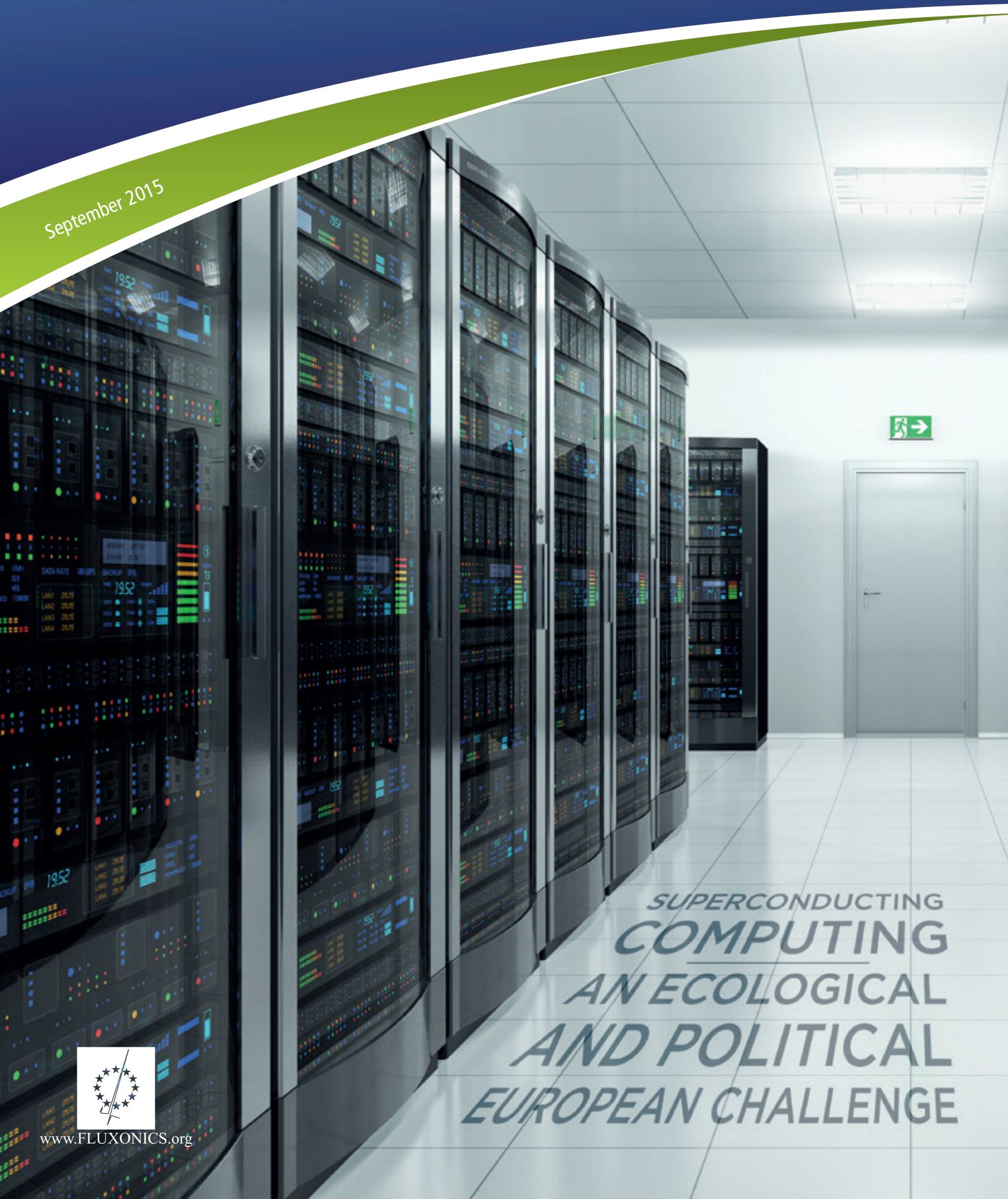


FLUXONICS

WHITE PAPER

A SPECIAL EDITION OF THE FLUXONICS NEWSLETTER

September 2015



*SUPERCONDUCTING
COMPUTING
AN ECOLOGICAL
AND POLITICAL
EUROPEAN CHALLENGE*



www.FLUXONICS.org

ECOLOGICAL CONTEXT OF HIGH PERFORMANCE COMPUTING

1 Driven by Internet traffic, cloud computing, smartphones usage and intensive data crunching applications, the number of supercomputers is growing and growing. Energy consumption by large-scale computing systems becomes more and more a non-negligible financial burden. Cooling the cloud is on the business agenda and becomes an ecological issue. Industry and data centres face the need of energy efficiency to reduce running costs: these account from 10 to 15% of the total costs of building and running a data centre for 15 years [1]. In 2010, routers and servers consumed worldwide between 1.1 % and 1.5 % (1.7 % and 2.2 % in the United States) of the total energy production. In 2011 it was estimated to be 31 GW of electric power [2].

2 Better high performance computers are crucial to run better predictive models for climate change and weather forecast, to better understand the formation of the early Universe and subatomic physics, to model earthquakes with scenarios for short term disaster response, to calculate multiple drug interactions, to model cells, for genetics, biotechnologies, to simulate

brain functions, to improve the energy efficiency of cars and planes, and more generally, for better engineering of objects used in our daily life.

3 However, high-performance computing (HPC) is a dual-use technology. Besides the rosy side mentioned above, one should not hide the fact that it is also, more and more, an instrument of power. Utilization for the so-called «financial services» requires also the regulation and if possible the control of high frequency trading. Security, independence and intelligence are also on governmental agendas. Hence the political interest in this issue, illustrated by the recent executive order of the White House “Creating a National Strategic Initiative” (July 29, 2015) in which the Complex Cryogenic Computing programme, which is already on track, is now embedded.

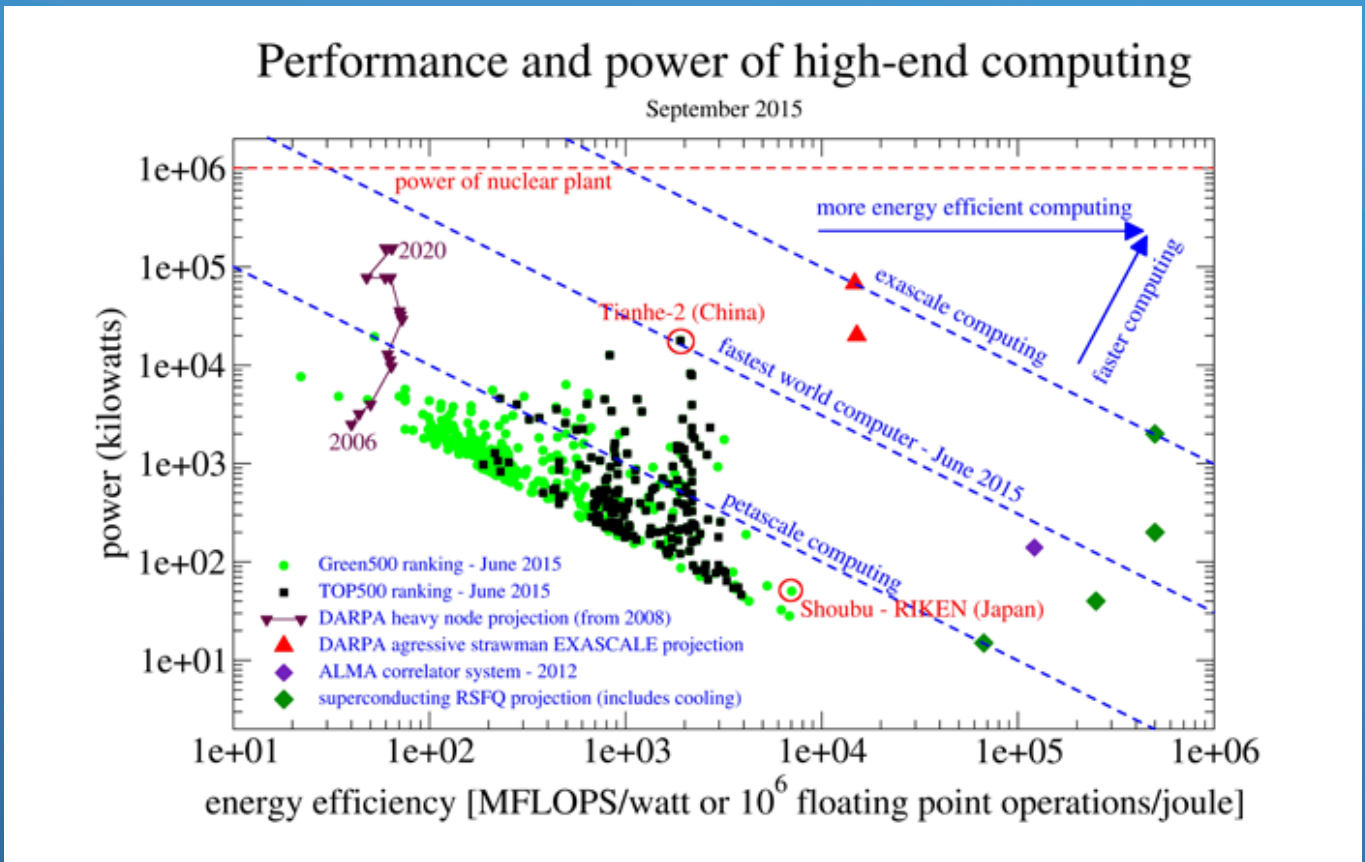
4 It is common knowledge that the doubling of the density of transistors per unit area of electronics chips roughly every two years, known as the Moore’s law [3], has so far been the main factor behind this evolution. This doubling has been accompanied by an equivalent reduction of the power consumption per device, which is known as Dennard scaling [4], to keep constant the power dissipated by the chip. Indeed, the power dissipation has increased from about 1 watt/cm² in 1985 to reach about 130 watts/cm² in 2005. During the same period clock frequencies of processors increased from about 10 MHz in 1985 to 3 GHz in 2005, corresponding roughly to a 40% increase of frequency each year for two decades.



Data centre cooling system

[1] R. Ascierto & A. Lawrence, «Will energy prices power US datacenter growth or short-circuit energy efficiency?», 451 Research's Market Insight Service- DCT Data Center Technology, 2013. <https://451research.com/report-short?entityId=76124&tmpl=print&refer=marketing> [2] D. S. Holmes, A. L. Ripple, and M. A. Manheimer, «Energy-Efficient Supercomputing – Power Budgets and requirements», IEEE Trans. Appl. Supercond. 23 (2013) 1701610. [3] Gordon Moore, «Cramming more components onto integrated circuits», Electronics, 38(8), 1965. [4] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, «Design of ion-implanted MOSFETS with very small physical dimensions», IEEE Journal of Solid State Circuits 9(5), pp. 256-268, 1974. [5] <http://www.green500.org/sites/default/files/subramaniam-tgi-hppac12.pdf> [6] T. Smalley, «Performance per What?», Bit Tech. - http://www.bit-tech.net/columns/2007/05/20/performance_per_what/1 [7] International Technology Roadmap For Semiconductors - <http://www.itrs.net> [8] Characteristics of low-carbon data centres - <http://www.koomey.com/post/54013825367> [9] E. Masanet, A. Shehabi and J. Koomey, «Characteristics of low-carbon data centres», Nature Climate Change, vol. 3, pp. 627-630, July 2013. DOI: 10.1038/NCLIMATE1786 [10] World Economic Forum Annual Meeting 2015 - <http://www.weforum.org/sessions/summary/beyond-moores-law> [11] L. Gammaitoni, D. Chiuchiu, M. Madami and G. Carloti, «Towards zero-power ICT», Nanotechnology 26(22), 222001, 2015. [12] <http://www.landauer-project.eu> [13] K. K. Likharev and V. K. Semenov, «RSFQ logic/memory family: A new Josephson-junction digital technology for sub-terahertz-clock-frequency digital systems», IEEE Trans. on Appl. Supercond., vol. 1, pp. 3-28, Mar. 1991. [14] ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems, DARPA IPTO, September 28, 2008 - <http://www.cse.nd.edu/Reports/2008/TR-2008-13.pdf> [15] No exascale for you! - an interview with Berkeley lab's Horst Simon - http://www.hpcwire.com/hpcwire/2013-05-15/no_exascale_for_you_an_interview_with_nesc_s_horst_simon.html [16] Obama orders effort to build first exascale computer, ScienceMag, July 30, 2015. <http://news.sciencemag.org/technology/2015/07/obama-orders-effort-build-first-exascale-computer> [17] A. Fujimaki, M. Tanaka, T. Yamada, Y. Yamanashi, H. Park, and N. Yoshikawa, «Bit-serial single flux quantum microprocessor CORE», IEICE Trans. Electron. E91-C (2008) 342. [18] M. H. Volkman, A. Sahu, C. J. Fourie, and O. A. Mukhanov, «Implementation of energy efficient single flux quantum digital circuits with sub-aj/bit operation», Supercond. Sci. Technol., vol. 26, no. 1, p. 015002, Jan. 2013. [19] H.J.M. ter Brake, et al., «SCENET Roadmap for Superconductor Digital Electronics», Physica C, Vol. 439, Issue 1, pp. 1-42, 2006. [20] S. Anders, et al., «European roadmap on superconductor electronics: status and perspectives», Physica C, vol. 470, Issues 23-24, pp. 2079-2126, Dec. 2010; doi:10.1016/j.physc.2010.07.005 [21] Standard Helium Liquefier/Refrigerator - http://www.linde-kryotechnik.ch/public/datenblaetter/l280-lr280_datenblatt_eng.pdf [22] Ranking the World's most energy-efficient supercomputers - <http://www.green500.org/lists/green201506> [23] The TOP500 most powerful commercially available computer systems - <http://top500.org/lists/2015/06/> [24] Atacama Large Millimeter/submillimeter Array (ALMA) – In search of our cosmic origins - <http://www.almaobservatory.org> [25] Powerful Supercomputer Makes ALMA a Telescope - <http://www.nrao.edu/pr/2012/almacorrelator/> [26] The Square Kilometre Array - Exploring the Universe with the world's largest radio telescope - <https://www.skatelescope.org> [27] FLUXONICS - <http://www.fluxonics.org> [28] Marc A. Manheimer, «Energy-efficient superconducting computing coming up to speed», November 2013 - <https://www.nsa.gov/research/tmw/tnw202/article5.shtml> [29] Marc A. Manheimer, «Cryogenic Computing Complexity Program: Phase 1 Introduction», - IEEE Trans. Appl. Supercond., 25(3), 1301704, June 2015 - <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=7029597>

PERFORMANCE AND POWER



The projected performance of superconducting RSFQ is on the lower right side of this diagram; energy savings are not obtained at the detriment of energy efficiency.

Power is now limiting growth in computing performance. It is assessed by the number of logical operations performed per second, and per watt of power consumption at the mains supply [5]. This measures the average energy efficiency of the system for a required computing task. Performance per watt is expressed in Millions or billions (Giga) of FLoating-point Operations Per Second (FLOPS) per watt (MFLOPS/W or GFLOPS/W). It is directly connected to the processors clock frequency that gives the speed at which logical operations are done on the chip. For the fastest systems petascale computing corresponds to 10^{15} FLOPS, while exascale computing refers to 10^{18} FLOPS.

However the «performance per watt» metric can be sometimes misleading since a system that runs slowly (a few operations per second) but consumes a low power can give a higher number of FLOPS/W (a higher «performance per watt») than a faster system consuming more power. For instance the power consumption of portable devices, related to battery life, is a crucial metric that has to be considered independently from the intrinsic performance of the device, so that proper trade-offs can be made. Inversely, an improved performance per watt, associated to a faster system, can sometimes hide a higher absolute power consumption [6].

For High Performance Computing, the goal is a better performance per watt associated to a faster upgraded system that requires the same power, so that the improvement of energy efficiency is associated to an increase of speed only, for the same energy need.

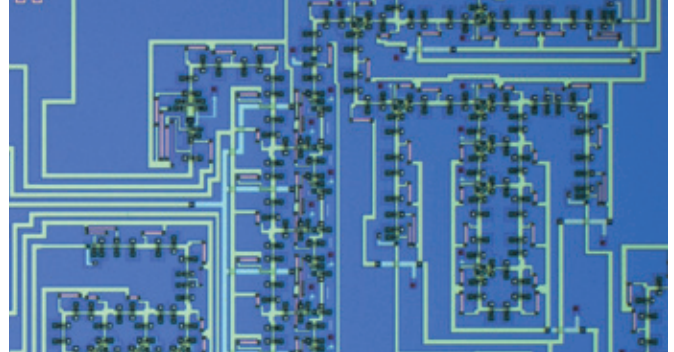


BEYOND MOORE'S LAW

This golden track is reaching its limits:

- Since 2005 there has been a strong slowdown of clock frequency improvement, with an expected increase of only 4% per year due to device scaling limits closer to atomic size, assessed in the 2011 International Technology Roadmap for Semiconductors (ITRS) [7]. In the same time the number of transistors went up from a few thousands per chip in 1985 to more than 3 billion today, with a goal of more than 6 billion by 2020.
- Dennard scaling has also slowed down since the beginning of the 21st century, primarily because of current leakage of smaller devices, 130 watts/cm² is indeed a limit set by the physical properties of air cooling.

Many techniques are being used to minimize the power consumption, like frequency and voltage control of devices, development of multicore microprocessors, or use of new dielectrics. None can overcome the power consumption problem, due to fundamental physical reasons, but, even if they will not provide disruptive solutions, these techniques will enable further technology scaling for a few more years up to about 2022. Efficient IT devices are the key to greener data centres. As stated by Koomey [8-9] «*IT efficiency (which includes higher utilization and performance improvements as well as purchasing efficient hardware) is the most important issue on which to focus. (...) just switching an inefficient data center to low carbon electricity isn't a good choice, because it uses up scarce low carbon electricity that could otherwise be used elsewhere.*»



RSFQ digital circuit

Consequently, research and developments efforts are currently devoted towards improving the energy efficiency of Information and Communication Technology (ICT) components and systems. Such measures can take place at three different levels: architecture level, system level, and device level. Any substantial improvement on the device level has a huge effect because it will be multiplied by the large number of logical switching elements within an integrated circuit (typically millions to billions).

Besides, the development of concepts for operating conventional devices in a regime of low power loss, and also the development of novel devices for unconventional computing, bears a promising potential for improving energy efficiency. Several alternative physical devices, like carbon electronics (graphene, carbon nanotubes), quantum computing [10], or reversible computing [11-12], are being studied in what is usually named beyond Moore activities.

New types of information representation have also been introduced for performing digital functions, replacing the electrical charge as information carrier. An attractive example of this approach is given when, instead of the electrical charge on electrodes within a transistor, single amounts of magnetic flux –so-called flux quanta– are used for representing binary information. This leads to the concept of Rapid Single Flux Quantum electronics – abbreviated as RSFQ [13]. This principle and demonstrations have been known for more than 25 years. As a few foundries for fabrication on a research level already exist worldwide and the understanding on how to construct such circuits is well developed, the research need has been de-prioritized over most of the last decade leading to a stagnation in funding.

RSFQ SUPERCONDUCTING DIGITAL ELECTRONICS

Nevertheless, due to the real need of progress in energy efficient computing for the ICT infrastructure, the interest for this technology has risen in the recent years, in particular with the need for exascale computing [14-16]. A practical proof-of-concept on an interesting scale of complexity has been brought in 2008 by

researchers in Japan where large-scale integrated microprocessors and network switches have been demonstrated in RSFQ technology. These circuits operate with clock frequencies above 20 GHz while consuming only milliwatts of power [17]. Considered on the chip itself, this essentially means a huge improvement in energy efficiency



Data centre building



Inside data centre

of more than 4 orders of magnitude [18]. However, as in any ICT system, cooling is a mandatory issue and as the quantization of magnetic flux is observed in superconductors, RSFQ systems must be cooled as well. For a proper assessment of the energy efficiency on system scale, these auxiliary components have to be taken into account as well.

Nowadays cooling systems in the temperature range required for superconducting circuits are commercially available with a high reliability enabling continuous operation of more than 3 years without maintenance [19-20]. The efficiency of cooling scales with the system size: the bigger the system to be cooled, the better the efficiency. It typically ranges from 5000 watts of electricity consumption per watt of cooling power at liquid helium temperature for small systems to 400 watts of electricity consumption per watt of cooling power for bigger systems [21]. In other words larger systems are more favorable concerning the overall energy efficiency. Globally the gain in energy efficiency of superconducting electronics at the device level, taking cooling into account, lies between 10 and 100, but the main advantage relies in the low required power consumption.

While, in the past, significant technical obstacles prevented serious exploration of superconducting computing, recent innovations have created foundations for a major breakthrough. These include new families of superconducting logic without static power dissipation and new ideas for energy efficient cryogenic memory. A superconducting computer also promises a simplified cooling infrastructure and a greatly reduced footprint. A recent study [2] found that RSFQ-based high-performance computing systems from the 10-petaFLOP class will be on a very favourable efficiency of 250 GFLOPS/W, including the energy required for cooling. This can be compared with the number of 7.03 GFLOPS/W for one of today's most energy-efficient supercomputers with a processing power of 33.8 petaFLOPS [22-23].

Superconducting electronics is a disruptive beyond Moore approach. Its low power requirements, its ballistic transfer of information on-chip at the speed of light and the availability of reliable and compact commercial cryocoolers derived from applications driven by the space sector and large scientific experiments make it a good candidate for high-end computing.

CONCLUSION: A POLITICAL CHALLENGE FOR EUROPE

Of the few suggestions for mid-term high-performance computing beyond Moore's law, the superconducting solution offers mature and well proven devices. The European expertise is at par with major players in the game...but the deployment phase has never been supported at the convenient level. Europe's competitiveness in this field has strongly suffered in the last years with the exception of low-temperature cryogenics.

The issue is not to pay lip service to a lobby by keeping understaffed teams alive but to launch a preindustrial deployment program- presumably outside traditional academia. It will be organized around its major technical equipment: a foundry 24/7 fully dedicated to superconductive devices allowing mass production of devices to build up a reliable production. The time is ticking for launching a European initiative starting with manpower formation in an interdisciplinary approach with code developers and chip engineers to define the milestones of this initiative.

SUPERCONDUCTING TECHNOLOGY PLATFORM FOR COMPLEX COMPUTING SYSTEMS

As the energy demands of today's high-performance computers have become a critical challenge, different activities and research programs have been launched world-wide to overcome this obstacle. *"Computers based on superconducting logic integrated with new kinds of cryogenic memory will allow expansion of current computing facilities while staying within space and energy budgets, and may enable supercomputer development beyond the exascale,"* said Marc Manheimer, program manager of Cryogenic Computer Complexity (C3) at the Intelligence Advanced Research Projects Activity (IARPA) in the US.

General use supercomputers and data centres are an obvious driver for future computing needs. However the required technological developments encompass a much wider range of applications. Such is the case for instance for the largest international astronomical project in existence: the Atacama Large Millimeter/submillimeter Array (ALMA) [24]. The processing of data has required the development of customized machine, one of the most powerful on Earth, made of 134 millions of processors for correlating data from 50 antennas [25]. Higher requirements are needed for the developments of the Square Kilometre Array (SKA) [26], the largest radio telescope on Earth. Astronomy, and in general Big Science, is in need of ever more high-performance systems. Ultimate performance requires often the cooling of systems, where superconductors are used for their low power need and quantum sensitivity. Next computing challenges also concern the real-time processing of data produced by large instruments or smart sensors, some equipped with complex imagers.

Rapid Single Flux Quantum (RSFQ) logic offers an extremely attractive high-speed and low-power computing solution. Besides, the technology is compatible with superconducting ultrasensitive sensors used for quantum computing or readout of imagers, making possible what otherwise requires hybrid integration of several technologies, as shown in the 2013 International Technology Roadmap for Semiconductors (ITRS) [7].

While the fabrication of small RSFQ circuits was mainly done on a research level so far [27], the use of RSFQ circuits for real computing requires fabrication of large-area, high-density, superconductive circuits with reasonable yield. The final goal is the development of a petaflops-scale computer. For the former Hybrid Technology MultiThreaded (HTMT) petaflops project in the US, it was estimated that 4,096 RSFQ processors operating at a clock frequency of 50 GHz at least, comprised of roughly 37,000 chips containing a total of 100 billion Josephson junctions would be required. This ambitious goal can be only reached in a step-by-step pre-industrial development. Five topical areas were identified for the development of such a superconductive computer:

- architecture & design
- memory and processors
- manufacturing superconductive circuits
- interconnects & input / output
- system integration

Developments in all areas are needed. Among them, cryogenic RAM has been the most neglected superconductor technology. Three attractive candidates are presently investigated, namely orthogonal spin transfer torque cells optimized for operation at liquid helium temperature, cryogenic spin Hall effect cells, and cells based on Josephson junctions with properties modified by magnetic layers [28-29].

Manufacturing superconductive circuits of the required complexity and in the required volume requires fabrication processes comparable to semiconductor industry in a clean room of class 3 or 4 (ISO 14644). The technological requirements for key parameters such as feature size, linewidths, complexity, processes, etc correspond to semiconductor technology of the mid-1990s. The estimated costs are around 200M€ for a five-year period. This time will be needed to develop and to establish the reliable fabrication of prototype circuits containing from 10,000 over 100,000 to 1,000,000 Josephson junctions per chip. The circuits will be fabricated on 200 mm Silicon wafers in sub micrometer technology (junctions diameters of 0.5 μm at the beginning and 90 nm in the long term using deep-UV steppers for lithography and chemical-mechanical polishing (CMP) for planarization).