We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,200 Open access books available
116,000 International authors and editors
125M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected. For more information visit www.intechopen.com
Chapter

Terahertz Sources, Detectors, and Transceivers in Silicon Technologies

Li Zhuang, Cao Rui, Tao Xiaohui, Jiang Lihui and Rong Dawei

Abstract

With active devices lingering on the brink of activity and every passive device and interconnection on chip acting as potential radiator, a paradigm shift from “top-down” to “bottom-up” approach in silicon terahertz (THz) circuit design is clearly evident as we witness orders-of-magnitude improvements of silicon THz circuits in terms of output power, phase noise, and sensitivity since their inception around 2010. That is, the once clear boundary between devices, circuits, and function blocks is getting blurrier as we push the devices toward their limits. And when all else fails to meet the system requirements, which is often the case, a logical step forward is to scale these THz circuits to arrays. This makes a lot of sense in the terahertz region considering the relatively efficient on-chip THz antennas and the reduced size of arrays with half-wavelength pitch. This chapter begins with the derivation of conditions for maximizing power gain of active devices. Discussions of circuit topologies for THz sources, detectors, and transceivers with emphasis on their efficacy and scalability ensue, and this chapter concludes with a brief survey of interface options for channeling THz energy out of the chip.

Keywords: terahertz, detector, source, transceiver, silicon

1. Introduction

The significance of terahertz electronics is self-evident for readers of this book. The general consensus among silicon THz circuit designers (!) is that silicon will be the dominant technology for the lower end of the THz spectrum (300 GHz to around 1 THz) in light of recent breakthroughs of silicon circuits in terms of effective isotropic radiated power (EIRP), phase noise, and receiver sensitivity. For many applications, silicon circuits are on par or even superior to III/V compound technologies and optical-based techniques in this frequency range now. This chapter aims to introduce the reader to the fascinating world of silicon THz circuit design through a step-by-step approach: We examine conditions for extracting the most power gain out of a given active device. Popular topologies for silicon sources, detectors, and transceivers are discussed next, and this chapter concludes with a brief survey of THz interface options for efficient energy transfer between circuits and the outside world.
2. How to have THz power and radiate it too

Due to the excessive loss and scarcity of power gain for silicon devices in the THz region, one should strive to extract the most power out of a given device during the whole design phase. This involves making sure that the device is working under the optimum condition (i.e., the device is embedded in the right impedance environment for maximum power gain), the topology of the circuit is optimum for the intended application, and the power is transferred from the circuit through the most efficient interface. This section gives an overview of these areas.

2.1 Power gain maximization for a given active device

The active devices in THz circuits are connected to the rest of the circuits through passive elements, such as capacitors, inductors, and transmission lines. The overall circuit performance is decided both by the active device and these passive elements. Thus, to maximize the circuit performance, a “divide-and-conquer” approach is the logical choice. That is, we first find the “best” active device in a given technology under certain constraints such as power consumption or noise performance. We then decide the best passive network into which the device should be embedded. The problem is there is no such thing as “pure” active device; passive elements are always present in a given active device. Mason [1] has thus defined a figure of merit for active devices:

\[
U = \frac{|Y_{21} - Y_{12}|^2}{4(G_{11}G_{22} - G_{12}G_{21})}
\]

(1)

\(G_{ij}\) is the real part of \(Y_{ij}\) in Eq. (1). The above FOM is called Mason’s invariant \(U\), since it is invariant to passive embedding environments that are linear, lossless, and reciprocal [1, 2].

A device is active if its \(U\) is larger than one, which means this device is capable of providing real power. The maximum oscillation frequency (\(f_{\text{max}}\)) of a device is defined as the frequency where its \(U\) equals one, i.e., beyond which frequency it is no longer active. The maximum power gain of this device embedded in the two-port also drops to unity at the maximum oscillation frequency (\(f_{\text{max}}\)). \(U\) is also the maximum power gain of the device after unilateralization, that is, when \(Y_{12}\) is made to zero. Generally speaking, higher \(U\) means higher power gain at a given frequency.

For a given two-port shown in Figure 1(b), the power gain is defined as

\[
P_L = \frac{P_L}{P_{IN}} = \frac{-0.5 \times \Re(V_1 \times I_2^*)}{0.5 \times \Re(V_1 \times I_1^*)} = \frac{\Re(V_2 \times V_2^* \times Y_l^*)}{\Re(V_1 \times I_1^* \times Y_{IN}^*)} = |A_V|^2 \frac{\Re(Y^*_l)}{\Re(Y_{IN}^*)}
\]

(2)

\(P_L\) and \(P_{IN}\) are the real power delivered to the load and to the two-port. \(V_1, I_1,\) and \(V_2, I_2\) in Figure 1 are the voltage and current at port 1 and port 2, respectively. \(A_V\) is the voltage gain of the two-port. \(Y_l\) and \(Y_{IN}\) represent the source and load admittance presented to the device. \(I_l\) and \(V_l\) form the Norton equivalent circuit of the signal source.

For an unconditionally stable two-port (it does not oscillate for any passive load and source admittance) at a given frequency, the power gain could be maximized by biconjugate matching at the input and output. Conjugate matching is achieved when the load admittance is equal to the conjugate of the source admittance at a given node; biconjugate matching means that this condition is satisfied both at the input and the output port. For a given two-port, its maximum power gain is
where $K$ is the stability factor defined as

$$
G_{\text{max}} = \frac{|Y_{21}|}{|Y_{12}|} \times \left( k - \sqrt{k^2 - 1} \right)
$$

Biconjugate matching is possible when $K$ is equal or greater than 1. For a given two-port, it is unconditionally stable when the following conditions are simultaneously satisfied:

$$
g_{11} \geq 0 \quad (5)$$

$$
g_{22} \geq 0 \quad (6)$$

$$
K \geq 1 \quad (7)
$$

Keep in mind that unlike $U$, which is invariant to the embedding network, $G_{\text{max}}$ is sensitive to its environment. We can modify the embedding environment to make $G_{\text{max}}$ larger. It can be shown that the maximum $G_{\text{max}}$ for a given device is $\sqrt{U + \sqrt{U - 1}}$. For a detailed discussion about $U$ and $G_{\text{max}}$, please refer to [3–6] which present ways of designing the embedding network for maximizing $G_{\text{max}}$ while maintaining stability under process variations. The basic idea is to utilize feedback to generate negative resistance, such as adding capacitive degeneration to CE (common emitter) amplifiers or adding inductance to the base node of CB (common base) amplifiers.

Another important THz circuit is oscillators. Here, we need to make a distinction between amplifiers and oscillators. The former is one kind of driven circuit, the output of which is controlled by its input. Oscillators belong to the group of autonomous circuit, which generates time-varying signals without time-varying stimulus.
By definition, amplifiers operate below \( f_{\text{max}} \) to provide power gain. For oscillators, the situation is more complicated: The power gain of active devices within the oscillator should be greater than unity to start the oscillation. As the oscillation amplitude grows, the power gain of the two-port (including device parasitic and loading) gradually compresses to unity when the circuit reaches steady-state oscillation. Thus, it could be argued that devices within oscillators operate at the \( f_{\text{max}} \) of its embedded two-port in steady-state oscillation. To take the gain compression into account, large-signal parameters should be used for the analysis.

Unlike analog circuit designers who deal exclusively with voltage and current gains, microwave circuit designers are more comfortable with power gain. Momeni [7] has thus shown a refreshing view about the optimum voltage gain and phase shift for a given two-port to oscillate at \( f_{\text{max}} \):

\[
A_V = \frac{V_2}{V_1} = A_{\text{OPT}} = \sqrt{\frac{G_{11}}{G_{22}}}
\]

(8)

\[
\phi = \angle \frac{V_2}{V_1} = \phi_{\text{OPT}} = (2k + 1)\pi - \angle (Y_{12} + Y_{21}^*)
\]

(9)

\( A_{\text{OPT}} \) and \( \phi_{\text{OPT}} \) are the optimum voltage gain and phase shift for the two-port at \( f_{\text{max}} \). Equations (8) and (9) are derived assuming no clipping to the power rails occurs inside the circuit. If clipping happens, another set of equations apply for \( A_{\text{OPT}} \) and \( \phi_{\text{OPT}} \).

It can be shown that biconjugate matching automatically satisfies Eqs. (8) and (9) at \( f_{\text{max}} \).

Under biconjugate matching, \( G_p \) reaches unity at \( f_{\text{max}} \). Equation (2) can be rewritten as

\[
|A_V| = |A_{\text{OPT}}| = \sqrt{\frac{\text{Re}(Y_{\text{IN}}^*)}{\text{Re}(Y_{\text{IN}}^{\text{OPT}})}} = \sqrt{\frac{\text{Re}(Y_{S,\text{OPT}})}{\text{Re}(Y_{L,\text{OPT}})}}
\]

(10)

\( Y_{S,\text{OPT}} \) and \( Y_{L,\text{OPT}} \) are the optimal source and load admittance for biconjugate matching:

\[
Y_{S,\text{OPT}} = \frac{|Y_{12}Y_{21}| \times \sqrt{K^2 - 1} + j \times (\text{Im}(Y_{12}Y_{21}) - 2 \times \text{Re}(Y_{22}) \times \text{Im}(Y_{11}))}{2\text{Re}(Y_{22})}
\]

(11)

\[
Y_{L,\text{OPT}} = \frac{|Y_{12}Y_{21}| \times \sqrt{K^2 - 1} + j \times (\text{Im}(Y_{12}Y_{21}) - 2 \times \text{Re}(Y_{11}) \times \text{Im}(Y_{22}))}{2\text{Re}(Y_{11})}
\]

(12)

By substituting Eqs. (11) and (12) into (10), we have

\[
|A_{\text{OPT}}| = \sqrt{\frac{G_{11}}{G_{22}}}
\]

(13)

We are now in the position to derive \( \phi_{\text{OPT}} \). First, we have to define the net power flowing into the two-port shown in Figure 1(b):

\[
P = V_1 \times I_1^* + V_2 \times I_2^*
\]

(14)

At \( f_{\text{max}} \), the power gain of the two-port drops to unity, which means the real power consumed by the two-port equals the real power generated. Thus, the real part of Eq. (14) equals zero:
\[ P_R = |V_1|^2 \times \left[ G_{11} + G_{22} \times |A_V|^2 + \text{Re} \left( Y_{12}^* A_V^* + Y_{21}^* A_V \right) \right] = 0 \]  

(15)

Substituting Eq. (13) into (15), we have

\[ 2 \sqrt{G_{11} G_{22}} + \text{Re} \left[ (Y_{12} + Y_{21}^*) e^{i\phi} \right] = 0 \]

(16)

Since the Mason’s invariant equals unity at \( f_{\text{max}} \), Eq. (1) could be rearranged as

\[ \frac{|Y_{12} + Y_{21}^*|^2}{4G_{11} G_{22}} = 1 \]

(17)

Equation (16) thus equals

\[ 2 \sqrt{G_{11} G_{22}} \times \left[ 1 + \cos \left( \phi + \angle \left( Y_{12} + Y_{21}^* \right) \right) \right] = 0 \]

(18)

We have

\[ \phi_{\text{OPT}} = (2k + 1)\pi - \angle \left( Y_{12} + Y_{21}^* \right) \]

(19)

Since the above derivation is restricted to \( f_{\text{max}} \), it would be interesting to observe the possible deviations of Eqs. (8) and (9) with respect to the two-port’s voltage gain and phase shift under biconjugate matching when operating at frequency below \( f_{\text{max}} \). A SiGe HBT transistor is used as an example. The emitter width and length of the transistor is 0.12 and 2.5 \( \mu m \), and the emitter current density is biased for peak \( f_{\text{max}} \). The source and load admittance are adjusted for biconjugate matching under each frequency evaluated (Figures 2 and 3).

It is clear that Eqs. (8) and (9) are only strictly valid at \( f_{\text{max}} \), but the optimum phase shift calculated with Eq. (9) tracks reasonably well with the results obtained with biconjugate matching over a wide frequency range.

Figure 2.
Comparison of the phase shift of a SiGe HBT transistor under biconjugate matching and the optimum phase shift calculated with Eq. (9).
2.2 Circuit topology for THz sources, detectors, and transceivers

Among the many potential benefits offered by THz application, the large bandwidth available is the most obvious one. However, a lot of design issues need to be addressed in order to truly harness this bandwidth potential. We discuss this problem in terms of $\text{SNR}$ at the receiver:

$$\text{SNR} = \frac{P_r}{kTBF} = \frac{\lambda^2}{4\pi d^2} \times \frac{P_t G_t G_r}{kTFL}$$

where $B$ is the receiver bandwidth. For communications, we would like $B$ to scale with frequency. For imaging applications, sometimes $B$ is not that important once it reaches certain value as only the range resolution scales with $1/B$. The cross-range resolution scales inversely with wavelength $\lambda$. Thus, we lump SNR and $B$ together for trade-offs. $P_t$ and $P_r$ are the power transmitted and received by the transmitter (Tx) and receiver (Rx). $G_t$ and $G_r$ are the gain of the transmitting and receiving antenna. $K$ is the Boltzmann constant. $T$ is the ambient temperature. $d$ is the distance between the transmitter and receiver. $F$ is the noise factor of the receiver. $L$ represents the loss in the Tx and Rx system, and we assume it scales with $f^{0.5}$ to $f^{1}$.

Assuming constant drive power, $P_t$ for terahertz transmitters approximately follows a $P_t \propto 1/f^2$ relationship since the maximum unilateral gain $U$ follows a -20dB/decade slope above $f_{\text{max}}/2$. $F$ generally scales with $f$. $\text{SNR} \times B$ then scales with $1/f^{3.5}$ to $1/f^6$, which is a really disheartening result. This partly explains why the current silicon THz links are usually demonstrated with link distances ranging from centimeters to meters.

Before leaving this chapter in despair, we can try to manipulate Eq. (20) a little bit further:
The first term is by all means beyond our control, and we do not want to change the second term for now. So, what can we do about the last term? It happens that if we were to keep the two-antenna size constants while increasing the frequency, $G_t$ and $G_r$ each come with a nice $\lambda^2$ on the denominator. Equation (21) thus equals

$$\text{SNR} \times B = \frac{1}{16\pi k} \times \frac{\lambda^2}{Td^2L} \times \frac{P_t G_t G_r}{F}$$

Equation (21) thus equals

$$\text{SNR} \times B = \frac{1}{16\pi^2k} \times \frac{\lambda^2}{Td^2L} \times \frac{4\pi A_{p_t} \varepsilon_t P_t}{\lambda^2} \times \frac{G_r}{F} = \frac{A_{p_t} \varepsilon_t P_t}{4\pi k Td^2L} \times \frac{G_r}{F}$$

Equation (22) could be rewritten as

$$\text{SNR} \times B = \frac{A_{p_t} \varepsilon_t N_t P_{te}}{4\pi k Td^2L} \times \frac{N_r G_{re}}{F_{\epsilon}}$$

where $N_t$ and $N_r$ are the numbers of transmitters and receivers. $P_{te}$ is the output power for each transmitter. $G_{re}$ is the gain of the receiving antenna for each receiver. For active phased arrays, antenna elements with their corresponding transmitters and receivers are evenly distributed with a pitch of about $\lambda/2$. Thus, $N_t$ and $N_r$ are proportional to $1/\lambda^2$. It is proven that for phased array with lossless combing, the $G/F$ term in Eq. (22) scales with $N_r$ [10]. Assuming constant $\varepsilon_t$ and $G_{re}$ with respect to $f$, we see that $\text{SNR} \times B$ scales with $f^{-0.5}$ to $f^0$.

The six orders of magnitude difference of $\text{SNR} \times B$ deduced from Eqs. (20) and (23) give us a hint of the size of the design space for silicon THz systems.

2.2.1 THz Sources

When talking about silicon THz sources, a plethora of options is available that varies in functionality, complexity, and performance. For incoherent imaging applications, the most important metrics are output power and efficiency, whereas for spectroscopy, the bandwidth is the most important specification. Perhaps the most demanding application is for THz communications, for which output power, power efficiency, tuning range, phase noise, harmonics, and spurious suppression are all important parameters. This subsection aims to give a brief and incomplete introduction to what has been done in this area.

THz signal can be generated either by frequency multipliers or by on-chip oscillators.

In multipliers, the MOS or bipolar transistor is driven heavily to generate highly nonlinear current. The intended frequency component is then extracted with other components filtered. If efficiency is important, the active device should be conjugate matched for the fundamental and the intended harmonic. The impedance presented to the device at other harmonics is usually short or open circuit to maximize energy transfer between the fundamental and the intended harmonic. But we should not be overzealous about this goal; usually taking care of the first two or three harmonics is enough since the higher harmonics are insignificant. The transistor also has to be biased correctly for maximum harmonic generation. For MOS transistor, the conduction angle is specified. Like power amplifiers, efficient
MOS multiplier works in the class AB, B, or C region depending on the frequency, multiplication factor, and input power. For bipolar transistor, this efficiency is a function of $V_{be}$ (or collector current density) [11].

Relationship of phase noise between the harmonic and the fundamental for multipliers is [11]:

$$S_{out}(\Delta \omega) = N^2 S_{fund}(\Delta \omega) + S_{harm}(\Delta \omega) + S_{amp}(\Delta \omega)$$  \hspace{1cm} (24)

where $S_{out}(\Delta \omega)$ and $S_{fund}(\Delta \omega)$ represent the spectral density of phase fluctuations for the harmonic and fundamental signal with $\Delta \omega$ radians offset from the carrier. It is in dBc/Hz form. $N$ is the frequency multiplication ratio. The last two terms in Eq. (24) represent the added noise from the harmonic-generating device and the ensuing amplifiers (if any). The combined value is usually less than 3 dB for reasonably designed circuits.

Multipliers are usually compact and broadband, but they are not as efficient as (well designed) oscillators. A 90–300 GHz transmitter based on distributed quadrupler is designed for spectroscopy and imaging [12]. It resembles the distributed amplifier (DA) in that the input and output capacitance of active device are absorbed in the input and output transmission line. Differential quadrature signal is used to drive two groups of quadrupler diff-pairs, the current of which is then combined to cancel the second harmonic. As another example, quadrupler is used in an 8-element 400 GHz transmitter phased array to replace power amplifier [13]. This also simplifies phase shifter design since the fundamental signal only needs to be shifted within 90 degrees as the phase shift is multiplied by four.

In oscillators, the transistor is made unstable by intentionally introducing positive feedback around it. Steady-state oscillation occurs at the frequency where the open-loop transfer function equals –1 (Barkhausen’s criteria). Since the $f_{max}$ of most silicon devices is below 300 GHz, harmonic generation is employed. The fundamental oscillation frequency is usually around 100 GHz for better phase noise, since larger oscillation amplitude and hence better SNR are easier to obtain at lower frequencies.

A high efficiency and scalable $4 \times 4320$ GHz oscillator array is built in SiGe BiCMOS technology [9]. The oscillator shown in Figure 4(a) oscillates at 160 GHz and is optimized for optimum transistor voltage gain and phase shift as discussed in Section 2.1. $Y_1$ and $Y_2$ represent the source and load admittance for the transistor. A transmission line with impedance $Z_0$ and electrical length $\theta_{TL}$ is used to introduce

![Figure 4](image.png)

**Figure 4.** Topology of (a) self-sustained oscillator and (b) conventional cross-coupled oscillator and its equivalent circuit at second harmonic.
feedback. The transmission line spans approximately a quarter wavelength at
the second harmonic to transform the relatively small impedance of the gate node to
high values at the drain node. Since harmonic signals are in the current form,
boosting the output impedance at the harmonic frequency substantially improves
output harmonic power. The DC-to-THz radiation efficiency is 0.54%. Early reports
of THz oscillators based on push-push oscillators are less efficient than this partly
due to the insufficient output impedance at second harmonic. As is shown in
Figure 4(b), the cross-coupled pair in a push-push oscillator is effectively a
diode-connected transistor at second harmonic.

Frequency tuning of oscillators is usually done by varying the capacitance of
varactor in the oscillation tank. Higher oscillation frequency translates to
smaller capacitance, which is problematic for small varactors as its parasitic
capacitance would swamp the variable capacitance. This would severely constrain
the oscillator’s tuning range.

Shown in Figure 5(a) and (b) are the cross section of a NMOS varactor and its
small signal model with its source AC grounded. To increase its Q factor, the
minimum channel length is usually employed, which further increases the overlap
capacitance \( C_{ov} \) between the two terminals.

A straightforward way to increase the tuning ratio of the varactor is to place an
inductor to partially absorb the parasitic capacitance. A 300 GHz differential Clapp
push-push VCO with 8.5% tuning range and phase noise of −85 dBc at 1 MHz offset
is reported in [14]. Its simplified schematic and equivalent small signal circuit for
calculation of the input impedance seen at the base is shown in Figure 6. Note that
the base resistance, the depletion capacitance between base and collector, and the
output resistance of the transistor are ignored.

The input impedance seen from the base is

\[
Z_B = -\frac{1}{j\omega C_{pi}} \left( 1 + \frac{g_m + j\omega C_{pi}}{1/R_{var} + j\omega C_{eff}} \right)
\]  

(25)

where \( C_{eff} \) is

\[
C_{eff} = C_{var} \times \left( 1 - \frac{1}{\omega^2 L_{deg} C_{var}} \right) = C_{var} \times \left( 1 - \frac{\alpha_0^2}{\omega_0^2 \alpha_{OSC}} \right)
\]

(26)

\( R_{var} \) represents the loss in varactor; its conductance is \( 1/Q \) of the varactor at the
evaluation frequency of the quality factor. The resonant frequency \( \omega_0 \) for \( C_{var} \) and
\( L_{deg} \) is set below the main oscillation frequency.

The equivalent series resistance and capacitance for \( Z_B \) is

![Diagram](image-url)

**Figure 5.**
Cross section of NMOS varactor (a) and equivalent circuit with grounded source (b).
Another interesting property of the circuit is that the oscillation frequency for common mode is intentionally set to the second harmonic. $C_r$ in Figure 6(a) and (c) forms a series resonator with the tank, raising the second harmonic voltage seen at the base significantly. This boosts the second harmonic generation. The output power for the two versions of the VCO is 0.6 dBm and 0.2 dBm, respectively.

As is evident from Eq. (27), the negative resistance seen at the base of the capacitively degenerated transistor could be used to mitigate the loss of the varactor [15]. This property is used in [16] to build a 300 GHz triple-push VCO. The tuning range is 8%, and the phase noise is $-101.9 \text{ dBc/Hz}$ at 1 MHz offset for the 100 GHz main loop. That translates to a phase noise of $-80.28 \text{ dBc/Hz}$ at 1 MHz offset for 300 GHz assuming noiseless multiplication.

An interesting observation is that inductors actually have better quality factor than varactors at higher frequency. A carefully designed inductor has a Q of 15–20 at 100 GHz, whereas the quality factor for varactor is around 2–5 at that frequency. It would be nice if we can replace the varactor with a high-quality metal-insulator-metal (MIM) or metal-oxide-metal (MOM) capacitor; we then need to figure out how to tune the inductance of these nice inductors. Figure 7 shows one such circuit [17] which is also based on Clapp oscillators.

We know that inductance at the base generates a negative resistance and a positive inductance seen from the emitter, as LNA designers can attest to [18]. Careful derivation of $Z_E$ leads to the following results [14]:

$$
Z_E \approx \frac{1}{g_m} - \frac{\omega^2 L_{BB}}{\omega_T} + \frac{\omega^2 R_{BB}}{\omega_T^2} + j\omega \left( \frac{R_{BB}}{\omega_T} - \frac{1}{g_m \omega_T} + \frac{\omega^2 L_{BB}}{\omega_T^2} \right)
$$

Thus, we can tune the inductance by varying the $g_m$ of the transistor. For bipolar transistors, $g_m$ equals the emitter current divided by the thermal voltage $V_T$ (26 mV in room temperature).

The impedance at the emitter is mapped to the resonant tank through the transformer formed by $L_E$ and $L_T$ in Figure 7(b). The currents of the two active inductors are controlled by tail current source $Q_5$. The second harmonic current is extracted through $L_C$, and the degeneration resistor $R_{EE}$ is used to improve
common-mode rejection. Two versions of this VCO with different oscillation frequency are built; the tuning range are 3.5 and 2.8%, respectively. The harmonic power suffers due to the inclusion of $R_{EE}$, the output power at 201.5 and 212 GHz are $-7.2$ and $-7.1$ dBm, respectively. But the phase noise performance is very good, with $-87$ and $-92$ dBc at 1 MHz offset for the two VCOs. This translates to $-83.5$ and $-89$ dBc at 1 MHz offset if they oscillate at 300 GHz, which is comparable to the circuit shown in Figure 6 that generates 0 dBm at 300 GHz.

This raises interesting question as smaller output power usually means inferior phase noise performance. One possible explanation is that the noise current at the second harmonic in Q1 and Q2 of Figure 6 generates large noise voltage at the emitter since these nodes are open circuit due to resonance, thus amplifying the noise current at second harmonic. It should be noted that the phase noise could be improved substantially by breaking the noise current path at second harmonic [19].

A salient feature of Clapp oscillator is the inherent isolation of the load from the tank. This helps to preserve the quality factor of the tank, leading to a better phase noise and less load pulling (variation of oscillation amplitude and frequency caused by load variation). The problem of low output impedance at second harmonic is also mitigated in this topology as the base is isolated from the drain.

One starts to wonder if there is a way to further improve the phase noise performance. We know there is a trade-off between noise performance and power consumption, but we are kind of stuck here: We need larger transistor to burn larger power, but the larger capacitance of the device means smaller inductors in the tank, which complicates the design and ultimately degrades the $Q$. We need a larger design space here.

One way to do that is to build an array of $N$ oscillators and lock them together. Theoretically the phase noise would drop by $N$ as the SNR increases by $N$, and the output power would also increase by $N$. Better still, if we distribute them evenly with a certain pitch (normally $\lambda/2$) and radiate the power out collectively, the energy would focus in certain direction, improving the EIRP by $N^2$. Tousi et al. shows one such design [20]:

Each individual oscillator shown in Figure 8 is a cross-coupled push-push oscillator. It is designed for optimum fourth harmonic generation by making sure that the gate is isolated from the drain at the fourth harmonic. The second harmonic is rejected by the narrow band on-chip antenna. Each oscillator is coupled to other oscillators as shown in Figure 8(a) through active phase shifters. Figure 8(a) forms...
a unit cell through which a 2-D oscillator lattice could be formed as shown in Figure 8(b). One nice feature of this design is the use of phase shifter as coupling elements, as this provides us with a new way of tuning the frequency of these injection-locked oscillators [21]:

$$\frac{\Delta f_0}{f_0} = k \sin (\Delta \phi(t))$$ (30)

Equation (30) is derived from the Adler’s equation under locked conditions. $\Delta f_0$ is the frequency difference of the injected signal and the free-running frequency $f_0$ of the slave oscillator. $\Delta \phi(t)$ is the phase difference. $K$ is a factor relating to the quality factor of the oscillation tank, amplitude of the injected signal, and the free-running oscillation amplitude of the tank.

Since the total phase shift through the loop in Figure 8(a) is $2k\pi$, the phase shift through the oscillator plus that of the phase shifter is constrained to a set of fixed values. The oscillation frequency of the whole array is changed if we apply a common shift to all the phase shifter in Figure 8(b). Interesting thing happens when we tune the phase shifter connected with an individual oscillator. If we apply a common shift to the four shifters, this disrupts the local loop momentarily as the instantaneous phase of the affected oscillator jumps to a new value to accommodate the change. This ability of changing the phase of radiating elements individually turns this design into a phased array. Measurements show that the beam could be steered $\pm 50$ degree in the E plane and $\pm 45$ degree in the H plane. The frequency tuning range is 2.1%. The peak EIRP is 17.1 dBm, and the phase noise is $-93$ dBc at 1 MHz offset for this 338 GHz array, which show substantial improvements over single oscillators.

A brief comparison between multipliers and harmonic injection-locked VCOs is also in place: The former is generally compact and wideband. They add negligible phase noise if designed properly as dictated by Eq. (22). The biggest issue is the harmonics, which leads to annoying LO spurs that leads to spurious emission and corrupts received signals. The latter comes with higher efficiency and much higher harmonic rejection ratio, but the bandwidth is limited. The close-in phase noise is
dominated by the source just like the multiplier, but the far-out phase noise is dominated by the VCO. For mm-wave frequency synthesizers, the two options generally achieve comparable phase noise performances [11]. Since we have to use multipliers to get from mm-wave to THz anyway, this same conclusion holds for THz. For communication applications, it is advisable to use PLL-locked or injection-locked VCOs to generate relatively high LO frequency and use multipliers to boost it to THz frequency (N-push VCOs does this in one place).

2.2.2 THz detectors

THz detectors utilize the nonlinearity of active devices to directly rectify THz signal to DC. A lot of devices could be used, like diode-connected NMOS transistor [22], CE (common emitter) [23] or CB (common base) connected [24, 25] SiGe HBT, CMOS-compatible Schottky diode [26], or P+/n-well diode [27]. THz reception with detectors is incoherent, that is, only the amplitude information is recovered at the receiving side, which limits THz detectors almost exclusively to incoherent THz imaging applications. The strength of THz detectors lies within their simplicity: They do not need LO (local oscillator) signal to do the THz downconversion. The received THz signal self-mix themselves to DC through even-order nonlinearity of the device. This makes scaling extremely easy, as only low-frequency routing is needed, whereas LO-driven mixer needs cumbersome and power-hungry LO tree which quickly becomes unmanageable when the array gets large. A 1024 pixel NMOS detector array [22] in 65-nm CMOS process showcases the impressive scalability of THz detectors.

However, this flexibility comes with a price: The gain and noise performance of detectors is quite limited, and the specification “responsivity” and “noise equivalent power (NEP)” are used in place of conversion gain and noise figure. The responsivity is defined as the voltage output divided by received power, and NEP is defined as the output noise voltage density divided by responsivity.

The bandwidth of imaging application is usually below 1 MHz; thus, technologies with lower 1/f noise corner frequencies like SiGe HBT or P+/n-Well diode are preferred. The 1/f noise corner frequencies for SiGe HBT and P+/n-well diode are below 1 kHz and 10 kHz, respectively, whereas for NMOS transistor or Schottky diodes, the numbers are well above 1 MHz. For SiGe HBTs, it is shown that CB-connected topology has higher responsivity than CE-connected topology when operating above $f_{\text{max}}$ [24].

The principle of THz imaging with detectors largely follow their optical counterpart: They use THz lenses to do the focusing. The problem is that THz wavelength is 2–3 orders of magnitude larger than visible lights, thus large and bulky THz optics are required for reasonable imaging resolutions. They require a lot of effort to set up the imaging setup with the invisible THz radiations [25]. This is the innate deficiency with incoherent imaging. Coherent imaging with THz transceivers could get rid of those optics.

2.2.3 THz transceivers

For transceivers working at lower frequencies, the transmitter and receiver are usually integrated on one chip, and they share the common RF port through switches or duplexers (bandpass filters tuned for simultaneous transmit and receive on different bands). Up to now, neither option is satisfactory for fully integrated silicon THz circuits.

One solution for integrated THz transceiver is to share the antenna and figure out ways to isolate the Tx (transmitter) and Rx (receiver). Park et al. have shown a
fully integrated 260 GHz transceiver based on shared leaky-wave antenna [28]. The leaky-wave antenna resembles a lossy transmission line (TL); thus, the Tx and Rx ports could be placed on either end of the antenna. When the transmitter is working, the receiver is turned off and terminates the TL on its side. The same holds true for the receiving mode. The problem with the leaky-wave antenna is that they are relatively long (1.2 mm or 2.5 $\lambda$ in this design). Statnikov et al. [29] have shown a fully integrated 240 GHz frequency-modulated continuous wave (FMCW) radar transceiver based on shared dual-polarization antenna. A quadrature hybrid coupler is used as a polarizer for the dual-polarization antenna and duplexer for the Tx and Rx. Isolation of the Tx and Rx depends on the orthogonality of left hand circular polarized (LHCP) and right hand circular polarized (RHCP) waves. The Tx and Rx interface with two orthogonal port of the branch-line coupler and are isolated from each other. In Tx mode, the branch-line coupler excites the LHCP mode of the antenna. When the transmitted wave hits a target and bounces back, it changes to RHCP and is subsequently routed to the receiver through the coupler. This scheme is not directly applicable for point-to-point communications, just like frequency-division duplexing (FDD)-based transceivers could not communicate directly with each other.

Another solution is to use two antennas. For FMCW radars, the leakage from the TX to the Rx results in strong interferences around DC [30]. This raises the noise floor in the range spectrum. With area permitting, the Tx and Rx antenna should be separated further apart for better isolation. The measured crosstalk between the two antennas with a separation of about 1.8 mm in a 160 GHz FMCW radar transceiver is below 31 dB [31]. This isolation might be adequate for FMCW radar applications, but it is still wanting for communications.

Transceiver-based THz imaging makes coherent imaging possible, as both the magnitude and phase information of the signal from targets are retained. With both information available, it is possible to get rid of bulky THz optics by sampling the THz field directly and do the focusing digitally. The THz field is usually sampled on a 2-D plane with different THz frequencies; this is fulfilled by raster scanning a FMCW transceiver (or the sample). For a given point in space, the round-trip phase delay from the transceiver to that point is a function of its position and sampling frequency. By raster scanning the transceiver or the sample under different frequencies, its phase delay variation is orthogonal to every other point in the sampling space. This forms the basis for the 3-D imaging through the back-projection algorithm. 3-D imaging based on SiGe FMCW transceivers is reported by several groups [32, 33], showcasing the great potential for low-cost THz imaging applications.

For communication applications, the modulation scheme plays a major role in deciding the transceiver architecture. Low-complexity modulation schemes like on-off keying (OOK) and binary phase shift keying (BPSK) lead to robust and power-efficient design, but the spectrum efficiencies are relatively low. Modulation schemes like 32 QAM and 128 QAM lead to much higher spectrum efficiency, but they are quite demanding on linearity and phase noise performance, and they require image-rejection architectures as the spectra of QAM are asymmetric around the carrier. The upper sideband (USB) and lower sideband (LSB) of the spectra become each other’s own image when converted to baseband, and image-rejection is needed to avoid signal corruption. Image-rejection modulation/demodulation is difficult in THz range as I/Q mixers are required. It is very difficult to guarantee phase and amplitude matching for the I/Q LO signal for adequate image-rejection at THz frequency.

A 210 GHz fundamental transceiver chipset with OOK modulation is demonstrated in a 32-nm SOI CMOS process [34]. Ideally speaking, power amplifier (PA)-based fundamental operation is more power-efficient than frequency multipliers.
This helps to boost efficiency of the whole system as PAs are usually the most power-hungry circuits in transceivers. Perhaps the most difficult part of this design is controlling the oscillator pulling effect. Since the PA works at the same frequency as the on-chip VCO, significant coupling could occur between PA and VCO. The injection-locking effect would impact the phase noise performance heavily. The on-chip antenna used in this design only makes things more difficult. To improve the VCO performance, a stacked cross-coupled VCO topology is used to boost oscillation amplitude, improving its robustness in response to interferences.

A 240-GHz direction-conversion transceiver in SiGe BiCMOS technology is demonstrated with BPSK capability. BPSK is a constant-envelope modulation, which means the PA could be driven to saturation for better power efficiency. The spectra of BPSK modulation are symmetric around its carrier (symmetrically modulated), making direct conversion easier to implement as no image-rejection is needed. A 30 GHz LO signal is supplied to this transceiver, and on-chip ×8 multipliers are used for the 240 GHz LO generation. This helps to alleviate the detrimental effect of LO spur caused by multipliers since they are separated by twice the baseband bandwidth (15 GHz). An on-chip antenna with 1-dB bandwidth of 33 GHz is achieved partly due to the local back etching (LBE) technology used. The silicon substrate below the antenna is removed, resulting in a low-loss air cavity below the antenna. The transceiver link is tested with 15 cm separation, and an impressive 6-dB bandwidth of 35 GHz is obtained. A 25 Gbps wireless link is demonstrated by this transceiver with no equalization. One problem with direct conversion using no I/Q demodulation is that the demodulated signal’s SNR is dependent on the phase difference between the Tx LO and Rx LO. A phase shifter is used in this test in case manual tuning is required to boost the SNR.

A 230 GHz direct-conversion 16-QAM 100-Gbps wireless link is demonstrated in a 40 nm CMOS process [36]. The I/Q mixer directly interfaces with on-chip antenna to avoid bandwidth limitation introduced by LNA. On-chip LO multiplier chain is used to convert the external 13.75–16 GHz LO to 220–256 GHz. The baseband bandwidth is around 14 GHz; this poses challenge as the spacing of LO spur is comparable to this bandwidth. This leads to spurious modulation that overlaps with desired signal. Nevertheless, 100 Gbps with an EVM of 17% is demonstrated.
the 32-QAM modulation with an EVM of 8.9% is achieved with 105 Gbps. No on-chip antenna is used as this chip is intended to drive high-power THz devices like traveling-wave tubes.

2.3 THz interface for efficient power transfer

The THz interface serves as a gateway between the circuit and the outside world. The efficiency of this interface greatly impacts the performance of the overall system. A simple derivation of the transceiver’s link budget would highlight the importance of this interface (Figure 9):

\[ P_R = P_T - I L_2 - I L_0 - I L_1 \]

where \( P_R \) is power received at the receiver, \( P_T \) is the output power of the transmitter, and \( I L_2 \) and \( I L_1 \) are the loss for THz interface at the transmitter and the receiver. \( I L_0 \) is the loss associated with the propagation medium. If the medium is free space, which is often the case, \( I L_0 \) equals

\[ I L_0 = 20 \times \log_{10} \left( \frac{\lambda_0}{4\pi d} \right) + G_{TX} + G_{RX} \]

\( \lambda_0 \) is the free-space THz wavelength, \( d \) is the propagation distance, and \( G_{TX} \) and \( G_{RX} \) are the gain of antenna employed at the transmitter and receiver. Note that Eqs. (31) and (32) are in dB form. Also keep in mind that Eq. (32) is only valid under far-field conditions, that is, the radiation field seen by the receiver is not reactive and could be approximated by plane waves. Common criterion for far-field condition is that the receiver is separated from the transmitter by at least \( 2D^2/\lambda \), where \( D \) is the maximum overall dimension of the transmitting antenna.

To maximize \( P_R \), we have to minimize the loss at the interface and increase the gain of the antennas. This section covers both areas.

For THz silicon chips, grounded coplanar waveguide (GCPW) is the most prevalent medium for on-chip THz routing. It is a combination of coplanar waveguide with microstrip line. This configuration has several merits: First, the ground plane of the microstrip line shields the signal from the electrically thick silicon substrate, which reduces loss and prevents the signal from leaking into substrate modes; the coplanar waveguide makes interface with outside world easier, be it through flip-chip bonding or on-chip probing since the ground conductor lies in proximity with the signal trace.

A 90–300 GHz transmitter and a 115–325 GHz receiver are flip-chip bonded to a liquid-crystal polymer (LCP) substrate [12]. This connects the chip to the 100–280 GHz Vivaldi antenna on the LCP substrate. Such wideband antenna is extremely difficult to realize on-chip. As another example, a CMOS 300 GHz transmitter chip is flip-chip bonded to a GCPW-to-WG transition module [38] implemented on a
low-cost glass epoxy PCB. Once transitioned to waveguide interfaces, the chip could be interfaced to a plethora of THz components like horn antennas and high-power amplifier modules. The packaging loss is 8 dB, which includes the transition loss (flip-chip bonding and the GCPW-to-WG), impedance mismatch, and loss in the epoxy material. It should be noted that gold stud bumping is used in both cases, which is compatible with conventional wire bonders and is quite convenient for R&D labs.

An effective way to lower transition loss is to radiate the THz signal directly from the chip. Since the high permittivity silicon substrate readily traps the THz radiation, there are basically two lines of thoughts regarding on-chip antenna design. One is to accept this coupling and take this into account while designing antennas; the other is to eliminate the substrate mode altogether.

The first approach tries to make use of the electrically thick antenna to improve the antenna bandwidth. Metal reflector are placed underneath the substrate to reflect energy back, which is often the top layer of the PCB under the chip. To make sure the reflected power adds constructively with the surface radiation, the chip should be \((2k + 1)/4\lambda\) thick \([34, 39, 40]\). The phase delay of the reflected wave increases linearly with frequency, which limits the bandwidth. Artificial magnetic conductor (AMC) reflector in place of the solid metal reflector on the PCB is used to compensate this phase shift \([41]\), which extends the on-chip antenna bandwidth substantially. The antenna efficiency is within 0.5 dB of the peak efficiency between 200 and 300 GHz. Note that the chip should be 1/2\(\lambda\) thick in this case as AMC introduces zero phase shift at the center frequency.

Although promising, allowing reflections within the substrate increases couplings between antenna elements and on-chip passive components. This makes circuit performance sensitive to antenna location and substrate dimensions (both lateral dimension and chip thickness), which is undesirable for designing arrays.

There are two widely adopted approach to eliminate the substrate mode. The first approach involves attaching a high-k dielectric lens on the back side of the chip, and the antenna radiates through the lens \([22, 42–44]\). This approach offers high directivity and improves efficiency. The most obvious drawback is the need for nonstandard packaging, which could be quite costly. An intuitive way of understanding why this structure can eliminate the substrate mode can be found in \([8]\).

The second approach is more straightforward: The substrate is shielded from the radiating element using one or a few low-level metal layers. This eliminates the possibility of coupling with substrate modes, but it limits the bandwidth and radiation efficiency severely since the radiating element is only a few microns away from the ground due to the restriction of silicon backend process. The radiating element thus forms a high Q tank with the ground, and the bandwidth of the antenna is on the order of a few percent \([13, 20, 25]\). The obvious way to increase bandwidth and radiation efficiency is to lower the Q, which could be accomplished by increasing the volume of the resonant tank. One solution is to add a dielectric superstrate above the radiating element \([45, 46]\), which diverts some of the electric fields from the dielectric layer below the radiating element, increasing the resonance volume considerably. The radiation efficiency increases with the superstrate till the onset of the TE1 mode, which limits the thickness of the superstrate to \(\lambda_0/4\sqrt{\varepsilon_r} - 1\) where \(\varepsilon_r\) is the dielectric constant of the superstrate. Dielectric resonator antenna (DRA) is another option \([47, 48]\). This antenna composes of a dielectric resonator on top of the chip and a feed element in the top metal layer of the chip. The feed element is used to excite resonance inside the resonator, through which the THz signal radiates. The size of the resonator and relative position of feed could be adjusted for intended oscillation frequency and resonance mode. For a 270 GHz microstrip
antenna, the gain is enhanced by 4 dB, and the 3-dB gain bandwidth is extended by 100% when an dielectric resonator is used [49].

For phased array applications, the patch antenna with ground shield is the most straightforward approach.

3. Summary

Silicon THz circuit design is an active research area open to innovations on multiple levels. We need better passive components, better circuits, and the most important thing is we need to come up with better ways of building THz arrays. Scaling is the key for significantly boosting the performance of silicon THz systems as we venture into this last untapped spectrum [50].

Author details

Li Zhuang*, Cao Rui, Tao Xiaohui, Jiang Lihui and Rong Dawei
Key Lab of Aperture Array and Space Applications, He Fei, China

*Address all correspondence to: lizhuang@mail.ustc.edu.cn

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
References


[44] Pfeiffer UR et al. A 0.53 THz reconfigurable source module with up to 1 mW radiated power for diffuse illumination in terahertz imaging applications. IEEE Journal of Solid-State Circuits. 2014;49(12):2938-2950


[48] Li C, Chiu T. 340-GHz low-cost and high-gain on-Chip higher order mode dielectric resonator antenna for THz applications. IEEE Transactions on
Terahertz Science and Technology.
2017;7(3):284-294
