Abstract—This paper presents a W-band 2\times2 focal-plane array (FPA) for passive millimeter-wave imaging in a standard 0.18 \mu m SiGe BiCMOS process \( f_T/f_{max} = 200/180 \) GHz. The FPA incorporates four Dieke-type receivers representing four imaging pixels. Each receiver employs a direct-conversion architecture consisting of an on-chip slot folded dipole antenna, an SPDT switch, a low noise amplifier, a single-balanced mixer, an injection-locked frequency tripler (ILFT), an IF variable gain amplifier, a power detector, an active bandpass filter and a synchronous demodulator. The LO signal is generated by a shared Ka-band PLL and distributed symmetrically to four local ILFTs. The measured LO phase noise is \( \approx 93 \) dBc/Hz at 1 MHz offset from the 96 GHz carrier. This imaging receiver (without antenna) achieves a measured average responsivity and noise equivalent power of 285 MV/W and 8.1 fW/Hz\( ^{1/2} \), respectively, across the 86–106 GHz bandwidth, which results a calculated NETD of 0.48 K with a 30 ms integration time. The system NETD increases to 3 K with on-chip antenna due to its low efficiency at W-band. MMW images have been generated in transmission mode. This work demonstrates the highest integration level of any silicon-based systems in the 94 GHz imaging band.

Index Terms—Dieke radiometer, direct conversion, focal-plane array, frequency generation, LO distribution, millimeter-wave, on-chip antenna, passive imaging, SiGe, W-band.

I. INTRODUCTION

Silicon technologies have been adopted as the primary platform for development of millimeter-wave (MMW) systems for the target applications of short-range high data-rate wireless communication, automotive radar, sensing and imaging. Within the MMW frequency range (30–300 GHz), there are propagation windows located near 35, 94, 140, 220 GHz [1], where the atmospheric absorption is relatively low; not only in clean air but also through smoke, dust, fog, and clothing. This notion makes a passive mm-wave (PMMW) imaging system an ideal candidate for various applications such as remote sensing, security surveillance (e.g., concealed clothing. This notion makes a passive mm-wave (PMMW) imaging system an ideal candidate for various applications such as remote sensing, security surveillance (e.g., concealed weapon detection at the airport), non-destructive inspection for biological tissues as well as industrial process control [1], [2]. Additionally, the non-invasive nature of passive imaging avoids any public health concern in medical and security applications. PMMW imaging systems operating near the 94 GHz frequency window provide reasonable balance among capability of currently available silicon technology, chip size, spatial resolution, and atmospheric attenuation.

III-V technologies have been commonly used to realize MMW radiometers or passive imaging receivers that are based on multi-chip modules [3], [4]. Recently, benefiting from the aggressive feature size scaling, silicon technology has shown the capability for implementation of W-band passive imaging receivers with fine image and temperature resolution [5]–[8]. However, these efforts are limited to a single receiver/pixel. To reduce the scanning time and enable video rate real-time imaging, a focal-plane array (FPA) could be used with an array of detectors located at the focal-plane of a focusing system. In this paper, we present a W-band 2 \times 2 FPA in a commercial 0.18 \mu m SiGe BiCMOS process \( f_T/f_{max} = 200/180 \) GHz.

II. DIEKE RADIOMETER SYSTEM

The radiometer collects the radiated power from the target object and produces an output voltage proportional to the incident power. The sensitivity or the minimum detectable temperature change of an imaging receiver is characterized by noise equivalent temperature difference (NETD), which is given by (1) for a total-power radiometer in the presence of gain fluctuation [9]:

\[
NETD = T_s \sqrt{\frac{1}{B\tau} + \left(\frac{\Delta G}{G}\right)^2}
\]

where \( T_s \) is the system noise temperature, \( B \) is the bandwidth, \( \tau \) is the integration time which is typically less than 30 ms (standard video imaging rate) in real-time imaging application, and \( \Delta G/G \) denotes the total gain fluctuation. To build a useful imaging system, the NETD needs to be below 1 K, while less than 0.5 K is preferred for good imaging quality [3]. The gain fluctuation problem can be solved by periodically chopping above the gain fluctuation frequency using Dicke architecture [10]. Fig. 1 shows the block diagram of a Dicke radiometer that employs two synchronized SPDT switches: the one at the front-end switches between the antenna and a reference load, while the one after the detector demodulates the signal by multiplying it with \( \pm 1 \) in the opposite phase with respect to the front-end switch. In addition to the gain fluctuation problem, the 1/\( f \) noise is another source of low frequency disturbance, affecting the system NETD in a similar way. Therefore, the

Manuscript received October 05, 2011; revised May 22, 2012; accepted May 22, 2012. This work was supported in part by an NSF grant under contract ECCS-1002294.

Z. Chen was with the University of California at Irvine, Irvine, CA 92697 USA. He is now with the Beijing Institute of Technology, Beijing, China (e-mail: zhiming@cuit.edu).

C.-C. Wang is with Peregrine Semiconductor, San Diego, CA 92121 USA.

H.-C. Yao is with HTC Corporation, Taiwan.

P. Heydari is with the University of California at Irvine, Irvine, CA 92697 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2012.2209775
chopping frequency also needs to be higher than the $1/f$ noise corner frequency.

Power detector serves as the core of an imaging pixel. The common figure-of-merit to evaluate performance of a power detector is noise-equivalent power (NEP) $[W/\sqrt{Hz}]$ defined by (2) as detector's output rms noise voltage $v_n [V/\sqrt{Hz}]$ divided by responsivity, $R$, which is a measure of detector gain and equals to the output DC voltage divided by the input RF power.

$$NEP = \frac{v_n}{R} = \frac{v_n}{V_{amp, DC}} \frac{1}{P_{in, RF}}$$

The NEP and responsivity definitions can be generalized to any power-detecting system, for instance, a power detector preceded by pre-amplification gain stage. Besides (1), there are also other ways to calculate NETD, as reported in [5], [6] and shown in (3) and (4). Note that the gain fluctuation term is not included in (3) and (4) for simplicity.

$$NETD = \frac{T^2}{2\tau} + \left(\frac{NEP_{det}}{k_BGB}\right)^2 \frac{1}{2\tau}$$

$$NETD = \frac{NEP_{sys}}{k_BB/\sqrt{2\tau}}$$

where $NEP_{det}$ and $NEP_{sys}$ are the noise-equivalent power of the detector (detector NEP) and imaging RX (system NEP), respectively, $G$ is the total gain preceding the power detector, and $k_B$ is the Boltzmann constant. Other parameters carry the same meaning as in (1).

For a stand-alone detector acting as a simple imaging RX without any pre-amplification, the NEP rather than NF is the proper measure for noise performance, since the square-law detector is essentially a non-linear circuit. Therefore, (4) should be used to calculate NETD. Note that in this case, $NEP_{sys}$ equals $NEP_{det}$. For a direct detection imaging RX consisting of a pre-amplification gain stage (e.g., LNA) and a power detector [5]–[7], both the noise temperature (or NF) of the pre-amplification stage and noise from the detector (measured by $NEP_{det}$) contribute to overall system noise. Therefore, the system NETD is obtained either from system NEP, $NEP_{sys}$, using (4) or by superposing the noise contribution from the pre-amplification stage and the detector using (3). For a frequency conversion type imaging RX like this work, where the detector noise is suppressed by high front-end gain, the first term under the square-root of (3) (representing noise from the front-end) dominates and thus (3) can be simplified to (1). Note that a factor of 2 due to Dicke switch operation needs to be added to all NETD calculations [9].

III. System Architectures

The NETD expression above indicates that the imaging receiver mandates low noise, high pre-detection gain, and wide bandwidth. Given the fact that the operation frequency is half of $f_{\text{max}}$ for this design, in order to meet the stringent design requirements of the imaging receiver, the direct conversion architecture [8], [11] is adopted with the LO frequency being placed in the middle of the RF band. Because the input signal is, in fact, broadband noise that contains no phase information, no I/Q path is needed to fulfill downconversion. The design requirement for the IF amplifier is also relaxed since the IF bandwidth is reduced to one half of the RF bandwidth. Moreover, in direct conversion architecture, the detector operates at IF frequency instead of MMW frequency, which leads to lower detector NEP due to higher responsivity and lower output noise. And the lower detector NEP would in turn reduce the required pre-detection gain for the system to be limited by the front-end noise rather than the detector noise.

The $2 \times 2$ FPA imager architecture is presented in Fig. 2 [12]. Each RX employs a local frequency tripler whose input is fed by a central on-chip 32 GHz PLL shared among the four RXs with the PLL’s output placed at the center of the chip. This layout floorplan minimizes the routing distance of the LO signal to all constituent RXs and also enables fully symmetrical LO distribution network. An advantage of the proposed scheme is that the LO signal can be generated/distributed at one-third of the 96 GHz operation frequency. More detailed discussions on the LO path design will be provided in Section V. Each RX (cf. Fig. 2) consists of a slot folded dipole antenna (SFDA), an SPDT switch, a four-stage LNA, a single-balanced mixer, an injection-locked frequency tripler (ILFT), an IF VGA, a power detector, an active bandpass filter and a synchronous demodulator. A 1 MHz off-chip clock controls the SPDT switch to toggle between the reference resistor and the on-chip antenna, and then the modulated MMW signal is amplified by the LNA and down-converted to IF band (0.1 GHz–10 GHz). The gain of the IF stage can be adjusted for different input power levels to increase the RX’s dynamic range. The energy detector detects the average power level during each half period of the modulation clock. The output waveform from the detector is a 1 MHz envelop signal which is then amplified by the active bandpass

Fig. 1. Block diagram of a Dicke radiometer.
IV. READER DESIGN AND IMPLEMENTATION

A. Slot Folded Dipole Antenna (SFDA)

The small antenna form factor at MMW frequency makes it possible for on-chip antenna integration [13]–[16]. The antenna-on-chip solution turns out to be an attractive option for a multi-pixel FPA imager [17], since it eliminates the complicated, lossy, and expensive MMW packaging. Despite the benefits of integrating antenna on-chip, the combination of high dielectric constant and low resistivity (8 \( \varepsilon_{\text{silicon}} \)) of silicon substrate poses a major challenge to on-chip antenna design. The bulk silicon substrate absorbs most of the radiation energy into the undesired surface-wave mode due to its high permittivity [18], [19], and the substrate’s low resistivity leads to electric field loss which is found to be the dominant loss factor [20]. Several methods [14]–[16] have been proposed to solve this problem, however, these solutions either require complex post-fabrication processes [14], [16] or use additional off-chip components [15]; which make the packaging even more challenging, reduce the yield, and increase the overall cost.

Fig. 3(a) depicts the 3D-view of the SFDA with CPW feed line used in this design. The folded-slot structure is favored since it provides wide bandwidth [21], [22], CPW friendly interface, low metallic loss, and high isolation. The advantages of using an SFDA can be better understood by comparing with its complementary counterpart, i.e., a folded dipole antenna (FDA), shown in Fig. 3(b). The FDA is a differential-type antenna that requires balanced excitation from a differential feed line (cf. Fig. 3(b)). Although this differential-type antenna makes it possible to realize a fully differential RX, the front-end circuits (SPDT switch and LNA) use single-ended design for the purpose of power/area saving and easy characterization. An additional balun, for this reason, is needed at the FDA interface.
which degrades both gain and NF. By interchanging conductive and dielectric material, an FDA is converted to an SFDA, which is excited by a pair of balanced slots instead of balanced metal lines in an FDA. When a signal travels along a single-ended CPW line, differential electromagnetic field is generated in the two slots between the signal line and the two ground sidewalls. These two slots can thus provide balanced excitation to the SFDA antenna (cf. Fig. 3(a)). In other words, the CPW feed line provides single-ended interface (through the metal trace) to the front-end circuits, and from the other side, provides differential interface (through the two slots) to the SFDA. As shown in Fig. 3(a), the ground walls surrounding the slot consist of all metal layers shorted by array of vias in parallel. This configuration leads to several advantages. First, it helps to meet metal density rules for all layers in the vicinity of the antenna. Second, the conductive loss from metals becomes significantly lower compared to FDA, since the current flows through a much wider path. Third, the ground wall confines the electromagnetic field and prevents mutual coupling between antenna and front-end circuits. In contrast, the radiation pattern of FDA is susceptible to interference from surrounding metal lines or other circuit components.

To further improve the antenna efficiency, a patterned deep trench mesh is embedded in the substrate underneath the antenna, as shown in Fig. 4(a). In this way, the conductive bulk silicon substrate is decomposed into an array of isolated small squares near its surface. The deep trench available from the technology is \(7 \mu m\)–\(10 \mu m\) deep from the substrate’s surface and is made of high resistivity material. The main advantage offered by the deep trench mesh is that it effectively increases the substrate resistance and thus reduces the substrate’s electric field loss, which is the key factor for antenna efficiency. Note that although the depth of the deep trench is much smaller than that of the substrate (280 \( \mu m \)), it is still effective in improving antenna efficiency. This is because the electric field losses are more
pronounced near the substrate’s surface. To examine the effect of deep trench lattice on antenna performance and validate the above analysis, two SFDAs test structures have been fabricated and tested. The two SFDAs have exactly the same design except that one employs deep trench lattice underneath. The measurement results in Fig. 4(b) demonstrate an increase in antenna gain by an average value of 2 dB as a result of using deep trench lattice. The simulated antenna gain and directivity are also shown in Fig. 4(b) which agree fairly well with measurements. In simulation, the SFDA exhibits an efficiency of 16% with deep trench. The deep trench lattice also reduces coupling between antenna and other circuits through silicon substrate. Fig. 5 shows the simulated radiation pattern of the SFDA.

B. SPDT Switch

The SPDT switch, shown in Fig. 6(a), is based on a λ/4 T-line approach similar to the one in [23]. By turning on and off the shunt NMOS transistor connected to port 1 (or 2), the corresponding branch shifts between isolated and thru state. Since the insertion loss of the switch is directly added to the noise figure, it needs to be minimized for the sake of improving system NETD. In order to reduce insertion loss, we need to maximize the off-state impedance of the transistor. This impedance can be modeled using a parallel R-C network [23], where the capacitance is tuned out by a shorted T-line stub acting as an inductor and the off-state resistance is highly dependent on the substrate resistance which can be increased by two layout techniques: (1) surrounding the NMOS device by deep trench; (2) reducing the number of substrate contacts and placing the contacts several micrometers away from the transistor. Shown in the Fig. 6(b) is the CPW line used in the SPDT switch design with its geometry labeled on the figure. The characteristic impedance is 50 Ω, and the signal line combines M5&M6 in parallel to reduce loss. The simulated loss at 95 GHz is 0.8 dB/mm. The CPW line uses a ground plane (M1) underneath and two ground side-walls to isolate the signal line from the environment.

C. LNA and Mixer

Fig. 7(a) shows the schematic of the four-stage LNA employing cascode topology for high gain and high reverse isolation. In order to reduce footprint, lumped inductors (rather than bulky T-lines) are used for output as well as interstage matching, while the input matching is realized using T-line to avoid layout discontinuity at the interface between the T-line-based SPDT switch and the LNA. The first LNA stage is biased at minimum NF current density of 4 mAM/μm², whereas the last three stages are biased for highest available gain at 95 GHz corresponding to a current density of 8 mAM/μm². The input matching and the size of the first-stage transistor are optimized for noise performance with satisfying impedance matching. Fig. 7(b) plots the simulated optimum source impedance (Zs) and input impedance (Zi) traces from 80 to 110 GHz on the Smith Chart, which indicates that an excellent noise matching is achieved. Capacitance of the GSG pad (20 fF) is absorbed into the input matching network of the stand-alone LNA fabricated separately. Interstage conjugate matching is performed by a split T-section inductive load and a series 50 fF capacitor to maximize power gain. Fig. 7(c) shows 3-D drawing of the split T-section inductive load (inside the dashed ellipse of Fig. 7(a)) comprised of two spiral inductors and one microstrip line inductor, where the use of microstrip line segment is mandated by layout arrangement to provide enough distance between adjacent stages. Fig. 7(d) shows one of the LNA stages including the biasing circuit for common-base transistor and decoupling circuit for the power supply. The bias voltage for transistor Q2 is generated by a resistor voltage divider from Vdd. The decoupling network for the common-base transistor consists of C1, C2 and R1, where C2 (500 fF) is the main decoupling metal-insulator-metal (MIM) capacitor and it is put as close as possible to the base of Q2 in the layout to minimize the parasitic inductance and ensure stability. Resistor R1 (20 Ω) further improves stability of the circuit. C1 is made of multiple capacitors (both MIM cap and MOS cap) in parallel to filter out the noise from the bias circuit. The local decoupling network for the power supply consists of C3 (500 fF), C4 (1 pF) and R2 (10 Ω). The self-resonant frequency of the MIM capacitors used in the design is simulated.
carefully with the EM tool. For example, the capacitance value for $C_3$ from the design kit is 240 fF, whereas its actual value at 96 GHz, taking into account the parasitic inductance, is 500 fF. The self-resonant frequency of $C_3$ is 130 GHz.

A single-balanced topology is employed for the downconverting mixer due to its simplicity and better noise performance [8]. The schematic of the single-balanced mixer is shown in Fig. 8. Since the IF frequency band (100 MHz–10 GHz) is far from the LO frequency (96 GHz), LO feedthrough is greatly suppressed, to the first order, at the mixer’s output and further LO rejection is provided by the two-stage IF amplifier. A 45 pH inductor realized by a CPW line is inserted between the collector of $G_{sw}$ stage and the emitters of the LO switching pair to increase conversion gain of the mixer over a wide bandwidth.

Fig. 7. (a) Schematic of the four-stage LNA, (b) Simulated input noise and impedance matching of the LNA from 80 to 110 GHz, (c) LNA interstage matching network, (d) Schematic of one LNA stage w/ bias and decoupling network.
is the center frequency of the oscillator, \( f_0 \)) with the same transistor size and biasing \( \) operating in triode region and exceeds a certain value (around 550 mV), \( \) and are combined in current-doubling configuration, \( \) represents the equivalent input resistance, \( \) in push-push configuration so as to cancel second-order harmonic of \( \) the first-order harmonic. The DC components generated by the active bandpass filter out the undesired harmonics generated by the HBT transistors. As the control voltage \( V_{c1} \) increases from 0 V, these two transistors migrate from deep triode region to saturation region, and along the way, the tuning is achieved. However, the gain flatness becomes poor when \( V_{c1} \) exceeds a certain value (around 550 mV), because transistors \( M_{y1}, M_{y2} \) do not act as linear “resistors” outside the deep triode region. In order to obtain \( > 12 \) dB tuning range and maintain flat frequency response within the 10 GHz bandwidth, another voltage controlled resistor – realized by transistor \( M_{c1} \) biased in deep triode region – is used for emitter degeneration. The input of the amplifier is AC-coupled with a corner frequency of 100 MHz to remove the 1/f noise and DC offset voltage from the mixer. Fig. 9(b) plots the simulated frequency response of the two-stage variable gain amplifier, which achieves a tuning range of 8–22 dB and the gain degradation of less than 0.5 dB within the 10 GHz bandwidth.

Fig. 10 shows the schematic of the power detector following the IF amplifier. The detector core circuit comprises the transistor pair \( Q_{1} \)–\( Q_{2} \) in push-pull configuration so as to cancel the first-order harmonic. The DC components generated by the second-order harmonic of \( Q_{1} \)–\( Q_{2} \) are combined in current domain and converted to voltage (\( V_{\text{out}} \)) by the load resistor. A replica stage (\( Q_{3} \)–\( Q_{4} \)) with the same transistor size and biasing is also employed to provide the reference DC output voltage, so that the detector exhibits zero output voltage when no input signal is present. The load resistor together with the input capacitance of next stage forms a first-order 10 MHz lowpass filter to filter out the undesired harmonics generated by the HBT transistors.

### E. Baseband Circuitry

The baseband circuitry consists of an active bandpass filter and a chopper acting as a demodulator. The active bandpass filter has a bandwidth of 0.1 MHz–10 MHz and in-band maximum gain of 25 dB, which captures up to ninth harmonics of the Dicke modulation frequency (1 MHz) and rejects the out-of-band noise and harmonics from the detector output. The chopper is based on a passive mixer structure [8]. The four transistors are biased at zero voltage, conducting no DC current, and therefore, exhibit negligible 1/f noise.

### V. LO Generation and Distribution

#### A. Architecture

As described in Section III, the LO path utilizes a common 32 GHz PLL and four local ILFTs instead of a fundamental PLL running at 96 GHz. This method readily lowers the operation frequency of the LO distribution network and reduces the power consumption compared to a 90 GHz LO generation/distribution network [25]. As for the frequency tripler design, the injection-locking topology is preferred over a harmonic-based structure since it can achieve higher power efficiency [26]. Furthermore, this method offers better phase noise than directly designing a 96 GHz PLL. In order to compare the two methods in terms of phase noise, we first start with a 32 GHz PLL circuit. The first method is to scale up the PLL frequency by a factor of three, while the second method involves cascading a frequency tripler after the 32 GHz PLL. For the first method, the phase noise of a general oscillator can be expressed by Leeson’s model [27], [28], i.e.,

\[
S_{\Delta \phi, \text{out}} = 2 \frac{|V_n|^2}{V_o^2} \left( \frac{\omega_o}{2Q\Delta \omega} \right)^2 \Delta \omega \ll \omega_o \tag{5}
\]

where \( S_{\Delta \phi, \text{out}} \) denotes the output phase noise power spectral density, \( \omega_o \) is the center frequency of the oscillator, \( Q \) is the quality factor of the tank, \( \Delta \omega \) is offset frequency, \( V_o \) is the oscillation amplitude, \( V_n \) represents the equivalent input referred noise voltage. Equation (5) states that for ideal frequency tripling of an oscillator, the phase noise will degrade by 9.54 dB (\( 20\log_{10}3 \)). However, this ideal number is based on the assumption that other parameters such as Q factor, \( V_o \), and \( V_n \) do not change with frequency, which is not the case, in practice. As the operation frequency increases, especially when it goes into W-band range, the \( V_o \) and \( Q \) (usually limited by varactor’s Q) tends to drop, while the intrinsic device noise parameter \( |V_n| \) increases dramatically [29]. As a result, the phase noise degradation due to scaling up a VCO operation frequency from Ka-band to W-band would be much more than the ideal 9.54 dB. For the same reasons, it is extremely difficult to achieve good phase noise for a fundamental PLL running above 90 GHz (half of \( f_{\text{max}} \) for the BiCMOS technology in use), as corroborated in [25] and [30]. As for the second method, the phase noise at the output of an injection-locked frequency tripler can be expressed by (6) [31]:

\[
S_{\text{out, ILFT}}(\Delta \omega) = \frac{3^2}{1 + \left( \frac{\Delta \omega}{\omega_o} \right)^2} S_{\text{inj}}(\Delta \omega) \\
+ \frac{1}{1 + \left( \frac{\Delta \omega}{\omega_o} \right)^2} S_{\text{harm}}(\Delta \omega) + \frac{\left( \frac{\Delta \omega}{\omega_o} \right)^2}{1 + \left( \frac{\Delta \omega}{\omega_o} \right)^2} S_{\text{VCO}}(\Delta \omega) \tag{6}
\]
where \( S_{\text{out,ILFT}}, S_{m_1}, S_{\text{harm}}, \) and \( S_{VCO} \) are the phase noise spectral density of the ILFT’s output, injected signal, harmonic generator, and free-running VCO, respectively. Equation (6) implies that \( S_{VCO} \) has a highpass nature, while both \( S_{m_1} \) and \( S_{\text{harm}} \) have a lowpass nature, and the noise contribution of \( S_{\text{harm}} \) is nine times more than \( S_{m_1} \). Therefore, the first term dominates the output phase noise for small offset frequency compared to the corner frequency, given as 

\[
\omega_p \approx \left( \frac{I_3}{I_0} \right) \left( \frac{\omega_0}{2Q} \right) \quad [31],
\]

where \( I_0 \) and \( I_3 \) represent oscillation current and the amplitude of third-harmonic injection current, respectively. In this design, the tripler’s output phase noise is determined by the injected signal (i.e., Ka-band PLL output) for frequency offset up to around 100 MHz. The above analysis indicates that the phase noise degradation after the ILFT can be very close to theoretical minimum value, which is also verified by several reported works [31]–[33], as well as the measurement results presented in Section VI.

**B. Ka-Band PLL**

The Ka-band PLL is comprised of a differential Colpitts VCO with emitter degeneration, a frequency divider chain with the division ratio of 256, a phase/frequency detector (PFD) [34], a charge pump (CP), and a 3rd-order loop filter (cf. Fig. 2). The divider chain utilizes ECL circuits for asynchronous divide-by-32 and TSPC circuits for divide-by-8. The VCO employs 3-bit digital band selection in addition to analog varactor control. Programmable PFD delay, CP current, and loop BW compensate for model inaccuracy and PVT variation. The PLL consumes 65 mW from 1.8 V/2.5 V supply.

**C. LO Distribution Network**

The 32 GHz LO generated by the Ka-band PLL is distributed to local tripler within each RX through a symmetric CPW T-line H-tree network without using any bulky 32 GHz power splitters, as shown in Fig. 11. Since the distance \( L \) from the output of the 1st stage buffer (buf1) to the input of the 2nd stage buffers (buf2 & 3) is 220 \( \mu \)m which roughly equals to \( \lambda_0/20 \) at 32 GHz, this interconnect can still be treated as a lumped element which is part of the tuned load of the first buffer stage whose schematic is also shown in Fig. 11. Buf2 & 3 are both output matched to 50 \( \Omega \) differentially for maximum power delivery to the T-line network. From simulation, the LO power level at the output of buf2 (or 3) and input of the ILFT are 0 dBm and 2 dBm, respectively, ensuring a locking range of > 5.5 GHz for the ILFT. The entire LO distribution network is designed to be fully differential and consumes 90 mW.

**D. ILFT**

The schematic of W-band ILFT circuit is shown in Fig. 12, which consists of two parts: (1) a pair of harmonic generating transistors \( Q_3 \& Q_4 \); and (2) an injection-locked oscillator (ILO). The ILO is based on differential Colpitts oscillator with a free-running frequency close to three times of the input fundamental frequency, and therefore, exhibits a loop gain greater than unity for the 3rd harmonic component only. Injection-locking operation is realized by feeding the 3rd harmonic of the input signal generated by \( Q_3 \& Q_4 \) into the emitters of ILO. Transistors \( Q_3 \& Q_4 \) reuse part of the DC current of the tank and are biased in Class-AB regime for maximum 3rd harmonic generation [26]. The 96 GHz output signals are taken out from the collectors of \( Q_1 \& Q_2 \) through a cascode buffer stage to minimize leakage of the 32 GHz injection signal at the output. Differential operation is achieved by connecting...
two inter-digitated metal-oxide-metal capacitors (C3 & C4) back-to-back across the emitters of Q1 & Q2. To attain compact layout yet avoid mutual electromagnetic and substrate noise coupling, ground-shielded CPW (GCPW) lines (T1 ~ T6) are used at base, collector, and emitter to provide the tank, the load, and part of the degeneration inductances. Additional emitter degeneration inductance is realized by spiral inductor (L1 & L2) to save area. The GCPW structure is favored, because it can realize small inductance with good modeling accuracy and adequate quality factor.

One observation from simulation is that the loss introduced by the buffer to the collector nodes of Q1 & Q2 has a noticeable impact on the tank’s Q factor due to the coupling via Cμ (base-to-collector capacitance) of these transistors. The same
phenomenon was also observed in [35] where the authors chose source/emitter as the output node, instead. As a qualitative explanation, in the presence of $C_{be}$, the resistive part of the buffer’s input impedance, $R_{buf}$ (mainly due to the base resistance of the buffer’s input transistor) adds loss to the tank which may quench the oscillation. For quantitative analysis, the input impedance seen into the base of a Colpitts oscillator is derived based on a general circuit model shown in Fig. 13. Note that the base current has been ignored for this derivation.

$$Z_{in} = \frac{(1 + sC_{be}Z_c) \cdot \{1 + g_mZ_c + sC_{be}Z_e\}}{sC_{be}1 + sC_{be}Z_c + sC_{be}1 + g_mZ_c + g_mZ_e + sC_{be}Z_e}. \quad (7)$$

If the effect of $C_{be}$ is ignored, the above equation is simplified to:

$$Z_{in} = \frac{1}{sC_{be}} + Z_c + \frac{g_mZ_e}{sC_{be}} \quad (8)$$

as shown in [36] for the input impedance of a source/emitter degenerated transistor. But this simplification gives non-negligible errors. In the ILFT design (cf. Fig. 12), $C_{be}$ stands for $C_{1}$; $C_{be}$ equals the sum of $C_1$ and $C_{1e}$ (base-to-emitter capacitance of $Q_1$ & $Q_2$); $Z_c$ and $Z_e$ are the emitter degeneration impedance and the load impedance, respectively, expressed by:

$$Z_c = \frac{1}{s(C_3 + C_{par})(1 + sL_c/C_3 + C_{par})}; \quad (9)$$

$$Z_e = \frac{\frac{sR_{buf}L_c}{R_{buf} + sL_e + s^2R_{buf}L_c(C_{c} + C_{buf})}}{L_{buf}}. \quad (10)$$

Where $L_e$ is the total emitter degeneration inductance from $L_3$ and $L_5$, $C_{par}$ includes all the parasitic capacitances associated with emitter node contributed by $Q_1$ and $Q_2$, $C_{ce}$ is the collector-to-substrate capacitance of $Q_1$, $L_c$ is the load inductance from $T_3$, $R_{buf}$ and $C_{buf}$ are the equivalent shunt resistance and capacitance of the buffer. In order to avoid the loss introduced by the buffer, capacitive degeneration is employed in the buffer design to reduce the real part of the buffer’s input impedance. The schematic of the buffer is shown in Fig. 14(a). By choosing the resonant frequency of $L_{DEG}C_{DEG}$ lower than the operation frequency (i.e., $1/2\pi \sqrt{L_{DEG}C_{DEG}} < 96$ GHz), the emitter degeneration impedance ($L_{DEG}$ in parallel with $C_{DEG}$) becomes capacitive at 96 GHz, and the real part of the buffer’s input impedance is written as:

$$R_t(Z_{in,buf}) = R_bQ_{be} - \frac{g_mQ_{be}}{\omega^2C_{eff}C_{ne}} \quad (11)$$

where $R_bQ_{be}$ is the base resistance of $Q_b$ (cf. Fig. 14(a)), $C_{eff}$ is the effective emitter degeneration capacitance.

$$C_{eff} = C_{DEG} \left(1 - \frac{1}{\omega^2L_{DEG}C_{DEG}}\right). \quad (12)$$

Fig. 14(b) plots the simulated $\text{Re}(Z_{in,buf})$ with and without capacitive emitter degeneration. The combination of negative input resistance from the buffer and avoidance of using low Q varactors improves tank’s Q factor of the ILFT, and therefore, decreases the required DC current to sustain the oscillation. The ILFT and output buffer consume 50 mW and 25 mW from a 2.5 V supply, respectively. To the authors’ best knowledge, this work presents the first implementation of an injection-locked-based frequency multiplier in SiGe BiCMOS process.

VI. EXPERIMENTAL RESULTS

A. Chip Fabrication

The FPA imaging chip has been fabricated in a 0.18 $\mu$m SiGe BiCMOS process (TowerJazz SBC18H2) with six metal layers [7]. The HBT device has a minimum width of 0.15 $\mu$m with $f_T = 200$ GHz and $f_{max} = 180$ GHz. Fig. 15(a) and (b) show the die photo of the $2 \times 2$ FPA imager as well as a zoom-in view of a single pixel. The entire imaging chip occupies an area of $3 \times 3.5 \text{mm}^2$, including all the pads. A chip-on-board assembly is used to characterize the performance of the FPA imaging chip. Note that no MMW signals travel through the bond-wires thanks to the on-chip antenna integration. To facilitate characterization of the $2 \times 2$ imaging system, breakout circuits of on-chip antenna, SPDT switch, LNA, single receiver front-end (excluding antenna and baseband) and PLL with tripler have been fabricated and measured, separately. All the building blocks are tested through wafer probing. For the LO breakout circuit, all DC pads are bonded to a PCB board, while the RF output is wafer probed. The reference signal for the PLL is provided by a signal generator through wafer probing to measure the PLL tuning range. As for the phase noise measurement, the reference signal is provided by a crystal oscillator on board through wirebonding.

B. Measurement Results of Building Blocks

1) Antenna: The stand-alone antenna is characterized by placing a transmitting horn antenna with a gain of 24 dBi in the peak-gain direction of the antenna-under-test (AUT). The distance ($L$) between the horn and the AUT is set to 50 cm to ensure the satisfaction of far-field condition. Fig. 16 depicts the measured antenna impedance matching and gain over a wide
Fig. 14. (a) Schematic of the tripler’s buffer, (b) Simulated real part of the buffer’s input impedance w/ and w/o capacitive degeneration.

Fig. 15. (a) Die photo of the 2 × 2 FPA, (b) Die photo of a single pixel.

Fig. 16. Measured antenna gain (in peak-gain direction) and return loss.

frequency range, where the antenna gain is obtained using Friis transmission equation.

2) Receiver Front-End: Fig. 17(a) plots the measured insertion loss and input-to-output isolation of the SPDT switch. The switch achieves −2.8 dB minimum insertion loss and better than −20 dB isolation using 0.18 μm CMOS transistors. Fig. 17(b) shows the input and output matching of the switch under ‘through’ state. Occupying only 380 × 300 μm² (excluding pad), the LNA break-out circuit exhibits a measured peak gain of 25 dB, 3-dB bandwidth of 20 GHz (86–106 GHz), and a minimum NF of 8.3 dB, as shown in Fig. 18(a). Fig. 18(b) plots the simulated and measured LNA input and output return losses. The measured gain and noise figure of a single receiver front-end (from SPDT switch down to IF amplifier) is plotted in Fig. 19 versus IF frequency. The IF amplifier is set to high gain mode for this measurement. A maximum in-band gain of 44 dB and minimum DSB-NF of 11.3 dB are reported.
3) **LO**: The test chip for Ka-band PLL and ILFT has been fabricated in the same technology to characterize the performance of LO generation circuit (see [26]). The single-ended 96 GHz output spectrum measured at the output of the ILFT is shown in Fig. 20, which indicates a reference spur level of $-54$ dBc. The Ka-band PLL achieves a measured tuning range of 30.3–33.8 GHz and the tuning range measured after the ILFT is 92.8–98.1 GHz, which is limited by locking range of the ILFT. Fig. 21 shows the input sensitivity measurement of a stand-alone ILFT. The ILFT has a free-running frequency of 95.5 GHz, and exhibits an input sensitivity of $-14$ dBm. Phase noise performance was measured using an 125 MHz voltage controlled crystal oscillator as reference signal. Shown in Fig. 22, the phase noise at 1 MHz offset measured at the output of the PLL and ILFT are $-103$ dBc/Hz and $-93$ dBc/Hz, respectively.
C. System Measurements

The responsivity of the receiver is measured from the input port of the SPDT switch to the output of the power detector through on-wafer probing on the breakout circuit (front-end+detector). It is the same breakout circuit that was used to measure the front-end gain and NF above where the detector was disabled during the front-end measurement. The gain of the baseband amplifier is not included in the RX responsivity measurement, because the baseband op-amp following the power detector amplifies the signal level as well as the output noise at the same time. As a result, the baseband gain does not help to lower the system NEP, at all. During responsivity measurement, the W-band input signal is provided by a low frequency signal generator followed by a 6 active frequency multiplier and the input power is controlled using a waveguide variable attenuator. The input RF signal is modulated with a 1 MHz square-wave by controlling the front-end SPDT switch to be above the noise corner frequency. Shown in Fig. 23(a) is the simulated and measured responsivity across the RF frequency (without on-chip antenna), which indicates a peak responsivity of 395 MV/W and an average responsivity of 285 MV/W over the 20 GHz RF bandwidth. By tuning the gain of the IF amplifier, the average system responsivity can be lowered from 285 MV/W to 12 MV/W. The output noise is measured at the output of the detector using spectrum analyzer with 1 Hz resolution bandwidth, and the result is shown in Fig. 24. From the plot, the output noise profile exhibits a $1/f$ corner frequency of around 40 kHz after which it reaches a constant noise floor of 2.3 μV/Hz$^{1/2}$. Fig. 23(b) shows the simulated and measured system NEP with respect to RF frequency. The measured NEP is based on the measured responsivity and output noise at 1 MHz which is far above the gain fluctuation frequency and $1/f$ noise corner frequency of the SiGe process. The imaging receiver (without on-chip antenna) achieves a minimum and average NEP of 5.8 fW/Hz$^{1/2}$ and 8.1 fW/Hz$^{1/2}$, respectively. It is worth mentioning that the baseband circuits (including the active bandpass filter and chopper) after the detector do not affect or degrade the system noise performance. This is because the simulated baseband input-referred noise is 58 nV/Hz$^{1/2}$ which is negligible compared to the measured system output noise after the detector (2.3 μV/Hz$^{1/2}$).
Fig. 23. Measured and simulated system (a) responsivity and (b) NEP versus input frequency.

Fig. 24. Measured output noise versus frequency.

In Fig. 23(a), the simulated responsivity is slightly higher than measured results, because the measured LNA gain is about 1 dB lower than simulated value. The simulated output noise exhibits a 1/f corner frequency of 20 kHz and a constant noise floor of 2 $\mu$V/Hz$^{1/2}$ above 1/f corner, which is close to the measured value (40 kHz and 2.3 $\mu$V/Hz$^{1/2}$). The lower 1/f noise in simulation may be attributed to the inaccurate 1/f noise model of the transistor and resistor.

Fig. 25 indicates the measured system responsivity versus input RF power. This linearity measurement has been carried out at 95 GHz. From the plot, the input $P_{1 dB}$ of the receiver,
defined as the input power when the system responsivity drops by 1 dB, is $-57$ dBm which is good enough for passive imaging applications. The total system NETD is 0.48 K, which is calculated using (4) based on the measured average NEP with 30 ms integration time. A factor of 2 due to Dicke switching is taken into account in the NETD calculation. It is noteworthy that the radiometer bandwidth in the NETD formula is the pre-detection bandwidth, which is, the 10 GHz IF bandwidth instead of the 20 GHz RF bandwidth. This is because the advantage of DSB NF has already been accounted in the system noise temperature and cannot be exploited again in the NETD calculation [8], [37].

The imaging functionality is tested under transmission mode using the setup shown in Fig. 26, where an incoherent source is employed to illuminate the object. The integrated on-chip Dicke switch eliminates the need for lock-in amplifier and mechanical chopping system used in previously reported silicon-based imaging tests [17], [38]. The reason of using an active imaging setup is because the loss introduced by the on-chip antenna degrades the system NETD. For an imaging RX without antenna, the system responsivity is defined as $V_{out}/P_{r}$, where $P_{r}$ is the power delivered to the imaging receiver, as shown in Fig. 27. While with the on-chip antenna, the system responsivity equals $V_{out}/P_{pixel}$, where $P_{pixel}$ is the power collected by the receiving antenna. According to the definition of the antenna efficiency ($\varepsilon_r$), $P_{r} = P_{pixel} \times \varepsilon_r$. Therefore, the system responsivity with antenna is lowered by 6.25 times ($1/\varepsilon_r$). As a result, both the system NEP and the system NETD increases by 6.25 times with on-chip antenna. In conclusion, the system NETD is 3 K with on-chip antenna compared to 0.48 K without on-chip antenna. Therefore, with the use of on-chip antenna, this chip cannot operate in radiometer mode due to the low antenna efficiency (i.e., 16%) and an off-chip source is needed to increase the SNR. The low efficiency of the on-chip antenna is due to the inherent low substrate resistivity of the silicon technology. The antenna efficiency can be improved to as much as 50% in CMOS SOI technology [39]. The input SNR of the imaging setup can be calculated using the following equation [40]:

$$SNR = \frac{P_{pixel}}{NEP \cdot \sqrt{B_{LP}}} = \frac{P_{pixel}}{NEP \cdot \sqrt{\frac{1}{2T}}}$$ (13)

where $P_{pixel}$ is the power per pixel collected by the antenna, which can be expressed by [39]:

$$P_{pixel} = \frac{P_t \cdot G_t}{4\pi R^2} \cdot A_{pixel}$$ (14)

$$A_{pixel} = D_e \frac{\lambda^2}{4\pi}$$ (15)

where $P_t$ is the transmitted power (0.1 mW), $G_t$ is the gain of transmitting antenna (24 dBi), $D_e$ is the directivity of receiving antenna, $A_{pixel}$ is the effective antenna area and $R$ is the distance between transmit and receive antenna (50 cm). $A_{pixel}$ is calculated to be 2.55 mm$^2$ based on the simulated antenna directivity (cf. Fig. 4(b)) and operation frequency (90 GHz). Due to the lack of free-space power meter for direct measurement, $P_{pixel}$ is estimated using (14). The resulting SNR given by (13) is 50 dB. For this measurement, we use a relative large power (0.1 mW) for the illumination source to increase the SNR and lower the gain of the IF stages to prevent saturating the receiver.
The W-band signal (90 GHz) is radiated by a waveguide horn antenna and the radiated power is controlled using a waveguide-based variable attenuator before the horn antenna. Two lenses made of high density polyethylene (HDPE) are employed to collimate and focus the radiation onto the object which is placed right at the focal point. Another lens system is used to collimate and refocus the beam onto the FPA chip. The diameter and focal length of the lens are 76.2 mm and 75 mm, respectively. The MMW images of two metallic objects hidden inside an envelope is constructed by mechanically scanning the object with 5 mm step size. The output voltage of each pixel is read from the oscilloscope and mapped to corresponding grayscale intensity.

As discussed above, the system NETD is limited by the low efficiency of the on-chip antenna. One way to improve the system’s NETD is to increase the integration time at the expense of reducing frame rate. In other words, we can trade off frame-rate for antenna efficiency at a constant sensitivity. Shown in Fig. 28 is the passive image of a coin taken with 270 ms integration time, which corresponds to an NETD of 1 K. This measurement is done using the similar setup as shown in Fig. 26 except that the W-band source and the first two lenses are removed. The measured performance of the imaging receiver is summarized in Table I. Table II compares this work with other state-of-the-art silicon-based W-band radiometers.

VII. CONCLUSIONS

A 94 GHz 2 × 2 focal-plane array featuring on-chip antenna and LO generation/distribution has been designed and implemented in a 0.18 µm SiGe BiCMOS process. The on-chip integration of antenna and Dicke switch avoid any high frequency transitions from or to the chip. Measurements of the fabricated prototype exhibit excellent results. The imaging receiver (without on-chip antenna) achieves a measured average responsivity and noise equivalent power of 285 MV/W and 8.1 fW/Hz1/2, respectively. The minimum temperature sensitivity of the Dicke radiometer is 0.48 K (without antenna) which is comparable to multi-chip imaging modules made in III-V technologies. Taking into account losses of the on-chip antenna, the average NEP and system NETD increase to 50.6 fW/Hz1/2 and 3 K, respectively. Overall, a high level of integration is achieved with small form factor. For the first time in silicon technology, a PLL has been integrated with a receiver at 90 GHz. This work demonstrates the feasibility of using silicon technology for multi-pixel PMMW imaging and opens up possibilities for future deployment of low-cost wafer-scale imagers.

ACKNOWLEDGMENT

The authors thank TowerJazz Semiconductor for chip fabrication. They thank Dr. V. Jain of SaberTek for helpful discussions and suggestions. They are indebted to the anonymous reviewers for their diligent comments and suggestions. Technical support from Sonnet Software is highly appreciated.

REFERENCES

Chun-Cheng Wang (S’07) received the B.S. and M.Eng. degrees in electrical and computer engineering from Cornell University, Ithaca, NY in 2003 and 2004 respectively, and the Ph.D. degree in electrical engineering from University of California, Irvine in 2012. From 2004 to 2007 he was with Realtek Semiconductor Corp. as an Analog/RF Design Engineer, where he was involved in the design of 802.11 a/b/g/n CMOS radio. In 2012, he joined Peregrine Semiconductor, San Diego, CA as a Senior Engineer. His research interests are in RF and millimeter-wave (mm-wave) integrated circuit design for wireless communications, automotive radars, and imaging applications. During his Ph.D., he interned at Fujitsu Laboratories of America, Sunnyvale, CA and Broadcom Corp., Irvine, CA, both involved in the design of 60 GHz CMOS radio.

Dr. Wang was the recipient of the Mindspeed Fellowship in 2011, the Center of Pervasive Communications and Computing (CPCC) Fellowship Award in 2010, the School of Engineering Research and Travel Grant Award in 2009 and the EECS Department Fellowship in 2007 at University of California, Irvine. He is a member of Golden Key Honor Society andEta Kappa Nu.

Hsin-Cheng Yao received the B.S. degree in electrical engineering in 2005 from University of Illinois at Urbana-Champaign and the M.S. degree in electrical engineering in 2007 from the University of Southern California, Los Angeles. From 2007 to 2010, he was a research assistant in the NCIC Lab at University of California, Irvine, where he was involved in the design of RF and millimeter-wave integrated circuit design for wireless communications and imaging applications. In 2012, he joined HTC Corporation, Taiwan, as an Engineer.

Payam Heydari (S’98–M’00–SM’07) received the B.S. and M.S. degrees (with highest honors) in electrical engineering from the Sharif University of Technology in 1992 and 1995, respectively. He received the Ph.D. degree in electrical engineering from the University of Southern California in 2001.

In August 2001, he joined the University of California, Irvine, where he is currently a Professor of Electrical Engineering. His research interests include the design of high-speed analog, radio-frequency (RF), and mixed-signal integrated circuits. He is the (co)-author of one book and over 90 journal and conference papers.

The Office of Technology Alliances at UCI has named Dr. Heydari one of 10 outstanding innovators at the university. Dr. Heydari is the co-recipient of the 2009 Business Plan Competition First Place Prize Award and Best Concept Paper Award both from Paul Merage School of Business at UC-Irvine. He is the recipient of the 2010 Faculty of the Year Award from UC-Irvine’s Engineering Student Council (ECS), the 2009 School of Engineering Fariborz Maseeh Best Faculty Research Award, the 2007 IEEE Circuits and Systems Society Guillemin-Cauer Award, the 2005 NSF CAREER Award, the 2005 IEEE Circuits and Systems Society Darlington Award, the 2005 UCI’s School of Engineering Teaching Excellence Award, the Best Paper Award at the 2000 IEEE International Conference on Computer Design (ICCD), the 2000 Honorable Award from the Department of EE-Systems at the University of Southern California, and the 2001 Technical Excellence Award in the area of Electrical Engineering from the Association of Professors and Scholars of Iranian Heritage (APSIH). He was recognized as the 2004 Outstanding Faculty of Electrical Engineering from the Association of Professors and Scholars of Iranian Heritage (APSIH).

Dr. Heydari is a Guest Editor of IEEE JOURNAL OF SOLID-STATE CIRCUITS.

Zhiming Chen (S’08) received the B.Eng. degree in electronic engineering from Tsinghua University, Beijing, China, in 2007, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Irvine, in 2009 and 2012, respectively.

He was a summer intern at Finisar Corporation, Sunnyvale, CA, in 2008. He was an intern at Broadcom Corporation from June 2010 to November 2011, where he was engaged in 60 GHz phased-array radio development. In Spring 2012, he joined the Faculty of Beijing Institute of Technology, China. His research interests include radio frequency (RF) and millimeter-wave (MMW) integrated circuits design.

His research interests are in RF and millimeter-wave (mm-wave) integrated circuit design for wireless communications, automotive radars, and imaging applications. During his Ph.D., he interned at Fujitsu Laboratories of America, Sunnyvale, CA and Broadcom Corp., Irvine, CA, both involved in the design of 60 GHz CMOS radio.

Dr. Wang was the recipient of the Mindspeed Fellowship in 2011, the Center of Pervasive Communications and Computing (CPCC) Fellowship Award in 2010, the School of Engineering Research and Travel Grant Award in 2009 and the EECS Department Fellowship in 2007 at University of California, Irvine. He is a member of Golden Key Honor Society and Eta Kappa Nu.