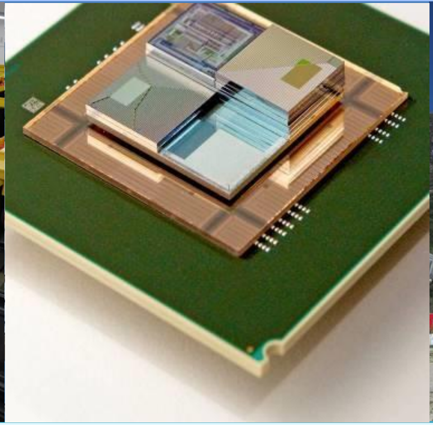
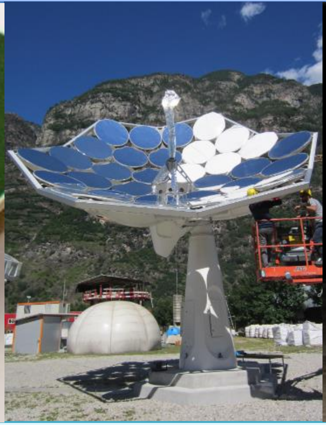
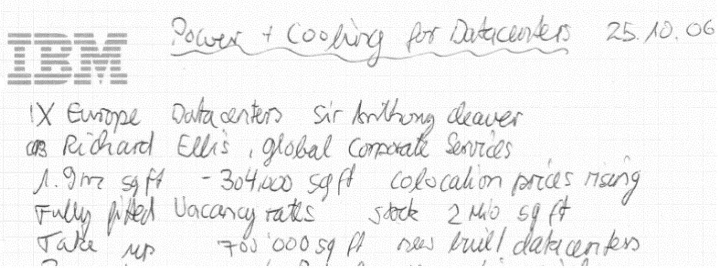
Ultimately Dense and Efficient Future Computers

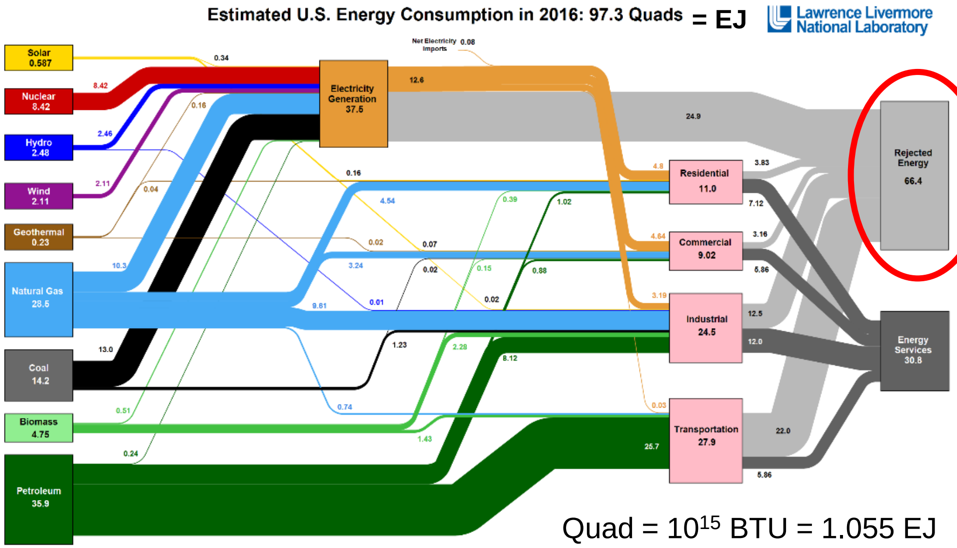
Bruno Michel, Mgr. Smart System Integration, IBM Research – Zurich, Switzerland
Member US National Academy of Engineering and Member IBM Academy of Technology
https://www.zurich.ibm.com/st/ https://www.zurich.ibm.com/st/smartsystem/

Reason for Energy and Efficiency Research



➔ Paradigm Change 1: Human energy consumption is main cause for global warming

➔ Paradigm Change 2: Running a datacenter is more expensive than buying a datacenter



Quad = 10^{15} BTU = 1.055 EJ

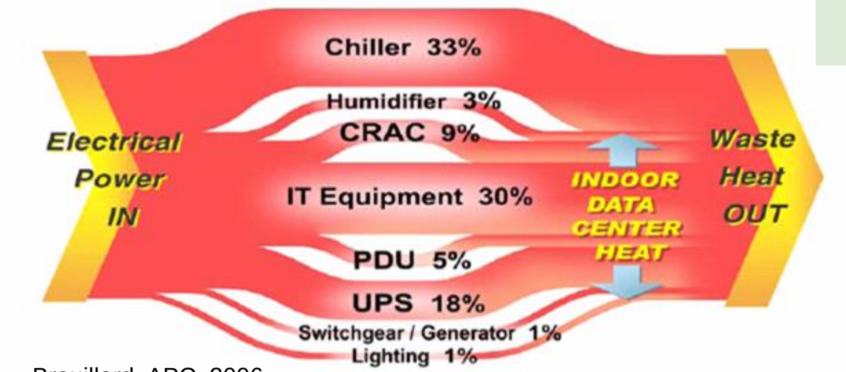
All energy carriers have risks.
The only risk free "energy" is efficiency

- **Coal and Oil, Greenhouse** due to CO_2
- **Gas, Greenhouse** due to CO_2 and 30x due to methane
- **Hydro, flooding,** limited due to rainfall patterns
- **Geothermal, fracking can trigger earthquakes**
- **Solar, radiative forcing** (not included in diagram)
- **Nuclear, accidents** and heating of nuclides
- Wind, light flicker and sound
- Waves limited geographically

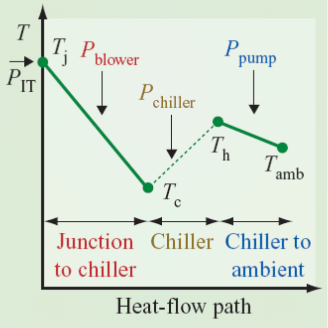
Bruno Michel, bmi@zurich.ibm.com

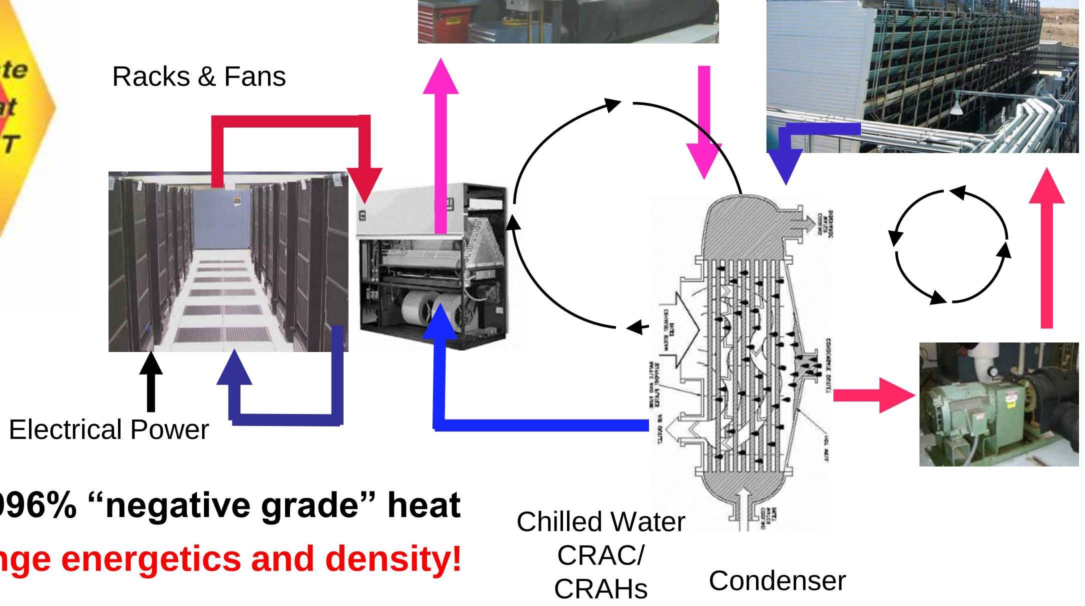
Low Datacenter Efficiency

- **Efficiency is 0.000’004%**
- **Volume used for compute is <1 ppm (part per million)**



Brouillard, APC, 2006





- **A computer is an inefficient “joule heater” “producing” 10-20ºC “heat”.**
  The output is 0.000’004% ICT and 99.999’996% “negative grade” heat
- **Consequence: We have a lot to do to change energetics and density!**

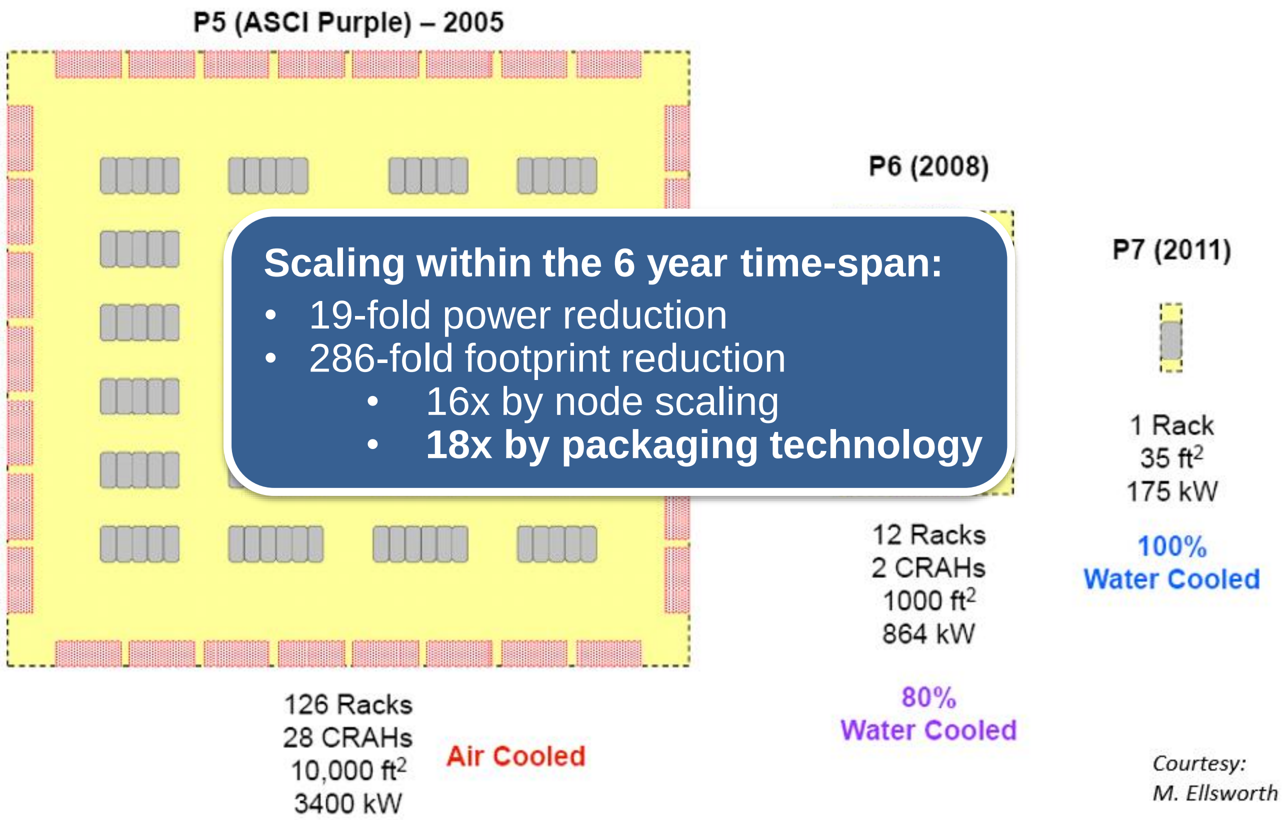
Bruno Michel, bmi@zurich.ibm.com

Agenda

- Thermal resistance key to improve efficiency
- Zero Emission Datacenters – from Idea to largest European Computer in 5.5 Years
- Volumetric density scaling to replace Moore's law and transform IT industry
- Information Industry reinvents itself every 12 years
- Wearables for healthcare and Human Centric Sensing and Computing

Bruno Michel, bmi@zurich.ibm.com

Data Center Evolution … @ 100 TFLOPS



Scaling within the 6 year time-span:

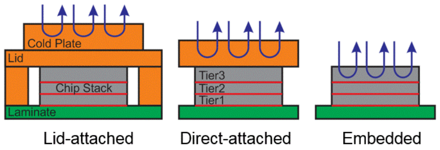
- 19-fold power reduction
- 286-fold footprint reduction
    - 16x by node scaling
    - **18x by packaging technology**

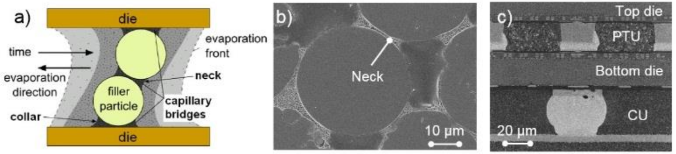
Courtesy:
M. Ellsworth

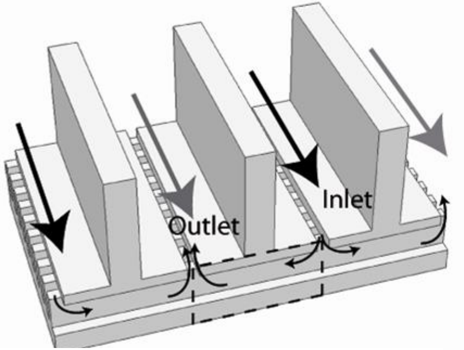
Bruno Michel, bmi@zurich.ibm.com

How are Thermal Resistances Massively Reduced?

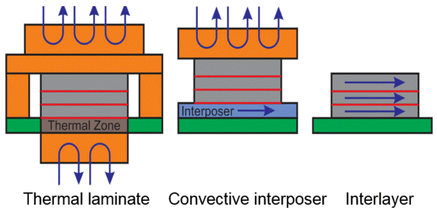
Back-side cooling evolution

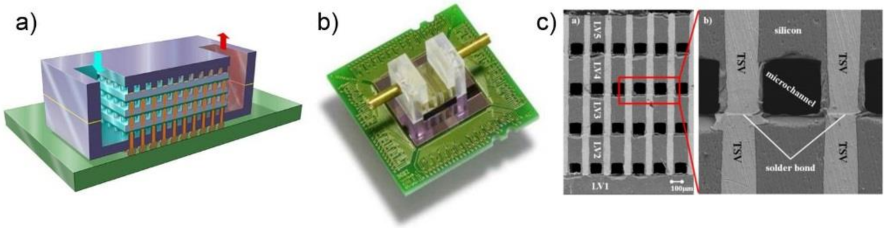


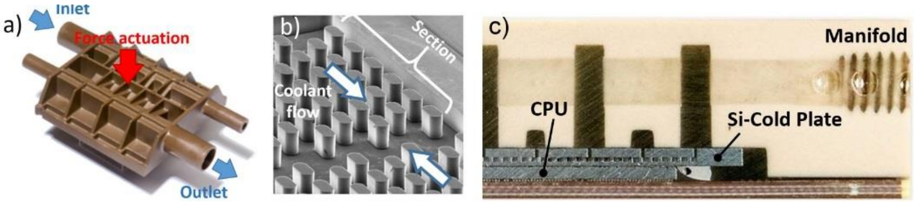


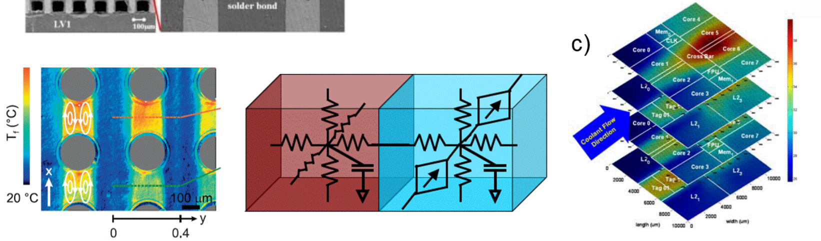


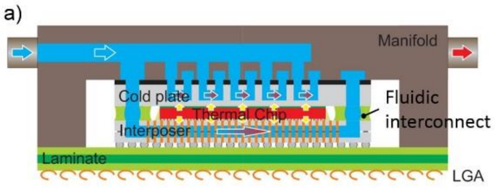
Disruptive cooling approaches

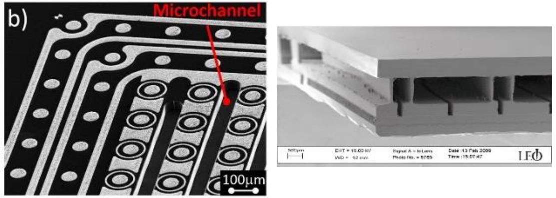


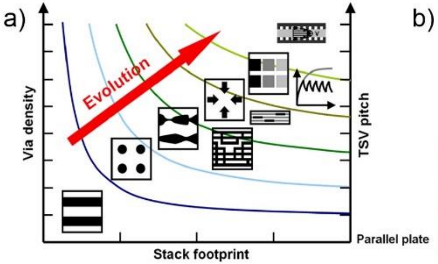








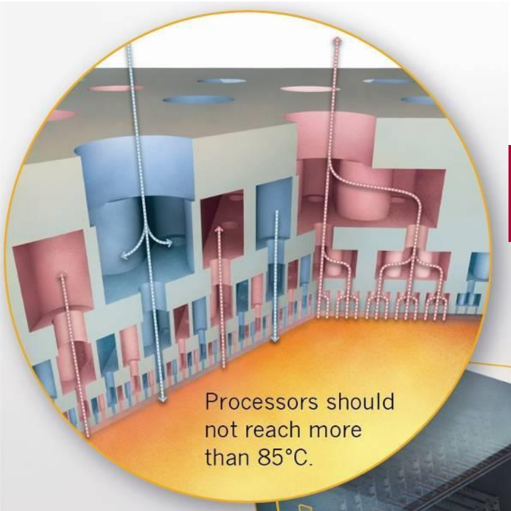


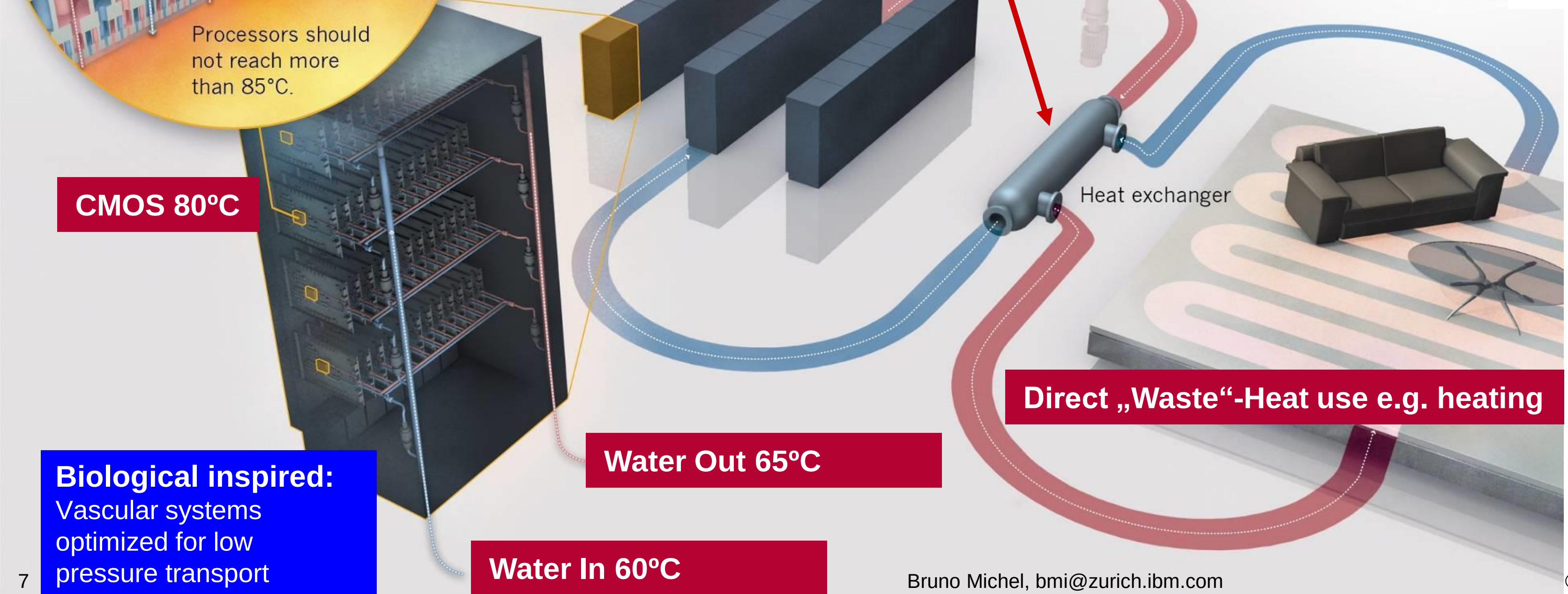


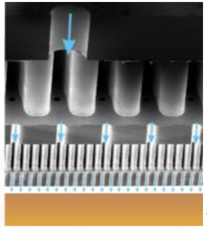
Bruno Michel, bmi@zurich.ibm.com

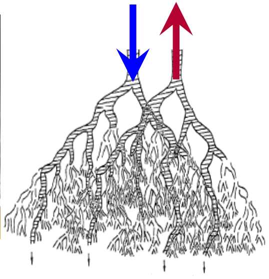
Biological Concepts Transform IT Industry

Hot-Water-Cooled Zero Emission Datacenters „Aquasar"









Micro-channel liquid coolers

Heat exchanger

Water Pump

CMOS 80ºC

Direct „Waste"-Heat use e.g. heating

Water Out 65ºC

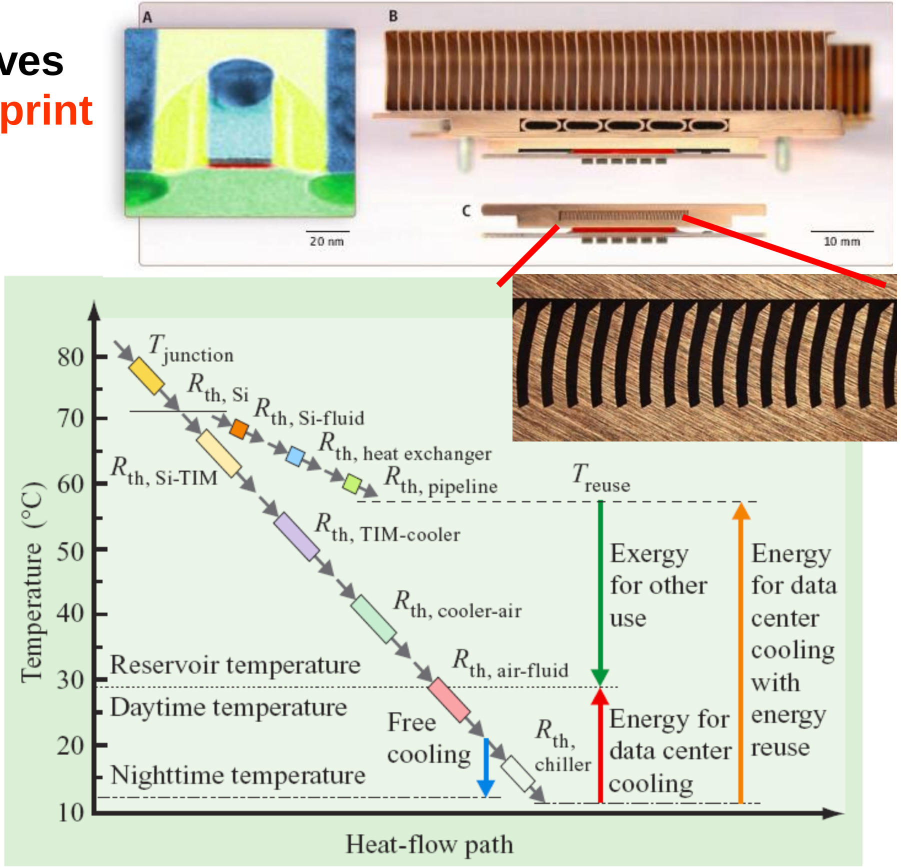
Water In 60ºC

Biological inspired:
Vascular systems optimized for low pressure transport

Bruno Michel, bmi@zurich.ibm.com

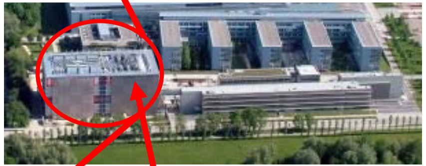
Zero-Emission Data Centers

- **High-performance chip-level cooling improves energy efficiency AND reduces carbon footprint**
  - Cool chip with **DT = 20ºC** instead of 75ºC
  - **Save chiller energy:** Cool with T > 60ºC **hot water**
  - **Re-use:** Heat 700 homes with 10 MW datacenter
- **Need carbon footprint reduction**
  - EU, IPCC, Stern report targets
  - Chillers use ~50% of datacenter energy
  - Space heating ~30% of carbon footprint
- **Zero-emission concept valuable in all climates**
  - Cold and moderate climates: **energy savings** and **energy re-use**
  - Hot climates: Free cooling, desalination
- **Europe: 5000 district heating systems**
  - Distribute 6% of total thermal demand
  - Thermal energy from datacenters absorbed



Bruno Michel, bmi@zurich.ibm.com

SuperMUC at Leibniz Rechenzentrum



CoolMUC-2 (2016)

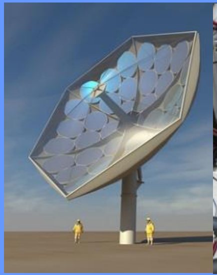
- **9288** IBM System x iDataPlex dx360 M4
  - **43997256** Components
  - **8.08 m^2** CMOS **4.22x10^{13}** transistors
  - **74304** Samsung 4 GB DIMMs
- **11868** IB Fibre Cables **192640 m**
- Cooling
  - **34153 m** Copper
  - **18978** Quick Connects
  - **7.9 m^3** Water
- Mass **194100 kg**

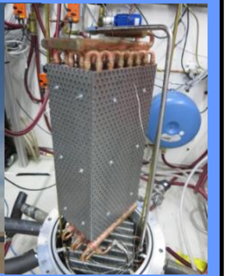
- IBM System x / Lenovo NeXtScale **DWC nx360 M5**

Bruno Michel, bmi@zurich.ibm.com

Mega Trends with Implications

Datacenter Carbon Footprint



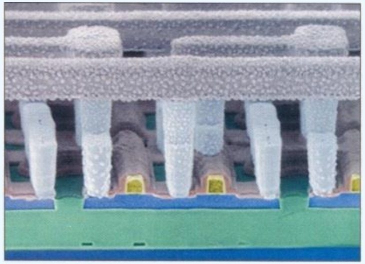


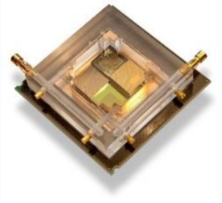
Renewables Energy Cost and Climate

Big Data

Cognitive Computing

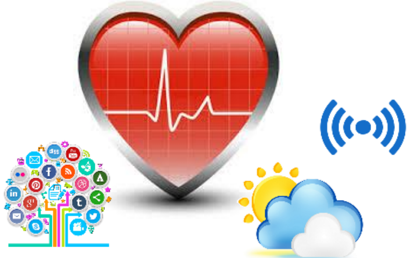
The End of Transistor Scaling





Dense and Efficient Systems Roadmap

The Internet of Things IoT

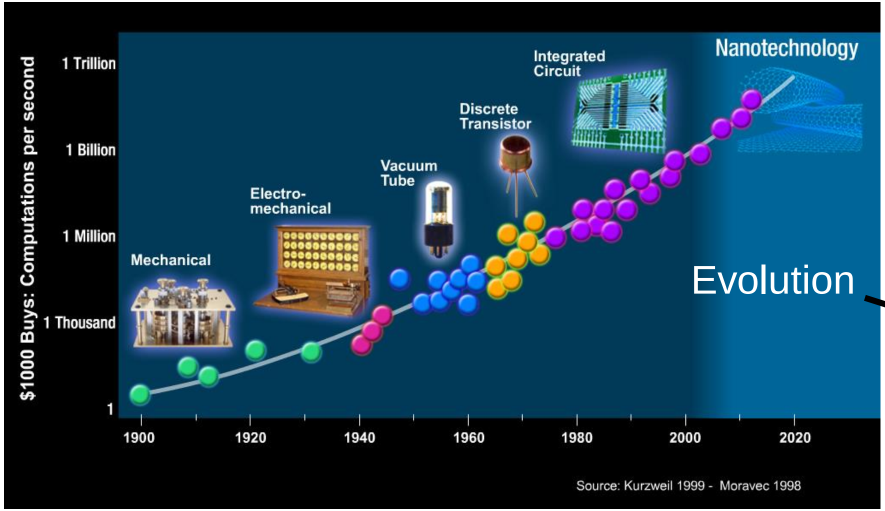


IBM Watson

Wearables and Healthcare

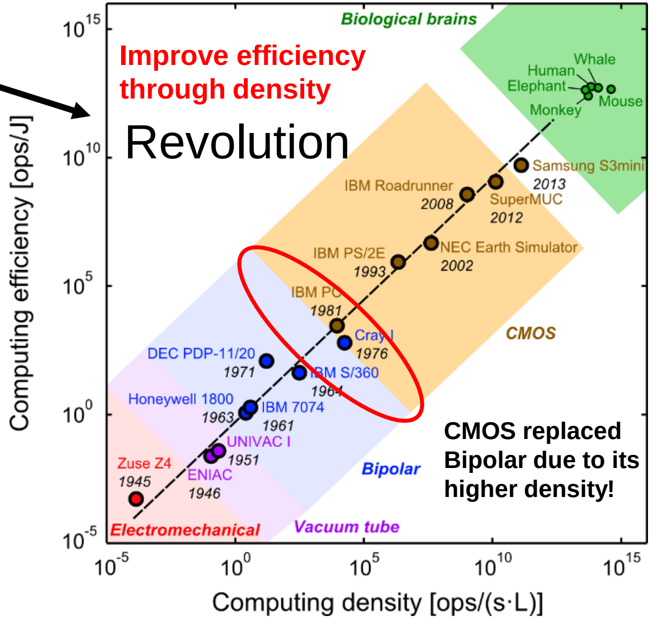
Bruno Michel, bmi@zurich.ibm.com

(R)evolution of Information Technology



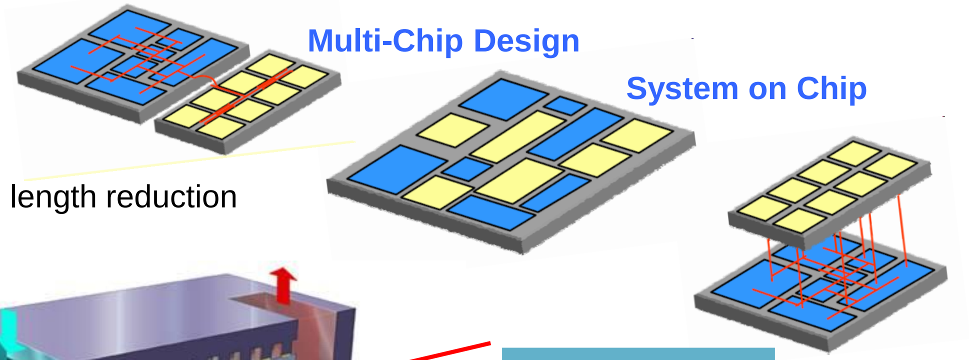
- Density and efficiency on log-log line
  - Brain is 10^4 times denser AND 10^4 times more efficient
- Independent of switch technology
  - No jumps mechanical – tube – bipolar – CMOS – neuron
- Communication as main bottleneck
  - Memory proximity lost in current computers (1300 clock access)
  - Detrimental for efficiency

- **Device centric viewpoint (left)**
  ➔ Device performance dominates
  - Power depends on better device performance
- **Density centric viewpoint (below)**
  ➔ Communication efficiency dominates
  - Power and memory proximity depend on size
  - Dominant for large systems (>Peta-scale)



Bruno Michel, bmi@zurich.ibm.com

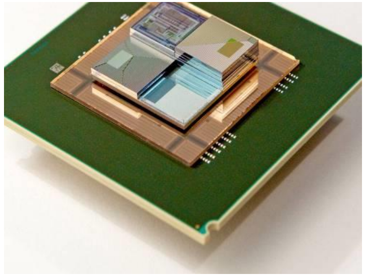
Paradigm Change: Vertical Integration

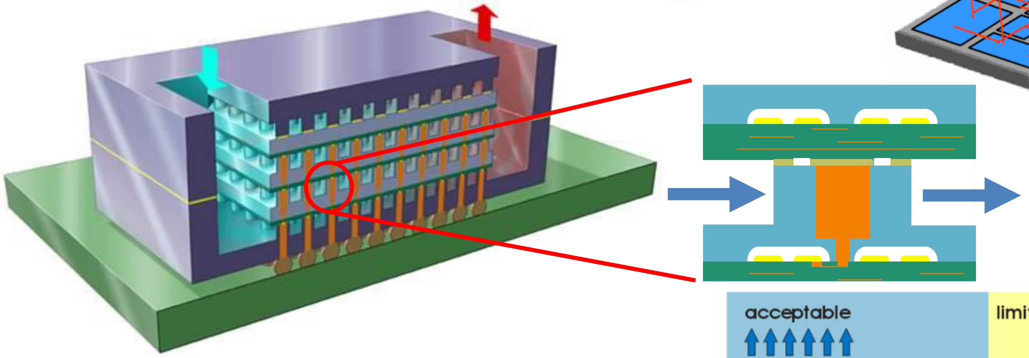


Global wire length reduction

Brain: synapse network

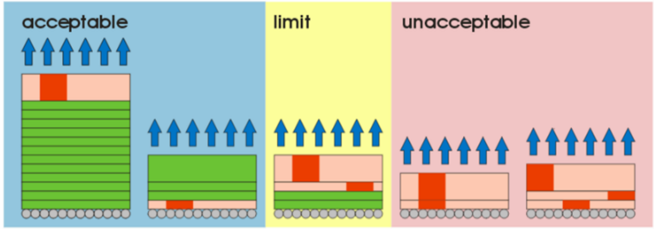
3D Integration

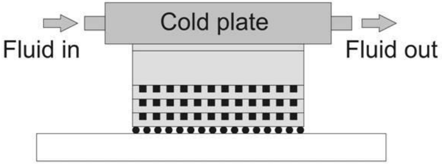




Through silicon via
electrical bonding
and water insulation
scheme

- A large fraction of energy in computers is spent for data transport
- Shrinking computers saves energy



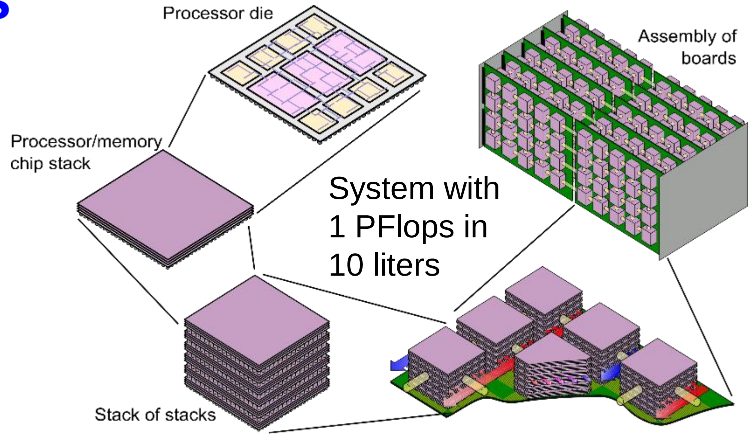


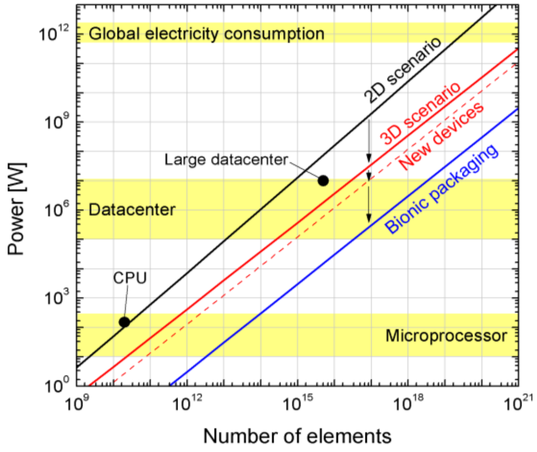
Heat removal limit constrains electrical design

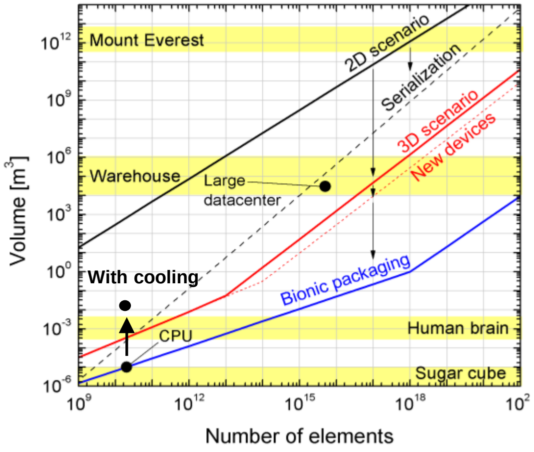
Bruno Michel, bmi@zurich.ibm.com

Scaling to 1 PFlops in 10 Liters

- **Efficiency comparison**
  - 1PFlops system currently consumes ~10MW
  - 0.1 PF ultra-dense system consumes 20 W
- **Ultra-dense Bionic System**
  - Stack ~10 layers of memory on logic
  - Stack several memory-logic stacks to stack of stacks
  - Combine several blocks of stacks to MCM (MBM)
  - Combine MCMs to high density 3D system
- **Key enabling technologies**
  - Interlayer cooling
  - Electrochemical chip power supply
- **Impact**
  - **5’000x smaller power**
  - 50’000’000x denser
  - Scalability to zetascale







P. Ruch, T. Brunschwiler, W. Escher, S. Paredes, and B. Michel, “Towards 5 dimensional scaling: How density improves efficiency in future computers”, IBM J. Res. Develop. 55 (5, Centennial Issue), 15:1-15:13 (2011).

Bruno Michel, bmi@zurich.ibm.com

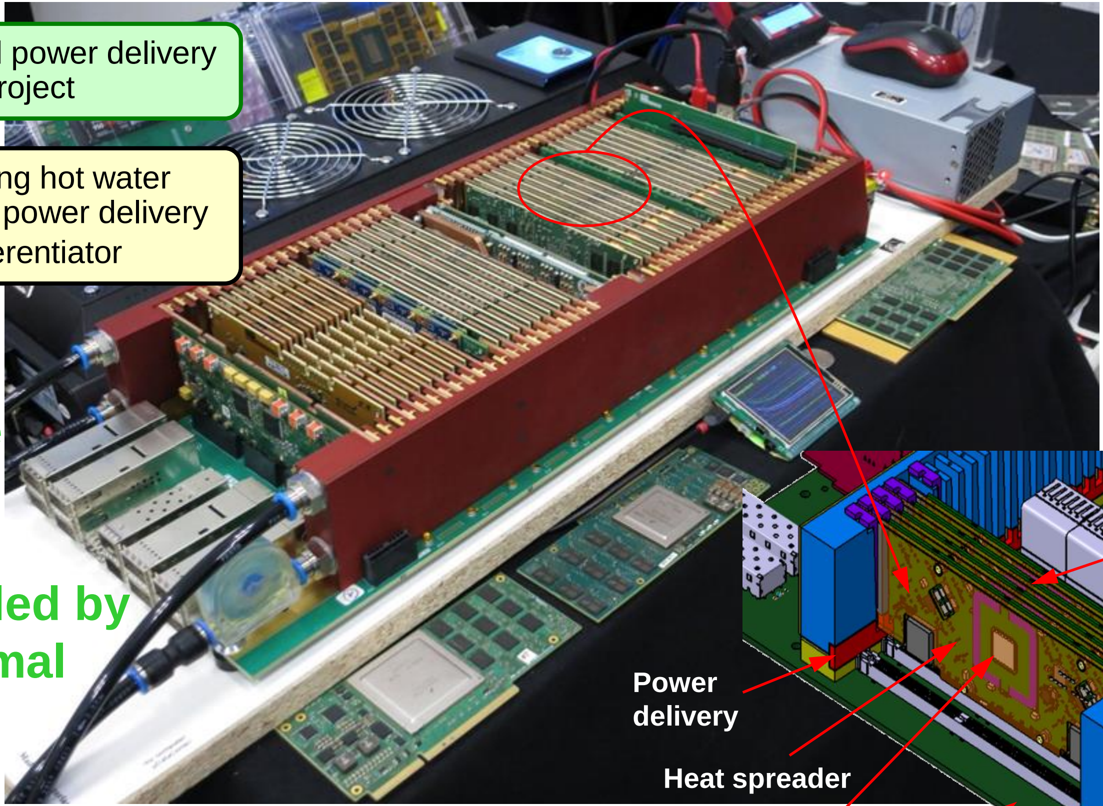
Datacenter in a Box: Top down Densification

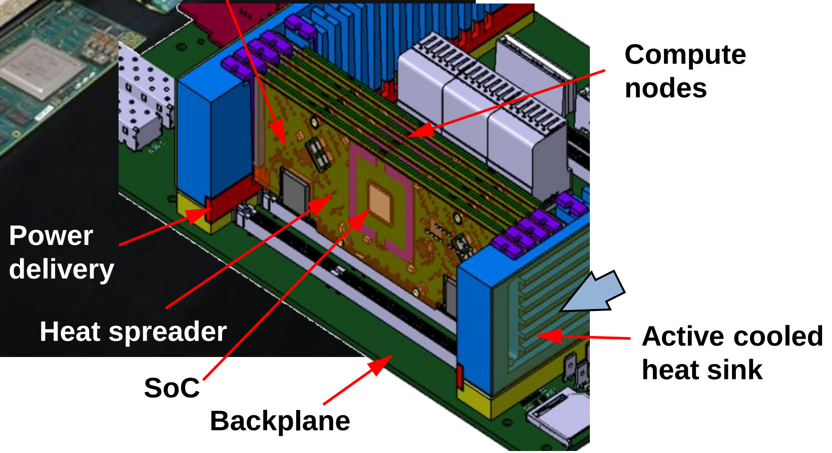
Leading cooling and power delivery concept for Dome project

Increase density using hot water cooling structure for power delivery
➔ Density: Key differentiator

1000x denser and 10x more efficient!!

Density enabled by reduced thermal and electrical resistance



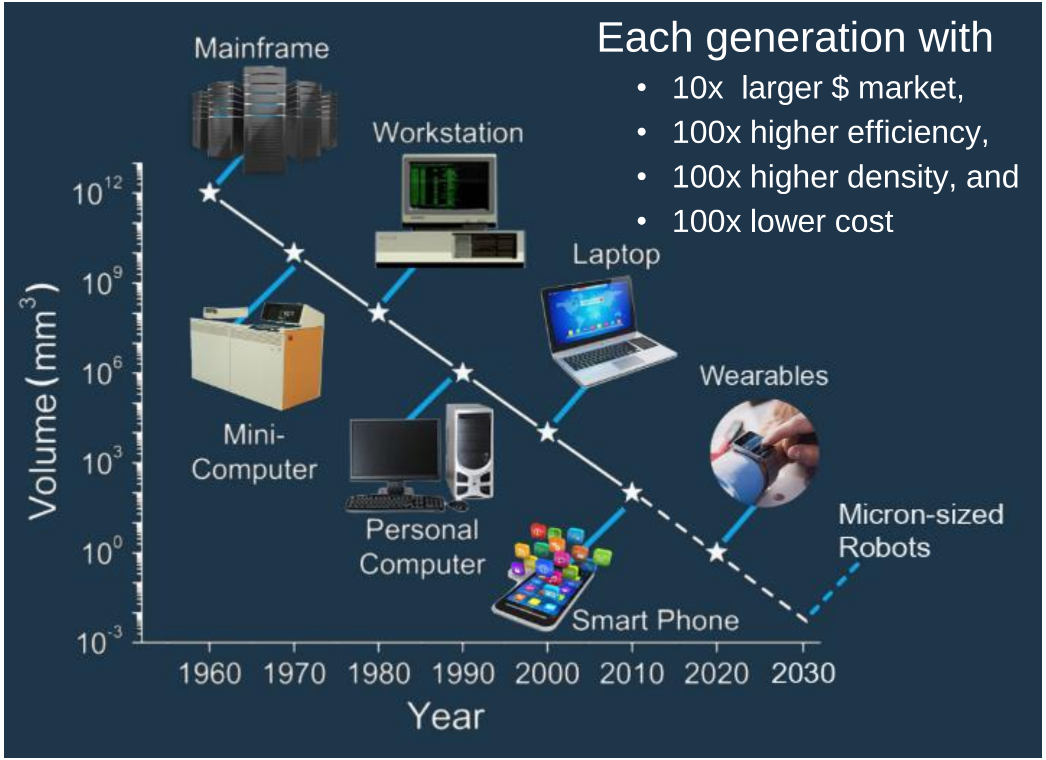


Target system
- T4240 1.8GHz node
  - 12 cores 2 threads
  - 48GB of DRAM
  - **220 GFLOP**
- **2U rack unit**
  - 128 nodes
  - 3072 Threads
  - 6 TB DRAM
  - **28 TFLOP**

Bruno Michel, bmi@zurich.ibm.com

Bell's Law Demands more Integration

- Every 12-15 years restart of a new generation
- Hardware cost fraction decreases from 100% (mainframe) to <10% while adding functionality
- Serve healthcare (wearables), industry (work safety)
- Strengthen AI with cloud, intelligent edge, and sensors
- Efficient data acquisition/labelling, reduce learning time, and quicker valuation of trained system
- Improved autonomy, functionality, latency, reliability, and privacy



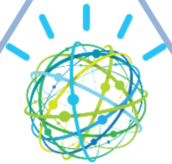
- **Data is not transferred to cloud but AI tools are transferred to the edge to improve**

Bruno Michel, bmi@zurich.ibm.com

IoB Platform Building blocks

Humans as largest data source and largest consumer of cognitive services

***modular, flexible and scala-ble** platform that adapts to use cases and strengthens base of AI pyramid*



Data Science
- SML Algorithms
- Deep Learning
- microservices

Data Collection/Preparation
- Gathering protocol
- Reliable Data
- Data Storage
- Data cleaning
- Segmentation, labeling
- Feature selection

Human in Center and Pyramid Base

- Physiology / Medicine
- Psychology / Behavioral change
- Acceptance / Usability / User experience

User Interface
- *Data*
- *Thresholds and trends*

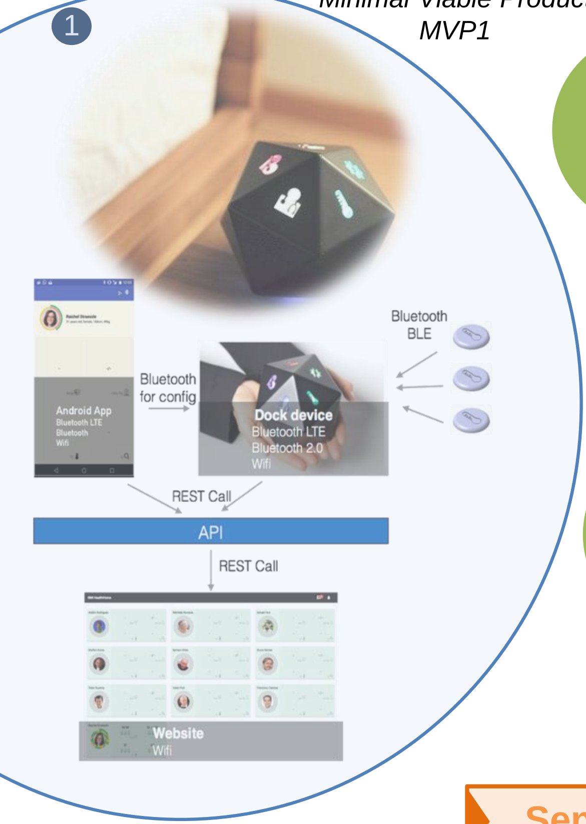
Cloud Services
- *Storage*
- *Big data science*

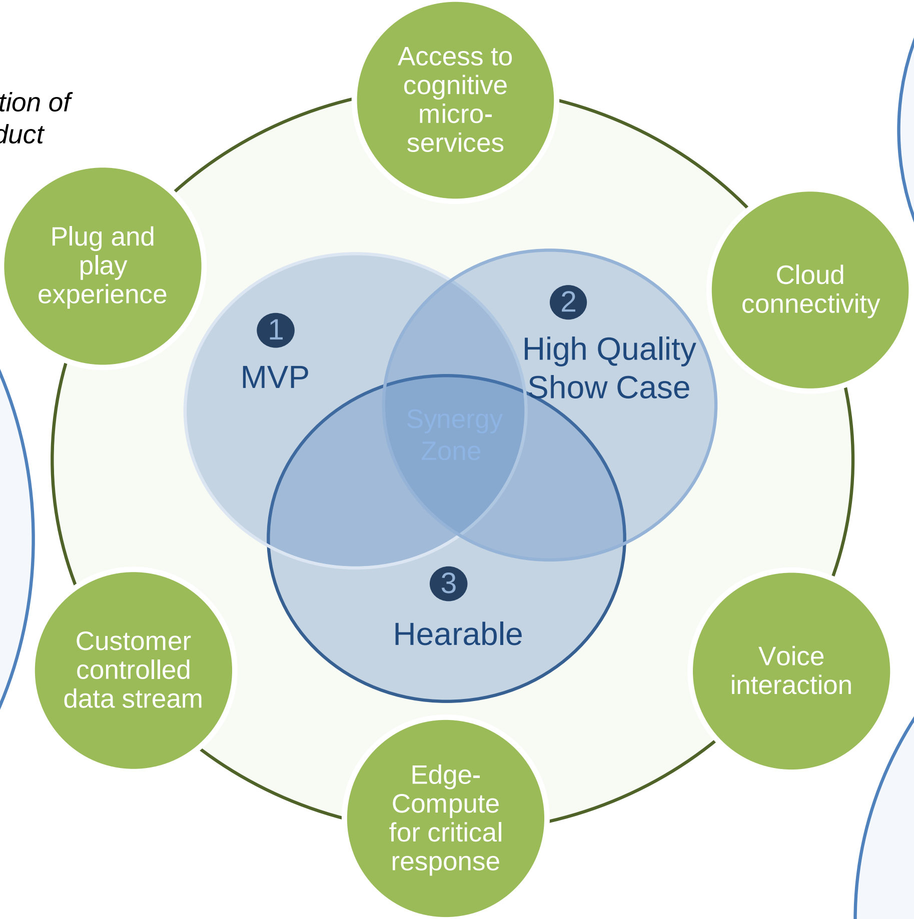
Data Hub and Sensors
- *Edge*
- *Communication*
- *Alerts*
- *Commercial and new sensors*

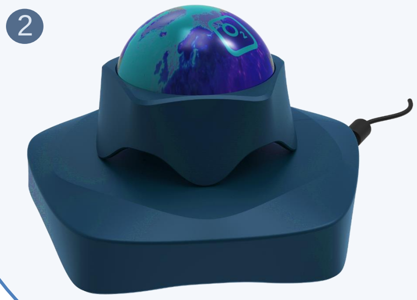
Use Cases
- *eCompanion – Hospital of the Future*
- *CAir – Chronic Disease Mgmt.*
- *FireFighter – Stress Detection*

eCompanion

Functional demonstration of Minimal Viable Product MVP1







In fabrication

Prototype assembly ongoing

Sensor data fusion and feedback to user

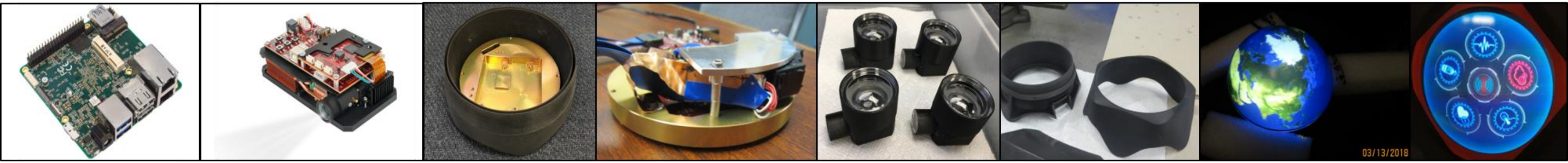
Prototype of High Quality Show Case

User experience is as part of the human-machine system

Hardware
- Spherical projection screen
- Microphone array
- Speaker

Soft Layer
- Connected by BLE to sensors and by WiFi to cloud
- Watson Assistant

Applications
- Patient’s home
- Hotel rooms
- Waiting rooms
- Weather stations

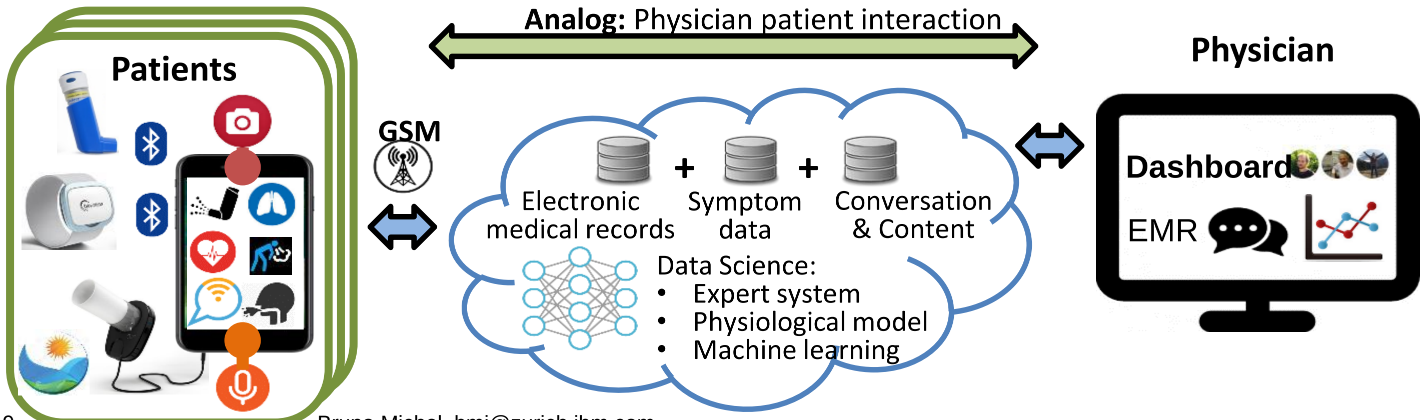


PoC: Firemen Stress Monitoring and Coaching of COPD patients

Vital signs monitoring

- ECG sensors record R-R intervals and HRV (heart rate variability)
- Physical stress by pulse and emotional stress by HRV
  - Use on firemen, police, athletes, workplace etc.
  - Use machine learning for classification
  - Integrated acquisition, labelling, and learning
  - Include breath, sweat etc.
  - Project name: **DeStress**

- Bi-directional electronic patient – physician communication, continuous patient symptom and activity tracking
- Prediction of exacerbation risk, personalized and context based communication by virtual agent “chat-bot”.



Recording:
- Medication adherence
- Lung function
- Cough intensity
- Sputum color
- Activity and sleep
- Vital-signs
- Environmental parameters

Bruno Michel, bmi@zurich.ibm.com

Results: Mental and Physical Stress Classification with HRV

- 4 classes: (0) No stress, (1) Mental stress, (2) Physical stress, (3) Physical **&** mental stress
  - 20 different subjects, male and female, from unfit till very fit
- Test starting with 3 minutes relaxing
  - IQ test (25 min)   Reflex test   Stroop Color Tests
  - Memory training   Arithmetic test
- Test for physical and mental stress
  - 3 relaxing - fitness test - fitness & IQ test

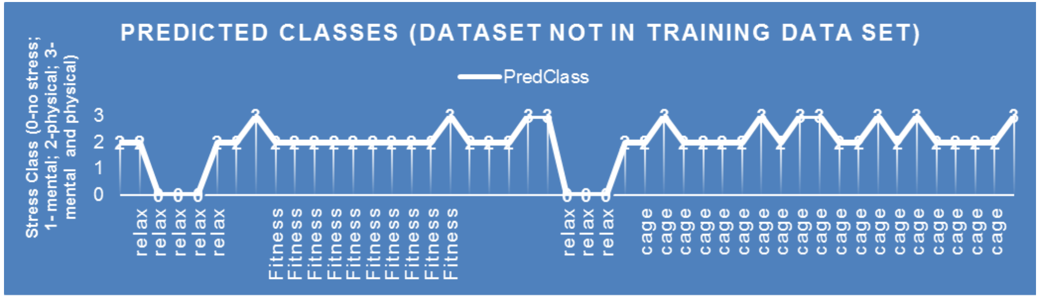
$$precision = \frac{TP}{TP + FP}$$

$$true\ positive\ rate = recall = \frac{TP}{TP + FN}$$

$$F\text{-}score = \frac{2 * precision * recall}{precision + recall}$$

	C5			SVM		
Red flag = overfitting	Total precision	Total recall	Total F-score	Total precision	Total recall	Total F-score
All HRV parameters	0,859	0,859	0,859	0,808	0,805	0,808
All HRV & rest parameters	0,915	0,909	0,914	0,785	0,784	0,785
Time domain parameters	0,813	0,810	0,813	0,612	0,595	0,608
Time domain & rest parameters	0,814	0,821	0,815	0,584	0,582	0,583

- Distinguish mental stress from parallel physical stress
- Best algorithm: C5 decision tree, with >80 % precision and recall
- Detected all phases for unknown subjects



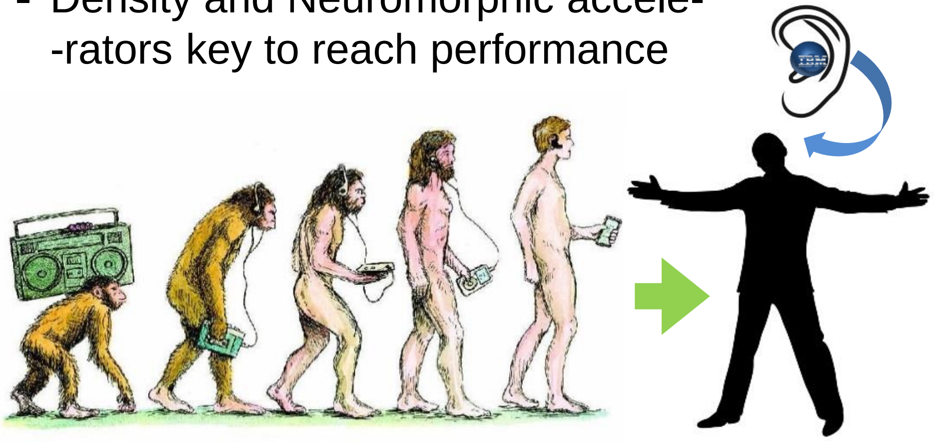
Bruno Michel, bmi@zurich.ibm.com

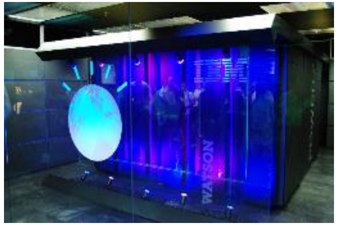
Hearables – All in One Devices

- Augmented interaction using emotional awareness.
  - Hands free and invisible
  - Multi function, smart and cognitive
- Human - Human vocal communication;
- Seamless communication and symbiotic human computer interaction
- Hospital of the Future, Cognitive Companion for the Elderly
- Fitness, Wellness, Chronic Disease Monitoring
- Density and Neuromorphic accele- -rators key to reach performance

Earplug in movie *"Her"*

Annapurna Pictures

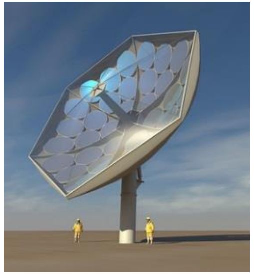




Bruno Michel, bmi@zurich.ibm.com

Summary

- **Thermal resistance key to improve efficiency and reduce carbon footprint**
  - Reduce losses to unusable heat by reduced thermal resistances
  - Application in datacenters and all thermally mediated energy conversion processes
  - Thermal and electrical packaging can take lead in many more areas



- **Zero Emission Datacenters –** from Idea to largest European Computer in 5.5 Years
  - 100 PY investment in reducing convective and conductive thermal resistances
  - Thermal and electrical packaging can take lead in many more areas
  - Top down miniaturization: Datacenter in a box as first step

- **Volumetric density scaling to replace Moore's law and transform IT industry**
  - Big Data and Cognitive Computing Drive demand for more efficient IT hardware
  - Brain inspired packaging with combined power supply and cooling ➔ Bionic packaging
- **Wearables for healthcare and Human Centric Sensing and Computing**
  - **Bell's law:** Mainframe computing – PCs – mobile computing – wearable computing
  - Cognitive companion and the augmented human

Thank you very much for your kind Attention!

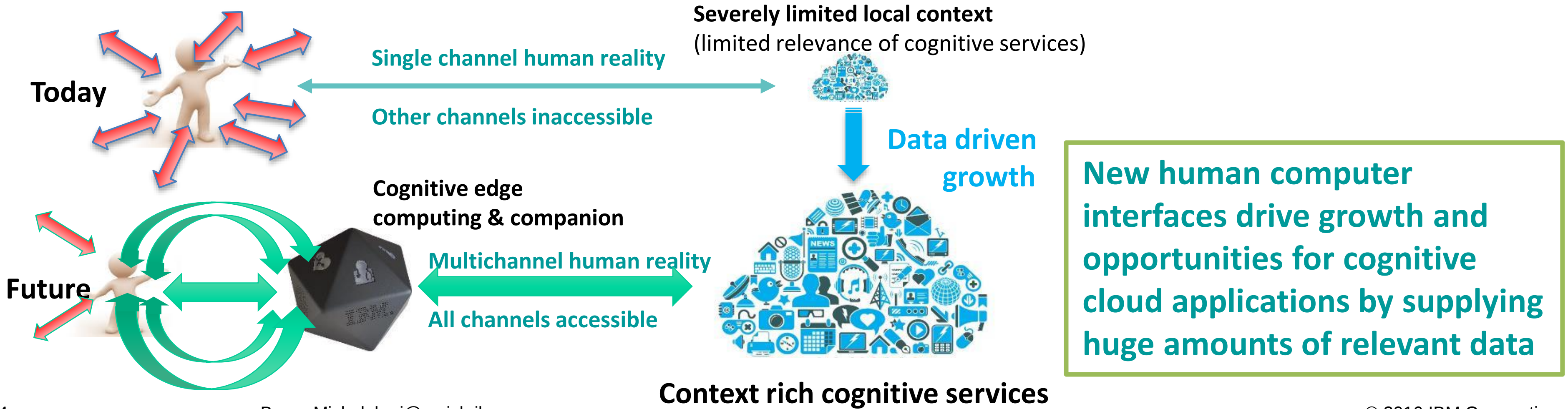
Acknowledgement

- Ingmar Meijer, Patrick Ruch, Thomas Brunschwiler, Stephan Paredes, Werner Escher, Yassir Madhour, Jeff Ong, Gerd Schlottig and many more for Aquasar and SuperMUC design
- Sebastian Gerke, Rahel Straessle, Arvind Sridhar, Ismael Faro, Yuksel Temiz, Neil Ebejer, Theodore G van Kessel, Keiji Matsumoto, Emanuel Loertscher, Frank Libsch, Hyung-Min Lee, John Knickerbocker, Jonas Weiss, Jonathan E Proesel, Kang-Wook Lee, Mehmet Soyuer, Minhua Lu, Mounir Meghelli, Norma E Sosa cortes, Paul S Andry, Roy Yu, Shriya Kumar, Sufi Zafar, Teodor K Todorov, Yves Martin …. and many more
- PSI Tobias Rupp and Thomas Schmidt
- ETH Severin Zimmermann, Adrian Renfer, Manish Tiwari, Ozgur Ozsun, Dimos Poulikakos
- EPFL Yassir Madhour, John Thome, and Yussuf Leblebici
- Funding: IBM FOAK Program, IBM Research, CCEM Aquasar project, Nanotera CMOSAIC project, SNF Sinergia project REPCOOL, NRP 70 project THRIVE, EUFP7 project CarrICool, EUFP7 project HyperConnect, DARPA IceCool, KTI / DSolar. IBM Research Frontiers Institute: www.research.ibm.com/frontiers

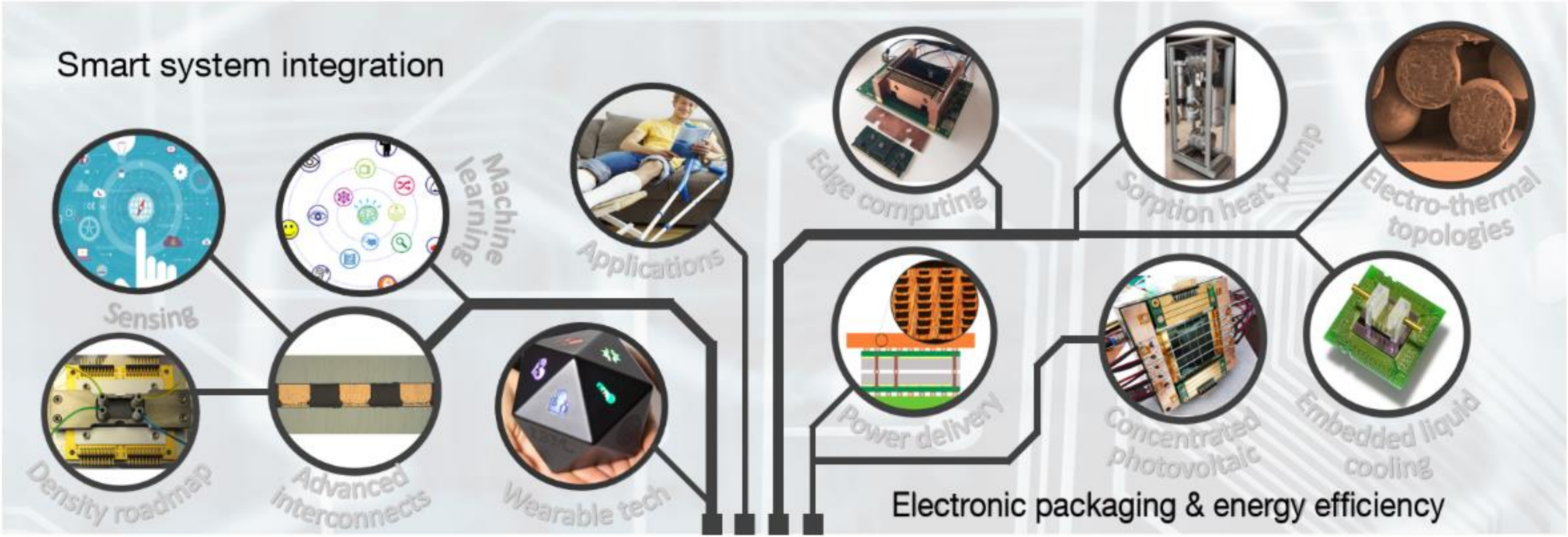
Bruno Michel, bmi@zurich.ibm.com

Motivation and Vision

- Cost and accessibility of healthcare, blockbuster drugs not personalized, cost, reliability, accessibility
- Stress umbrella link to human wellbeing needs to be relevant to the people
- **Human Centric Sensing and Computing:** Context key for relevant personalized cognitive services
- Personalized cognitive services in preventive medicine / coaching; work safety; wellbeing; elderly care
- Miniaturization for low-cost non-intrusive monitoring to reduce cost in acute and preventive medicine
- **Move intelligence to the edge instead of data to the cloud**.



Bruno Michel, bmi@zurich.ibm.com



- **Links:**
- **Smart System Integration** https://www.zurich.ibm.com/st/smartsystem/
- **Functional Electronic Packaging** https://www.zurich.ibm.com/st/electronicpackaging/
- **Efficiency and Green Technologies** https://www.zurich.ibm.com/st/energy_efficiency/

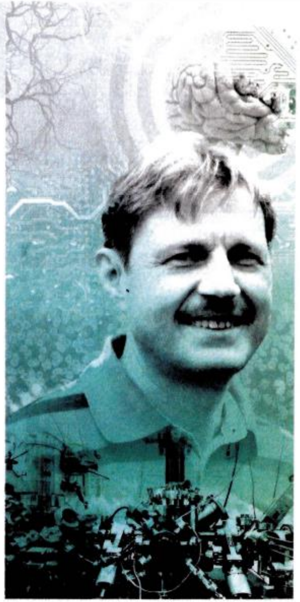
Bruno Michel, bmi@zurich.ibm.com

Economist, March 12, 2016: The Future of Computing

- http://www.economist.com/news/leaders/21694528-era-predictable-improvement-computer-hardware-ending-what-comes-next-future

- **Project Syndicate**
- Enabling cognitive computing with new density scaling roadmap
- https://www.project-syndicate.org/commentary/computing-will-mimic-human-brain-by-bruno-michel-2016-06

Brain scan | Bruno Michel



IBM's head of advanced micro-integration reckons biology holds the key to more energy-efficient chips

IN 2011 a supercomputer called Watson, built by IBM, beat two top-rank human champions at "Jeopardy!", an American quiz show. This caused great excitement. Unlike chess, a game of abstract reason and logic, "Jeopardy!" is full of puns, *double entendres* and wordplay; just the sort of thing meant to stump computers. But Bruno Michel, who heads the advanced micro-integration group at IBM's research lab in Zurich, says it was not a fair fight. "Do you know how much power [Watson] consumes? At the time it was something like 80kW. That's thousands of times more than the humans it was playing against."

Dr Michel argues that computers are extremely inefficient machines, both in terms of the electricity they consume and the space they take up. A typical desktop machine or a server in a data centre, he reckons, uses only about 0.0001% of its volume to crunch numbers, and perhaps 1% to shuttle the results around. The rest is mostly empty space. The laws of physics set a limit to how efficiently information can be processed, but modern computers perform at only about 0.00004% of that theoretical maximum, he calculates. So for now a big data centre consumes tens of megawatts of power, almost all of which is transformed, literally, into hot air.

Moore's law used to keep a lid on electricity usage even as computing capacity raced ahead because smaller transistors needed less power. That is no longer true. "Nowadays the cost of buying a computer or a data centre is less than the cost of running it for a few years," says Dr Michel. "That's a paradigm shift." Data centres already consume 2% of all the electricity produced in the world.

Dr Michel's benchmark for efficiency is evolution; his original training was in mechanical engineering, but he fell in love with biology after reading a genetics textbook. After earning a PhD in biochemistry and biophysics from the University of Zurich, he joined IBM's Zurich lab to work with the scanning tunnelling microscope developed there in 1981. It won its inventors a Nobel prize, allowing scientists to see and manipulate individual atoms. That job led to a project on manufacturing technology for flat-screen displays. "I get fascinated by new things, and then I want to start working on them," he says. "But my advice is: if you want to work in a new field, don't be driven by creativity—be driven by impact."

That is how he got involved in his current work on more energy-efficient chips. "In the middle of the past decade there was a panic in the chip industry—soon we won't be able to keep these things cool," he notes. At the same time energy policy was becoming more important as climate change moved up the political agenda.

Biology's secret weapon, he thinks, is the spidery, fractally branching network of blood vessels that supply energy to the brain, allowing most of its volume to be turned over to useful data-processing tasks. As near as neuroscientists can tell, a mammalian brain uses about 70% of its volume for moving information around, 20% for processing it and the remaining 10% to keep everything in the right place and supplied with nutrients. In doing all these things, a human brain consumes about 20 watts of power. That makes it roughly 10,000 times more efficient than the best silicon machines invented by those brains, Dr Michel reckons.

One of his favourite charts compares the density and efficiency of brains with a string of computing technologies going back to the second world war. All of them fall on a straight line, suggesting that to match the energy efficiency of the brain, scientists will have to emulate its density.

He is now working on a project to build an electronic version of the blood that channels energy to biological brains (see main article). "It was something like 200 years after the invention of the steam engine before mechanical engineering began to catch up with biology in terms of efficiency," he says. "It would be good if computing could accomplish the same thing in half the time."

The Economist March 12th 2016 9

References (Packaging, Cooling)

- J. Zuercher, T. Brunschwiler, et al., “Nanoparticle assembly and sintering towards all-copper flip chip interconnects,” Proceedings of ECTC, San Diego, 2015.
- J. Zuercher, T. Brunschwiler, et al., “All-copper flip chip interconnects by pressureless and low temperature nanoparticle sintering,” Proceedings of ECTC, Las Vegas, 2016.
- T. Brunschwiler, J. Zuercher, S. Zimmmermann, B. Burg, G. Schlottig, R. Straessle, et al., “Review on percolating and neck-based underfills for three-dimensional chip stacks,” *Journal of Microelectronics and Electronic Packaging*, 138(4), 2016.
- J. Zürcher, X. Chen, B. R. Burg, T. Brunschwiler, S. Zimmermann, et al., “Enhanced percolating thermal underfills achieved by means of nanoparticle bridging necks,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, PP(99), 1-11, 2016.
- T. Brunschwiler, A. Sridhar, C. L. Ong, and G. Schlottig, “Benchmarking study on the thermal management landscape for 3D ICs: From back-side, to volumetric heat removal,” *Journal of Electronic Packaging*, 138(1), 2016.
- G. Schlottig, W. Escher, V. Khanna, T. Brunschwiler, et al., “Lid-integral cold-plate topology: Integration, performance, and reliability,” *J. of Electr. Packag*, 138 (1), 2016.
- T. Brunschwiler, O. Ozsun, G. Schlottig, et al., “Dual-side heat removal by micro-channel cold plate and silicon-interposer with embedded fluid channels,” Proc. 6th Electronic System-Integration Technology Conference (ESTC), Grenoble, France, pp. 1-6, 2016.
- T. Brunschwiler, B. Michel, S. Paredes, U. Drechsler, W. Cesar, G. Toeral, Y. Temiz, Y. Leblebici, “Validation of porous-media prediction of interlayer cooled 3D-chip stacks,” Proc. IEEE International Conference on 3D System Integration, 3DIC , San Francisco CA, p. 1-10, 2009.
- C. L. Ong, S. Paredes, A. Sridhar, B. Michel, T. Brunschwiler, “Radial hierarchical microfluidic evaporative cooling for 3-D integrated microprocessors,” Proc. 4th European Conference on Microfluidics, Limerick, 2014.
- Y. Madhour, B. Michel, et al., “Modeling of two-phase evaporative heat transfer in three-dimensional multicavity high performance microprocessor chip stacks,” *J. Electr. Packag.*, 136(2), 2014.
- A. Sridhar, M. Sabry, T. Brunschwiler, and D. Atienza, “3D-ICE: Fast compact transient thermal modeling for 3D-ICs with inter-tier liquid cooling,” International Conference on Computer-Aided Design, San Jose CA, 2010.
- A. Sridhar, Y. Madhour, D. Atienza, T. Brunschwiler and J. Thome, “STEAM: A fast compact thermal model for two-phase cooling of integrated circuits,” 32nd IEEE/ACM International Conference on Computer-Aided Design (ICCAD), San Jose, CA, 2013.
- T. Andersen, P. A. M. Bezerra, F. Krismer, J. W. Kolar, A. Sridhar, T. Brunschwiler, T. Toifl, C. Rabot, Z. Pavlovic, C. O’Mathuna, S. Gaborieau, C. Bunel, “CarrICool: Interposer supporting optical signaling, liquid cooling, and power conversion for 3D chip stacks,” Proc. 4th Int’l Power Supply on Chip Workshop (PwrSoC 2014), Boston, MA, USA, 2014.
- A. Sridhar, T. Brunschwiler, T. Toifl, R. Christen and J. Smajic, “Exploration of inductor-based hybrid integrated voltage regulator architectures,” Proc. 6th Electronic System-Integration Technology Conference (ESTC), Grenoble, France, pp. 1-6, 2016.
- P. A. M. Bezerra, T. Andersen, F. Krismer, J. W. Kolar, “Multi-objective Optimization of Fully Integrated Voltage Regulators: Switched Capacitor and Inductor-Based Converters,” ECPE PowerSoC Workshop (µPE: Powering Low-Power Systems), Ismaning/Munich, Germany, 2015.
- P. A. M. Bezerra et al., “Modeling and multi-objective optimization of 2.5D inductor-based Fully Integrated Voltage Regulators for microprocessor applications,” 2015 IEEE 13th Brazilian Power Electronics Conference and 1st Southern Power Electronics Conference (COBEP/SPEC), Fortaleza, pp. 1-6, 2015.
- R. K. Aljameh, P. A. M. Bezerra, F. Krismer, J. W. Kolar, A. Sridhar, T. Brunschwiler, T. Toifl, B. Michel, “Control and implementation aspects of a multiphase inductor-based FIVR in 14 nm bulk CMOS for microprocessor applications,” Proc. 5th International Workshop on Power Supply On Chip (PwrSoC2016), Madrid, Spain, October 3-5, 2016.

References (Wearables, Hot Water Cooling, and Density Scaling)

- G. Bell, “Bell’s Law for the Birth and Death of Computer Classes,” *Communications of the ACM* 51(1), 86–94 (2008).
- B. Michel, “The Future of Computing,” Project Syndicate, 30 June 2016.
- “European Commission seeks input on a Reflection and Orientation paper on Smart Wearables,” European Commission Reflection and Orientation Paper, 2 December 2016.
- P. Ruch, T. Brunschwiler, W. Escher, S. Paredes, and B. Michel, “Toward five-dimensional scaling: How density improves efficiency in future computers,” *IBM J. Res. Develop.* 55(5), 15:1-13 (2011).
- P. Ruch, T. Brunschwiler, S. Paredes, I. Meijer, and B. Michel, “Roadmap towards Ultimately-Efficient Zeta-Scale Datacenters,” Design, Automation, and Test Conference (DATE), March 18-22, 2013, Grenoble, France.
- M. Cossale, S. Paredes, R.P. Luijten, and B. Michel, “Combined power delivery and cooling for high density, high efficiency microservers,” 21st Intl. Workshop on Thermal Investigations of ICs and Systems “THERMINIC” (2015).
- S. Zimmermann, I. Meijer, M.K. Tiwari, S. Paredes, B. Michel, D. Poulikakos, “Aquasar: A hot water cooled data center with direct energy reuse,” *Energy* 43(1), 237–245 (2012).
- S. Zimmermann, M.K. Tiwari, I. Meijer, S. Paredes, B. Michel, D. Poulikakos, “Hot water cooled electronics: Exergy analysis and waste heat reuse feasibility,” *Int’l J. Heat and Mass Transf.* 55(23), 6391–6399 (2012).
- A. Renfer, M.K. Tiwari, R. Tiwari, F. Alfieri, T. Brunschwiler, B. Michel, “Microvortex-enhanced heat transfer in 3D-integrated liquid cooling of electronic chip stacks,” *Int’l J. Heat and Mass Transf.* 65, 33–43 (2012).
- P. Ruch, T. Brunschwiler, W. Escher, S. Paredes, B. Michel, “Toward five-dimensional scaling: How density improves efficiency in future computers,” *IBM J. Res. Develop.* 55(5), 15:1–15:13 (2011).
- A. Kubilay, S. Zimmermann, I. Zinovik, B. Michel, D. Poulikakos, “Compact thermal model for the transient temperature prediction of a water-cooled microchip module in low carbon emission computing,” *Numerical Heat Transfer, Part A: Applications* 59(11), 815–835 (2011).
- G. I. Meijer, “Cooling Energy-Hungry Data Centers,” *Science* 328, 316 (2010).
- T. Brunschwiler, G.I. Meijer, S. Paredes, W. Escher, and B. Michel, “Direct Waste Heat Utilization from Liquid-Cooled Supercomputers,” Proc. 14th Int’l Heat Transfer Conf., IHTC14–23352 (2010).
- W. Escher, B. Michel, and D. Poulikakos, “A Novel High Performance, Ultra Thin Heat Sink for Electronics,” *Int’l J. Heat and Fluid Flow* 31(4), 586–598 (2010).
- W. Escher, T. Brunschwiler, B. Michel, and D. Poulikakos, “Experimental Investigation of an Ultra-thin Manifold Micro-channel Heat Sink for Liquid-Cooled Chips,” *Int’l J. Heat Transf.* 132(8), 081402–11 (2010).
- P. Kasten, S. Zimmermann, M. Tiwari, B. Michel, and D. Poulikakos, “Hot Water Cooled Heat Sinks for Efficient Data Center Cooling: Towards Electronic Cooling With High Exergetic Utility,” *Frontiers in Heat and Mass Transfer* 1, 023006 (2010).
- T. Brunschwiler, B. Smith, E. Ruetsche and B. Michel, “Toward Zero-emission Data Centers through Direct Reuse of Thermal Energy,” *IBM J. Res. Develop.* 53(3), Paper 11 (2009).
- G.I. Meijer, T. Brunschwiler, S. Paredes, and B. Michel, “Using Waste Heat from Data Centres to Minimize Carbon Dioxide Emission,” *ERCIM News* 79, 23–24 (2009).
- W. Escher, B. Michel, and D. Poulikakos, “Efficiency of optimized bifurcating tree-like and parallel microchannel networks in the cooling of electronics,” *Int’l J. Heat and Mass Transf.* 52, 1421–1430 (2009).

 Bruno Michel, bmi@zurich.ibm.com

The Big Data and Cognitive Challenge

- Big Data in 2030 to 2040 means: ~1000x as much data as now
- End of Transistor Scaling means: We will not get more efficient chips
- Cognitive Compute Era means: We work ~100x more intensive with the data
- Currently ICT industry consumes ~3% ww energy and ~10% ww electricity

- 1000x more data times 100x more intense compute (at constant efficiency)
- Results in ~100'000x more compute and thus ~100'000 times more energy!

- Current computers are operated at ~1% of maximal efficiency
  since they run at <<10% load and power is not proportional to load
- Cloud compute delivery and workload optimization allows an efficiency improvement of ~100x

- But we still need 1'000x more energy; since ICT consumes 10% ww electricity
  we need ~100x more electrical power stations to fully enable world wide use
  of big data and cognitive computing! (in a worst case scenario)

- Big data neither works with current computers nor with current roadmaps!
- We clearly need major breakthrough innovations!

Bruno Michel, bmi@zurich.ibm.com

Technologies Involved

Efficient data aggregation

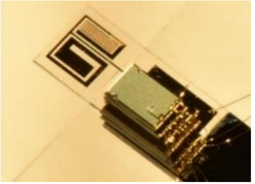
- Power harvesting and power management is key in particular for ECG, EEG, GPS, Computation & Communication technology
- Data quality and fidelity

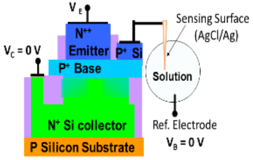
New efficient sensing technologies

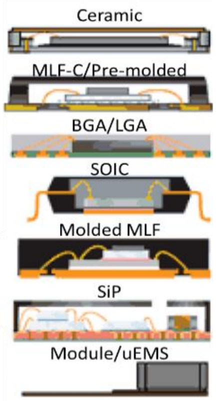
- Multi-sensing, compactness, S/N, selectivity and specificity
- CMOS compatible bipolar junction transistor for integrated sensing
- In-sensor neural network fingerprinting

Smallest and lowest cost computer technologies

- Extended packaging technologies from mobile and MEMS: SiP and 3D system integration
- Flexible, lightweight, thin, and large-area, bio-compatible integration
- Building on existing strategic activity







Raise energy 'umbrella'

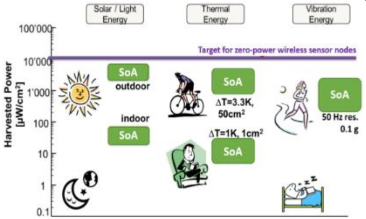
- Miniature power conversion with integrated micro-batteries
- Energy harvesting from heat, light, vibration, chemicals
- Energy transmission to energy poor sensors

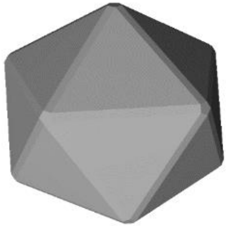
Neuromorphic accelerators

- Neuromorphic accelerators, analog frontends, and analog computing
- approximate and probabilistic computing

Efficient security and edge computing

- Edge computing with hypervisor shifts pareto optimum higher
- Energy efficient security tools against intrusions



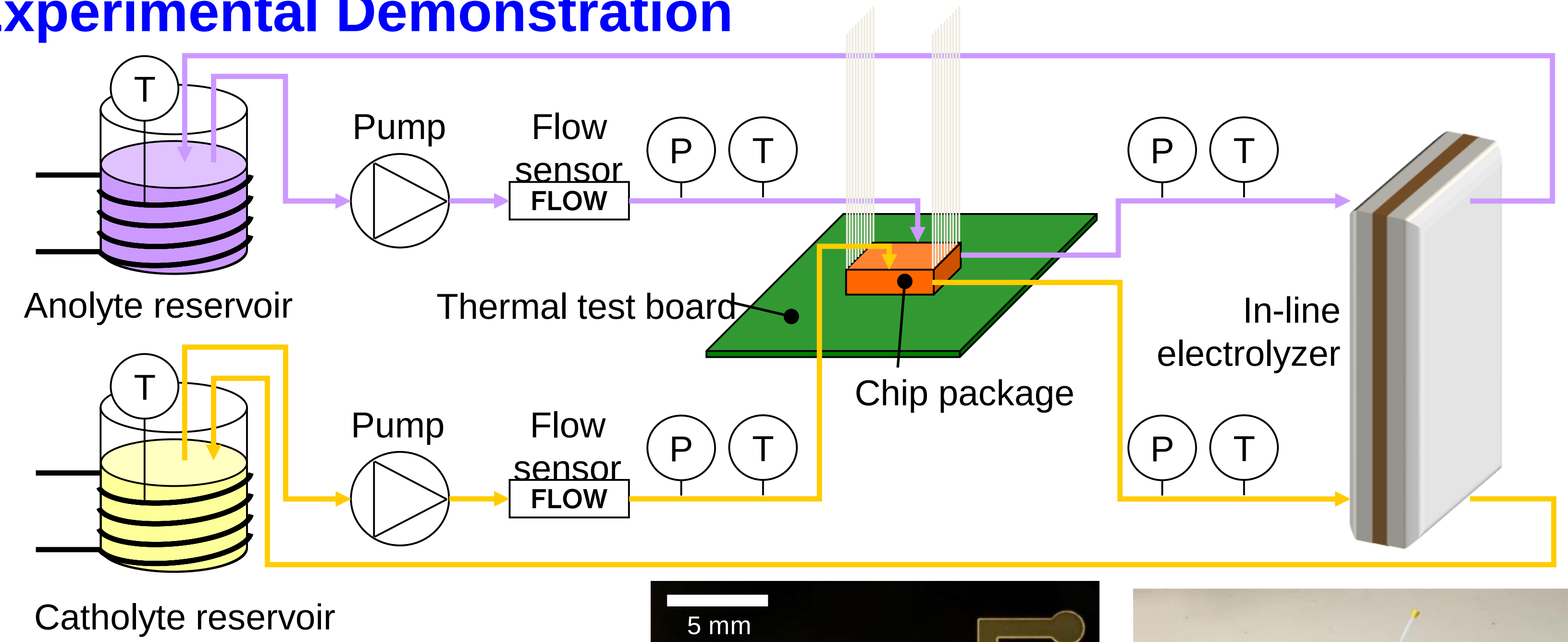


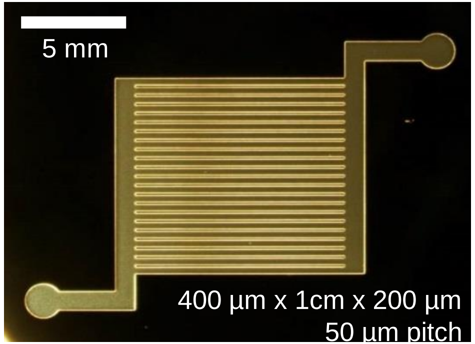
Bruno Michel, bmi@zurich.ibm.com

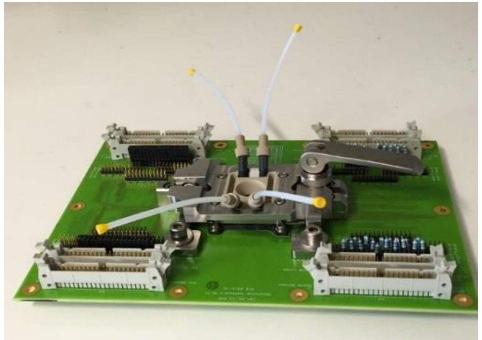
References (Solar, Heat Pump, Electronic Blood)

- S. Paredes, P. Ruch, E. Lörtscher, F. Malnati, A. Pedretti, and B. Michel, "First Outdoor Results of the 'Sunflower' HCPVT System," CPV 12, Frankfurt, Germany, April 25-27 (2016).
- S. Zimmermann, H. Helmers, M.K. Tiwari, S. Paredes, B. Michel, M. Wiesenfarth, A.W. Bett, D. Poulikakos, "A high-efficiency hybrid high-concentration photovoltaic system," *Int'l J. Heat and Mass Transf.* 89, 514-521 (2015).
- S. Paredes, B.R. Burg, P. Ruch, E. Loertscher, F. Malnati, M. Schmitz, T. Cooper, M. Cucinelli, D. Bonfrate, A. Mocker, A. Bernard, A. Steinfeld, G. Ambrosetti, A. Pedretti, and B. Michel, "Receiver-Module-Integrated Thermal Management of High-Concentration Photovoltaic Thermal Systems," IEEE PV Specialist Conference, New Orleans (2015).
- B. Burg, P. Ruch, S. Paredes, and B. Michel, "Placement and Efficiency Effects on Radiative Forcing of Solar Installations," AIP Conf. Proc. 1679, 090001-9 (2015).
- B.R. Burg, A. Selviaridis, S. Paredes, and B. Michel, "Ecological and economical advantages of efficient solar systems," AIP Conference Proceedings 1616, 317-320 (2014).
- V. Garcia-Heller, S. Paredes, C.L. Ong, P. Ruch, and B. Michel, "Exergoeconomic Analysis of High Concentration Photovoltaic Thermal Co-generation System for Space Cooling," *Renewable & Sustainable Energy Reviews* 34, 8-19 (2014).
- S. Zimmermann, H. Helmers, M.K. Tiwari, S. Paredes, P. Neves, B. Michel, and D. Poulikakos, "HCPVT Receiver Package Using Advanced Liquid Cooling for High Energy Efficiencies," AIP Conf. Proc. 1556, 248-254 (2013).
- C.L. Ong, W. Escher, S. Paredes, A.S.G. Khalil, and B. Michel, "A novel concept of energy reuse from high concentration photovoltaic thermal (HCPVT) system for desalination," *Desalination* 295, 70-81 (2012).
- L. Huber, P. Ruch, R. Hauert, G. Saucke, S. Kumar Matam, B. Michel, M.M. Koebel, "Monolithic nitrogen-doped carbon as a water sorbent for high-performance adsorption cooling," *RSC Advances* 6, 25267-25278 (2016) .
- S. Saliba, P. Ruch, W. Volksen, T.P. Magbitang, G. Dubois, B. Michel, "Combined influence of pore size distribution and surface hydrophilicity on the water adsorption characteristics of micro- and mesoporous silica," *Microporous and Mesoporous Materials* 226, 221-228 (2016).
- J. Ammann, P.W. Ruch, A. Studart, B. Michel, "Gas transport measurements in porous monolithic adsorbents," Poster at Sorption Friends, Milazzo, Sicily (2015).
- P. Ruch, S. Saliba, C.L. Ong, Y. Al-Shehri, A. Al-Rihaili, A. Al-Mogbel, B. Michel, "Heat-Driven Adsorption Chiller Systems for Sustainable Cooling Applications," Proceedings of the 11th International Energy Agency Heat Pump Conference, Montreal, Canada (2014).
- A. Al-Mogbel, P. Ruch, A. Al-Rihaili, S. Al-Ajlan, P. Gantenbein, A. Witzig, B. Michel, "The Potential of Solar Adsorption Air-Conditioning in Saudi Arabia: A Simulation Study," 5th International Conference on Solar Air-Conditioning, Bad Krozingen, Germany (2013).
- P. Ruch, J. Ammann, I. Meijer, B. Michel, "Aus warm mach kalt — Effizienzsteigerung in Rechenzentren mit thermisch getriebenen Wärmepumpen," *Bulletin* 6, 30-33 (2016).
- J. Marschewski, L. Brenner, N. Ebejer, P. Ruch, B. Michel, D. Poulikakos, "3D-printed fluidic networks for high-power-density heat-managing miniaturized redox flow batteries," *Energy Environ. Sci.*, Advance Article (2017).
- J. Marschewski, R. Brechbuehler, S. Jung, P. Ruch, B. Michel, D. Poulikakos, "Significant heat transfer enhancement in microchannels with herringbone-inspired microstructures," *International Journal of Heat and Mass Transfer* 95, 755-764 (2016).
- J. Marschewski, S. Jung, P. Ruch, N. Prasad, S. Mazzotti, B. Michel, D. Poulikakos, "Mixing with herringbone-inspired microstructures: overcoming the diffusion limit in co-laminar microfluidic devices," *Lab On A Chip* 15, 1923-1933 (2015).
- A. Sridhar, M.M. Sabry, P. Ruch, D. Atienza, B. Michel, "PowerCool: Simulation of integrated microfluidic power generation in bright silicon MPSoCs," IEEE/ACM International Conference on Computer-Aided Design (ICCAD), pp. 527-534 (2014).
- M.M. Sabry, A. Sridhar, D. Atienza, P. Ruch, B. Michel, "Integrated microfluidic power generation and cooling for bright silicon MPSoCs," Design, Automation and Test in Europe (DATE) Conference and Exhibition (2014).

Experimental Demonstration

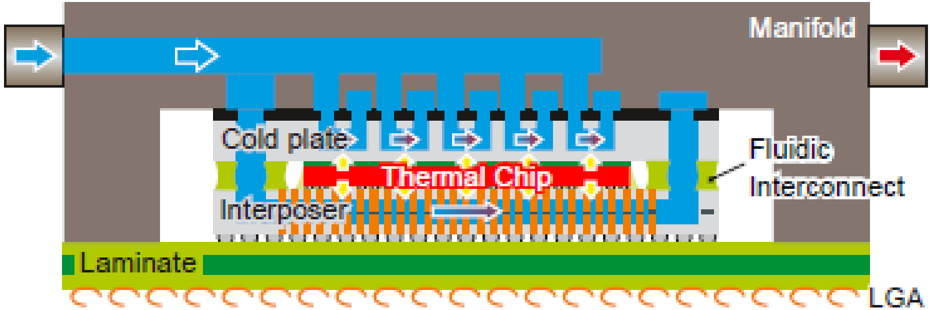


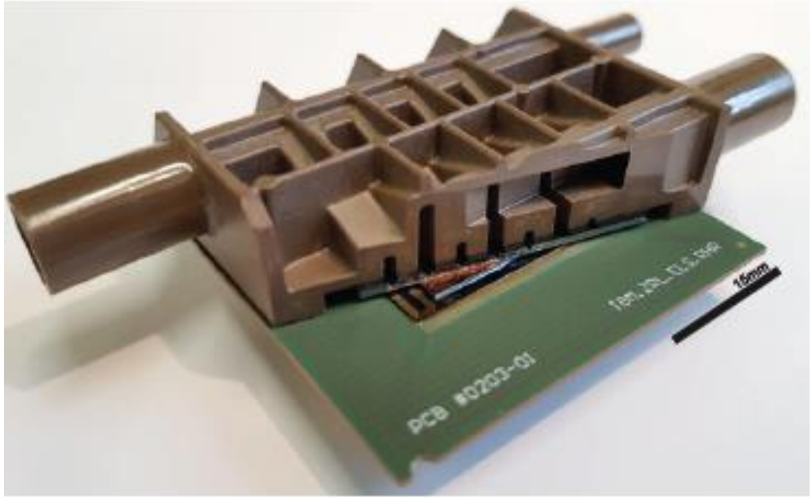


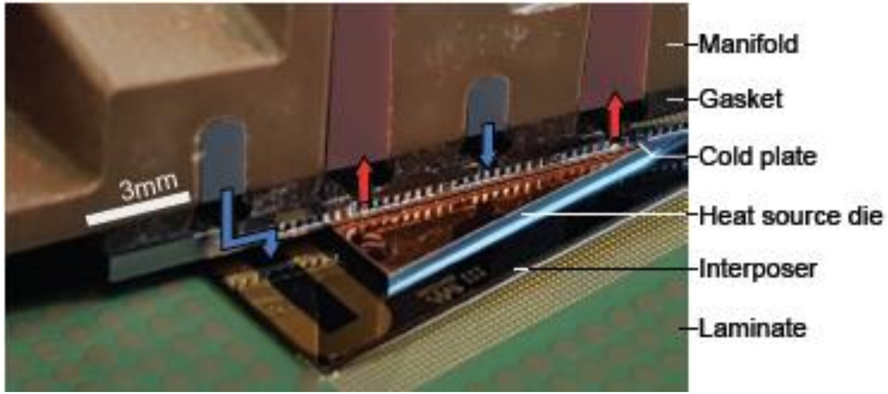


Bruno Michel, bmi@zurich.ibm.com

Dual-Side Heat Removal







- Thermally viable solution for a highpower die at the bottom of the stack,
- Free up silicon area in the other tiers, because CPU power and signaling wiring from the carrier do not have to be passed on through the other tiers.
- A pyramid-shape chip stack can also be supported.

Bruno Michel, bmi@zurich.ibm.com

What we can learn from Computer History

- Integration density and efficiency wins over performance (again)
- “Toys” have and will become supercomputers
- Managing high areal and volumetric power densities is the key to enable a new density roadmap
- Human Centric Sensing and Computing: Computers that are closer to humans can do much more useful things and develop into a cognitive companion
- Call to action: Future of Computing will be driven by “smart” system-level integration
- Main driver: **Electrical and thermal packaging to reduce resistances**
- Resistances are dissipative elements and key culprits for efficiency loss

Figure. Jeff Holter with his original 38-kg radio-electrocardiograph recording device in 1947.

Bruno Michel, bmi@zurich.ibm.com

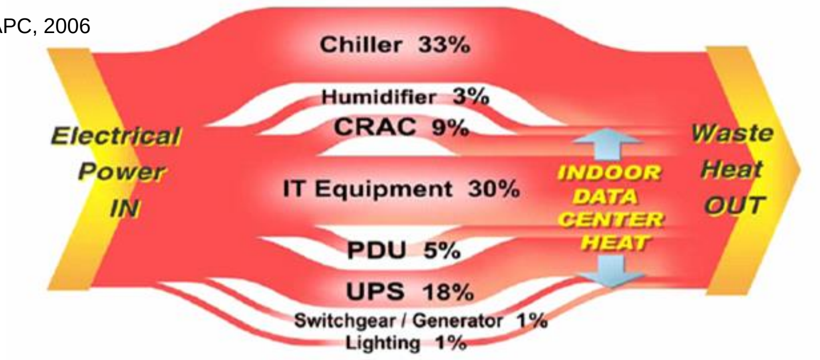
Questions

- **How efficient is a current computer in %?**
- **How much of the volume of a computer is used for the compute function?**

Answers

- **Efficiency is 0.000’004%**
- **Volume used for compute is <1 ppm (part per million)**

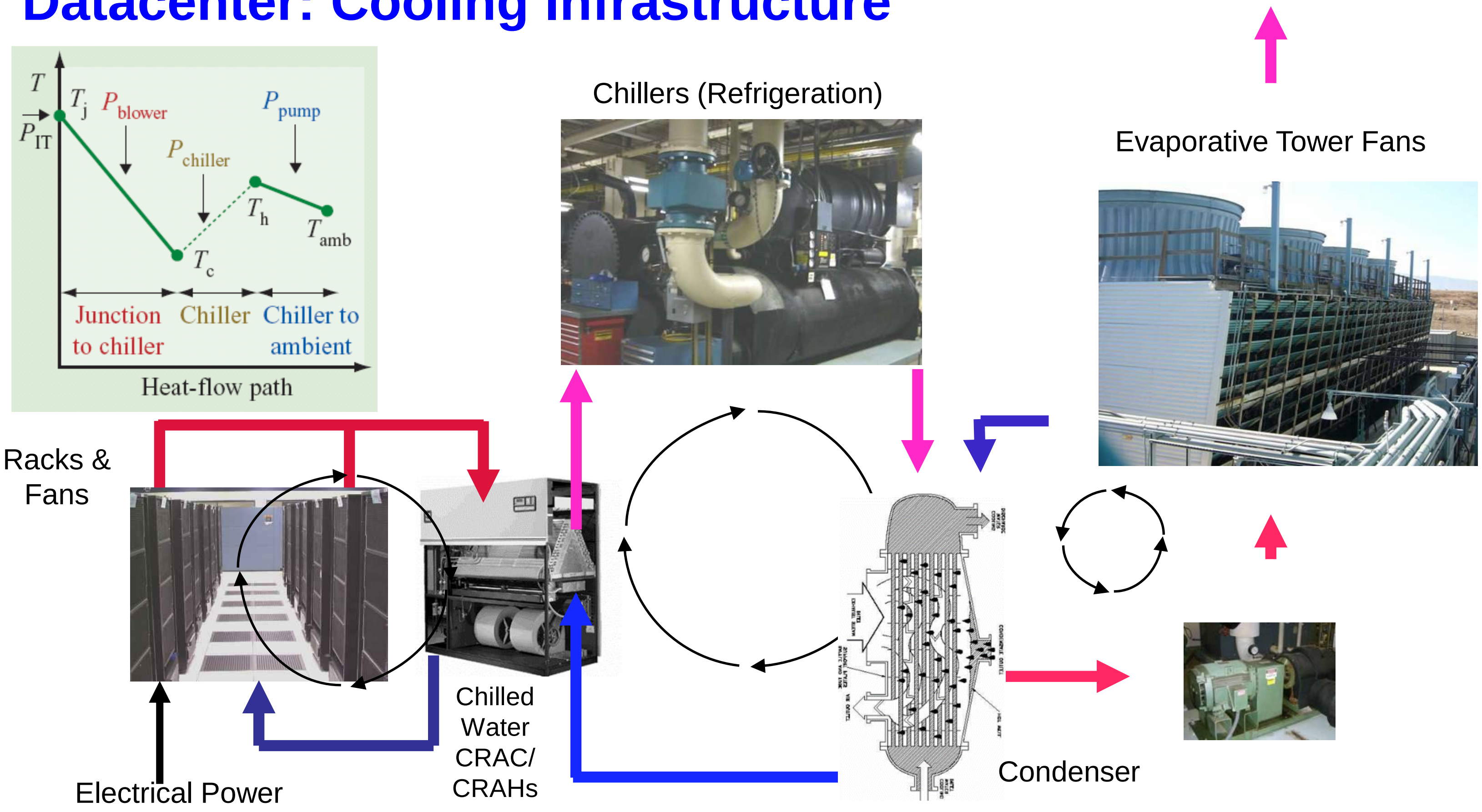
Brouillard, APC, 2006



- **A computer is an inefficient “joule heater” since it produces 10-20ºC “heat”. The output is 0.000’004% ICT and 99.999’996% “negative grade” heat**
- **Consequence: We have a lot to do to change energetics and density!**

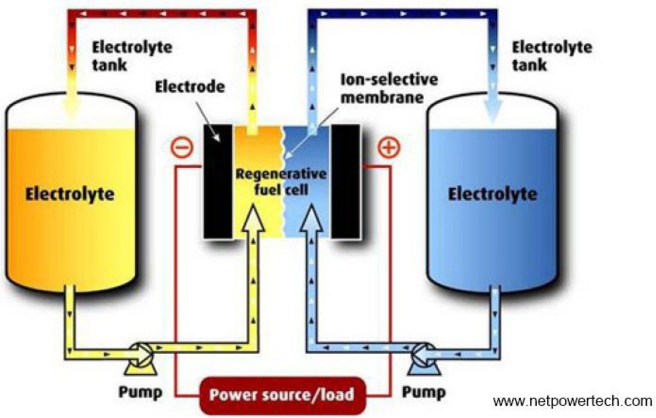
Bruno Michel, bmi@zurich.ibm.com

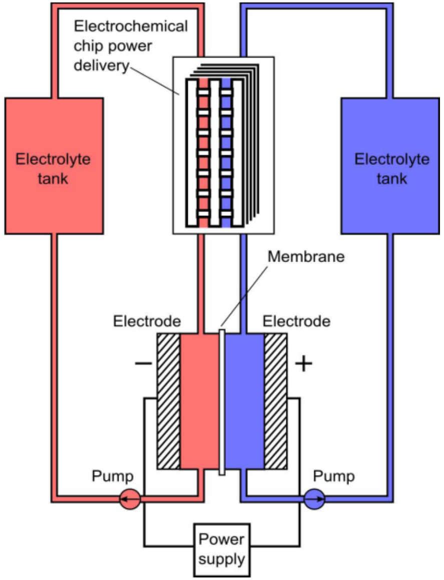
Datacenter: Cooling Infrastructure



Bruno Michel, bmi@zurich.ibm.com

Electrochemical Redox Flow Batteries





Characteristics
- Soluble redox species
- Inert electrodes
- Independent energy and power properties
- Single charge and discharge unit

Technology benefits
- No changes in electrode active surface area
- Deep discharge and high power possible
- No electrode lifetime limitations

Electrochemical chip power supply
- Single macroscopic charging unit
- Multiple chip-level discharge units
- Satisfies congruent demand for power delivery and heat removal

Bruno Michel, bmi@zurich.ibm.com

High Quality Healthcare by Wearables & Contextual Patient Model

Patient

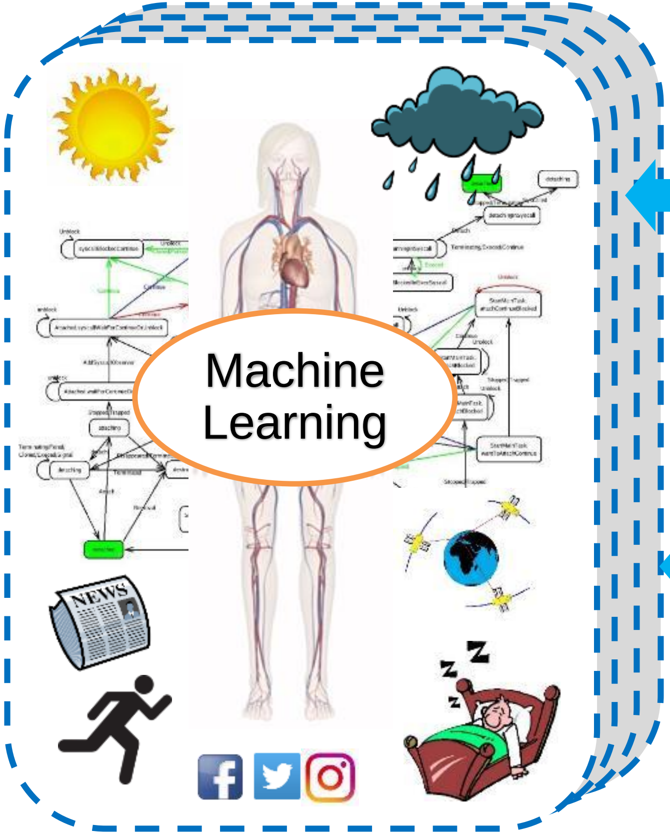
Mobile Sensing

Wearables

- Vital Parameters
- Bio-markers
- Social-parameters
- Environment
- Health record

Measure

Contextual Patient Models



Correlate between patients
Learn and improve model
Include Historical Data

Detect

Reasoning

Predict

Serve/Improve/Save

Advise

Alert

EMERGENCY RESPONSE TEAM

Science

Bruno Michel, bmi@zurich.ibm.com