Filling the Gap

The definition for the frequency range of a terahertz wave varies in the literature. Some refer to terahertz as the spectrum between 300 GHz ($\lambda = 1$ mm) and 3 THz ($\lambda = 0.1$ mm) to distinguish it from its millimeter-wave counterpart (30–300 GHz). Others adopt a decimal pattern and a loosely defined frequency range from 100 GHz to 10 THz. Based on the latter definition, probably the earliest documented endeavor to reach terahertz frequency was published in a paper in *Physical Review* in 1923 [1]. In this nearly century-old work, authors Nichols and Tear generated a 0.17-THz signal using a spark-gap oscillator, detected it with a radiometer vane, and analyzed the spectrum through a Boltzmann interferometer.

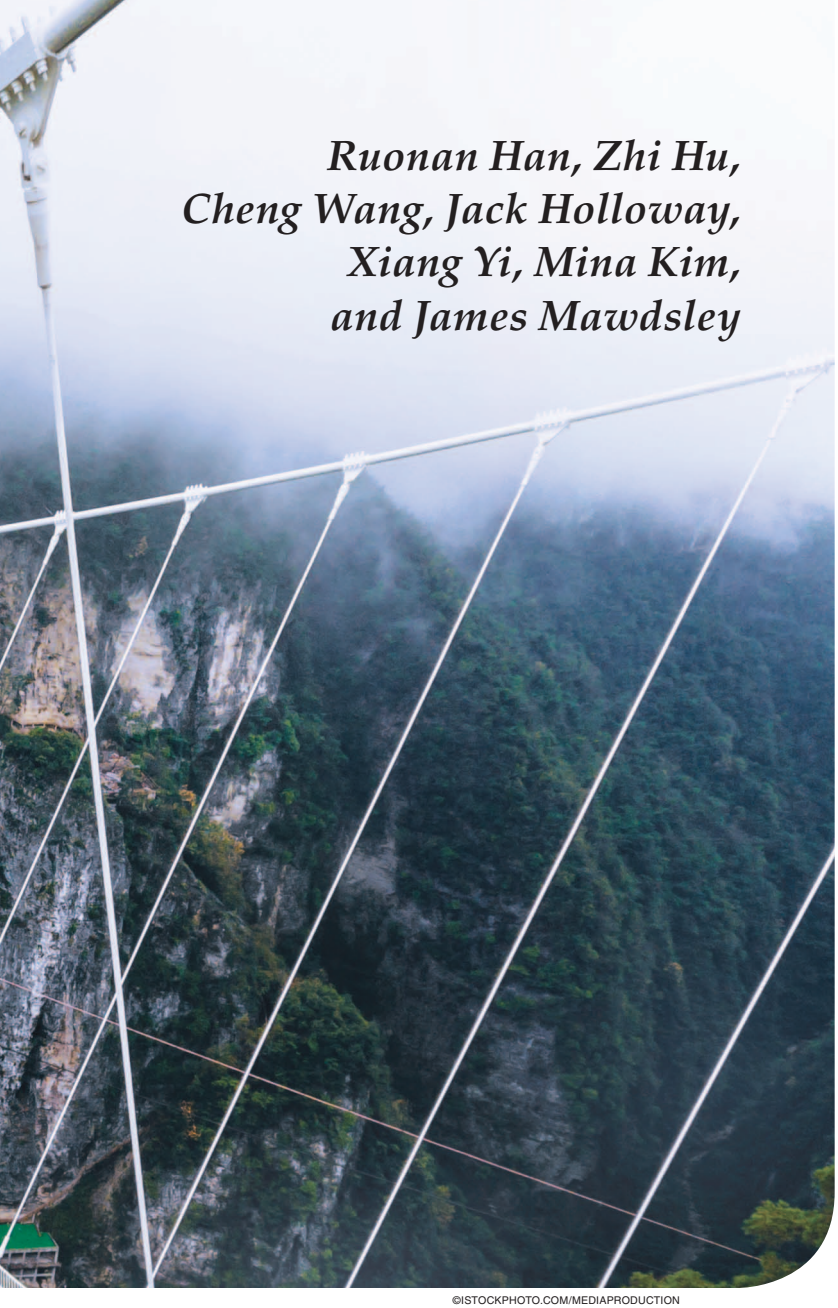
It is particularly interesting to revisit the motivation of the study [1]: "There is doubtless little need of any further experimental proof of the essential identity

Ruonan Han (ruonan@mit.edu), Zhi Hu (zhihu@mit.edu), Cheng Wang (wangch87@mit.edu), Jack Holloway (holloway@mit.edu), Xiang Yi (xiangyi@mit.edu), Mina Kim (minahkim@mit.edu), and James Mawdsley (jpmawd@mit.edu) are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, United States.

Digital Object Identifier 10.1109/MMM.2019.2891379

Date of publication: 7 March 2019

Ruonan Han, Zhi Hu, Cheng Wang, Jack Holloway, Xiang Yi, Mina Kim, and James Mawdsley



between heat waves and electric waves, yet for the sake of completeness, further effort to bridge the existing gap between these two spectra seems warranted." Similar to many other research pursuits, the journey of filling the so-called terahertz gap started from pure scientific curiosity!

In any case, how about one century later? The bad news is that the gap still exists, and, in fact, the significant challenges in building terahertz hardware and the scarce application of such hardware turn the terahertz spectrum into the widely acknowledged last frontier. The good news is that this area remains rich in highly diverse research opportunities. For example, regarding the generation of terahertz signals for electronic approaches alone, numerous areas have already been actively pursued, such as micromachined vacuum amplifiers and oscillators, resonant-tunneling diodes, III–V compound semiconductor transistors and diode multipliers, and CMOS/BiCMOS integrated circuits. This is in sharp contrast to the microwave and millimeter-wave electronics regime, where silicon integrated circuits play a dominant role and billions of such low-cost chips are actively running in every walk of life.

At the 2008 International Solid-State Circuits Conference (ISSCC), two CMOS oscillators at 410 GHz [2] and 324 GHz [3] marked the debut of CMOS terahertz circuits. Their output power was merely at the nanowatt level [2], [3], but, since then, remarkable progress has been made. Are we approaching a flourishing of terahertz-integrated electronics in silicon? In this article, we tackle this question through a series of discussions on some accomplishments and challenges in this field.

Are Silicon Devices Ready?

The End of Moore's Law in Terahertz Silicon Chips

Probably the most daunting problem of silicon terahertz circuits is the low speed of silicon MOSFETs and bipolar transistors. The scaling of CMOS devices is driven almost exclusively by the requirements of higher density and efficiency in digital systems. Unfortunately, the maximum oscillation frequency f_{max} of transistors does not scale up simultaneously. In fact, according to the measured f_{max} data of CMOS/silicon–germanium (SiGe) devices (including metal stacks) across a wide range of technology nodes [5], the high-frequency performance of CMOS peaks at the 45–65-nm nodes ($f_{max} \approx 320$ GHz) and then degrades with the device scaling.

Similarly, SiGe heterojunction bipolar transistors (HBTs) reach their peak at the 130-nm node ($f_{max} \approx 450$ GHz). As a result, amplifier design is possible only below ~0.2–0.3 THz, and signal sources above such a threshold rely only on harmonic generation originated from device nonlinearity. Not surprisingly, the efficiency in both cases is low. CMOS circuits' advantages of low cost and monolithic integration with baseband analog/digital blocks, while always attractive, gradually appear insufficient to justify their application in

terahertz hardware. What is worse, the data from [5] indicate that the end of Moore's law has already come about in silicon terahertz electronics, and more advanced CMOS/SiGe devices are unlikely to help fill the terahertz gap.

Powerful Chips Formed by Weak but Collegial Devices

So why do we still bet on silicon chips in light of all the deficiencies just stated? One important reason is that silicon devices, when forming a large-scale, collaborative, and flexible cluster, can sometimes be more powerful than many other terahertz technologies. These clusters fall into two categories:

- those that enable functions and applications that other technologies simply fail to deliver
- those that allow for coherent combinations of individual devices, leading to performance on a par with or even better than their III–V counterparts and at a much smaller scale.

For the former category, we present several examples in the section "Bridging the 2D Terahertz Gap in the System and Application Domains," including a rapid spectrometer with 20 simultaneous spectral-sensing channels (Figure 1). For the latter category, one typical example is terahertz radiator arrays, where on-chip terahertz sources (mostly harmonic oscillators) with integrated antennas share identical frequencies and phases, and their radiated waves are combined coherently and quasi-optically in free space. This approach, assisted by the high yield and repeatability of CMOS/SiGe transistors, enables the generation of high-intensity terahertz beams.

Shown in the survey in Figure 2, multiple milliwatts of total radiated power can now be generated from a CMOS chip in the low terahertz regime (~0.2–0.4 THz); this marks an improvement of five orders of magnitude compared to what the field started with a decade ago [2], [3]. This has been accomplished without a significant speed increase in silicon devices. Meanwhile, some of these radiated arrays also achieved additional functions, such as beam steering [6], [7] and integrated phase locking [8].

As we aggressively scale up the operation frequency of a large radiator array, other opportunities emerge. A typical radiator array with sidelobe suppression possesses an interelement pitch of about a half-wavelength ($\lambda/2$). As a result, a larger number of radiators n_{radiator} with increased output frequency can be implemented within a fixed chip area, leading to a more collimated and intense beam in the far field ($I_{\text{rad}} \propto n_{\text{radiator}}^2$). Figure 3 illustrates what is expected to happen if such coherent radiator arrays are implemented within a practical chip size of 10 mm^2 and at different frequencies. At 1 THz, a beamwidth of only a few degrees is expected, the result of a high array density of ~100 radiators/mm^2! Such a terahertz laser chip could effectively overcome path loss and extend sensing and communication distance (probably still within tens of meters, however).

But here is the caveat: even if the fastest SiGe HBTs ($f_{\max} \approx 450$ GHz) are used, we could reach 1 THz possibly only by 1) building a 250-GHz oscillator and 2) extracting its fourth harmonic signal—both highly challenging tasks that could lead to quite low efficiency. What is worse, the oscillator, frequency filter, and antenna should fit into a tight $(\lambda_{1\text{THz}}/2)^2$ space, which is only ~0.1 × 0.1 mm^2 in the on-chip dielectric (silicon dioxide

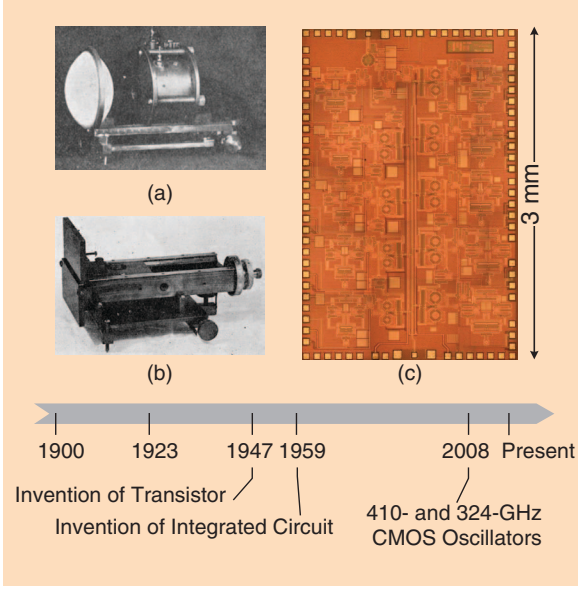


Figure 1. *The evolution of terahertz electronics: from (a) a spark-gap oscillator (f_{out} = 0.17 THz) [1] and (b) interferometer to (c) single-chip integration of 10 transceivers ~0.2–0.3 THz, on a CMOS chip [4].*

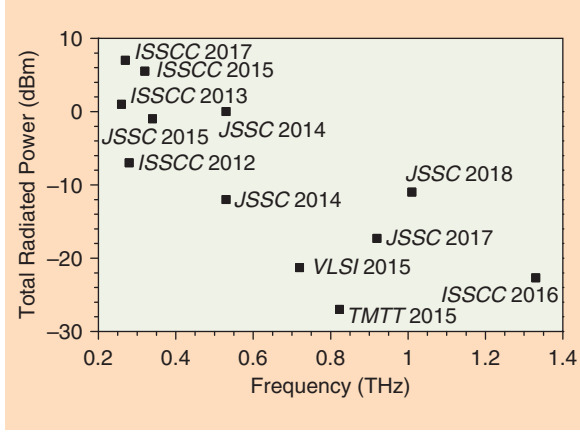


Figure 2. *State-of-the-art coherent silicon radiation sources [9]. JSSC: IEEE Journal of Solid-State Circuits; TMTT: IEEE Transactions on Microwave Theory and Techniques; VLSI: Symposium on Very-Large-Scale Integrated Circuits.*

and silicon) environment. The length of a 250-GHz resonator is already $\sim\lambda_{1\text{THz}}$! It now becomes evident that, to manufacture silicon devices operating in the concerted way we desire, more secret ingredients are needed.

Large-Scale, High-Density Array: A 1-THz Radiation Source in Silicon

Next, we show how the previous problems are solved in a 1-THz radiator array using a 130-nm SiGe HBT technology [Figure 4(a)] that integrates 91 coherent radiators within a 0.1-mm^2 chip area [9]. The key enabler is a multifunctional slot resonator structure that performs fundamental oscillation ($f_0 = 250$ GHz); interradiator 2D coupling; signal filtering at f_0, $2f_0$, and $3f_0$; and radiation at $4f_0$ simultaneously. In each array element, two HBTs drive two pairs of slot resonators and oscillate differentially at f_0. Next, standing waves at all harmonics also exist inside the slots, and normally occurring in-phase interference of these waves causes strong radiation. The loss associated with such radiation at f_0, $2f_0$, and $3f_0$ severely weakens the fundamental oscillation and frequency upconversion to $4f_0$. Fortunately, that problem is solved in our case, where 1) the slot resonators ($\lambda_{0.25\text{THz}}$/4-in length) are folded to form two squares with side lengths of $\lambda_{1\text{THz}}/2$; and 2) all elements tightly abut, as shown in Figure 4(a).

Now, as we examine the electrical field distribution in these slots [Figure 4(b)], we see that all standing waves in the vertical slot sections (e.g., in *BC*) are always canceled by their out-of-phase counterparts in adjacent array elements; hence, there is no radiation. More interestingly, as shown in Figure 4(c), the standing waves at $\sim f_0$–$3f_0$ are also mutually canceled in all horizontal slot sections, either horizontally adjacent (e.g., *AB* and *AB*′ at f_0 and $3f_0$) or vertically adjacent (e.g., *AB* and *CD* at $2f_0$). In contrast, at $4f_0 = 1$ THz (the desired output frequency), the waves in all of the horizontal slots are in phase, so that effective radiation occurs. Through geometric manipulation of the multiharmonic standing waves, we turn the oscillator's resonator into a frequency-selective radiation structure without having dedicated filters or antennas. Hence, the complexity and high-frequency loss of on-chip passives decrease significantly. Also, at 1 THz, each slot pair, such as *AB*–*AB*′ and *CD*–*CD*′, behaves as a dipole slot antenna, and the pitch of the formed antenna array is $\sim\lambda_{1\text{THz}}/2$—exactly what we strove for in the previous discussion.

Similar to many other research pursuits, the journey of filling the so-called terahertz gap started from pure scientific curiosity!

A radiator chip is implemented with 13 × 7 HBT-driven antennas inside only a 1-mm^2 area [9]. That well matches the density predicted in Figure 3. In the measurement, the chip generates a total radiated power of ~80 μW at 1 THz while consuming 1.1 W of dc power. The output half-power beamwidth, as a result of the large-scale coherent array, is only ±4° (directivity of ~24 dBi); this leads to a high equivalent isotropically radiated power of 20 mW. Such performance is already on a par with what photoconductive optoelectronics could deliver [13] but without the need for fast laser pumping. Last, it is noteworthy that the presented

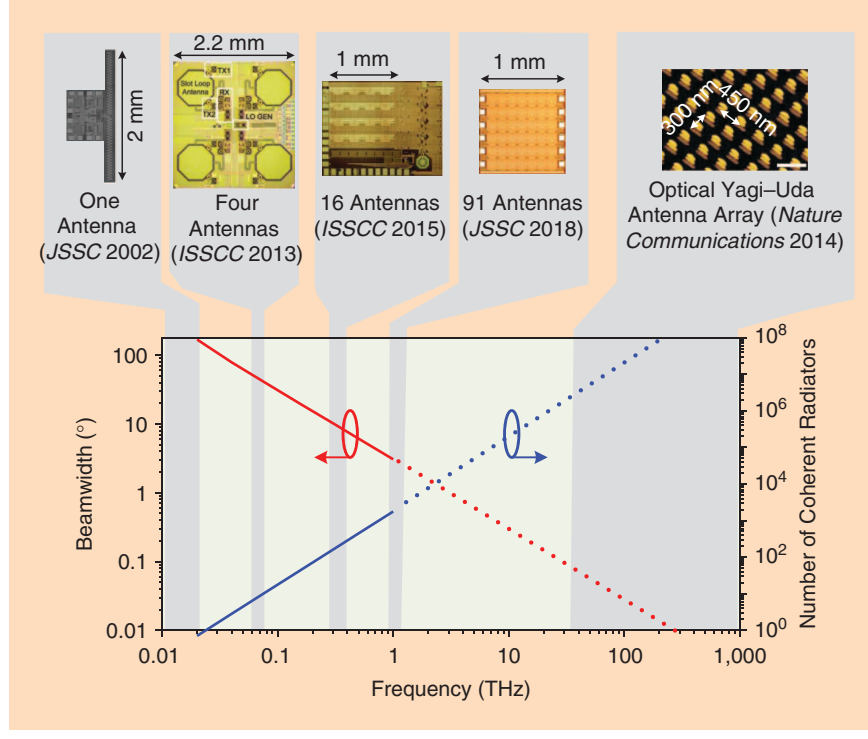


Figure 3. *An estimation of the number of coherent radiators integrated on a 10-mm^2 chip area and the corresponding beamwidth across the RF-to-optical spectrum. In the calculation of the wavelength, a dielectric environment with an equivalent relative dielectric constant of 2 is assumed. The works shown are [8]–[12], which match the array densities.*

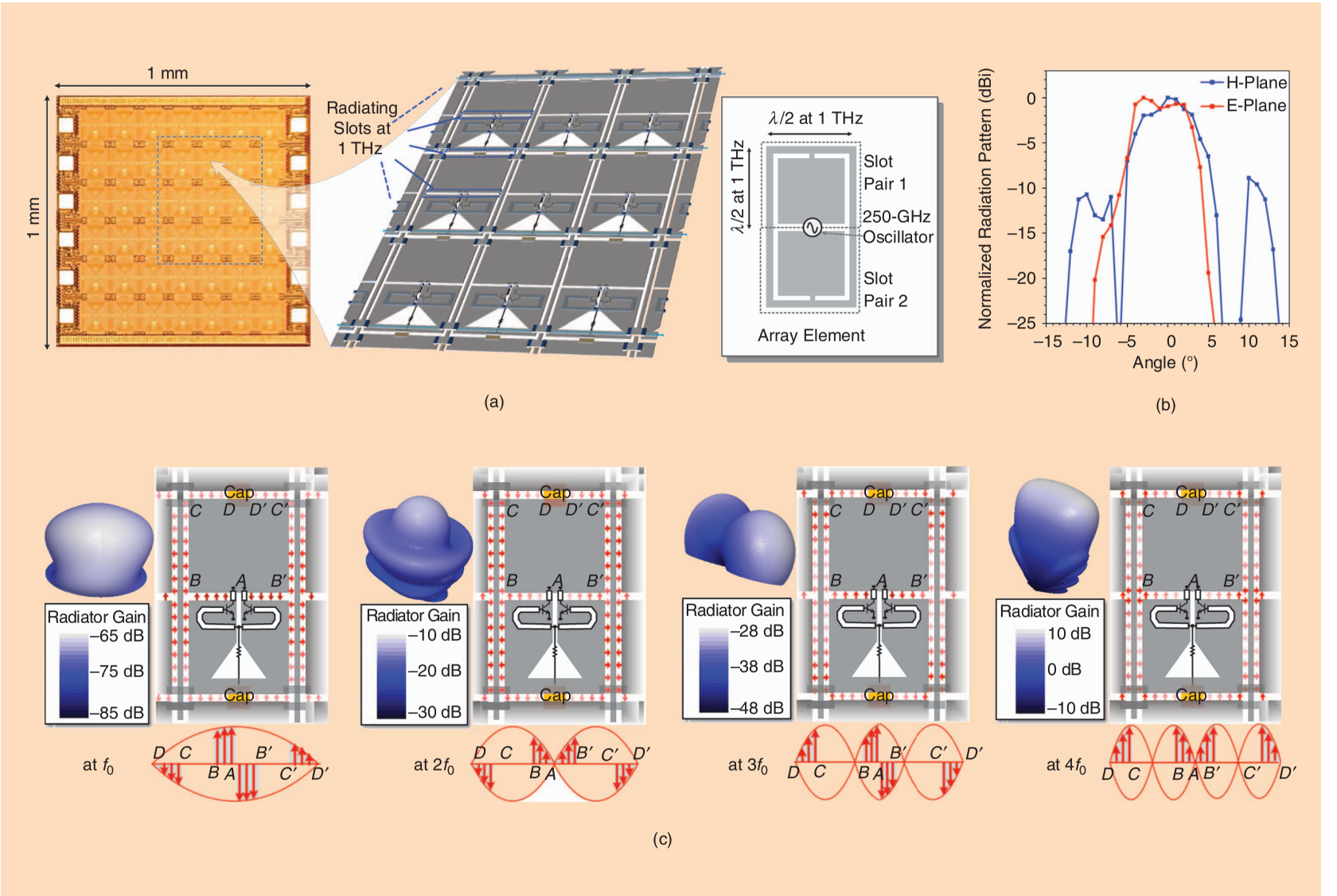


Figure 4. *(a) A large-scale, high-density 1-THz SiGe radiator array. (b) The measured radiation beam patterns of the chip. (c) Theoretical electrical-field distributions in each single unit at different harmonics (f_0, $2f_0$, $3f_0$, and $4f_0$) along with the resultant antenna gains simulated by a high-frequency structure simulator [9]. Cap: capacitance.*

multifunctional electromagnetic design is not the only ingredient; other techniques involving optimal device oscillating conditions [14]–[16], return-path gap structure [8], and so forth are also crucial and adopted in the circuit.

So, are silicon devices ready? Based on this discussion, we see that what they may achieve, at the component level at least, is not pessimistic at all. But to better answer this question, we need to zoom out and investigate the systematic and application levels.

Bridging the 2D Terahertz Gap in the System and Application Domains

The concept of the terahertz gap is traditionally linked to inferior performance of basic components in the terahertz realm, namely, the low generated power of terahertz sources and low sensitivity of terahertz detectors. While this certainly reveals fundamental physical limitations related to material and device properties, the deficiency in another two critical domains—systems and applications—is often overlooked. As a result, people have spent tremendous effort with terahertz components to fill the gap, with the notion that, when such tasks are completed, terahertz systems can be readily constructed by connecting the components in ways similar to their lower-frequency counterparts. Meanwhile, the projected applications fall almost entirely under three areas: broadband wireless communication, noninvasive imaging, and spectroscopy. The problem is that the path from terahertz components to terahertz systems is not as straightforward as one might expect, because of the many hurdles, such as excessive loss of interconnects among components, incompatibility of component materials, low yield, and integration scale.

In addition, studies of many terahertz applications, often based on laboratory-scale demonstrations and targeting long-range scenarios, have led to pessimistic conclusions about their practicality [17]. Last, more studies on the interplay between terahertz systems and applications are critically needed in terms of 1) how the unique features and challenges of a terahertz application affect the system architecture and 2) how the progress and new capabilities of terahertz systems, e.g., those in small and cheap CMOS chips, enable applications beyond the scope outlined previously.

Thus, the terahertz gap is, in reality, 2D (Figure 5)—and its system and application sections require dedicated efforts to be completely filled. For this purpose, CMOS serves as an ideal platform because, from day one, there are at least no concerns for the integration of components, economic practicality of the systems, and such. Next, a few terahertz CMOS systems are presented as examples of such efforts.

Silicon devices, when forming a large-scale, collaborative, and flexible cluster, can sometimes be more powerful than many other terahertz technologies.

CMOS Terahertz Rotational Spectrometer: High Parallelism in Broadband Sensing

Among the many proposed applications of terahertz spectroscopy, some—such as backscatter-based, stand-off detection of explosives and hazardous materials—appeared to be attractive but later faced severe problems associated with atmospheric absorption, pressure broadening, and the like. [17]. In comparison, some other spectroscopy operations under controlled conditions, such as rotational-mode gas sensing, still hold great promise. Because of the electric/magnetic dipoles of polar molecules, waves with certain frequencies can excite the rotation of molecules. The resonance frequency is determined by the molecular structure. The quality factor (Q) of the Doppler-limited transition lines in a transmission-mode spectroscopic setup with low gas pressure is typically $\sim 10^5$–10^6. Such unique fingerprints of molecules, therefore, enable so-called absolute specificity in molecular identification [18].

Because of the quantum nature of rotational energy levels, a polar molecule possesses many discrete transition lines, which normally span from microwave to terahertz and follow a periodic pattern. For example, the spectral periods of carbonyl sulfide (OCS) and hydrogen cyanide (HCN) are $\Delta f_{\mathrm{OCS}} = 12.16$ GHz and $\Delta f_{\mathrm{HCN}} = 88.61$ GHz, respectively. Compared to the

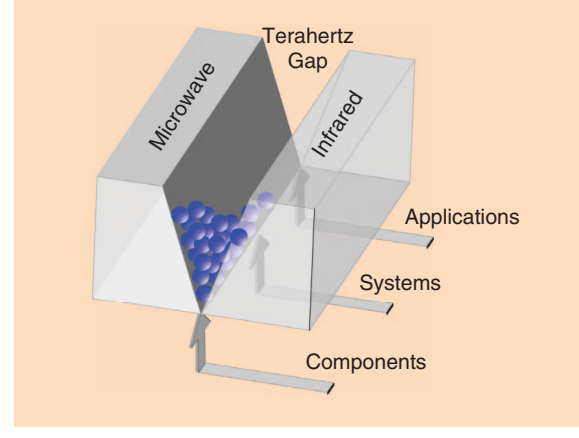


Figure 5. *The terahertz gap is two dimensional. In addition to stand-alone signal sources and detectors with operational frequencies between microwave and infrared, large-scale and high-performance systems as well as new applications enabled by these systems are in even more critical demand.*

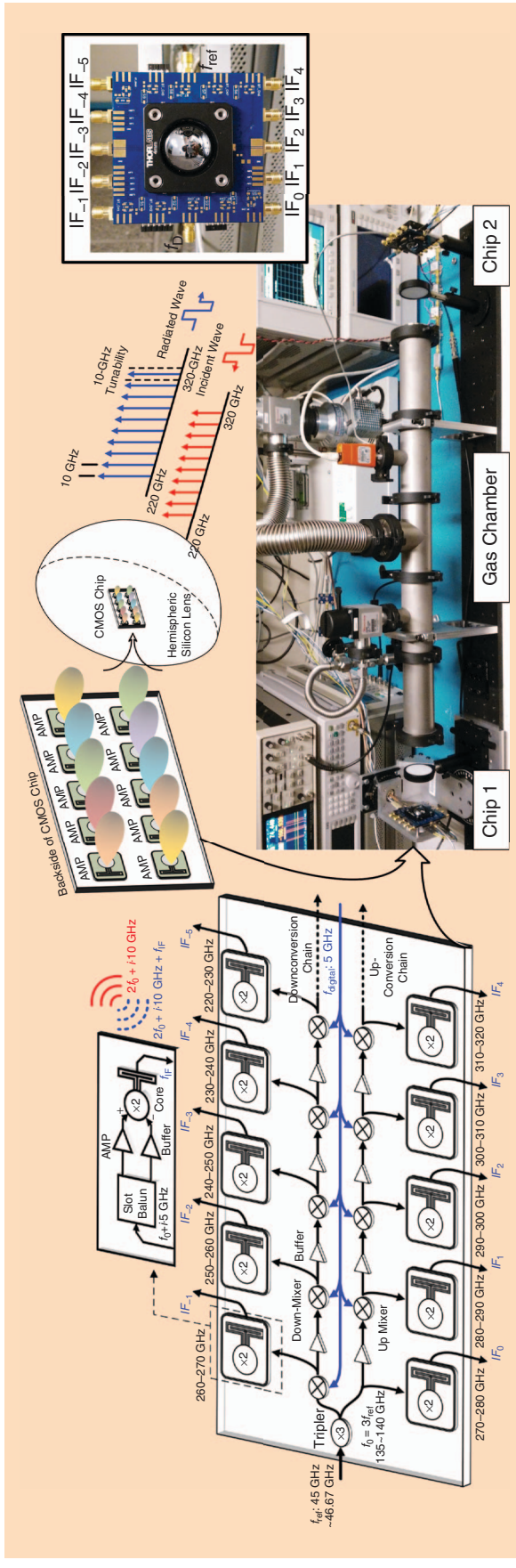


Figure 6. *Dual-terahertz comb spectroscopy for gas sensing: the chip architecture and measurement setup [20].*

rotational spectroscopy performed in the microwave regime, rotational spectroscopy in the terahertz regime offers a few advantages. First, transition lines at terahertz have higher absorption intensity because of the larger number of degenerate substates. As a result, we can perform spectroscopy in a much shorter gas cell. For example, to miniaturize the spectrometer, we may use a single-mode hollow waveguide to hold the gas; its cross-sectional area scales down quadratically with the wave frequency. Second, in the terahertz frequency range, the available bandwidth is much broader (given the same fractional bandwidth of electronics); hence, there is a wider range of molecular detection. This is particularly useful for detecting molecules (e.g., HCN) with a large rotational spectral period Δf. A bandwidth of up to ~100 GHz is typically preferred. Previously, a tabletop-size spectrometer using solid-state terahertz multipliers and receivers and a gas preconcentration (gain of ~10^5) was demonstrated, with a parts per trillion-level sensitivity [19].

There has been a growing interest in building terahertz spectrometers using CMOS technologies. It is expected that such compact and low-cost chip-scale sensors can make high-precision, wide-range analysis of exhaled volatile organic compounds (VOCs) possible. The concentration of different VOCs in human breath correlates with metabolism disorders; thus, VOC analysis is an effective, noninvasive approach for health monitoring. Recently, SiGe and CMOS spectrometer chips based on single-tone, continuous-wave operation have been demonstrated [21], [22].

In the meantime, however, a few critical challenges have emerged. The first relates to a bandwidth/efficiency tradeoff. Most, if not all, terahertz circuits rely on high-Q resonance to achieve optimal performance (e.g., highest power); this inevitably leads to a narrow operational bandwidth. Therefore, single-tone spectrometers have either reduced bandwidth or degraded power and sensitivity (especially on the two band edges). The second challenge stems from a tradeoff between resolution and scanning time. Because of the narrow molecular lines, the spectrometer needs to scan with high resolution (<100 kHz). If the total covered bandwidth is as large as ~100 GHz, this will entail a few hours of scanning time (assuming an integration time of 10 ms per point).

One viable solution, enabled by the excellent integration capability of CMOS, is parallel spectral sensing. This has been recently demonstrated in a dual-frequency comb spectrometer [4] (die photo given in Figure 1). Shown in Figure 6, two identical CMOS chips are used in the spectrometer. Each chip

transmits 10 equally spaced signal tones (i.e., a comb spectrum), located in the 220–320-GHz band, while receiving the other 10 tones from another chip. The two comb spectra are tunable but always keep a subgigahertz offset. As a result, with only a 10-GHz tuning range, the counterpropagating waves can seamlessly cover the entire 100-GHz bandwidth with their 20 signal tones. Each CMOS chip integrates 10 interlocked terahertz transceivers to realize these functions. An input frequency f_{ref}, ranging between ~45 GHz and 46.67 GHz, is tripled on the chip first, then successively up-/down-converted by 5 GHz at each stage, and finally doubled to ~220–320 GHz inside each active molecular probe (AMP) module. The AMP doubler not only radiates the output signal, but also downconverts the received signal tone (from another chip) to a subgigahertz intermediate frequency (IF) through subharmonic mixing. The 20 IF signals are stitched in the back end.

Because the 100-GHz-wide band is now scanned by 20 tones simultaneously, the total scanning time is greatly reduced. The comb architecture also reduces the bandwidth requirement for all blocks to only ~5%. That allows us to incorporate high-Q resonator structures and even positive feedback (in the AMP doubler) to enhance the device gain [20]. As a result, in the measurement, the chip provides a total radiated power of 5.2 mW, and no power degradation is observed at the edges of the band. The measured noise figure is ~15 dB at the lower end of the band and ~20 dB at the higher end. In gas-sensing experiments, we obtained all transitions of a gas mixture (V_{OCS}:V_{CH_3CN} = 1:60) within the entire ~220–320-GHz band (Figure 7). The results also well match the existing molecular spectroscopy database [23]. The signal-to-noise ratio (SNR) of the measured OCS lines reaches 89 dB, which indicates a detection sensitivity of 11 parts per million (ppm)/$\sqrt{\tau}$ for OCS and acetonitrile (CH_3CN) and a 3 ppm/$\sqrt{\tau}$ for HCN (τ is the integration time in seconds) [24]. The parallel sensing scheme allows for a longer integration time. If a preconcentration similar to that in [19] is used, the sensitivity is expected to improve to the sub-parts-per-billion level, which is sufficient for the previously discussed diagnostics via breath analyses.

Last, the gas tube in Figure 6 can be replaced with a folded, single-mode waveguide. As we show in the following section, the gas tube is then shrunk to bottle-cap size, leaving the vacuum and preconcentration systems to be the remaining bottlenecks for spectrometer miniaturization. In that regard, a battery-sized scroll pump has been developed and recently used in miniaturized mass spectrometers [25].

Time Keeping: CMOS Molecular Clock

In addition to molecular identification, another significance of terahertz rotational spectroscopy is that

The path from terahertz components to terahertz systems is not as straightforward as one might expect, because of the many hurdles.

it provides a fully electronic approach for standard CMOS chips to correlate their low-frequency clock with a highly stable physical constant, namely, the resonance frequency of certain molecules. Therefore, with a closed-loop configuration that automatically aligns its terahertz probing frequency with the transition line center of certain encapsulated molecules, a highly stable clock signal is obtained. Such a molecular clock scheme opens up new opportunities for building compact, low-cost frequency references with atomic clock-grade stability [26].

Previously, chip-scale atomic clocks (CSACs) using the coherent-population trapping principle were developed [27], [28] that provide long-term stability of 10^{-11} with ~120 mW of power [28]. These CSACs have been utilized in defense equipment, underwater oil-detection sensors, and so forth. Further expansion of their applications is, however, encumbered by their high cost (>US$1,000), as a result of their complicated electrooptical construction.

Recently, we reported the first CMOS prototype of a chip-scale molecular clock, which uses OCS gas as its reference [26], [30]. OCS molecules have rotational transition lines every ~12.16 GHz, and the one at ~231.061 GHz is used in our clock. Shown in Figure 8(a), the clock includes a transmitter and receiver chipset that probes low-pressure (10-Pa) OCS gas inside a folded, single-mode waveguide. The transmitter is

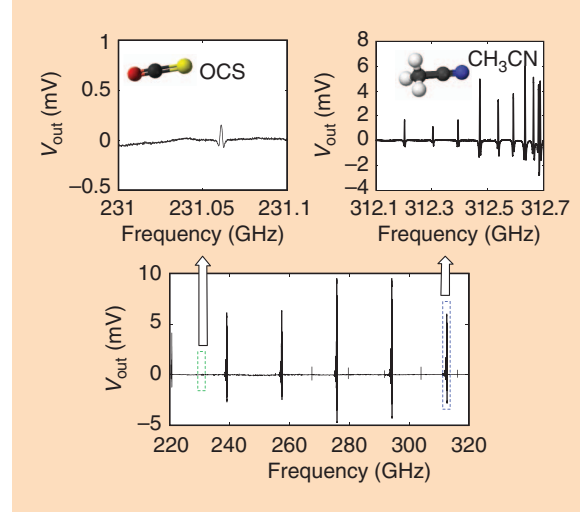


Figure 7. *The measured transition lines of an OCS and acetonitrile (CH_3CN) mixture (V_{OCS}:V_{CH_3CN} = 1:60) [24].*

based on a terahertz Δ–Σ fractional-N phase-locked loop (PLL), which locks its 231-GHz output to an 80-MHz signal generated by a voltage-controlled crystal oscillator (VCXO). Through an automatic adjustment of the PLL division ratio, the transmitter output is frequency-shift keying (FSK), modulated with a modulation frequency f_m of 16 kHz and a deviation frequency Δf of 1 MHz. The chip therefore periodically probes the absorption at the two slopes of the OCS transition line [Figure 8(b)]. If the mean probing frequency f_c is not aligned with the center of the transition line (f_0), an absorption imbalance occurs.

When the terahertz signal is eventually injected into a terahertz square-law detector inside the clock receiver, the imbalance is turned into a square wave V_{error} at f_m. The amplitude and phase of V_{error} are

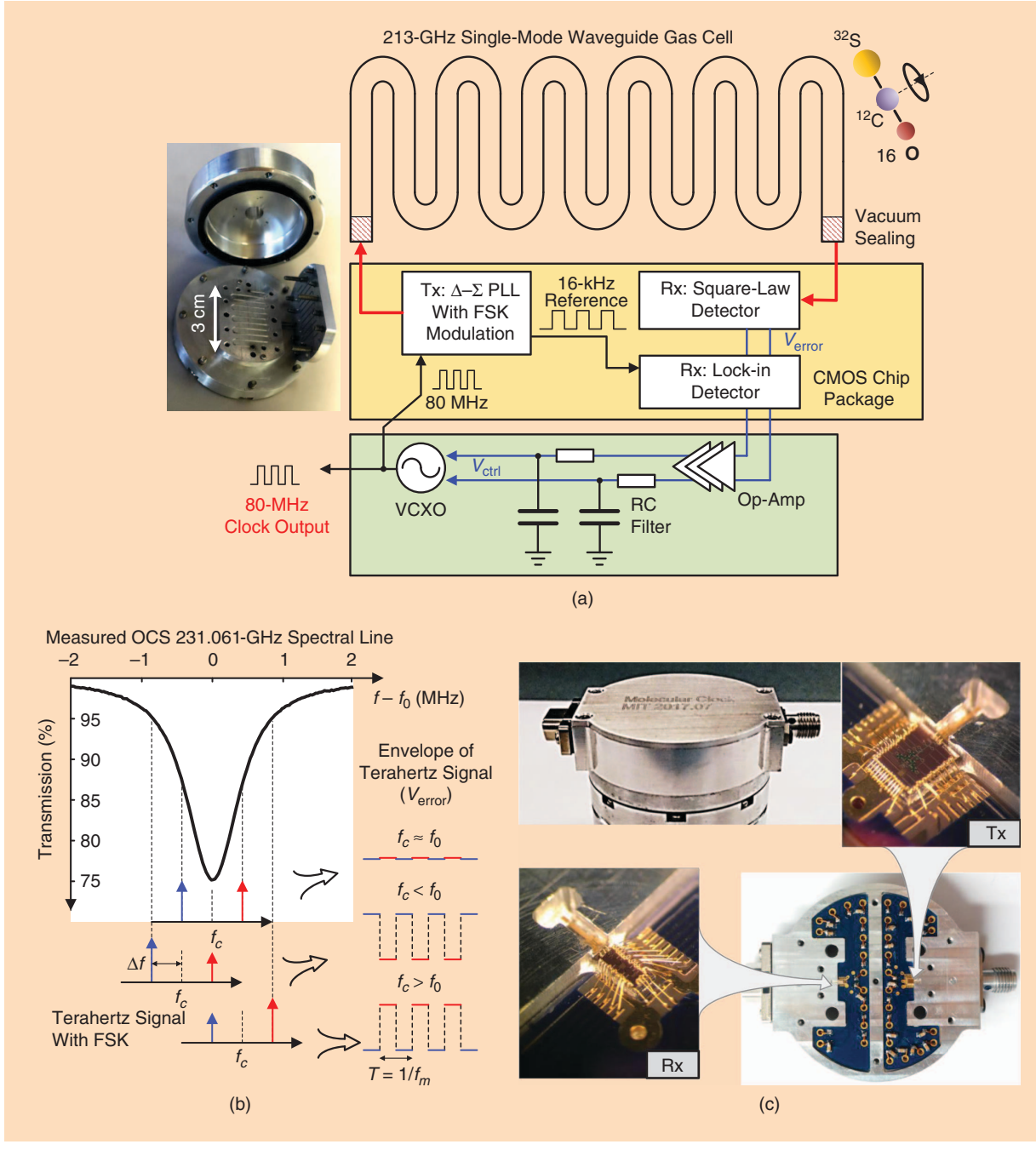


Figure 8. *A CMOS molecular clock: (a) the chip architecture and an interior view of the gas cell; (b) the molecule probing principle, based on an FSK-modulated signal; and (c) photos of the clock package [29]. Rx: receiver; Tx: transmitter; Op-amp: operational amplifier.*

determined by $f_0 - f_c$ [Figure 8(b)]. Subsequently, with another integrated lock-in detection at f_m and low-pass filtering, V_{error} is turned into a control signal V_{ctrl} used to dynamically tune the VCXO and keep the V_{error} zero (i.e., $f_0 = f_c$).

The short-term stability of the clock is determined by the product of 1) the Q of the spectral line ($Q \sim 2 \times 10^5$) and 2) the SNR of the measured transition curve. Our experimental studies in [26] show that, to achieve a high SNR without saturating the molecular states, only ~100 μW of probing power is needed, leading to reduced power consumption in comparison with other terahertz microsystems. The measured dc power of the clock chipset is only 66 mW. The clock package, shown in Figure 8(c), is the size of a matchbox. Ultimately, the molecular clock size is limited by the volume of the OCS, which is approximately 83 mm^3. Currently, the connections of the chips to the waveguide are via bond wires; this, however, causes ~20 dB of total coupling loss. Therefore, the SNR of the OCS transition curve obtained by the clock degrades from the expected value of ~70 dB to 53 dB. Nevertheless, the measured Allan deviation of the chip-scale molecular clock, shown in Figure 9, still reaches 3.8×10^{-10}, with an averaging time τ of 1,000 s. With improved chip–waveguide coupling, the Allan deviation should reach near 10^{-11}, which has been experimentally verified by a laboratory-scale clock using vector network analyzer frequency extenders with a waveguide interface [26]. The long-term (e.g., $\tau = 1{,}000$ s) stability performance is currently degraded by gas-cell leakage and is ultimately limited by the temperature/pressure-induced shift of the OCS transition (see details in [26]).

The chip-scale molecular clock is expected to significantly reduce the construction complexity and cost of high-stability frequency references. A monolithic integration of a molecular clock (not including the gas cell) in a CMOS system on chip is now made possible and can lead to ultracompact systems for navigation and sensing under GPS-denial conditions.

On-Chip Terahertz Imaging: Toward Large-Scale Sensitive Arrays

Early silicon-based terahertz imagers are based on a focal-plane array of square-law detectors [31], [32]. Unlike terahertz sources, these detectors rely on passive self-mixing of the signal inside a nonlinear device and are therefore much less constrained by the cutoff frequency (f_T)/f_{max} of devices. That is why terahertz imaging has been demonstrated even with MOSFET detectors in a 0.25-μm CMOS process [31] and with Schottky-diode detectors in a 0.13-μm CMOS process [32]. It is, however, noteworthy that the sensitivity (i.e., the noise-equivalent power [NEP]) provided by most, if not all, of these square-law detectors ranges

The terahertz gap is, in reality, 2D—and its system and application sections require dedicated efforts to be completely filled.

from tens to hundreds of NEP units, and no significant improvement has been observed in recent years. For a practical bandwidth of 1 kHz, a detector with a 100-pW/$HZ^{1/2}$ NEP has a sensitivity of only 3.2 nW. While this may be sufficient for short-range, transmissive-mode imaging, orders-of-magnitude improvement is required for applications like reflective-mode, standoff screening. Nevertheless, one significant advantage of CMOS terahertz imagers, in comparison with other detector technologies, is that they can readily scale to a large array size, which increases the imaging speed. For example, in [33], a 32×32 array of 1-THz detectors is demonstrated.

One approach to enhance the sensitivity is to move from square-law to heterodyne detection. By mixing the weak input signal with a strong local oscillator (LO) signal, the output baseband signal is much larger than that generated via self-mixing in a square-law detector. State-of-the-art terahertz heterodyne receivers have achieved noise figures of ~20 dB (such as the one in [4]), which is equivalent to a sensitivity of 0.4 fW (assuming a practical bandwidth of 1 kHz).

Therefore, what is the next challenge? Similar to constructing square-law imagers, building a large-scale heterodyne array is highly desirable. Such a sensor not only can directly replace the focal-plane array while providing better sensitivity but also can enable electronic steering of its narrow-beam response using the interference of output IF signals. The latter may eliminate the slow and bulky raster scanner used in high-resolution terahertz systems. Unfortunately, what we have ignored so far is that a terahertz heterodyne

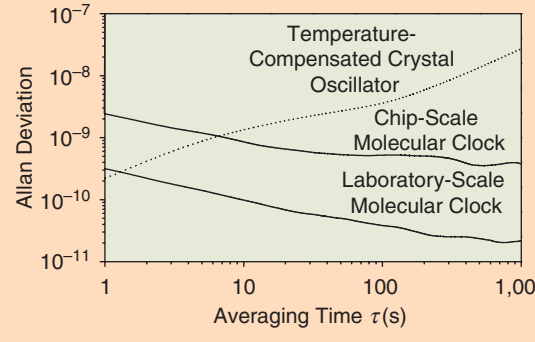


Figure 9. *The measured Allan deviation of molecular clocks [29].*

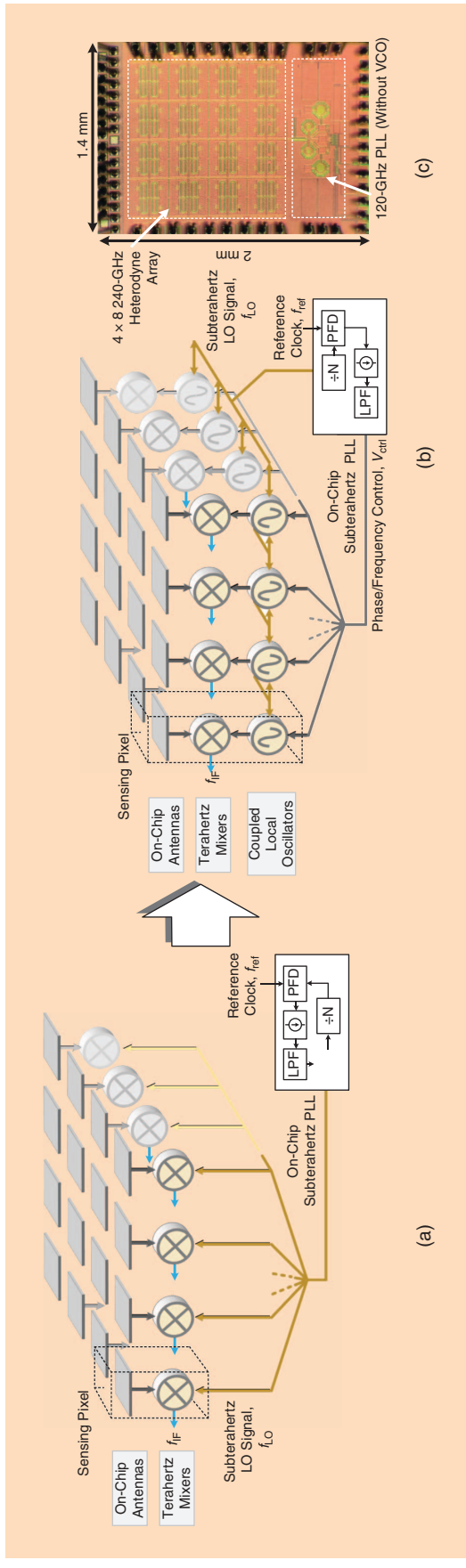


Figure 10. *A comparison between (a) the conventional centralized array architecture and (b) the decentralized architecture. (c) A CMOS prototype for the latter [34]. LPF: low-pass filter; PFD: phase-frequency detector; VCO: voltage-controlled oscillator.*

receiver, because of the additional LO generator, typically occupies a much larger area. This prevents the implementation of large-scale heterodyne arrays with a half-wavelength unit pitch, which is required to suppress sidelobes in beam steering.

A solution is provided in [35] with an array where eight antenna-fed subharmonic mixers share a single LO generator via an LO-distribution network. But the sensitivity compared with square-law detectors is improved by only some 10 times ($P_{NEP} = 71$ pW when the bandwidth is 1 kHz). One key problem stems from the adopted centralized architecture (Figure 10): the LO power shared by each mixer is weak, causing higher conversion loss. In particular, when the array is scaled up, the situation further deteriorates, and the LO network will eventually become excessively complicated and lossy. Therefore, approaches for increased scalability and density of heterodyne arrays are needed.

To address such a demand, a decentralized array architecture is presented [34]. Shown in Figure 10, an LO is incorporated in each pixel and is also coupled to its neighboring peers, so that a synchronized LO network is formed. The network is part of a PLL as well, which references to an external low-frequency clock. This architecture ensures that the pixel conversion loss is independent of the array size and that it decreases the coupled LO phase noise when the array scales up. Compared to the previous centralized scheme, it also eliminates the high-frequency global signal (only a megahertz-level LO tuning signal is distributed to all of the units). However, the aforementioned problem related to limited pixel area reoccurs in this highly scalable architecture.

In [34], a self-oscillating subharmonic mixer, which performs downconversion within the LO, is adopted to increase space utilization. Similar to the 1-THz radiator array in the "Are Silicon Devices Ready?" section, a modified, multifunctional structure is also used to simultaneously perform fundamental resonance, harmonic-signal reception, and interunit coupling. The design is prototyped in a 65-nm CMOS chip shown in Figure 10(c), where a 4×8 array of 240-GHz antenna-fed heterodyne units is integrated inside a ~1.7-mm^2 area (the total area, including the built-in 120-GHz PLL, is 2.8 mm^2). The measured sensitivity (1-kHz bandwidth) of the array pixel is 58 fW, with a ~5-dB variation among all of the pixels. The dc power of the chip is 0.98 W [36]. Shown in Figure 11, this receiver array, compared to other square-law arrays of similar scale and density, exhibits an orders-of-magnitude sensitivity improvement. Sensors with >1,000 pixels and subfemtowatt (1-kHz bandwidth) sensitivity should be possible in the future, which may replace cryogenically cooled bolometers and make many long-range terahertz imaging applications, such as real-time standoff screening, more practical.

Terahertz Communications Using CMOS: Wireless Meets Wireline

CMOS/BiCMOS wireless transceivers have been playing critical roles in microwave and even millimeter-wave communication systems. With the rapid progress in terahertz silicon circuits, transceivers with terahertz carriers have been gaining increasing attention in the past few years. Shown in Figure 12 is a summary of recent high-speed (data rate ≥10 Gb/s) CMOS wireless transceivers. Our observation is that, other than the 300-GHz CMOS transmitter reported in [37], the data rate supported by the majority of terahertz transceivers and transmitters is still lower than 30 Gb/s. In comparison, the data rate of sub-100-GHz transceivers has ramped up rapidly, recently reaching 120 Gb/s, with an excellent energy efficiency of 2.3 pJ/b [38].

Not surprisingly, one major hurdle for terahertz transceivers in CMOS is the limited high-frequency performance of transistors. In particular, for signal generation above ~200 GHz, heavily driven nonlinear circuits with harmonic output are almost the only solution; this inevitably leads to low output power (and, hence, a very short transmission distance) and makes a high-order data modulation scheme difficult, if not impossible. As a result, it is still not clear if CMOS terahertz transceivers can catch up with their millimeter-wave counterparts in the near future. However, with the development of III–V/CMOS heterogeneous integration processes, the situation may potentially change.

Thus, do terahertz CMOS chips play an irreplaceable role in communication? Very likely, if we extend our investigation of terahertz communication from the wireless domain to the wireline realm. For short-reach links, such as those for interchip communications, although a terascale data rate is required, such throughput is normally realized by the aggregation of a high-density but low-speed copper wire array (e.g., the interconnects in a packaging interposer). For long-reach communications, copper interconnects are unsuitable because of their excessive

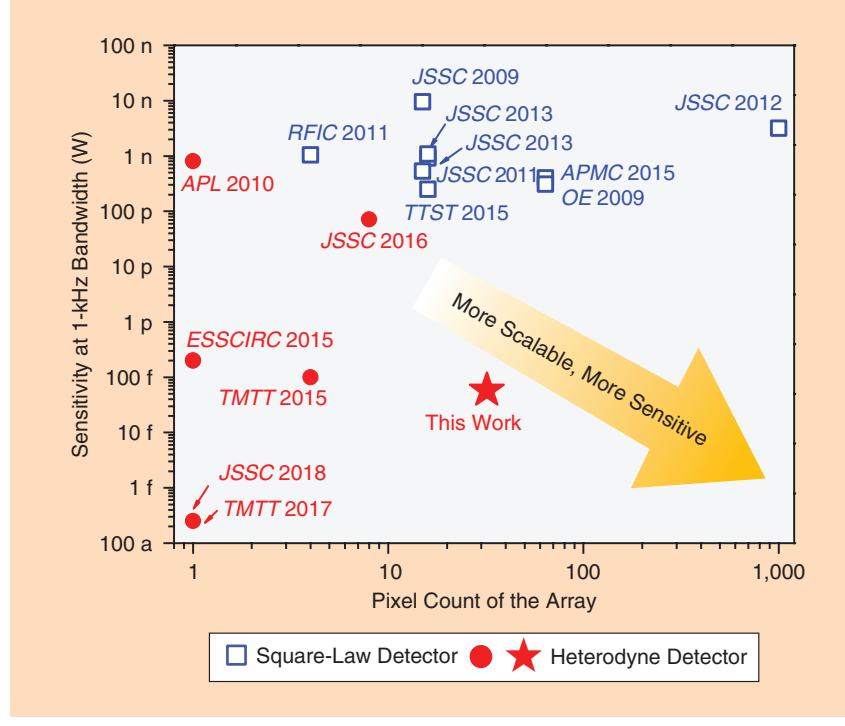


Figure 11. *The heterodyne array in [34] in comparison with other terahertz imaging arrays in silicon. APL: Applied Physics Letters; ESSCIRC: European Solid-State Circuit Conference; RFIC: Symposium on Radio-Frequency Integrated Circuits; TTST: IEEE Transactions on Terahertz Science and Technology; APMC: IEEE Asia-Pacific Microwave Conference; OE: Optical Express.*

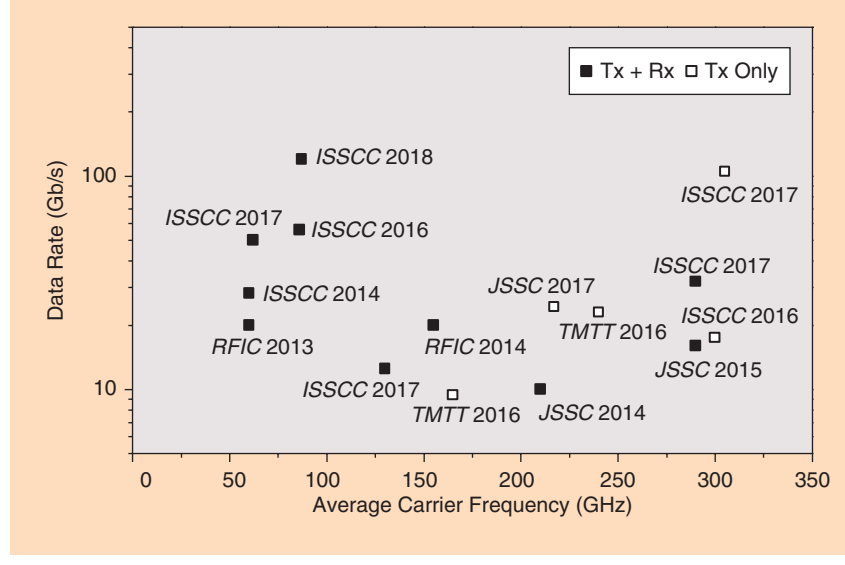


Figure 12. *A comparison of recent high-speed wireless transceivers (Tx + Rx) and transmitters in silicon.*

Research in silicon-based terahertz circuits is undergoing a transition from the exploration of their feasibility to their systematic/application practicality.

loss and dispersion; but, fortunately, the smaller number of links needed justifies the use of the more expensive optical interconnects. The situation becomes awkward, though, for the midrange links (a few meters) among adjacent server racks or blades inside a rack: the transmission distance prevents high data rates and efficiency if using copper links, while the link density needed is still too great to use optical fibers in an economical way.

To fill this gap in interconnect solutions, there is an increased interest in terahertz transmission through dielectric waveguides. Prior research has shown that, compared with metal cable, dielectric waveguides exhibit much smaller loss and dispersion [40] while readily offering bandwidths of tens to hundreds of gigahertz. As a result, the midrange links may be realized by two banks of channelized terahertz transceivers coupled through a dielectric waveguide. Because of the confinement of the terahertz wave, the requirement for transmitter output power is much more relaxed compared to the wireless case, making CMOS terahertz transceivers an attractive choice for cost-sensitive systems.

At present, the data rate demonstrated by the prototypes [41], [42] of this concept is still within 12 Gb/s, although it is imaginable that a larger throughput may be realized using higher-speed circuitry, such as the transmitter in [37], or combining multiple transceivers (using different carrier frequencies) via a channelizer. Another challenge lies in the ultrabroadband coupling between the chip and the dielectric waveguide.

In [43], we presented a fully integrated coupler using a leaky traveling-wave substrate-integrated waveguide (SIW) structure. Shown in Figure 13, the coupler turns the input signal into a traveling wave with a horizontal electric field facing toward a dielectric waveguide placed on top of the silicon chip. The field distribution well matches that of the wave mode inside the waveguide, which enables effective coupling. The SIW width is also tapered down, leading to a gradually increased cutoff frequency and wave impedance. As a result, the signal wave is squeezed from the coupler to the waveguide as it travels in the longitudinal direction; this shortens the coupler length needed for a certain bandwidth. In our experiments, Rogers polymer strips ($\epsilon_r = 6$) with various lengths and bends are used to connect standard 130-nm BiCMOS chips. The measured back-to-back overall link loss is ~11 dB. The loss and group delay stay flat across the entire 220–270-GHz band [Figure 13(c)]. Compared to other schemes, this coupler design offers a straightforward planar integration of chip and waveguide for the terahertz link systems.

Conclusions

Research in silicon-based terahertz circuits is undergoing a transition from the exploration of their feasibility to their systematic/application practicality. As described in this article, chip-scale systems utilizing a large number of collegial devices are enabling applications that are nonconventional compared to those of the earlier terahertz community and include miniature clocks, beam-steerable imaging, and interchip links. This work opens up new opportunities that will change the current landscape of terahertz research and accelerate the commercialization pace of terahertz

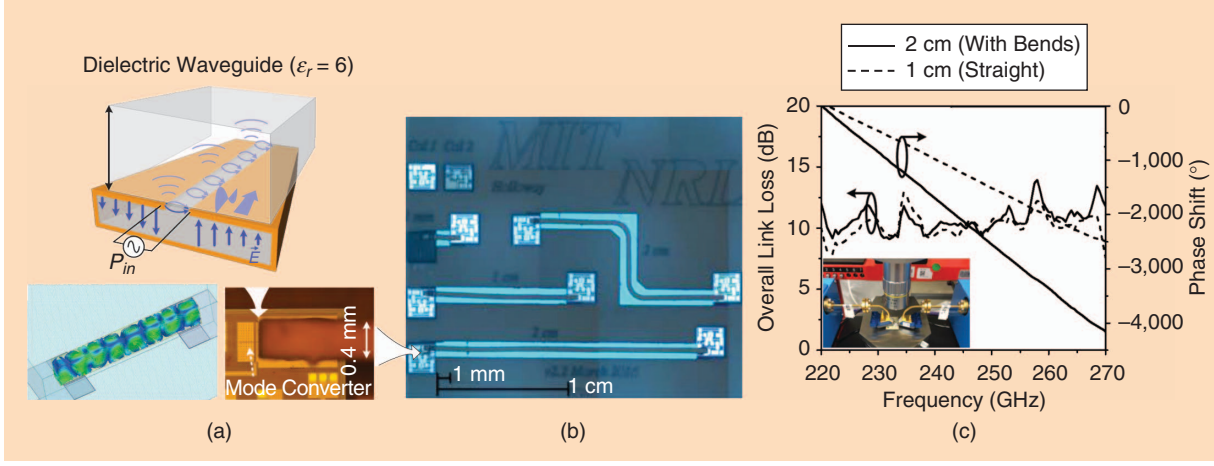


Figure 13. *A fully integrated chip-to-dielectric waveguide coupler: (a) the design, (b) test structures, and (c) the measured back-to-back loss and dispersion [39].*

hardware. While not perfect, silicon-based integrated circuits are our best bet to push the field forward.

References

[1] E. F. Nichols and J. D. Tear, "Short electric waves," *Phys. Rev.*, vol. 21, pp. 587–610, June 1923.

[2] E. Seok et al., "A 410GHz CMOS push-push oscillator with an on-chip patch antenna," in *Proc. Int. Solid-State Circuit Conf. (ISSCC)*, San Francisco, 2008, pp. 472–473.

[3] D. Huang, T. LaRocca, L. Samoska, A. Fung, and M.-C. F. Chang, "324GHz CMOS frequency generator using linear superposition technique," in *Proc. Int. Solid-State Circuit Conf. (ISSCC)*, San Francisco, 2008, pp. 476–477.

[4] C. Wang and R. Han, "Rapid and energy-efficient molecular sensing using dual mm-wave combs in 65nm CMOS: A 220-to-320GHz spectrometer with 5.2mW radiated power and 14.6-to-19.5dB noise figure," in *Proc. Int. Solid-State Circuit Conf. (ISSCC)*, San Francisco, 2017, pp. 18–20.

[5] R. L. Schmid, A. Ulusoy, S. Zeinolabedinzadeh, and J. Cressler, "A comparison of the degradation in RF performance due to device interconnects in advanced SiGe HBT and CMOS technologies," *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1803–1810, 2015.

[6] K. Sengupta and A. Hajimiri, "A 0.28THz 4x4 power-generation and beam-steering array," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, 2012, pp. 256–257.

[7] Y. Tousi and E. Afshari, "A high-power and scalable 2-D phased array for terahertz CMOS integrated systems," *IEEE J. Solid-State Circuits*, vol. 50, no. 2, pp. 597–609, 2015.

[8] R. Han et al., "A 320GHz phase-locked transmitter with 3.3mW radiated power and 22.5dBm EIRP for heterodyne THz imaging systems," in *Proc. Int. Solid-State Circuit Conf. (ISSCC)*, San Francisco, 2015, pp. 446–447.

[9] Z. Hu, M. Kaynak, and R. Han, "High-power radiation at 1-THz in silicon: A fully scalable array using a multi-functional radiating mesh structure," *IEEE J. Solid-State Circuits*, vol. 53, no. 5, pp. 1313–1327, 2018.

[10] B. A. Floyd, C.-M. Hung, and K. K. O, "Intra-chip wireless interconnect for clock antennas, receivers, and transmitters," *IEEE J. Solid-State Circuits*, vol. 37, no. 5, pp. 543–552, 2002.

[11] L. Kong, D. Seo, and E. Alon, "A 50mW-TX 65mW-RX 60GHz 4-element phased-array transceiver with integrated antennas in 65nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf.*, San Francisco, 2013, pp. 234–235.

[12] D. Dregely, R. Taubert, J. Dorfmüller, R. Vogelgesang, K. Kern, and H. Giessen, "3D optical Yagi-Uda nanoantenna array," *Nature Commun.*, vol. 2, article no. 267, Apr. 2011. doi: 10.1038/ncomms1268.

[13] C. W. Berry, N. Wang, M. R. Hashemi, M. Unlu, and M. Jarrahi, "Significant performance enhancement in photoconductive terahertz optoelectronics by incorporating plasmonic contact electrodes," *Nature Commun.*, vol. 4, pp. 1610–1622, 2013.

[14] O. Momeni and E. Afshari, "High power terahertz and millimeter-wave oscillator design: A systematic approach," *IEEE J. Solid-State Circuits*, vol. 46, no. 3, pp. 583–597, 2011.

[15] R. Spence, *Linear Active Networks*. New York: Wiley, 1970.

[16] R. Han and E. Afshari, "A CMOS high-power broadband 260-GHz radiator array for spectroscopy," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3090–3104, 2013.

[17] C. M. Armstrong, "The truth about terahertz," *IEEE Spectr.*, vol. 49, no. 9, pp. 36–41, 2012.

[18] I. R. Medvedev, C. F. Neese, G. M. Plummer, and F. C. De Lucia, "Submillimeter spectroscopy for chemical analysis with absolute specificity," *Optics Lett.*, vol. 35, no. 10, pp. 1533–1535, 2010.

[19] C. F. Neese, I. R. Medvedev, G. M. Plummer, A. J. Frank, C. D. Ball, and F. C. De Lucia, "Compact submillimeter/terahertz gas sensor with efficient gas collection, preconcentration, and ppt sensitivity," *IEEE Sensors J.*, vol. 12, no. 8, pp. 2565–2574, 2012.

[20] C. Wang and R. Han, "Dual-terahertz-comb spectrometer on CMOS for rapid, wide-range gas detection with absolute specificity," *IEEE J. Solid-State Circuits*, vol. 52, no. 12, pp. 3361–3372, 2017.

[21] K. Schmalz et al., "245-GHz transmitter array in SiGe BiCMOS for gas spectroscopy," *IEEE Trans. THz Sci. Technol.*, vol. 6, no. 2, pp. 318–327, 2016.

[22] N. Sharma et al., "200-280GHz CMOS RF front-end of transmitter for rotational spectroscopy," in *Proc. Symposia VLSI Technology and Circuits*, 2016, pp. 1–2.

[23] HITRAN Online: The HITRAN database. Accessed on: Oct. 1, 2018. [Online]. Available: http://hitran.org/

[24] C. Wang and R. Han, "Molecular detection for unconcentrated gas with ppm sensitivity using dual-THz-comb spectrometer in CMOS," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 3, pp. 709–721, 2018.

[25] C. H. Chen et al., "Design of portable mass spectrometers with handheld probes: Aspects of the sampling and miniature pumping systems," *J. Amer. Soc. Mass Spectrometry*, vol. 26, no. 2, pp. 240–247, 2015.

[26] C. Wang, X. Yi, J. Mawdsley, M. Kim, Z. Wang, and R. Han, "An on-chip fully-electronic molecular clock based on sub-THz rotational spectroscopy," *Nature Electron.*, vol. 1, no. 7, pp. 421–427, 2018.

[27] S. Knappe et al., "A microfabricated atomic clock," *Appl. Phys. Lett.*, vol. 85, no. 9, pp. 1460–1462, 2004.

[28] R. Lutwak et al., "The chip-scale atomic clock: Prototype evaluation," in *Proc. 39th Annu. Precise Time and Time Interval (PTTI) Meeting*, 2007, pp. 269–290.

[29] C. Wang et al., "Chip-scale molecular clock," *IEEE J. Solid-State Circuits*, to be published. doi: 10.1109/JSSC.2018.2880920.

[30] C. Wang, X. Yi, M. Kim, Y. Zhang, and R. Han, "A CMOS molecular clock probing 231.061-GHz rotational line of OCS with sub-ppb long-term stability and 66-mW DC power," in *Proc. Symposia VLSI Technology and Circuits*, Honolulu, HI, 2018, pp. 113–114.

[31] E. Öjefors, U. R. Pfeiffer, A. Lisauskas, and H. G. Roskos, "A 0.65 THz focal-plane array in a quarter-micron CMOS process technology," *IEEE J. Solid-State Circuits*, vol. 44, no. 7, pp. 1968–1976, 2009.

[32] R. Han et al., "A 280-GHz Schottky diode detector in 130-nm digital CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 11, pp. 2602–2612, 2011.

[33] H. Sherry et al., "A 1kpixel CMOS camera chip for 25fps real-time terahertz imaging applications," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2012, pp. 252–253.

[34] Z. Hu, C. Wang, and R. Han, "Heterodyne sensing CMOS array with high density and large scale: A 240-GHz, 32-unit receiver using a de-centralized architecture," in *Proc. IEEE Radio Frequency Integrated Circuits Symp.*, Philadelphia, PA, 2018, pp. 252–255.

[35] C. Jiang et al., "A fully integrated 320 GHz coherent imaging transceiver in 130 nm SiGe BiCMOS," *IEEE J. Solid-State Circuits*, vol. 51, no. 11, pp. 2596–2609, 2016.

[36] Z. Hu, C. Wang, and R. Han, "A 32-unit 240-GHz heterodyne receiver array in 65-nm CMOS with array-wide phase locking," *IEEE J. Solid-State Circuits*, to be published.

[37] K. Takano et al., "A 105Gb/s 300GHz CMOS transmitter," in *Proc. 2017 IEEE Int. Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, 2017, pp. 308–309.

[38] K. K. Tokgoz et al., "A 120Gb/s 16QAM CMOS millimeter-wave wireless transceiver," in *Proc. IEEE Int. Solid-State Circuits Conf.*, San Francisco, CA, 2018, pp. 168–170.

[39] J. Holloway, L. Boglione, T. Hancock, and R. Han, "A fully integrated broadband sub-mmwave chip-to-chip interconnect," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2373–2386, 2017.

[40] C. Yeh, F. Shimabukuro, and P. H. Siegel, "Low-loss terahertz ribbon waveguides," *Appl. Opt.*, vol. 44, no. 28, pp. 5937–5946, 2005.

[41] Y. Ye, B. Yu, X. Ding, X. Liu, and Q. J. Gu, "High energy-efficiency high bandwidth-density sub-THz interconnect for the 'Last-Centimeter' chip-to-chip communications," in *Proc. IEEE MTT-S Int. Microwave Symp.*, 2017, pp. 805–808.

[42] N. V. Thienen, Y. Zhang, and P. Reynaert, "Bidirectional communication circuits for a 120-GHz PMF data link in 40-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 53, no. 7, pp. 2023–2031, 2018.

[43] J. W. Holloway, L. Boglione, T. M. Hancock, and R. Han, "A fully-integrated broadband THz chip-to-chip interconnect," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2373–2386, 2017.