

SPG MITTEILUNGEN COMMUNICATIONS DE LA SSP

AUSZUG - EXTRAIT

Progress in Physics (53)

The physics of data centers – beyond the transistor node

Patrick Ruch, Thomas Brunschwiler, IBM Research – Zurich

This article has been downloaded from:
http://www.sps.ch/fileadmin/articles-pdf/2016/Mitteilungen_Progress_53.pdf

© see http://www.sps.ch/bottom_menu/impressum/

Progress in Physics (53)

The physics of data centers – beyond the transistor node

Patrick Ruch, Thomas Brunschwiler, IBM Research – Zurich

Introduction

In 1974, Robert Dennard et al. described the physics of how to shrink a field-effect transistor (FET) while maintaining its performance characteristics (Dennard et al., 1974). The beauty of the so-called Dennard scaling was that the power consumption of a FET per unit area was constant, so that populating the same microprocessor area with more FETs of smaller dimensions would make the system faster, but not more power-hungry. This conclusion was certainly valid for the FETs developed in the early 1970s; however, substantial advances in the photolithographic techniques, gate materials and transistor geometries used by the semiconductor industry have reduced the gate length to less than 20 nm and the equivalent gate oxide thickness to about 1 nm. This evolution toward atomically thin gate oxides has resulted in increased gate leakage currents, whereas the reduction in supply voltage to ~ 1 V has resulted in a substantial increase in the subthreshold leakage currents (Roy et al., 2003). These leakage currents are responsible for a considerable amount of power consumption even when the chip is idle (Nowak, 2002). The combination of these physical limits at the nanometer scale and limited power delivery and heat dissipation at the centimeter scale has stalled traditional transistor scaling and resulted in a saturation of clock speed and chip power (Borkar & Chien, 2011; Shalf et al., 2009).

Outside of the classical FET, there is still significant room for improvement of microprocessor performance and energy efficiency. For example, the metallic interconnects forming the short-range on-chip wiring suffer from increased surface scattering after scaling beyond the 50 nm node (Davis et al., 2001), whereas global interconnects spanning across the chip do not scale well because the chip size has remained roughly constant (Borkar, 2013). The disparity in interconnect scaling and transistor scaling means that the majority of power supplied to a chip is spent on data transmission via interconnects rather than computation via FETs (Moore & Greenfield, 2008). Moreover, the scaling of interconnects for data transmission suffers at the chip edge, where solder interconnects at a fairly large pitch (i.e., more than $100 \mu\text{m}$) bridge to the printed circuit board (PCB). This limits communication bandwidth and results in a competition between

interconnects allocated to the power supply and those used for data transmission (Stanley-Marbell et al., 2011).

These developments have spurred interest in alternative approaches to value creation in the design of computers. They include key enabling technologies, such as 3D packaging (Knickerbocker et al., 2008; Ramm et al., 2010), advanced heat dissipation (Brunschwiler et al., 2008) and power delivery (Andersen et al., 2014; Chang et al., 2010). Further, the thermodynamics of the entire datacenter facility needs to be revisited (Brunschwiler et al., 2009; Shah et al., 2008) to ensure efficient use of energy and counteract the increasing electricity consumption of computing systems worldwide (Koomey, 2011; Van Heddeghem et al., 2014). Our contribution describes a selection of scientific and technological innovations to address the above challenges. Overall, the physics of computing systems spans 12 orders of magnitude in characteristic length scales (Figure 1), thereby representing a formidable range of scientific and technological challenges.

The ideal chip package – power in, heat out, and abundant bandwidth

Wiring congestion has evolved into an important challenge in modern microprocessors. The availability of vertical electrical interconnects, so-called through-silicon vias (TSVs), enables the stacking of integrated circuit (IC) dies to overcome the chip-edge interconnect and wiring limitations (Figure 2a) (Erdmann et al., 2014; Khurshid & Lipasti, 2013). In this way, the bandwidth, latency and energy efficiency for die-to-die communication can be reduced by an order of magnitude thanks to proximity and improved interconnect densities.

However, the interface area for power delivery and heat removal to and from the chip stack stays invariant, whereas the electrical and thermal interfaces scale proportionally to the number of dies in the chip stack. Thus, novel materials and technologies are key to support the scalability of power delivery and heat removal. We will discuss two novel material formulations supporting current 3D chip-stack topologies with power delivery through the front and heat removal through the back-side of a chip stack. Both benefit from

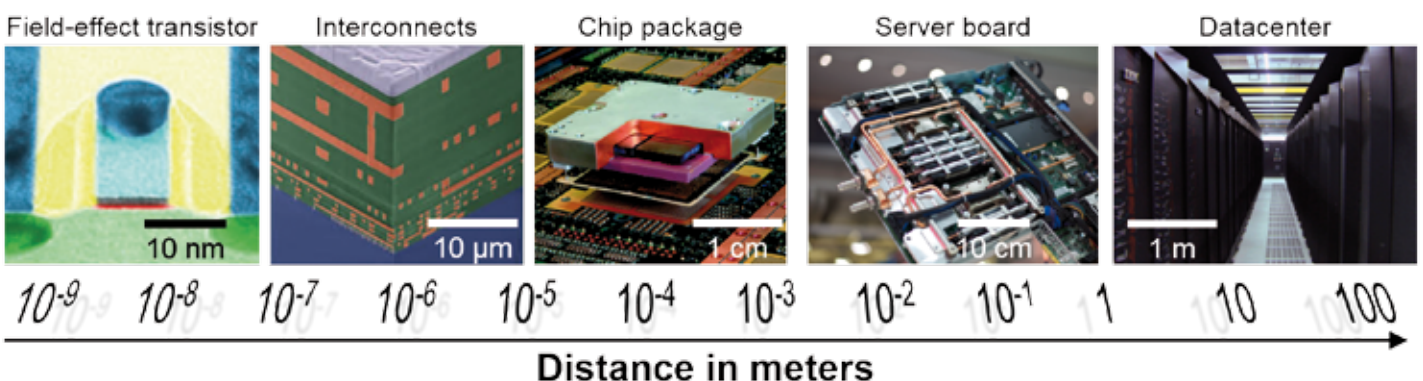


Figure 1: Characteristic length scales of computing systems.

nanoscale effects to improve thermal or electrical transport across interfaces. A topology change towards volumetric heat removal and power delivery through embedded fluid networks, providing access to coolants and electrolytes into the chip stack is also discussed, with the main focus on electrochemical power delivery.

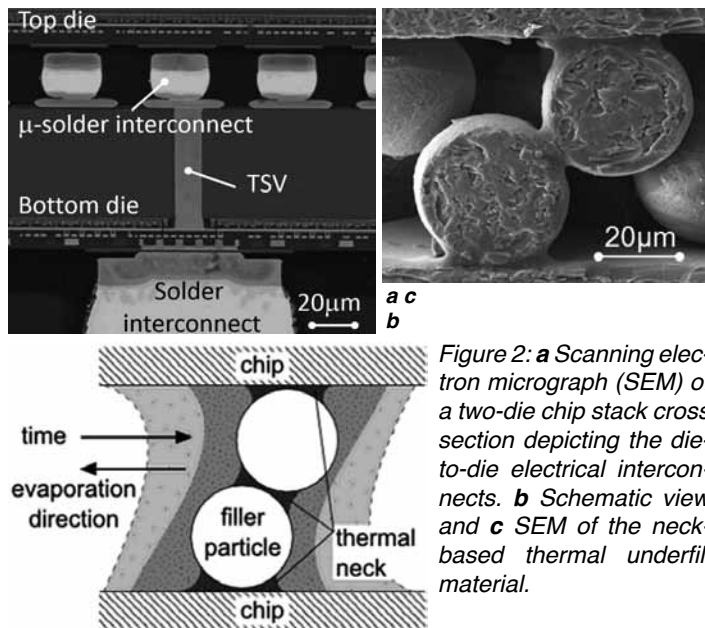


Figure 2: **a** Scanning electron micrograph (SEM) of a two-die chip stack cross section depicting the die-to-die electrical interconnects. **b** Schematic view and **c** SEM of the neck-based thermal underfill material.

To mitigate thermal gradients within the chip stack, a neck-based composite material with enhanced thermal conduction can be infiltrated between the microsolder interconnects, which impose a natural thermal barrier owing to the low area fill fraction of the solder interconnects (Brunschwiler et al., 2012). The infiltration procedure is performed in three main steps: first, micron-sized alumina filler particles are accumulated between the IC dies using centrifugal forces. Second, an aqueous nano-suspension is induced into the percolating particle bed by capillary action. During the subsequent evaporation procedure of the dispersant, capillary bridges, such as in wet sand, form between the filler particles and the concentration of alumina nanoparticles in the suspension increases. As a result, the nanoparticles assemble and form quasi-areal contacts, so-called necks in the point contact regions of the filler particles. Third, the remaining pores are filled with an epoxy, which is cured to provide the required mechanical integrity and chip-stack reliability. Thermal transport between the IC dies could be improved three-fold by means of the neck-based thermal underfill compared with state-of-the-art materials, thus supporting heat dissipation to the heat sink attached to the back-side of the chip stack (Zürcher, et al., 2015).

The ultimate heat-removal approach, which scales with the number of dies in the chip stack, is the introduction of microchannels between the active dies (Figure 3) (Brunschwiler et al., 2008). Coolants such as water can be pumped by a pressure gradient through the microchannels, which have a hydraulic diameter of $\sim 50 \mu\text{m}$, to remove the dissipated heat by convection in close proximity of the transistors. Die-to-die communication can still be maintained by embedded TSVs in the microchannel walls. Sealing rings around TSVs prevent electrical shorting through the liquid. The dissipation of power densities of up to 3.9 kW/cm^3 was demonstrated on lab-scale test vehicles, while maintaining the thermal

budget of 85°C maximal junction temperature (Brunschwiler et al., 2010).

The increased power levels supported by the discussed heat-removal methods require also novel power-delivery concepts and technologies to achieve a balanced system performance. Currently, an array of lead-free solder interconnects is used as interface between the PCB and the bottom-most die in the chip stack. During a reflow process with temperatures above the liquidus temperature of the solder (e.g. 260°C for SnAgCu alloys), the copper pads of the mating components are wetted, resulting in electrical interconnects with a threshold current of up to 200 mA, defined by electromigration (Tong et al., 2013). The high joining temperature in combination with the two disparate thermal-expansion coefficients of the mating components (i.e., 2.3 ppm/K for the silicon die and $\sim 18 \text{ ppm/K}$ for the PCB) result in mechanical shear above the solder yield strength, causing plastic deformation. Moreover, during the joining process, a brittle intermetallic compound is formed at the pad-solder interface. Both effects can cause interconnect failure, compromising the lifetime of the server product (Wang et al., 2012).

To overcome the limitations of solder interconnects, all-copper interconnects formed at low temperatures are being proposed (Zürcher et al., 2015) (Figure 3b). A copper paste consisting of a bi-modal distribution of copper nano- and micro-particles with a diameter 15 nm and $3 \mu\text{m}$, respectively, is formulated in an organic dispersant. To prevent the oxidation of particles after their synthesis, surface functionalization is applied in the reactor in inert gas conditions. By means of doctor-blading, a thin film of the paste with

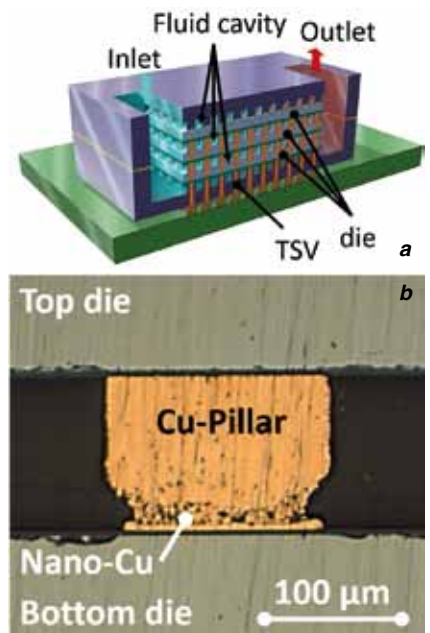


Figure 3: **a** Isometric view of an inter-layer-cooled 3D chip stack with embedded fluid channels.

b Optical micrograph of an all-copper interconnect cross section.

a thickness of $20 \mu\text{m}$ results. On the IC die, an array of copper pillars is formed by electrodeposition as substitution of the solder-ball array. The copper paste is transferred to the tips of the copper pillars of the IC die by means of a dipping process in a flip-chip bonder. Subsequently, the die is aligned and placed on the PCB. The paste now bridges the area between the copper pillar and the pad on the PCB. Then the evaporation of the organic binder, the desorption of the surface functionalization, and the annealing of the copper nano- and micro-particle agglomeration are done in an oven at temperatures as low as 200°C (i.e., below the liquidus temperature of copper, which is 1084°C) in a formic-acid atmosphere to reduce any residual copper oxide. As a result, an all-copper electrical interconnects is formed at substantially lower joining temperatures, without formation

of intermetallic compounds and with better electromigration resistance than solder joints. Moreover, these interconnects support the power-density scaling, while maintaining system reliability.

A fundamental change in the power-delivery architecture in the form of a liquid-cooling/redox-flow hybrid fluid network has been proposed based on the integrated microchannels in the chip stack, which previously only served to dissipate heat from the volume (Ruch et al., 2011). Adding redox-active molecules or ions to an aqueous solution enables convective transport of chemical energy along with the cooling water, equivalent to grid-scale redox flow batteries (Skylas-Kazacos et al., 2011). The conversion of chemical energy into electrical energy via an electrochemical cell embedded in the fluid network allows the multiple voltage conversion stages used in the conventional power-delivery chain in a datacenter (Pratt & Kumar, 2007) to be bypassed. Further, as the heat dissipation and the power demand on a chip are locally congruent, the concept of using one fluid network for both cooling and power delivery appears attractive in terms of energy efficiency.

It is illustrative to compare the energy penalty of charge transport via electrical wires with the convective transport of ions in a liquid conduit. For a current flowing through a wire, the main loss mechanism is Joule heating, whereas for laminar fluid flow through microchannels, it is friction loss. The critical diameter D_{crit} for which the Joule heating in a wire is equivalent to the friction loss in a pipe is derived as

$$D_{crit} = \frac{1}{fczF} \sqrt{32\mu\sigma} \quad (1)$$

where σ is the electrical conductivity of the wire and μ the viscosity of the fluid. For convective transport of a redox species at a volumetric flow rate v_L and a concentration c , an equivalent current $I = czFv_L f$ can be calculated, where z is the number of electrons exchanged in the redox reaction and F is Faraday's constant. The coefficient f represents the fraction of redox molecules converted during the electrochemical discharge ($0 \leq f \leq 1$).

For $\mu = 20$ mPa·s (concentrated electrolyte solution), $\sigma = 6 \cdot 10^7$ S/m (copper), $c = 1$ mol/L and $z = 1$, $D_{crit} = 64/f \mu\text{m}$. For a channel with diameter $D < D_{crit}$, charge transport is more efficient in wires than in fluid channels, whereas for $D > D_{crit}$, convective ion transport in fluids is more efficient. Assuming a conversion rate of 20% of the redox molecules in solution ($f = 0.2$) (Kjeang et al., 2007), charge transport in fluid channels is favorable only in hydraulic diameters

> 320 μm . This magnitude of channel diameter corresponds well to the microchannels used in liquid cooling.

Electrolytic charging of redox electrolytes, on-chip testing of power delivery and heat removal are carried out in an electrochemical fluid loop containing an electrochemical flow cell and a custom-designed thermal test board (Figure 4a). Miniaturized redox flow cells comprising a polymeric semi-permeable membrane or a nanoporous separator together with electrochemically active carbon-fiber electrodes integrated in silicon have been demonstrated (Figure 4b) (Ruch et al., 2015). Preliminary work has shown that an on-chip all-vanadium redox flow battery can be used to power up the caches with an average power consumption of 1 W/cm² (Sabry et al., 2014).

Datacenter thermodynamics– the case for heat recovery, and new perspectives for utilization

The annual electricity consumption in datacenters has been estimated to account for about 30 billion USD or 330 TWh, which represents 2% of the worldwide electricity consumption (Meijer, 2010). The power consumption of 61 standalone datacenters was found to be on average 1.9 times greater than the power consumed by their computing hardware alone (ENERGY STAR, 2014). The overhead is mainly due to the compression chillers, which provide cold air, and to electrical losses in the power-delivery chain. The reason for the high cooling-energy need is the cost-effective but low-performing thermal design of air heat sinks. A measure of the efficacy of a heat sink is its thermal resistance $R_{th} = \Delta T/\dot{q}'$, where ΔT is the mean difference between the chip surface temperature and the cooling fluid and \dot{q}' is the heat flux density. A typical air-cooled heat sink exhibits $R_{th} \approx 0.6$ K cm²/W, and the corresponding temperature difference between chip and coolant for a chip-level power density of 100 W/cm² is 60 K. To maintain chip temperatures below 85°C for reasons of reliability, the temperature of the air entering the heat sink must therefore be cooled to below 25°C.

Besides the direct cost associated with active cooling, there are also thermodynamic implications that affect the overall efficiency of the system. According to the Carnot theorem, the maximum amount of work that can be extracted between the temperature level of the chip surface, T_j , and that of the ambient air, T_a , is

$$\dot{W}_{max} = \eta_c \dot{q}' A = \left(1 - \frac{T_a}{T_j}\right) \dot{Q}, \quad (2)$$

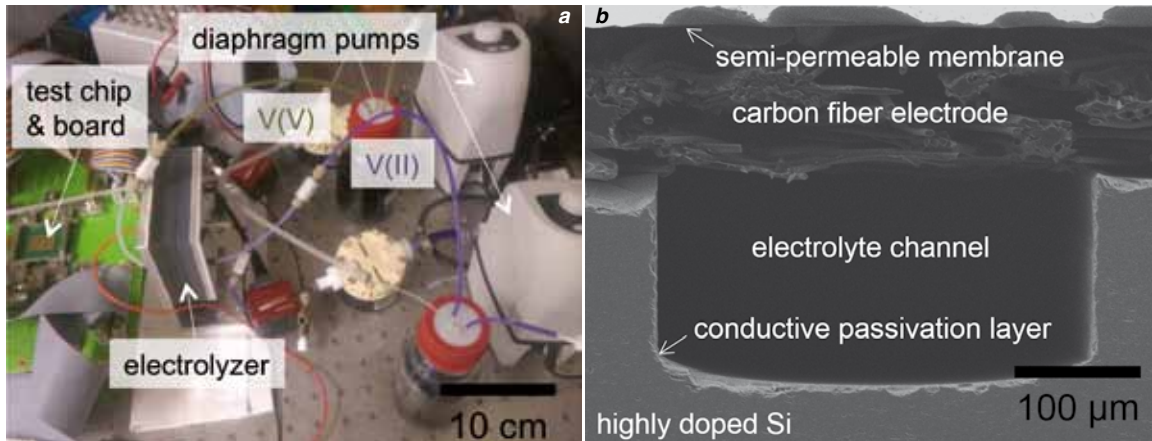


Figure 4: **a** Experimental setup used to generate and test redox electrolytes and cell designs for on-chip electrochemical power delivery. The solutions in the photograph are 1 M V(V) (yellow) and 1 M V(II) (purple), each in 2 M H₂SO₄. **b** SEM of a cross section of an electrochemical half-cell integrated on silicon.

where η_c is the Carnot efficiency, A the chip area and \dot{Q} the total power dissipation. The quantity \dot{W}_{\max} is sometimes referred to as the exergy content of the dissipated heat (Shah et al., 2006). A fundamental difference between exergy and energy is that the former is not conserved. The upper bound to the work that can be extracted from the heat that is actually recovered in the cooling fluid is

$$\dot{W} = \left(1 - \frac{T_a}{T_j - R_{th}\dot{Q}}\right)\dot{Q}, \quad (3)$$

The exergy loss due to thermal resistance is obtained by subtracting Eq (3) from Eq. (2). When $(T_j - R_{th}\dot{Q}) < T_a$, Eq. (3) yields $\dot{W} < 0$, i.e., additional work has to be invested by means of a heat-pump cycle to enable heat rejection to T_a (Brunschwiler et al., 2009).

Based on the above, providing heat sinks with reduced R_{th} has a two potential merits. First, the expression $(T_j - R_{th}\dot{Q})$ can be made greater than T_a to enable heat rejection to ambient without requiring an additional heat-pump cycle, which is referred to as *free cooling*. Second, the exergy content of the recovered heat can be increased, which potentially enables further utilization of the heat rather than discarding it into the environment.

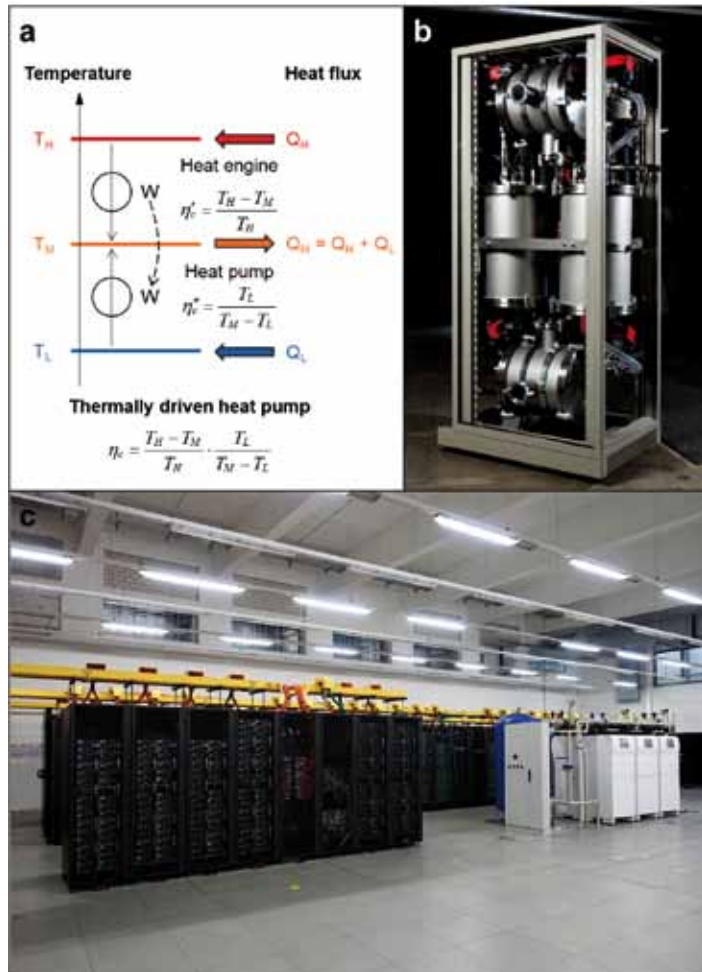


Figure 5: **a** Thermodynamic principle of a thermally-driven heat pump operating with three temperature levels. The Carnot efficiencies of the individual heat transformation stages and the overall Carnot efficiency are given. **b** Experimental adsorption heat pump test rig at IBM Research – Zurich comprising two vertical adsorption chambers, a condenser chamber and an evaporator chamber. The rack height is 2 m. Photograph courtesy of SPF Rapperswil. **c** CoolMUC-2 HPC cluster (left) providing the driving heat for six adsorption chillers (right). Photograph courtesy of SorTech AG.

To this end, direct liquid-cooled heat sinks have been developed with $R_{th} < 0.1$ Kcm²/W (Escher et al., 2010). For fluid outlet temperatures of 60°C, a direct use of the recovered heat in space-heating systems has been demonstrated (Zimmermann et al., 2012). Furthermore, a thermally-driven heat-pump cycle can provide direct conversion of the recovered heat into cooling (Ziegler, 2009). In contrast to a conventional heat-pump cycle based on mechanical compression, there are three characteristic temperature levels for thermally-driven heat pumps: high (T_H), medium (T_M) and low temperature (T_L). The operation can be described thermodynamically as a combination of a heat engine and a heat-pump cycle (Figure 5a).

The heat recovered from datacenters employing servers with low R_{th} can be supplied at temperature T_H to adsorption heat pumps, which rely on the reversible physisorption of a vapor on a solid desiccant (Aristov, 2014). The heat uptake at the low temperature T_L can be used to provide cooling for non-liquid-cooled datacenter components, whereas the T_M level corresponds to the outside ambient, which acts as the final heat sink. High sorption rates and evaporation of liquid water at sub-ambient temperatures are achieved by implementing the desiccant within evacuated sealed vessels (Figure 5b).

Recently, the largest combination of a computing cluster with adsorption chillers to date was put into operation at Leibniz Rechenzentrum (LRZ). An HPC cluster is direct water-cooled, and the recovered heat is supplied at up to 60°C to six adsorption chillers with a nominal cooling power of up to 16 kW each (Figure 5c). The chilled water is supplied to data-storage cabinets. The thermal coefficient of performance (COP_{th}), which describes the ratio Q_L/Q_H (cf. Figure 3a), is on the order of 0.6. However, the more relevant quantity in the context of waste-heat usage is the electrical COP (COP_{el}), which describes the ratio Q_L/W_E , where W_E is the electricity consumption of the system. Compared with mechanical compression chillers, which are normally used to provide datacenter cooling, adsorption-chiller systems in the field have a COP_{el} that is at least twice as high. In other words, the electricity consumption for the cooling infrastructure is at least halved as a result of the switch from compression chillers to waste-heat-driven adsorption chillers, which is enabled by the implementation of heat sinks with low exergetic losses.

Conclusions

The physics of miniaturization has essentially determined the roadmap for the success of the semiconductor industry for more than 50 years, and now several emerging trends highlight that advances in computing technology are not just restricted to the chip. For example, the disparity in transistor scaling and the interconnectivity at the chip edge have triggered the need for alternative power-delivery approaches. The same holds true for novel chip architectures relying on 3D packaging. Further, the unabated increase in chip-level power dissipation for computationally intensive systems calls for high-performance cooling solutions that minimize thermal resistance. A co-design of power-delivery and heat-dissipation structures appears essential to fully support the scalability of computing resources and interconnects for data transmission. Finally, with the cost of purpose-built

computing hardware dropping and concerns about resource utilization increasing, the energy efficiency of computing systems has become increasingly important. As discussed herein, it is clear that to address these challenges holistically, various physical principles need to be applied to the world of computing at various length scales. There still is ample opportunity for innovation and scientific exploration across many disciplines to advance computing.

References

- Andersen, T. M., Krismer, F., Kolar, J. W., Toifl, T., Menolfi, C., Kull, L., Morf, T., Kossel, M., Brändli, M., Buchmann, P., Francese, P. A. (2014). A Sub-ns Response On-Chip Switched-Capacitor DC-DC Voltage Regulator Delivering 3.7 W/mm² at 90% Efficiency Using Deep-Trench Capacitors in 32 nm SOI CMOS. In *IEEE International Solid-State Circuits Conference 2014* (pp. 90–91).
- Aristov, Y. (2014). Concept of adsorbent optimal for adsorptive cooling/heating. *Applied Thermal Engineering*, 72(2), 166–175.
- Borkar, S. (2013). Role of interconnects in the future of computing. *Journal of Lightwave Technology*, 31(24), 3927–3933.
- Borkar, S., & Chien, A. A. (2011). The future of microprocessors. *Communications of the ACM*, 54(5), 67–77.
- Brunschwiler, T., Michel, B., Rothuizen, H., Kloter, U., Wunderle, B., Oppermann, H., & Reichl, H. (2008). Interlayer cooling potential in vertically integrated packages. *Microsystem Technologies*, 15(1), 57–74.
- Brunschwiler, T., Paredes, S., Drechsler, U., Michel, B., Cesar, W., Leblebici, Y., Wunderle, B., Reichl, H. (2010). Heat-removal performance scaling of interlayer cooled chip stacks. *12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*.
- Brunschwiler, T., Schlottig, G., Ni, S., Liu, Y., Goicochea, J. V., Zürcher, J., & Wolf, H. (2012). Formulation of percolating thermal underfills using hierarchical self-assembly of micro- and nanoparticles by centrifugal forces and capillary bridging. *International Microelectronics and Packaging Society (IMAPS)*, 749–759.
- Brunschwiler, T., Smith, B., Ruetsche, E., & Michel, B. (2009). Toward zero-emission data centers through direct reuse of thermal energy. *IBM Journal of Research and Development*, 53(3), 1–13.
- Chang, L., Montoye, R. K., Ji, B. L., Weger, A. J., Stawiasz, K. G., & Dennard, R. H. (2010). A fully-integrated switched-capacitor 2:1 Voltage converter with regulation capability and 90% efficiency at 2.3A/mm². *IEEE Symposium on VLSI Circuits, Digest of Technical Papers*, 55–56.
- Davis, J. A., Venkatesan, R., Kaloyeros, A., Beylansky, M., Sour, S. J., Banerjee, K., Saraswat, K.C., Rahman, A., Reif, R., Meindl, J. D. (2001). Interconnect limits on gigascale integration (GSI) in the 21st century. *Proceedings of the IEEE*, 89(3), 305–324.
- Dennard, R. H., Gaensslen, F. H., Yu, H.-N., Rideout, V. L., Bassous, E., & Andre R. LeBlanc. (1974). Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9(5), 256–268.
- ENERGY STAR Score for Data Centers in the United States. (2014). Retrieved from https://www.energystar.gov/sites/default/files/tools/Data_Centers.pdf. Last accessed 08.02.2016.
- Erdmann, C., Lowney, D., Lynam, A., Keady, A., McGrath, J., Cullen, E., Breathnach, D., Keane, D., Lynch, P., De La Torre, M., De La Torre, R., Lim, P., Collins, A., Farley, B., Madden, L. (2014). A heterogeneous 3D-IC consisting of two 28 nm FPGA die and 32 reconfigurable high-performance data converters. *IEEE Journal of Solid-State Circuits*, 50(1), 258–269.
- Escher, W., Michel, B., & Poulikakos, D. (2010). A novel high performance, ultra thin heat sink for electronics. *International Journal of Heat and Fluid Flow*, 31(4), 586–598.
- Khurshid, M. J., & Lipasti, M. (2013). Data compression for thermal mitigation in the hybrid memory cube, *IEEE 31st International Conference on Computer Design, (ICCD)* (pp. 185–192).
- Kjeang, E., Proctor, B., Brolo, A., Harrington, D., Djilali, N., & Sinton, D. (2007). High-performance microfluidic vanadium redox fuel cell. *Electrochimica Acta*, 52(15), 4942–4946.
- Knickerbocker, J., Andry, P., Dang, B., Horton, R., Interrante, M., Patel, C., Polastre, R. J., Sakuma, K., Sirdeshmukh, R., Sprogis, E. J., Sri-Jayantha, S. M., Stephens, A. M., Topol, A. W., Tsang, C. K., Webb, B. C., Wright, S. L.. (2008). Three-dimensional silicon integration. *IBM Journal of Research and Development*, 52(6), 553–569.
- Koomey, J. G. (2011). Growth in Data Center Electricity Use 2005 To 2010. Retrieved from <http://www.koomey.com>. Last accessed 08.02.2016.
- Meijer, G. (2010). Cooling Energy-Hungry Data Centers. *Science*, 328(5976), 318.
- Moore, S., & Greenfield, D. (2008). The next resource war: computation vs. communication. *Proceedings of the 2008 International Workshop on System Level Interconnect Prediction*, 81–85.
- Nowak, E. (2002). Maintaining the benefits of CMOS scaling when scaling bogs down. *IBM Journal of Research and Development*, 46(2), 169–180.
- Pratt, A., & Kumar, P. (2007). Evaluation of Direct Current Distribution in Data Centers to Improve Energy Efficiency. *The Datacenter Journal*, (March).
- Ramm, P., Klumpp, A., Weber, J., & Taklo, M. M. V. (2010). 3D System-on-Chip technologies for More than Moore systems. *Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems*, 16(7, SI), 1051–1055.
- Roy, K., Mukhopadhyay, S., & Mahmoodi-Meimani, H. (2003). Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits. *Proceedings of the IEEE*, 91(2), 305–327.
- Ruch, P., Brunschwiler, T., Escher, W., Paredes, S., & Michel, B. (2011). Toward five-dimensional scaling: How density improves efficiency in future computers. *IBM Journal of Research and Development*, 55(5), 15:1–13.
- Ruch, P., Ebejer, N., Sridhar, A., & Michel, B. (2015). Power delivery and thermal management of electronic packages using redox flow systems, *The International Flow Battery Forum (IFBF)*, Glasgow.
- Sabry, M. M., Sridhar, A., Atienza, D., Ruch, P., & Michel, B. (2014). Integrated microfluidic power generation and cooling for bright silicon MPSoCs. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 1–6.
- Shah, A. J., Carey, V. P., Bash, C. E., & Patel, C. D. (2006). An Exergy-Based Figure-of-Merit for Electronic Packages. *Journal of Electronic Packaging*, 128(4), 360.
- Shah, A. J., Carey, V. P., Bash, C. E., & Patel, C. D. (2008). Exergy Analysis of Data Center Thermal Management Systems. *Journal of Heat Transfer*, 130(2), 021401.
- Shalf, J., Bashor, J., Patterson, D., Asanovic, K., Yelick, K., Keutzer, K., & Mattson, T. (2009). The MANYCORE Revolution: Will HPC LEAD or FOLLOW? *SciDAC Review, Fall*, 40–49.
- Skyllas-Kazacos, M., Chakrabarti, M. H., Hajimolana, S. a., Mjalli, F. S., & Saleem, M. (2011). Progress in Flow Battery Research and Development. *Journal of The Electrochemical Society*, 158(8), R55.
- Stanley-Marbell, P., Cabezas, V. C., & Luijten, R. (2011). Pinned to the walls: impact of packaging and application properties on the memory and power walls. *Proceedings of the 17th IEEE/ACM International Symposium on Low-Power Electronics and Design (ISLPED '11)*, 51–56.
- Tong, H. M., Lai, Y. S., & Wong, C. P. (2013). *Advanced Flip Chip Packaging*. New York: Springer, ISBN 978-1-4419-5768-9..
- Van Heddeghem, W., Lambert, S., Lannoo, B., Colle, D., Pickavet, M., & Demeester, P. (2014). Trends in worldwide ICT electricity consumption from 2007 to 2012. *Computer Communications*, 50(0), 64–76.
- Wang, Y., Chae, S. H., Dunne, R., Takahashi, Y., Mawatari, K., Steinmann, P., Bonifield, T., Jiang, T., Im, J., Ho, P. S. (2012). Effect of intermetallic formation on electromigration reliability of TSV-microbump joints in 3D interconnect, *Proceedings of the Electronic Components and Technology Conference (ECTC)* (pp. 319–325).
- Ziegler, F. (2009). Sorption heat pumping technologies: Comparisons and challenges. *International Journal of Refrigeration*, 32(4), 566–576.
- Zimmermann, S., Meijer, I., Tiwari, M. K., Paredes, S., Michel, B., & Poulikakos, D. (2012). Aquasar: A hot water cooled data center with direct energy reuse. *Energy*, 43(1), 237–245.
- Zürcher, J., Chen, X., Burg, B. R., Zimmermann, S., Hong, G., Studart, A. R., Potasiewicz, G., Warszynski, P., Brunschwiler, T. (2015). Enhanced thermal underfills by bridging nanoparticle assemblies in percolating micro-particle beds, *Proceedings of the International IEEE Conference on Nanotechnology (IEEE Nano 2015)* (pp. 577–580).
- Zürcher, J., Yu, K., Schlottig, G., Baum, M., Taklo, M. M. V., Wunderle, B., Warszynski, P., Brunschwiler, T. (2015). Nanoparticle assembly and sintering towards all-copper flip chip interconnects, *Proceedings of the Electronic Components and Technology Conference (ECTC)* (pp. 1115-1121)