IEEE Open Journal of the
Solid-State Circuits Society

Received 14 July 2021; revised 28 August 2021; accepted 1 September 2021. Date of publication 8 September 2021;
date of current version 22 September 2021.

Digital Object Identifier 10.1109/OJSSCS.2021.3110748

Terahertz Integrated Circuits and Systems for High-Speed Wireless Communications: Challenges and Design Perspectives

PAYAM HEYDARI (Fellow, IEEE)

(Invited Paper)

Nanoscale Communication Integrated Circuits Labs, Center for Pervasive Communication and Computing,
University of California at Irvine, Irvine, CA 92697, USA
CORRESPONDING AUTHOR: P. HEYDARI (e-mail: jssc.heydari@gmail.com)

This work was supported in part by the Faculty Innovation Fellow award sponsored by UC-Irvine Beall Applied Innovation.

ABSTRACT This paper presents challenges and design perspectives for terahertz (THz) integrated circuits and systems. THz means different things to different people. From International Telecommunication Union (ITU) perspective, THz radiation primarily means frequency range from 300 − 3000 GHz. However, recently, a more expansive definition of THz has emerged that covers frequencies from 100 GHz to 10 THz, which includes sub-THz (100 − 300 GHz), ITU-defined THz frequencies. This definition is now commonly used by communication theorists, and since this paper is intended for people with a wide variety of expertise in system and circuit design, we have adopted the latter definition. The paper brings to the open unmitigated shortcomings of conventional transceiver architectures for multi gigabit-per-second wireless applications, unfolds challenges in designing THz transceivers, and provides pathways to address these impediments. Furthermore, it goes through design challenges and candidate solutions for key circuit blocks of a transceiver including front-end amplifiers, local oscillator (LO) circuit and LO distribution network, and antennas intended for frequencies above 100 GHz.

INDEX TERMS Terahertz (THz), mm-Wave, high-speed, transceiver, CMOS, SiGe BiCMOS, wireless communications, 5G, 6G, next-G, MIMO, phased-array.

I. INTRODUCTION

THE EXPANDED definition of terahertz (THz) band from 100 GHz −10 THz has emerged as part of the professional and public consciousness due to emergence of exciting applications including active and passive sensing/imaging as well as forthcoming generations of high data-rate wireless communications [1]. In the area of wireless communications, which is the scope of this article, mobile networks with nomadic distributed base-stations using unmanned aerial vehicles (UAVs) are expected to become progressively more prevalent in the future society as complementary part of ever-evolving wireless networks, connecting billions of people across the globe and an even higher number of immobile/mobile cyber devices scattered in the environment. At the same time, the data rate supported by mobile devices keeps increasing with the deployment of next-generation networks, e.g., 5G, and upgrades of existing infrastructures. Consequently, enormous amounts of data traffic will be generated on a daily basis and data exchange between base stations and the backbone network through conventional backhaul links will quickly become a bottleneck. On the other hand, the operation of mobile terminals continues to be hindered by ever-increasing interference problems in a congested environment.

Fig. 1, from [2], shows the data-rate growth over the years for three communication protocols, namely, cellular, WLAN, and short wireline links, projecting linear growth fueled by the user demand.

This plot implies that continuing growth of world's population together with worldwide access to Internet and general public tendency to use bandwidth-intensive applications are major driving force for enhancement (and revamping) of wireless infrastructure so as to meet these demands. Indeed, COVID-19 pandemic outbreak in 2019-2022 period and the explosion of on-line video-communication services has only accelerated this demand. The end-users need for wider

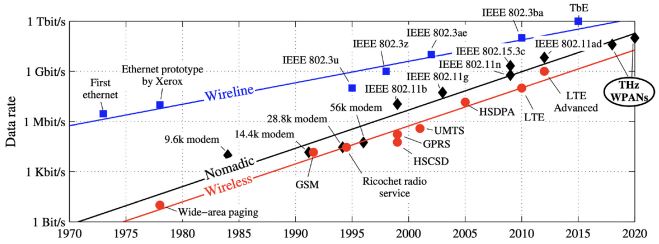


FIGURE 1. Data rate vs. the year for three communication protocols, short wired links, wireless LAN, and cellular communication [2].

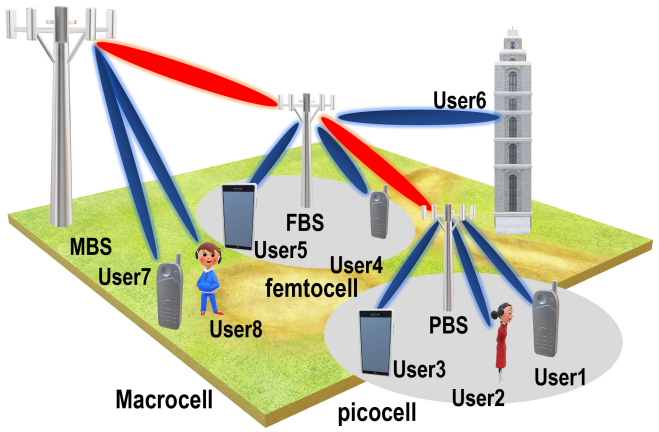


FIGURE 2. A generic distributed network comprising small cells and local base-stations that can also act as relay nodes.

bandwidth calls for a paradigm shift in the way wireless infrastructure is being designed and deployed to enable spectrally efficient wireless communication [2], [3]. Along with the demand for more bandwidth comes the desire for more computing and storage resources provided by large-scale data farms. The wired networks in such data-centers face severe over-subscription and hot-spot problems [4], [5]. On-demand flexible wireless links greatly alleviate these issues.

One way of conforming with the need for reliable high data-rate connectivity involves the deployment of distributed base stations with massive number of antennas ($>$100) providing high-speed wireless access to multi co-channel users (Fig. 2). Considered as an alternative solution to the currently used centralized wireless networks, each distributed base station in this scenario acts as a relay node within a large relay network, as shown in Fig. 2. To establish a point-to-point or point-to-multi-point line-of-sight wireless link, the relay nodes should exchange large data volumes rapidly, justifying an essential need for 50+ Gbps transceivers.

The use-scarcity of mm-Wave frequencies from 30- to 300−GHz, especially high side of this range from 100− to 300−GHz has motivated research and development teams across the globe to investigate future wireless communication networks for data rates beyond what is achievable by 5G [6]–[37]. At the first glance, it may appear that operating in the THz band should alleviate an important design concern associated with conventional wireless links, namely, a demand for very sophisticated modulation schemes (e.g., 1024 or higher-order quadrature amplitude modulation (QAM)) in order to boost the communication speed at the commonly used RF carrier frequencies (i.e., $\leq$ 10 GHz). As an example, binary phase-shift keying (BPSK) modulation of high-speed data with 5−GHz bandwidth on a 120−GHz carrier frequency can potentially yields 5 Gbps data rate. This means that operation in THz frequency range can provide wide RF spectrum with a fairly small fractional bandwidth (e.g., $\sim$10%), which is quite attainable by integrated transceivers fabricated in standard silicon technologies.

Although operation at even higher frequencies (e.g., above 300 GHz) can offer wider bandwidth and thus higher channel capacity, however, the limited transistor's maximum operation frequency (f_{max}) sets a performance upper-bound. Heuristically, frequencies up to $\approx f_{max}/2$ would be a range where the active devices exhibit sufficient gain and generate acceptable RF power, as will be discussed in Section II-B. Referring to the performance of the most advanced commercially available silicon (Bi)CMOS transistors, this range lies somewhere within 100−300 GHz frequency range [38]–[40]. This assertion indicates that the use of modulation schemes offering high spectral efficiency in conjunction with a wideband radio/modem should instigate a pathway toward deployment of high data-rate THz transceivers. Even though high-speed wireless radios using conventional homodyne or heterodyne architectures have been disclosed lately [6]–[19], [24]–[37], [41], the inputs/outputs of these radios are still in the form of modulated baseband or intermediate frequency (IF) signals. To procure raw information bits, high-speed and high-resolution mixed-signal blocks are needed. The sampling rates of these data converters need to be at least two or four times the baud-rate of the modulated baseband and IF signals, respectively, so as to ward off aliasing issue. As will be explained in Section II-D, reducing the complexity of data converters and back-end DSP is of utmost importance, as it facilitates low-power and cost-effective high-speed wireless links for mass consumer market. One can argue that designing an integrated ultra-high data-rate (e.g., above 50 Gbps) wireless transceiver would be practically impossible due to excessive amount of power - as high as 10 W - consumed by data converter and baseband units unless totally new architecture-level solutions are explored. This power consumption problem will only exacerbate if a multi-antenna architecture rather than a single-element transceiver at 100+ GHz is to be designed.

One effective way of boosting the data-rate, link reliability, co-channel user service, and combating path loss is to employ multi-antenna architectures [42]–[46]. Increasing the number of antennas results in channel hardening and reduction of small-scale fading (less multi-path and Doppler spread), which in return simplifies baseband signal processing algorithms. Various configurations of multi-antenna architectures provide: (1) multiplexing gain to enhance link capacity through concurrent transmission of parallel

data/user streams, (2) diversity gain to improve reliability of wireless links especially in non line-of-sight (NLOS) scenarios through transmission of copies of the same data stream, (3) antenna gain to combat path loss, integrated wide-band noise, and co-channel interference through breamforming in LOS or directed NLOS scenarios. The modern multi-antenna system should provide multi-functionality beyond beamsteering or signal-to-noise-ratio (SNR) improvement of a conventional phased-array system. We will discuss this notion, in more details, in Section III-A. This paper makes an attempt to study THz transceivers from both system- as well as circuit-level perspectives.

II. KEY PERFORMANCE PARAMETERS

When it comes to the design of THz transceivers, three major performance parameters should be taken into consideration. Operation in THz frequencies offers huge untapped frequency bandwidth. However, increasing bandwidth per user incurs several design challenges, which will be illustrated in Section II-A. Link reliability, coverage, and throughput are important parameters that mandate careful attention at THz frequencies. Section II-B will briefly go through the concept of multi-antenna communication as an effective way of increasing throughput, reliability, and coverage and its design challenges. Moreover, wireless communication at high frequencies should combat propagation loss, which increases exponentially with frequency. Section II-C will discuss the communication range and its associated design challenges. The communication speed and link performance can further improve with the aid of higher complexity digital modulation schemes. Section II-D will briefly discuss design challenges associated with high order modulation for THz transceivers.

A. BANDWIDTH INCREASE AND DESIGN CHALLENGES

Shannon theory predicts that wider bandwidth linearly increases the channel capacity in the bandwidth-limited regime, and has negligible impact on capacity in the power-limited regime. While bandwidth increase would be a straightforward way to boost data rate, other factors constrain its benefit when one accounts for challenges of wideband transceiver design. A wideband design requires the transmitter/receiver RF chains to satisfy high performance over wide bandwidth. For transmitter side, this includes satisfying high gain, high power and efficiency, high linearity, low error vector magnitude (EVM). Likewise, the receiver front-end should demonstrate low receiver sensitivity, low noise and high gain/linearity over wide bandwidth.

It is noteworthy that RF design is often revolved around narrowband operation about a tuned center frequency. Exceeding above 20% fractional bandwidth calls for circuits that should employ high-order bandpass matching circuits. Shown in Fig. 4(a)–(b) are examples of a narrowband tank and wideband double-tuned circuits, respectively, with their respective magnitude and phase responses. Clearly, the phase response of the double-tuned circuit is highly nonlinear

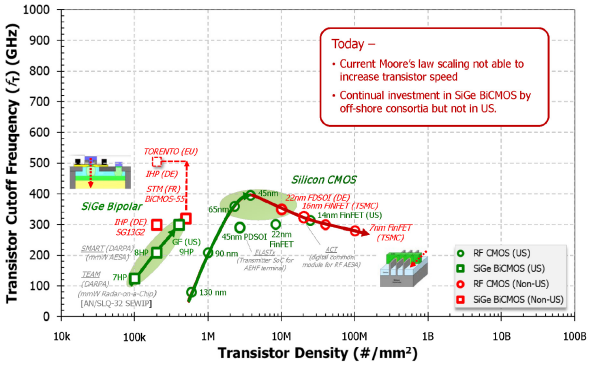


FIGURE 3. Transistor cutoff frequency vs. transistor density (*Courtesy of DARPA*).

compared to that of the narrowband tank circuit particularly when the two resonance frequencies are far away from one another for wider bandwidth. The phase response can no longer be assumed linear or constant, thereby introducing phase distortion. Adding to these challenges is the notion that constituent active devices within a wideband circuit exhibit frequency-dependent characteristic and nonlinearity, which leads to large distortion and in-band noise integration (and thus, SNR degradation) over a wide bandwidth. Therefore, circuit design with large fractional bandwidth more than 30% may not be a proper design strategy for data-rate increase.

One obvious way of keeping the fractional bandwidth within this range, while boosting the absolute bandwidth, is to increase carrier frequency. Besides, achieving wide bandwidth (while maintaining a relatively small fractional bandwidth), increasing the carrier frequency leads to smaller passive size and dimension. Most notably, the antenna size and spacing will decrease, making it possible to design multi-antenna architectures with large array size that improve diversity and spatial multiplexing. Despite advantages, one cannot keep increasing the carrier frequency due to a number of reasons related to principle of propagation and silicon technology limitation. Transceivers implemented in III-V semiconductor technologies with $f_{max} > 1$ THz have achieved ultrahigh data rates at low-THz range of frequencies [12]. Nonetheless, these technologies are not deemed suitable for integrated systems incorporating large antenna array due to low yield and integration density as well as high fabrication cost. Silicon technologies, on the other hand, present much higher level of integration and may be considered as platform of choice for mass-marketing of ultra-high-speed transceivers. CMOS/BiCMOS transmitters/receivers operating in THz band have been demonstrated by prior work [12], [16], [18]–[21], [23]. However, increasing the carrier frequency is limited by the device f_{max}. It should also be pointed out that MOS transistor's cutoff frequency does not keep increasing with device scaling and seems to be peaking at 45 nm feature size, as has been clearly indicated in the plot of Fig. 3. Importantly, Fig. 3 also reveals that SiGe BiCMOS, with its cutoff frequency gracefully increasing with device density, seems to be a better technology

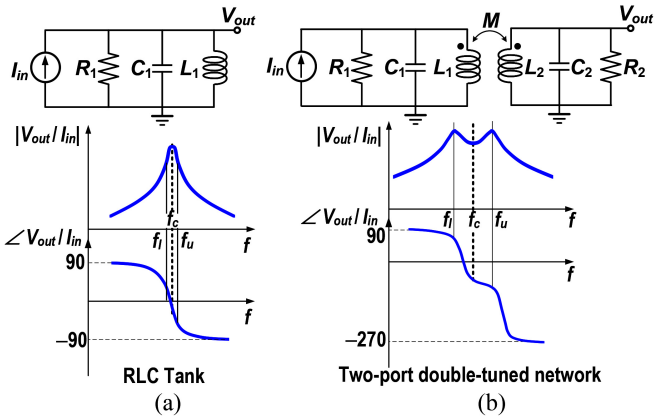


FIGURE 4. (a) Narrow-band tank circuit and its magnitude and phase responses. (b) High-order double-tuned network and its magnitude and phase responses.

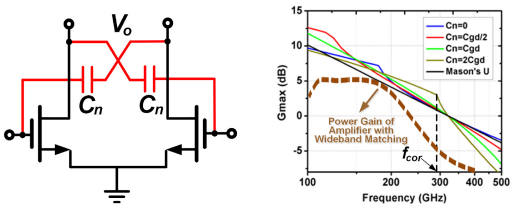


FIGURE 5. Simulated device G_{max} vs. frequency for different values of neutralizing capacitors, and the frequency response of an above 100−GHz amplifier.

for integrated THz transceiver compared to a silicon CMOS process.

Based on what has been stated, operation around and below $f_{max}/2$ is considered to be a sweet spot for THz wireless systems so that high performance (e.g., high-gain, low-noise, and high output power and power-added efficiency) can still be achieved. Given an f_{max} of around 350–500 GHz for commercially available silicon technologies, frequency range around 100–200 GHz could be spectral range of interest for tens of Gbps wireless speeds.

Moreover, designing THz transceivers with 20-30% fractional bandwidth would entail yet another challenge, which is best comprehended by looking at a frequency response of a neutralized 32-nm SOI differential-pair device for different C_n values and its Mason's U [16] in Fig. 5. In light of the degrading magnitude response, a matching circuit of higher-order is needed to establish a flat frequency response over the bandwidth of interest. A more detailed study laying down steps to design a wideband THz amplifier will be provided in Section IV-A.

B. INCREASE IN THROUGHPUT/RELIABILITY AND DESIGN CHALLENGES

Increasing the capacity and reliability of wireless communications systems through the use of multiple antennas, first discovered in [47], has been an active area of research for the past 30 years. The well-known MIMO capacity of an N-element transceiver is [48]:

$$C = BW \log_2 \det\left[\mathbf{I}_N + \frac{E_s}{N \times N_0 BW}\mathbf{H}\mathbf{H}^H\right] \quad (1)$$

where C denotes the capacity, $\mathbf{I}_N$ is the identity matrix of size N, det [.] indicates the matrix determinant, and $\mathbf{H}$ is the channel transfer function, representing transfer functions h_{ij} from the j^{th} transmit antenna to the i^{th} receive antenna. It is noteworthy that multi-antenna architectures can also be used to obtain array and diversity gain in addition to capacity gain. Diversity combining exploits the fact that independent signal paths have a low probability of experiencing deep fades simultaneously. Thus, the idea behind diversity is to send the same data over independent fading paths. These independent paths are combined in such a way that the fading of the resultant signal is reduced, leading to higher reliability of communication. For an N-element multi-antenna transceiver, a maximum diversity gain of N^2 can be achieved and the average probability of error decreases with $1/SNR^{N^2}$ [49], thereby improving EVM. The fading problem for line-of-sight communication established by THz transceivers is not an issue, thus, the diversity-gain attribute may not be as essential as in RF frequencies.

In low-SNR THz channels, increasing the capacity is limited by both the transmitter output power and receiver integrated noise. Adopting beamforming multi-antenna architectures to transmit sharp beams with highly directional antenna gains enhances the capacity by improving the SNR, thus making it possible to employ high-order modulation schemes to achieve higher spectral efficiency. While the SNR improves directly with array size, the beam also becomes increasingly more directive, leading to narrow beam antenna pattern.

On the other hand, in high-SNR THz channels with high diversity or rank order, exploiting multiplexing gain via propagation of independent signal streams through multiple distinct paths in different spatial and polarization domains can further enhance the channel capacity or multi-user service. In general, knowing the channel state information on the transmit side (CSIT), one can find optimum power allocation across antennas. To approach the capacity limit, a knowledge of CSIT is required. CSI acquisition can, however, be very costly. Furthermore, increasing the number of antennas for mm-Wave channels requires expensive spectral resources during CSI determination. At mm-Wave frequency, CSI error due to pilot contamination is highly suppressed and CSI can be acquired based on ray-tracing model by estimation of AOD (angle of departure), AOA (angle of arrival), and paths gains. Additionally, mm-Wave channel exhibits spatial/angular sparsity and the number of resolvable paths for both indoor and outdoor communication is very low (i.e., less than four) resulting in a low-rank channel response matrix. Therefore, due to low number of detectable paths, parameterized techniques such as interference cancelation (interference of clusters on each other) in addition to MUSIC algorithm [50] can be utilized to distinguish between

multi-paths (parallel data streams). Thus, DOA (direction of arrival) and LS (least squared) methods can be used to estimate paths directions and paths gains, respectively.

At THz frequencies, a single-element wireless link operates in the power-limited or low SNR regime. As mentioned above, to foresee the advantages of MIMO spatial multiplexing on capacity increase, we can employ beamforming to form beams, and thus, increasing the SINR. This requires large transceiver array, and hence the notion of massive MIMO. Indeed, as N grows to be large value (e.g., 128-element array), the MIMO channel capacity in the absence of CSIT approaches $C = N \times BW \log_2 [1+E_s/(N \times N_0 BW)]$ and hence grows linearly with N. One can leverage the benefits of both spatial multiplexing and beamforming concurrently, as will be discussed in Section III-A. For example, it is possible to use multiple beams, where each beam employs beamforming to increase SNR in power-limited situations, while also providing unique data streams on each of the beams using the same carrier frequency [51]. As mentioned above, beamforming leads to highly directive radiation, which makes the transmitter-receiver re-alignment a challenging task.

C. COMMUNICATION RANGE AND DESIGN CHALLENGES

Communication range, R, is markedly affected by carrier-frequency scaling in multiple ways. First, the path loss, PL, increases at a rate proportional to square of frequency, f, and R. The received signal power P_{RX} predicted by the Friis transmission equation (2) is degraded by the path loss and polarization mismatch.

$$P_R = PLF \times \frac{P_T G_T G_R c^2}{(4\pi R f)^2} \tag{2}$$

where P_R, P_T, G_R, G_T denote the received/transmitted powers and received/transmitted antenna gains, respectively. c is the speed of light and PLF is the antenna polarization loss factor. Although increasing transmit power and transmit/receive antenna gains help increase the range, each comes with its won constraints and limitations.

To understand the underlying challenges behind increasing transmit power, suppose the last power amplification stage employs a differential neutralized circuit surrounded by input (output) matching networks with impedance transformation ratios of n_{in} (n_{out}) and the power losses L_{in} (L_{out}), as shown in Fig. 6. The saturated output power, P_{sat}, of this stage is readily calculated, as follows:

$$P_{sat} = \frac{n_{in}^2 V_{1dB,\max}^2}{8\mathrm{Re}[Z_s]} \frac{G_{max}}{L_{in}} \tag{3}$$

where G_{max} denotes the maximum available power gain (MAG), which, for a neutralized device holding unconditional stability with K-factor $K_f > 1$, is equal to $G_{max} = |Y_{21}/Y_{12}|(K_f - (K_f^2-1)^{0.5})$. G_{max} for a neutralized differential

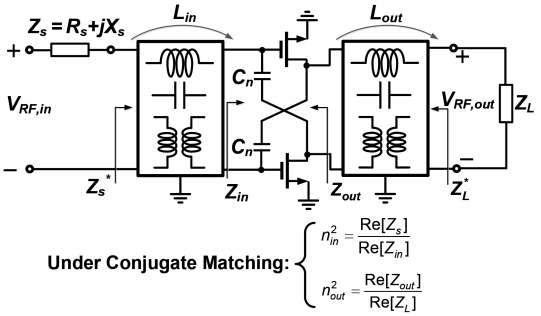


FIGURE 6. A general schematic of a neutralized differential stage with input/output matching networks.

pair in terms of device f_T and f_{max} is:

$$G_{max} \approx \frac{\left(\frac{f_T}{f}\right)\sqrt{\frac{f^2}{f_T^2}\left(\frac{4f_{max}^2}{f^2}-1\right)^2 + \left(1+\frac{C_{gs}}{C_{gd}}\right)^2}}{K_f + \sqrt{K_f^2-1}} \tag{4}$$

where $f_T = g_m/[2\pi(C_{gs} + C_{gd})]$, $f_{max} = (1/4\pi)(g_m/[r_g C_{gd}(C_{gs} + C_{gd})])^{0.5}$. Equation (4) is still valid for slightly over-neutralized device when C_n is slightly bigger than C_{gd}. P_{1dB} in (3) for a quasi-differential pair, like the one in Fig. 6, assuming short-channel MOS model, is [52]:

$$V_{1dB,\max} = \frac{2\sqrt{1-10^{-1/20}}}{3} \frac{1}{\frac{\mu_0}{2v_{sat}L} + \theta} \tag{5}$$

where μ_0 is the low-field mobility, θ is the fitting parameter accounting for mobility degradation, and v_{sat} is the saturated velocity. Equations (3), (4), (5) present trade-offs between P_{sat}, operation frequency, and device dimension. Specifically, they imply that at high frequencies, the capability of the power amplifier to generate suffiently high P_{sat} is limited by lack of gain, G_{max}, and its linearity. On the other hand, linearizing the device by increasing the device channel-length adversely affects the MAG, thus compromising P_{sat}.

One way of increasing the transmit power, and thus increase the range, is to increase overall antenna gain though transceiver array. The amount of radiated power increase at the broadside of an N-element array could be as high as N times the single-element antenna [53]. This, however, comes at the cost of a highly directive radiation.

D. HIGH-ORDER DIGITAL MODULATION AND DESIGN CHALLENGES

Increasing the modulation complexity improves spectral efficiency, thereby resulting in higher data rate for a given specific bandwidth. If so effective, why not keep increasing the modulation complexity (e.g., 2048QAM, 4096QAM and etc) at 100+ GHz center frequencies? To better understand the underlying challenges, we should revisit the structure of modern transceivers handling high order modulation schemes. Modulation and demodulation in state-of-the-art transceivers are done in digital domain, which means

TABLE 1. High-performance DACs.

Ref.	Speed – Resolution	SFDR [dB]	BW [GHz]	Power [mW]	CMOS Node
[?]	56GS/s – 6b	34@28GHz	>40	750	65nm
[?]	64GS/s – 8b	41@14GHz	20	620	20nm
[?]	65GS/s – 8b	41@8GHz	>13	750	40nm
[?]	100GS/s – 8b	27@24.9GHz	>13	2500	28nm

TABLE 2. High-performance ADCs.

Ref.	Speed – Resolution	ENOB@ Nyquist	ENOB@ DC	Power [mW]	CMOS Node
[?]	64GS/s – 8b	5.3b	6.3b	950	20nm
[?]	72GS/s – 8b	4.8b	6.2b	235	14nm
[?]	56GS/s – 8b	5.2b@27.1GHz	–	702	28nm
[?]	128GS/s – 5b	4@32GHz	4.6b	1800	55nmSiGe

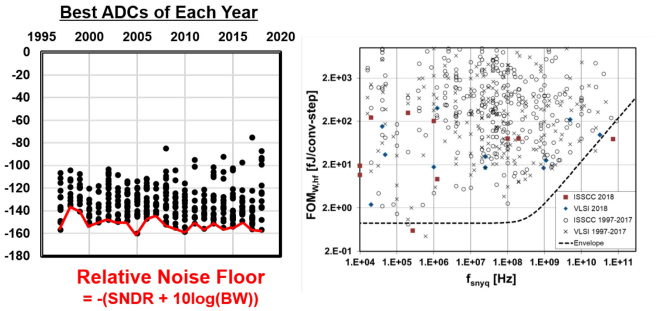


FIGURE 7. Latest CMOS ADC trend showing the relative noise floor of all published ADCs from 1995 to 2018.

the entire mixed-signal, analog baseband, and RF chain should be able to process modulated signals with high peak-to-average power ratio (PAPR) and dynamic range. Furthermore, realization of higher-order modulation requires (a) local oscillators with lower phase noise, (b) data converters with higher resolution, and (c) RF chain with high dynamic range (accounting for both high sensitivity and linearity), while operating at THz frequencies.

Generation and processing of high-speed high-order modulation particularly exert stringent requirement on the mixed-signal (i.e., analog-to-digital converter (ADC) on the receive side and digital-to-analog converter (DAC) on the transmit side). The sampling rate of a Nyquist-rate data converter is chosen to be 5- to 6-times the baud-rate, in practice, to improve SNR and bit error-rate. Furthermore, the required data-converter resolution is increased with the modulation complexity, which becomes increasingly more challenging to attain at higher data rates. For instance, a 16QAM receiver targeting a bit-error rate (BER) of 10^{-4} requires a minimum resolution of around 6 bits to capture degradation due to thermal noise and component mismatch. Assuming that this receiver is designed to operate at 50 Gbps, the sampling rate of the data converter, designed to operate at 5 times the baud rate, is 62.5 GS/s. To better appreciate the challenges of designing such data converter for a low-power transceiver, we have summarized state-of-the-art high-speed DAC and ADC performances in Tables 1 and 2. It is noteworthy that the power dissipation reported for all these data converter prototypes exclude the I/O and clock buffers, digital calibration and clock generation circuits. Moreover, these prototypes do not include on-chip memory. Nevertheless, the effective number of bit (ENOB) and power dissipation are considerably compromised at higher sampling speeds. All these requirements should be met at the reported 50+ Gbps data rate.

Besides trade-off between signal-to-noise+distortion ratio (SNDR) and speed, the data converter's power consumption represents another major issue for a THz transceiver, as power dissipation is a super-linear function of frequency. As an example, according to a recently published work [55] summarized in Table 2, the power consumption of a CMOS 64 GS/s ADC – excluding input/output interfaces and clock generator – with an ENOB of only 5.3 b (at the Nyquist rate) is 0.95 W.

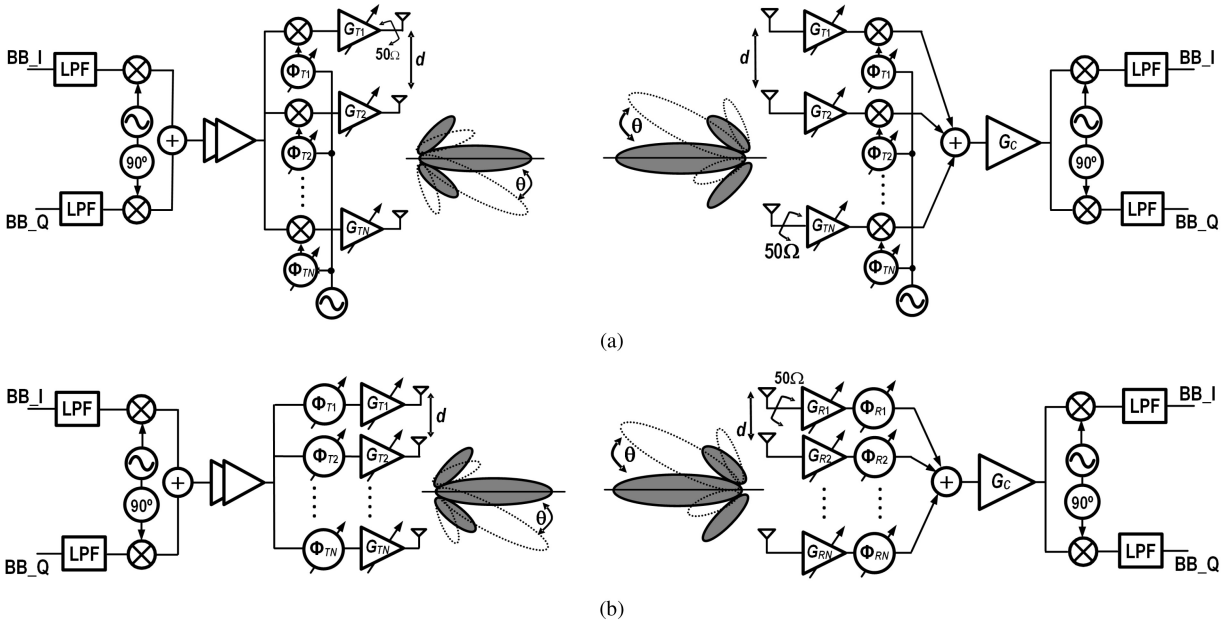
More importantly, a study in [61] predicts that power consumption of clock generators for ADCs will be increasing quadratically with their speed and resolution, thus becoming significant at the speed of interest. For example, the lower-bound of power consumption of a phased-locked-loop (PLL) clock generator fed by 50-MHz external crystal oscillator with an excellent phase noise of –150 dBc/Hz at 1 MHz offset for a 6-bit 62.5 GS/s ADC is around 1.8 W. Adding this value to the power dissipation of the core ADC, the mixed-signal block alone can consume multi-watts of power, rendering conventional direct-conversion or low-IF transceivers impractical for THz wireless communication. Depicted in Fig. 7 are comprehensive surveys of published (Bi)CMOS ADCs [62] in two forms. In the plot demonstrating relative noise floor achieved by the ADCs appeared in publications between 1995 to 2018, a lower-limit of –160 dBc can be identified. The plot of Walden figure-of-merit versus Nyquist frequency of the same groups of ADCs shows a lower limit for the power per conversion step which becomes worse with sampling rate. This explicitly means that higher sampling rate ADCs are incapable of achieving an arbitrarily high SNDR and ENOB. It is also noteworthy that technology scaling does not seem to mitigate this trade-off between bandwidth and SNDR.

III. ARCHITECTURE LEVEL TECHNIQUES

Section II elaborated that increasing bandwidth, modulation order, and transmit power face fundamental barriers. This section will go through two architectures – namely (a) multi-antenna systems and (b) transceivers implementing modulation/demodulation directly in analog/RF domains – that overcome issues and challenges discussed in Section II.

A. MULTI-ANTENNA ARRAY ARCHITECTURES

The smaller size of passive components at THz frequencies makes it possible to think of having integrated multi-antenna transceiver arrays. Improvement of capacity and reliability of wireless communication systems through the use of multiple antennas has been an active area of research for over 25 years. Discovered by Paulraj and Kailath [47], MIMO



(a)

(b)

FIGURE 8. **(a) Phased array based on LO phase shifting. (b) Phased array based on RF phase shifting.**

wireless systems are now part of current standards and have been widely deployed for public use [63].

Much of the research effort on multi-antenna transceiver design has been centered around beamforming through phased-array implementations [64]–[70]. Implementation of the first phased-array system in silicon incorporated the LO phase shifting [64]–[67], as shown in Fig. 8(a). The primary advantage of this architecture is that the phase-shifters are placed away from the RF path. Therefore, they process single-tone LO signals rather than an RF signal and their insertion loss will have no effect on the RF power. Alternatively, a phased-array transceiver with RF phase-shifting in Fig. 8(b) relaxes the LO distribution network, which can be a considerable design challenge for large array sizes [68]–[71]. Phased-arrays, however, offer limited features of a multi-antenna architecture such as improvement of capacity and diversity.

An all-digital beamforming system [72], [73], shown in Fig. 11(a), enables multi-beam communication with the highest adaptability and data-rate. The digital beamforming (DBF) approach offers three major advantages [73]: (1) high magnitude and phase resolution can be achieved by digital precoding. (2) A DBF array can be used to superpose multiple beams for several data streams, thereby resulting in higher capacity. (3) For multicarrier signals, such as orthogonal-frequency-division-multiplexing (OFDM) signals, the fully DBF architecture can realize independent beamforming precoding at each subcarrier or resource block to obtain extraordinary performance at a wide signal bandwidth. On the other hand, a digital beamforming array demands dedicated RF chains for each antenna element and high dynamic-range for RF front-end. Moreover, the need for additional signal processing to facilitate multi-beam transmission as well as interference management in a multi-user environment mandates the requirement of digital baseband precoding and combining.

A possible approach to address the complexity and excessive power consumption issues in a conventional MIMO system is the code modulated path sharing multi-antenna (CPMA) architecture [74], [75], in which code multiplexing is used to combine the signals emerging from multiple antennas into a single RF/IF/baseband/ADC path. The primary advantages of CPMA include (1) a significant reduction in area and power consumption, and (2) amelioration of crosstalks and power losses of large LO routing/distribution network used in massive-MIMO transceivers. Fig. 9 shows an exemplary block diagram of the CPMA transceiver applicable for a massive MIMO base-station [42]. The CPMA architecture subdivides the entire antenna array of size N into groups of M antennas, and uses code multiplexing to combine the M signals onto a single RF path. The individual signals are easily extracted using a code-demodulation in digital domain. Likewise, on the transmit side and prior to transmission, the signal for a given antenna is extracted using an RF code demodulator. The CMOS development and integration of this idea in [75] used mutually orthogonal Walsh-Hadamard code sequence due to its ease of implementation and its ability of achieving maximum MIMO capacity [74]. This M-fold reduction of RF/IF paths and ADC/DACs, however, at the price of M-fold increase in bandwidth and sampling-rate ADC/DACs if orthogonal codes are being used. It turns out that non-orthogonal codes can provide an acceptable trade-off between capacity and bandwidth [74]. Fig. 10 demonstrates the plot of capacity versus code correlation coefficient, ρ ($\rho = 0$ for orthogonal codes), of a 4×4 MIMO system four SNR values. It is evident from

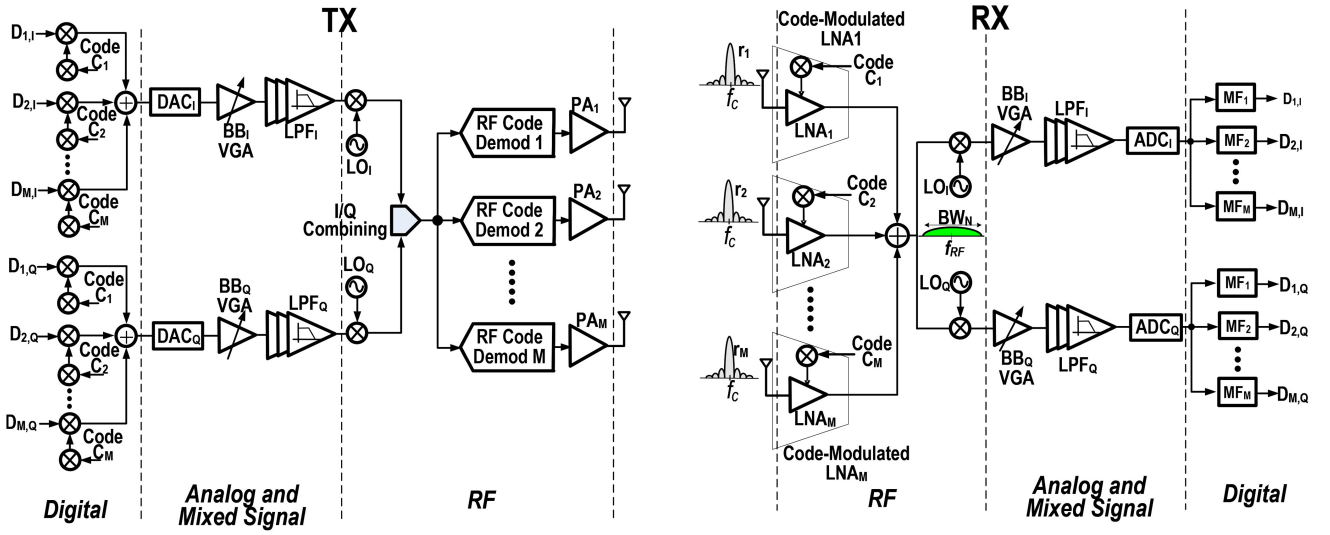


FIGURE 9. Conceptual block diagram of the code-modulated path sharing transceiver architecture.

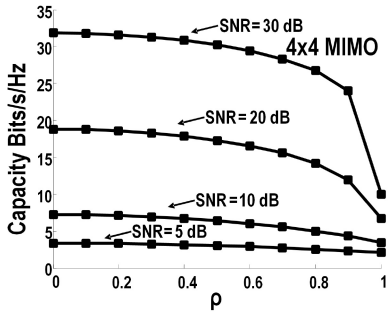


FIGURE 10. Capacity versus ρ for various SNRs [74].

the plots of Fig. 10 that at high SNR's the capacity is affected by ρ more substantially, whereas at low SNR's the capacity is noise limited and is less affected by ρ.

Alternatively, a hybrid architecture, depicted in Fig. 11(b), with both analog beamforming and digital MIMO coding has been pursued recently, as it reduces the complexity of the digital baseband with a smaller number of up/downconversion chains in systems with massive number of antennas, thereby emerging as a viable candidate for both outdoor and indoor mm-Wave/THz communication. The multi-beam digital baseband processing with analog beamforming facilitate both multiplexing and beamforming gain. Hybrid architectures can be designed to receive (or transmit) all data-streams from all antennas in Fig. 11(b) (when $N = N_{RF}$), or receive (or transmit) only a subset of data-streams, N_{RF} with $N_{RF} < N$, per each antennas leading to a sub-array system. In Fig. 11(b), the complex weighting coefficients are generally defined as $W_{i,k} = A_{ik}e^{j\phi_{ik}}$ for $i \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, N_{RF}\}$. ϕ_{ik} and A_{ik} are realized by RF phase shifters and variable gain attenuators and/or amplifiers (VGAs), respectively. The RF phase shifters are used for main-lobe steering, whereas the RF VGAs enable spatial filtering of interference by placing the null locations of each beamforming path toward the directions of the interference incident angles. The number of required RF chains N_{RF} in a hybrid architecture is strictly lower-limited by the number of parallel data streams K, while beamforming gain is determined by N_{RF} complex weighting coefficients emerging to each antenna in Fig. 11(b). In retrospect, a full-array realizes the function of an all-digital architecture. The number of signal processing paths (from the digital baseband to the antenna front-end) for the sub-array is equal to $N_{RF} \times N$ and for full-array is equal to N^2. On the other hand, beamforming gain of the sub-array is N_{RF}/N of the full-array. Therefore, a trade-off exists between signal processing complexity and beamforming gain of hybrid architectures. A recent circuit implementation of a hybrid architecture was presented in [76]. It utilizes Cartesian combining concept to enable 2-stream reception. One important consideration in this design is that its implementation requires 8 splitters, 20 combiners, and 12 mixers for a 2-stream reception. These large number of signal paths introduce electromagnetic cross-talks due to many cross-overs between these paths. Later, [77] unfolded a partially-overlapped beamforming-MIMO architecture capable of achieving higher beamforming and spatial multiplexing gains with lower number of elements compared to conventional architectures. Reference [77] showed that overlapping the clusters in an N-element hybrid allows us to allocate larger number of antennas per cluster, thereby resulting in higher beamforming gain compared to the corresponding N-element sub-array.

B. LOW-POWER DIRECT-RF-(DE)MODULATION TRANSCEIVERS

As discussed in Section II-D, research works in ultra-high speed transceivers have not addressed an essential question: what are the power efficient solutions for DAC/ADC and DSP components of integrated transmitter and receiver chipsets that can handle data rates above 50 Gbps? More

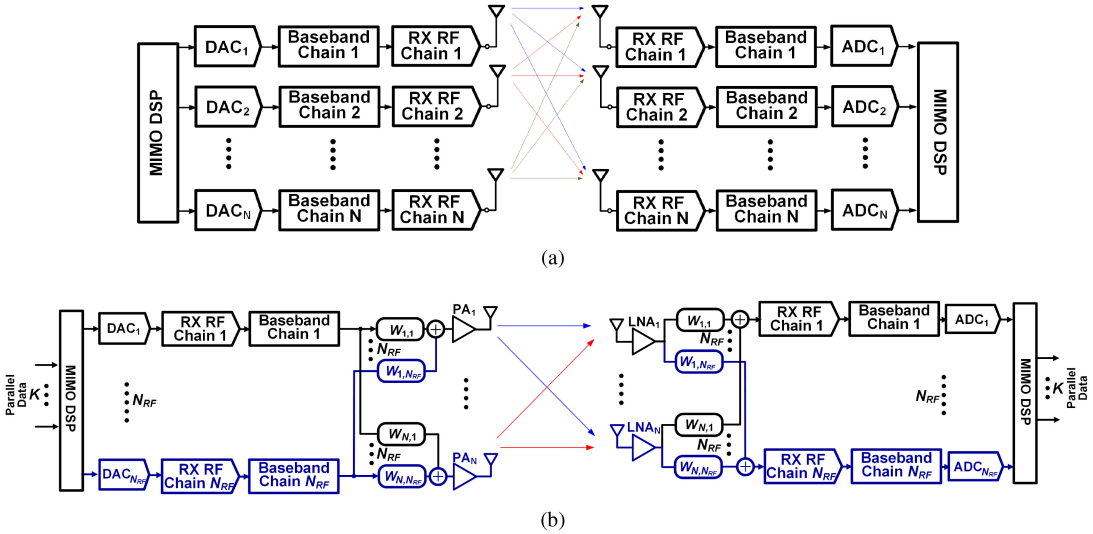


FIGURE 11. Beamforming MIMO architectures: (a) digital array and (b) hybrid-beamforming array.

precisely, in prior work targeting these applications, the entire back-end and mixed-signal processings are carried out by an expensive commercial Arbitrary Waveform Generator (AWG) and a real-time oscilloscope to generate high-power sub-channelized modulated signals off-chip feeding the transmit side, and to equalize/demodulate/synchronize the RF signal followed by extraction of the baseband stream on the receive side. A front-end with external AWG and real-time scope is certainly not practical let alone being a power-efficient solution. The situation will only become far more severe if we think of extending a single element transceiver to a multi-antenna architecture for the same ultra-high data-rate application domain. While the achieved data-rates by prior work (e.g., [12]) are impressive, there is no discussion on how to implement high-speed signal generation and (de)modulation. In fact, the mixed-signal design challenges for these systems are unresolved. Specifically, while transceivers incorporating higher order modulations operate at smaller RF bandwidth for a given data rate, they require significantly higher resolution and higher DAC/ADC sampling rate compared to the signal baud-rate.

References [20], [21], [23] entertained the idea of high-order (de-)modulation directly in RF domain. Shown in Fig. 12(a) is the block diagram of two commonly used transmitter architectures, i.e., direct conversion and heterodyne. Notable in both structures is the fact that (de)modulation should be handled by DSP. As pointed out in Section II-D, DAC and DSP, handling most of the signal processing including pulse-shaping and high-order modulation (e.g., 64QAM), should operate at 50+ GS/s for 100 Gbps wireless communication. Notwithstanding is the fact that DAC resolution increases with modulation order, adding higher complexity (and thus higher power dissipation). Similar issues will arise if conventional direct-conversation or heterodyne receiver schemes in Fig. 13(a) are used for high-speed wireless communication.

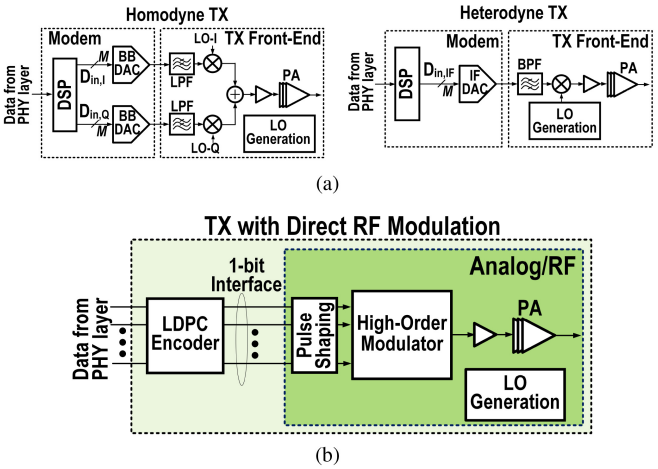


FIGURE 12. (a) Block diagrams of conventional direct-conversion and heterodyne transmitter architectures. (b) Block diagram of the proposed transmitter realizing high-order modulation directly in analog/RF domain.

On the other hand, delegating the modulator/demodulator function to the analog/RF domain, as indicated in 12(b), will lead to whole new generation of architectures that are amenable to higher speeds. In fact, assuming a power-efficient transmitter/receiver solution with such capability exists, great advantages readily come to fruition: (1) Power-hungry high resolution and high data rate ADC and DAC will be removed. (2) The complexity of the baseband blocks will be significantly relaxed.

1) DESIGN CHALLENGES: PULSE SHAPING, EQUALIZATION, CARRIER SYNCHRONIZATION

Pulse shaping is commonly employed in a wireless transmitter to mitigate the problem of excessive bandwidth. As stated in [23], pulse shaping with root-raised cosine (RRC) filters requires multi-bit resolution DAC with a sample rate of more than twice the baud rate. Putting aside the daunting

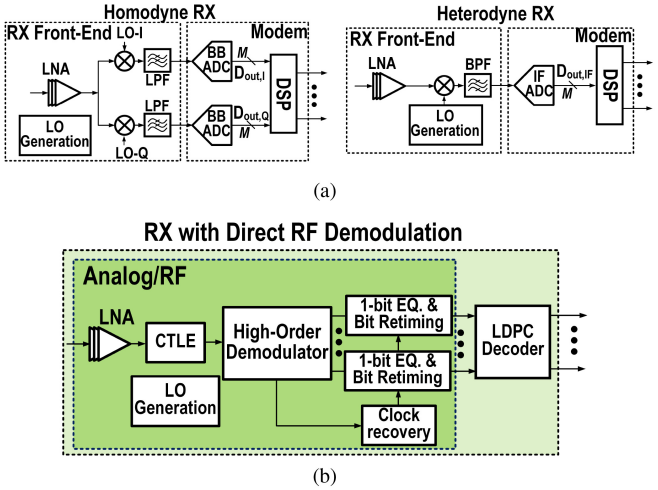


FIGURE 13. (a) Block diagrams of conventional direct-conversion and heterodyne receiver architectures. (b) Block diagram of the proposed receiver realizing high-order demodulation directly in analog/RF domain.

challenges of building such DACs and digital filters at ultra-high speeds, the fundamental role of pulse shaping is revisited first. Correlation detection, or equivalently matched filtering, is at the core of any communication system to maximize SNR before the decision making circuitry in the receivers. Any pulse shape (not limited to RRC) theoretically have the same SNR performance under the same noise power density N_0 [78] so long as pulse shapes are matched in the transmitter and receiver. This degree-of-freedom allows us to design a pulse-shaping filter in analog domain (*cf.* Fig. 12(b)).

Equalization in any type of communication modality is crucial, as the channel impairments severely degrade the BER. The bandwidth of free space channel is typically wide enough and the bandwidth limitation primarily comes from transmitter front-end. Therefore, circuit techniques that can achieve very flat frequency response across a wide bandwidth are the key in pushing the limit of data rate. In the context of the proposed direct modulation/demodulation-based transceiver architecture, analog equalizers and bit recovery/retiming circuits succeed the RF demodulator, as shown in Fig. 13(b), in a similar way as in broadband wireline receivers [79]. The clock recovery circuit in Fig. 13(b) generates the clock signal for symbol synchronization and retiming. One distinction worth mentioning is that, in the context of wireless systems, complex-domain equalization techniques may be necessary to account for the asymmetry in lower and upper sidebands induced by RF building blocks such as PA. All in all, design considerations on equalization may not be one-dimensional when multiple practical limitations are accounted for. Similar to a wireline transceiver, we can employ both transmitter side and receiver side equalization. On the transmit side, a low-order feedforward equalizer can be used to equalize channel and attenuate pre-cursor intersymbol interference (ISI). On the RF-demodulation-type receiver side handling high-order modulation, the bit extraction and recovery following I/Q downconversion will be

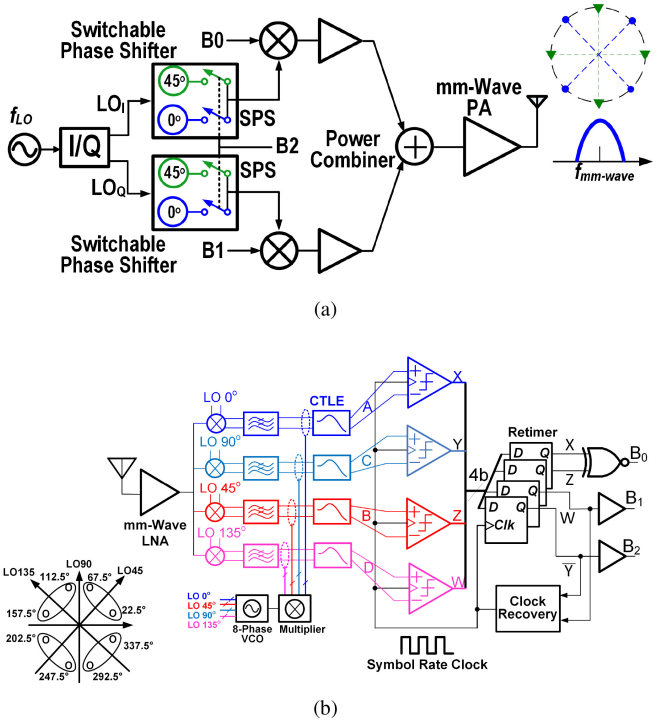


FIGURE 14. (a) Conceptual block diagram presenting RF-8PSK modulation concept and simplified direct RF-8PSK modulation transmitter [21]. (b) The signal space of 8PSK symbols partitioned using 8 LO phases, and conceptual block diagram of an RF-8PSK receiver incorporating multi-phase RF-correlation [20], sign-check comparison, carrier synchronization, and symbol recovery.

performed on a multi-level PAM (PAM-M) signal [20]. This means the equalization should be able to improve horizontal and vertical eye-opening, which indicates that an M-step bit-by-bit equalization can be a plausible method.

2) CASE STUDIES

We briefly go through three case studies, namely, and RF-8PSK transmitter, an RF-8PSK receiver, and an RF-QAM transmitter.

RF-8PSK transmitter: Aiming for a bits-to-RF transmitter and an RF-to-bits receiver that can overcome the aforementioned challenges, we investigate direct RF modulation and demodulation, which has so far been employed only for OOK and QPSK schemes, to construct higher order modulations. One immediate extension would be an RF-8PSK modulator and demodulator. To do so, we start with a QPSK modulator and introduce additional level of phase modulation to the QPSK output so as to create two versions, a QPSK and an Offset-QPSK constellation, and enable one of these schemes based upon the status of the third bit (*cf.* Fig. 14(a)) [21]. To avert wideband RF phase-shifters at THz frequencies (e.g., 170 GHz in [21]) and mitigate their significant insertion loss in the RF path, this additional phase-modulation is deployed in the LO path and placed prior to quadrature mixers, as shown in see Fig. 14(a). Specifically, two switchable phase-shifters (SPSs) realized by passive all-pass filters vary the phase of both I and Q signals to construct 8PSK modulation

in RF domain. The I/Q SPS phase-shifts are controlled by the third input bit stream, B_2, to take on one of two values, 45° or 0°. Fig. 14(a) constitutes the core of a bits-to-RF transmitter presented in [21]. Here, to generate higher local oscillator power, we used two separate SPSs, although a single SPS can alternatively be employed prior to I/Q generation circuit.

RF-8PSK receiver: Likewise, 8PSK demodulation can be carried out directly in RF/analog domain, which mitigates the use of ultra-high-speed high-resolution data converters and sophisticated digital demodulation using off-the-shelf FPGA or backend DSP [20]. Shown in Fig. 14(b) is the conceptual block diagram of the RF-8PSK receiver. Fundamentally based on an advanced version of a direct conversion scheme, this receiver employs an 8-PSK demodulator, which is directly realized in RF/analog domain, thereby obviating the need for ultra-fast DSP and mixed-signal building blocks. The RF demodulator employs four correlation-based detectors driven by multi-phase LO signals with 45° phase differences. The decision circuitry is comprised of ultrahigh-speed comparators and simple logic circuits, only. The RF correlator is essentially a mixer followed by a lowpass filter (LPF). As shown in Fig. 14(b), the 2-D signal space is partitioned into eight angular areas, where each symbol is located in the middle and has maximum Euclidean distance toward the boundaries. To achieve this optimum detection, the received signal phase reference and the LO phase are purposely offset by 22.5°. As a result, the error tolerance in detecting the symbols will be maximized. Only the polarity of LPF output is needed to determine symbols. Assuming Gray-coding for the three-bit symbols, we can easily decode the bits with much simpler hardware. This will help us explore low-complexity, yet high-speed, baseband circuits which can resolve the symbols. The three bits B_2, B_1, B_0 of each symbol can easily be extracted with simple logic circuits from the re-timed outputs of the RF correlators. The bits are easily derived from sign-check comparator outputs, i.e., $B_2 = \bar{Y}$, $B_1 = W$, and $B_0 = X \bar{\oplus} Z$. There is no need for any explicit ADC in this design, and symbols are detected using only sign-check comparators performing essentially BPSK decision and simple logic functions [20]. The carrier synchronization is achieved using an extension of a Costas loop, whereby the downconverted signals out of multi-phase correlators are low-pass-filtered, and the outputs are then fed to multipliers to detect the phase and appropriately adjust the voltage-controlled oscillator (VCO) phase. The symbol and clock recovery are accomplished using a clock recovery circuit and data retimers, similar to the way is done in high-speed wireline receivers.

RF-4^{M}QAM transmitter: QPSK was observed by prior work to be amenable to analog implementation at ultra-high data rates [18], [19]. Can we construct higher order QAM modulations using a QPSK scheme with easily realizable operations/functions in RF domain? Starting with a simpler form of 4^{M}QAM, e.g., 16-QAM, we explore a modular approach to generate this constellation from a QPSK scheme.

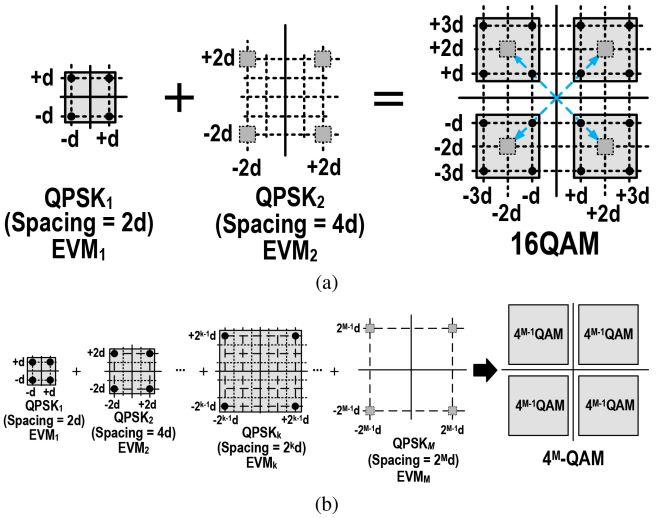


FIGURE 15. (a) 16QAM constellation is realized vectorially summing two QPSKs with a magnitude ratio of 2. (b) Generalized RF-4^{M}QAM constellation [23].

At the first glance, a 16QAM constellation is clearly comprised of four QPSKs across four quadrants of the complex plane (see Fig. 15(a)). An alternative perspective is to start with a QPSK cluster, and replicate it around four distinct origins, indicated by gray rectangles in Fig. 15(a). To construct a 16QAM constellation directly in RF domain, only two QPSK clusters are needed, QPSK1 with a symbol spacing of $2d$ and QPSK2 with 4d symbol spacing. QPSK2 is responsible for mapping the $(0, 0)$ origin to four origins located at $(-2d, -2d), (-2d, +2d), (2d, -2d), (2d, 2d)$. A Cartesian vector summation of QPSK1 symbols with those of QPSK2 in the complex plane will generate four new random symbols around each new origin at each quadrant, thereby resulting in generation of 16QAM constellation [23], [80]. Similarly, an RF-64QAM constellation is directly realized by replicating an RF-16QAM constellation across four quadrants and around four origins obtained by another QPSK cluster, QPSK3, with symbol spacing of $8d$. In general, to construct a high-order 4^{M}QAM scheme, this procedure will employ M QPSKs with symbol spacing of $2kd$ $(k = 1, \ldots, M)$. Once M QPSK clusters are generated in RF domain, this iterative procedure only requires scaling and vector summation, which are easily implemented in analog (or RF) domain [23]. Therefore, M QPSK signals having constant magnitude ratio of two are combined in order to build RF-4^{M}QAM constellation (shown in Fig. 15(b)). A conceptual block diagram of an RF-16QAM transmitter is shown in Fig. 16.

Assuming a random data stream, the error vectors in each QPSK constellation satisfy a two-dimensional Gaussian distribution with no correlation. Since each QPSK signal is constructed from randomly independent bits, the overall error vector power in an RF-4^{M}QAM constellation after addition is shown to be the weighted summation of EVMs of constituent M QPSK signals. Considering a special case where all QPSK signals exhibit the same EVM, the high-order QAM EVM

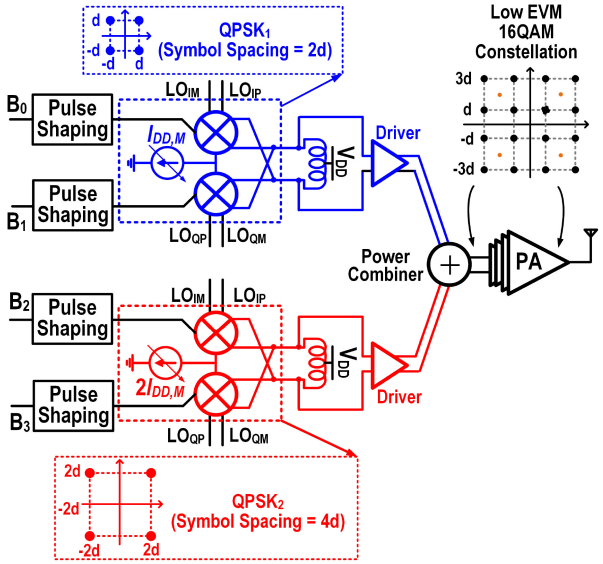


FIGURE 16. Conceptual block diagram of bits-to-RF RF-16QAM transmitter.

will be equal to the QPSK EVM [23]. This attribute brings along a number of advantages: (1) The QPSK signal can be generated using only symbol-rate timing with negligible degradation [23]. This readily relaxes the speed requirement of the mixed-signal interface by more than half compared to conventional DAC-based transmitter. (2) With no high-resolution high-speed DAC being present in this transmitter, high frequency linearity bottleneck is dramatically alleviated. It is because the amplitude linearity is no longer critical in achieving a low EVM for a constant-amplitude QPSK signal. (3) A precise magnitude ratio of two between any two side-by-side QPSK signals, as indicated in Fig. 15(b), is easily obtained by fine-tuning the DC bias current of each QPSK modulator, as shown in Fig. 16. This notion reveals yet another advantage compared to the current-source trimming associated with a DAC circuit. Only DC bias tuning is used in the RF-QAM modulation scheme to maintain the magnitude ratio of two between QPSK modulators instead of high-speed RF switching in a high-speed DAC within a conventional transmitter with its modulation being realized in digital domain. (4) Processing QPSK signals mandates much relaxed linearity requirement compared to a 4^{M}-QAM in a conventional transmitter. This has significant implication on design of the front-end power amplifier operating above 100 GHz. To leverage this attribute, the PA at the power combiner output in Fig. 16, which amplifies a high-PAPR 4^{M}-QAM signal, can be removed. Instead, each QPSK path employs a PA circuit prior to power combiner, which can now handle a low-PAPR constant-amplitude signal.

In summary, this RF-4^{M}QAM method shows elevated performance at much lower power consumption at THz frequencies compared to digital realization of QAM in conventional transmitters, because ultra-high-speed low-EVM QPSK signals can readily be constructed [19]. Finally, efforts are currently under way to explore power-efficient RF-4^{M}QAM demodulators.

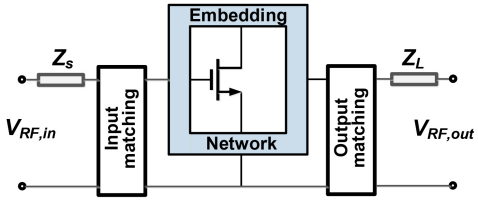


FIGURE 17. Conceptual block diagram of an amplifier employing an embedded transistor.

IV. CORE BUILDING BLOCKS FOR THz TRANSCEIVERS

A. THz AMPLIFIER DESIGN

The amplifiers designed at THz frequency range should be able to exhibit high performance at center frequencies beyond 100 GHz frequency range, while covering 20-30% fractional bandwidth. The power gain G_p of an amplifier in terms of its MAG, G_{max}, is readily calculated to be:

$$G_p = G_{max}\left(1-|s_{11}|^2\right)\left(1-|s_{22}|^2\right) \tag{6}$$

Around f_{max} frequencies, source and load conjugate matching at the amplifier's center frequency f_c is critical due to low available gain of transistors. This leaves little or no room to realize matching networks aiming to widen the amplifier's bandwidth on high side of the passband all the way to the upper corner frequency f_H, where G_{max} is dropping from its value at f_c. Recently, a few silicon-based THz amplifiers have been reported [16], [81], [82]. Great efforts have been made to improve the power gain of a THz amplifier, introducing all kinds of "embedding network" to the core device, as exemplified in generic block diagram representation in Fig. 17 [83], [84]. To gain a better insight into the THz amplifier design, we will look into the effects of active and passive (mainly due to matching networks) components, separately.

Actives: Although modern silicon technologies provide transistor devices with close to half-THz f_{max}, G_{max}-limitation at around $0.5f_{max}$, loosely defined as near-f_{max} frequencies, is a bottleneck for a THz amplifier design. Several powerful approaches to design amplifiers with power gain close to the MAG were proposed [16], [81]–[84]. Most notably, the gain-plane approach provides graphical representation where the contours of constant power gains are plotted within $\mathrm{Im}[U/A]-\mathrm{Re}[U/A]$ plane (U is Mason U [16], and $A=Y_{21}/Y_{12}$ is the maximum stable gain (MSG)), as demonstrated in Fig. 18. Any embedding network surrounding the main amplifier will be representing a locus crossing these constant gain contours [85], [86], as also shown in Fig. 18.

Our analysis in [83] proved that maximum value of MAG, proved to be equal to $\max[G_{max}]=(2U-1)+2\sqrt{U(U-1)}$, is achieved if and only if the imaginary part of A is zero and the device operates at the edge of unconditional stability region, i.e.,

$$\begin{cases}\mathrm{Im}\left[\frac{U}{A}\right]=0\\ K_f=1\end{cases} \tag{7}$$

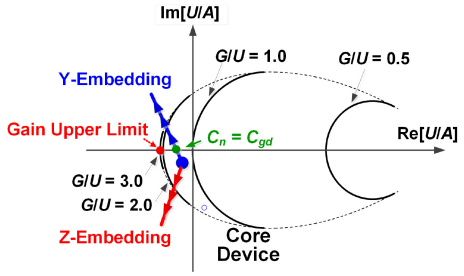


FIGURE 18. Constant-gain contours in the gain-plane together with moving direction of gain-state point of the device due to Y- or Z-embedding networks.

FIGURE 19. Two basic types of embedding network for a transistor (a) Y-embedding (b) Z-embedding.

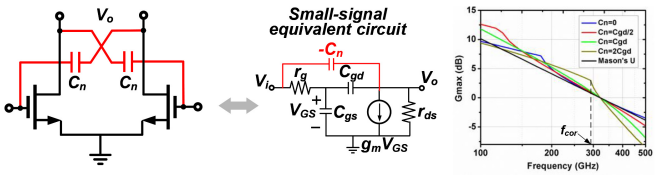


FIGURE 20. Block diagram of a differential pair with cross connected capacitor pair, C_n, and simulated plot of G_{max} vs. frequency for four C_n values along with the simulated Mason U.

Equation (7) sets forth the necessary and sufficient condition for an RF amplifier to attain the theoretical upper limit of its power gain. Moreover, two implications can be inferred from Eq. (7). First, it corroborates the commonly known intuitive approach, which primarily maintains that pushing the device towards its instability region will result in higher power gain. It is because the maximum power gain always occurs at the edge of the unconditional stability region. Second, setting $K_f = 1$ is not sufficient for the amplifier to reach its maximum power gain. This is because the imaginary part of U/A must also be zeroed, meaning that the phase of Y_{21} and Y_{12} must be the same.

Two basic types of embedding networks for a single device can be conceived; Y- and Z-embedding (see Figs. 19(a)–19(b)). Usually, reactive elements (e.g., inductors or capacitors) are used to realize these networks. Acting as a local-shunt feedback, the Y-embedding network can readily be characterized using Y-parameters. More precisely, Y_f is added to both Y_{11} and Y_{22}, while being subtracted from Y_{12} and Y_{21} [Y_f is the admittance of Y-embedding network]. Similar observation is made for the Z-embedding network, which acts as a local series feedback. The most widely used Y-embedding network is a pair of cross-connecting capacitors, C_n, in a differential pair in Fig. 20, acting as a neutralizing network [87].

An important notion, which is often missed in the design of THz amplifiers, is how to maintain gain flatness and small group delay variation across a frequency band as wide as 20-30% of the center frequency. A layout-parasitic-extracted differential pair with W/L=32μm/32nm in a 32 nm CMOS SOI process under four distinct neutralization capacitors, C_n, were simulated to study G_{max}, and the simulation result is shown in Fig. 20. For $K_f \leq 1$, the simulated G_{max} is indeed equal to MSG, A, and the corner frequency, f_{cor}, of the circuit of Fig. 20 is readily derived to be:

$$f_{cor} = \frac{1}{2\pi} \frac{\sqrt{\left[\frac{2}{g_m r_{ds}}\left(1+\frac{C_{gs}}{C_{gd}}\right) - \frac{C_X}{C_{gd}} + 1\right]^2 - 1}}{\left|r_g(C_{gs}+C_{gd}) - C_X/g_m\right|} \tag{8}$$

where $C_X = C_n - C_{gd}$. A study of the gain plots in Fig. 20 reveals that the gain is falling proportionally with $\mathcal{A}(K_f)$ $\left(\text{where } \mathcal{A}(K_f) = (K_f - \sqrt{K_f^2 - 1})\right)$ for frequencies above f_{cor}. This faster gain drop implies that the frequency range of interest for the wideband amplifier should be selected below f_{cor}. This is because for a transceiver handling high-order modulation scheme, the flatness of the front-end amplifier's frequency response is important. Otherwise, a degradation of EVM due to amplitude fluctuation is to be expected. In addition, the power gain will be significantly compromised to achieve a flat frequency response above f_{cor}. Therefore, for THz wideband amplifier design, the upper limit, f_H, of the frequency response is set to be equal to f_{cor}, i.e., $f_H = f_{cor}$.

Assuming that $f_H = f_{cor}$ and $K_f < 1$ for $f < f_{cor}$, the embedded device will become conditionally stable. In this case, the deployment of matching networks having frequency-dependent loss at the input and output ports of the amplification stage should increase the overall stability factor so that $K_f \to 1$. As a consequence, the core amplification stage with matching networks can be designed to operate at the boundary of unconditional and conditional stability across the amplifier's frequency band. This is clearly shown in Fig. 21 that demonstrates a graphical illustration of the design methodology for wideband THz amplifier.

In practice, the input and output impedances of a THz amplifier are determined by the preceding stage as well as the antenna load in the case of a front-end power amplifier, which cannot be an ideal resistive 50 Ω across the band. Moreover, the Bode-Fano criterion also specifies fundamental limitations for wideband input/output matching networks [88]. Therefore, a reliable THz amplifier design should guarantee stability under any loading condition across the entire operating frequency range, or equivalently, $K_f \geq 1$ across the entire bandwidth. The matching network plays an important role in satisfying this condition, as will be discussed next.

Matching Network: Reference [16] introduced over-neutralization to achieve gain-boosting in amplifiers designed to operate at near-f_{max} frequencies. It is, however, noteworthy that over-neutralization essentially is a narrowband technique. This means it helps boost the power gain, while

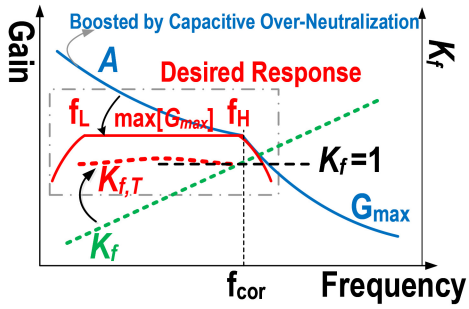


FIGURE 21. Illustration of the design methodology for a wideband THz amplifier showing how to determine and adjust gain, bandwidth, and stability.

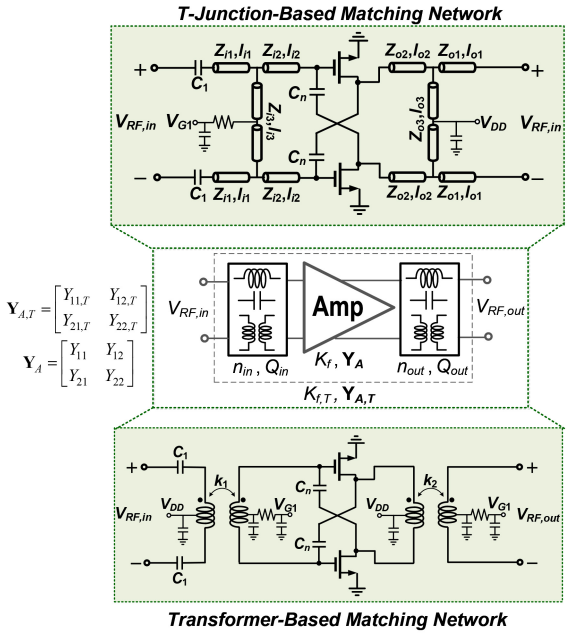


FIGURE 22. System block diagram of the amplifier with input and output matching network with two examples, namely, a neutralized differential amplifier with T-junction matching network and a neutralized amplifier with transformer-based matching network.

shrinking the bandwidth. The input-output and inter-stage matching networks play a critical role in widening the frequency range. Fig. 21 shows an illustration of the amplifier design methodology [89]. The input/output matching networks mainly perform two tasks, i.e., (1) they re-shape the frequency response with acceptable reflection and insertion losses, and (2) they stabilize the gain stage so that the entire amplifier will become unconditionally stable. As mentioned before, the gain flatness and stability are considered to be the key performance targets in amplifier design with 20-30% of fractional bandwidth. As such, the matching networks and their associated loss are designed so as to satisfy these performance targets. Fig. 22 shows a differential amplification stage with input and output matching circuits where n_{in} and n_{out} represent input and output transformation ratios. $\mathbf{Y}_{A,T}$ and $K_{f,T}$ are the overall network Y-parameter matrix and stability factor, respectively. To guarantee stability and maximum gain requirements, the amplifier with matching networks is designed so that $K_{f,T} \geq 1$. The matching loss can be quantified using the network Q. The circuit analysis of the circuit in Fig. 22 leads to the following relationship between network Q and the core amplifier K-factor, K_f [89].

$$Q_{in}Q_{out} = \frac{\cos\frac{1}{2}(\angle Y_{21} + \angle Y_{12} + \Phi_{MN})}{\cos\frac{1}{2}(\angle Y_{21} + \angle Y_{12})} \sqrt{\frac{1}{K_f}\left|\frac{Y_{21,T}}{Y_{21}}\right|\left|\frac{Y_{12,T}}{Y_{12}}\right|} \tag{9}$$

Specifically, (9) provides a clear relationship between matching loss that leads to bandwidth of interest and K_f.

Two approaches can be taken into consideration to design matching networks at THz frequencies. One approach uses T-junction matching networks to accomplish impedance matching. The microstrip-based structures are easily modeled using electromagnetic (EM) simulation tools. Shown in Fig. 22 is the schematic of a neutralized amplifier with microstrip T-section input and output matching networks. The major issue with the T-section network involves piecewise partial rotations of the immittance on the Smith-chart, which essentially results in narrowband matching. The same figure also shows the same core amplification stage with interstage transformer matching networks. The transformer-based interstage matching network can be viewed as a double-tuned passive network. The transformer bandwidth is a function of coupling coefficient and loaded quality factors of primary and secondary sides, and can increase by as much as 40% of the center frequency at the expense of larger in-band ripples. It is thus conceivable that front-end amplifiers intended for high data rate wideband applications to use a version of transformer-based matching.

B. THz LO GENERATION AND DISTRIBUTION

As the operation frequency increases, the implementation of low phase-noise LOs with adequate tuning range and output power becomes increasingly challenging. For a transceiver array with large array size, distribution of the LO across all the constituent transceivers adds another level of design hurdle. To address the LO distribution challenge, the core synthesizer can employ a subharmonic PLL at 1/Mth of the desired LO frequency. The LO distribution network will then carry LO signal at much lower frequency which is then boosted to the desired range with the aid of local frequency multipliers [90]. We will briefly discuss the LO generation and distribution in this section.

At the core of the PLL lies the VCO and divider chain determining the phase-noise, tuning range, and output power [52]. At THz frequencies, a multi-port oscillator circuit with multiple active devices exciting a multi-port passive structure, as a shared resonator, can produce much higher oscillation power and efficiency than current oscillators [91]. However, the use of multiple independent elements to excite a multi-port passive structure can lead to multiple stable oscillation states [92]–[94]. Inherent to such systems, this

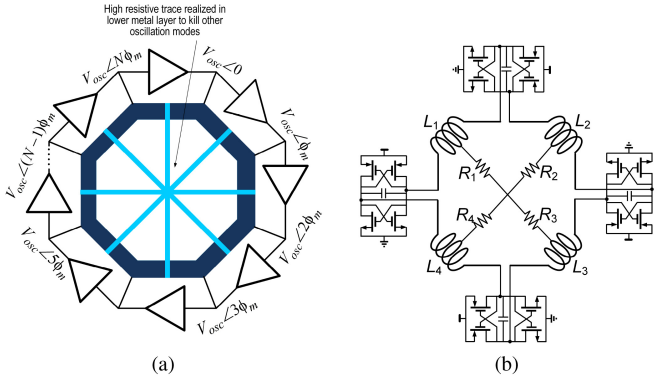


FIGURE 23. (a) Generic circularly symmetric oscillator. (b) A quad-core example presented in [95].

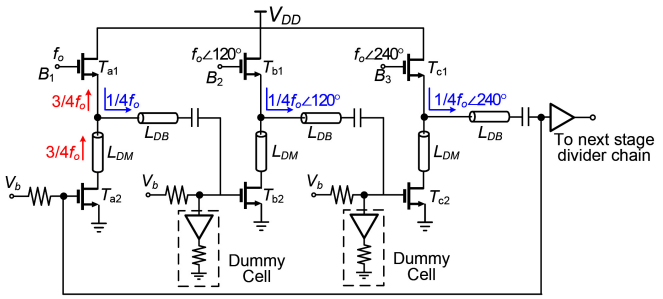


FIGURE 24. An example of multi-injection frequency divider, a triple-push divide-by-four frequency divider based on the design in [96].

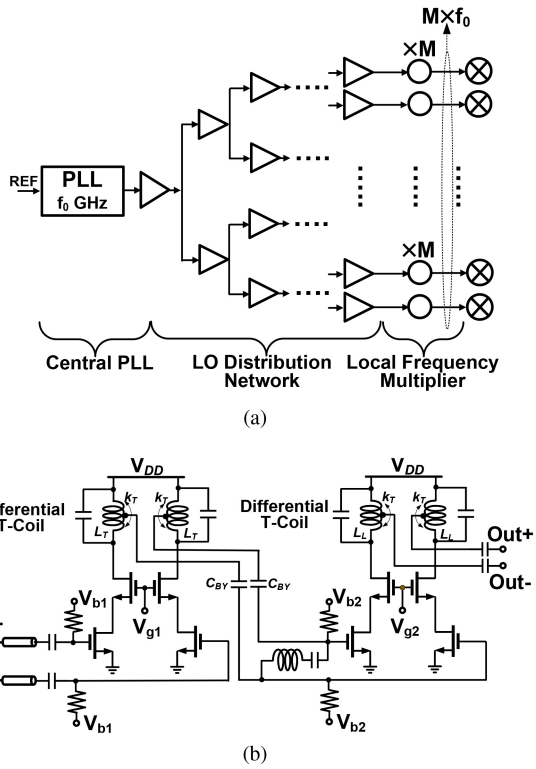


FIGURE 25. (a) An LO distribution network, (b) a broadband harmonic-based tripler.

attribute calls for a careful study that can quantitatively discover oscillation conditions that result in different oscillation states within a multi-port structure consisting of multiple active devices and passive networks. Reference [91] conducted a comprehensive study of the multi-port oscillators. Among multi-port oscillators, the ones with circular geometry are considered to be viable topology for low-noise THz oscillators [95]. Assuming an N-port circularly-symmetric passive network $\mathbf{Y}$, an N-port circularly symmetric excitation network $\tilde{\mathbf{Y}}$ exists which can potentially generate oscillation once it is connected to this circularly-symmetric passive network. The low phase-noise attribute of the circularly symmetric multi-port oscillator is easily understood once it is viewed as a N coupled oscillators. It is commonly known that a well-designed system of N-coupled oscillators should have $20\log_{10} N$ lower phase noise than a single oscillator. Another critical advantage of a multi-port circularly symmetric oscillator within the context of a phase-locked-based LO design is that next stage frequency divider, whose performance can be as critical as the VCO, can be designed to employ multi-injection architecture. One example of a multi-injection frequency divider, showcasing the MOS implementation of the circuit presented in [96], is depicted in Fig. 24, where three phases of a three-stage VCO output are injected to nodes B_k, $1 \leq k \leq 3$. In this circuit, the amplifying stages (T_{a2}, T_{b2}, T_{c2}) with the transmission lines L_{DM} and L_{DB} form the divider's three-stage ring and transistors T_{a1}, T_{b1}, T_{c1} act as the three mixing cells. The three-phase input signals coming from preceding three-stage VCO are fed to the gate terminals of the mixing cells, and subsequently, mixed with the loop's 3rd harmonic signals. The mixer's outputs ($1/4f_0\angle 0°$), $1/4f_0\angle 120°$, $1/4f_0\angle 240°$) flow back to the loop at three injection points.

One example of LO distribution network and a sub-THz harmonic-based frequency tripler are shown in Fig. 25(a) and 25(b), respectively. In this example, the lower frequency, f_0, (e.g., 13 GHz) is realized by a PLL. The LO distribution network also operating at much lower frequency (e.g., 13 GHz) carries the signal to K transceivers. The LO frequency is locally boosted to the desired frequency (e.g., 117 GHz) by a cascade of local frequency multipliers (e.g., two frequency multiplers (e.g., triplers) in Fig. 25(a)). Though the in-band phase noise in this synthesizer is magnified by $20\log_{10} M$ due to frequency multiplication, this degradation would be offset by the term in Leeson's equation, leaving the improvement in Q-factor, output swing and tuning range at low frequencies as added bonuses to overall phase noise improvement. Due to its inherently wideband characteristic, a harmonic-based frequency multiplication approach is preferred for wideband LO generation. Following the design principle presented in [90], one example of a high frequency harmonic-based frequency tripler is shown in Fig. 25(b). The circuit is comprised of two differential cascode stages driving broadband T-coils and an interstage matching network, where the first cascode stage is biased in

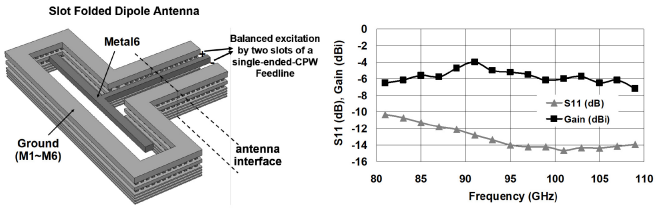


FIGURE 26. (a) The SFDA 3D view and (b) its simulation results [101].

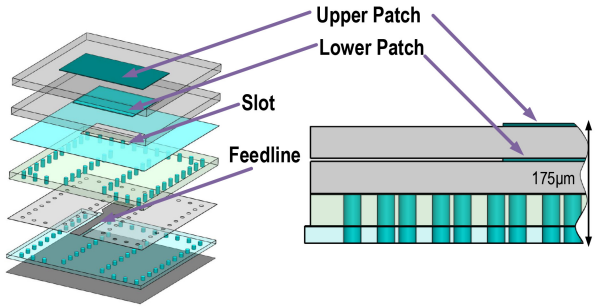


FIGURE 27. Schematic and side views of the proposed antenna [103].

the class-C region to maximize the 3rd-harmonic generation efficiency (as was shown in our prior work [97]) and the second stage acts as a wideband amplifier.

In short, the LO generation scheme in Fig. 25(a) offers several advantages. Operation at $1/M^{th}$ of the LO allows (1) the distribution network to be scalable to large arrays, and (2) more variety of synthesizer architectures including all-digital or fractional-N PLLs to be considered.

C. THz ANTENNA DESIGN

The antenna design at beyond 100−GHz frequency range, while covering a fractional bandwidth of 30%, is quite challenging. An on-chip antenna [16], [98]–[100] greatly simplifies the antenna-transceiver interface, as it is integrated alongside the rest of the system. This means the antenna and the feedlines can be co-designed and co-optimized with the front-end modules. Due to limited elevation of the metal stack, the on-chip antenna suffers from poor antenna gain and efficiency. For instance, [101] designed an on-chip slot-folded dipole antenna (SFDA) with coplanar-waveguide (CPW) feed line at W-band (Fig. 26(a)). To improve the antenna efficiency, a patterned deep trench mesh with a depth of $7 - 10$ μm from the substrate was embedded in the substrate underneath the antenna. In spite of using the deep trench lattice, this SFDA exhibited an efficiency of 16% and peak antenna gain of -4 dBi (Fig. 26(b)).

On the other hand, [102] implemented an aperture-stacked patch (ASP) antenna on printed circuit board (PCB) for the same frequency band. The simulated average gain was 6.3 dBi across 75− to 100−GHz, while the antenna efficiency varied from 88% to 93% [102].

The above two examples provide a clear snapshot about performance achievable by off-chip and on-chip antennas at W-band frequency range. A performance comparison between these two structures favors the off-chip antenna over its on-chip counterpart. It is, however, noteworthy that as the frequency is increased beyond 100 GHz, the antenna interface poses far greater challenge than at lower frequencies. More precisely, the transition VIA structure used by [102] to realize the interface reaches its performance limit at higher frequencies. The stringent physical requirement of the VIA structure imposed by limitation of the PCB fabrication at beyond 100 GHz undermines its use. We have extended the design of a PCB antenna to above 100 GHz frequency range [103]. This design was composed of two rectangular stacked patches fed by a rectangular slot coupled to a stripline feed. The 3-dimensional and the side views of the antenna structure are shown in Fig. 27.

In contrast to low-frequency board antennas where the via locations do not impact the performance, in this design, the via arrangement should be optimized to minimize the excitation of unwanted surface waves. The substrate layer material in this flexible printed circuit (FPC) technology has an ϵ_r of 2.6, which reduces the loss associated with the dielectric material compared to on-chip counterparts. By having the larger antenna at the top, the fringing fields of both antennas will have well-defined connections to the ground layer, thereby increasing the gain and bandwidth of the radiation. The simulated antenna achieves a -10 dB bandwidth of 44.9 GHz with an average realized gain of 5.7 dBi and average efficiency of 73.9% across the bandwidth [103].

Alternative approaches such as copper-pillar or direct waveguide interface to the antenna structure would be more amenable to higher frequencies. Recently, the radio-on-glass technology has showed very promising performance at D-band [104].

V. CONCLUSION

This paper presented an overview of challenges behind the design and implementation of THz integrated systems and circuits. Several key performance parameters including the bandwidth; communication range; link reliability and throughout; modulation and spectral efficiency were outlined. At the system-level design, we discussed multi-antenna architecture as an inevitable choice to meet performance requirement in a high-speed wireless link setting. Started with conventional LO- and RF-phase-shifting phased array schemes, we briefly discussed MIMO transceivers incoporting digital and hybrid beamforming. Next, we argued that modern transceiver architectures are fundamentally incapable of addressing unresolved challenges to achieve 50+ Gbps data rates. Delegating (de-)modulation to the digital back-end as commonly done requires high-resolution and high-speed data converters that are impossible to realize in silicon technologies. In addition, methods such as channel bonding often lead to unacceptable amount of power dissipation. We made an argument in favor of novel transmitter and receiver architectures incorporating direct-modulation and direct demodulation in RF domain applicable for beyond-5G communications. Three Examples were presented, namely an

RF-8PSK transmitter, an RF-8PSK receiver, and RF-16QAM transmitter. At the circuit-level design, we briefly studied THz amplifiers, LO generation and distribution networks, and antenna designs made several important observations and design guidelines.

ACKNOWLEDGMENT

The author would like to thank all former and current students in the Nanoscale Communication Integrated Circuits (NCIC) Labs including Peyman Nazari, Zheng Wang, Huan Wang, Hossein Mohammadnezhad, Zhiming Chen, Chung-Cheng Wang, and Zisong Wang for their outstanding contributions to many topics of this work. The author would also thank Dr. Hamidreza Aghasi and Hedayatullah Maktoomi for their joint collaboration on off-chip antenna design in [103].

REFERENCES

[1] K.-C. Huang and Z. Wang, "Terahertz terabit wireless communication," *IEEE Microw. Mag.*, vol. 12, no. 4, pp. 108–116, Jun. 2011.

[2] T. Kürner and S. Priebe, "Towards THz communications—Status in research, standardization and regulation," *J. Infrared Millim. THz Waves*, vol. 35, no. 1, pp. 53–62, 2014.

[3] T. Nagatsuma *et al.*, "30-Gbit/s wireless transmission over 10 meters at 300 GHz," in *Proc. 39th Int. Conf. Infrared Millim. THz waves (IRMMW-THz)*, Tucson, AZ, USA, 2014, pp. 1–2.

[4] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 1, pp. 92–99, 2010.

[5] C. Terzi and I. Korpeoglu, "60 GHz wireless data center networks: A survey," *Comput. Netw.*, vol. 185, 2021, Art. no. 107730. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128620313207

[6] M. Abbasi *et al.*, "Single-chip 220−GHz active heterodyne receiver and transmitter MMICs with on-chip integrated antenna," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 2, pp. 466–478, Feb. 2011.

[7] I. Kallfass *et al.*, "All active MMIC-based wireless communication at 220 GHz," *IEEE Trans. THz Sci. Techn.*, vol. 1, no. 2, pp. 477–487, Nov. 2011.

[8] I. Kallfass *et al.*, "64 Gbit/s transmission over 850 m fixed wireless link at 240 GHz carrier frequency," *J. Infrared Millim. THz Waves*, vol. 36, no. 2, pp. 221–233, Jan. 2015.

[9] K. Katayama *et al.*, "A 300 GHz CMOS transmitter with 32-QAM 17.5 Gb/s/ch capability over six channels," *IEEE J. Solid-State Circuits*, vol. 51, no. 12, pp. 3037–3048, Dec. 2016.

[10] H. Takahashi, T. Kosugi, A. Hirata, K. Murata, and N. Kukutsu, "10-Gbit/s quadrature phase-shift-keying modulator and demodulator for 120−GHz-band wireless links," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 12, pp. 4072–4078, Dec. 2010.

[11] S. Carpenter *et al.*, "A *D*-band 48-Gbit/s 64-QAM/QPSK direct-conversion I/Q transceiver chipset," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 4, pp. 1285–1296, Apr. 2016.

[12] H. Hamada *et al.*, "300-GHz. 100-Gb/s InP-HEMT wireless transceiver using a 300−GHz fundamental mixer," in *IEEE MTT-S Int. Micrw. Symp. Dig.*, Philadelphia, PA, USA, Jun. 2018, pp. 1480–1483.

[13] H.-J. Song, J.-Y. Kim, K. Ajito, N. Kukutsu, and M. Yaita, "50-Gb/s direct conversion QPSK modulator and demodulator MMICs for terahertz communications at 300 GHz," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 3, pp. 600–609, Mar. 2014.

[14] M. Fujishima, M. Motoyoshi, K. Katayama, K. Takano, N. Ono, and R. Fujimoto, "98 mW 10 Gbps wireless transceiver chipset with D-band CMOS circuits," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2273–2284, Oct. 2013.

[15] K. Takano *et al.*, "17.9 a 105Gb/s 300GHz CMOS transmitter," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2017, pp. 308–309.

[16] Z. Wang, P.-Y. Chiang, P. Nazari, C.-C. Wang, Z. Chen, and P. Heydari, "A CMOS 210-GHz fundamental transceiver with OOK modulation," *IEEE J. Solid-State Circuits*, vol. 49, no. 3, pp. 564–580, Mar. 2014.

[17] K. K. Tokgoz *et al.*, "13.3 a 56Gb/s W-band CMOS wireless transceiver," in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2016, pp. 242–243.

[18] S. V. Thyagarajan, S. Kang, and A. M. Niknejad, "A 240 GHz fully integrated wideband QPSK receiver in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 50, no. 10, pp. 2268–2280, Oct. 2015.

[19] S. Kang, S. V. Thyagarajan, and A. M. Niknejad, "A 240 GHz fully integrated wideband QPSK transmitter in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 50, no. 10, pp. 2256–2267, Oct. 2015.

[20] H. Mohammadnezhad, H. Wang, A. Cathelin, and P. Heydari, "A 115–135-GHz 8PSK receiver using multi-phase RF-correlation-based direct-demodulation method," *IEEE J. Solid-State Circuits*, vol. 54, no. 9, pp. 2435–2448, Sep. 2019.

[21] P. Nazari, S. Jafarlou, and P. Heydari, "A CMOS two-element 170-GHz fundamental-frequency transmitter with direct RF-8PSK modulation," *IEEE J. Solid-State Circuits*, vol. 55, no. 2, pp. 282–297, Feb. 2020.

[22] H. Wang, H. Mohammadnezhad, D. Dimlioglu, and P. Heydari, "A 100-120GHz 20Gbps bits-to-RF 16QAM transmitter using 1-bit digital-to-analog interface," in *Proc. IEEE CICC*, Austin, TX, USA, Apr. 2019, pp. 1–4.

[23] H. Wang, H. Mohammadnezhad, and P. Heydari, "Analysis and design of high-order QAM direct-modulation transmitter for high-speed point-to-point mm-Wave wireless links," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 3161–3179, Nov. 2019.

[24] D. Fritsche, P. Stärke, C. Carta, and F. Ellinger, "A low-power SiGe BiCMOS 190-GHz transceiver chipset with demonstrated data rates up to 50 Gbit/s using on-chip antennas," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3312–3323, Sep. 2017.

[25] C. Jiang, A. Cathelin, and E. Afshari, "A high-speed efficient 220-GHz spatial-orthogonal ASK transmitter in 130-nm SiGe BiCMOS," *IEEE J. Solid-State Circuits*, vol. 52, no. 9, pp. 2321–2334, Sep. 2017.

[26] N. Sarmah *et al.*, "A fully integrated 240-GHz direct-conversion quadrature transmitter and receiver chipset in SiGe technology," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 2, pp. 562–574, Feb 2016.

[27] J. Grzyb, P. R. Vazquez, N. Sarmah, W. Forster, B. Heinemann, and U. Pfeiffer, "High data-rate communication link at 240 GHz with on-chip antenna-integrated transmitter and receiver modules in SiGe HBT technology," in *Proc. 11th Eur. Microw. Integr. Circuit Conf.*, Paris, France, Mar. 2017, pp. 1369–1373.

[28] N. Sarmah, P. R. Vazquez, J. Grzyb, W. Foerster, B. Heinemann, and U. R. Pfeiffer, "A wideband fully integrated SiGe chipset for high data rate communication at 240 GHz," in *Proc. 11th Eur. Microw. Integr. Circuit Conf.*, London, U.K., Oct. 2016, pp. 181–184.

[29] J. Grzyb, P. R. Vazquez, N. Sarmah, B. Heinemann, and U. R. Pfeiffer, "A 240 GHz high-speed transmission link with highly-integrated transmitter and receiver modules in SiGe HBT technology," in *Proc. IEEE 42nd Int. Conf. Infrared Millim. THz Waves*, Cancun, Mexico, Aug. 2017, pp. 1–2.

[30] P. Rodriguez-Vázquez, J. Grzyb, B. Heinemann, and U. R. Pfeiffer, "Performance evaluation of a 32-QAM 1-meter wireless link operating at 220–260 GHz with a data-rate of 90 Gbps," in *Proc. IEEE Asia–Pac. Microw. Conf.*, Kyoto, Japan, Nov. 2018, pp. 723–725.

[31] P. Rodriguez-Vázquez, J. Grzyb, N. Sarmah, B. Heinemann, and U. R. Pfeiffer, "Towards 100 Gbps: A fully electronic 90 Gbps one meter wireless link at 230 GHz," in *Proc. IEEE 48th Eur. Microw. Conf.*, Madrid, Spain, Sep. 2018, pp. 369–372.

[32] P. R. Vazquez, J. Grzyb, N. Sarmah, B. Heinemann, and U. R. Pfeiffer, "A 219–266 GHz fully-integrated direct-conversion IQ receiver module in a SiGe HBT technology," in *Proc. 12th Eur. Microw. Integr. Circuit Conf.*, Nuremberg, Germany, Oct. 2017, pp. 261–264.

[33] M. Elkhouly, Y. Mao, C. Meliani, F. Ellinger, C. Schyett, "A 245 GHz ASK modulator and demodulator with 40 Gbits/sec data rate in 0.13 μm SiGe BiCMOS technology," in *IEEE MTT-S Int. Micrw. Symp. Dig.*, Seattle, WA, USA, Jun. 2013, pp. 1–3.

[34] S. Lee *et al.*, "300-GHz CMOS-based wireless link using 40-dBi cassegrain antenna for IEEE standard 802.15. 3d," in *Proc. IEEE Int. Symp. Radio Freq. Integr. Technol. (RFIT)*, Hiroshima, Japan, 2020, pp. 136–138.

[35] H. Hamada *et al.*, "300-GHz-band 120-Gb/s wireless front-end based on InP-HEMT PAs and mixers," *IEEE J. Solid-State Circuits*, vol. 55, no. 9, pp. 2316–2335, Sep. 2020.

[36] K.-S. Choi, D. R. Utomo, K.-M. Kim, B.-H. Yun, S.-G. Lee, and I.-Y. Lee, "29.7 a 490GHz 32mw fully integrated CMOS receiver adopting dual-locking FLL," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, San Francisco, CA, USA, 2020, pp. 452–454.

[37] S. Ooms and P. Reynaert, "A 120-GHz wireless link using 3D-printed lens with flexible dielectric fiber feed and 28-nm CMOS transceiver," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 142–145, 2020.

[38] R. Carter *et al.*, "22nm FDSOI technology for emerging mobile, Internet-of-Things, and RF applications," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2016, pp. 1–4.

[39] S. Ong *et al.*, "A 22nm FDSOI technology optimized for RF/mmWave applications," in *Proc. IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, Philadelphia, PA, USA, 2018, pp. 72–75.

[40] T. Zimmer *et al.*, "SiGe HBTs and BiCMOS technology for present and future millimeter-wave systems," *IEEE J. Microw.*, vol. 1, no. 1, pp. 288–298, Jan. 2021.

[41] M. H. Eissa *et al.*, "Wideband 240-GHz transmitter and receiver in bicmos technology with 25-Gbit/s data rate," *IEEE J. Solid-State Circuits*, vol. 53, no. 9, pp. 2532–2542, Sep. 2018.

[42] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[43] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[44] N. Ebrahimi, P.-Y. Wu, M. Bagheri, and J. F. Buckwalter, "A 71–86-GHz phased array transceiver using wideband injection-locked oscillator phase shifters," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 2, pp. 346–361, Feb. 2017.

[45] Y. Yang, O. D. Gurbuz, and G. M. Rebeiz, "An eight-element 370–410-GHz phased-array transmitter in 45-nm CMOS SOI with peak EIR of 8–8.5 dBm," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4241–4249, Dec. 2016.

[46] F. Golcuk, T. Kanar, and G. M. Rebeiz, "A 90-100-GHz 4 × 4 SiGe BiCMOS polarimetric transmit/receive phased array with simultaneous receive-beams capabilities," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 8, pp. 3099–3114, Aug. 2013.

[47] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," U.S. Patent 5 345 599, Sep. 1994. [Online]. Available: https://patents.google.com/patent/US5345599A/en

[48] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[49] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[50] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antenn. Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[51] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2nd Quart., 2016.

[52] B. Razavi, *RF Microelectronics*, vol. 2. New York, NY, USA: Prentice-Hall, 2012.

[53] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA: Wiley, 2016.

[54] Y. M. Greshishchev *et al.*, "A 56GS/s 6b DAC in 65nm CMOS with 256×6b memory," in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2011, pp. 194–196.

[55] J. Cao *et al.*, "A transmitter and receiver for 100Gb/s coherent networks with integrated 4×64GS/s 8b ADCs and DACs in 20nm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, Feb 2017, pp. 484–485.

[56] *Factsheet LEIA 55–65 GSa/s 8-bit DAC*, Fujitsu Semicond., Kanagawa, Japan, 2019.

[57] H. Huang, J. Heilmeyer, M. Grözing, M. Berroth, J. Leibrich, and W. Rosenkranz, "An 8-bit 100-GS/s distributed DAC in 28-nm CMOS for optical communications," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 4, pp. 1211–1218, Apr. 2015.

[58] L. Kull *et al.*, "A 24–72-GS/s 8-b time-interleaved SAR ADC with 2.0–3.3-pJ/conversion and > 30 db SNDR at nyquist in 14-nm CMOS FinFET," *IEEE J. Solid-State Circuits*, vol. 53, no. 12, pp. 3508–3516, Dec. 2018.

[59] K. Sun, G. Wang, Q. Zhang, S. Elahmadi, and P. Gui, "A 56-GS/s 8-bit time-interleaved ADC with ENOB and BW enhancement techniques in 28-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 54, no. 3, pp. 821–833, Mar. 2018.

[60] A. Zandieh, P. Schvan, and S. P. Voinigescu, "Design of a 55-nm SiGe BiCMOS 5-bit time-interleaved flash ADC for 64-Gbd 16-QAM fiberoptics applications," *IEEE J. Solid-State Circuits*, vol. 54, no. 9, pp. 2375–2387, Sep. 2019.

[61] B. Razavi, "Lower bounds on power consumption of clock generators for ADCs," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Seville, Spain, Oct. 2020, pp. 1–5.

[62] B. Murmann. (2021). *ADC Performance Survey*. Accessed: Jun. 26, 2021. [Online]. Available: https://web.stanford.edu/~murmann/adcsurvey.html

[63] A. Behzad *et al.*, "A fully integrated MIMO multiband direct conversion CMOS transceiver for WLAN applications (802.11n)," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2795–2808, Dec. 2007.

[64] X. Guan, H. Hashemi, and A. Hajimiri, "A fully integrated 24-GHz eight-element phased-array receiver in silicon," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2311–2320, Dec. 2004.

[65] A. Natarajan, A. Komijani, and A. Hajimiri, "A fully integrated 24-GHz phased-array transmitter in CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2502–2514, Dec. 2005.

[66] A. Babakhani, X. Guan, A. Komijani, A. Natarajan, and A. Hajimiri, "A 77-GHz phased-array transceiver with on-chip antennas in silicon: Receiver and antennas," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2795–2806, Dec. 2006.

[67] A. Natarajan, A. Komijani, X. Guan, A. Babakhani, and A. Hajimiri, "A 77-GHz phased-array transceiver with on-chip antennas in silicon: Transmitter and local LO-path phase shifting," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2807–2819, Dec. 2006.

[68] K.-J. Koh and G. M. Rebeiz, "0.13-μm CMOS phase shifters for X-, Ku-, and K-band phased arrays," *IEEE J. Solid-State Circuits*, vol. 42, no. 11, pp. 2535–2546, Nov. 2007.

[69] K.-J. Koh, J. W. May, and G. M. Rebeiz, "A millimeter-wave (40–45 GHz) 16-element phased-array transmitter in 0.18-μm SiGe BiCMOS technology," *IEEE J. Solid-State Circuits*, vol. 44, no. 5, pp. 1498–1509, May 2009.

[70] W. Shin, B.-H. Ku, O. Inac, Y.-C. Ou, and G. M. Rebeiz, "A 108–114 GHz 4×4 wafer-scale phased array transmitter with high-efficiency on-chip antennas," *IEEE J. Solid-State Circuits*, vol. 48, no. 9, pp. 2041–2055, Sep. 2013.

[71] B. Sadhu *et al.*, "A 28-GHz 32-element TRX phased-array IC with concurrent dual-polarized operation and orthogonal phase and gain control for 5G communications," *IEEE J. Solid-State Circuits*, vol. 52, no. 12, pp. 3373–3391, Dec. 2017.

[72] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[73] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, "Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3403–3418, Jul. 2018.

[74] A. Jahanian, F. Tzeng, and P. Heydari, "Code-modulated path-sharing multi-antenna receivers: Theory and analysis," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2193–2201, May 2009.

[75] F. Tzeng, A. Jahanian, D. Pi, and P. Heydari, "A CMOS code-modulated path-sharing multi-antenna receiver front-end," *IEEE J. Solid-State Circuits*, vol. 44, no. 5, pp. 1321–1335, May 2009.

[76] S. Mondal, R. Singh, A. I. Hussein, and J. Paramesh, "A 25–30 GHz fully-connected hybrid beamforming receiver for MIMO communication," *IEEE J. Solid-State Circuits*, vol. 53, no. 5, pp. 1275–1287, May 2018.

[77] H. Mohammadnezhad, R. Abedi, and P. Heydari, "A millimeter-wave partially overlapped beamforming-MIMO receiver: Theory, design, and implementation," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 5, pp. 1924–1936, May 2019.

[78] J. G. Proakis and M. Salehi, *Digital Communications*. Boston, MA, USA: McGraw-Hill, 2008.

[79] B. Razavi, "Prospects of CMOS technology for high-speed optical communication circuits," *IEEE J. Solid-State Circuits*, vol. 37, no. 9, pp. 1135–1145, Sep. 2002.

[80] K. Miyauchi, S. Seki, and H. Ishio, "New technique for generating and detecting multilevel signal formats," *IEEE Trans. Commun.*, vol. 24, no. 2, pp. 263–267, Feb. 1976.
[81] O. Momeni, "A 260GHz amplifier with 9.2dB gain and −3.9dBm saturated power in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, USA, 2013, pp. 140–141.
[82] D.-W. Park, D. R. Utomo, B. Yun, H. U. Mahmood, J.-P. Hong, and S.-G. Lee, "Design of high-gain sub-THz regenerative amplifiers based on double-g_{max} gain boosting technique," *IEEE J. Solid-State Circuits*, early access, Jul. 5, 2021, doi: 10.1109/JSSC.2021.3092168.
[83] Z. Wang and P. Heydari, "A study of operating condition and design methods to achieve the upper limit of power gain in amplifiers at near-f_{max} frequencies," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 2, pp. 261–271, Feb. 2017.
[84] H. Bameri and O. Momeni, "A high-gain mm-Wave amplifier design: An analytical approach to power gain boosting," *IEEE J. Solid-State Circuits*, vol. 52, no. 2, pp. 357–370, Feb. 2017.
[85] A. Slnghakowinta and A. R. Boothroyd, "Gain capability of two-port amplifiers," *Int. J. Electron.*, vol. 21, no. 6, pp. 549–560, 1966.
[86] R. Spence, *Linear Active Networks*. London, U.K.: Wiley, 1970.
[87] P. Heydari, "Neutralization techniques for high-frequency amplifiers: An overview," *IEEE Solid-State Circuits Mag.*, vol. 9, no. 4, pp. 82–89, Nov. 2017.
[88] R. M. Fano, "Theoretical limitations on the broadband matching of arbitrary impedances," *J. Franklin Inst.*, vol. 249, no. 1, pp. 57–83, 1950.
[89] Z. Wang, H. Wang, and P. Heydari, "CMOS power-amplifier design perspectives for 6G wireless communications," in *Proc. IEEE Midwest Symp. Circuits Syst.*, Aug. 2021, pp. 1–4.
[90] C.-C. Wang, Z. Chen, and P. Heydari, "W-band silicon-based frequency synthesizers using injection-locked and harmonic triplers," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 5, pp. 1307–1320, May 2012.
[91] P. Nazari, S. Jafarlou, and P. Heydari, "Analysis and design of a millimeter-wave cavity-backed circularly polarized radiator based on fundamental theory of multi-port oscillators," *IEEE J. Solid-State Circuits*, vol. 52, no. 12, pp. 3293–3311, Dec. 2017.
[92] K. Kurokawa, "The single-cavity multiple-device oscillator," *IEEE Trans. Microw. Theory Techn.*, vol. 19, no. 10, pp. 793–801, Oct. 1971.
[93] A. Mirzaei, M. E. Heidari, R. Bagheri, S. Chehrazi, and A. A. Abidi, "The quadrature LC oscillator: A complete portrait based on injection locking," *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 1916–1932, Sep. 2007.
[94] E. Laskin, P. Chevalier, A. Chantre, B. Sautreuil, and S. P. Voinigescu, "165-GHz transceiver in SiGe technology," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1087–1100, May 2008.
[95] D. Murphy and H. Darabi, "A 27-GHz quad-core CMOS oscillator with no mode ambiguity," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3208–3216, Nov. 2018.
[96] P.-Y. Chiang, Z. Wang, O. Momeni, and P. Heydari, "A silicon-based 0.3 THz frequency synthesizer with wide locking range," *IEEE J. Solid-State Circuits*, vol. 49, no. 12, pp. 2951–2963, Dec. 2014.
[97] Z. Chen and P. Heydari, "An 85-95.2 GHz transformer-based injection-locked frequency tripler in 65nm CMOS," in *Proc. IEEE MTT-S Int. Microw. Symp.*, Anaheim, CA, USA, 2010, pp. 776–779.
[98] K. Sengupta and A. Hajimiri, "A 0.28 THz power-generation and beam-steering array in CMOS based on distributed active radiators," *IEEE J. Solid-State Circuits*, vol. 47, no. 12, pp. 3013–3031, Dec. 2012.
[99] H. Jalili and O. Momeni, "A 0.46-THz 25-element scalable and wideband radiator array with optimized lens integration in 65-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 55, no. 9, pp. 2387–2400, Sep. 2020.
[100] S. Kong, K. M. Shum, C. Yang, L. Gao, and C. H. Chan, "Wide impedance-bandwidth and gain-bandwidth terahertz on-chip antenna with chip-integrated dielectric resonator," *IEEE Trans. Antennas Propag.*, vol. 69, no. 8, pp. 4269–4278, Aug. 2021.
[101] Z. Chen, C.-C. Wang, H.-C. Yao, and P. Heydari, "A BiCMOS W-band 2× 2 focal-plane array with on-chip antenna," *IEEE J. Solid-State Circuits*, vol. 47, no. 10, pp. 2355–2371, Oct. 2012.
[102] S. Shahramian, M. J. Holyoak, and Y. Baeyens, "A 16-element W-band phased-array transceiver chipset with flip-chip PCB integrated antennas for multi-gigabit wireless data links," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3389–3402, Jul. 2018.
[103] M. H. Maktoomi *et al.*, "A GSG-excited ultra-wideband 103-147 GHz stacked patch antenna on flexible printed circuit," in *Proc. IEEE Antennas Propag. Symp.*, 2021, p. 1.
[104] M. Elkhouly *et al.*, "D-band phased-array TX and RX front ends utilizing radio-on-glass technology," in *Proc. IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, Los Angeles, CA, USA, 2020, pp. 91–94.

PAYAM HEYDARI (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Sharif University of Technology in 1992 and 1995, respectively, and the Ph.D. degree from the University of Southern California in 2001.

He is the Chancellor's Professor of Electrical Engineering and Computer Science with the University of California at Irvine (UCI), Irvine. His research covers the design of THz/millimeter-wave/RF and analog integrated circuits. He is the Director of Nanoscale Communication IC Labs. His group was among the first who introduced the design of millimeter-wave integrated circuits in silicon technologies. They demonstrated the world's first fundamental frequency D-band CMOS transceiver, the world's highest radiated power, and highest efficiency sub-THz circularly-polarized radiator in silicon employing a multiport cavity-backed structure. He has coauthored of two books, one book chapter, and more than 170 journal and conference papers.

Dr. Heydari is a recipient of the 2021 IEEE SSCS Innovative Education Award and the Best Invited Paper Award at the 2021 IEEE Custom Integrated Circuits Conference (CICC). He was a recipient of the 2016–2017 UCI's School of Engineering Mid-Career Excellence in Research, the 2014 Distinguished Engineering Educator Award from Orange County Engineering Council, the 2009 Business Plan Competition First Place Prize Award and the Best Concept Paper Award both from Paul Merage School of Business at UCI, the 2010 Faculty of the Year Award from UCI's Engineering Student Council, the 2009 School of Engineering Best Faculty Research Award, the 2007 IEEE Circuits and Systems Society Guillemin-Cauer Award, the 2005 IEEE Circuits and Systems Society Darlington Award, the 2005 National Science Foundation CAREER Award, the 2005 Henry Samueli School of Engineering Teaching Excellence Award, the Best Paper Award at the 2000 IEEE Int'l Conference on Computer Design, and the 2001 Technical Excellence Award from the Association of Professors and Scholars of Iranian Heritage. He was recognized as the 2004 Outstanding Faculty with the EECS Department, UCI. His research on novel low-power multipurpose multiantenna RF front-ends received the Low-Power Design Contest Award at the 2008 IEEE Int'l Symposium on Low-Power Electronics and Design. He currently serves both as an Associate Editor IEEE *Journal of Solid-State Circuits* and IEEE SOLID-STATE CIRCUITS LETTERS. He gave Keynote Speech to IEEE GlobalSIP 2013 Symposium on Millimeter Wave Imaging and Communications, served as an Invited Distinguished Speaker to the 2014 IEEE Midwest Symposium on Circuits and Systems, and gave a Tutorial at the 2017 International Solid-State Circuits Conference (ISSCC). He was a Distinguished Lecturer of the IEEE Solid-State Circuits Society (SSCS) from January 2014 to January 2016, and is currently a Distinguished Microwave Lecture of the IEEE Microwave Theory and Techniques Society from January 2019 to December 2022. He was selected as the inaugural Faculty Innovation Fellow by the UCI's Beall Applied Innovation. He is a member of Technical Program Committee (TPC) of ESSCIRC and IMS. Formerly, he was a member of TPC of the ISSCC, CICC, CSICS, and an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS. He was an AdCom Member of the IEEE SSCS from 2018 to 2020.