The Future of High Frequency Circuit Design

Invited Paper

Ali Hajimiri
California Institute of Technology, Pasadena, CA 91125, USA

Abstract[1]

The cut-off wavelengths of integrated silicon transistors have exceeded the die sizes of the chips being fabricated with them. Combined with the ability to integrate billions of transistors on the same die, this size-wavelength cross-over has produced a unique opportunity for a completely new class of holistic circuit design combining electromagnetics, device physics, circuits, and communication system theory in one place. In this paper, we discuss some of these opportunities and their associated challenges in greater detail and provide a few of examples of how they can be used in practice.

Introduction

A significant milestone went unnoticed a few years ago when the cut-off wavelengths[2] of silicon transistors exceeded the dimensions of typical chips implemented in those processes (Fig. 1). This cross over has important implications in terms of the way high-frequency mm-wave circuits should and will be architectured and designed in the coming decades. This is particularly important considering the remarkable and unparalleled growth and expansion that the field of integrated circuits has experienced over the last 60 years [1]-[4]. Today, it is possible to integrate more than a billion functional transistors capable of operating at mm-wave frequencies on a single silicon die. This (practically) unlimited number of high-speed transistors combined with the crossing over of the chip-sizes and the cut-off-wavelengths have produced a truly unique opportunity that cannot be ignored and must be fully harnessed.

Very complex integrated circuits capable of operating at high frequencies have been designed and implemented by applying the traditional lumped circuit concepts based on Kirchhoff's voltage and current laws (KVL and KCL). This has been possible because transistor dimensions are still much smaller than the wavelength in most of these applications. Full-scale and/or simplified electromagnetic field solvers are often used to account for the detailed electromagnetic behavior of passive elements, such as inductors, capacitors, interconnects, and to produce lumped circuit equivalents that can be used in conjunction with the active devices. Despite this, the integrated circuit design is mostly done using traditional lumped design approaches augmented with selective application of electromagnetic field solvers. This traditional lumped circuit approach is illustrated in Fig. 2. A tremendous amount of effort has been placed into creating reliable, scalable computer aided design (CAD) tools to design, verify, and manage very complex circuits made this way. This level of abstraction has enable very sophisticated mixed-mode chips to be made over the last decade.

At higher frequencies, where the chip dimensions become comparable with the wavelength, it becomes necessary to account for the distributed effects at least in the inter-connects. This has been the approach adopted by the microwave monolithic integrated circuit (MMIC) designers coming from a microwave background. Perhaps the most practical example of this hybrid approach is where individual building blocks (e.g., amplifiers, mixers, oscillators, etc.) are design as lumped circuits with accurately modeled reactive elements. These blocks are connected to each other with transmission lines, as depicted in Fig. 3. These transmission lines are distributed circuits with dimensions comparable to the wavelengths of interest and are thus modeled as such. Nonethe-

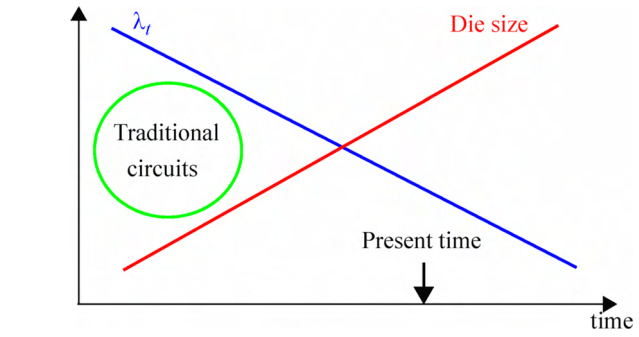

Fig. 1 Trends in transistor cut-off wavelengths and die sizes.

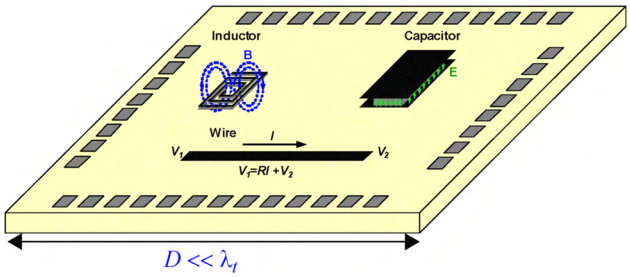

Fig. 2 Classical circuit view of a high frequency integrated circuit.

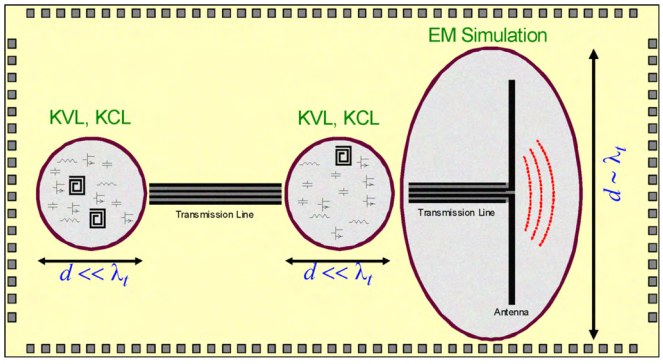

Fig. 3 The monolithic microwave integrated circuit (MMIC) approach to high frequency silicon integrated circuits.

1. This work has been supported in part by Lee Center for Advanced Networking. Parts of this work appeared in [25] and [26] previously.

2. The cut-off wavelength is defined as $\lambda_t = c/f_t$, where f_t is the unity current gain of the transistors and c is the speed of light in the medium.

978-1-4244-4353-6/09/$25.00 ©2009 IEEE

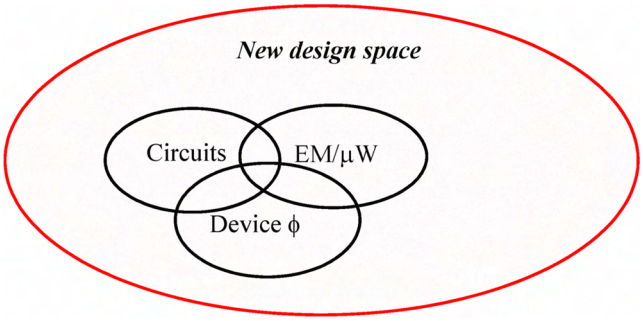

Fig. 4 Traditional boundaries of different disciplines and the new design space available with new co-integration approaches.

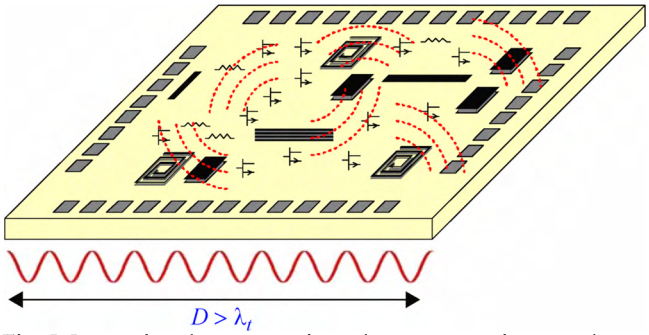

Fig. 5 Interactions between various elements on an integrated circuit above the size wavelength cross-over.

less, the individual building blocks are designed separately in a lumped fashion. Even if an antenna is present, it is usually design separately and connected to the rest of the circuits either directly or using a transmission line, as shown in Fig. 3. Although this hybrid approach can overcome some of the basic challenges of modeling these circuits at higher frequencies, it is still more of a patch to make working circuits and does cover a small subset of the design space. While essentially the only way to get working circuits up until now, there are several major disadvantages to such an approach. In addition to resulting in sub-optimal solutions due to the limited design space (Fig. 4), the major disadvantage of this methodology is that many new and potentially beneficial architectural variations are simply overlooked since they occur in the broader design space.

While this hybrid approach has produced some very interesting and useful results in the past [7]-[11], it is far from optimum in terms of what is achievable considering the remarkable plethora of new opportunities offered by a true combination of the practically unlimited number of on-chip active and passive elements operating at wavelengths comparable to or greater than their collective dimensions (Fig. 5).

To be able to take full advantage of the numerous opportunities offered by the die-size wavelength cross-over, the designers must unleash their imagination and retool for a new era of fundamentally different design methodology. This can be done by removing the artificial borders between different levels of abstraction, such as electromagnetics, antenna, propagation, device physics, as well as analog and digital circuit and system design. These levels of abstraction, which were establish to partition the overall system design into tractable tasks, are no longer suitable to take full advantage of the unparalleled new opportunities at mm-wave

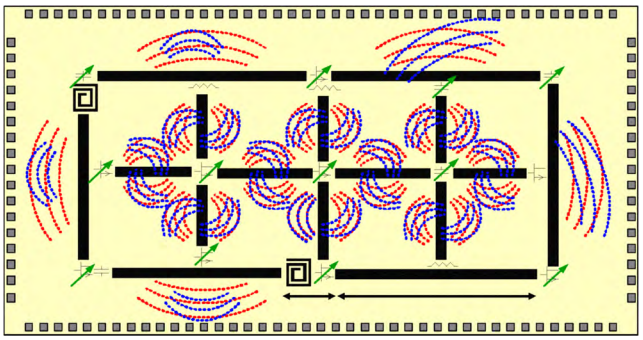

Fig. 6 The holistic design approach to harness the new opportunism offered by cross-disciplinary integration of electromagnetics, circuits, and devices on a single silicon chip.

frequencies and above created by the dimension-wavelength cross-over of Fig. 1.

In this new holistic design approach (Fig. 6), no artificial lines are drawn between electromagnetics, radiating elements (antennas), passive and active devices, analog and digital circuit design, and device physics. This larger design space enables the designer to consider all the possibilities and to devise truly novel architectures and topologies with better performance and new features not offered by the conventional approaches discussed earlier. While it is apparent how the holistic codesign approach of Fig. 6 is beneficial and can result in more optimum results, it is not at all obvious how, in general, such a system can be conceptualized, designed, verified, and tested. As is always the case, new opportunities bring new challenges along with them and as such provide relevant problems to the theorists and CAD tool developers. Nonetheless, this new holistic design approach has tremendous opportunities, as demonstrated by the example of such design approach given in subsequent sections.

Many different existing and new applications can benefit from such high-frequency complex circuits implemented in silicon. There are many sensing and ranging applications using mm-wave frequencies. For instance, low-cost, highly integrated short-, mid- and long-range high-resolution radars at 24GHz, 77GHz, and 94GHz, which can be used in a variety of applications such as automotive radar and have received a great deal of interest in the recent years [5][6][9]. Automotive radars can be used in a broad range of applications, e.g., early warning and brake priming, self-parking, global traffic control, low-visibility driving aid, autonomous cruise control, and collision avoidance.

On the communication side, in addition to the more traditional multi-gigabit data transmission at 60GHz, it is possible to create new direction-dependent modulation schemes such as near-field direct-antenna modulation [24] that make it possible to create fundamentally secure and power efficient data streams with the possibility of concurrent full-rate transmission in several directions.

In addition to the communication and ranging applications, mm-wave silicon-based circuits can find numerous applications in security and medical imaging as well as biochemical sensors. These applications provide a range of new possibilities, such as low-cost medical imaging, bio-molecular point-of-care diagnostics, and sequencing.

We will discuss some of the challenges faced by designers in the design of such systems with several practical examples in the rest of this paper.

Silicon Integration Challenges

Despite their numerous benefits, integration of mm-wave systems on a silicon substrate faces several challenges, some of which will be discussed in this section.

The conductive substrate of conventional silicon integrated circuits can cause energy loss due to magnetically induced eddy currents or electromagnetic radiation by intended (or unintended) on-chip antennas. This is compounded by the additional energy loss due to the finite conductivity of the metal structures used in passive devices. Even the ohmic energy loss can be substantially larger at mm-waves frequencies because of the small skin depth. For instance, copper has a skin depth of roughly 300nm at 60GHz. This problem is exacerbated in processes where chemical mechanical polishing (CMP) is used since it results in higher surface roughness, which combined with the small skin depth results in substantial energy loss in the system. The current handling capability of the metal lines is further degraded by the additional "fill" and "cheese" rules for wide metal strips [25].

The high dielectric constant of most semiconductor materials, such as silicon (*e.g.*, ε_r=11.7 for silicon) is another limiting factor. The dielectric waveguide formed by the rectangular silicon substrate can sustain undesirable propagating modes which depend on its dimensions. Many of these natural modes are in the mm-wave frequencies. This results in an alternative mechanism for energy loss for on-chip components. The energy leaks into the substrate modes and is dissipated by the substrate ohmic loss, or is radiated in undesirable directions [5][25]. This high-frequency electromagnetic energy can also easily couple back to other on-chip components creating possibly strong parasitic coupling between unrelated elements on the same die. This necessitates special attention to be paid to the overall design to avoid such coupling, utilizing techniques such as fully-differential signaling to minimize the generation and pick-up of high-frequency signals.

An important challenge to an efficient on-chip antenna with a well-defined radiation pattern is the high dielectric constant of the substrate which causes a considerable portion of the electromagnetic energy to be absorbed into the substrate instead of being radiated into the air [12][5]. The impact of this effect is exacerbated by the conductivity of the substrate in the case of silicon. As will be discussed further later in this section this is a very important issue whenever we try to integrate radiating elements, such as antennas, on a silicon die.

Power generation presents a very serious challenge in small feature size silicon technologies. This is mainly due to the lower transistor breakdown voltages resulting from the scaling process and the shrinking of the depletion regions in the transistors, which necessitates the use of a lower power supply voltage. Unfortunately, this is in direct conflict with the maximum power that can be generated using conventional power amplifier techniques, which makes it almost inevitable to use parallel structures and novel power combing approaches, as described in [13][14] and [28]-[31]. These approaches in turn are more susceptible to different energy loss mechanism in passive devices that were discussed earlier.

The modeling of the transistors themselves (as well as the passive devices) becomes more challenging at these higher frequencies. Not only the smaller parasitic components within the models are more prone to error and hence special attention must be paid to guarantee reliable results, but also the impact of transistor-to-transistor variations is felt more at the smaller dimensions necessary for higher frequency operation.

The high frequency chip interface to the outside presents another challenge, which does not scale well with frequency. External connections to the chip need to be rethought completely. For instance, even if we were to connect the mm-wave input (or output) of the chip to a perfectly matched transmission line on an adjacent printed circuit board using a wirebond, it would present an impedance of $12\Omega+j360\Omega$ at 60GHz (modeling a 1mm long wirebond as a 1nH inductor with a Q of 30). This makes it extremely difficult to achieve a proper impedance match using conventional approaches. Additionally, such an element can easily turn into an unintended radiative element at mm-wave frequencies, creating very serious problems at the system level. Even a solder ball in flip-chip technology has a non-trivial impedance to match. If we were to make the interface with wirebond, every effort should be made to make it into a structure emulating a transmission line, as in [9]. On the other hand, these interface challenges make the totally-integrated holistic solutions similar to that concept shown in Fig. 6 much more suitable for these applications by completely eliminating the high frequency electrical interface to the outside world and replacing it with an electromagnetic radiating interface, where the chip is the radiating element itself.

As clear from this discussion, the direct application of lumped (Fig. 2) or monolithic microwave integrated circuits (Fig. 3) approaches to design of silicon-based integrated circuits is not optimum. The direct generalization of lumped approach would not be particularly successful, as in the absence of proper treatment of the complex electromagnetics issues involved in such systems, the designers will not be able to come up with an implementation that closely matches the expected results. Also, the classical microwave integrated circuit techniques are not highly scalable since they are based on the underlying assumptions of a small number of well-characterized transistors and highly specialized interconnects. Thus, it is clear that to take full advantage of this significant opportunity, one needs to start at the system architecture level and devise new holistic methodologies and topologies.

Integrated Radiating Elements

At the most basic level, on-chip antennas make it possible for the input and output signals to be radiated directly into and out of the chip with no need for an electrical interface at mm-wave frequencies. In the more general scheme of Fig. 6, the ability to design radiating elements with desirable characteristics and predictable behavior is of central importance. While we focus on the on-chip antenna design challenges in this section, it should be noted that many of the challenges faced in the design of on-chip radiating elements are exemplified in the special case of on-chip antennas.

Successful implementation of an on-chip antenna with acceptable gain and efficiency is a non-trivial task due to the high dielectric constant and conductivity of the substrate. To elucidate these challenges, consider a dipole antenna placed at the interface of air and a semi-infinite region of a dielectric material with a relative permittivity of ε_r, as shown in Fig. 7. The dielectric presents a lower electromagnetic impedance than the air and as a result more of the radiated electromagnetic power goes into the dielectric. The percentage of the power going in each direction as a function of the ε_r is shown in Fig. 8 [12][5]. As can be seen, for

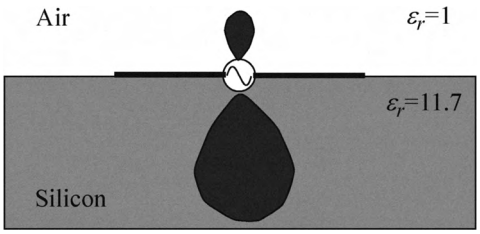

Fig. 7 Dipole antenna at the air-silicon boundary

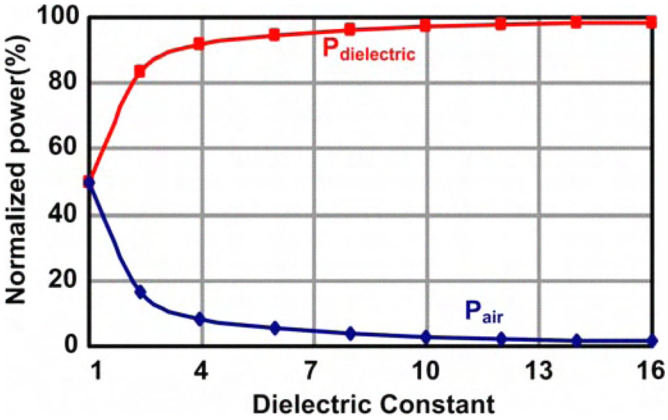

Fig. 8 Percentage of the radiated power into the air and dielectric vs. relative permittivity of the dielectric for a dipole antenna at the boundary [5].

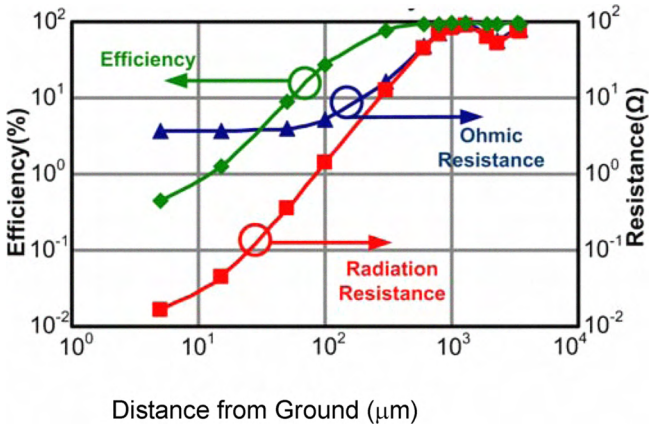

Fig. 9 The efficiency, radiation and loss resistance of a dipole vs. spacing from the ground plane.

silicon with an ε_r of approximately 11.7, more than 95% of the power will end up going into the silicon, as opposed to 5% for the air.

To avoid this problem, one may decide to use a ground plane in the lowest metal layer (*e.g.*, M1) while placing the antenna in the top-most layer. While this may sound appealing initially, it is not very practical due to the typical spacing between the top and bottom metal layers (*e.g.*, in the 10-15μm range in today's technologies). If one were to attempt this approach, the resulting antenna would have such a small radiation resistance due to the close proximity of the ground plane that its efficiency would be extremely low (e.g., <1%). The trade-off between these different parameters is shown in Fig. 9 [5].

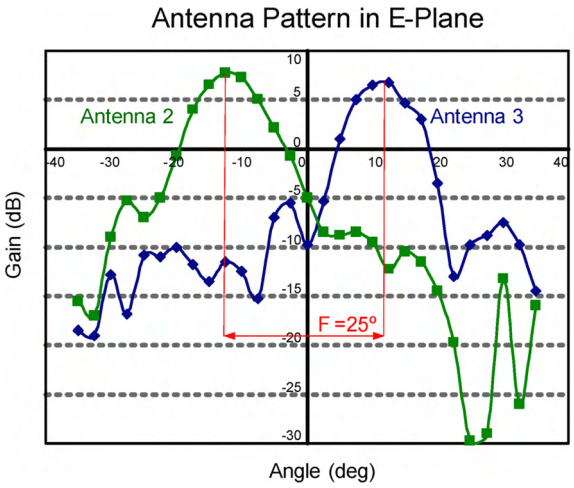

Fig. 10 Measured patterns of two adjacent antennas on a silicon chip with a backside lens.

Placing the ground plane directly underneath the substrate does not improve this. Although the spacing may be high enough for the radiation resistance to become acceptably large, the entire substrate will act as a waveguide with multiple propagation modes, some of which will typically fall very close to the frequency of interest. The energy does get coupled into these modes, where part of it is lost to heat in the conductive substrate and the remainder is radiated in undesirable directions.

One approach to solve this problem is to take advantage of the energy coupling into the substrate and radiating it from the backside of the chip. To do this, we must suppress the substrate modes and direct the energy into the desired radiative mode. This can be achieved by changing the "shape" of the substrate electromagnetically. One way to do this is to use a dielectric lens [12] on the backside of the chip which couples the EM energy into a dominant radiative mode [15]. In one example implementation, to reduce the antenna metal loss, three bottom layers are connected in parallel with vias to form the dipole antennas as close to the substrate as possible. To reduce the substrate loss, the fabricated chip is thinned down to 100μm. Due to the layout constraints, antennas are placed at the chip edge and to maintain a uniform dielectric constant substrate underneath the antennas, an undoped silicon slab with the same thickness of silicon chip is abutted to the chip. For mechanical stability, a 500μm silicon wafer is used to hold the chip and the silicon lens is attached to the backside of the undoped wafer. A 2-axis spherical far field measurement technique is utilized to measure the radiation pattern while a W-band horn antenna is used to irradiate the integrated dipoles. The 3-D measured patterns of two middle antennas are shown in Fig. 10. A maximum peak gain of about +8dBi is achieved in this measurement. As seen in Fig. 10, the peaks of two antennas occur at the two different directions. The chip requires no high frequency electrical connection to the outside world.

Near-Field Digital Antenna Modulation

Conventional radio transmitters modulate the desired information at the baseband and then up-convert and amplify that information by modulating the desired signal to create the desired phase and amplitude for the symbols transmitted. This requires upconversion mixers and perhaps more importantly a linear (or linearized) PA for non-constant-envelope modulations. The modulated signal

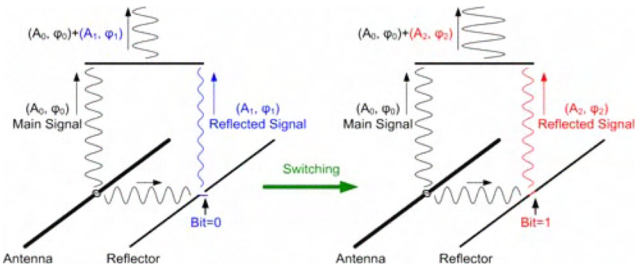

Fig. 11 Single bit modulation of the transmitted signal by digital switching the reflector in the near field.

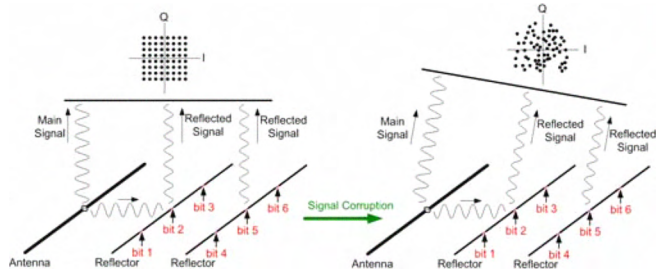

Fig. 12 Multiple reflectors with multiple switches to create different constellation points in a) the intended direction and b) in the unintended one.

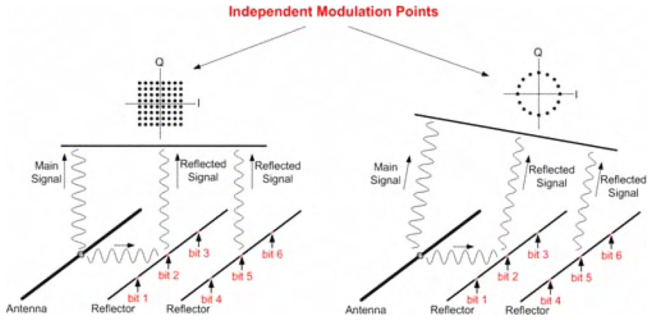

Fig. 13 Multiple reflectors with multiple switches creating different constellation points in different directions.

is fed to an antenna that radiates the modulated signal in all directions, albeit with different gain factors and time delays. Even in the case of a directional antenna, although a stronger signal is sent in the desired direction, essentially the same data constellation and information is sent in other directions that can be intercepted by an unintended receiver with good sensitivity.

The near-field digital antenna modulation technique is a fundamentally different approach to modulating and transmitting data that combines digital circuitry and electromagnetics to modulate the signal directly at the antenna through digital manipulation of the electromagnetic boundary conditions [24]. The digital antenna modulation makes it possible to simplify the rest of the transmitter by eliminating the need for upconversion mixers and linear power amplifiers. It can be used to create a secure communication channel by sending the desired data only in the intended direction, while a scrambled data stream is transmitted in other directions. Also, it can be used to *concurrently* transmit two (or multiple) streams of completely independent data in different directions both at the maximum rate, increasing the transmitted data rate significantly. As we will see next its very simple building block allows it to be used for transmitting data at very high rates not limited by the bandwidth of the upconversion chain.

Fig. 11 shows the basis principle behind the digital antenna modulation technique, where an a dipole antenna in this case is driven by a CW signal of constant amplitude and phase. There is also a conductive metal line with comparable dimensions to the wavelength (a reflector) next to the antenna which can be shorted or opened at some point along its length using a switch. The reflected signal interferes with the main signal radiated by the antenna in a given direction. The amplitude and phase of the reflected signal depend on the boundary conditions that the reflector imposes and can be varied by turning the digital switch on or off. The two states of the switch result in two different phases and amplitudes in the direction of interest hence generating two distinct points in the I-Q plane. This provides a simple and crude one-bit digital modulation without changing the output power or phase of the PA driving the antenna hence allowing it to operate at its highest efficiency.

Although the single-reflector single-switch configuration of Fig. 11 provides a basic binary modulation, it offers limited control over where the two modulation points fall. More control over the constellation point (i.e., transmitted phase and amplitude) can be obtained by introducing multiple reflectors each with multiple switches, as shown in Fig. 12a. A large enough number of reflectors and switches in close proximity of the antenna makes it possible to create more constellation points and eventually achieve full coverage across the constellation.

It is noteworthy that a set of switch combinations that generates constellation points of the given modulation in the intended direction generates different constellation points in a different direction, as shown in Fig. 12b. This can be used to scramble the signal further in the unintended directions to implement a secure communication link by preventing an undesired eavesdropper from demodulating and recovering the signal.

For a large number of switches and reflectors, there are numerous different switch combinations that generate almost the same constellation point in a given direction. However, these different switch combinations can be used to generate distinct points in other directions simultaneously, allowing concurrent transmission of two independent streams of data at full rate in two different directions, as depicted in Fig. 13.

The viability of this technique is demonstrated via an integrated circuit implementation [24] with an on-chip dipole antenna with 10 reflectors (5 on each side), each with 9 tuned MOS switches along its length, resulting in a total number of 90 switches and a extremely large 2^{90} ($\sim 10^{27}$) switching combinations, as illustrated in Fig. 14. The very large number of switching combinations provides numerous ways to generate a desired point on the constellation (phase and amplitude) in a given direction, providing many additional degrees of freedom that can be used for concurrency or security. The ability to simultaneously send independent information to several directions at full rate using a single transmitter is not achievable using the conventional transmitter architectures.

It is self-evident at this point that the ability to switch the antenna reflectors in the near field using a large number of very fast switches is a direct result of our ability to integrate, antennas, RF, and digital circuits on the same substrate. This is an example of the holistic co-design of electromagnetic structures with analog

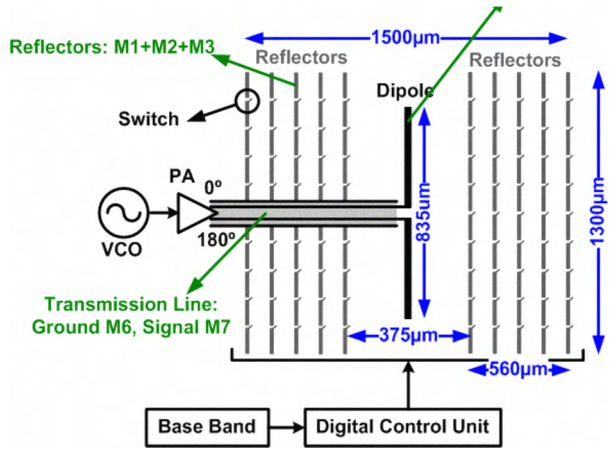

Fig. 14 The practical implementation of the digital antenna modulation scheme.

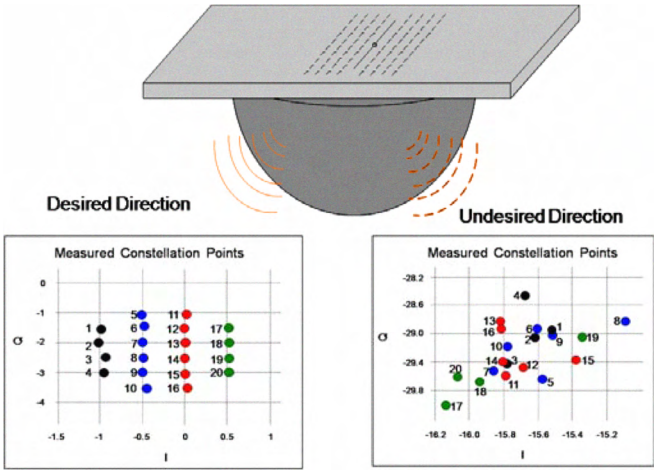

Fig. 15 The direct antenna modulation chip with a dielectric lens; a sample of the measured constellation points in the intended and unintended directions.

and digital circuitry. In practice, designing such system pose challenges not only at the architecture and circuit levels, but also with respect to the simulation tools and methodology.

Making a high quality and fast integrated switch for use in mm-wave frequencies is challenging. Achieving a small on-impedance requires a relatively large transistor size, which itself limits the maximum achievable off-impedance due to its large parasitic capacitance. This can be alleviated by resonating the transistor parasitic capacitance using a transmission line connected between its drain and its source.

A transmitter at 60GHz using the direct antenna modulation scheme with an on-chip dipole antenna and reflectors is implemented in silicon. The antenna, the ten reflectors, the ninety switches, and the digital control circuitry for the switches occupy a die area of roughly 1.5mm x 1.5mm. The power is radiated from the backside of the chip with the aid of a hemispherical dielectric lens [5], as shown in Fig. 15.

An experimental demonstration of this technique is also shown in Fig. 15, where we show 20 points generated purely by switching the reflectors using 20 different switch combinations on the left hand side of Fig. 15. It should be noted that the input power and

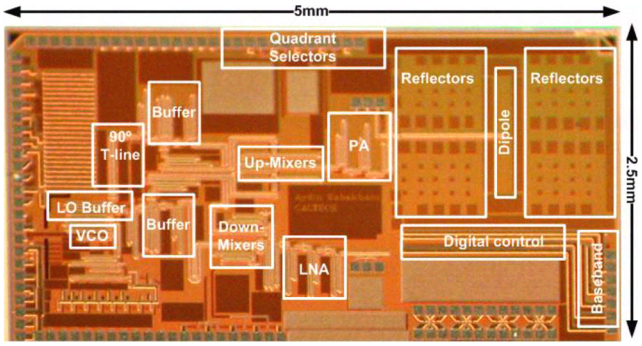

Fig. 16 The die micrograph of the direct antenna modulation transmitter.

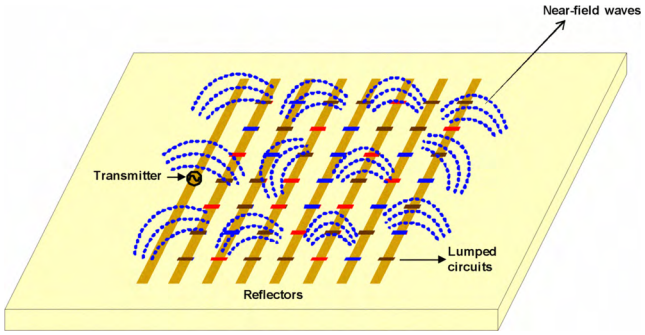

Fig. 17 A generalization of the near-field direct antenna modulation where the ports are connected to variable impedance where the near- and by extension the far field patterns can be changed continuously.

phase of the signal driving the antenna has not been changed to achieve the modulation. In addition, on the right side of Fig. 15, the received signal constellation at a different angle (about 90° off) for the same set of switch combinations is plotted. It is obvious that these points are completely scrambled, making it practically impossible for a receiver at the undesirable angle to recover this signal, no matter how sensitive it may be. This is in contrast with a traditional transmitter that sends essentially the same modulated signal, albeit with different amplitude scales and phase epochs, in all directions. This measurement demonstrates the feasibility of the idea for security of communication links. The antenna is driven by a three stage fully-differential PA with an output power of at least +7dBm. The die photo is shown in Fig. 16.

This concept can be generalized to the case, where instead of switches we use variable impedances at the ports using lumped circuit, as shown in Fig. 17. In this case, it is possible to change the near field electromagnetic profile in a continuous fashion, making it possible to produce any desired far-field pattern.

Integrated Phased Arrays

Integration of a complete phased array system in silicon results in substantial improvements in cost, size, and reliability. It also offers various opportunities to perform on-chip signal processing and conditioning without having to go off-chip, leading to additional savings in cost and power. The multiple signal paths, operating in harmony, provide benefits at the system and circuit level.

Multiple antenna phased-arrays imitate a directional antenna whose bearing can be controlled electronically [16]-[22]. This electronic steering makes it possible to emulate antenna proper-

ties such as gain and directionality, while eliminating the need for continuous mechanical reorientation of the actual antennas. Additionally, the parallel nature of a phased array antenna transceiver alleviates the power handling and noise requirements for individual active devices used in the array. This makes the system more robust to the failure of individual components. In the past, such systems have been implemented using a large number of microwave modules, adding to their cost and manufacturing complexity [20][21].

A phased-array transmitter or receiver consists of several signal paths each connected to a separate antenna. The antenna elements of the array can be arranged in different spatial configurations [18]. The array can be formed in one, two, or even three dimensions, with one or a two dimensional array being more common.

The principle of operation of a phased-array is similar for both receivers and transmitters. In a phased-array receiver, the radiated signal arrives at different times at each of the spatially separated antennas. The difference in the time of arrival of the signal at different antennas depends upon the angle of incidence and the spacing between the antennas. An ideal phased-array receiver compensates for the time delay difference between the signals from different antennas and combines the signals coherently to enhance the reception from the desired direction(s), while rejecting emissions from other directions. Similarly, in a phased-array transmitter, the signals in different elements are delayed by different amounts so that the outgoing signals add up coherently only in the desired direction(s). Incoherent addition of the signals in other directions results in lower radiated power in those directions. Thus in a phased-array based system, the transmitter generates less interference in receivers that are not targeted. Furthermore, the receiver is also capable of nulling out some interferers as long as they do not originate from the same direction as the signal. Additionally, for a given power level at the receiver, the power that has to be generated is lower in a phased-array transmitter than in an isotropic transmitter.

In a transmitter with n elements if each element radiates P Watts, the total power that will be seen at the receiver in the desired direction is n^2P Watts. The n^2 improvement comes from the coherent addition of the signals in amplitude in the desired direction. For example, in a four element transmitter, the total power radiated in the beam direction is 12dB higher than the power radiated by each element.

In receivers, the advantages of a phased array include better sensitivity and higher interference rejection capabilities. For a given receiver sensitivity, the output *SNR* sets an upper limit on the noise figure of the receiver. The noise figure, *NF*, is defined as the ratio of the total output noise power to the output noise power caused only by the source [23]. In an actual phased array receiver implementation, compared to the output *SNR* of a single-path receiver, the output SNR of the array is improved by a factor as high as n, depending on the noise and gain contribution of different stages [8][9]. For instance, if the noise from the antennas is uncorrelated, an 8-path phased-array can improve the receiver sensitivity by 9dB.

Thus, in a system based on phased arrays at the transmitter and receiver, the higher SNR and lower interference increases channel capacity. Furthermore, the directivity of the transmit-receive pairs can result in higher frequency reuse ratios, leading to higher network capacity.

For narrowband systems, the true-time delay necessary in each element of a phased-array can be approximated by a phase-shift.

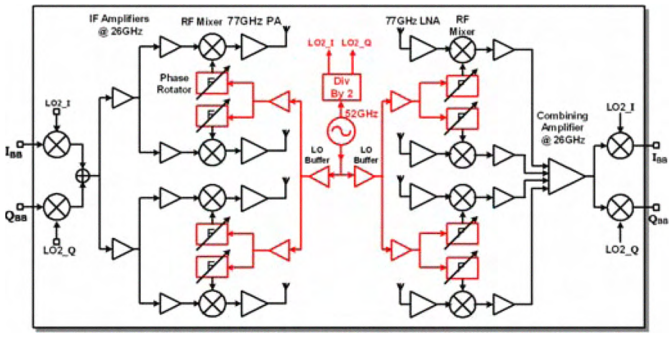

Fig. 18 The architecture of the fully-integrated 4x2 77GHz phased array transceiver with integrated antennas.

This approximation leads to some signal dispersion, due to the non-constant group delay, which increases as the bandwidth of the signal increases. This dispersion translates to a higher BER in communication systems and lower resolution in radar systems [8].

The phase shift necessary in each element of a phased-array can be achieved at RF, at baseband/IF or in the LO path. In integrated implementations, there are several advantages to using LO path phase shifting, as the gain in each element of the transmitter or receiver is less sensitive to the amplitude variations at the LO ports of the mixers [8][9].

The various phases of the LO necessary in the LO-phase shifting approach can be generated in a central fashion by generating all the necessary phases at one place (e.g, a multi-phase oscillator), as in [8][9], or in a decentralized fashion by distributing only one phase of the LO signal and generating the remaining phases locally using a phase rotator for each LO path, similar to [5][6].

A fully integrated transceiver 4-element phased array transceiver with on-chip antennas has been designed and fabricated in a 0.13μm SiGe BiCMOS process [5][6]. The receiver consists of the complete down-conversion path with low-noise amplifier (LNA), frequency synthesizer, phase rotators, combining amplifiers, and on-chip dipole antennas, as shown in As shown in Fig. 18, A distributed active combining amplifier at an IF of 26 GHz is used to perform the signal combining. A 52-GHz first LO is generated on chip and is routed to different elements, where it can be phase shifted locally by the phase rotators. The local LO-path phase-shifting scheme enables a robust distribution network that scales well with increasing frequency and number of elements while providing high-resolution phase shifts. Measurements indicate a single-element LNA gain of 23 dB and a noise figure of 6.0 dB. Each of the four receive paths has a gain of 37 dB and a single-path overall noise figure of 8.0 dB. Each on-chip antenna has a gain of +8 dBi. Each element of the 2-step upconversion transmitter generates +12.5 dBm of output power at 77 GHz with a bandwidth of 2.5 GHz leading to a 4-element effective isotropic radiated power (EIRP) of 24.5 dBm. Each on-chip PA has a maximum saturated power of +17.5 dBm at 77 GHz. The entire phased array transceiver occupies an area of 3.8mm x 6.8mm, as shown in the die photo of Fig. 19

Conclusions

For the first time, it is possible to have a very large number of extremely fast and reliable transistors capable of operating at mm-wave frequencies in close proximity of structures comparable to the wavelengths of interest on the same silicon substrate, where electromagnetics, analog, and digital circuits can be intertwined deeply. This unique combination makes is possible to con-

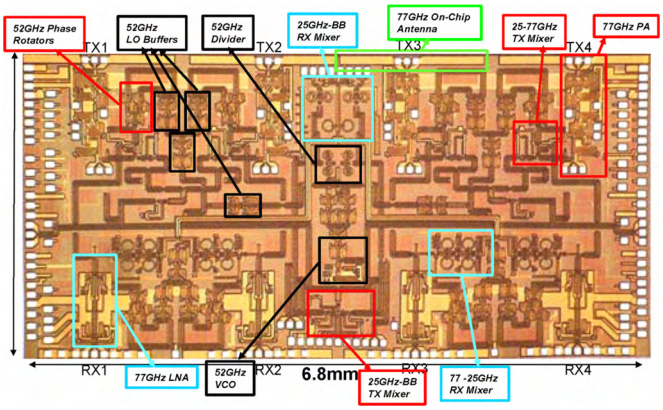

Fig. 19 Die photo of the phased-array transceiver with on-chip antennas

sider novel holistic approaches to communication and radar systems operating at mm-wave frequencies only limited by our imagination and creativity.

Acknowledgments

The author would like to particularly thank Dr. A. Babakhani, as well as F. Bohn, A. Chang, K. Sengupta, S. Bowers, Dr. H. Wang, Prof. D. B. Rutledge, Prof. S. Weinreb of Caltech and Dr. A. Natarajan, Dr. I. Aoki, Dr. S. Kee, Dr. A. Komijani, Dr. X. Guan, Dr. Y. Wang, Prof. H. Hashemi, Prof. J. Buckwalter, and Prof. E. Afshari formerly of Caltech for their numerous contribution to Caltech's mm-wave activities. We have benefitted from the support of Caltech's Lee Center for Advance Networking, Raytheon Company, National Science Foundation, and DARPA Trusted Foundry Program.

References

[1] J. S. Kilby, United States Patent Number 3,138,743, Filed Feb. 6, 1959, Granted June 23, 1964.

[2] J. S. Kilby, United States Patent Number 3,261,081, Filed Feb. 6, 1959, Granted July 19, 1966.

[3] R. F. Stewart, United States Patent Number 3,138,747, Filed Feb. 12, 1959, Granted June 23, 1964.

[4] G.E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, No. 8, pp. 114-117, April 1965.

[5] A Babakhani, X Guan, A Komijani, A Natarajan, and A. Hajimiri, "A 77-GHz Phased-Array Transceiver with On-Chip antennas in silicon: Receiver and Antennas," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2795-2806, Dec. 2006.

[6] A Natarajan, A Komijani, X Guan, A Babakhani, and A. Hajimiri, "A 77-GHz Phased-Array Transceiver with On-Chip antennas in silicon: Transmitter and Local LO-Path Phase Shifting," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2807-2819, Dec. 2006.

[7] X. Guan and A. Hajimiri, "A 24-GHz CMOS front-end," *IEEE J. Solid-State Circuits,* vol. 39, no. 2, pp. 368-373, Feb. 2003.

[8] A. Hajimiri, A. Komijani, A. Natarajan, X. Guan, H. Hashemi, "Phased array systems in silicon," *IEEE Communications Magazine*, vol. 42, Issue 8, pp. 122-130, Aug. 2004.

[9] H. Hashemi, X. Guan, A. Komijani, and A. Hajimiri, "A 24-GHz SiGe phased-array receiver - LO phase shifting approach," *IEEE Trans. Microwave Theory Tech,* vol. 53, Issue 2, pp. 614-626, Feb. 2005.

[10] C. Doan, S. Emami, A. Niknejad, and R. Brodersen, "Millimeter Wave CMOS Design," *IEEE J. Solid-State Circuits,* vol. 40, no. 1, pp. 144-155, Jan. 2005.

[11] B. Floyd, s. Reynolds, U. Pfeiffer, T. Zwick, T. Beukema, and B. Gaucher, "SiGe Bipolar Transceiver Circuits Operating at 60GHz," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 156-167, Jan. 2005.

[12] D. B. Rutledge, *et al*., "Integrated-circuit antennas," *Infrared and Millimeter-Waves*, K. J. Button, Ed. New York: Academic, 1983, vol. 10, pp. 1–90.

[13] I. Aoki, S.D. Kee, D.B. Rutledge, and A. Hajimiri, "Fully integrated CMOS power amplifier design using the distributed active-transformer architecture," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 371-383, March 2002.

[14] E. Afshari, H. Bhat, X. Li, A. Hajimiri, "Electrical funnel: A broadband signal combining method," *Dig. International Solid-State circuits Conference,* Feb. 2006.

[15] A. Babakhani and A. Hajimiri, "mm-Wave Phased Arrays in Silicon with Integrated Antennas," *Proc. Antenna and Propagation Symposium*, June 2007.

[16] V. Aulock, W. H., "Properties of phased arrays," *Proc. IRE*, vol. 48, No. 10, pp. 1715-1728, Oct. 1960.

[17] R. C. Hansen, Ed., *Significant Phased Array Papers*. Norwood, MA: Artech House, 1973.

[18] R. S. Elliott, *Antenna Theory and Design*. Englewood Cliffs, NJ: Prentice-Hall, 1981.

[19] M. Golio, Ed., *The RF and Microwave Handbook*, Session 6.9, CRC Press LLC, 2000.

[20] D. Parker, D. C. Zimmermann, "Phased arrays – part I: theory and architectures," *IEEE Trans. Microwave Theory Tech*, vol. 50, No. 3, pp 678-687, Mar. 2002.

[21] D. Parker, D. C. Zimmermann, "Phased-arrays – part II: implementations, applications, and future trends," *IEEE Trans. Microwave Theory Tech.,* vol. 50, No. 3, pp 688-698, Mar. 2002.

[22] B. Kane, L. Geis, M. Wyatt, D. Copeland, and J. Mogensen, "Smart Phased Array SoCs: A Novel Application for Advanced SiGe HBT BiCMOS Technology," *Proc. of IEEE.*

[23] "IRE standards on electron tubes: Definition of terms," *Proc. IRE*, vol. 45, pp. 983-1010, July 1957. 57 IRE 7. S2.

[24] A. Babakhani, D. B. Rutledge, and A. Hajimiri, "A Near-Field Modulation Technique Using Antenna Reflector Switching," to appear in *ISSCC Digest of Technical Papers*, Feb. 2008.

[25] A. Hajimiri, "mm-Wave Silicon ICs: Challenges and Opportunities," *IEEE Custom Integrated Circuits Conference*, Sept. 2007.

[26] A. Hajimiri, "Holistic Design in mm-Wave Silicon ICs," *IEICE Transactions on Electronics* E91-C(6):817-828, 2008.

[27] L. Zhu, "Guided-Wave Characteristics of Periodic Coplanar Waveguides with Inductive Loading: Unit-Length Transmission Parameters," IEEE Trans. on Microwave Theory and Techniques, vol. 51, no. 10, pp. 2133-2138, Oct. 2003.

[28] A. Komijani, A. Natarajan, and A. Hajimiri, "A 24GHz, +14.5dBm Fully-Integrated Power Amplifier in 0.18µm CMOS," *IEEE J. of Solid-State Circuits*, Vol. 40, no. 9, pp 1901-8, Sept. 2005.

[29] A. Komijani and A. Hajimiri, "A Wideband 77GHz, 17.5dBm Power Amplifier in Silicon," *IEEE J. of Solid-State Circuits*, Vol. 40, no. 9, pp 1901-8, Sept. 2005.

[30] E. Afshari, H. S. Bhat, A. Hajimiri, and J. E. Marsden, "Extremely Wideband Signal Shaping Using One-and Two-Dimensional Nonuniform Nonlinear Transmission Lines," *Journal of Applied Physics*, vol. 99, 2006.

[31] I. Aoki, S. Kee, R. Magoon, R. Aparicio, F. Bohn, J. Zachan, G. Hatcher, D. McClymont, and A. Hajimiri, "A Fully-Integrated Quad-Band GSM/GPRS CMOS Power Amplifier," *ISSCC Digest of Technical Papers*, Feb. 2008.