

Nanoscale CMOS for mm-Wave Applications

Ali M. Niknejad, S. Emami, B. Heydari, M. Bohsali, E. Adabi
Berkeley Wireless Research Center
University of California at Berkeley, Berkeley, CA 94720

(Invited Paper)

Abstract—Aggressive technology scaling of CMOS has culminated in a low-cost high volume commercial process technology with $F_t > 150$ GHz and $F_{max} > 200$ GHz. This paper discusses the key trends in CMOS scaling that have led to this level of performance and attempts to predict the performance down to 45 nm. The design of active and passive components in CMOS for power gain and low noise are discussed in detail and unique features of CMOS technology are highlighted. Experimental results derived from a 60 GHz amplifier in 90nm CMOS and a complete 60 GHz front-end receiver in 130nm CMOS are reported.

I. INTRODUCTION

A concerted effort by academia, industry, and worldwide research organizations has resulted in aggressive and steady technology scaling of CMOS technology. This scaling was fueled by the demand for digital computation and memory and thus the technology has primarily evolved to serve these markets. Even though specialized versions of bulk CMOS processes have been adapted to serve the communication sector for RF and microwave applications, the core devices are still digital in character. Silicon technology has been favored for scaling due to the simplicity of fabrication, which results in part from the native SiO_2 oxide and self-aligned gate formation. Scaling was initially guided by Dennard's Law [1], which called for constant field scaling, which in turn required lowering the supply voltage and also scaling the gate oxide thickness. Eventually constant field scaling was violated and high field effects, such as velocity saturation, led to additional innovations such as non-uniform doping (halo implants) and lightly doped drain contacts extensions (LDD). Despite many dire predictions about the scaling of bulk CMOS technology, the shrinking has continued down to the 45 nm technology node.

But today's CMOS is not "pure" in the traditional sense as the technology has evolved from requiring a handful of materials (Si, O, Al, and a few dopants and rare earth metals) fabricated using optical lithographic techniques to a sophisticated process using a wide variety of materials, high-K gate dielectrics, metal gates, low-K interconnect dielectrics, seven or more metal layers incorporating Cu, and various technologies that introduce strain into the channel for enhanced mobility (SiGe drain/source or capping layers). Today devices with $L_g < 32$ nm (sub optical) are routinely fabricated for high volume production. Beyond this gate length, new device structures have been proposed, such as multi-gate (FinFET) and SOI transistors. In this paper we limit focus to the bulk CMOS transistor with experimental results down to the 90nm node.

The technology scaling of the channel gate length L_g has resulted in raw performance benefits, particularly in the speed of the transistor. But a big penalty has been paid in other performance metrics, in particular the device output conductance g_o which translates into low intrinsic gain ($A_v = g_m/g_o \sim 10$) and extremely low supply voltages. Fortunately the supply scaling has stopped at 1V, but even at this level there are severe limitations in the achievable dynamic range of amplifiers and other key building blocks. Naturally this also translates into poor output power capability. Finally, in all technology nodes the bulk substrate is conductive which results in higher losses in passive components such as inductors and transmission lines. Fortunately in many mixed-mode process nodes the substrate is only moderately conductive ($\rho \sim 10\Omega\text{-cm}$) to minimize substrate coupling, which is quite tolerable in many applications. Despite these shortcomings, CMOS technology has transitioned from a low performance digital process to the prevalent technology. The low gate leakage has also been a key for realizing switched capacitor analog circuits and CMOS is the technology of choice for mixed-signal circuits. In the RF regime CMOS has proved viable as a low cost alternative to SiGe and GaAs for consumer applications up to 5GHz. Niche applications such as power amplifiers still favor technologies with higher breakdown voltages but cost considerations often dominate performance in consumer applications.

In our research project we have explored CMOS for mm-wave frequencies up to 100GHz. This paper chronicles this effort beginning with the performance of active and passive components, and culminating in the discussion of a complete front-end receiver operating at 60GHz.

II. CMOS MM-WAVE ACTIVE DEVICES

NMOS devices are naturally preferred to PMOS devices due to the higher mobility. To realize the largest possible gain, devices are biased in strong inversion. To assess the capability of CMOS, we analyze key performance metrics for mm-wave applications.

A. F_t and F_{max} Scaling

In analog and digital applications it is common to quote the process F_t , the unity gain current frequency, as a performance metric. For analog circuits, this limit represents the gain bandwidth product for a single stage amplifier. Digital circuits typically operate up to a fraction of F_t , and maximization of this metric has dominated process and technology engineers. But for mm-wave frequencies, the device F_t is only part of the story. In reality, we would like to find the activity limits of the

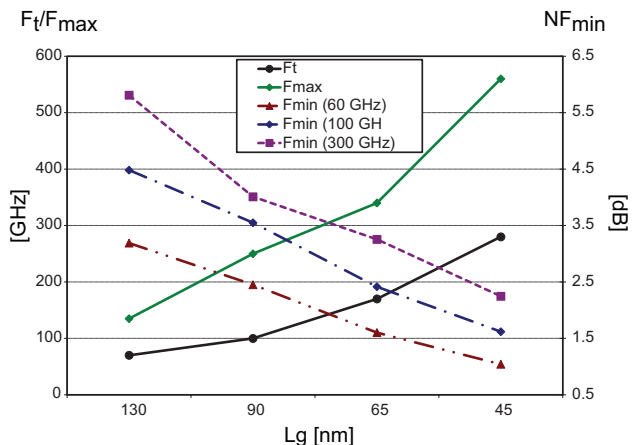


Fig. 1. The speed F_t/F_{max} and noise scaling trends of CMOS.

technology (F_{max}) or the maximum frequency of oscillation. While F_t is mainly a function of bias and technology, the device F_{max} is highly layout sensitive and determined by the device loss, such as the gate resistance, source/drain resistance, and substrate parasitics. It can be shown that the device F_{max} is predicted by [7]

$$F_{max} \approx \frac{F_t}{2\sqrt{R_g(g_m C_{gd}/C_{gg}) + (R_g + r_{ch} + R_S)g_{ds}}}$$

which shows that F_{max} can be larger than or smaller than F_t . We have found that proper layout in bulk CMOS technology allows one to realize $F_{max} \approx 2 \sim 3F_t$. A simple multi-finger layout is used to minimize gate and substrate resistances and ample contacts are employed to minimize the drain/source resistance. The multi-finger layout also results in lower drain/source area due to junction sharing.

Even though F_t continues to improve, as shown in Fig. 1, F_{max} does not necessarily improve. Source/drain junction depths are decreasing to minimize “drain” control of the channel, and interconnect metal thickness is decreasing as well. This means that interconnect metal and contact resistance (vias) are increasing and often the extrinsic device parasitics, as opposed to the intrinsic device characteristics, can dominate the performance. The limit to the finger width of the unit transistor is set by line edge roughness limits imposed by lithography, which is about 0.5μ . In Fig. 1 we have made some optimistic predictions and $F_{max} > F_t$ is maintained with scaling. The ITRS guidelines are used for F_t [2] whereas F_{max} is predicted based on measurements up to 90nm and calculated as a factor of 2 times larger than F_t for higher frequencies. The likely introduction of a metal gate at 45nm should help to offset the additional sources of losses with scaling.

To reduce important external parasitics, the array layout shown in Fig. 2, dubbed the “round table” layout, is preferred over a traditional multi-finger FET device [3]. Each sub-element consists of 10 fingers of a double-contacted 1μ width device. These elements are connected as shown using a multitude of contacts at the gate and source to minimize parasitics. The measured and de-embedded s-parameters and data-fit small-signal model of the device clearly demonstrate

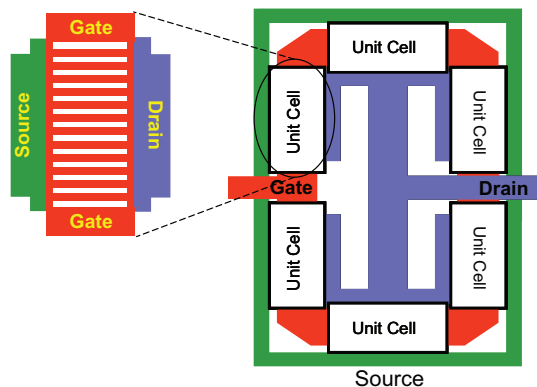


Fig. 2. Layout of an NMOS FET in a “round table” array.

an improvement in the gain of the device. Compared to the standard layout, an MSG improvement of 1dB is obtained. The extrapolated F_{max} of the device is 300 GHz, whereas the actual cross-over frequency, predicted by an accurate small-signal model, occurs at 200 GHz, substantially higher than the F_{max} of the standard layout device (140 GHz).

B. Minimum Achievable Noise

As technology has advanced, the CMOS technology minimum noise figure F_{min} has dropped significantly, approximately like (f/F_t)

$$F_{min} = 1 + 2 \left(\frac{f}{F_t} \right) \sqrt{g_m R_g \frac{\gamma}{\alpha}}$$

The above is calculated based on Pospiezalski’s noise model [4], which has resulted in a good match with CMOS devices in mm-wave frequencies. In the limit that we use short transistor fingers to minimize the gate resistance R_g , the noise is bounded by the NQS resistance seen at the gate $R_g = 1/5g_m$, which makes the fundamental noise just dependent on the technology parameters

$$F_{min} > 1 + 2 \left(\frac{f}{F_t} \right) \sqrt{\frac{1}{5} \frac{\gamma}{\alpha}}$$

Based on this simplistic prediction, the F_{min} is as low as 3.72 dB at 300 GHz in the 45nm node (Fig. 1). Meanwhile, no commercial or research results have demonstrated low noise CMOS amplifiers approaching the F_{min} of the device above 24 GHz. Our recent research efforts at 30 GHz has produced a functional 90nm LNA with a noise figure of 4.2 dB [5], somewhat competitive with some of the best technologies available today. Measurement results at mm-wave frequencies (60 GHz) on the 130 nm transistor also reveal that the minimum achievable noise figure is between 3-4 dB, and 90nm technology appears to be about 1 dB lower. This level of performance is sufficient for many applications up to 100 GHz. The remaining challenge is to build an amplifier that can actually match for low noise.

III. SILICON MM-WAVE PASSIVE DEVICES

A. Inductors and Transformers

Inductors are typically realized as planar spirals in the RF frequency range. These structures are easily scaled into the mm-wave frequency range where a single loop “ring” can provide enough reactance (~ 100 pH) with measured quality factors approaching 15-20 at 60 GHz. We can also view the ring inductor as shorted variable spacing differential transmission line. A ring inductor based resonator (150 pH inductor resonating with MIM capacitor) has been measured. The resonator occupies an area of $150^2 \mu\text{m}^2$ and has a loaded measured quality factor over 20.

Transformers are also easily implemented using planar lateral coupling or with 3D structures taking advantage of the many available metal layers. Transformers with insertion loss below 1 dB are possible (1:1 turn ratio) which conveniently provide dc isolation, biasing, and reactance for impedance matching. Transformers are especially desirable in fully differential circuits for these stated reasons. The area of a transformer is comparable to the area of a coupling capacitor at these frequencies.

B. Capacitors

Many CMOS processes offer high density MIM capacitors. But the availability of six or more metal layers in a typical 90nm process means that very high density structures can be realized with the metallization at no added cost. Many structures are possible but a simple “finger” comb using all available high density metal layers strapped in shunt is a typical approach. Ample vias are employed to improve the sidewall capacitance. These capacitors are effective as bypass capacitors as long as the finger lengths are kept short to increase the self-resonant frequency. For supply/ground bypassing, the inductance of the structure can be lowered by using a solid ground shield.

MOS varactors are also commonly employed for bypass and as VCO tuning capacitors in the RF frequencies. At mm-wave frequencies they have more limited utility due to the low Q (3-5) and low self-resonant frequencies. The lack of a high quality variable tuning element is a severe limitation of mm-wave CMOS which favors an architecture employing a lower frequency VCO and frequency multipliers.

C. Transmission Lines

An ideal T-line is inherently scalable in length and it fully characterized by Z_0 , α , and β . The dominant quasi-TEM mode can be modeled as a simple electrical lossy transmission line. Co-planar lines are preferred over microstrip lines due to higher obtainable inductive quality factors. This is important since transmission lines are used extensively to resonate MOS capacitors. Furthermore, since lateral dimensions are determined by lithography, the characteristic impedance of the co-planar line is more predictable and constant over process. The layout of the T-line must inherently suppress higher order and undesired modes from propagation. This is easy in a modern CMOS process with “bridges” that connect the coplanar ground planes to suppress unwanted modes.

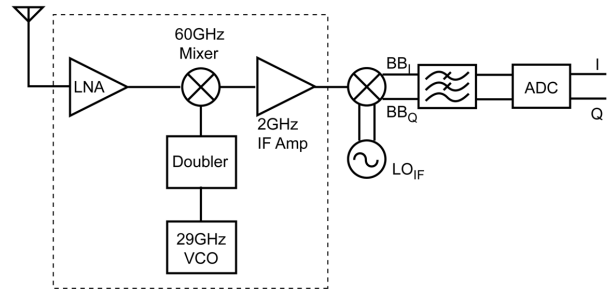


Fig. 4. Block diagram of an integrated 60GHz receiver.

Short sections of transmission lines can be used as inductors. Compared to spiral inductors, the reactance is more predictable, less influenced by surrounding structures, and provides better isolation. Inductive quality factors as high as 19 have been measured for coplanar lines. The inherent two-dimensional layout and scalability of transmission lines also simplifies simulation and compact modeling.

IV. EXPERIMENTAL RESULTS

A. 60 GHz Amplifier

The 40μ round table device forms the core of a two-stage 60 GHz amplifier shown schematically in Fig. 3 [3]. Since the device is only conditionally stable, the matching network is carefully designed to produce an unconditionally stable amplifier matched to 50Ω . Co-planar transmission lines are used extensively in the matching and interconnection networks. Custom models are employed for all active and passive devices. The chip is fabricated in a digital CMOS process and no analog options were employed. The amplifier has 12 dB of gain while consuming only 10.4mW of power from a 1V supply. The measured output 1-dB compression is +4 dBm, resulting in a power efficiency of 24%, making it suitable as a pre-driver or PA. The simulated noise figure is 6dB based on extracted noise models from measurements. The overall power consumption is low compared to previously reported 60GHz amplifiers [7] [6].

B. 60 GHz Front-End Receiver

A highly-integrated 60 GHz CMOS front-end receiver consisting of an LNA, a quadrature balanced down-converting mixer, a 30 GHz VCO, and a frequency doubler was fabricated in a 130-nm standard digital CMOS technology (Fig. 4). Transit frequencies (F_t) and maximum frequencies of oscillation (F_{max}) of 85 GHz and 135 GHz, respectively, were measured for NMOS devices in this process [7]. To predict the mm-wave behavior of CMOS transistors, a large-signal BSIM3 transistor model with external parasitics was extracted prior to the design of the circuits [8].

Fig. 5 depicts the die photo of the integrated front end. The chip area is about 7mm^2 including pads. On-wafer measurements were performed using a Cascade Microtech probe station. An Anritsu 37397C VNA was used for S-parameter measurements. The input return loss at the RF port is better than 15dB. The front-end has a conversion gain of 11.8dB at

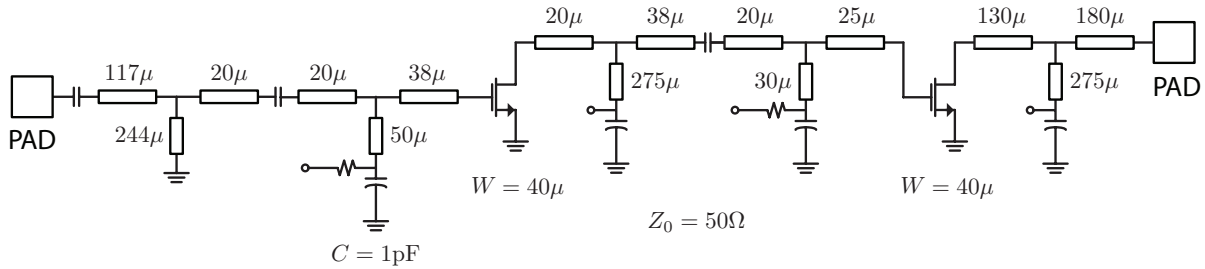


Fig. 3. Schematic of 60GHz amplifier.

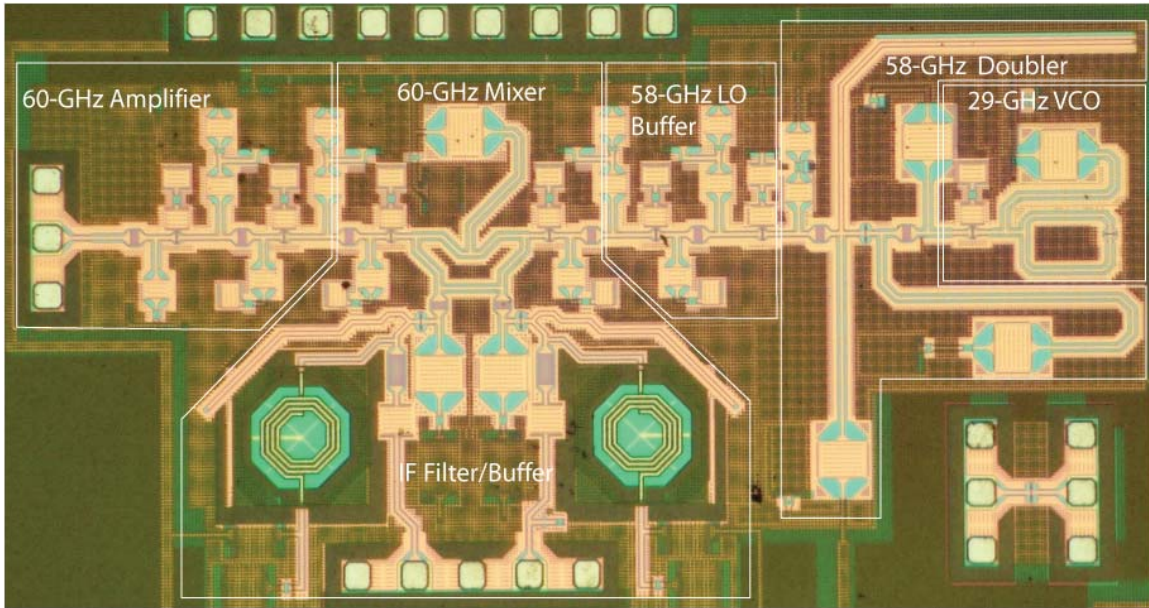


Fig. 5. Die photo of integrated 60GHz receiver.

60 GHz. The RF and LO frequencies were varied to obtain the frequency characteristics of the mixer for fixed IF frequency of 2 GHz. The measured input-referred 1-dB compression point of the front-end is -15.8 dBm at 60 GHz. Noise figure (NF) measurement of the integrated front-end was performed using a Millitech WR-15 noise source, WR-15 waveguide probes, and an Agilent N8973A NF measurement system. The measured NF of the downconverter is 10.4dB at 60 GHz. The total power dissipation of the integrated front-end is 64mA from a 1.2V supply.

V. CONCLUSION

In this paper we have shown that CMOS technology is capable of operation in the mm-wave band. Despite the speed advantage of SiGe, reported CMOS amplifiers and receivers are competitive in terms of power consumption and performance. Key passive components such as inductors, capacitors, resonators and transmission lines are easily integrated into the technology with reasonable quality factors and small footprints. Given the continued scaling of the technology, there is no doubt that researchers will continue to push CMOS mm-wave performance, making it not only a viable alternative but rather the preferred low-cost mm-wave technology.

VI. ACKNOWLEDGEMENT

The authors thank STMicroelectronics for chip fabrication. This work was funded by the DARPA TEAM program and co-sponsored by the UC-Micro program and BWRC member companies.

REFERENCES

- [1] R. Dennard, et al., "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE Journal of Solid State Circuits*, vol. SC-9, no. 5, pp. 256-268, Oct. 1974.
- [2] *International Technology Roadmap for Semiconductors*, 2005 Edition.
- [3] B. Heydari, M. Bohsali, E. Adabi, A. M. Niknejad, "Low Power High Performance mm-Wave Components up to 104 GHz in 90nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 200-201, 597, Feb. 2007.
- [4] M. W. Pospieszalski, "On the measurement of noise parameters of microwave two-ports," *IEEE MTT-S International Microwave Symposium Digest*, p.456-458, April 1986.
- [5] E. Adabi, B. Heydari, M. Bohsali and A. M. Niknejad, "30 GHz CMOS Low Noise Amplifier," to appear at *RFIC 2007*.
- [6] C.-M. Lo, C.-S. Lin, H. Wang, "A Miniature V-band 3-Stage Cascode LNA in 0.13m CMOS," *ISSCC Dig. Tech. Papers*, pp. 322-323, Feb. 2006.
- [7] C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Brodersen, "Design of CMOS for 60 GHz applications", *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2004.
- [8] S. Emami, C. H. Doan, A. M. Niknejad, R. W. Brodersen, "Large-signal millimeter-wave CMOS modeling with BSIM3," *RFIC Digest of Papers*, 2004 pp. 163-166.