W-Band Direct-Modulation >20-Gb/s Transmit and Receive Building Blocks in 32-nm SOI CMOS

Hasan Al-Rubaye, Student Member, IEEE, and Gabriel M. Rebeiz, Fellow, IEEE

Abstract—This paper presents Gb/s-speed circuit building blocks in 32-nm CMOS SOI, for a > 20-Gb/s Cartesian direct-modulation W-band transmitter. Transmitter systems nonidealities and performance limitations are discussed, and circuit design techniques and analyses are presented. The transmitter employs two 2-b high-speed RF digital-to-analog converters driven in quadrature, 20-dB gain W-band LO drivers, and 30-Gb/s high-speed digital retimers and deserializers, and is capable of supporting BPSK/PAM4/QPSK/16-QAM modulation schemes, at a saturated output power $P_{\text{sat}}$ of +4 dBm. A maximum data rate of 20 Gb/s was achieved when operating in QPSK mode, 4 Gb/s in 16-QAM mode, and 12 Gb/s in both BPSK and PAM4 modes. The chip occupies 1.4×0.8 mm$^2$, and consumes 110 mW in BPSK/PAM4 modes and 220 mW in QPSK and 16-QAM modes, resulting in the state-of-the-art $9^\text{th}$ and $15^\text{th}$-order BPSK and PAM4 $\Sigma^\Delta$ efficiencies, respectively. A mixer-first wideband W-band receiver that includes a passive mixer and a wideband transimpedance amplifier is also presented. Measurements of the receiver chip demonstrated its capability to downconvert and amplify highly complex modulated waveforms (> 256-QAM), and at high data rates, up to 60 Gb/s in 64-QAM, which proves the feasibility of building high dynamic-range mm-wave receivers with bandwidth greater than 30 GHz. The receiver chip was also built in 32-nm CMOS SOI, occupying a core area of 0.18×0.1 mm$^2$.

Index Terms—32-nm, amplifier, CMOS, digital predistortion, digital transmitter, direct modulation, error vector magnitude (EVM), filter, mm-wave, QAM, RF digital-to-analog converter (RF-DAC), SOI, transimpedance transmitter (TIA), true single-phase clocked (TSPC), W-band.

I. INTRODUCTION

The continuous scaling of CMOS technology has led to the emergence of a relatively new class of digitally assisted analog and RF circuit design techniques. Digital-like power amplifiers and modulators at RF [1]–[7] and mm-wave [8], [9] frequencies are one such example. This is driven by the increasing transition frequency $f_T$ of CMOS devices, which enables the design of high-performance digital switches. Analog and RF circuit designers, on the other hand, are forced to cope with analog building blocks, which suffer from poor linearity due to short-channel effects and downscaling of supply voltages. Digitally assisted techniques at RF frequencies leverage technology scaling and continue to improve as switches get faster, while offering attractive systems-on-chip integration where RF and analog blocks are seamlessly integrated with digital baseband circuitry.

More specifically, the increasing need for high data rate (Gb/s) wireless links has created a new class of direct-modulation RF transmitters, which offer greater bandwidths than traditional heterodyne or zero-IF transmitters. Direct-modulation RF transmitters perform the digital-to-analog conversion and modulation directly at the RF frequency, using RF digital-to-analog converters (RF-DACs), which maximize the available modulation bandwidth [4], [10]. Heterodyne transmitters require additional IF and RF filtering stages, while zero-IF transmitters require a low-pass filtering stage after the baseband DAC to remove the images above the Nyquist frequency. These filters are either implemented as active Gm–C stages—which suffer from limited linearity and lead to an increased power consumption—or as passive filters built from lumped or distributed R/L/C components that tend to be lossy and bulky and are difficult to implement with high quality factor on CMOS back ends, resorting to off-chip components. On the other hand, RF-DACs minimize or eliminate the need for filtering stages, which limit the bandwidth and degrade the error vector magnitude (EVM) as will be discussed in Section II, also leading to reduced chip area and cost. Moreover, using an RF-DAC enables higher integration by employing the baseband data generator together with the transmitter on the same chip.

The output spectrum of an RF-DAC contains aliases at multiples of the baseband clock $f_s$ modulated with the carrier frequency. Without oversampling, non-return-to-zero (NRZ) baseband data will result in a zero-order-hold interpolation at the output, which translates to a sinc frequency response. The sinc function nulls occur at multiples of the clock rate away from the carrier frequency, and the first peak is 13 dB lower in magnitude. Alternatively, higher order interpolations or filtering can be synthesized by running the sampling clock higher than the Nyquist rate, at the expense of increased power consumption [3]–[7], [11]. Adequate filtering can also be obtained using the output RF matching network, especially at high sampling clock frequencies where the clock harmonics fall out of band.

This paper presents the design of building blocks for the proposed mm-wave transmitter solution shown in Fig. 1. The transmitter topology is based on a Cartesian architecture consisting of two RF-DACs that are driven in quadrature and current-combined at the output. In order to minimize
the transmit EVM and meet spectral mask requirements, a calibration scheme is proposed where the transmitter output is coupled and downconverted with a wideband receiver to determine required predistortion settings for the $N$-bit RF-DACs. The predistortion settings are then stored in an on-chip memory, which serves as a lookup table in the operation mode. Furthermore, reconfigurable high-speed digital retiming and deserializing circuitries are included to deserialize a single high-speed bit stream into $N$-bit streams, depending on the RF-DACs resolution. The transmitter chip presented in this paper is based on 2-b RF-DACs; it achieves BPSK, PAM4, QPSK, and 16-QAM modulations, depending on the data rate and the RF-DACs mode of operation; and it is implemented in 32-nm SOI CMOS process. The goal of this paper is to investigate system architectures and circuit design techniques that lead to the highest possible data throughput at $W$-band while maintaining a bit-error rate (BER) < $10^{-3}$. Section II presents system analysis and simulations that dictated the design decisions followed. Section III introduces the circuit implementations of the 2-b RF-DAC, $W$-band LO driver, wideband receiver, and high-speed baseband digital drivers. Section IV presents transmitter and receiver link measurements, and Section V concludes this paper.

II. SYSTEM ANALYSIS

A. System Architecture

A Cartesian system architecture is employed in this paper, as shown in Fig. 1. Cartesian systems are more favorable for high-performance mm-wave systems, due to the relative simplicity of generating and combining quadrature signals. More importantly, Cartesian transmitters do not suffer from bandwidth expansion that is experienced in polar transmitters [9], [10], which require oversampling in the phase path, thus limiting the maximum achievable baud rate. Moreover, outphasing transmitters suffer from similar bandwidth expansion issues, since the transformation needed to synthesize the phase information is inherently a non-linear operation.

B. BER and EVM Requirements

Since one of the applications of wideband mm-wave transceivers is in replacing fiber-optics in last-mile links and data centers, an upper limit of pre-forward-error-correction (FEC) BER of < $10^{-3}$ is imposed in this paper, which guarantees passing FEC limits [12], thus rendering the link error-free after implementing FEC. For M-ary QAM modulation, BER and SNR are related by [13]

$$P_b(M) = \frac{2}{\log_2 M} P_b(\sqrt{M}(1 - \frac{1}{2}P_b(\sqrt{M})))$$  \hspace{1cm} (1)

where $P_b$ is the bit-error probability, and $P_b(\sqrt{M})$ is

$$P_b(\sqrt{M}) \approx M - 1 \cdot Q\left(\sqrt{\frac{\log_2 M \cdot BW}{M - 1}} \cdot SNR\right)$$

where $BW$ is the signal bandwidth at baseband, $BR$ is the bit rate, and $Q(x)$ is the standard $Q$-function. The expression can be further simplified by assuming that the bandwidth $BW$ occupied is equal to the baud rate, leading to

$$P_b(\sqrt{M}) \approx 2\sqrt{M - 1} \cdot Q\left(\frac{3}{M - 1} SNR\right).$$  \hspace{1cm} (3)

EVM is the most widely accepted standard to quantify the performance of wireless transmitters nowadays, and therefore it is desirable to relate EVM directly to BER, then use EVM to derive the required design specifications for the circuit blocks. EVM is defined as the ratio of the root-mean-square (rms) constellation error to the peak constellation symbol. This is the most adopted definition in industry [14], [15] and is used in vector signal analysis (VSA) softwares in commercial oscilloscopes. Assuming an SNR-limited link, such that any error in the received symbols is caused by additive white Gaussian noise, then the SNR represents the average signal power to the average constellation error power. Since EVM is measured with respect to the peak constellation symbol, EVM and SNR are related by the maximum-to-average constellation ratio, such that

$$EVM = 100\% \times \sqrt{\frac{\sqrt{M} + 1}{3 \cdot SNR \cdot (\sqrt{M} - 1)}}$$  \hspace{1cm} (4)

which leads to the simulation results given in Fig. 2 that show the EVM values for a given BER requirement and a choice of modulation format. It can be observed that the EVM requirements are highly relaxed for constant envelope modulations, such as BPSK and QPSK, which make them fitting for very high-frequency mm-wave and sub-terahertz transmitters (>100 GHz), where there is an abundance of frequency bandwidth but a limited output power available. Higher order AM-based m-QAM modulations, such as 16-QAM, 32-QAM, and 64-QAM, offer improved spectral efficiencies, at the cost of a rapid increase in SNR and linearity requirements. For example, the EVM requirement for

---

Fig. 1. Block diagram of proposed $W$-band transmitter. Shaded circuit blocks are presented in this paper.
Fig. 2. BER versus EVM for popular m-QAM modulations.

1e-6 BER changes from 20% for QPSK to 7% for 16-QAM. Modulation formats beyond 64-QAM demand stringent phase noise and dynamic-range requirements, as will be shown in Sections II-D and IV-C.2, and are perhaps best suited for transmitters with operating frequencies below 30 GHz.

An alternative definition of EVM normalizes the error vectors to the average symbol energy [16], which has the advantage of relating EVM to the SNR simply by

\[ EVM_{\text{rms,avg}} = \frac{100}{\sqrt{\text{SNR}}} \]

and is independent of the modulation format. \( EVM_{\text{rms,peak}} \) and \( EVM_{\text{rms,avg}} \) are related by

\[ EVM = EVM_{\text{rms,avg}} \times \frac{((M)^{1/2} + 1)/((M)^{1/2} - 1))}{(3((M)^{1/2} - 1))^{1/2}}. \]

Throughout this paper, unless otherwise stated, EVM refers to the first definition that normalizes to the peak symbol energy \( EVM_{\text{rms,peak}} \).

C. Link Margin Analysis and Choice of Digital Modulation Technique

The difficulty of efficiently generating modulated waveforms, at mm-wave carrier frequencies with sufficiently high SNR to support high-order modulation formats, is exacerbated in Cartesian mm-wave RF-DACs, since there is 3 dB of inevitable loss when combining the quadrature outputs. Therefore, to achieve the lowest joule per bit energy efficiency, the transmitter in this paper is designed to support modulation formats up to 16-QAM—while maintaining an SNR of at least 20 dB (Fig. 2). The required transmit power \( P_{\text{TX}} \) can be calculated using

\[ P_{\text{TX}} = P_{\text{RX}} + \text{FSPL} - (G_{\text{TX}} + G_{\text{RX}}) \]  \hspace{1cm} (5)

where \( P_{\text{RX}} \) is the required power at the receiver in order to maintain the targeted SNR over 30 GHz of bandwidth. Assuming a receiver noise figure of 20 dB, \( P_{\text{RX}} \) is estimated to be equal to -29 dBm. \( \text{FSPL} \) is the free-space loss and is 71 dB at a 1-m distance, and \( G_{\text{TX}} \) and \( G_{\text{RX}} \) are the transmit and receive antenna gains, respectively. Using (5) and assuming waveguide horn antennas with 20 dB of gain, the minimum required transmitter output power is found to be 2 dBm. In conclusion, RF-DACs with a minimum resolution of 2 b and 2-dBm output power are needed for the transmitter, while a wideband design is required in the receiver with relaxed noise figure requirements.

D. EVM Degradation Sources and Design Challenges

There has been a number of contributions studying the effect of system non-idealities on the transmit EVM at mm-wave frequencies [17]. Sources of non-idealities, such as non-linearity, in-band noise, IQ imbalance, and LO leakage, have been studied extensively, and a number of design techniques were developed to tackle these challenges. This paper focuses on more fundamental and less explored limitations of the EVM.

1) Gain Ripple and Group Delay: In order to meet mask requirements and maximize spectral efficiency in wireless transmitters, pulse shaping filters are implemented, and such filters cause intersymbol interference (ISI) in the transmitted symbols. However, the overall BER is minimally affected, since a complementary filter is implemented in the receiver to equalize the magnitude and phase distortion caused by the transmit filter.

However, for very wideband applications, it proves useful to model the effect of the transmit and receive channel frequency responses on the EVM. The channels can be modeled as non-ideal analog frequency filters with a certain magnitude ripple and non-linear phase response (non-constant group delay). The filter’s non-ideal impulse response \( H_f \) leads to an increase in the ISI, thus degrading the EVM for a given SNR, such that [18]

\[ EVM_{\text{rms,avg}} = \sqrt{\sum_{k=-\infty}^{\infty} \sum_{k \neq 0} \frac{|H_f(t_0 + kT_s)|^2}{|H_f(t_0)|^2}}. \]
Using an approximate, yet more accessible expression, the EVM is given by [19]

\[
EVM_{\text{rms,avg}} = \sqrt{\Delta a_{\text{rms}}^2 + \left[\tan(\Delta \Phi_{\text{rms}})\right]^2} \quad (7)
\]

where the \(\Delta a_{\text{rms}}\) is the rms value of the magnitude ripple, and \(\Delta \Phi_{\text{rms}}\) is the rms value of the deviation of the phase response from an ideal linear phase response, over the frequency of interest.

2) Frequency Synthesizer Phase Noise: There has been several research efforts in building W-band on-chip PLLs in CMOS and SiGe BiCMOS for radio and radar applications [20]–[24]. A significant performance edge has been documented in SiGe BiCMOS, due to lower transistor noise and lower \(1/f\) flicker noise corner. This is in comparison with FET transistors where the corner frequency of highly scaled FETs can reach tens of megahertz. The LO phase noise results in a frequency spreading of the carrier frequency that can be modeled as a Gaussian distribution. As a result, the EVM degradation is given by [16]

\[
EVM_{\text{rms,avg}} = \sqrt{2 - 2e^{-\sigma^2/2}} \quad (8)
\]

where \(\sigma\) is the rms phase error in radians and is derived from the integrated LO phase noise. For small values of \(\sigma\) phase error and \(N\) being the division ratio needed to bring the VCO frequency down to the reference signal, if \(PN_{\text{avg}}\) is the average phase noise of the reference in dBc/Hz within the PLL bandwidth \(BW_{\text{PLL}}\), then the \(EVM_{\text{rms,avg}}\) can be approximated as

\[
EVM_{\text{rms,avg}} \approx \sigma \approx \sqrt{2N^210PN_{\text{avg}}/10BW_{\text{PLL}}} \quad (9)
\]

For a PLL bandwidth of 10 MHz and ignoring the phase noise contribution of the dividers in the loop, then to achieve lower than 3% \(EVM_{\text{rms,avg}}\), a phase noise lower than \(-94\) and \(-104\) dBc/Hz at 1- and 10-MHz offsets are needed, respectively. Recently published SiGe BiCMOS PLL designs are capable of meeting these performance metrics [21], [25]. Unfortunately, mm-wave CMOS PLL designs still lag in phase noise performance [20], [24], which is one of the fundamental reasons why modulation schemes higher in order than 64-QAM are not popular in mm-wave CMOS transceivers.

3) LO Leakage: There are two sources that lead to LO leakage or carrier feedthrough in direct-conversion transmitters: IQ dc offsets in the modulator, which translate into shifting in the origin point of the constellation diagram and direct LO leakage from the LO to the RF port in the modulator—due to finite isolation between the two ports or due to layout asymmetry in the upconversion mixer. LO leakage resulting from either source can be corrected for by applying a differential dc offset to the I and Q upconversion mixers. This requires a 2-D calibration, since the differential dc offsets required for the I and Q mixers are usually not equal [26].

For switching-mode modulators or RF-DACs, this can be achieved using digital predistortion by applying the appropriate digital words that represent the effect of a differential dc offset across the positive and negative terminals of the baseband switching stage in the modulator. Typical LO leakage values at mm-wave frequencies range from \(-30\) to \(-24\) dBc, which translate to 3%–6% \(EVM_{\text{rms,avg}}\)

\[
EVM_{\text{rms,avg}} = \sqrt{10LO_{\text{lk}}/10} \quad (10)
\]

where \(LO_{\text{lk}}\) is the carrier leakage and is measured relative to the average power of the transmitted signal. The average transmitted power can be calculated using the peak-to-average power ratio of the corresponding m-QAM waveform. DC offsets in the receiver, resulting from LO leakage in the transmitter, can be removed with a high-pass filter (e.g., \(1/f\) block) or using digital baseband filtering. Nevertheless, LO leakage degrades the dynamic range of the transmitter and directly impacts its EVM and BER.

### III. BUILDING BLOCKS

A. 2-b RF-DAC

A major performance limitation in the design of high-speed modulators is the dependence of the their input and output impedances on the digital code word, which leads to distorting the output waveform at every switching event and, as a result, a degradation in the transmit EVM [10]. This paper proposes the 2-bit RF-DAC design shown in Fig. 3(a) to mitigate the aforementioned issue. The fixed current source at the bottom of the cell ensures that the biasing condition of the switching quad remains unchanged, thus fixing the output impedance of the RF-DAC and making it independent of the digital-word transmitted [27].
For an $N$-bit RF-DAC, the switching transistors are size-segmented in a binary fashion. This is achieved in a single transistor layout by grouping the gate fingers into $N$ groups with a progressively increasing number of fingers in each group that follows a binary $2^N$ sequence. For $N$ baseband inputs: $b_0, b_1, \ldots, b_{N-1}$, phase modulation is achieved by the switching of the most significant bit $b_{N-1}$. Afterward, amplitude modulation is achieved by switching the remaining bits $b_0, b_1, \ldots, b_{N-2}$. The RF-DAC output power can be written as

$$P_{\text{out}} = P_{\text{sat}} + 20 \log \left| \sum_{k=0}^{N-1} \frac{(-1)^{b_k} 2^k}{2^N - 1} \right|$$

where the sign bit $b_{N-1}$ determines the phase of the output waveform, which can take one of two values: $0^\circ$ or $180^\circ$, and $P_{\text{sat}}$ is the saturated output power. As an example, for a 4-b RF-DAC, the minimum constellation symbol (occurs at 0001 or 1110) is approximately 24 dB below $P_{\text{sat}}$, which is in agreement with the classical 6 dB per bit resolution in baseband DACs. As mentioned before, the dynamic range is limited by factors such as the SNR of the RF-DAC and the LO leakage.

The design shown in Fig. 3(a), which is an alternative constant-current RF-DAC topology to the one presented in [8] and [10], allows the segmented transistors to be driven by the 1-V thin-oxide SOI CMOS transistors rather than the thick-oxide transistors, thus permitting the DAC to switch at higher speeds. Another advantage of this circuit is that it allows the modulator to operate deep in saturation, without the need for back-off. Therefore, the switching quad transistors are driven differentially by an LO amplifier and into saturation. Since it is not necessary to back off to produce amplitude modulation, the modulator can be made efficient at peak output power.

This topology can be thought of as an outphasing amplifier [Fig. 3(b)], where the amplifiers are biased in Class-A and are $180^\circ$ out-of-phase at all times, which leads to poor efficiency for any constellation outputs other than the peaks. It does, however, allow for the fastest operation, since the output impedance is fixed and is not dependent on the digital word, which would otherwise cause amplitude distortion and phase distortion in the EVM constellation.

The 2-b RF-DAC design implemented in this paper is shown in Fig. 4(a). Fig. 4(b) presents the simulated and measured small-signal parameters of the transmitter chain, along with the measured saturated output power of 3–5 dBm at 85–105 GHz. The $S_{21}$ of the transmitter is measured by setting the digital input to a static value of $(D_0 = 1, D_1 = 1)$, thus effectively enabling a single arm of the Gilbert cell. The output matching network of the modulator is reduced to low- $Q$ shunt 150-pH inductors on both sides of the differential output. The minimal matching at the output of the RF-DAC is used to achieve the highest possible bandwidth with minimal gain ripple and phase distortion, at the expense of reduced output power and efficiency. Measurements show 6% peak efficiency for the modulator stage and 1.6% efficiency for the entire I/Q transmitter.

B. High-Speed Digital

The digital blocks are based on true single-phase clocked (TSPC) logic [28], [29]. TSPC logic is capable of operating at higher speeds compared with static CMOS logic and at a more moderate power consumption than the CML logic family.

1) TSPC Latch: The basic building block for the TSPC logic is the latch shown in Fig. 5. When the CLK is high (1 V), the latch is in the transparent mode and can be modeled as two inverters in series; and the input logic is buffered to the output. When the CLK is low (0 V), both inverters are disabled, and the latch is in hold mode. Only the pull-up PMOS transistors are still active, while the pull-down NMOS transistors are disconnected. As a result, the latch output remains at the same logic level irrespective of the input logic $D$. An edge-triggered TSPC flip-flop can be built by adding a second latch, which is activated at the opposite clock level from the first latch. The flip-flop is used as a retimer to align all the high-speed baseband data with the respect to the clock.

2) TSPC Divider: The TSPC digital divider consists of nine transistors as shown in Fig. 6. A closer look at the circuit reveals that it resembles a flip-flop in feedback. When $CLK = 0$, the three-transistor circuit from the input to node
A is a negative D-latch with inverted output. When CLK = 1, the six-transistor circuit from node A to the output is a positive D-latch. Therefore, with the feedback connection, the circuit is capable of dividing digital clocks at higher speeds than traditional CMOS digital dividers, due to the small capacitive loads arising from the low transistor count in the topology.

3) Deserializer Design: To drive the baseband inputs of the 2-b RF-DAC, the NRZ pseudorandom binary sequence (PRBS) data are fed into the chip from an external source. It is desirable to be able to dynamically change the operation mode of the RF-DAC and, consequently, the choice of the modulation format for the transmitter. The digital driver in Fig. 7 accepts an NRZ PRBS data stream and retimes it the with clock while—in parallel—also deserializes the incoming data into two streams, each at half of the data rate. Two switches select between passing the deserialized data or passing the two identical retimed data streams which effectively turns the 2-b RF-DAC into a 1-b RF-DAC. This capability allows the transmitter to seamlessly change its modulation scheme.

Measurement results in Figs. 8 and 9 show that the digital driver is capable of retiming and demultiplexing PRBS31 data streams faster than 30 and 20 Gb/s, respectively. This performance makes the digital blocks in this design sufficient to support 60- and 40-Gb/s data rates in QPSK and 16-QAM operating modes in the transmitter.

C. W-Band LO Driver

To drive the RF-DAC into saturation at reasonable LO power levels, the RF-DAC is preceded by an LO amplifier consisting of three differential cascode stages (Fig. 10). All the LO amplifier stages are biased in class-A and they operate under saturation, so as to reduce the I/Q amplitude imbalance introduced by the I/Q coupler. Transistor sizes in each stage are incrementally increased by a factor of 1.5 to ensure saturating the amplifier stages while providing sufficient output power to drive the RF-DAC. Fig. 11 presents the simulated and...
measured S-parameters of the LO amplifier. Small-signal gain > 10 dB is obtained from 90–110 GHz with a peak gain of 20 dB at 100 GHz. The differential amplifier consumes 40 mW from a 1-V supply. To generate the quadrature signal, a 50-Ω-matched varactor-tuned transformer-based quadrature coupler [30] is used (Fig. 12). The accumulation-mode MOS varactors are externally controlled with analog voltages. Inevitably, the coupler produces quadrature amplitude errors, which are corrected by the saturated LO amplifiers. Any AM–PM distortion resulting from saturating the driver amplifiers can be corrected by retuning the varactors.

D. W-Band Wideband Receiver

This section discusses the design procedure for the wideband W-band receiver. Since the receiver is meant to be used in the feedback path (Fig. 1), it should be capable of downconverting and amplifying wideband modulated signals centered around any carrier frequency within the band. Ideally, the receiver will cover the entire W-band (75–110 GHz), within its 3-dB bandwidth.

1) Wideband Passive Mixer: Passive mixers have become the standard design choice for RF receivers. This is due to the poor linearity of active mixers and their high 1/f noise, an issue which is exacerbated in short-channel devices as discussed earlier. Passive mixers offer superior linearity especially if their output is presented with a low impedance, forcing them to operate in current mode. Furthermore, they exhibit little 1/f noise, since no dc current flows through the devices, a useful feature for direct-conversion systems. Passive mixers at mm-wave frequencies, however, are not as common due to their high conversion loss. A passive mixer is employed in this paper, motivated by the low ON-resistance of the devices in 32-nm SOI and by the inherently wider bandwidth offered by passive mixers when compared with active mixers. Assuming the parasitic capacitance of the switches resonate with the inductance presented by the input balun [Fig. 13(a)] and since it is not possible to generate a square LO waveform at W-band frequencies, the single-ended input impedance for the passive mixer architecture can be expressed as [31]

$$Z_{in}(\omega) \simeq R_{SW} + \frac{1}{16}(Z_{TIA}(\omega) - \omega_{LO})$$

where $R_{SW}$ is switch ON-resistance and $Z_{TIA}$ is the input impedance of the transimpedance amplifier (TIA). The impedance looking into the receiver is then

$$Z_{in,RX}(\omega) = R_p + R_s + 2Z_{in}(\omega)$$

where the factor of 2 accounts for the differential input impedance of the passive mixer, which is dominated by the switch resistance. In fact, it is desirable to choose the switch sizes such that $Z_{in,RX}(\omega) \sim 50 \Omega$ to achieve a wideband impedance match. It is also desirable to reduce the contribution from the TIA input impedance, that is to suppress the image current appearing $\omega_{LO} - \omega_{RF}$. The impedance created at the image can cause different high- and low-side conversion gains [31]—an unwelcome effect as discussed in Section II-D.1. For $L_p = L_s \sim 250 \mu$H, which are the inductor values needed to resonate out the reactive part of the mixer input impedance, and for a quality factor $Q = 12$ at 90 GHz, then $Z_{diff}(\omega) \sim 24 \Omega$ is needed. This translates into $R_{SW} \sim 12 \Omega$ or a 30-μm device width in 32-nm SOI including layout parasitics. The mixer conversion gain simulation in Fig. 13(b) shows 10 dB of loss over the entire bandwidth, with symmetric upper and lower sideband gains around 94-GHz carrier frequency. The high conversion loss is expected, since about 3.9 dB of loss arises from the mixing operation [32], and an additional loss of ~ 3 dB is attributed to the ON-resistance of the switches, which form a voltage divider with the TIA input resistance leading to $20 \log_{10}(Re(Z_{TIA})/(R_S + Re(Z_{TIA})))$ of loss and ~ 2-dB loss of the input balun. Furthermore, if there is time period $t_{ov}$ where both switches are ON, (i.e., overlap in the differential driving), then the conversion gain is degraded by $20 \log_{10}(\cos(\pi t_{ov})/(2\Omega_{LO})))$ [33]. The switching transistors are biased near threshold (0.3 V) to minimize $t_{ov}$ without significantly increasing the required LO power to fully switch on the transistors. The simulated mixer IP1dB is +13 dBm, which allows the mixer to sample a high-dynamic-range signal without distortion.

2) Wideband IF-Amplifier: Traditionally, PMOS devices needed to be sized larger in order match the transconductance of NMOS devices for a given bias current, which increased the output capacitance and consequently affected the bandwidth of the CMOS inverter design. Another concern with the inverter is the smaller available voltage headroom, since the devices are biased at half the supply voltage. The closing gap in speed between NMOS and PMOS devices [34] led to the revival of the CMOS inverter stage (Fig. 14). Moreover, the simulation in Fig. 15 reveals that both the NMOS and PMOS devices in 32-nm CMOS SOI maintain a high $f_T$ as the drain-source voltage $V_{DS}$ is reduced, as long as the device remains in saturation. This is in agreement with
the observation that nanoscale FETs have weak dependence on $V_{DS}$ as long as they are biased at peak $f_T$ condition (0.2–0.4 mA/μm) [35].

The inverter also offers superior linearity performance when compared with an NMOS stage with a resistive load, since the combination of the NMOS and PMOS devices helps maintain a constant transconductance for different input swings. Furthermore, the inverter leads to the same noise for half the bias current, since the additional noise by the PMOS device is suppressed by an increased effective transconductance $g_{m,\text{eff}} = g_{m,\text{p}} + g_{m,\text{n}}$. The noise figure of a single inverter stage is given by [33]

$$F = 1 + \frac{R_f + Z_{in}}{g_m R_f + 1} \left( \frac{\gamma g_m + 1/Z_L}{Re(Z_{in})} \right)^2$$

where $R_f$ is the feedback resistance, $Z_{in}$ is the input impedance, and $Z_L$ is the load impedance. Recognizing that
gain $S_{21} > 20$ dB. Cadence simulations show a minimum NF value of 3.5 dB. The discrepancy at higher frequencies can be attributed to ignoring the effects of the reactive elements such as $C_{gs}$ and $C_{gd}$ in the matching networks, leading to optimistic predictions from the analysis. Inductors can be used at the input and output matching networks to resonate out the capacitances. No resonant elements are used in this particular design in favor of compactness and to maintain a wide bandwidth.

**IV. MEASUREMENTS**

**A. RF-DAC Measurements**

A breakout of the RF-DAC shown in Fig. 4(a) was measured by applying an external W-band LO signal using the VDI AMC-335 multiplier chain through a WR-10 waveguide probe (Fig. 17). The PRBS31 data and clock inputs are obtained from a Tektronix 30-Gb/s two-channel programmable pattern generator PPG3002. The modulated RF output is then downconverted using an external Quinstar QMB waveguide-balanced mixer to a 10-GHz IF. The modulated IF output is then directly applied to a real-time 100-GS/s DPO73304DX Tektronix scope with 33 GHz of analog bandwidth. The constellation diagram, frequency spectrum, and eye diagrams are then observed using the SignalVu VSA software.

Fig. 18(a) presents a 1-Gb/s RF-DAC output spectrum in PAM2/BPSK mode, after downconversion to 10 GHz, over 14-GHz frequency span. The spectrum follows the sinc function discussed earlier, with clear nulls at multiples of the 1-GHz sampling clock. Fig. 18(c) and (d) present the eye diagrams at maximum data rates obtained in PAM2/PAM4 modes, both at 12 Gb/s, with SNR values of 9 and 13 dB, respectively. Fig. 18(b) gives a summary of the measured EVM in BPSK and PAM4 modes, for different data rates.

**B. Transmitter Measurements**

1) Evaluating Transmitter Impairments: The measurement setup in Fig. 19(a) depicts the I/Q transmitter configuration as a single-sideband (SSB) transmitter, to evaluate the quadrature phase and amplitude imbalance, by means of measuring the sideband suppression or image rejection ratio (IRR) of the SSB transmitter. A 100-MHz tone is generated using an arbitrary waveform generator (AWG) that is controlled in software by
MATLAB, such that the first channel of the AWG generates the 100-MHz sinusoidal tone and the second channel generates the Hilbert transform (quadrature signal in the case of a single tone) of the signal in channel 1. The two channels are synchronized and phase adjusted such that the two quadrature signals arrive at the I and Q input data pads of the die with correct phases. The output of the transmitter is then downconverted to an intermediate frequency in order to display the output spectrum on a spectrum analyzer. Fig. 19(b) depicts an example spectrum where the lower sideband is suppressed, and the IRR and LO leakage can be readily calculated.

Fig. 19(c) shows the measured sideband suppression and LO leakage as a function of the LO frequency. Sideband suppression is measured by tuning the varactors of the I/Q coupler, presented in Section III-C, each time the LO frequency is changed, in order to set the optimum coupler tuning voltages that achieve the highest sideband suppression value. SSB suppression better than 30 dB over the entire band is achieved. Furthermore, Fig. 19(c) shows that the LO leakage can vary from $-30$ to $-18$ dBc, which can lead to a worst

MATLAB, such that the first channel of the AWG generates the 100-MHz sinusoidal tone and the second channel generates the Hilbert transform (quadrature signal in the case of a single tone) of the signal in channel 1. The two channels are synchronized and phase adjusted such that the two quadrature signals arrive at the I and Q input data pads of the die with correct phases. The output of the transmitter is then downconverted to an intermediate frequency in order to display the output spectrum on a spectrum analyzer. Fig. 19(b) depicts an example spectrum where the lower sideband is suppressed, and the IRR and LO leakage can be readily calculated.

Fig. 19(c) shows the measured sideband suppression and LO leakage as a function of the LO frequency. Sideband suppression is measured by tuning the varactors of the I/Q coupler, presented in Section III-C, each time the LO frequency is changed, in order to set the optimum coupler tuning voltages that achieve the highest sideband suppression value. SSB suppression better than 30 dB over the entire band is achieved. Furthermore, Fig. 19(c) shows that the LO leakage can vary from $-30$ to $-18$ dBc, which can lead to a worst
case \( EVM_{\text{rms,avg}} \) of approximately 13% (10). Additionally, measurements demonstrate that the carrier feedthrough shows an increasing trend with the carrier frequency. This is to be expected, since the capacitive coupling through the gate–drain capacitance in the switching quad transistors becomes the dominant mechanism for the leakage at mm-wave frequencies.

2) Digital I/Q Modulator Measurements: A similar measurement setup is used to measure the I/Q transmitter chip [Fig. 20(a)]. Two PRBS data sequences representing I/Q data and a full-rate clock are obtained from the PPG and enter the chip through a 50-GHz GSGS probe, using 2.4-mm coaxial cables. External wideband phase shifters are used to compensate for any phase mismatch between the different cables.

Fig. 20 presents QPSK and 16-QAM real-time EVM, constellation and eye diagram measurements for a 95-GHz LO carrier, at different data rates and up to a maximum of 20 Gb/s in QPSK and 4 Gb/s in 16-QAM. At 20-Gb/s QPSK, the RF output occupies 20 GHz of bandwidth (85–105 GHz) and is downconverted with an 85-GHz LO source to a 10-GHz IF. The constellation shows an EVM of 24%, which corresponds to 3e-5 of expected BER value (Fig. 2). This is, to the best of our knowledge, the highest data rate ever reported for a digital CMOS transmitter below saturation and also due to the limited isolation between the I and Q paths, which degrades the EVM and the effective number of bits of the I and Q RF-DACs to less than 2 b per DAC. The transmitter has 4 dBm of output power and consumes 220 mW. Table I presents a comparison with the state-of-the-art W-band transmitters in silicon.

In order to test the transmitter performance in a practical physical link setting, the setup in Fig. 21 was constructed. The output of the transmitter was connected to a WR-10 waveguide horn antenna through a waveguide probe. A waveguide-based W-band receiver connected to another horn antenna was placed half-meter away from the transmitter, and in its line of sight [Fig. 22(a)]. The downconverted IF output was then connected to a real-time scope, and the EVM values for QPSK constellations at different data rates were measured. As expected, the maximum data rate fell to 12 Gb/s [Fig. 22(b)], limited by the SNR of the link.
TABLE I
COMPARISON WITH mm-WAVE TRANSMITTERS BELOW 100 GHz IN SILICON-BASED TECHNOLOGIES

<table>
<thead>
<tr>
<th>Reference</th>
<th>8</th>
<th>9</th>
<th>30</th>
<th>36</th>
<th>37</th>
<th>38</th>
<th>39</th>
<th>40</th>
<th>This Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>45 nm SOI</td>
<td>40 nm CMOS</td>
<td>0.18 μm BiCMOS</td>
<td>65 nm CMOS</td>
<td>45 nm SOI</td>
<td>0.13 μm BiCMOS</td>
<td>65 nm CMOS</td>
<td>40 nm SOI</td>
<td>45 nm SOI</td>
</tr>
<tr>
<td>Frequency [GHz]</td>
<td>90</td>
<td>60</td>
<td>88</td>
<td>68/102</td>
<td>94</td>
<td>77</td>
<td>60</td>
<td>78.5</td>
<td>100</td>
</tr>
<tr>
<td>Modulation</td>
<td>ASK</td>
<td>OOK</td>
<td>16QAM</td>
<td>32QAM</td>
<td>16QAM</td>
<td>64QAM</td>
<td>QPSK</td>
<td>16QAM</td>
<td>64QAM</td>
</tr>
<tr>
<td>Integration Level</td>
<td>On Wafer</td>
<td>On Wafer</td>
<td>On Wafer</td>
<td>On Wafer</td>
<td>On Wafer</td>
<td>On Wafer</td>
<td>On Wafer</td>
<td>On Wafer</td>
<td>On Wafer</td>
</tr>
<tr>
<td>Peak Data Rate [Gbps]</td>
<td>15</td>
<td>3.5/7</td>
<td>10/18.75 (1 m link) (Ext. AWG)</td>
<td>56 (Ext. AWG)</td>
<td>1</td>
<td>18</td>
<td>3.5/6</td>
<td>3</td>
<td>10/4 (189 mm link)</td>
</tr>
<tr>
<td>Post [dBm]</td>
<td>19</td>
<td>5.3</td>
<td>&gt; 6</td>
<td>-8.4</td>
<td>&gt; 10</td>
<td>9</td>
<td>9.6</td>
<td>12</td>
<td>16.9</td>
</tr>
<tr>
<td>TX Power Consumption [W]</td>
<td>0.89</td>
<td>0.04*</td>
<td>0.25</td>
<td>0.26</td>
<td>2.1 (0.8*)</td>
<td>- 0.4</td>
<td>0.38</td>
<td>0.1</td>
<td>1.27/0.93</td>
</tr>
</tbody>
</table>

* RF-DAC/PA core power consumption only.

C. Receiver Measurements

1) TIA Measurements: The S-parameters of the CMOS inverter-based TIA, presented in section III-D.2, were measured with a VNA up to 67 GHz. The results are shown in Fig. 23(a). The TIA exhibits 22 dB of peak gain and a small-signal 3-dB bandwidth of 25 GHz. Using a Keysight 346C (10 MHz–26.5 GHz) noise source and E4448A spectrum analyzer, the noise figure was also measured [Fig. 23(b)]. Measured noise figure is ~1 dB higher than simulated, which is attributed to the inaccuracy of the transistor noise model. The high noise figure at dc frequencies is due to flicker noise and the ac coupling used, while the noise figure increase at higher frequencies is predicted by (15) and (16), since the effective input impedance decreases (Fig. 16). Furthermore, large-signal probe measurements revealed an OP1dB of 0 dBm. The TIA consumes 18 mW from a 1-V supply.

Simulated and measured group delay [Fig. 23(c)] show good agreement and flat group delay response up to 60 GHz. Using (7), the group delay is expected to lead to 0.2° rms phase error or an equivalent degradation in EVM of 0.35%. No significant gain ripple is observed, and therefore the maximum baud rate supported by the TIA is limited by its 3-dB bandwidth.

Moreover, eye diagram measurements were carried out, to further validate the measurement results of the TIA noise figure, gain and bandwidth. A 65-GSa/s Keysight M8195A AWG was used to generate multi-level baseband signals that are fed to the TIA. The TIA output was then connected to a 33-GHz Keysight real-time oscilloscope, and the eye diagrams were captured in Fig. 24. Good eye openings were obtained up to 40 Gb/s with NRZ input waveforms and 30 Gb/s with PAM-4.
2) Receiver Measurements: A breakout test cell of the receiver in Fig. 14 was measured. The receiver is comprised of the wideband passive mixer and the wideband TIA discussed earlier, and is therefore expected to have sufficiently wide bandwidth to cover the entire W-band. To verify, the conversion gain was first measured [Fig. 25(a)]. The measurement reveals that the receiver exhibits a very wide (>30 GHz) bandwidth with approximately 10 dB of conversion gain. The ripple in the measured conversion gain is a measurement artifact, and is due to measurement uncertainties in the power meter reading and in the output power of the W-band VDI multiplier used to generate the input RF signal. Furthermore, the noise figure was measured [Fig. 25(b)], and large-signal measurements reveal a −10-dBm input P1dB, limited by the TIA.

In order to prove that the receiver has sufficient bandwidth and dynamic range to downconvert wideband modulated signals, as intended in the system proposed in Fig. 1, modulated QAM signals are generated using the AWG M8195A at an IF of 10 GHz, the output is mixed with an 85-GHz signal, and the RF signal centered at 95 GHz is fed into the RF input of the receiver. The signal is then downconverted and amplified using the on-chip receiver. The receiver is capable of downconverting > 50-Gb/s PRBS9 data sequence with low EVM values. EVM measurements are shown in Fig. 26.

It is worth noting that the phase noise of a continuous wave (CW) signal from the Keysight E8267D signal generator used in the experiments is $\sim -110$ dBc/Hz at 100-KHz offset, in the 10–20 GHz frequency region [41]. This imposes a lower EVM limit of approximately 1% (9), where the multiplication factor $N = 8$ to arrive to W-band frequencies. This leads to a BER close to or greater than 1e-3 for 1024-QAM or higher order m-QAM modulation schemes. For reference, 1024-QAM and 2048-QAM constellation measurements are shown in Fig. 25(e) and (f) at 10 and 5.5 Gb/s, respectively, with 1e-2 expected BER values.

V. CONCLUSION

A state-of-the-art W-band transmitter and receiver chips using 32-nm CMOS SOI were presented. The 94-GHz transmitter consists of two current-combined RF-DACs that are driven in quadrature, and is capable of supporting multiple modulation schemes including BPSK, PAM4, QPSK, and 16-QAM with maximum data rates of 12, 12, 20, and 4 Gb/s, respectively, all with EVM values that are expected to achieve $< 10^{-3}$ of uncoded BER. The receiver chip demonstrated greater than 30 GHz of bandwidth enabled by a wideband passive mixer and a CMOS-inverter-based TIA design. Receiver measurements proved its capability to support highly complex modulated waveforms (> 256-QAM) and at very high data rates, up to 60 Gb/s in 64-QAM. Future work may include increasing the resolution of the mm-wave RF-DACs, in order to support more spectrally efficient modulation formats, and to implement digital predistortion and filtering.

ACKNOWLEDGMENT

The authors would like to thank Hsin-Chang Lin for his assistance with the IF amplifier design.

REFERENCES


Hasan Al-Rhubay (S’11) received the B.Sc. degree in electrical engineering from the University of Toronto, Toronto, Canada, in 2013 and the M.Sc. degree from the University of California at San Diego, La Jolla, CA, USA, in 2015, where he is currently pursuing the Ph.D. degree. He has previously held internships with Advanced Micro Devices, Toronto, ON, Canada, and Nokia Bell Labs, Murray Hill, NJ, USA, where he was involved in research and development efforts in mm-wave phased array systems. His current research interests include the design of millimeter-wave-integrated circuits and systems.
Gabriel M. Rebeiz (S’86–M’88–SM’93–F’97) received the Ph.D. degree from the California Institute of Technology, Pasadena, CA, USA.

From 1988 to 2004, he was with the University of Michigan, Ann Arbor, MI, USA. His group has optimized the dielectric-lens antenna, which is the most widely used antenna at millimeter-wave and terahertz frequencies. His group also developed 6–18-GHz, 30–35-GHz, 40–50-GHz, 77–86-GHz, and 90–110-GHz 8- and 16-element phased arrays on a single silicon chip, the first silicon phased-array chip with built-in-self-test capabilities, the first wafer-scale phased arrays with on-chip antennas, and the first SiGe millimeter-wave silicon passive imager chip at 85–105 GHz. His group also demonstrated high-performance RF MEMS tunable filters at 0.7–6 GHz, RF MEMS phase shifters at 1–100 GHz, and high-power high-reliability RF MEMS metal-contact switches. As a consultant, he helped develop 24- and 77-GHz single chip SiGe automotive radars, phased arrays operating at X- to W-band for defense and commercial applications (SATCOM, automotive, and point-to-point), digital beamforming systems, and several industrial RF MEMS switches. He has graduated 64 Ph.D. students and 22 post-doctoral fellows. He is currently a Distinguished Professor and the Wireless Communications Industry Endowed Chair Professor of Electrical and Computer Engineering with the University of California at San Diego (UCSD), La Jolla, CA, USA. He also leads a group of 20 Ph.D. students and post-doctoral fellows in the areas of millimeterwave radio frequency integrated circuits (RFICs), tunable microwaves circuits, RF MEMS, planar millimeterwave antennas, and terahertz systems. He has authored or co-authored over 670 IEEE publications, and authored the book RF MEMS: Theory, Design and Technology (Wiley, 2003).

Dr. Rebeiz is a member of the National Academy. He has been a Distinguished Lecturer of IEEE MTT-S, the IEEE Antennas and Propagation Society (AP-S), and the IEEE Solid-State Circuits Society. He was a National Science Foundation Presidential Young Investigator, a URSI Koga Gold Medal recipient, and the 2003 IEEE Microwave Theory and Technique Society (IEEE MTT-S) Distinguished Young Engineer. He was a recipient of the IEEE MTT-S 2000 and 2014 Microwave Prize, the IEEE MTT-S 2010 Distinguished Educator Award, the IEEE AP-S 2011 John D. Kraus Antenna Award, the 2012 Intel Semiconductor Technology Council Outstanding Researcher in Microsystems, the R&D100 2014 Award for this work on phased-array automotive radars, the 2014 IEEE Daniel E. Noble Field Medal for his work on RF MEMS, and the IEEE AP-S 2015 Harold A. Wheeler Applications Prize Paper Award. He was a recipient of the 1997–1998 Eta Kappa Nu Professor of the Year Award, the 1998 College of Engineering Teaching Award, and the 1998 Amoco Teaching Award given to the best undergraduate teacher at the University of Michigan, and the 2008 Teacher of the Year Award of the Jacobs School of Engineering, UCSD. His students have received a total of 22 best paper awards from the IEEE MTT-S, RFIC, and AP-S conferences. He has been an Associate Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES.