Asian Scientist

Issue 02
July 2017

SUPERCOMPUTING ASIA

SUPERCHARGING HEALTHCARE

TRANSFORMING MEDICINE WITH TECH

+

THE OCEAN IS STARTING TO BOIL

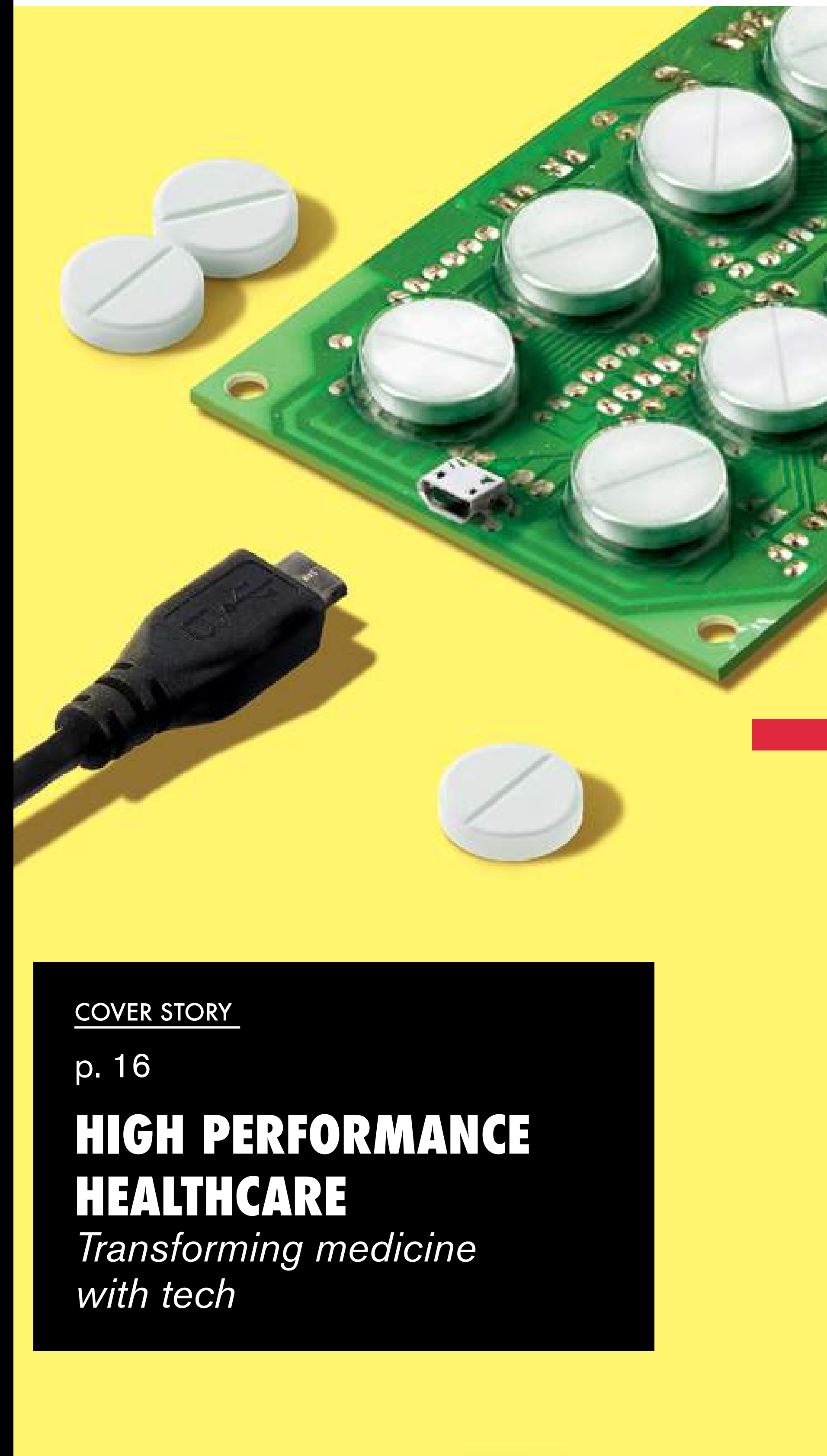
WHEN YOUR DOCTOR KNOWS IT ALL

HOW TO WIN THE NOBEL PRIZE OF SUPERCOMPUTING

MCI (P) 166/04/2017

CONTENTS

Issue 02
July 2017



EDITOR'S NOTE

When it comes to data, the healthcare system is a treasure trove. From decades' worth of patient data to meticulous records of every drug sold, each country's healthcare system easily generates terabytes of data every year.

While not all healthcare problems can be solved with high performance computing, supercomputers can play a big role in unlocking the value of all that data (*High Performance Healthcare*, p. 16). In fact, countries across Asia are already reaping the benefits, as we find out in our interview with IBM's Ms. Farhana Nakhooda (*When Your Doctor Knows it All*, p. 26).

In this issue, we hear how a new form of computer arithmetic looks set to change the way we compute (*The Ocean is Starting to Boil*, p. 22) and what it takes to win the prestigious Gordon Bell Prize (*How to Win the Nobel Prize of Supercomputing*, p. 30) from our interviewees John Gustafson and Fu Haohuan, respectively.

And if all these exciting developments have gotten you excited about the potential of supercomputers, check out our special feature on how to build one of your very own (*Build Your Own Supercomputer for Less Than $1,000*, p. 10).

Enjoy!

Rebecca Tan
Managing Editor
Supercomputing Asia

SUPERCOMPUTING ASIA

EDITORIAL ADVISORY COMMITTEE
Prof. Tan Tin Wee
Nebojsa Novakovic
Dr. John Gustafson
Yves Poppe

EDITOR-IN-CHIEF
Prof. Juliana Chan

MANAGING EDITOR
Dr. Rebecca Tan

EDITOR
Dr. Shuzhen Sim

ASSISTANT EDITOR
Dr. Jeremy Chan

CONTRIBUTORS
Jerry Lim
Jon Lau
Cindy Lim
Prof. Kenneth Ban

SALES & MARKETING
Teo Yiling

DESIGN & LAYOUT
Lam Oi Keat
Allan Chan

PUBLISHED BY
Asian Scientist Publishing Pte Ltd

DISTRIBUTED BY
Pansing Distribution Pte Ltd

HEAD OFFICE
Asian Scientist Publishing Pte Ltd
71 Ayer Rajah Crescent #07-03
Singapore 139951
Tel: 65-6909-2279
Website: www.asianscientist.com/subscribe
Facebook: www.facebook.com/asianscientist
Twitter: www.twitter.com/asianscientist

To order reprints or e-prints, or request permission to republish Supercomputing Asia content, please contact admin@asianscientst.com

WHAT'S UP!

SUPERCOMPUTING FRONTIERS REBRANDS AS SUPERCOMPUTINGASIA

Come 2018, the National Supercomputing Centre Singapore will produce a 'super' conference that brings together supercomputing events ranging from student competitions to workshops and exhibitor tutorials.

Building on three years of success with Supercomputing Frontiers, SupercomputingAsia 2018 (SCA18) will host prominent keynote speakers in high performance computing across the industry, scientific and academic space. The inaugural annual conference, which will encompass an umbrella of notable supercomputing events, will be held in Singapore at Resorts World Convention Centre from 26–29 March 2018.

In addition to highlights such as updates from countries with the most powerful supercomputers and the latest industry happenings, SCA18 will also feature co-located events such as the Conference on Next Generation Arithmetic (CoNGA) and Asia Research Platform Workshops. There will also be industry-specific showcases on strategic verticals such as biomedical & health, advanced manufacturing & engineering and offshore & marine, among others.

For more information, visit sc-asia.org

WHAT
SupercomputingAsia 2018

WHEN
26–29 March 2018

WHERE
Resorts World Convention Centre, Singapore

HPC SUPERSTARS GATHER FOR THE 30TH EDITION OF SC

The International Conference for High Performance Computing, Networking, Storage and Analysis will be making its return to Denver, Colorado from 12–17 November 2017.

Since 1988, the SC Conference series has showcased the latest developments in the world of HPC via its highly competitive technical program. Within that program will be a number of talks, panels, research papers, poster presentations and even tutorials, all designed to keep the community up to date on the bleeding edge of HPC.

In addition, the Students@SC program will feature professional development programs, technical sessions and a plethora of other initiatives designed to help students transition from the university environment into the professional HPC world.

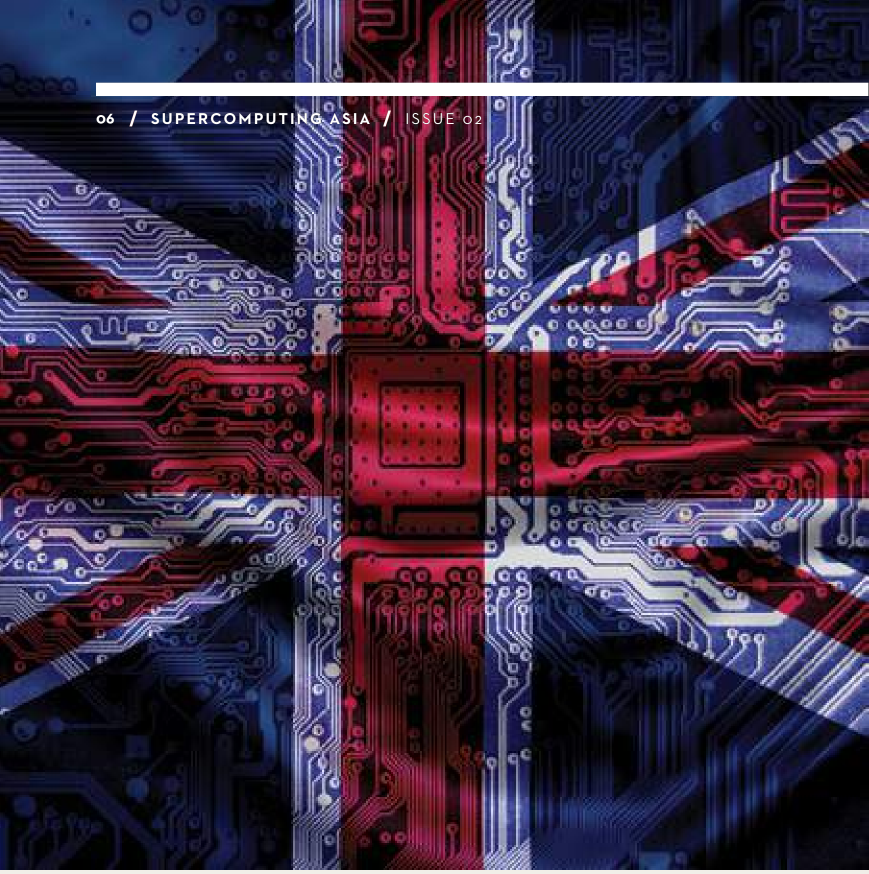
SC17 will also see the return of the SC Awards, which recognize industry individuals and teams for their contributions to the field. In addition to the highly anticipated ACM Gordon Bell Prize, other prizes include the IEEE-CS Seymour Cray Computer Engineering Award and the IEEE-CS Sidney Fernbach Memorial Award.

For more information, visit sc17.supercomputing.org

WHAT
SC17

WHEN
November 12–17 2017

WHERE
Denver, Colorado, US



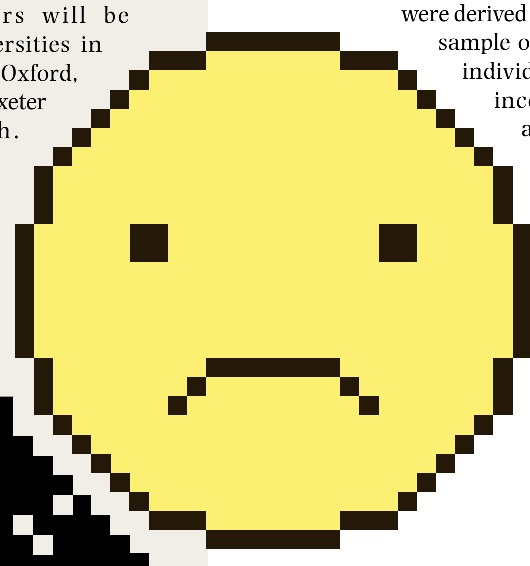
UK INVESTS £20 MILLION IN SIX NEW SUPERCOMPUTING CENTERS

As technical computing demands increase, governments and institutes are investing more heavily in HPC and supercomputing to keep up with the pace of innovation. The UK is no exception, having recently invested £20 million in six new supercomputer centers via the Engineering and Physical Sciences Research Council (EPSRC).

The new centers will provide users with access to GPU-accelerated and advanced RISC machine (ARM)-powered clusters, and will act as a bridge between university HPC centers and the main UK supercomputing framework, ARCHER.

"These centers will enable new discoveries, drive innovation and allow new insights into today's scientific challenges," said EPSRC CEO, Professor Philip Nelson. "Many universities are involved in the six centers, and these will give more researchers easy access to high performance computing."

The centers will be located in universities in Bristol, London, Oxford, Loughborough, Exeter and Edinburgh. The initiative was launched in Birmingham in March 2016.



SUPERCOMPUTER USED TO DETECT DEPRESSION

Cognitive neuroscientists at the Texas Advanced Computing Center (TACC) have trained the Stampede supercomputer to diagnose depression. Their findings have been published in *Psychiatry Research: Neuroimaging*.

The researchers used machine learning algorithms to analyze large datasets containing a variety of records, ranging from MRI scans of human brains to genetic markers and other biological factors, to enable the computer to predict at-risk patients based on historical population data.

"We looked at how to optimize [data from the brain scans] to diagnose people and detect certain vulnerabilities that might indicate future problems," said lead author Professor David Schnyer of the University of Texas.

In controlled trials, the researchers were able to detect depression in patients with 75 percent accuracy. The supercomputer is not quite ready to replace the psychiatrist, the researchers qualified, as these results were derived from a relatively small sample of less than a hundred individuals. Further studies incorporating genomics and other data are set to be run on Stampede 2, TACC's 18 petaFLOP supercomputer scheduled to be operational in mid-2017.

TSUBAME 3.0 SET TO BE MEAN, GREEN AI MACHINE

The Tokyo Institute of Technology (Tokyo Tech) Global Scientific Information and Computing Center (GSIC), along with SGI and NVIDIA, have begun construction of a next-generation supercomputer called TSUBAME 3.0, set to begin operations in the second half of 2017.

The theoretical performance of TSUBAME 3.0 is 47.2 petaFLOPS in 16-bit half precision mode or above. Once TSUBAME 3.0 is operating alongside the current TSUBAME 2.5, GSIC will be able to provide a total computation performance of 64.3 petaFLOPS in half precision mode or above, making it the fastest processing supercomputer center in Japan.

"Artificial intelligence is rapidly becoming a key application for supercomputing," said Mr. Ian Buck, general manager of accelerated computing at NVIDIA. "NVIDIA's GPU computing platform merges AI with HPC, accelerating computation so that scientists and researchers can tackle once unsolvable problems."

The computational power of TSUBAME 3.0 will not only be used for education and research within Tokyo Tech, but will also continue to provide "supercomputing for everyone" through the Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (JHPCN) and the High Performance Computing Infrastructure (HPCI).

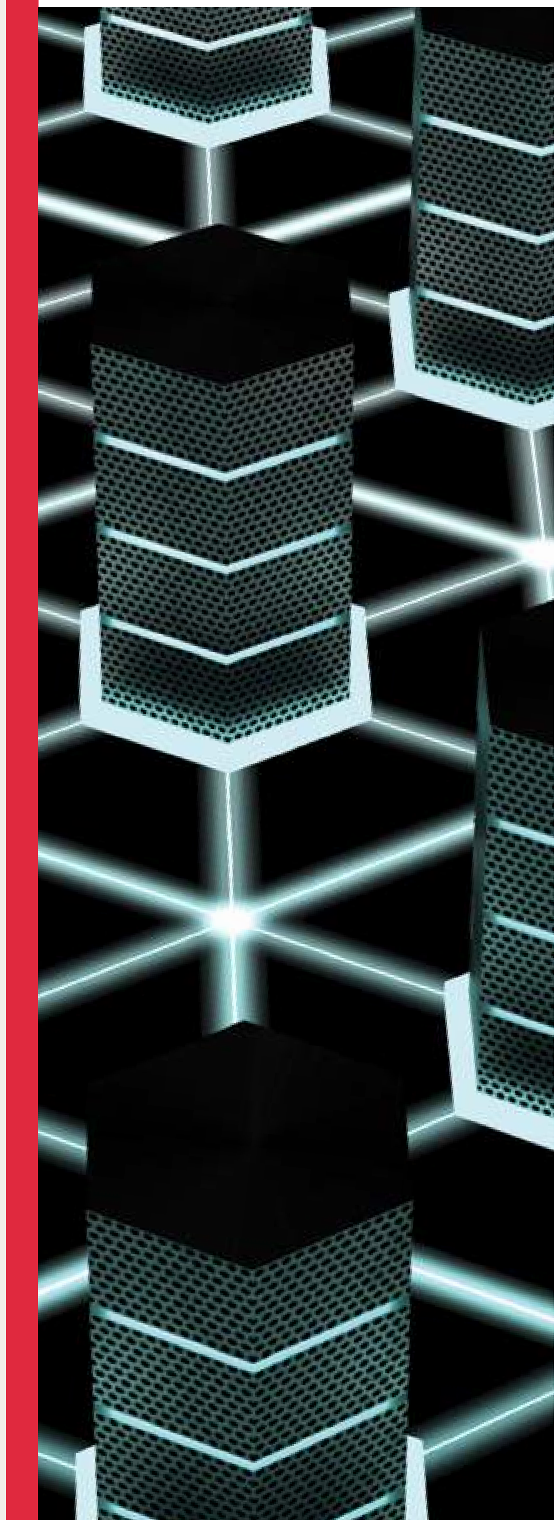
RIKEN TAKES A DEEP DIVE INTO DEEP LEARNING

The RIKEN Center for Advanced Intelligence Project has placed an order with Fujitsu for a new supercomputer, set to be one of Japan's most powerful machines for artificial intelligence research.

Called Deep Learning System, the platform is designed to accelerate R&D into AI technology. It possesses four petaflops or four quadrillion floating-point operations per second in total theoretical processing performance.

"The RIKEN Center for Advanced Intelligence Project's system will be the most powerful DGX-1 customer installation in the world," said Mr. Jim McHugh, vice president and general manager at NVIDIA. "Its breakthrough performance will dramatically speed up deep learning research in Japan, and become a platform for solving complex problems in healthcare, manufacturing and public safety."

The supercomputer, which went online in April 2017, will allow researchers to better manage healthcare for the elderly and aging infrastructure, as well as respond more efficiently to natural disasters.



Digital Dispatch

CAPTURING ASIA'S GENETIC DIVERSITY

Asia is home to 40 percent of the world's population, yet remains woefully underrepresented in global genetics surveys. The non-profit consortium GenomeAsia 100K, which aims to sequence 100,000 individual genomes, microbiomes and phenomes from Asian populations over the next three to four years, is on an ambitious mission to correct this imbalance.

The project is led by scientific chairman Stephan Schuster, a professor at Nanyang Technological University Singapore (NTU Singapore), the host institution of the consortium.

"Advances in sequencing, computing and mobile access mandates that we begin to study these underrepresented Asian populations," he said.

Over 2,000 genomes have already been analyzed on the Advanced Supercomputer for Petascale Innovation, Research and Enterprise (ASPIRE 1), Singapore's first petascale supercomputer run by the National Supercomputing Centre (NSCC). This pilot phase involved over 1.9 terabytes of data transfer between NTU and NSCC.

"The NSCC infrastructure and technical team have been instrumental in enabling the GenomeAsia 100K project to go forward as planned. They have provided the expertise and facilitated the processing of hundreds of terabytes of data from collaborators in multiple countries," Schuster added.

INDIA'S MET OFFICE GETS 10 PETAFLOP UPGRADE

Weather modeling is a processor-heavy game, to put it mildly, and India plans to stay on top of that game with the unveiling of their new 10 petaFLOP supercomputer in June this year.

Monsoons in India have a massive effect on the environment and people. Effects can range from transforming desert into grassland to unexpected droughts or floods. In addition, the monsoon is the main mechanism for the provision of fresh water to the country. Thus, predicting this phenomenon is a major priority for the government.

The computer will be jointly hosted by the Medium Range Weather Forecasting at Noida in Uttar Pradesh and the Indian Institute of Tropical Meteorology, Pune, and is expected to be ready for delivery in the middle of 2017.

INFINICLOUD 2.0: THE BEST OF BOTH WORLDS

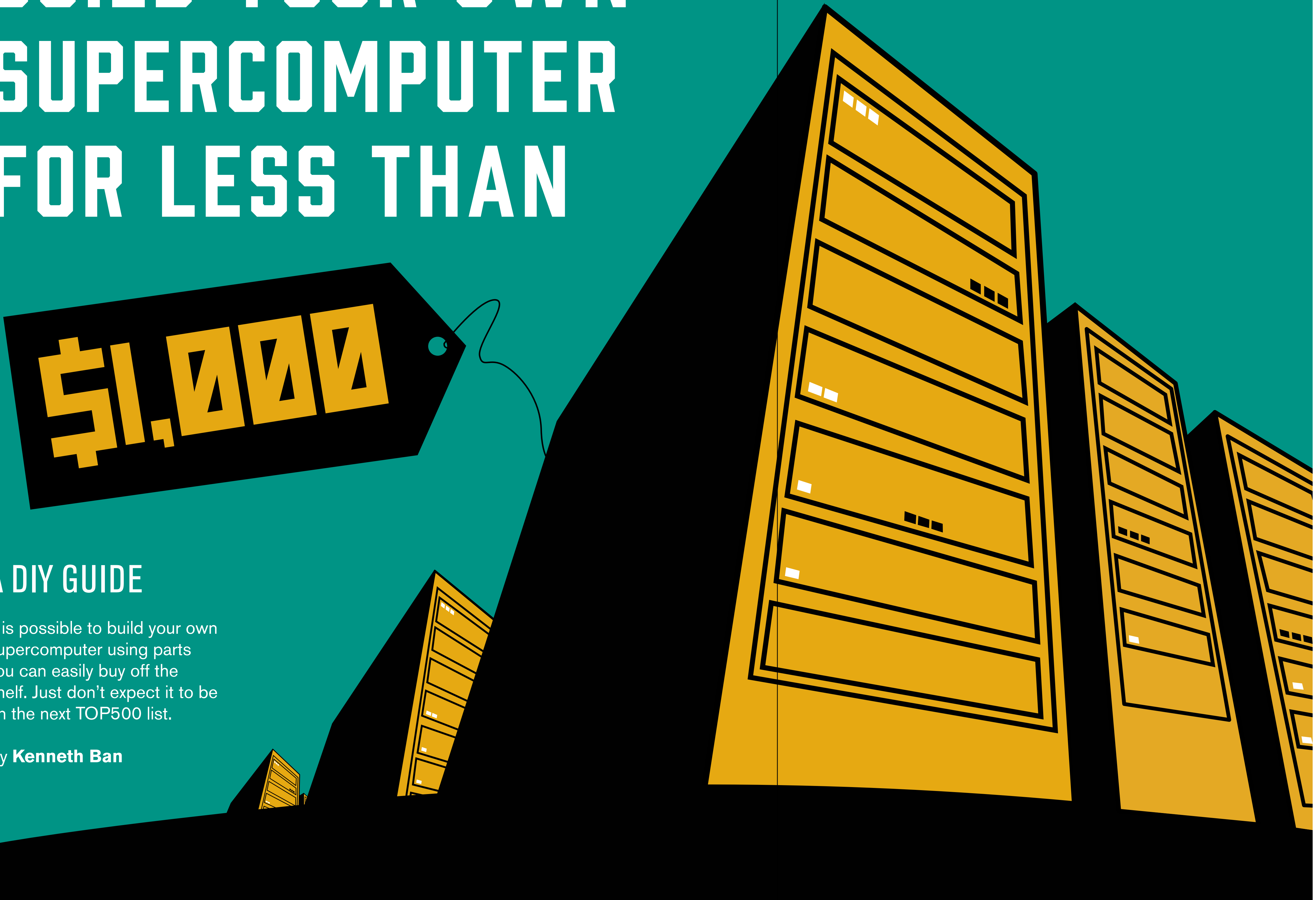
Building on the worldwide InfiniCortex architecture developed by Singapore's Agency for Science, Technology and Research (A*STAR)'s Computational Resource Centre (A*CRC), researchers in Singapore have demonstrated that distributed or cloud computing can be run on high performance systems.

"People working on high performance computing (HPC) generally do not use their systems for cloud computing because it requires a virtualization layer which degrades the performance and defeats the purpose of HPC," explained Kenneth Ban, an assistant professor at the National University of Singapore.

InfiniCloud 2.0, however, is an application that runs on the InfiniCortex network that can also bypass the virtualization layer. "This enables us to enjoy both the flexibility of cloud computing and the performance of supercomputing," Ban said.

Using InfiniCloud, the researchers were able to analyze genetic sequences in real time on systems distributed across four continents, unifying different platforms to run the analysis reproducibly. Ban envisions that a scaled-down version of InfiniCloud could be used in Singapore's hospitals to allow them to store their genomic data locally and securely while analyzing it on the National Supercomputing Centre's infrastructure.

BUILD YOUR OWN SUPERCOMPUTER FOR LESS THAN



A DIY GUIDE

It is possible to build your own supercomputer using parts you can easily buy off the shelf. Just don't expect it to be on the next TOP500 list.

By **Kenneth Ban**

Built at a cost of US$1.25 billion, the K supercomputer in Japan occupies 3,000 square meters of floor space and has an estimated operating cost of US$10 million every year. Although this kind of expense is out of reach save for the world's wealthiest individuals, it is actually not that difficult to build a simple supercomputer in your spare time.

The term 'supercomputer,' after all, refers more to how the machine is designed rather than how powerful it is. What makes machines like the K supercomputer so blazingly fast is that they break large and complex problems into smaller tasks that can be handled independently and simultaneously, by a method called parallelization.

In traditional computing, tasks are completed in sequence or serially, like a supermarket cashier serving customers one by one. Parallel computing is analogous to opening up more cashier lanes, improving the performance of the system by distributing the computing load across more nodes.

And since computing technology has now advanced to the point where the smartphone in your pocket is more powerful than a supercomputer from a decade ago, hardware components are now so widely and cheaply available that it's possible to build a functioning supercomputer for under a thousand dollars.

In this article, we'll walk you through the steps needed to build to do just that. Excited to get started? Let's go!

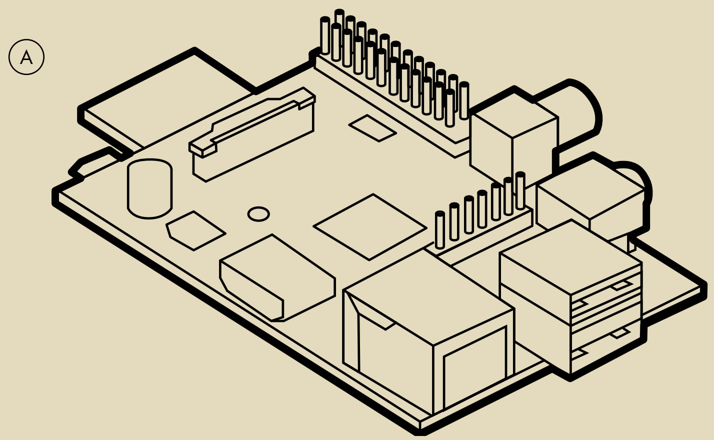
HARDWARE

The basic unit of a supercomputer is a core—the part of the machine where each individual computational task is performed, like a single cashier in our supermarket example. A single microprocessor or central processing unit (CPU) can contain one or more cores, and is usually connected to other components like memory, storage and external ports, which together make up a node.

The K computer, for example, uses a CPU that has eight cores. Individual nodes are then placed into a cabinet which holds a total of 102 nodes, 96 of them handling the computation and six of them dedicated to handling the input and output. The entire K computer system spans 864 cabinets, giving it a total of 705,024 cores [8 cores × 102 nodes/ cabinet × 864 cabinets]!

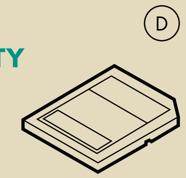
Our minimalist supercomputer is much, much more modest. All it requires is your existing personal computer which will act as the head node, compute nodes in the form of cheap Raspberry Pi computers, and credit card-sized miniature motherboards.

WHAT YOU WILL NEED:



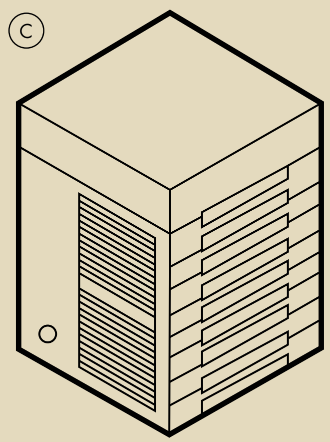
RASPBERRY PI COMPUTERS X 10

TOTAL: US$400



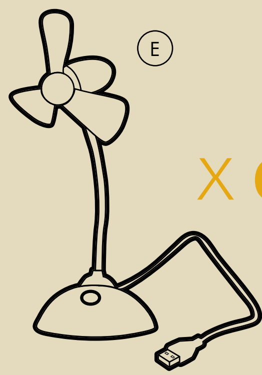
8GB CAPACITY SD CARDS X 10

TOTAL: US$80



RACK OR ENCLOSURE X 01

TOTAL: US$10



USB DESK FANS X 02

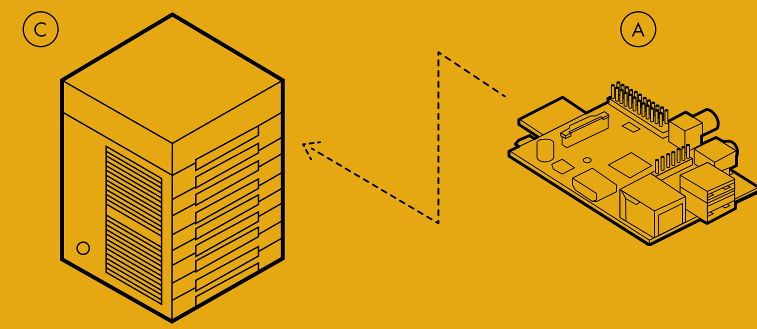
TOTAL: US$10

USB CHARGERS X 10

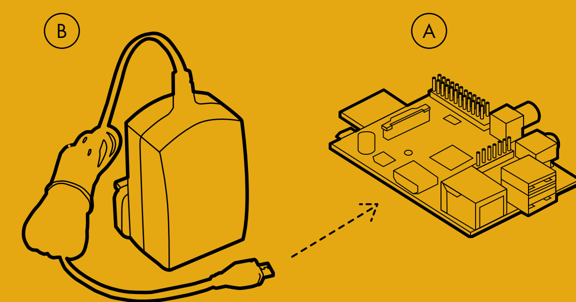
TOTAL: US$400

ASSEMBLY INSTRUCTIONS:

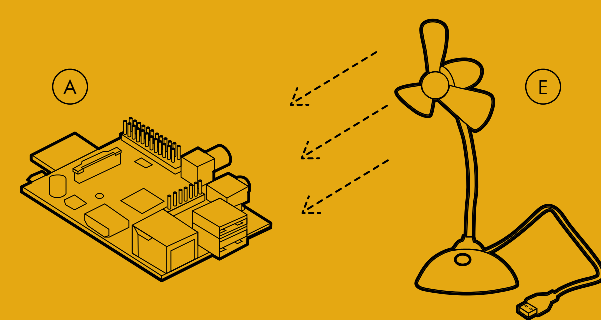
1 Assemble the Raspberry Pi computers in a rack that can accommodate all components



2 Plug in the handphone chargers which act as the power supply



3 Position the USB fan to maximize airflow



POWER HUNGRY

Obviously, machines on the TOP500 list of the world's most powerful supercomputers need much more power and sophisticated cooling systems than cellphone chargers and desk fans! The K computer, for example, uses ten million Watts of power each year, enough to power 10,000 homes. A typical cellphone charger only supplies about 3 Watts of power, but that's more than enough for a Raspberry Pi!

SOFTWARE

Even if you scale the project up and join thousands of nodes together instead of just two, you still will not have a supercomputer if all you have is the hardware. Software is essential to make sure that all those nodes are able to work on a problem together.

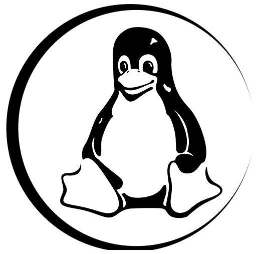
For the nodes to cooperate, they first need to all speak the same language. The first step, therefore, is to install the same operating system on all nodes. Next, you will need to set up a shared file system so that all nodes will be able to access the same data.

You will then need to write code for a job scheduler, which will dispatch the job to the different nodes. To solve certain types of problems, such as those commonly encountered in weather forecasting and molecular modeling, you will also need to give the nodes the ability to communicate with each other. This is called the message passing interface or MPI.

Problems that don't require MPI are called embarrassingly parallel, meaning that parallelizing the problem is trivial. An example would be asking the computer to count the number of times a word appears in a book. Parallelizing the problem in this cause would simply involve sending different pages to each node.

Problems that require MPI, on the other hand, are much more complex and communication-intensive. For example, predicting the weather in one region requires prediction of the weather patterns in other nearby regions, so even if the problem is split into smaller parts, the individual components will be dependent on each other and require information from each other to make calculations.

WHAT YOU WILL NEED:



LINUX OS

FREE

ASSEMBLY INSTRUCTIONS:

4 

Assign each node an IP address

5 

Install LINUX on all nodes

6 

Install your preferred LINUX graphical user interface (GUI) on the head node. CentOS is a popular version and freely available

7 

Set up a shared file system such as Lustre

8 

Add MPI calls to your application and let them cooperate

9 

Program the job scheduler

NETWORKING

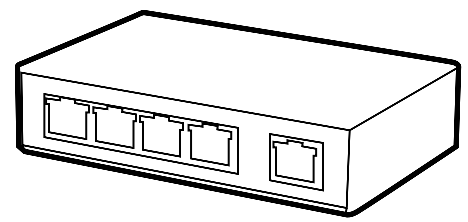
Here's where the magic happens, when all the nodes speaking the same language work in harmony. Apart from the software that we just discussed, the individual nodes need to be physically connected in a network.

If the network is thought of as a series of pipes linking the nodes, the bandwidth is the diameter of the pipes and the latency is the length. The speed of the network is determined by both parameters; if both the bandwidth and the latency are high, a large amount of data can flow but it does so slowly, as though over a longer distance.

The kinds of scientific problems that require supercomputers are those dealing with massive amounts of data, on the order of petabytes or more for the fastest supercomputers. At this scale, simply connecting two nodes over the internet is too slow. Instead, high bandwidth, low latency networks using InfiniBand or Ethernet interconnects are used.

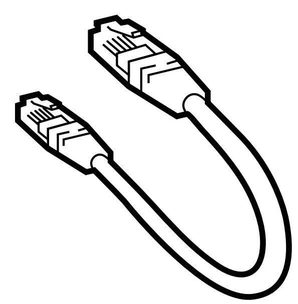
WHAT YOU WILL NEED:

F



ETHERNET SWITCH

US$60



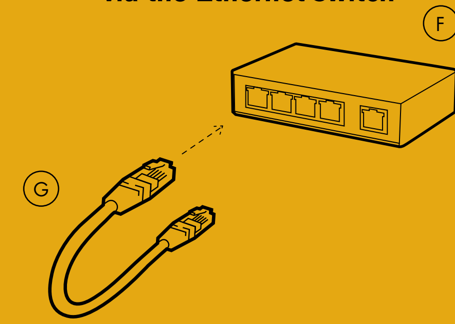
G

ETHERNET CABLES

US$30

ASSEMBLY INSTRUCTIONS:

10 **Connect the compute nodes to the head node via the Ethernet switch**



11 **Test and debug the network**

BIG LEAGUE SUPERCOMPUTING

The 'supercomputer' we've just built is clearly nowhere near the likes of the K supercomputer or other machines on the TOP500 list, but it does have the advantage of being significantly cheaper. To solve problems that supercomputers were designed to tackle, however, your best bet is probably to use your institution's or country's national resource.

Users in Singapore can register to use the ASPIRE1 supercomputer at the National Supercomputing Centre. Find out more at: https://www.nscc.sg/services/

HIGH PERFORMANCE HEALTHCARE

Transforming medicine with tech

Thanks to all the data that is already available, the healthcare sector is ripe for a data-driven transformation. Here's how supercomputers can make a difference.

By **Rebecca Tan**

No longer the preserve of academics, high performance computing has broken out of the ivory tower and become mainstream, transforming entire industries in its wake. Even healthcare, a traditionally cautious industry with legitimate concerns about the privacy of patient data, has not been left untouched.

At the Beth Israel Deaconess Medical Center in Boston, Massachusetts, for example, a supercomputer crunches 30 years' worth of data from over 250,000 patients to predict a patient's likelihood of dying over the next 30 days. In Asia, IBM's Watson supercomputer is helping doctors diagnose cancer across hospitals in Thailand, India and China. (Check out our interview with IBM's Farhana Nakhooda on p. 26!)

"Just as the effective use of data has transformed industries like banking and logistics, we should be able to do the same thing in healthcare to improve efficiency and the patient experience," said Professor Tai E Shyong of the Yong Loo Lin School of Medicine, National University of Singapore.

"These improvements do not need supercomputing per se, but there will be specific applications where we use large volumes of data to get the right services to the right patient at the right time," he said.

PUTTING THE PIECES TOGETHER

One area where data is expected to make a profound difference is in the emerging field of precision medicine, where genomics and other 'omics data are used to personalize treatments for different patients. Expecting it to revolutionize healthcare and the way it is delivered, the US government in 2016 set aside US$215 million for their Precision Medicine Initiative.

"Take the example of prostate cancer," said Dr. Saumya Jamuar, a consultant at KK Women's and Children's Hospital and clinical lead of the Singhealth Duke-NUS Institute of Precision Medicine. "If six individuals with prostate cancer walk into the clinic, only one of them is going to die from the disease. In the other five cases, the cancer is going to outlive the patient."

"It would have taken us much more time if we had used a traditional cloud-based server."

Dr. Saumya Jamuar
Co-Founder
Global Gene Corp

"If you treat all of them the same way, even though the majority have no added benefit from the treatment, they are all at risk of developing therapy-related complications. But how do you know which individual will be the one who responds well to treatment? This is one of the biggest frustrations we have as clinicians—how to choose the right therapy for the right patient?"

Thankfully, help is at hand. Scientists have identified genetic markers that have a specific pattern of expression in patients that require aggressive treatment, allowing doctors to identify and only treat those who are likely to respond to treatment, sparing the rest unnecessary surgery and expense.

Diagnosing individual patients, and identifying relevant gene markers in the first place, requires serious gene sequencing capabilities. Each individual human genome is about three billion base pairs long. To sequence a person's DNA, the entire genome is broken up into hundreds of millions of fragments of about 150 base pairs each. Computers are then used to stitch everything back together, a task akin to solving a three-billion-piece puzzle where multiple sets of the puzzle have been shaken together just to make matters more difficult.

"Supercomputers are essential for analyzing all this data. We need computational resources that are powerful enough to do the probabilistic calculations needed to do the alignments," continued Jamuar, who is also co-founder and chief scientific officer of Global Gene Corp, a genomics platform company focused on precision medicine.

"When we did an initial analysis of 300 samples, it generated about three terabytes of data. By parallelizing the analysis on the facilities provided by Singapore's National Supercomputing Centre, we were able to complete the analysis within a week. It would have taken us much more time if we had used a traditional cloud-based server."

RESTORING THE BALANCE

Supercomputers are set to play an even more important role as Global Gene Corp continues to scale up, Jamuar said. Founded in 2013 with the aim of diversifying the genetic sequences available worldwide, Global Gene Corp hopes to add tens of thousands of Asian genomes to the global pool, starting with South Asians.

"Worldwide, there is data from about 200,000 individuals that is publicly available; most of that data is from white Europeans. Asians, who make up 60 percent of the world's population, represented less than five percent of that pool when we first started," Jamuar shared with *Supercomputing Asia*. "Since then the numbers have improved slightly, but there's still a huge bias."

"Even the first version of the Exome Aggregation Consortium (ExAC), which is the largest consortium for genomic data with a dataset of over 60,000 individuals, had only 6,387 South Asians and 2,016 East Asians. The majority of the South Asians came from Pakistan, and the majority of East Asians were from Taiwan. What that has done is skewed our understanding of what a reference genome looks like," he added.

The company, which has already recruited more than 10,000 Indians to the study, has plans to do much more to rectify the imbalance.

"Precision medicine may be changing the way that healthcare is delivered in the West, but if we don't have the data for Asian populations, we are effectively excluding 60 percent of the world population. We want to change that, and this is what drives me, personally."

OF RACECARS AND READMISSIONS

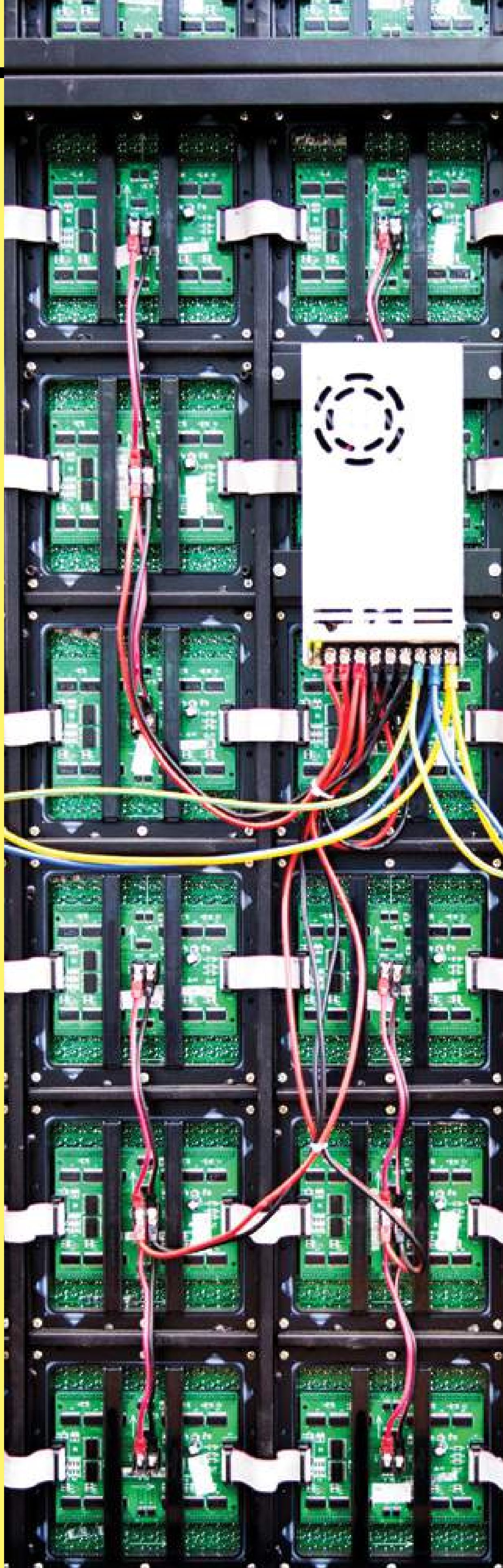
Gene sequencing and analysis is just one type of data useful for improving healthcare. Even more mundane measures like blood pressure and oxygen levels can tell us a great deal about patients. In fact, the supercomputer at the Beth Israel Deaconess Medical Center can use those parameters to accurately predict the probability of death 96 percent of the time.

Similarly, researchers in Singapore are beginning to embark on a project to use supercomputers to predict which patients will be readmitted to the hospital and when, shared Tai. According to a 2015 study by the US-based Center for Health Information and Analysis in Massachusetts, hospital readmissions cost the US government US$26 billion yearly, of which US$17 billion could have been saved in preventable cases.

Predicting which patients are likely to be readmitted requires processing a large amount of time-series data and producing the results fast, precisely what supercomputers are good at. "You can't put the data on a cluster and wait a week for the results; you need to know the answer while the patient is still in hospital so that something can be done about it," Tai continued. "Being able to get computationally-intensive analyses done in an efficient way is going to make a big difference."

The healthcare sector is getting help from an unexpected source: motorsports. McLaren Applied Technologies, the data science team that helps McLaren race cars perform at their peak, has partnered with the National Neuroscience Institute in Singapore to use predictive analytics to improve patient care.

"Over the course of a two-hour race, the McLaren team captures more than 12 billion real-time data points from over 300 sensors embedded on their race cars, running over a hundred simulations with that data," Tai said. "All this is time-series data; the same analytical methods can be applied to healthcare settings."



> **"We've only begun to scratch the surface."**
>
> Professor Tai E Shyong
> Yong Loo Lin School of Medicine
> National University of Singapore

STARTING CONVERSATIONS, BUILDING COLLABORATIONS

As transformative as the use of data can be, not all problems will require or are even best solved by supercomputers, Tai cautioned. "The question is: which problems are the most data intensive and will require high performance computing?" he asked.

"If the code isn't designed for high performance computing, it doesn't run any faster on the supercomputer than it does on a cluster in your own backyard," he shared with a laugh. "I think there's a lot of work that needs to be done; in particular, we need to understand how to optimally write our code to take advantage of parallelization."

While clinicians have a good understanding of the problems in healthcare that need to be addressed, they may not always understand what supercomputing is about and what the opportunities are. What is required, Tai said, is an ongoing conversation between clinicians and programmers.

"Making healthcare better takes a partnership," he said. "We need programmers who are not only capable of writing code that takes advantage of parallelization, but also capable of sitting down with either clinicians or investigators to design solutions collaboratively."

Once the problems have been identified, the next challenge is sharing the data for analysis. Genomic and healthcare data are typically stored at hospitals or medical centers, whereas supercomputing resources tend to be located offsite, necessitating high-speed and secure transfers. One approach has been to build networks like Singapore's InfiniCortex (see feature on p. 9) which would allow users to store data locally but analyze it in a distributed manner.

Another strategy would be to de-identify and aggregate the data before making it available online. This is what Global Gene Corp has done with its beacon network, which allows individual researchers, clinicians or country-level users to query whether gene variants are present in its database.

"It allows researchers to find each other and build collaborations," Jamuar said. "Improving healthcare can't be an individual effort, or the effort of a single company. It has to be a community effort, there has to be buy-in from four groups of people: the clinicians and researchers; the regulators, i.e., the government and related regulatory agencies; the payors, typically the insurance companies; and the patients themselves who have to be willing to participate."

UNLOCKING THE VALUE OF DATA

When it comes to healthcare, Tai noted, massive amounts of data are already there, collected over the years not for supercomputing purposes but simply for inventory management and billing.

"We've only begun to scratch the surface," said Tai. "We already have very detailed information on what services people consume, including 20 years of drug data for every Singaporean who has used the public health system. Exploiting all this data we're sitting on requires us to think a little bit about what the problems are, how we can solve them, and which of these are most effectively dealt with HPC," he added.

The future of healthcare looks set to be an exciting time, Jamuar stated, given the rapid advances in both technology and policy.

"The way I see it is that today's healthcare allows you to live long, but with diseases and disabilities. Data-driven healthcare not only allows you to live long, but also allows you to live healthy, and that is the ultimate goal of using supercomputers in healthcare," he concluded.

THE OCEAN IS STARTING TO BOIL

How posits are revolutionizing computing

Over the hundred years that they have been used, floating-point numbers have become so entrenched that replacing them would be like "boiling the ocean." But Dr. John Gustafson believes he has done just that.

By **Rebecca Tan**

For the first 30 seconds, all seemed to be going well with the first test launch of Ariane 5, a 700-million-dollar rocket developed by the European Space Agency. The boosters fired up right on cue, sending up plumes of smoke as the rocket followed its planned trajectory skyward and engineers on the ground heaved a collective sigh of relief. Just ten seconds later, Ariane 5 self-destructed, scattering smouldering debris across the launch site.

The costly error was eventually traced to a flaw in the on-board guidance system, a bug in the code caused by a calculation error. This error, according to applied physicist and mathematician Dr. John Gustafson, could have been completely avoided in a new system of arithmetic that he proposes.

WHEN FLOATING-POINTS FLOP

To understand what went wrong, you'll first need to understand how computers deal with numbers. Floating-point arithmetic is a system, developed over a hundred years ago, that represents real numbers by converting them into three components: sign, exponent and significant digits.

According to the rules determined by the Institute of Electrical and Electronics Engineers (IEEE), floating-point numbers are typically 32 or 64 bits long, with one bit representing the sign, eight or 11 bits giving the exponent, and the remaining bits for recording the significant digits (significand).

The advantage of floating-point numbers is that they allow users to deal with both very large and very small numbers by simply varying the exponent component. However, there is an inherent trade-off: The more bits you use to increase the range of numbers that can be represented (dynamic range), the fewer bits there are left to record the number accurately.

What are posits?

Posits are a new type of data that, like floating-point numbers, round if a number is inexact, but do so in a way that is simpler and more accurate.

Dr. John Gustafson
*Visiting Professor A*STAR Computational Resource Centre*

To get around this, programmers use more bits to represent floating-point numbers to get more precision for applications that require it, even if it is more computationally expensive to do so.

"In the case of Ariane 5, they were measuring speed with a 64-bit number, but feeding it into a guidance system that used 16-bit numbers," explained Dr. John Gustafson, a visiting professor at Singapore's Agency for Science, Technology and Research (A*STAR)'s Computational Resource Centre, who also holds a joint appointment with the National University of Singapore.

"Ariane's programmers specified the dynamic range, and chose poorly. What I'm pitching is to let computers manage their own accuracy and dynamic range to automatically avoid that kind of mistake."

THE POWER OF POSIT THINKING

The solution Gustafson has proposed involves entities named unums, which are the equivalent of adding a bit at the end of a number to indicate that a result is or is not exact, like writing the value of π as 3.14… indicating that there are more digits, instead of making the mistake of saying that π is 3.14 exactly. This allows computers to distinguish between when an infinite value is the correct answer and when it is in fact an overflow error like the case of Ariane 5. Recently, he created a form of unum called a posit that is designed as a drop-in replacement for floating-point numbers.

By introducing a component which Gustafson calls the 'regime,' posits allow the exponent and significand to vary in size, where bits saved by having a smaller exponent can be used to describe the significand with more precision. This unique feature gives posits accuracy where it is needed, at everyday numbers close to one where most calculations are likely to be done.

As a result, posits have a wider dynamic range than floating-point numbers of the same bit length and yet are more accurate. Gustafson has also found that in every case tested so far, posits can even get more accurate answers with fewer bits.

"That's what really motivated me—making the operands smaller and faster instead of making the transistors smaller. In effect, that would help us to continue Moore's Law for free, and more importantly, help us get to exascale computing without breaking the power budget," he said.

"If you put my arithmetic on a chip, it will take up less space in silicon and use less energy. And because things like our mobile devices and video games are constrained by how much you can get out of a watt, changing the arithmetic could have a big impact on battery life and miniaturization."

The best part about posits, however, might be the fact that they work like floating-point numbers and therefore can immediately replace them without needing to change the way things are currently done.

"Think about LED light bulbs; you can go to the store to get a light bulb that you can actually screw into the same socket as an incandescent light bulb and it immediately starts saving you energy, producing more light with less power. I believe posits can do the same for computation," Gustafson added.

> "But why are we walking when we could run so much faster?"

YOU CAN BOIL THE OCEAN

This ability to seamlessly replace floating-point numbers was the final piece of the puzzle that Gustafson has been working on for the last 35 years. When he presented an older version of the posit format to his colleagues while still a director at Intel Labs, he was told that it was futile to attempt to displace floating-point operations, that it was like trying to "boil the ocean."

"Essentially, they told me, 'Yes, there are better ideas out there, but look at all this existing infrastructure; we're stuck with it'," shared Gustafson, an inaugural Gordon Bell Prize winner who is no stranger to revolutions in computing, having pioneered parallel processing at a time when it was dismissed as impossible.

"I relish these situations. I've often played the maverick and been a disruptive influence on the computing business, and posits have all the same feel that parallel processing went through in the 1980s," he said. "The ocean is starting to boil."

Although the idea for posits was 35 years in the making, 32 of those years were false starts, Gustafson said. "Everything I tried had something that broke. It's only in the last three years while I've been here in Singapore that I've made the breakthrough, realizing that I could do a drop-in replacement for floating-point arithmetic."

RUN, DON'T WALK

In fact, Gustafson believes that posits will radically change the way we think about supercomputers. Currently, the world's most powerful computers ranked on the TOP500 list are judged based on how many floating-point operations per second (FLOPS) they are able to perform. Even though double precision 64-bit numbers are used, none of the machines are actually able to give the exact answer to the benchmark problem.

"They might get 0.99999 or 1.0001 instead of 1 and they'll say, 'Close enough.' With posits, I'm able to use one-fourth as many bits and get the exact answer," Gustafson said. "If accuracy were a criteria, using posits would destroy the rules of the LINPACK benchmark used to evaluate the TOP500."

"Using floating-point numbers is like taking part in a walking race, where they impose rules that prevent people from getting both feet off the ground. But why are we walking when we could run so much faster? Posits would let us do it the right way and go really fast."

Interestingly, posits are also particularly suitable for training neural networks, one of the hottest topics in artificial intelligence research. While typical computers take over a hundred clock cycles to calculate the sigmoidal curves used in neural network training, using posits could make the calculations a hundred times faster and use just eight bits.

VIVA LA REVOLUCIÓN

So while you may not even be aware when your computer switches from floating-point to posit arithmetic, you will be bound to feel the difference in terms of speed and battery life.

"I once calculated that people have put over 100,000 man-years of time into playing Angry Birds. If we converted that game from 32-bit floats to 16-bit posits, it would save thousands of barrels of oil!" Gustafson said, laughing.

But far beyond games, he predicts that posits will make a profound impact wherever calculations are necessary. The revolution, however, might take some time.

"It takes about ten years for an idea to get used everywhere in practice. This is the pattern and it doesn't change, no matter whether it is 1950 or 1980 or 2017," Gustafson contended.

"Singapore has been a wonderful place for me to develop these ideas, and I'd love to see Singapore become known as the place where arithmetic changed forever and revolutionized the world," he said. "And it will be."

COMING SOON

Gustafson is leading the formation of the Centre for Next-Generation Arithmetic at the Agency for Science, Technology and Research (A*STAR) in Singapore. Look out for our follow-up feature in the next issue!

WHEN YOUR DOCTOR KNOWS IT ALL

Paging Dr. Watson

IBM Watson is changing the practice of healthcare in both hospitals and homes with its ability to learn and apply huge volumes of information.

By **Sim Shuzhen**

When *Jeopardy!* champion Mr. Ken Jennings realized that he had been beaten by IBM's Watson supercomputer in a 2011 match, he bowed out with a touch of gallows humor. "I, for one, welcome our new computer overlords," he wrote by way of surrender, turning video screen into white flag.

Six years on, Watson has yet to achieve world domination. But its advanced machine learning capabilities, natural language processing skills and formidable computing power are being channeled towards a higher purpose: improving healthcare.

Watson can, for example, help oncologists find the best treatment options for cancer patients based on patient information, clinical observations and data from laboratory tests.

If this sounds like magic, it's not. With help from doctors at the Memorial Sloan Kettering Cancer Center in New York, Watson 'trained' by ingesting nearly 15 million pages of medical literature, Ms. Farhana Nakhooda, Director of Healthcare and Social Services at IBM Asia Pacific, tells *Supercomputing Asia*.

Based in the Watson Center in Singapore, Nakhooda draws from her training in biomedical research and her knowledge of IBM's portfolio to help the company's regional teams put together end-to-end solutions for their clients.

THE ONCOLOGIST'S RIGHT-HAND MAN

While Watson the *Jeopardy!* champion ran on servers, the Watson of today is a cognitive computing system that runs in the cloud. Human experts use machine learning techniques to train it to interpret and apply the huge volumes of information it ingests; once that is done, Watson continues to learn through interactions with users.

Rather than being a black box, Watson shows the user which journal articles it based its treatment recommendations on. In addition, unlike an internet search, it will request more information if it doesn't have enough to make a decision with sufficiently high confidence, said Nakhooda.

The physician, therefore, retains the final say. "It's a decision support system—if the doctor wants to do something different, it's entirely up to him or her," said Nakhooda.

Another Watson service that helps oncologists factor genetic data into their decisions is also being used in Asia, says Nakhooda. While DNA sequencing is increasingly being used to detect mutations that can then be targeted with specific treatments, analyzing terabytes of genomic data remains onerous and time-consuming. Watson automates this process, running the data against its accumulated knowledge, and coming back with treatment options.

HEALTHCARE EVERYWHERE

Watson's oncology expertise is now being used in a software-as-a-service (SaaS) format in Thailand, India, Korea and China. In Singapore, Watson has also been tested in intensive care units, where it uses real-time data from vital

sign equipment to help nurses predict which patients are prone to sepsis and organ failure, said Nakhooda.

But because it is device-agnostic, Watson can also be used for numerous healthcare applications outside of a clinical setting.

"To me that's where it gets very exciting and powerful—you could use any device in the home as a mode of communication, with Watson behind the scenes analyzing the information," said Nakhooda.

IBM's researchers in Japan, for example, have developed a personal assistant device designed to chat with elderly people in their homes. The goal: to use the conversation to detect early signs of dementia, and thus allow for early interventions, Nakhooda explained.

WHEN SPECIALISTS ARE SCARCE

Watson's oncology applications have started to gain traction in the medical community, especially now that studies are being done to validate the system's ability to recommend treatments, said Nakhooda.

For example, a study conducted by IBM and collaborators at India's Manipal Comprehensive Cancer Center found a high concordance between Watson's recommendations and those of a multidisciplinary tumor board of experts. Man and machine agreed 96 percent of the time on treatments for lung cancer, 81 percent for colon cancer and 93 percent for rectal cancer.

Doctors see a role for Watson in countries where specialists are hard to come by, says Nakhooda. India, for example, has one oncologist for every 1,600 patients. "The ability to give this tool to a more junior oncologist or to another doctor at least gives the hospital the ability to do something to treat the patient."

Watson is now a domain expert on cancer. But the Holy Grail, thinks Nakhooda, lies in the ability to recommend treatments for all manner of ailments. To appreciate how complex this problem is, consider that while a cough could be due to something as innocuous as the common cold, it might also herald the start of cancer.

"My ultimate dream would be for Watson to be used in developing countries, where there's very little clinical support. A nurse could put in some very basic information and get something back on what could be good for the patient," said Nakhooda.

For something like this to happen, Watson will have to learn and apply our entire corpus of medical knowledge—no mean feat even for a cognitive computing system.

> "My ultimate dream would be for Watson to be used in developing countries, where there's very little clinical support."

Ms. Farhana Nakhooda
Director of Healthcare & Social Services IBM Asia Pacific

HOW TO WIN THE NOBEL PRIZE OF SUPERCOMPUTING

A strong team with diverse strengths is what it takes, says the 2016 winner

Winning the ACM Gordon Bell Prize in High Performance Computing requires a team effort, says Professor Fu Haohuan, a member of the 12-man team that won the competition in 2016.

By **Rebecca Tan**

One week before Hurricane Sandy made landfall, it was the European Centre for Medium-range Weather Forecasting (ECMWF) that raised the alarm, accurately predicting the hurricane's trajectory: an unusual 'left hook' which would devastate the US, ultimately destroying over 650,000 homes and racking up an estimated US$60 billion in damages. The ECMWF prediction also showed up the shortcomings of their US counterpart, which took four days to come to the same conclusion.

Now, a team from China has developed a new weather simulation that could be a vast improvement over both systems. In 2016, the team was awarded the ACM Gordon Bell Prize in High Performance Computing for their work, becoming the first team from China to win the highly coveted award.

On the sidelines of the Supercomputing Frontiers conference held in Singapore from 13–16 March 2017, *Supercomputing Asia* caught up with Professor Fu Haohuan, a member of the winning team, to find out what led up to their historic achievement.

What makes weather and climate prediction so computationally intensive?

Fu Haohuan: Predictions only matter if they are accurate, and making accurate predictions about weather and climate systems is very challenging as they are complicated systems that are affected by many different factors.

To improve the accuracy of each element, you'd want to be running the respective sub-model at a higher resolution. For weather, we may have been using a resolution of 100 km or 20 km in the past, but now people are pushing the resolution down to 3 km or even 500 m.

In weather forecasting, people normally do not run a single model but instead use ensembles—different models with different initial conditions or parameters. Weather agencies typically run about 50 or 100, but if you increase the size of the ensembles you will get more accurate predictions.

As you can see, if we combine all these things, the computing requirement can be really high, easily filling up the largest systems in the world.

Your Gordon Bell Prize-winning work was for a fully implicit solver for nonhydrostatic atmospheric dynamics. Could you explain the difference between implicit solvers and explicit solvers?

FH: When we do the math, it's all about the equation $Ax=b$. For explicit solvers, you know A and you know x, so you just need to do the multiplication to get b. Because it is a straightforward calculation, it is easier to parallelize and get good efficiency.

Implicit solvers also use $Ax=b$, but you know A and b and now you want to compute the x. It is a more complicated calculation, but of course has its advantages, which make it worth doing. Implicit solvers are usually more stable and enable long-term simulations.

What makes it difficult to scale a fully implicit solver on a large system like Sunway TaihuLight?

FH: One of the main problems is that the equation system in our problem, like many other scientific problems, is a sparse matrix rather than a dense matrix. For this kind of computation, it is not easy to divide it into parallel tasks, unlike explicit solvers.

Usually, you are also largely bound by the memory bandwidth, because for sparse matrix operations, you read a lot of data but you don't do a lot of computation. This limits the performance quite a bit.

In our case, we had to solve both issues and try to make it scalable and also utilize the many-core architecture efficiently.

What was the biggest challenge you faced in developing this implicit solver?

FH: This was a difficult problem in pretty much every way. Thankfully, we have a strong team from many different backgrounds. The lead author for this study is Professor Yang Chao from the Chinese Academy of Sciences' Institute of Software. He worked on the numerical methods, making it scalable and suitable for a heterogenous system.

Professor Xue Wei from Tsinghua University worked on the message passing interface (MPI). He and I worked on the accelerators to find the most efficient way to run the different kernels on the many-core architecture of Sunway TaihuLight. Then we also have Professor Wang Lanning from Beijing Normal University, who is our climate scientist.

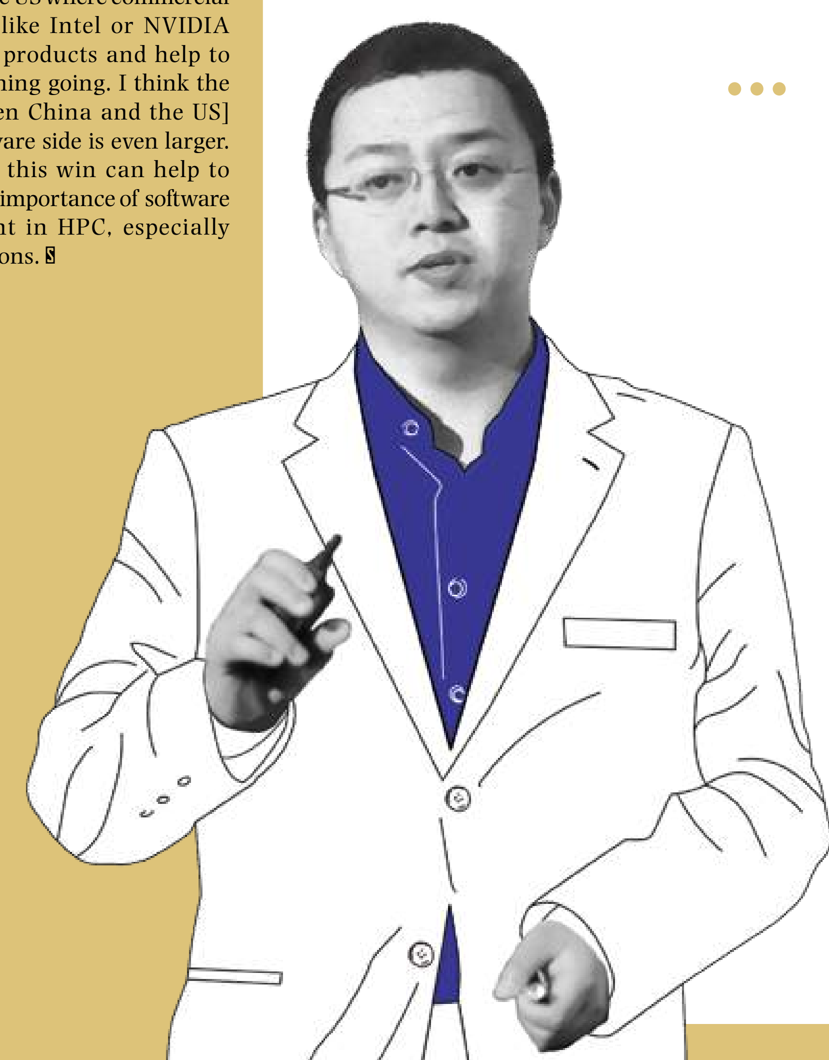
It is a really tough challenge. It is not a single's game but a team sport, like football or basketball; you need to have excellent players on different fronts to make it work.

How significant is this win for China?

FH: Winning the Gordon Bell prize is a very big breakthrough for Chinese researchers and a good recognition of the efforts of those working on HPC software and applications in China. Three of the six finalists were from China, so I was confident that one of us would win; we had a 50 percent probability!

Although China has nice hardware like Sunway TaihuLight, we do not have the ecosystem present in the US where commercial companies like Intel or NVIDIA create new products and help to keep everything going. I think the gap [between China and the US] on the software side is even larger. I hope that this win can help to promote the importance of software development in HPC, especially for applications. ▪

> "It's not a single's game but a team sport, like football or basketball; you need to have excellent players on different fronts to make it work."



A MEETING OF MINDS

Supercomputing Frontiers Singapore 2017

Since 2015, Singapore has played host to some of the biggest names in high performance computing each year at the Supercomputing Frontiers conference. Held at Biopolis from 13–17 March 2017, this year's edition featured keynote speeches from luminaries of the supercomputing world and talks by HPC practitioners on the latest updates from industry and academia. The conference also saw the launch of the National Supercomputing Centre Awards, which recognize excellence in Singapore's HPC sector.

1

2

3

4

5

6

1. Professor Gordon Bell (center), Researcher Emeritus at Microsoft, chatting with participants at a tea break on the first day of workshops.

2. Workshop participants discussing the day's events with Dr. Marek Michalewicz, Deputy Director of the Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw, Poland.

3. Associate Professor Tan Tin Wee, Organizing Chairman of Supercomputing Frontiers 2017 and Director of NSCC Singapore, addressing the full-house crowd at the opening ceremony.

4. Mr. Lim Chuan Poh (left), Chairman of Singapore's Agency for Science, Technology and Research (A*STAR), receiving the Supercomputing Frontiers Singapore Distinguished Service Award from Mr. Peter Ho (center), guest-of-honor and Chairman of the NSCC steering committee, and NSCC Director Associate Professor Tan Tin Wee (right).

5. From left to right: Professor Chan Eng Soon, CEO of the Technology Centre for Offshore and Marine Singapore (TCOMS), Mr. Aziz Merchant, Executive Director of the Keppel Offshore and Marine Technology Centre and Dr. Zhang Yongwei (right), Deputy Executive Director of A*STAR's Institute of High Performance Computing, receiving the Best Integrated Vertical Domain Award.

6. Dr. Raj Thampuran, Managing Director of A*STAR, was awarded the Supercomputing Frontiers Singapore Visionary Award.

7

8

9

10

12

11

7. From left to right: Mr. Peter Ho, Dr John Gustafson, Professor Satoshi Matsuoka, Professor Gordon Bell, Professor Fu Haohuan and Associate Professor Tan Tin Wee.

8. Professor Thom H. Dunning Jr., Director Emeritus of the Northwest Institute for Advanced Computing, US.

9. Professor Fu Haohuan, Deputy Director of the National Supercomputing Centre in Wuxi, China.

10. Dr. Alessandro Curioni, Vice President (Europe) and Director of the IBM Research Laboratory, Switzerland.

11. Professor Gordon Bell at a workshop session.

12. High school students from Singapore's National Junior College were given a chance to interact with leading HPC practitioners and showcase their own work.

13. The industry track saw presentations from leading figures including Ms. Lin Yonghua from IBM Research China.

14. At the scientific track, Dr. John Gustafson shared about the exciting possibilities of posits.

Read more about posit arithmetic in our feature story on Dr. Gustafson on p.22!

15. Dr. Du Xinxin from the Singapore-MIT Alliance for Research and Technology shared how his group is using HPC to teach autonomous vehicles how to 'see.'

16. Supercomputing Frontiers 2017 was proudly brought to you by the team at NSCC Singapore.

17. Endorsed by the team at Intel, Platinum Sponsor for SCF17.

18. Dr. Motoi Okuda from Japan's Research Organization for Information Science and Technology (RIST) giving an update on Japan's supercomputing efforts.

19. Dr. Marek Michalewicz, Deputy Director at the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM), University of Warsaw.

20. The conference was not all work and no play. Guests were treated to a dinner with a view at the networking reception held at The China Club.

21. Mr. John Shalf, Chief Technology Officer, Lawrence Berkeley National Labs, US.

22. NSCC team members and invited guests celebrating another successful run of Supercomputing Frontiers at the post-conference party held at IKKI Izakaya.

23. Mr. Yves Poppe, Consultant for NSCC, mingling with guests over drinks at the post-conference party.

Business Bytes

TENCENT CLOUD TO USE NVIDIA GPUs FOR CLOUD COMPUTE SERVICE

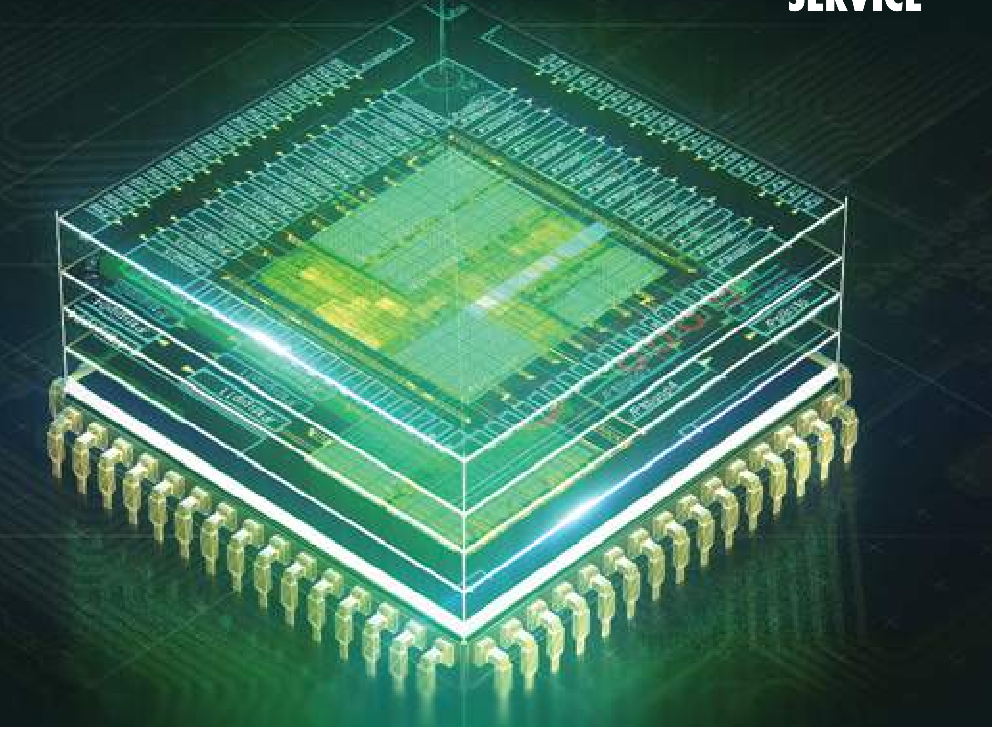
Tencent, China's largest internet company, has announced that it will be using NVIDIA Tesla GPU accelerators for their cloud compute platform, Tencent Cloud.

The partnership will see Tencent using the NVIDIA Tesla P100, P40 and M40 GPU accelerators and NVIDIA deep learning software to increase the processing capabilities of their HPC services. In doing so, Tencent hopes to enable researchers and developers to integrate AI into their products, which would give their simulations and analytics a much-desired boost in accessibility and computational power.

"Tencent Cloud GPU offerings with NVIDIA's deep learning platform will help companies in China rapidly integrate AI capabilities into their products and services," said Mr. Sam Xie, Vice President of Tencent Cloud. "Our customers will gain greater computing flexibility and power, giving them a powerful competitive advantage."

Tencent Cloud are planning to introduce servers with up to eight GPU accelerators in the first half of 2017.

IBM LAUNCHES CLOUD QUANTUM COMPUTING SERVICE



IBM have announced that they will be offering quantum computing services via their IBM Cloud Platform.

Dubbed IBM Q, the platform will enable classical computer users to make use of IBM's quantum computing processors, allowing for a wider range of problem solving and simulations which were hitherto unavailable to classical computing.

"IBM has invested over decades to growing the field of quantum computing and we are committed to expanding access to quantum systems and their powerful capabilities for the science and business communities," said Mr. Arvind Krishna, director for IBM Research. "Following Watson and blockchain, we believe that quantum computing will provide the next powerful set of services delivered via the IBM Cloud platform, and promises to be the next major technology that has the potential to drive a new era of innovation across industries."

Potential applications for quantum computing include drug and materials discovery, financial services, AI and cloud security, many of which will utilize quantum cryptography.

BASF BUYS ITS OWN SUPERCOMPUTER

BASF, the world's largest chemical company, has recently purchased a supercomputer from Hewlett Packard Enterprise (HPE). The purchase of the Apollo 6000 cluster will allow BASF to stay ahead of the game in chemistry research and development by giving the company over one petaFLOP of computational power, based on Proliant XL servers running Intel Xeon processors.

"The new supercomputer will promote the application and development of complex modeling and simulation approaches, opening up completely new avenues for our research at BASF," said Dr. Martin Brudermueller, chief technology officer at BASF. "The supercomputer was designed and developed jointly by experts from HPE and BASF to precisely meet our needs."

"In today's data-driven economy, high performance computing plays a pivotal role in driving advances in space exploration, biology and artificial intelligence," said Ms. Meg Whitman, President and Chief Executive Officer, HPE. "We expect this supercomputer to help BASF perform prodigious calculations at lightning-fast speeds, resulting in a broad range of innovations to solve new problems and advance our world."

CHINA INVESTS US$1.4 MILLION IN CAMBRICON DEEP LEARNING CHIPS

The Chinese Academy of Sciences (CAS) has set aside RMB 10 million (~US$1.4 million) to develop a brain-inspired processor chip specialized for deep learning.

Dubbed 'Cambricon,' the chip is expected to be the world's first processor that simulates human nerve cells and synapses to conduct deep learning, according to a statement issued by CAS. Cambricon is named after the Cambrian explosion, a sudden flowering of a great diversity of new life forms that began roughly 540 million years ago.

The investment from CAS will be used for basic research on both architectures and algorithms for the Cambricon system. Unlike existing neural networks which require thousands of GPU-based accelerators, Cambricon processors are designed to operate more efficiently and run on much less power.

According to an announcement made at the 3rd World Internet Conference held in China in 2016, the latest Cambricon-1A chip can handle 16 billion virtual neurons per second and has a peak capacity of two trillion synapses per second. This performance is double that of a conventional GPU but has a power consumption that is lower by one order of magnitude.

Super Snapshot

BUILDING A BRAIN-LIKE COMPUTER

Neuromorphic computing

Our brains are shaped by over **7 million** years of evolution

Human brains are **complex**…

- An average brain has 10^{11} neurons
- Each neuron forms about **7,000** connections

…but our brains are also extraordinarily **energy efficient**

- Our brains can do with just **20 watts** what would require megawatts on a supercomputer

Researchers are trying to emulate this complexity in silicon. This new interdisciplinary field, which combines computer science, biology, physics, electronic engineering and mathematics, is called **neuromorphic computing** or neuromorphic engineering.

Companies are racing to design brain-inspired chips. To date, efforts include:

- **Qualcomm**'s Zeroth program
- **IBM**'s TrueNorth brain-like microprocessor
- **Cambricon Technology**'s Cambricon-1A neuromorphic chip
- **Intel**'s yet-to-be-named design