Through the Looking Glass—The 2018 Edition

Trends in solid-state circuits from the 65th ISSCC

The ISSCC 2018 Program Committee (from left): Edith Beigné, Eugenio Cantatore, Anantha Chandrakasan, Laura Fujino, Stefano Pellerano, Alison Burdett, Axel Thomsen, Piet Wambacq, Byeong-Gyu Nam, Frank O'Mahony, Marian Verhelst, Makoto Ikeda, Leland Chang, Makoto Nagata, Un-Ku Moon, and David Blaauw.

Denis C. Daly,
Laura C. Fujino,
and Kenneth C. Smith

Digital Object Identifier 10.1109/MSSC.2017.2771103
Date of publication: 31 January 2018

The International Solid-State Circuits Conference (ISSCC) is the flagship conference of the IEEE Solid-State Circuits Society. This year, for the 65th ISSCC, the theme is "Silicon Engineering a Social World." Continued advances in solid-state circuits and systems have brought ever-more powerful communication and computational capabilities into mobile form factors. Such ubiquitous smart devices lie at the heart of a revolution shaping how we connect, collaborate, build relationships, and

share information. These social technologies allow people to maintain connections and support networks not otherwise possible; they provide the ability to access information instantaneously and from any location, thereby helping to shape world events and culture, empowering citizens of all nations, and creating social networks that allow worldwide communities to develop and form bonds based on common interests.

The ISSCC covers a spectrum of design approaches in various technical areas, which may be broadly categorized as analog systems, power management, analog-to-digital data conversion, communication systems, digital systems, and memory, along with innovative topics such as micromachines and microelectromechanical systems (MEMS), imagers, sensors, and biomedical devices, including forward-looking solutions that may be several years away from being commercialized.

The 11 ISSCC technical subcommittees annually update their analysis of industry trends for the benefit of the community at large. This article summarizes some of these views in selected technical areas. A more comprehensive trends document is available at www.isscc.org.

Analog Systems: Analog

Subcommittee Chair: Kofi A.A. Makinwa, Delft University of Technology, The Netherlands

Reduction in the energy consumption of analog systems—including sensors, amplifiers, and voltage and frequency references—has continued making gains to meet the demands of today's low-power systems. This trend is captured, for example, in the case of integrated temperature sensors by the movement toward the bottom right of the plot shown in Figure 1. Fast power-up and power-down of such systems are also necessary to facilitate the use of duty-cycling and achieve even higher energy efficiencies. Together, these trends portend a future in which portable devices can be powered indefinitely from sustainable sources, opening the door to the Internet of Everything (IoE), ubiquitous sensing, environmental monitoring, and medical applications.

The stringent clocking requirements of many types of battery-powered mobile and emerging IoE systems pose various challenges for system-level frequency references, including increased frequency stability, low noise, and strict temperature-coefficient (TC) control with a limited energy budget. To meet these challenges, the stability of integrated frequency references has been increasing steadily, as shown in Figure 2. This year at ISSCC 2018, a 7-MHz complementary–metal–oxide–semiconductor (CMOS) frequency reference is being demonstrated to achieve a TC of 3.85 ppm/°C, the lowest reported so far.

We also see an increase in the linearity and resolution of amplifiers for audio and precision amplifications, while still maintaining (or even improving) power efficiency.

Analog Systems: Power Management

Subcommittee Chair: Axel Thomsen, Cirrus Logic, Austin, Texas

While the manipulation and storage of information are efficiently performed digitally, the conversion and storage of energy are fundamentally performed with analog systems. The efficient control, storage, and distribution of energy are worldwide challenges and increasingly important areas for analog circuit research. Therefore, the key technologies for power management are predominantly analog.

For example, there is much interest in wireless power transmission for battery-charging applications, ranging from mobile handsets to medical implants. Increased performance in wireless power transmission is enabling more efficient power delivery over longer distances. There is also an explosion of technologies that allow energy to be collected from the environment in novel ways using photovoltaic, piezoelectric, or thermoelectric transducers, with a trend toward the use of multiple sources at the same time. A significant focus here is on analog circuits that can harvest submicrowatt power levels from multiple energy sources at tens of millivolts and so provide autonomy for IoE devices or supplement conventional battery supplies in portable devices. To achieve this, the attendant analog circuits must consume extremely low power so that surplus energy is available to charge a battery or supercapacitor. This trend is captured by the movement shown toward the top left of Figure 3.

Analog circuits also serve as a bridge between the digital and analog (real) worlds. Just as with actual bridges, analog circuits are often bottlenecks, and their design is critical to overall performance, efficiency, and robustness. In spite of this, digital circuits such as microprocessors have

driven the market, and semiconductor technology has been relentlessly optimized for digital circuits over the past 40 years to reduce their size, cost, and power consumption. Consequently, analog circuitry has proven increasingly difficult to implement using these modern IC technologies. For example, as transistor size has decreased, the range of analog voltages that transistors can handle, as well as their analog performance, has also decreased; at the same time, the variation observed in their analog parameters has increased.

These aspects of semiconductor technology explain two key divergent trends in analog circuits. One of these trends is to forgo the latest digital IC manufacturing technology and instead fabricate analog circuits in older technologies, which may be augmented to accommodate the high voltages demanded by increasing markets in medical, automotive, industrial, and high-efficiency lighting applications.

However, some other applications dictate full integration of analog and digital circuits in the most modern digital semiconductor process. For example, microprocessors with multiple cores can reduce their overall power consumption by dynamically scaling

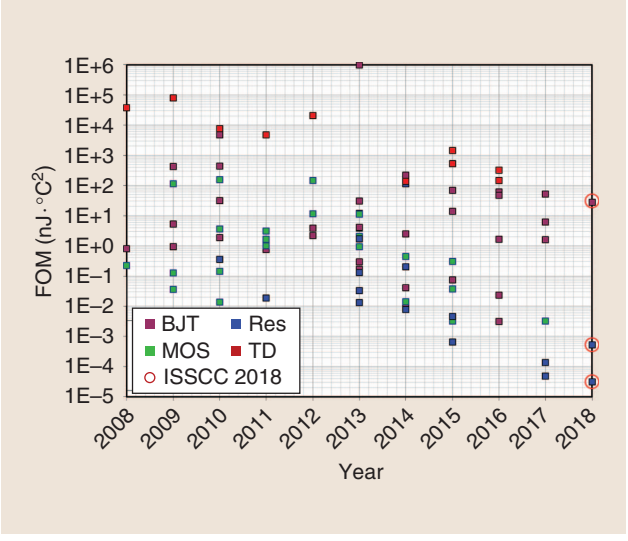

FIGURE 1: Various trends in the energy efficiency of integrated temperature sensors: resolution figure of merit (FOM) versus time. BJT: bipolar junction transistor; MOS: metal–oxide semiconductor; Res: resistor; TD: thermal diffusivity.

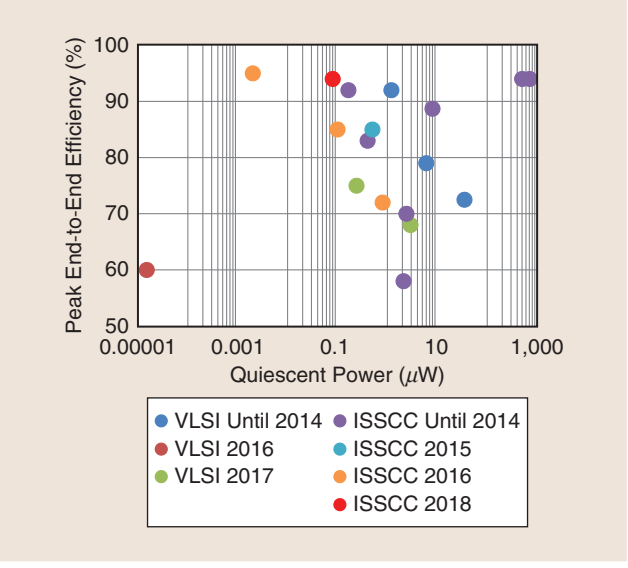

FIGURE 3: A comparison of integrated energy-harvesting systems showing peak end-to-end efficiency versus quiescent power. VLSI: very-large-scale integration.

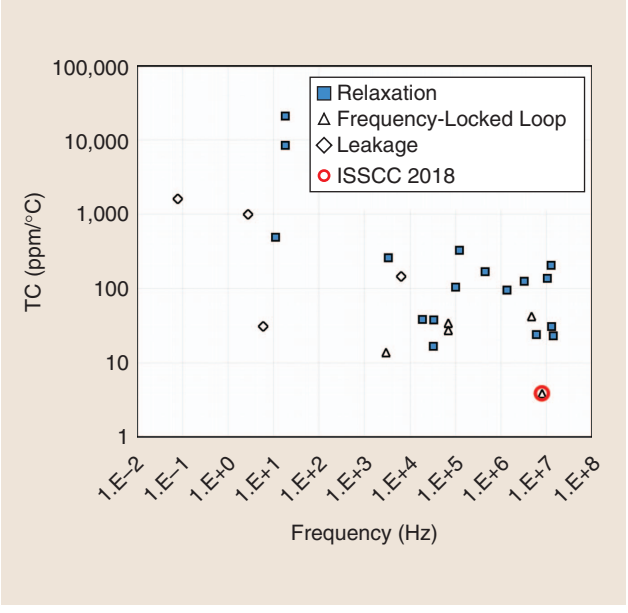

FIGURE 2: Various trends in the stability of integrated oscillators: TC versus frequency.

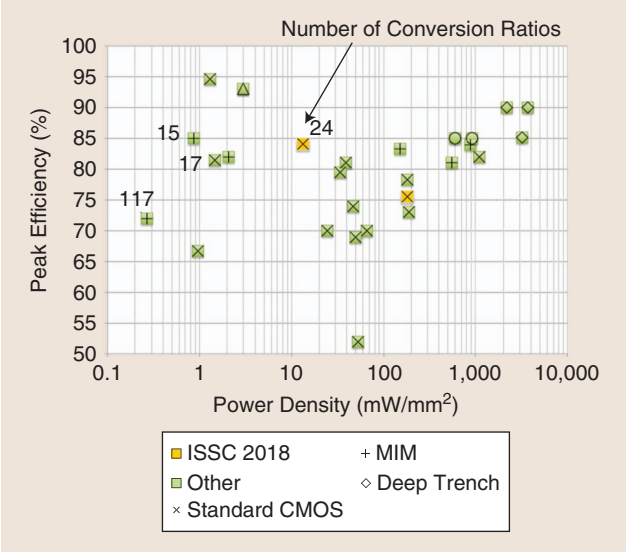

FIGURE 4: A comparison of integrated switched-capacitor power converters showing peak efficiency versus power density at peak efficiency. "Conversion ratios" refers to the relationship between output and input voltage. MIM: metal–insulator–metal.

operating voltage and frequency in response to time-varying computational demands. For this purpose, dc–dc voltage converters can be embedded alongside the digital circuitry, driving research into the delivery of locally regulated power supplies with increasing efficiency, decreasing die area, and increasing power density. Correspondingly, these recent advances are implemented in the latest low-cost standard CMOS technologies. Trends here are captured by the movement observed toward the top right of Figure 4.

This year at ISSCC 2018, we can see the performance of switched-capacitor power converters—with a large number of conversion ratios—making a significant step toward increasing their power density (Figure 4). A large number of conversion ratios are required in applications with high output and/or input voltage ranges, such as those needed for the IoE.

Analog Systems: Data Converters

Subcommittee Chair: Un-Ku Moon, Oregon State University, Corvallis, Oregon

Data converters form the key links between the analog physical world and the world of digital computing and signal processing prevalent in modern electronics. The need to faithfully preserve the signal across domains continues to pressure data converters to deliver more bandwidth and linearity while continuing to increase power efficiency. This year, ISSCC 2018 continues the trend of gigahertz digital-to-analog converters (DACs) with excellent linearity and noise performance. In analog-to-digital converters (ADCs) a combination of successive-approximation-register (SAR) and noise-shaping techniques is used to further improve speed, resolution, and power efficiency.

Figures 5–7 represent traditional metrics that capture innovative progress in ADCs. Figure 5 plots power dissipation relative to the Nyquist sampling rate (P/f_{snyq}), as a function of signal-to-noise-and-distortion ratio (SNDR), to provide a measure of ADC power efficiency. Note that a lower P/f_{snyq} metric on this chart represents a more efficient circuit. For low-to-medium-resolution converters, energy is primarily expended to quantize the signal; thus, the overall efficiency of this operation is typically measured by the energy consumed per conversion and quantization step. The dashed trend line in Figure 5 represents a benchmark of the 2-fJ conversion step.

Circuit noise becomes more significant with higher-resolution converters, necessitating a different benchmark proportional to the square of signal-to-noise ratio represented by the solid line. In Figures 5–7, designs published from 2009 to 2017 are shown in green, and older designs are shown in red, indicating a clear trend toward higher resolution at lower energy per conversion.

ISSCC 2018 designs are shown in blue, with superior performance moving toward the lower right of Figure 5. Figure 6 plots signal fidelity versus the Nyquist sampling rate normalized to power consumption. At low sampling rates, converters tend to be limited by thermal noise, independent of the sampling rate. Higher speeds of operation present additional challenges for maintaining accuracy in an energy-efficient manner, indicated by the roll-off versus frequency in the dashed line (Figure 6). The past ten years have resulted in an over 10-dB improvement in power-normalized signal fidelity, or a ten times improvement in speed for the same normalized signal fidelity. Of note, this year at ISSCC 2018, several designs improve

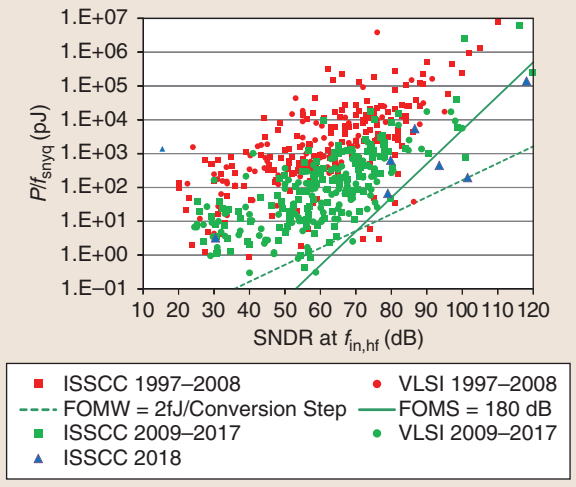

FIGURE 5: ADC power efficiency (P/f_{snyq}) as a function of SNDR. FOMW: FOM-Walden; FOMS: FoM-Schreier.

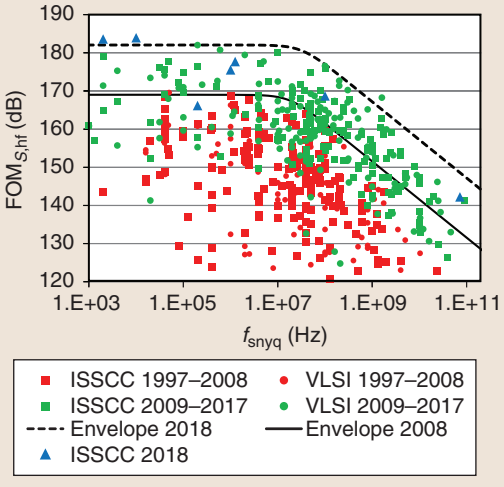

FIGURE 6: Power normalized noise and distortion versus the Nyquist sampling rate.

on the thermal-noise-limited efficiency using a combination of SAR and noise shaping.

Figure 7 shows achieved bandwidth as a function of SNDR. Sampling jitter or aperture errors coupled with an increased noise bandwidth make achieving both high resolution and high bandwidth a particularly difficult task. While, ten years ago, a state-of-the-art data converter showed an aperture error of approximately 1ps$_{rms}$, in recent years, designs with aperture errors below 100 fs$_{rms}$ have been published at a wide range of resolutions. ISSCC 2018 further advances the art with the first data converter having an input frequency above 30 GHz, using a 64-times time-interleaved SAR ADC.

Communication Systems: Radio Frequency

Subcommittee Chair: Piet Wambacq, IMEC, Leuven, Belgium

This year, ISSCC 2018 features ongoing advances in radio-frequency (RF) building blocks for broadband communications, millimeter-wave (mm-wave) sensing, and terahertz imaging. Topics include frequency generation; efficient transmit and robust receive techniques for emerging cellular narrowband IoE applications; full-duplex, fifth-generation (5G) massive multiple-input, multiple-output (MIMO); and backhaul link systems as well as real-time near-field imaging. Papers showcase advances in voltage-controlled oscillators (VCOs) and wideband mm-wave local-oscillator (LO)-generation blocks for 5G networks as well as in digitally intensive frequency synthesizers for communication and chirp-based radar-sensing applications.

Frequency Generation and Synthesis

In the area of VCOs, the trend is toward achieving low phase noise with architectural advances such as waveform-shaping LC tanks or multicore and multiple-resonator-based circuits. As illustrated in Figure 8, the phase noise of oscillators tends to increase at higher carrier frequencies. Achieving lower phase noise requires novel circuit topologies or spending more power. ISSCC 2018 presents a low-voltage inverse class-F VCO with a 196.2-dBc/Hz FOM and a 15-GHz BJT/CMOS (BiCMOS) VCO exhibiting –124-dBc/Hz phase noise at -MHz offset.

Wideband mm-wave LO-generation building blocks become essential for next-generation broadband communications systems. At ISSCC 2018, injection-locked frequency-multiplier-based techniques are demonstrated to support multiple bands in the mm-wave region. At subterahertz frequencies, stable sources are critical for imaging systems. ISSCC 2018 will present a BiCMOS 300-GHz oscillator with a very stable frequency output with no need for a phase-lock loop (PLL) or an off-chip crystal. The oscillator provides a peak output power of –13.9 dBm from 302 to 332 GHz.

In the field of frequency synthesizers, ISSCC 2018 will highlight advances in digital architectures enabling compact-area, low-power, low-noise implementations. A 653-μW fractional-N all-digital PLL for IoE applications will be presented. By leveraging an isolated constant-slope digital-to-time converter (DTC), it achieves an integrated jitter of 0.53 ps at 2.44-GHz output with a power consumption of 0.98 mW, corresponding to an FOM of –246 dB. Another development presented is a frequency synthesizer comprising an all-digital frequency-locked loop and a type-I PLL with an LC VCO occupying only 0.01 mm^2 in 65-nm CMOS, achieving a –254-dB jitter FOM from 4.6 to 5.6-GHz output.

An important trend is a rising interest in frequency synthesizers for fully integrated chirp-modulation-based radar systems. At ISSCC 2018, two papers will address this topic: one shows a 23-GHz digital bang–bang PLL for fast triangular and sawtooth chirp

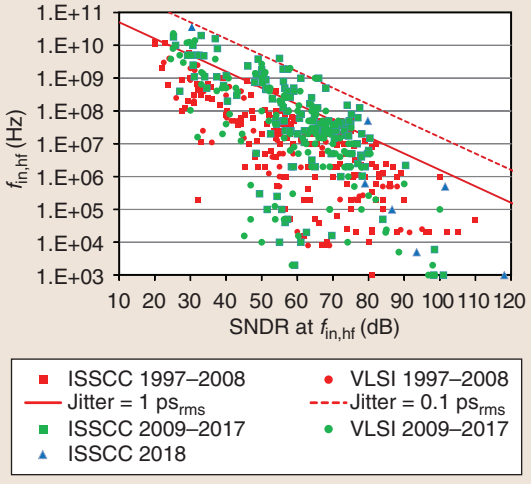

FIGURE 7: Bandwidth versus SNDR.

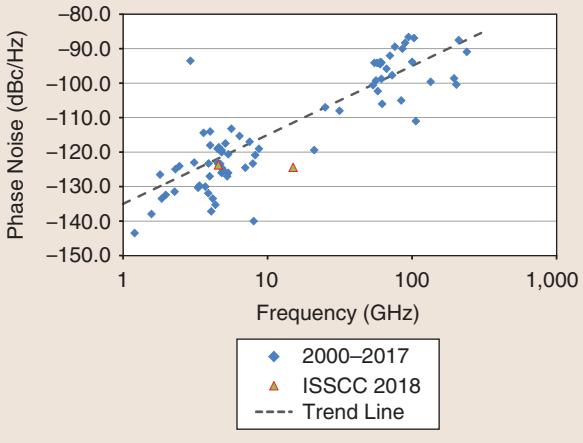

FIGURE 8: LC oscillator phase noise at 1-MHz offset versus carrier frequency.

modulation, and the other presents a 36.3–38.2-GHz fractional-N frequency-modulated continuous-wave chirp synthesizer with a bandpass delta-sigma DTC.

RF Techniques for Communication and Sensing

At ISSCC 2018, several papers will showcase techniques for integrating passive front-end filtering on an IC. A 3.4-dB noise figure, full-duplex hybrid coupler circulator with 40-dB transmitter self-interference suppression will be presented. In addition, an N-path-filtering-based 13th-order reconfigurable bandpass filter for surface-acoustic-waveless receivers will be described. In the area of terahertz imaging, a 0.56-THz sensing array for real-time near-field imaging in 0.13-μm silicon-germanium BiCMOS will be demonstrated.

Improving Output Power and Efficiency of mm-Wave Power Amplifiers

The output power and power efficiency of silicon-based power amplifiers (PAs) have been limited, due to their lossy output matching and combining networks. This year at ISSCC 2018, several mm-wave transmitters/PAs will be presented that are remarkable in both peak output power and efficiency, as plotted in Figure 9. In addition, a mm-wave radiator, with an on-chip antenna array targeting 77-GHz radar, achieves over 27-dBm output power in 65-nm CMOS. The codesign of a strongly coupled slot-antenna array and associated PAs effectively reduces the PA load impedance, achieving efficient spatial power sharing without significant distortion of the desired radiation pattern.

Harmonic impedance tuning is a key to maximizing peak-power efficiency. A 5-GHz CMOS PA in 65-nm CMOS achieves a record peak power-added efficiency (PAE) of 41% with a harmonically tuned class-F load network. In addition, the PA is made highly linear thanks to a novel transformer-based amplitude-/phase-modulation correction technique. Another 5G PA realizes wideband impedance matching at both fundamental and harmonic frequencies. It achieves 43.5% peak PAE and 19–29.5-GHz wide P_{out} 1-dB bandwidth.

Ubiquitous smart devices lie at the heart of a revolution shaping how we connect, collaborate, build relationships, and share information.

Communication Systems: Wireless

Subcommittee Chair: Stefano Pellerano, Intel, Hillsboro, Oregon

The insatiable need for big data communication calls for transceivers with high throughput achieved by higher digital modulation, carrier aggregation, MIMO with beamforming or beamsteering, and interference detection and cancellation. The crowded cellular spectrum demands flexible transceivers that exhibit very good linearity. This year at ISSCC 2018, a 14-nm mobile handset cellular transmitter with third-order counterintermodulation of −62.6 dBc will be demonstrated. Moreover, a universal 400-MHz–6-GHz full-band-capture macro base station transceiver will be presented that enables carriers to deploy a common IC solution in virtually all of the 2G/3G/4G/5G bands.

Scalable mm-wave phased arrays, with >100 elements, capable of high-resolution beamsteering, have been developed for >18-Gb/s point-to-point links. A further increase in network capacity by exploiting polarization-diversity has been demonstrated through integrated dual-polarization mm-wave MIMO and dual-polarization full-duplex links. Applications for integrated mm-wave radios include automotive radar. For this role, a fully integrated multi-element radar transceiver has been developed in CMOS, supporting beamforming and MIMO radar.

In the IoE space, Bluetooth low-energy (BLE) transceivers continue to achieve lower power consumption and demonstrate new features. These include an energy-harvesting BLE transmitter operating from just 0.2 V in 28-nm CMOS, as well as a single-crystal BLE transceiver using a 32-kHz reference in a 16-nm fin field-effect transistor (FinFET). Wake-up receivers with under 10-nW power consumption will also be presented at ISSCC 2018. These can be used as an alternative

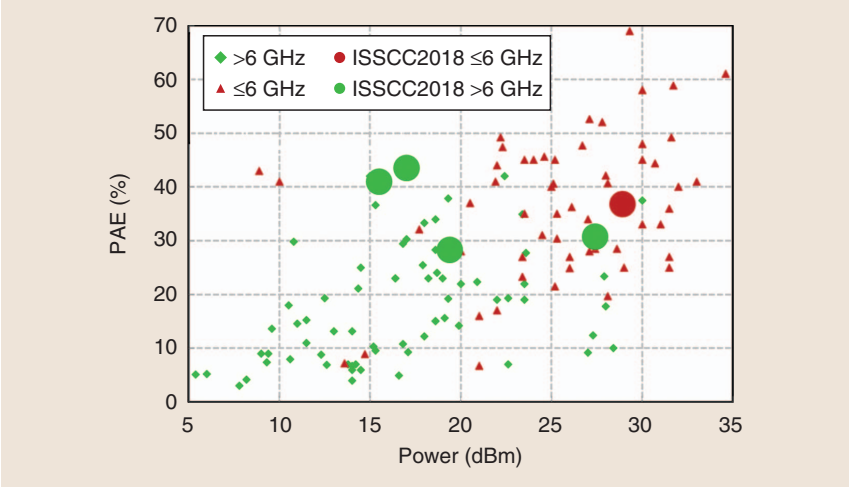

FIGURE 9: The peak PAE (%) versus maximum output power (dBm) for recent silicon-based PAs.

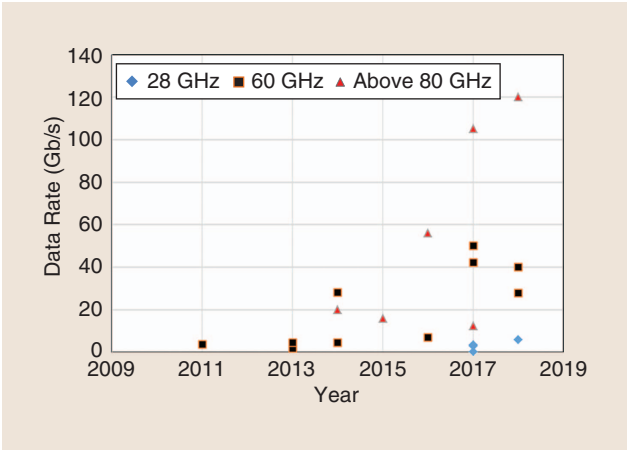

FIGURE 10: Wireless mm-wave data-rate trends.

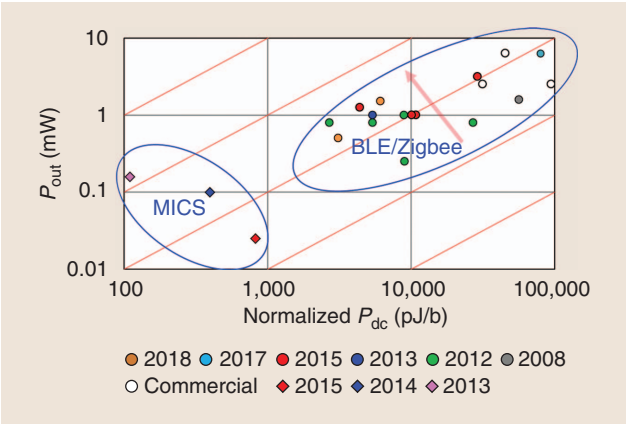

FIGURE 12: Ultralow-power 2.4-GHz and MICS-band wireless transmitter efficiency trends.

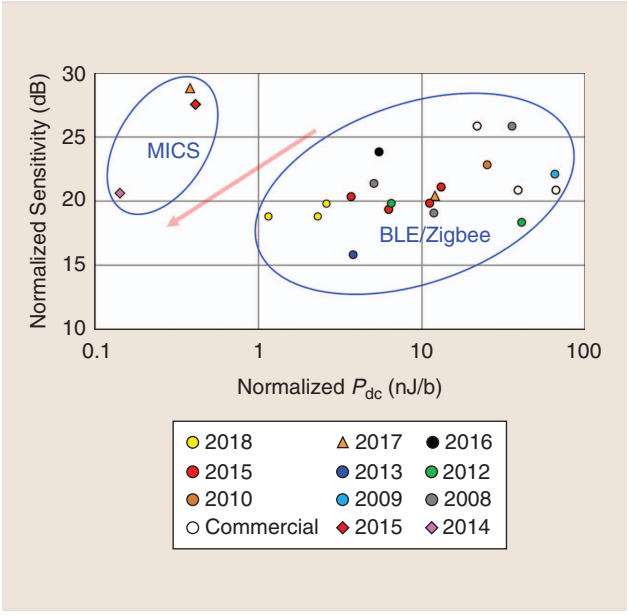

FIGURE 11: Ultralow-power 2.4-GHz and MICS-band wireless receiver sensitivity trends.

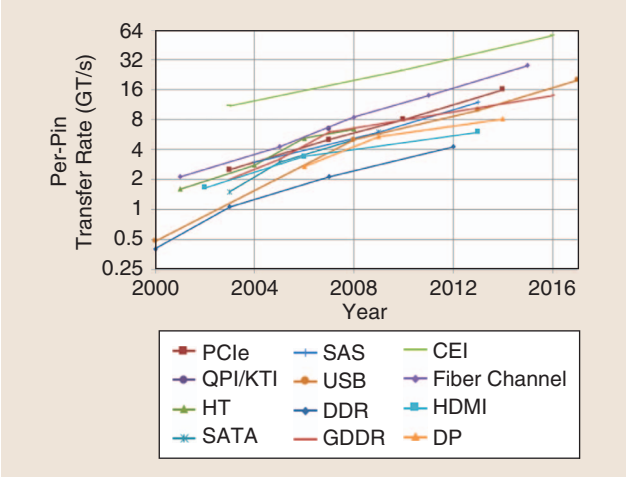

FIGURE 13: Per-pin data rate versus year for a variety of common I/O standards. PCIe: Peripheral Component Interconnect Express; QPI/KTI: QuickPath Interconnect/Keizer Technology Interconnect; HT: HyperTransport; SATA: Serial Advanced Technology Attachment; SAS: Serial Attached SCSI; USB: universal serial bus; GDDR: graphics DDR; CEI: common electrical I/O; HDMI: High-Definition Multimedia Interface; DP: DisplayPort.

to synchronization, enabling long battery life in wireless sensor nodes. For higher-data-rate applications, frequency-dependent in-phase/quadrature calibration is used in a wireless local area network (LAN) to support IEEE 802.11ax.

Figure 10 shows mm-wave wireless-data-rate trends over the past ten years. Figure 11 shows ultralow-power 2.4-GHz and Medical Implant Communication Service (MICS)-band wireless receiver sensitivity trends, and Figure 12 shows ultralow-power 2.4-GHz and MICS-band wireless transmitter efficiency trends. Because many different techniques are used to design the circuits, there is significant scatter in the graphs. However, in all cases, the arrows show the desired trend directions.

Communication Systems: Wireline

Subcommittee Chair: Frank O'Mahony, Intel, Hillsboro, Oregon

Over the past decade, wireline input/output (I/O) has enabled the remarkable scaling of computer systems both large and small, ranging from supercomputers to handheld electronics. During this time, aggregate I/O bandwidth requirements have increased at a rate of approximately two-to-three times every two years. Demand for bandwidth has been driven by applications such as memory expansion, graphics, chip-to-chip fabric, backplane, rack-to-rack, and LANs.

In part, this demand has been met by expanding the number of I/O pins per component. As a result, I/O circuitry consumes an increasing amount of area and power on today's chips. However, increasing bandwidth has also been enabled by rapidly accelerating the per-pin data rate. Figure 13 shows that per-pin data rate has approximately doubled every four years across a diverse variety of I/O standards, ranging from double data rate (DDR) to graphics to high-speed Ethernet.

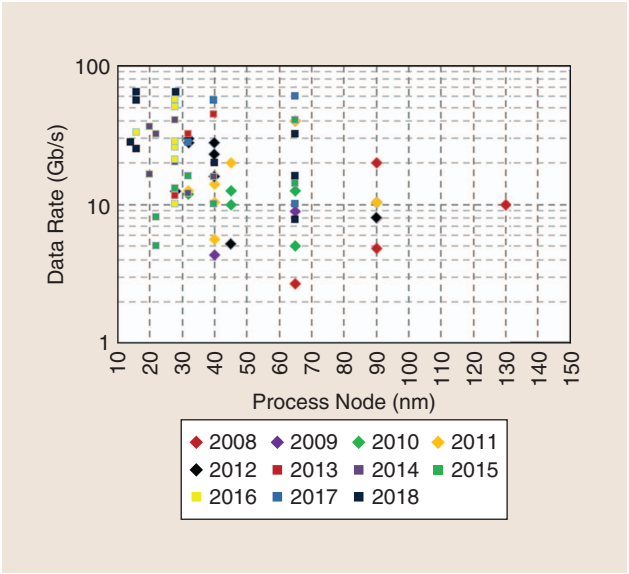

FIGURE 14: Data rate versus process node and year.

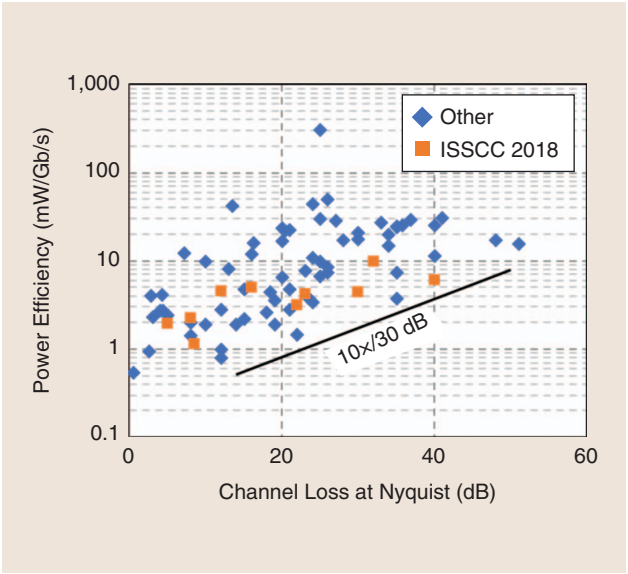

FIGURE 15: Transceiver power efficiency versus channel loss.

Figure 14 shows that the data rates for published transceivers have kept pace with these standards, enabled, in part, by process-technology scaling. However, continuing this remarkable I/O scaling trend will be thwarted by future difficulties in transistor scaling. Significant advances in both energy efficiency and signal integrity must be achieved to enable the next generation of low-power, high-performance computing systems. Papers at ISSCC 2018 include demonstrations of 112-Gb/s pulse–amplitude modulation with four-amplitude-level (PAM-4) transmitters in 10-nm and 14-nm CMOS as well as a power-efficient 56-Gb/s ADC-based transceiver in 16-nm CMOS.

Energy Efficiency and Interconnect Density

The power consumption of I/O circuits is a first-order design constraint for systems ranging from cell phones to servers. As the pin count and per-pin data rate for I/O have increased, so has the percentage of total power consumed by I/O. Technology scaling has enabled increased clock and data rates and offers improved energy efficiency, especially for digital components. However, simply increasing per-pin baseband data rates with existing circuit architectures and channels is not always a viable path, given fixed system-power limits. Figure 15 plots the energy efficiency (expressed in mW/Gb/s, which is equivalent to pJ/b) as a function of channel loss for recently reported transceivers. Looking at the most aggressive designs, the data indicate that the scaling factor between link power and signaling loss is slightly less than unity—and, in particular, that a 30-dB channel loss corresponds to a roughly ten times increase in pJ/b. Many recent advances have reduced power for high-speed link components through circuit innovation: these include low-power receiver equalization [decision feedback equalization (DFE) and continuous-time linear equalization (CTLE)], CMOS resonant clocking, low-swing voltage-mode transmission, and links with low-latency power-saving states. At ISSCC 2018, a 112-Gb/s PAM-4 transmitter in 10-nm CMOS that consumes only 1.16 pJ/b will be demonstrated. Also presented will be an 8-b DAC-based 112-Gb/s PAM-4 transmitter with a power efficiency of 2.55 pJ/b.

Electrical Interconnect

There has been increasing demand for very-high-data-rate communication across a wide variety of channels. Some types of channels, especially those related to medium-distance electrical I/O (such as server backplanes), must support high data rates over high-loss channels. For these links, the key to scaling has been improvements in equalization and clock/data recovery.

Recent high-speed transceivers use a combination of fully adaptive equalization methods, including transmit finite-impulse response (FIR), receive CTLE, and DFE, as well as receive FIR and/or infinite-impulse response FIR filter efficiency. As a result, recent receivers achieve data rates above 28 Gb/s across channels with up to 50 dB of loss.

ISSCC 2018 will present examples of transceivers that are starting to extend the equalization range of 56-Gb/s wireline communication. One paper describes a 28.05-Gb/s multistandard transceiver that compensates 40-dB loss with an energy efficiency of 6 pJ/b. Another paper presents a fully adaptive 19–56-Gb/s transceiver with a 3–7-b reconfigurable ADC that achieves power efficiency of 6.4 pJ/b over a 7.4-dB channel.

Optical Interconnect

As the demand for higher bandwidth has accelerated and electrical channel impairments have become increasingly severe with rising per-lane data rates, optical interconnects have become an increasingly attractive alternative to traditional electrical wireline interconnects. Optical communication has clear benefits for high-speed,

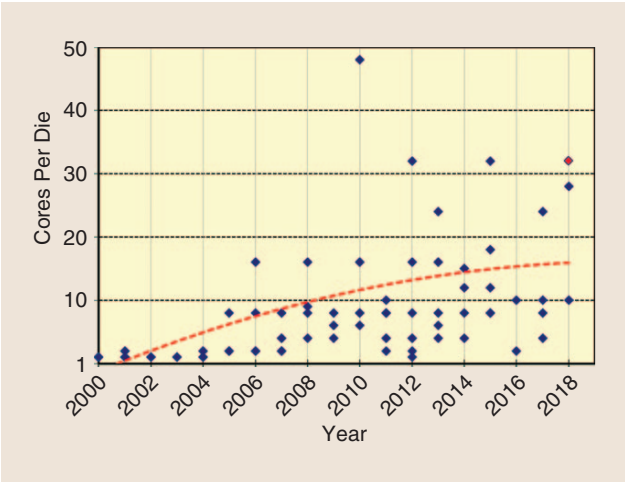

FIGURE 16: Core-count trends. (The red diamond designates an MCM.)

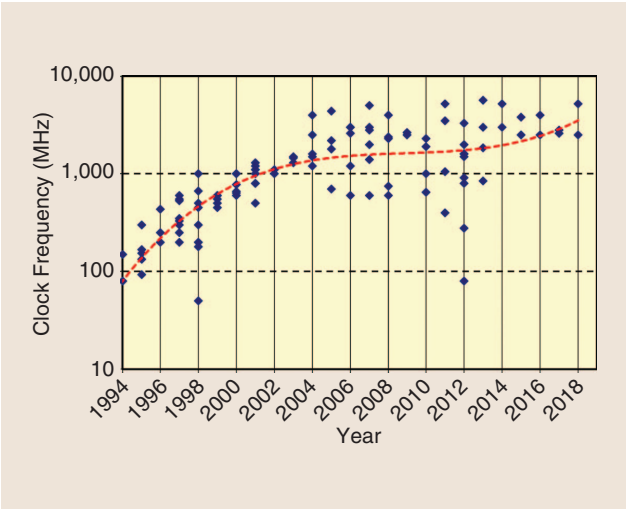

FIGURE 17: Clock-frequency scaling trends.

long-distance interconnects because it offers lower channel loss. Circuit-design and packaging techniques that have traditionally been used for electrical wireline are being adapted to enable integrated optical links with extremely low power. This has resulted in rapid progress in optical ICs for Ethernet, backplane, and chip-to-chip optical communication. At ISSCC 2018, a 56-Gb/s 128-mW burst-mode optical receiver with a rapid wake-up time of 6.8 ns will be presented.

Wireline Retrospective

Continuing to aggressively scale I/O bandwidth is essential but extremely challenging. Thus, innovations that provide higher performance and lower power must continue to be made to sustain this trend. Advances in circuit architectures, interconnect topologies, and transistor scaling are, together, changing how I/O will be performed during the next decade. The most exciting and promising of these emerging technologies for wireline I/O will be highlighted at ISSCC 2018.

Digital Systems: Digital Architectures and Systems

Subcommittee Chair: Byeong-Gyu Nam, Chungnam National University, Daejeon, South Korea

The dominant trend in high-performance central processing units (CPUs) is the increasing number of processor cores per die (Figure 16) and the amount of overall chip memory. For example, at ISSCC 2018, Intel will describe a server-class CPU integrating 28 cores using 14-nm tri-gate process technology. The processor has a two-dimensional mesh on-die interconnect fabric serving the cores.

Furthermore, multichip modules (MCMs) are emerging as a method to continue chip complexity growth, even while traditional process scaling slows down. For example, at ISSCC 2018, AMD will describe a system-on-chip (SoC) architected as a chiplet incorporating 8×86 cores, of which one to four would be configured on an MCM to service multiple markets including servers and both mainstream and high-end desktops. Utilizing a 14-nm FinFET process technology, this chiplet SoC incorporates over 4.8-B transistors on a 213-mm^2 die.

Meanwhile, clock-frequency growth has leveled off (Figure 17) and is no longer systematically increasing, which will require more innovation at the architectural level to maintain performance-growth requirements (Figures 18 and 19). For example, at ISSCC 2018, IBM will describe a z14 processor having 50% more L2 cache, two times the amount of L3 cache, and 25% more cores, running at 5.2 GHz (which is 200-MHz faster than previous designs) while maintaining the same power envelope. It also features significant microarchitectural updates for branch prediction, cache management, and cryptography.

The computational performance of mobile application processor (AP) SoCs has historically grown as silicon process technology advances (Figure 20). This year at ISSCC 2018, the leading SoCs remain in 10-nm process technology and, hence, do not show significant improvement in performance. Restricted by the slowdown in process-technology advancement, the maximum configuration shown has eight heterogeneous cores (similar to last year).

AP SoCs continue to add more features, including multimedia intellectual property cores and on-chip accelerators to enhance functionality (Figure 21). Programmable neural-network processors have appeared in AP SoCs to enhance machine-learning application capabilities. Dedicated neural-network accelerators execute machine-learning functions faster and with higher energy efficiency than generic CPUs and GPUs. Neural-network processor units (NNPUs) will be adopted in most future AP SoCs and thus establish an important future direction, consisting of an efficient software solution that utilizes heterogeneous computing combining CPUs, GPUs, and NNPUs.

Wired and wireless links continue to increase in bandwidth, trending toward a ten times increase in data rate

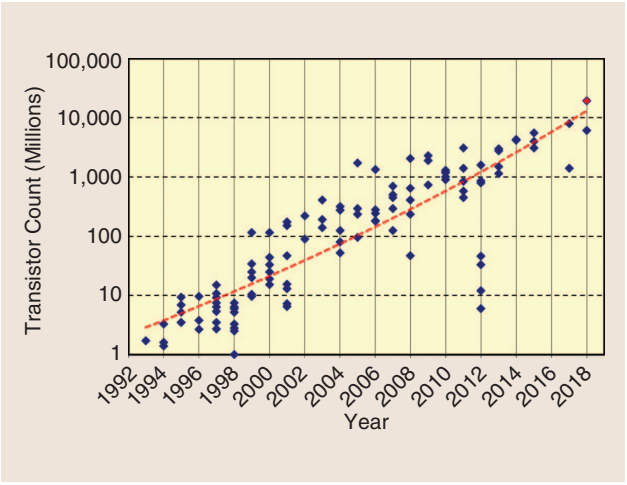

FIGURE 18: Chip-complexity scaling trends. (The red diamond designates an MCM.)

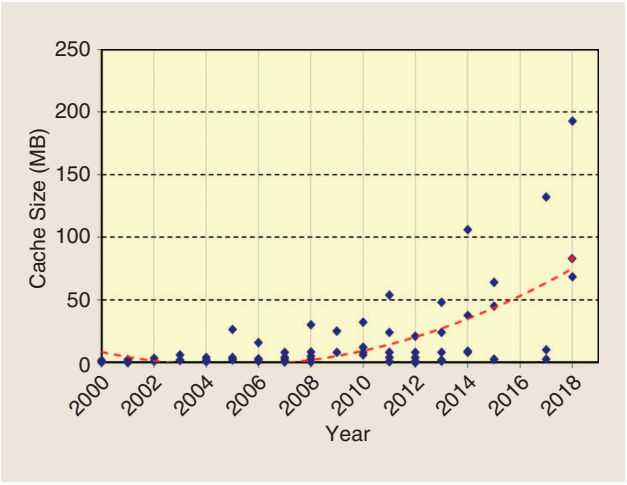

FIGURE 19: On-die cache-size trends. (The red diamond designates an MCM.)

every five years (Figure 22). Changes are modest this year relative to last year; however, the IEEE 802.11ax standard is on the near horizon. Both mm-wave and massive MIMO technologies are being actively studied to realize 5G communication. The explosion of IoE devices will require the evolution of narrowband wide-area networks.

Deep learning is a rapidly evolving topic at ISSCC 2018. The computational complexity of typical deep neural networks (DNNs) impedes their execution on resource-scarce mobile or wearable devices. Last year, several innovative solutions were introduced to enhance throughput and improve energy efficiency, mostly focusing on the efficiency of convolutional neural networks (CNNs). The current state of the art still faces two significant challenges: 1) a need to improve energy efficiency for ultralow-power applications and 2) solutions for efficient execution of fully connected nonconvolutional networks.

To realize improved energy efficiency, there is a trend toward reduced-precision networks, with binary networks as the extreme case (recently, the first binary neural-network accelerator appeared). This year, ISSCC 2018 pushes peak efficiency from several tens of trillions of operations per second (TOPS)/W (digital accelerators) to beyond hundreds of TOPS/W for a mixed-signal implementation. Several papers at ISSCC 2018 treat the energy efficiency of fully connected network acceleration. In such networks, the bottleneck is the memory load/stores, and, as such, innovative solutions include the fabrication of a three-dimensional (3-D) stack of processing and memory dies, as well as smart memory interfaces enhancing data reuse. These innovations clearly bring DNNs within reach of battery-operated devices.

Figure 23 illustrates deep-learning processor efficiency and throughput improvements for DNNs and CNNs presented at ISSCC 2018, as compared to the state of the art in 2017.

Digital Systems: Digital Circuits

Subcommittee Chair: Edith Beigne, CEA-LETI, Grenoble, France

The demand for higher performance across ubiquitously connected energy-constrained platforms ranging from the IoE to cloud data centers continues to drive innovations in all CMOS digital-circuit building blocks, with goals of improving energy-efficient

FIGURE 20: Clock-frequency trends for mobile CPUs.

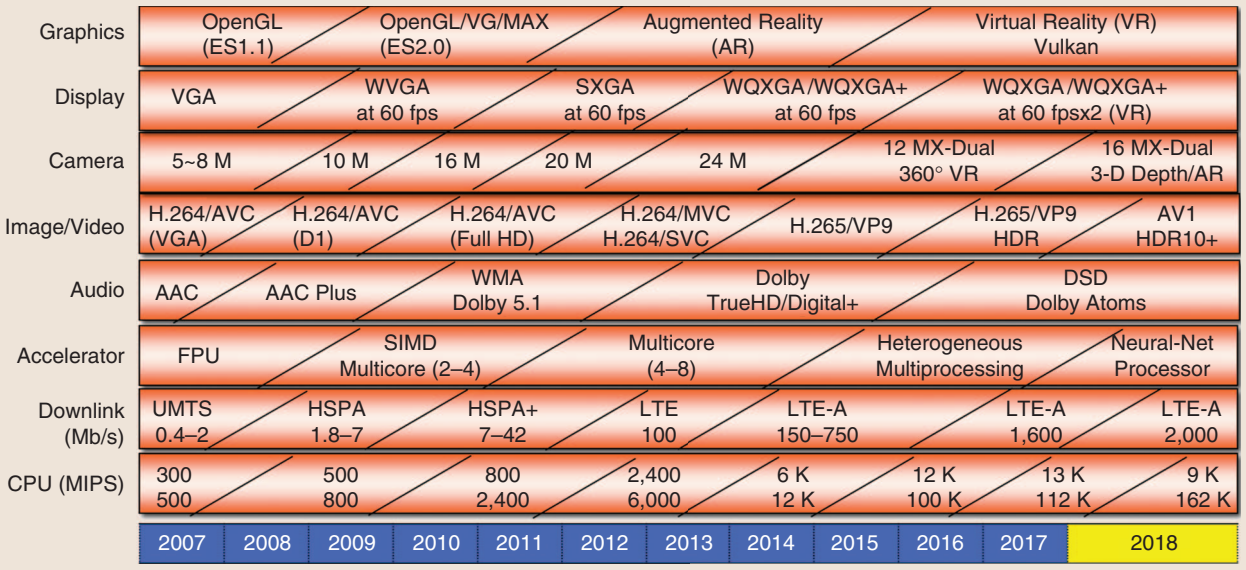

FIGURE 21: AP trends in smart phones. OpenGL: Open Graphics Library; VGA: video graphics array; WVGA: wide VGA; SXGA: superextended GA; WQXGA: wide quad extended GA; AVC: advanced video coding; HD: high definition; MVC: multiview video coding; SVC: scalable video coding; HDR: high dynamic range; AV1: AOMedia Video 1; AAC: advanced audio coding; WMA: Windows Media Audio; DSD: Direct-Stream Digital; FPU: floating-point unit; SIMD: single instruction/multiple data; UMTS: Universal Mobile Telecommunications System; HSPA: High-Speed Packet Access. MIPS: millions of instructions per second.

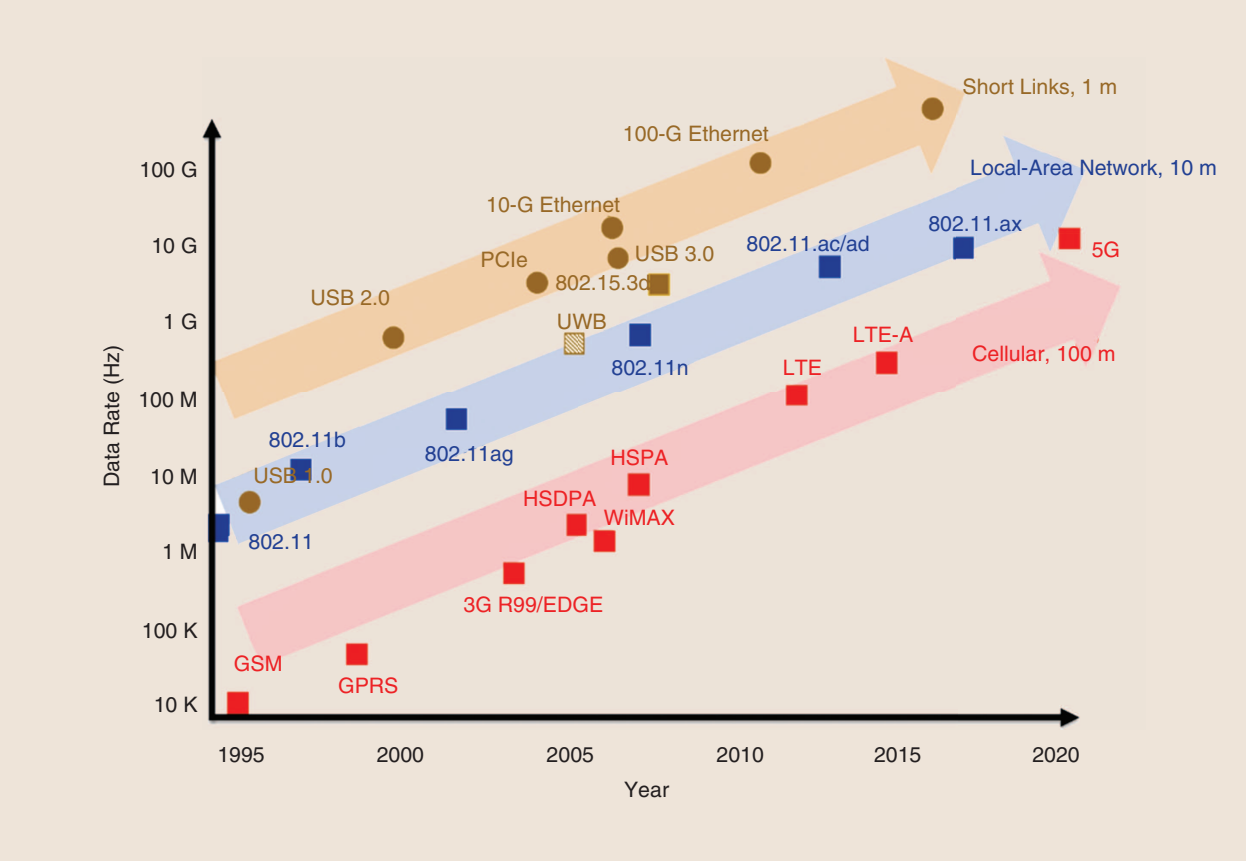

FIGURE 22: Data-rate trends for wired, wireless, and cellular communication. HSPA: High-Speed Packet Access; HSDPA: High-Speed Downlink Packet Access; GSM: Global System for Mobile Communications; GPRS: General Packet Radio Service; EDGE: Enhanced Data Rates for GSM Evolution: LTE: Long-Term Evolution; LTE-A: LTE-Advanced.

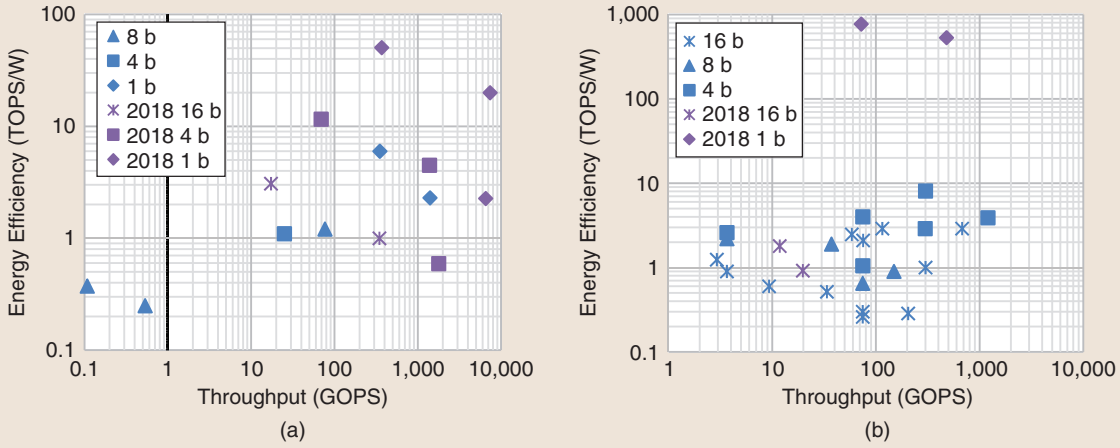

FIGURE 23: Deep-learning processor throughput and efficiency: (a) DNN and (b) CNN performance. GOPS: giga operations per second.

performance, lowering cost/design effort, and enhancing security. Classic technology scaling has slowed, and circuit-design efforts are exploiting technology features, such as body biasing and passive-device advances, to enable circuit innovation.

Energy-Efficiency Techniques and Integrated Voltage Regulators

Energy reduction remains a top priority as power density continues to grow. While traditionally off-chip, voltage regulators have increasingly been integrated on-chip to reduce cost. Low-dropout (LDO) linear regulators, switched-capacitor voltage regulators (SCVRs), and now, inductor-based buck voltage regulators (LCVRs) are integrated into scaled process nodes to enable faster and fine-grained dynamic voltage and frequency scaling (DVFS) of individual functional blocks. In turn, the low voltages supported in DVFS systems necessitate a move from analog-based LDOs to digital implementations.

Figure 24 shows the conversion efficiency and current density of these integrated voltage regulators, which continue to improve. Variation mitigation has also become a major trend in digital circuits to improve robustness and power efficiency across process, voltage, and temperature. Specific all-digital sensors and adaptive techniques are currently being proposed to mitigate these effects on-chip.

Synthesizable Digital PLLs for Low-Jitter Applications

Trends for PLLs include migration from analog to digital to include more functionality, cope with variability, and ease scaling to finer geometries. Demand for compact low-jitter PLLs is increasing. The use of more automated digital-design flows (such as synthesis and autoplace-and-route) dramatically reduces development costs but can degrade jitter, requiring new techniques to compensate. Figure 25 highlights metrics for PLLs and digital PLLs published at ISSCC over the past ten years (up to ISSCC 2018); the plot shows the relationship between reference (input) frequency and FOM, demonstrating the tradeoff between cost (higher reference frequency) and overall PLL performance.

Circuits for Hardware Security

With the increasing risk and cost of information theft, hardware-implemented security has become a common circuit component. Although focus on cryptographic implementation continues, cost-effective physically unclonable functions (PUFs) are now a focus area, such as in smart cards and consumer devices. True-random-number generators are also commonly leveraged to strengthen secret key

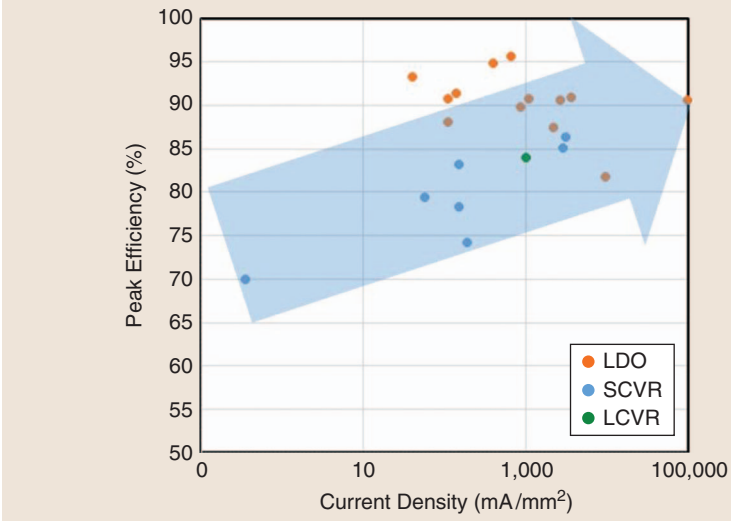

FIGURE 24: Integrated voltage regulators.

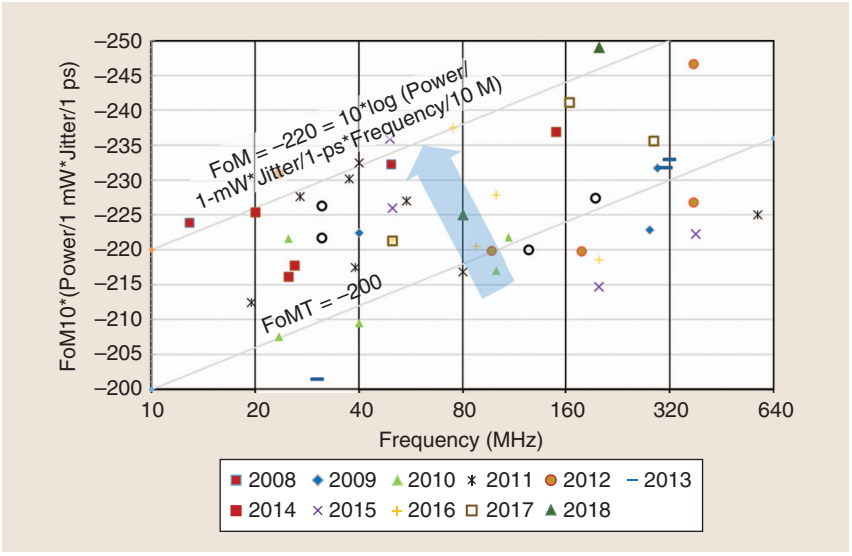

FIGURE 25: PLL and multiplying delay-clocked loop trends.

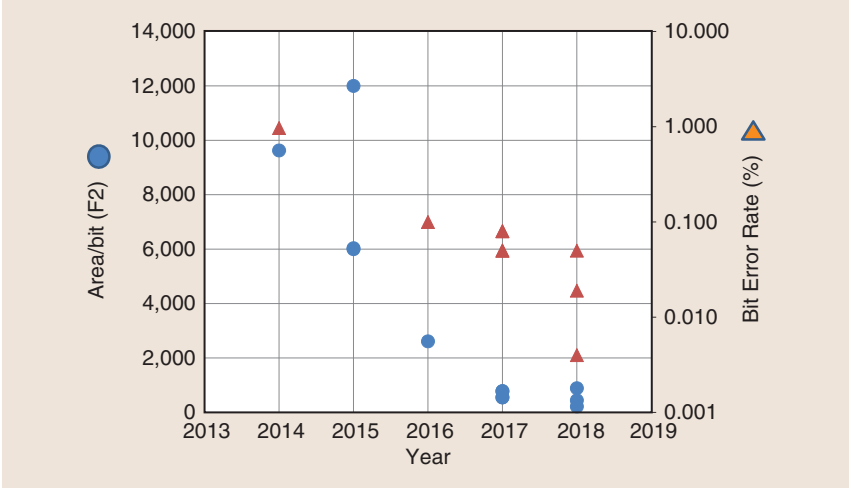

FIGURE 26: Area/bit and bit-error-rate trends for PUFs.

generation in cryptographic applications. Figure 26 illustrates trends in area/bit scaling and bit-error rates for PUFs published at ISSCC from 2014 to 2018.

Memory

Subcommittee Chair: Leland Chang, IBM, Yorktown Heights, New York

The demand for high-density, high-bandwidth, and low-energy memory systems continues in applications from high-performance computing to SoCs to wearables and the IoE. In embedded memory, record bit/cell-size static random-access memory (SRAM) in both 7-nm and 10-nm high-array-density SRAM have proved to be functional at the megabyte level. Dynamic RAM (DRAM) high-performance memory-interface technologies are enhanced for GDDR6 and high-bandwidth memory (HBM); meanwhile, record density DRAMs are implemented in 1xnm processes.

NAND Flash memory has extended capacity using both bit/cell techniques and number of layers in 3-D structures. A new era of compute-in-memory (CIM) for neural networks is emerging using both SRAM and resistive RAM (RRAM). In the meantime, spin-transfer-torque RAM (STTRAM) has progressed to operate at up to 120 °C with a one-transistor/one-magnetic-tunnel-junction (1T1MTJ) implementation.

Some current state-of-the-art implementations from ISSCC 2018 include the following:

- a 0.031-μm^2 and a 0.026-μm^2 SRAM bit cell in 10 nm and 7 nm, respectively, shown to be functional at a megabyte array level
- a 1-Mb 28-nm 1T1MTJ STTRAM array with 3.6-ns read access time at 120 °C, utilizing an offset-canceled sense amplifier and in situ write termination
- a 16-Gb GDDR6 implemented to operate at 18 Gb/s/pin with a per-bit-trainable single-ended DFE, ZQ-coded transmitter, and PLL-less clocking
- a 16-Gb low-power DDR (LPDDR4X) implemented in a 10-nm process with 81-mm^2 die size using in-DRAM error-correcting code, which achieves a data rate of 5 Gb/s/pin and self-refresh power of 0.1 mW/Gb
- a 1-Tb NAND Flash memory in 64 stacked layers using 4-b/cell technology with 5.63-Gb/mm^2 areal density
- a 512-Gb 3-b/cell 3-D Flash memory on a 96-word-line-layer technology with a string-based start bias-control scheme that achieves 7% shorter program time
- multiple compute in memory implementations for neural networks using SRAM and RRAM.

SRAM

For embedded high-speed applications, SRAM continues to be the memory of choice—from mobile to high-performance servers to IoE and CIM for neural networks. This year, ISSCC 2018 highlights 7-nm FinFET bit cells operating at 6 GHz and a 10-nm SRAM showing 23.6-Mb/mm^2 array density. SRAM arrays are also implemented for product-sum computation in memory operations. Figure 27 shows SRAM bit/cell area and V_{MIN} scaling trends.

High-Bandwidth DRAM

To maintain the optimal memory hierarchy ratio with respect to storage memory, DRAM continues to scale density, form factor, and bandwidth. This year, ISSCC 2018 presents the latest interface-standard benchmarks, including GDDR6, HBM for high-bandwidth, LPDDR4X for low-power applications, and DDR4 for computing with high density. Figure 28 shows DRAM bandwidth scaling over the past decade.

Nonvolatile Memories

Over the past decade, significant investment has been made in the emerging memories field to find an alternative to floating-gate-based nonvolatile memory (NVM). The emerging NVMs, such as phase-change memory (PRAM), ferroelectric RAM (FeRAM), STT magnetic RAM (STT-MRAM), and RRAM, are showing potential for achieving high cycling capability and lower power per bit in read/write operations. Figure 29 highlights 3-b/cell (TLC) NAND Flash and 4-b/cell (QLC) NAND Flash write throughput. Figure 30 shows a significant increase in NAND Flash capacity from 768 Gb to 1 Tb at ISSCC 2018. Such high areal densities are achieved through advances in 3-D vertical bit/cell-stacking technologies.

NAND Flash Memory

NAND Flash memory continues to advance toward higher density and lower power, resulting in low-cost storage solutions that are replacing traditional hard-disk storage with solid-state disks. In mass-production semiconductor industries, 3-D memory technology has been the mainstream for NAND Flash memory. At ISSCC 2018, both a 1-Tb 4-b/cell that achieves 5.63-mm^2/Gb areal density and a 512-Gb 3-b/cell 3-D NAND with 96 stacked word-line layers will be presented, continuing the trend toward satisfying the ever-growing demand for increased density requirements and lower manufacturing costs. Recent NAND developments at ISSCC 2018 show not only higher density but also higher performance (over 57-MB/s program throughput and 1.0-GB/s read throughput), all at lower voltages (1.8–1.2 V). Figure 31 shows the observed trend in NAND Flash capacities at ISSCC over the past 15 years.

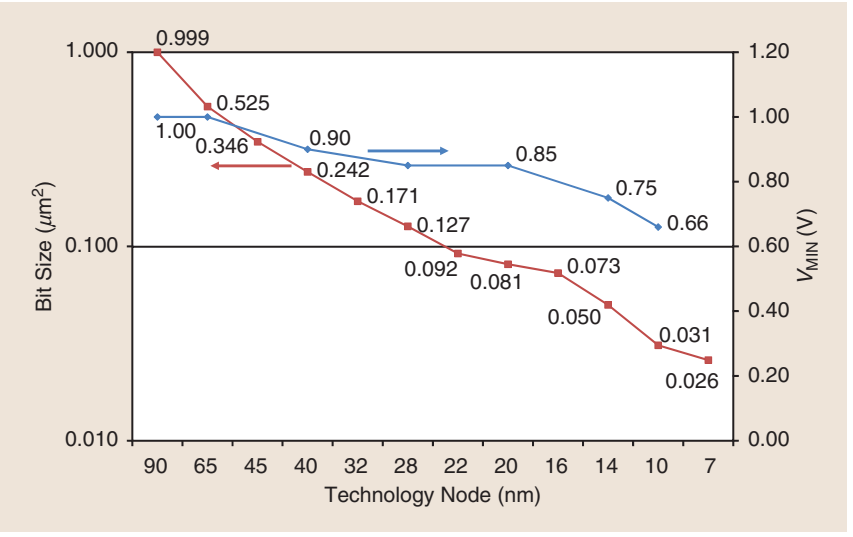

FIGURE 27: Bit-cell and V_{MIN} scaling trends for SRAM.

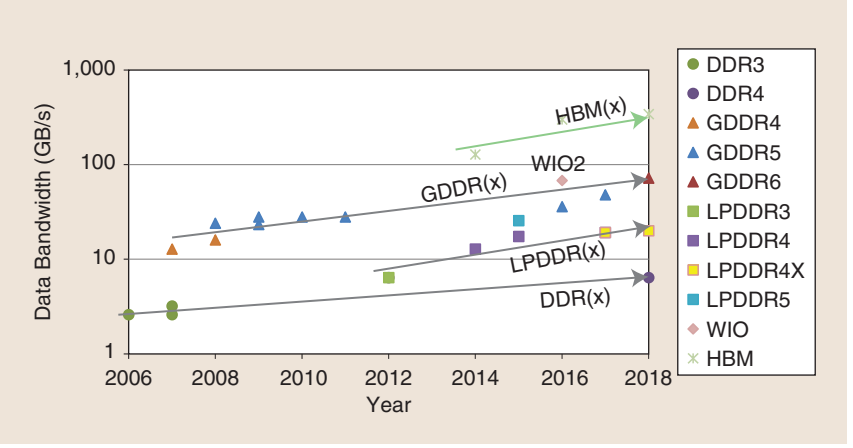

FIGURE 28: DRAM data-bandwidth trends. WIO: wide I/O.

Innovative Topics: Imagers

Subcommittee Chair: Makoto Ikeda, University of Tokyo, Japan

CMOS image sensors remain one of the fastest-growing segments of the semiconductor industry, with double-digit compound annual growth rates. Image sensors are required components in most mobile devices, which now include multiple front- and rear-facing cameras. The other applications that continue to drive the demand for image sensors include autonomous driving, smart security, wearables, gaming, VR, AR, the IoE, and biomedicine.

Because they enable increasing on-chip functionality without detracting from image quality, 3-D stacked image sensors have become more popular. At ISSCC 2018, six out of the ten papers on image sensors use a 3-D stacking process to preserve most of the top-layer area for light sensing, while keeping the readout and image-signal-processing circuits on the bottom layer. In one paper, a 14-b ADC is stacked under each pixel using a copper–copper hybrid bonding technology.

As seen at previous ISSCCs, the race to higher resolution and smaller pixels has slowed slightly but never stopped. At ISSCC 2018, a 0.9-μm pixel with a full-depth deep-trench isolation that reaches a performance level exceeding the current-generation 1.0-μm pixel in both the saturation level and readout noise will be presented. The broadcasting industry continues to push the array size toward, and beyond, the 8-k/4-k or 33-Mpixel resolution, with frame rates exceeding 120 f/s. At ISSCC 2018, a 33-Mpixel, 480 f/s CMOS image sensor will be presented.

We observe continued development of depth sensors, using either direct or indirect time-of-flight technology,

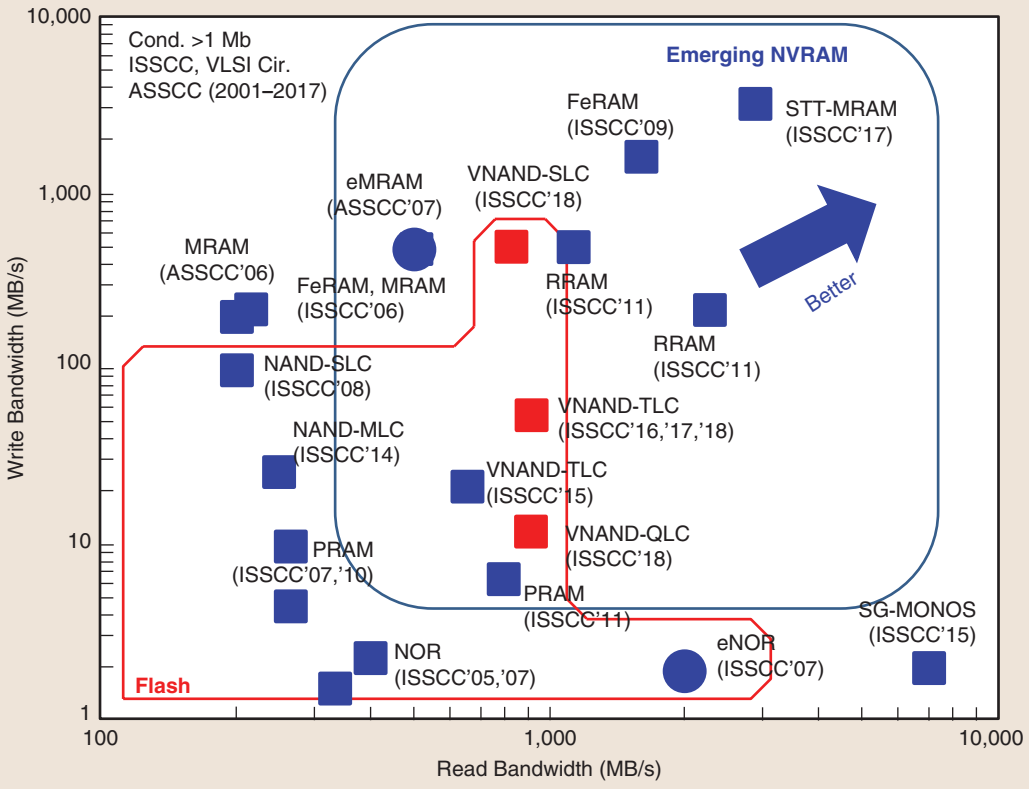

FIGURE 29: A read/write bandwidth comparison of NVMs. VNAND: vertical NAND; SLC: single-level cell; MLC: multilevel cell; SG-MONOS: split-gate metal–oxide–nitride–oxide–silicon.

to address emerging applications in VR, AR, and light detection and ranging (i.e., LiDAR) for autonomous driving. Four of the ten papers on image sensors at ISSCC 2018 will describe depth-sensing applications. Back-side illuminated and 3-D wafer stacking are now being used. Single-photon avalanche diodes and silicon photodiodes are being employed for direct time-of-flight measurement with extended range-imaging up to 200 m. Pixel resolution continues to scale for depth sensing beyond 1 Mpixel.

Innovative Topics: Medical

Subcommittee Chair: Makoto Ikeda, University of Tokyo, Japan

As illustrated at ISSCC 2018, both sensor and actuator systems for in-body implantable use continue to evolve toward more robust and energy-efficient operation. New circuit concepts permit recording of weak biopotential signals in the presence of real-life interference and under stringent power and size constraints, paving the way for autonomous robust systems that combine sensing and actuation as part of a single implantable device. This will allow therapy to be applied directly in a closed-loop fashion for the treatment of a variety of neurological disorders. Furthermore, technological approaches for multimodal very-high-density cellular interfacing continue to advance, with new applications in high-throughput drug screening.

The state of the art in biomedical ICs and systems is further advanced this year at ISSCC 2018, with a larger number of recording channels in neural interfaces (>16 K), integration of circuits for combined optogenetics and neurophysiology, increased dynamic range (>90 dB), increased common-mode and power-supply rejection (>100 dB), and improved noise/power efficiency (PEF <3).

Innovative Topics: Sensors

Subcommittee Chair: Makoto Ikeda, University of Tokyo, Japan

Sensors are a key building block for a wide variety of applications. They collect data and are an important part of the value chains that are enabling new features. Increasingly, processing is performed at the sensor interface, resulting in ever-more optimized systems.

MEMS inertial sensors (accelerometers and gyroscopes) are key components used in a wide variety of consumer and automotive products, where power consumption and reliability are primary requirements. A new frequency-modulated gyroscope employs rate chopping to reduce sensor offset drift. In addition, the fusion of pressure and inertial sensors is proved to enable a personal navigation system for operations in environments without global positioning system access.

New circuit techniques allow beamforming directly at ultrasound interfaces

with reduced power consumption and area. This will enable miniature 3-D ultrasound probes.

In an effort to enlarge the application area of user-touch interfaces and improve noise immunity, natural handwriting using a stylus with a capacitive touch controller is growing in importance. New touch-sensing architectures are being investigated to integrate an active stylus and an electrically coupled resonance stylus, both having pressure sensitivity.

Innovative Topics: Technology Directions

Subcommittee Chair: Makoto Nagata, Kobe University, Japan

Technology innovations bring the promise of enabling completely new system functionalities or substantially greater efficiency of existing ones. It is significant that harnessing such innovations to solve real-world problems requires thinking about technologies in the context of systems. Thus, with the current focus on silicon engineering in the social world, trends in technology directions focus on biomedical systems and in-memory computing for machine-learning applications.

Biomedical Systems

The human body is an incredibly complex system, yet innovations in medical devices and systems are allowing us to dive more deeply into the dynamic inner workings of the body. Over the past decade, the trend in biomedical IC development has focused primarily on building sensors and actuators, along with their corresponding electronic instrumentation, that interact with the body in the electrical domain: for example, circuits used to measure electrocardiograms, electroencephalograms, electromyograms, and so on, as well as stimulation of neural tissues, have all been explored in detail at ISSCC. This trend is continuing at ISSCC 2018, but now the research focus is on enhancing system-level functionality, for example, through multisensor interfaces that also infer hemodynamic processes, ultraminiaturized single-channel implants co-integrated with optical features, or integration of in-sensor machine-learning processors to help classify and use data in closed-loop systems. Machine learning can, in some cases, be an impactful technique to reduce the amount of data that needs to be wirelessly transmitted, potentially reducing overall system power due to lower radio energy costs.

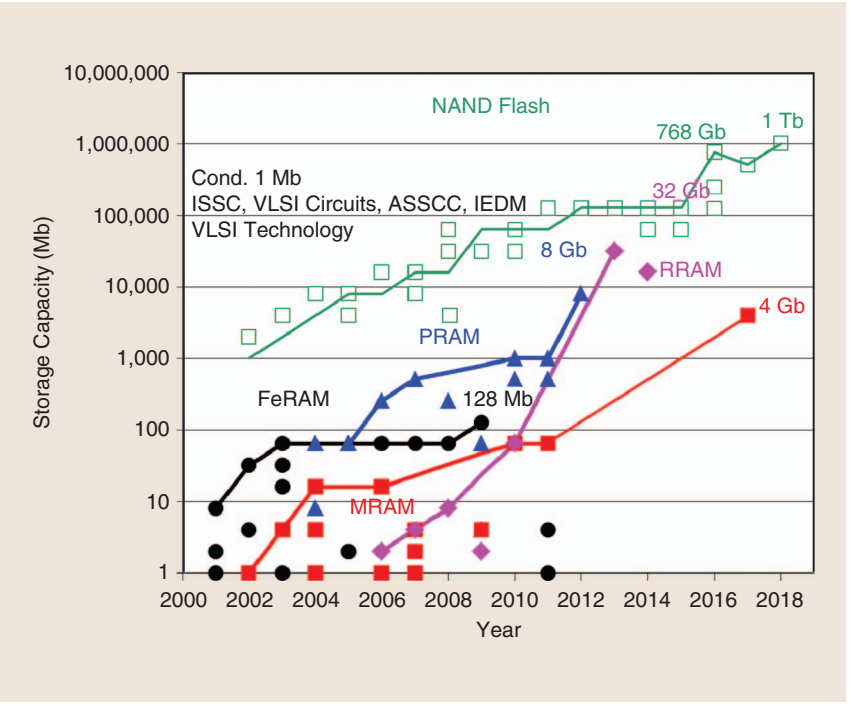

FIGURE 30: Memory-capacity trends for emerging NVMs. ASSCC: IEEE Asian Solid-State Circuits Conference; IEDM: International Electron Devices Meeting.

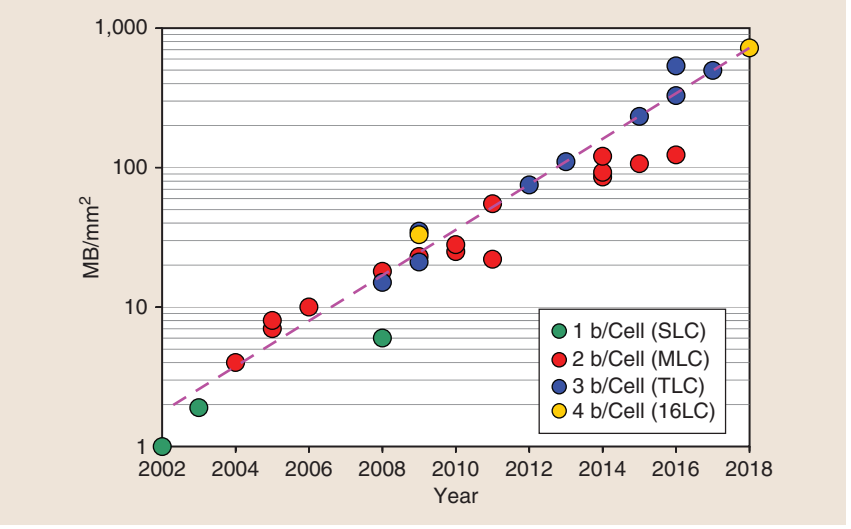

FIGURE 31: NAND Flash memory trends.

If in-sensor processing of data is infeasible (due to either energy or other application-level constraints), then wireless transmission of data must be extremely energy efficient to minimize the size of the energy source (such as a battery, energy harvester, and so on) and fit with anatomical size constraints. As a result, another trend in biomedical systems at ISSCC 2018 is the design of wireless communication

Another trend in biomedical systems at ISSCC 2018 is the design of wireless communication circuits that operate extremely efficiently in the presence of biological tissue.

circuits that operate extremely efficiently in the presence of biological tissue. Examples include innovations in body-coupled communication systems that operate in the gastrointestinal tract as part of a multicamera capsule endoscope and transcranial links that operate at up to 200 Mb/s.

Beyond sensing physical and electrophysiological parameters, an emerging trend in biomedical systems is to measure physiochemical parameters. Biosensors that respond to glucose, electrolytes, or gases can offer new views into the dynamic behavior of freely behaving biological subjects. In wearable and implantable applications, trends include enabling such functionality at nW-level power consumption or via bioenergy harvesting that exploits the energy naturally present in many of the metabolites being sensed. In very small systems, it may be possible to forgo conventional dc–dc converters if the voltage of the biofuel cell energy harvesters is sufficiently large or circuits are specifically designed to operate at low supplies (such as 0.3 V). In stationary applications, trends include increasing the resolution and dynamic range of correction circuitry to support long-term accurate operation.

Mixed-Signal and In-Memory Computing for Machine-Learning Applications

Machine-learning applications have ever-broader and more substantial impacts on increasing facets of society. In many cases, the applicability of such systems is limited by the energy they require and the performance they can achieve. Addressing these limitations has become a critical trend in technology directions. To overcome the limitations faced by traditional architectures, researchers in this area are taking advantage of innovations in unconventional architectures and emerging technologies. This has driven renewed interest in and perspectives on mixed-signal computation.

Unlike in previous years, when mixed-signal computation was primarily driven by its energy efficiency in the regime of low-resolution computation, now motivations at the architectural level are raising even greater promise for mixed-signal computation. This is particularly important in the area of memory access, which has posed one of the main limitations in current machine-learning systems and sensing: it represents a vital source of data on which machine-learning systems are being applied. Consequently, a key trend in technology directions has been mixed-signal computation in memory arrays and near sensors. This requires a confluence of circuits, architectures, and algorithms, opening new opportunities and perspectives through which integrated systems will impact society.

In line with these trends, ISSCC 2018 will feature two sessions representing the latest technological innovations in biomedical systems. One will include an invited presentation on the impact of sensors and electronic systems for food production and agricultural applications, while the other will emphasize in-memory computations for machine-learning applications.

Summary

According to the Semiconductor Industry Association, the industry generated US$338 billion in sales in 2016! In this environment, the ISSCC continues to be the premier technical forum for presenting advances and predicting trends in solid-state circuits and systems, as it has for the past 65 years. For more information, see the "The Origins of the ISSCC" by Dave Pricer on the early years of ISSCC in this issue.

Beyond this article, a complete trends document will be available at www.isscc.org. These trends are highlighted in papers that will be presented during the 65th ISSCC, held at the San Francisco Marriott Marquis on 11–15 February 2018. Attendance at ISSCC 2018 is expected to be around 3,000. Corporate attendees from the semiconductor and system industries typically represent about 60% of the attendees. We look forward to seeing you there!

Acknowledgments

We would like to acknowledge the ISSCC 2018 Technical Program Committee for providing original content for this article and each of the 11 subcommittee chairs (as referenced in each of the trends sections) for their leadership role. Without their collective efforts, this article would not have been possible.

About the Authors

Denis C. Daly (denis.daly@gmail.com) is currently vice president of Connected Sensors at Omni Design Technologies and the press coordinator for ISSCC 2018.

Laura C. Fujino (lcfujino@aol.com), ISSCC director of Publications and Presentations, is currently associated with the Department of Electrical and Computer Engineering at the University of Toronto, Canada, where she is a member of the Board of Advisors for Engineering Science.

Kenneth C. Smith is a professor emeritus and serves as chair of the Board of Advisors for the Division of Engineering Science, University of Toronto, Canada. Since 1975, he has volunteered for ISSCC in various capacities including as a member of the Program Committee, in press-related roles, and, for 40 years, as chair of the Awards Committee.