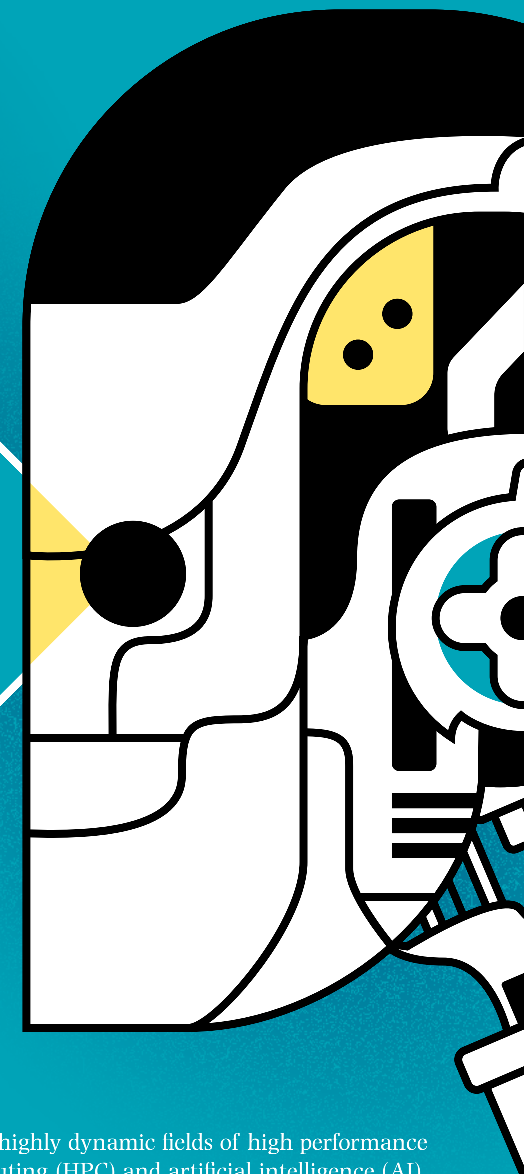
> **"It is not so much about data-driven design as design-driven data."**
>
> **Associate Professor Bige Tuncer**
> Singapore University of Technology and Design





STUDENTS TAKE THE STAGE

A spotlight on the next generation of HPC-AI talent



Given time, training and resources, students can achieve amazing feats, as demonstrated by the teams taking part in the 2019 APAC HPC-AI Competition.

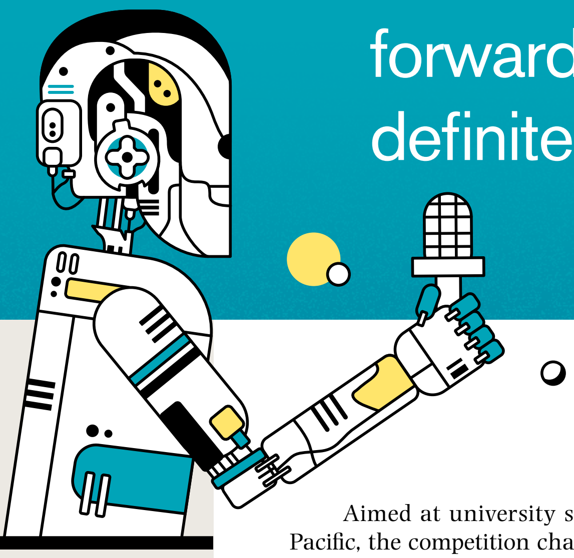
By **Rebecca Tan**

Illustrations by Shelly Liew/*Supercomputing Asia*

In the highly dynamic fields of high performance computing (HPC) and artificial intelligence (AI), it is easy to get caught up in the frenetic pace of change. But technology is ultimately only a tool. What is needed for HPC and AI to reach their true potential is users and developers who can take the technology forward; in other words, talent.

And yet, it can be daunting for young people hoping to enter the exciting fields of HPC and AI. While many learning resources are now available online, there is still nothing like hands-on experience and real-life datasets. To identify and support the next generation of HPC-AI professionals, the HPC-AI Advisory Council and the National Supercomputing Centre (NSCC) Singapore came together to organize the APAC HPC-AI Competition, now in its second year.

> “The world is facing a sea change in the way we will operate going forward, and that transformation will definitely be powered by HPC-AI.”
>
> —Professor Tan Tin Wee
> Chief executive, National Supercomputing Centre Singapore



Aimed at university students across the Asia-Pacific, the competition challenges teams to come up with optimized solutions for HPC and AI problems while staying within strict power constraints. This year saw 26 teams participating, ten more than the 16 teams that took part in the inaugural edition the year before. The 2019 competition also saw greater diversity, with new teams from India, Sri Lanka and Malaysia throwing their hats into the ring.

For the HPC component, teams had to run SWIFT, an application for simulating the effects of gravity and hydrodynamics on astrophysical objects such as planets and galaxies. Designed to simulate the entire universe, SWIFT is used by astronomers to test theories of how black holes form and how the universe evolved from the Big Bang up to the present day. Teams were tasked with running SWIFT over 32 nodes and assessed on how much they were able to speed up the performance of the application.

For the AI part of the competition, the students were asked to use machine learning to correctly identify images in the ImageNet dataset and judged on the accuracy they could achieve. Using TensorFlow, an open source software library developed by Google for deep neural networks, the teams were given 90 minutes of training time and up to 32 GPUs to classify the over 1.5 million images in the ImageNet dataset.

“The world is facing a sea change in the way we will operate going forward, and that transformation will definitely be powered by HPC-AI,” said Associate Professor Tan Tin Wee, chief executive of NSCC, addressing the students gathered at the APAC HPC-AI Competition award ceremony held on August 26, 2019.

“Even if you are not the top winner today, all of you are winners already for having made it through so many cycles of competition. You represent the crème de la crème of HPC-AI in the world; whatever you do, your future is extremely bright,” he said.

MEET THE WINNING TEAM: TEAM BOWHIVOLAR

Team members:
Chuang Chun-chieh, Hsieh Hsien-ting, Yang Chun-wei, Huang Cheng-yu, Hsu Zhong-wei and Liu Yen-fu

Coaches:
Hwang Chi-Chuan and Li Chao-Chin

From left to right: Associate Professor Tan Tin Wee; Yang Chun-wei; Liu Yen-fu; and Mr. Song Qingchun, senior director of market development at Mellanox.

Reaching a training speed of 11,720 images per second over 16 GPUs and improving the performance of SWIFT nearly three times helped the ‘Bowhivolar’ team from Taiwan’s National Cheng Kung University clinch the first place.

We caught up with a member of the winning team, Mr. Liu Yen-fu, to find out what is next for the bright, young students.

What made you want to join this competition?

Liu Yen-fu: As a first-year student majoring in engineering science, I am interested in programming and saw this competition as an opportunity to improve myself. I was also hoping that it would give me a chance to broaden my experiences overseas.

How do you feel about winning this challenge?

LY: I am very happy that our team performed so well. Of course, we joined hoping to get an award, but we were not expecting to win the first prize as this is an international competition entered by many of the top universities in the world.

What is next for you and Team Bowhivolar?

LY: As winners of this competition, we will be given the opportunity to participate in the International HPC-ISC Student Cluster Competition to be held in Germany in June 2020. For this competition, we will have to prepare our own cluster and we don’t have much money for it. Nonetheless, we are excited about the challenge!

For more information on the APAC HPC-AI Competition, please visit:
http://hpcadvisorycouncil.com/events/2019/APAC-AI-HPC/

WINNERS OF THE 2019 APAC HPC-AI COMPETITION

First place:
‘Bowhivolar’ from National Cheng Kung University, Taiwan
US$5,000

Second place:
‘AnonymousPanda’ from National Cheng Kung University, Taiwan
US$3,000

Third place:
National Tsing Hua University, Taiwan
US$1,500

Merit prize:
Fudan University, China
US$1,000

HPC honorable mention:
Nanyang Technological University, Singapore

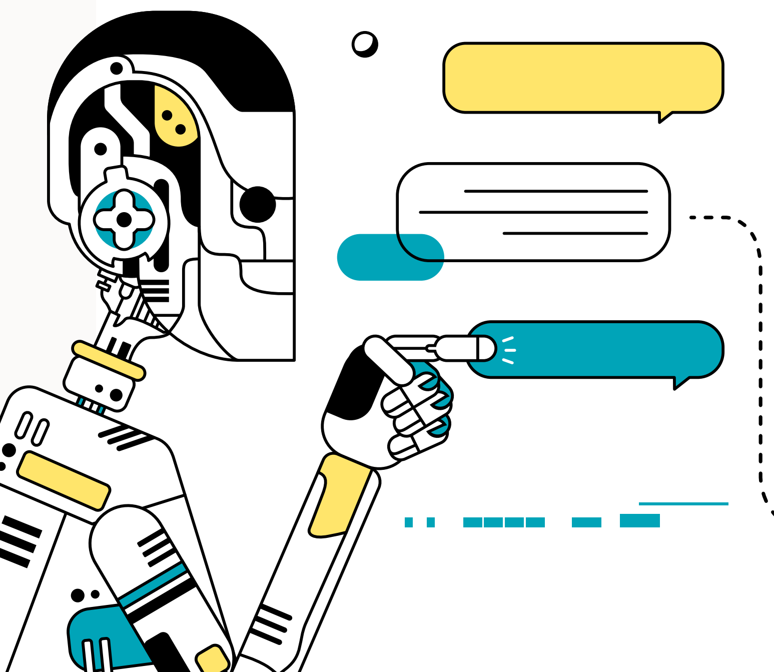
AI honorable mention:
Hong Kong University of Science and Technology, Hong Kong SAR

In addition to the awards presentation, the ceremony day also featured a technical forum which gave the students an opportunity to learn from academic and industry leaders, including one of the competition's judges, Professor Dhabaleswar K. (D.K.) Panda of Ohio State University. Panda is known throughout the HPC community for his work on MVAPICH, an open source software powering some of the fastest supercomputers in the world.

The Message Passing Interface (MPI), a standardized way for different computing nodes to pass instructions to each other, forms the very foundation of parallel computing. MPI libraries act as the middleware for HPC applications, allowing applications to seamlessly move from platform to platform. However, different implementations of MPI work with different types of networks, and when the InfiniBand network entered the market in 2000, there were no MPI libraries available to take advantage of its high bandwidth and low latency.

Sensing an opportunity, Panda and his team immediately got to work on an MPI library tailored for InfiniBand and other high performance networks. Since MVAPICH was demonstrated at the Supercomputing Conference in 2002, MVAPICH has been downloaded over half a million times by more than 3,000 organizations in 89 countries, underscoring its importance to the global HPC community. We caught up with Panda at the sidelines of the awards ceremony to find out more about MVAPICH, as well as his opinions on the convergence of HPC and AI.

Professor D.K. Panda
Ohio State University, USA



What motivated you to develop the MVAPICH open source software?

DK Panda: Direct memory access (DMA) allowed us to move data between the input/output and memory without involving the CPU. InfiniBand brought us the concept of remote direct memory access (RDMA), allowing us to transmit data to the memory from node to node with very few CPU cycles. However, when InfiniBand first came out, nobody knew how to redesign an MPI library using RDMA. We were the very first ones to come up with MPI libraries that could run over InfiniBand networks with very low overheads. This is what gave the network performance and scalability.

How has MVAPICH evolved since it was first launched?

DKP: We have continuously been enhancing MVAPICH over the last 18 years. Seven or eight years back, we worked on having a very tight integration with GPU clusters, combining RDMA with GPUs to deliver very good performance. We are keeping our eyes open for new developments, like when InfiniBand goes from 200 gigabits per second (Gbps) to 400 Gbps next year. Other than higher bandwidth, we are also developing support for features such as adaptive routing, which would allow the system to recover if a node fails.

Why are exascale systems a challenge for existing MPI libraries?

DKP: You can think of MPI like a car with many internal components. Let's say you want a regular car to become a Formula One race car. You would have to redesign the wheels, the transmission and so on. In the same way, when you take an MPI library from petascale to exascale, each and every component needs to be redesigned to take it to the next level. What runs on a 16-node system is not going to run the same way on an 8,000-node system.

We are continuously re-evaluating our designs to improve not just performance and scalability but also a very important third dimension: the memory footprint. Fat nodes are coming; the latest AMD 'Rome' CPU, for example, has 128 cores per node. But if you do the math, as the number of cores increase, the memory per core is decreasing, even though memory is also increasing. When designing the MPI library, we don't want it to consume too much memory. If the MPI library takes up 70 percent of the memory, then the application gets only 30 percent and the users will not be happy. So that is one big challenge we are working on right now.

What can HPC and AI practitioners learn from each other?

DKP: I think that there is a good momentum taking place in both fields right now, with each one trying to leverage the other. At the same time, there is also convergence taking place. The training phase of AI, in particular, is essentially a HPC problem, no different from weather forecasting, molecular dynamics or any other 'traditional' HPC problem.

In deep learning, the biggest challenge is for people to come up with solutions such that training can be done faster, so that whatever used to take years can be reduced to months, days or even hours. If you can go down to minutes or seconds, the whole field will change because people can try different kinds of solutions. As deep learning models become larger, you have to use HPC systems to make them run faster. So AI is a HPC problem and it is becoming dependent on HPC.

On the other hand, people are also trying to use deep learning or AI to make HPC simulations faster. For example, a molecular dynamics application might have normally taken seven days, but if you can analyze the patterns with deep learning, you might be able to identify the direction that is the most profitable and cut the execution time down to two days. I think that we will see a big revolution in this field over the next few years; all the traditional scientific applications will be accelerated through this principle.

Please share with us more about your recent work on libraries for big data and deep learning.

DKP: About eight years ago, my research group started to look at commonly used big data libraries like Hadoop and Spark. As each code base is different, you cannot just copy and paste our high performance MPI libraries into those architectures. So we took high performance ideas to Hadoop, Spark and so on, with our project called HiBD, high performance big data. Similarly, we are also trying to bring high performance to deep learning with HiDL. The objective of these projects is to exploit modern HPC solutions to scale out and accelerate big data and deep learning applications.

THE FUTURE IS NOW

Snapshots from SC19 in Denver, Colorado

Since its humble beginnings in 1988, the International Conference for High Performance Computing, Networking, Storage and Analysis—more conveniently known as SuperComputing or SC—has grown into the world's foremost conference for high performance computing. Centered around the theme 'HPC is Now,' the 2019 edition of the conference delved into the daily impacts of computational solutions that were once thought impossible three decades ago. SC19 was held in Denver, Colorado from November 17-22, 2019.

1


2

3


5


6

4

7

8

1. SC is back in Denver!
2. A record 13,300 eager members of the HPC community converged in Denver for SC19.
3. A warm welcome from the Fujitsu booth! Aside from displaying their latest products and projects, some exhibitors also took the opportunity to showcase their local culture.
4. Proudly bearing the country's national colors, the Singapore Pavilion highlighted its first petascale computing facility, the National Supercomputing Centre (NSCC).
5. SC19 chair Dr. Michela Taufer beams as she officially kicks off the conference. Taufer is the Jack Dongarra Professor in High Performance Computing in the Department of Electrical Engineering and Computer Science at the University of Tennessee Knoxville.
6. A dizzying array of tutorials and workshops tackling everything from parallel computing 101 to urgent decision making was offered at the conference.
7. Touted as the world's the world's fastest volunteer-built network, SCinet connects SC19 attendees with the wider HPC community by providing a platform to share the conference's activities and research all over the world.
8. In his keynote address, Dr. Steven Squyres of the Mars Exploration Rover Mission explained how HPC was central to the success of the mission's unmanned Mars exploration.

9. For students, another highlight of the conference was the job fair, where they could mingle with representatives from industry heavyweights such as NASA.

10. Sixteen teams competed in this year's Student Cluster Competition, where teams raced to assemble computing clusters and complete real scientific workloads over the course of 48 hours.

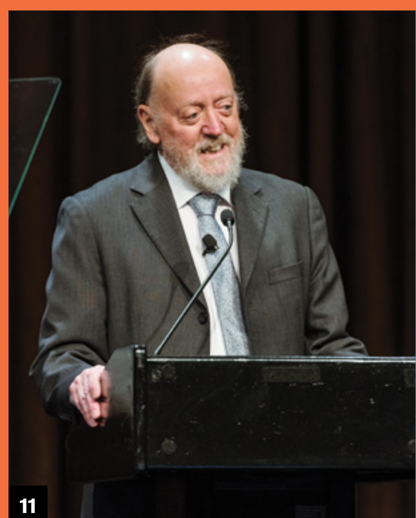
11. Professor Jack Dongarra—a leading light in the HPC community—gave a few remarks at the conference's closing award ceremony.

12. With its ninth gold medal in the Student Cluster Competition, Tsinghua University has established its reputation as the team to beat.

13. Not to be outdone by their counterparts from Tsinghua, the team from Singapore's Nanyang Technological University won the LINPACK benchmark challenge for the third year in a row, and garnered a respectable fourth place finish.

9


10


11

12

13


14


15

16

17

18

14. The 2019 ACM Gordon Bell Prize was awarded to a team from ETH Zurich for developing a simulation that maps heat in transistors.

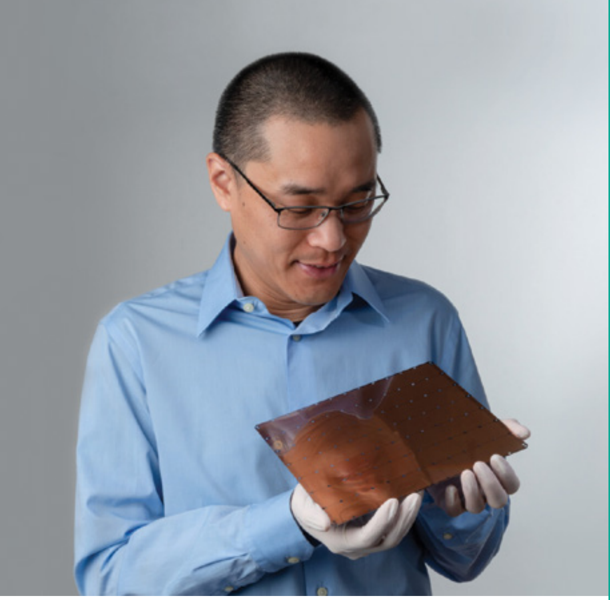
15. You're never too young to start learning about HPC! SC19 fostered an inclusive, family-friendly atmosphere by including childcare facilities and even allowing families to roam the exhibit hall during Family Day.

16. Lovingly referred to as the "SC Perennials," these scientists have attended every single SC conference since 1988. Among them was *Supercomputing Asia* editorial advisory committee member, Professor John Gustafson (first row, third from right).

17. Ms. Christine Cuicchi, director of the US Navy Department of Defense Supercomputing Resource Center and general chair of SC20, unveiled the location and theme of next year's conference.

18. Mark your calendars—SC20 will be held from November 15-20, 2020 in Atlanta, Georgia.

Business Bytes



CEREBRAS CRAMS OVER A TRILLION TRANSISTORS ON A SINGLE CHIP

Artificial intelligence (AI) startup Cerebras has produced the largest chip ever built, the Cerebras Wafer Scale Engine (WSE). The WSE accommodates 1.2 trillion transistors in 46,225 square millimeters. By processing information faster than existing microprocessors, the WSE shortens algorithm training times, currently the main bottleneck for many AI applications.

"Designed from the ground up for AI work, the Cerebras WSE contains fundamental innovations that advance the state-of-the-art by solving decades-old technical challenges that limited chip size—such as cross-reticle connectivity, yield, power delivery and packaging," said Mr. Andrew Feldman, founder and CEO of Cerebras Systems.

As WSE has both cores and memory on a single chip and 56.7 times more silicon than the largest GPU, it allows for more calculations and more efficient memory usage. The WSE houses 400,000 AI-optimized cores and 18 gigabytes of local SRAM memory. Cerebras' first buyer of this giant chip will be the US Department of Energy, which plans to incorporate them into systems at the Lawrence Livermore and Argonne National Laboratories.

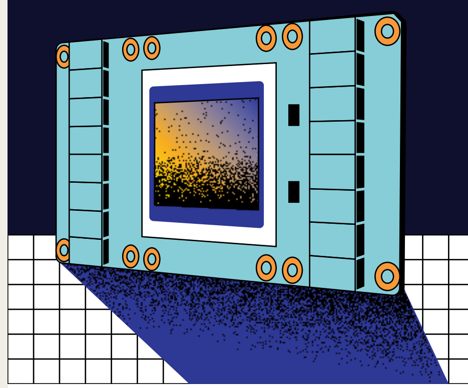
HPE ACQUIRES CRAY FOR US$ 1.4 BILLION

High performance computing (HPC) server market leader Hewlett Packard Enterprise (HPE) has cemented its position with its acquisition of Cray Inc., a storied supercomputing company. The transaction, completed in September 2019, was valued at approximately US$1.4 billion.

The acquisition gives HPE an edge over rivals Dell EMC and IBM, granting them access to Cray's existing XC and CS supercomputer range as well as their next-generation Shasta range featuring a new high-speed interconnect known as Slingshot. In 2019, Cray was awarded two large government contracts to build the fastest supercomputers in the US, Aurora and Frontier.

"Bringing together Cray and HPE establishes the most comprehensive end-to-end portfolio across compute, storage, software and services in the fast-growing high performance computing and artificial intelligence market segments," said Mr. Phil Davis, president of hybrid IT at HPE.

Collaboration between the two companies is expected to benefit R&D innovation and catalyze HPE's plans to build HPC and machine learning services on its cloud solution platform, GreenLake.

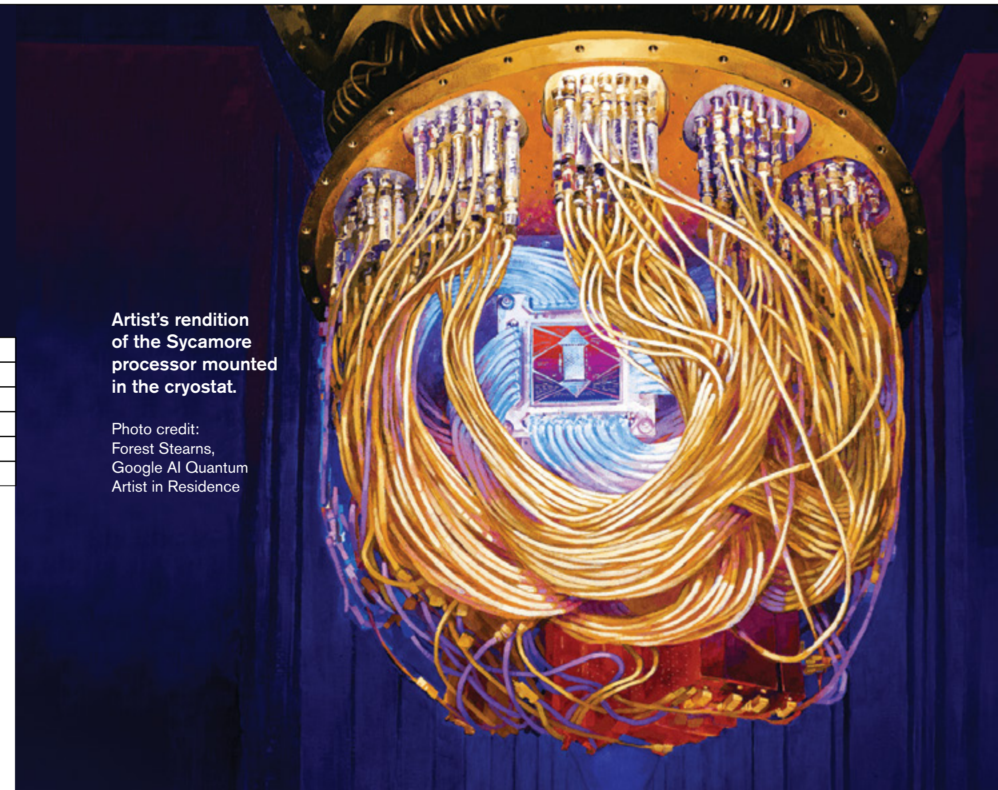


HPC NOW ON MICROSOFT AZURE

These days, users requiring HPC capabilities can choose to rent them on the cloud rather than build in-house capacity. Microsoft is the latest to join cloud service providers like Amazon Web Services and Google Cloud Platform in offering graphics processing unit (GPU)-accelerated computing. In partnership with NVIDIA, the company now offers access to up to 800 V100 Tensor Core GPUs over their Azure cloud platform.

According to the non-profit research organization OpenAI, the amount of computation required to train AI models doubles every three-and-a-half months. HPC, particularly GPU-accelerated computing, can help to meet this demand, allowing users to train their models in minutes or hours rather than days.

"Until now, access to supercomputers for AI and high performance computing has been reserved for the world's largest businesses and organizations," said Mr. Ian Buck, vice president and general manager of Accelerated Computing at NVIDIA. "Microsoft Azure's new offering democratizes AI, giving wide access to an essential tool needed to solve some of the world's biggest challenges."



Artist's rendition of the Sycamore processor mounted in the cryostat.

Photo credit: Forest Stearns, Google AI Quantum Artist in Residence

GOOGLE CLAIMS QUANTUM SUPREMACY

In a paper that was leaked to the public a month before being published in *Nature*, researchers at Google said that they have achieved the milestone of quantum supremacy, where their 54-qubit Sycamore quantum processor performed a task beyond the capabilities of even the most powerful supercomputers.

The benchmark task—sampling the output of a pseudo-random circuit—would have taken the state-of-the-art Summit supercomputer 10,000 years but was completed by Sycamore in just 200 seconds. Google's quantum computing competitors, IBM, contested the claim of quantum supremacy, contending that the task would have taken two and a half days and not 10,000 years as calculated by Google.

As the number of qubits that can be strung together increases, the computational power of quantum computers grows exponentially. However, qubits are currently error-prone, short-lived and difficult to work with, although this milestone demonstrates that scientists are making swift progress towards programmable qubits.

"When I entered this field in 1996, I wasn't sure I'd be alive by the time we got to this point," said Dr. Eleanor Rieffel, co-author of the study and a group leader at NASA's Quantum Artificial Intelligence Laboratory. "Now, we can play around with quantum algorithms we couldn't run before."

SCHNEIDER ELECTRIC DIPS INTO LIQUID COOLING COLLABORATION

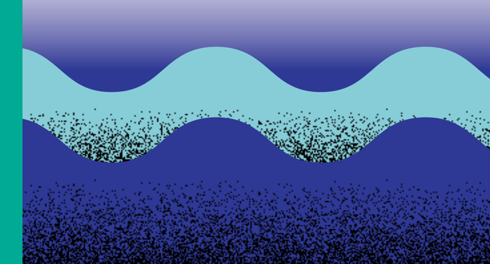
Energy solutions provider Schneider Electric has entered into a partnership with immersion cooling startup Icetope and technology integration company Avnet Integrated to develop liquid cooling technology. The partnership was announced at Schneider Electric's Innovation Summit held in Barcelona on October 2–3, 2019.

The new partnership builds on Schneider's existing relationship with Iceotope, which in 2014 received US$10 million in Series A funding from investors including Schneider. Schneider's interest in liquid cooling is driven by their prediction that it will become increasingly important to deal with the immense amount of heat generated by GPUs.

"Compute-intensive applications like AI and IoT are driving the need for better chip performance. Our quantitative analysis and testing of liquid cooling approaches shows significant benefits to the market," said Mr. Kevin Brown, chief technology officer and senior vice president of innovation and secure power at Schneider Electric.

Compared to traditional air-cooled solutions, liquid cooling is more efficient and the absence of fans and pumps make it silent. Furthermore, all components are sealed which makes the system more resilient to dust and smoke.

See our infographic on p. 38 for more information of the different types of cooling!



Super Snapshot

THE COST OF COOLING

With great power comes great heat

Today's supercomputers are more powerful—and hotter—than ever. Continued progress towards exascale computing requires innovative cooling techniques.

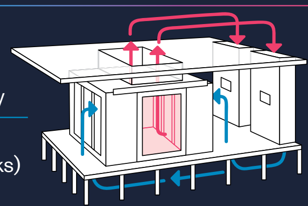
As anyone who has ever spent too long on a phone call knows, computers—even pocket-sized ones—generate heat. While a phone has anywhere between four to eight processors, Summit, the most powerful supercomputer in the world, has 2.4 million, including 27,648 extra-hot NVIDIA Tesla V100 GPUs. Clearly, USB-powered fans just aren't going to cut it.

In fact, cooling typically takes up half the energy costs of running a supercomputer, and heat transport can account for as much as 96 percent of the machine's volume. If the next generation of exascale computers are going to meet their power usage effectiveness goals, radical new cooling methods will need to be implemented.

Computer room air conditioners

Cheap to buy

Flexible (can move racks)



Unsuitable beyond 15kW

Expensive and inefficient to run

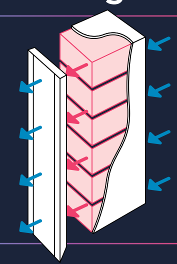
How it works:
Cold air is piped through a raised floor and collected through chimneys in the roof

Cool facts:
Air conditioning accounts for 10% of global energy consumption

Rear door heat exchangers

Easy to retrofit existing systems

More efficient as it is closer to the heat source



Still requires cooling

Leak risk

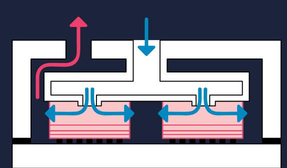
How it works:
Fans push hot air towards the back of the rack where it is cooled down by chilled water (~18°C)

Cool facts:
Water has 3,500 X the thermal capacity of air

Direct-to-chip, cold plate cooling

Warm water (>40°C) can be used, no chillers required

Waste heat can be used to heat buildings



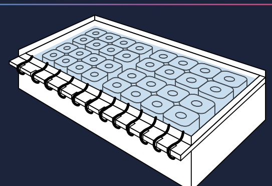
Expensive copper and piping required

How it works:
Water or a refrigerant is passed over a microchannel heat sink or cold plate in contact with the chip

Cool facts:
The Frontier supercomputer uses 5,900 gallons of water per minute for cooling

Immersion cooling

No exposure to oxygen or dust, prevents corrosion



Requires modification to components

Difficult to repair

How it works:
Entire system is immersed in a non-conductive coolant, usually oil

Cool facts:
Other than supercomputers like the Tsubame KFC, immersion cooling is popular with bitcoin miners

SCAsia Supercomputing 2020

Organised by the National Supercomputing Centre (NSCC) Singapore, SupercomputingAsia 2020 (SCA20) is an annual conference that encompasses an umbrella of notable supercomputing and allied events in Asia. SCA20 will be held from 24 to 27 February 2020. The key objective of the SupercomputingAsia conference is to promote a vibrant and relevant HPC ecosystem in Asia. Delegates will be able to gain access to visionary insights from thought leaders in academia and industry, optimum networking opportunities and the Supercomputing community in Asia.

Gathering the Best of HPC in Asia

www.sc-asia.org

24 to 27 February 2020
Singapore

SCA19 Highlights

Attended by more than **700** delegates from more than **26** countries

SCA19 **Guest of Honour, Mr Heng Swee Keat,** Minister for Finance and Chairman of the National Research Foundation Board

Get the latest HPC technology updates

Opening Keynote Speaker, Dr Christine Ouyang, Distinguished Engineer and Master Inventor, IBM

Optimal networking opportunities

WHAT'S IN SCA20?

INDUSTRY
HPC Technology Updates & Case Studies

SCIENTIFIC
Paper Presentations at Supercomputing Frontiers Asia

STRATEGIC SHOWCASE
Strategic verticals on the use of HPC

CO-LOCATED HPC EVENTS

TUTORIALS & EXHIBITIONS

Organised by

SG NSCC SINGAPORE
National Supercomputing Centre

For more information about the conference, please contact us at **SCA20@sc-asia.org.**

1 Fusionopolis Way, Connexis South Tower, #17-01, Singapore 138632

www.nscc.sg

NSCCSG

Information is correct at time of printing.