# From Oxymoron to Mainstream: The Evolution and Future of RF CMOS

Thomas H. Lee

Department of Electrical Engineering,
Stanford University, Stanford, California 94305-4070, USA

***Abstract* — Well into the latter half of the 1990s, few experts believed that CMOS would ever become a credible RF technology. A scant decade later, CMOS dominates in many RF applications below 10GHz. Recent developments show credible performance at millimeter-wave frequencies, and continued scaling promises near-THz operation within a decade. This paper tracks key milestones in the evolution of RF CMOS to identify broad themes that permit plausible speculation about possible futures.**

***Index Terms* — RF CMOS, low-noise amplifier, oscillator phase noise, ESD, millimeter-wave, terahertz electronics.**

## I. Introduction

Today, CMOS is so prevalent as an RF technology that it is hard for newcomers in particular to imagine a time when "RF CMOS" was regarded as oxymoronic (if not simply moronic). Aside from its notoriously poor transconductance per unit current, CMOS was thought to suffer so much from poor passive components and high noise that it would be forever an uncompetitive medium for realizing RF circuits. Yet, persistent work has overcome these deficiencies to such an extent that CMOS has actually become dominant in a great many RF applications. It is remarkable that the transition from "absurd" to "obvious" occurred in the space of a decade. It is perhaps even more remarkable that CMOS RF ICs for use at 60GHz are nearly commercial realities, and that possibilities of near-THz CMOS are now contemplated with seriousness. Tracing the history of this dramatic transition is interesting for more than simply nostalgic reasons. Identification of important enabling developments teaches valuable lessons that aid in predicting possible futures for RF CMOS specifically, and for wireless integrated circuit technology in general.

## II. Scaling as an Enabler for RF ICs

Scaling is such a well-known critical driver of progress that mentioning it might seem unnecessary. However, it is relevant to note that CMOS took time to scale to a point where it could provide the requisite performance at frequencies that correspond to those of compelling applications. An example of "too little, too soon" is a broadcast FM radio constructed in 2μm technology in the late 1980s [1]. This RF CMOS receiver – the world's first – had almost no impact on the field because of its lackluster performance for an uninteresting application that was already well served by superior incumbent technologies. Extrapolation to a more capable future simply required too great a leap of faith. Not surprisingly, the ISSCC paper selection committee found the work insufficiently valuable for presentation at the 1991 conference. Indeed, a complete CMOS RF receiver would not appear at ISSCC for six more years.

The time lag is completely understandable. Large-volume wireless applications then emerging (e.g. DECT cordless phones, pagers, cellular telephones) operate roughly in the 1GHz frequency range. Engineering practice is to operate devices at well below ft in order to obtain good performance, so work on RF CMOS ICs did not begin in earnest until after technology had advanced enough to provide ~10GHz transistors (see Fig. 1). Published results finally began appearing several years after that.

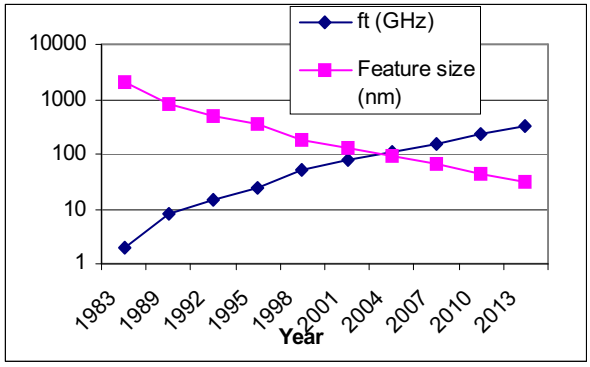


Fig. 1. Approximate ft and feature size by year (*Source*: Primarily various editions of the ITRS Roadmap).

As an *extremely* crude rule of thumb, the peak NMOS ft for modern technologies is approximately 10THz-nm/Lmin, so that the "practical" operating frequency is very roughly 1THz-nm/Lmin. Initial proofs-of-concept can debut at frequencies near ft (or more rigorously, fmax, but these quantities are generally within an octave of each other, so we will use one as a proxy for the other in this article). Practical commercialization can then follow perhaps a couple of generations later. In

conformance with this approximate rule is an early demonstration of 60GHz circuits in 130nm technology [2]. If this rule continues to hold, we can expect demonstrations of near-THz CMOS circuits within a decade.

## III. NQS Modeling Drives Advances in LNA Design

Scaling is necessary, but not sufficient. Modern wireless communications systems demand extraordinary dynamic range, which is bounded on the lower end by the achievable noise floor. Throughout much of the 1990s, few engineers had a correct understanding of what causes RF noise in MOSFETs, despite van der Ziel's relevant publication on JFET noise decades earlier (perhaps a case of "too much, too soon") [3]. Early CMOS LNAs were consequently quite noisy.

Most undergraduates are taught that the gate structure of a MOSFET is a capacitance. This low frequency first-order approximation may serve freshmen well, but it leads to fundamental misunderstandings if carried over to the RF regime without modification. The work of van der Ziel shows us that, *even if the gate electrode were made of superconductors*, a MOSFET would still have a dissipative input impedance, thanks to unavoidable non-quasistatic effects (see Fig. 2).

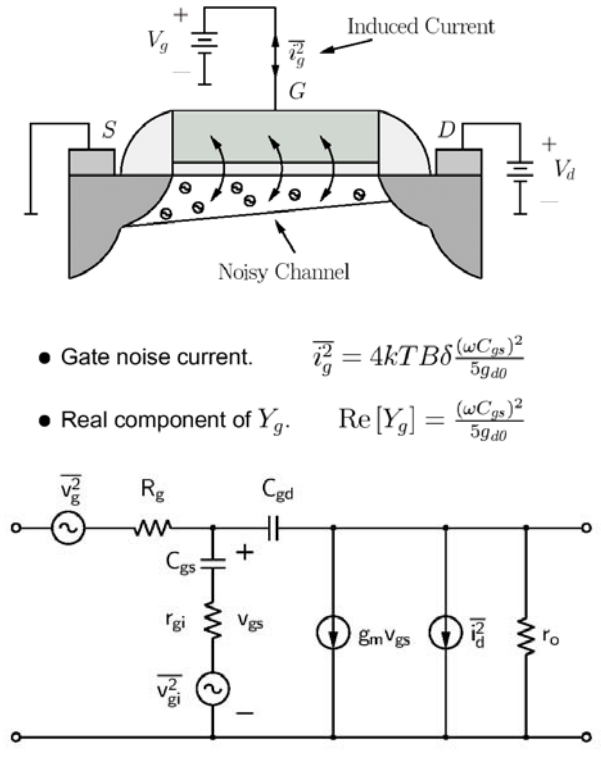


Fig. 2. *Top*: MOSFET cross-section showing how noisy channel charge induces a noisy gate noise current. Equations for this current, and its associated real gate impedance, are also shown. *Bottom*: One possible form of equivalent device model. Note carefully how the gate-source voltage is defined.

Not only does this real part bound the power gain (which a pure capacitance would not), it also produces a noisy gate current. The circuit design implications of accommodating the gate noise of van der Ziel's model were worked out by Shaeffer et al. in 1996-1997 [4][5]. Regrettably, an understanding of the fundamental role of gate dissipation in bounding noise and power gain is still not as widespread as it should be, even a decade later. For example, the following recent quote packages several tragic misapprehensions in two sentences:

> *CMOS devices take voltage, rather than power, as an input, and they have an almost purely capacitive input, so they can't absorb any power.* [G]*etting ... voltage gain using passive components that consume zero [power is] the most important factor in overcoming the noise of that device* [6].

In light of so recent a disappointing example, perhaps the reader will forgive a somewhat more detailed recapitulation of LNA design theory than would otherwise appear in an article of this kind.

Without a noisy gate current, the "passive voltage gain" strategy outlined in the citation would indeed succeed. Imagine, for example, simply resonating the gate capacitance with a suitable inductor. With a very small device and a large inductance, the network Q and voltage gain would be correspondingly large, permitting signal-to-noise ratios that are bounded only by the quality of the passive elements. Indeed, given the assumption of zero gate current noise, shrinking the device toward zero width while increasing the inductance to maintain resonance would asymptotically lead to zero noise factor, zero power dissipation, *but a nonzero gain*.

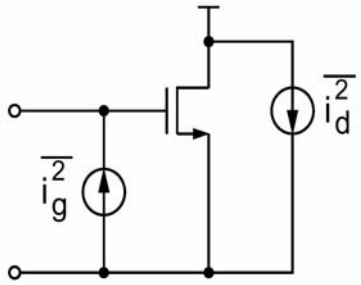


Fig. 3. MOSFET with gate and drain noise currents. Both mean-square current noise densities are proportional to device width.

Correctly accommodating noisy gate current fundamentally evades such an absurdity (see Fig. 3). Increases in inductance (decreases in capacitance) increase the impedance at resonance. The noisy gate current induces ever-larger noisy gate voltages as a consequence. Beyond a certain impedance level, this

induced noise component dominates. A definite optimum L/C ratio therefore exists for minimum noise figure. Furthermore, this optimum noise figure condition does not generally coincide with the condition that maximizes gain, so a trade-off exists between maximizing gain and minimizing noise figure. Shaeffer et al. identified the inductively-degenerated common-source topology (the vacuum-tube version of which had been analyzed by van der Ziel in the 1930s) as minimizing the impact of this trade-off. That is, this topology permits the near-simultaneous attainment of maximum gain and minimum noise figure.

Despite this achievement, a body of published data about noise in scaled devices (down to the 250nm generation) generated pessimism about the future. Some papers reported anomalously high noise in such short-channel devices, with speculations that the high electric fields within them heat carriers and thereby increase noise [7][8]. If true, continued scaling would deliver fast, but unacceptably noisy transistors. Fortunately, this "doomsday scenario" has not materialized, and recent publications have called some of this earlier work into question [9]. In any case, fears of a "noise catastrophe" seem to be wholly unfounded, as noise figures of approximately 1dB and below are now routinely achieved at gigahertz frequencies with deep-submicron CMOS.

## IV. Progress in Oscillator Phase Noise

Skepticism about RF CMOS extended to a belief that achieving low oscillator phase noise was impossible. The argument was that the notoriously poor 1/f device noise of MOSFETs would inevitably produce unacceptably high close-in phase noise, as was commonly observed in oscillators built in other FET technologies (e.g., GaAs MESFET).

The lack of good, simple phase noise models made evaluation of such statements extremely difficult. Extant simulation tools, though suitable for analysis of oscillator phase noise, provided little in the way of insight. In particular, such tools could not trace quantitatively the precise path by which 1/f device noise evolves into close-in phase noise.

The phase noise theory developed in response treats an oscillator as a general input-output system, in which the inputs are noise sources, and the output is the perturbed oscillation at a given node [10]. Evaluation of a given input-output relation generally reveals a violation of linear time-invariance. If phase is the output variable, the relationship is found to be well approximated as linear for the small noise signals that perturb practical oscillators. Though linear, these relations are periodically time *varying* in general. Most engineers are not conversant with LTV systems because curricula typically emphasize LTI analysis almost to the exclusion of all else. Fortunately linearity is sufficient to permit the continued exploitation of superposition, facilitating an evaluation of the collective effect of individual contributions to phase noise. Additionally, the validity of superposition also means that the response to an arbitrary input may still be deduced from an impulse response.

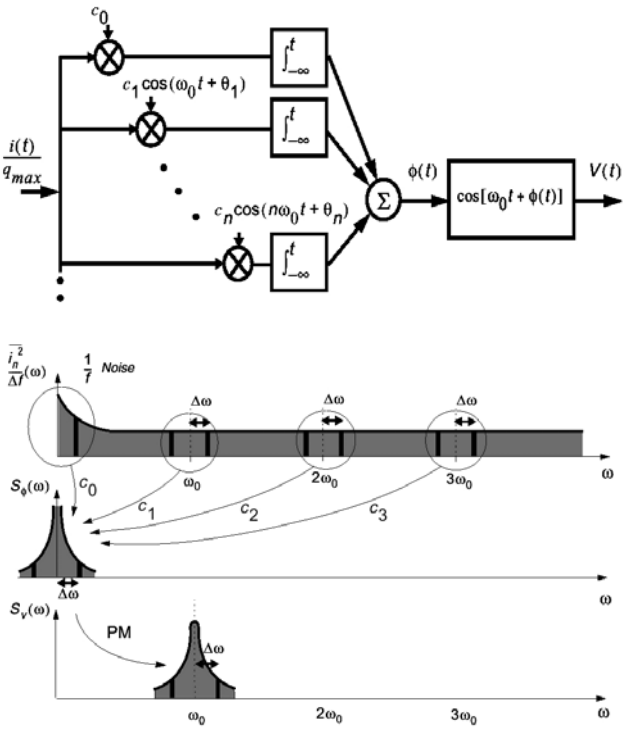


Fig. 4. *Top*: Equivalent block diagram of phase noise spectral conversion. Note similarity of each path to a superheterodyne receiver. *Bottom*: Illustration of spectral folding. The various $c_n$ are the Fourier coefficients of the normalized impulse response.

Time variation complicates matters somewhat because, in contrast with an LTI system, an LTV system may produce a response at a frequency that differs from that of the excitation (see Fig. 4). This behavior should be familiar, as it is a fundamental property of superheterodyne receivers. Consequently, this property is readily accommodated using the formal superposition integral in conjunction with the impulse response(s). In general, the latter needs to be determined through simulation (such as used in a related approach for evaluating mixer noise [11]), although an analytical form may be obtained for a few special cases.

In general there are multiple noise generators, so one must determine the impulse response connecting each noise source to the output needs to be determined. Compute the individual phase responses (which facilitates construction of an ordered list to identify dominant contributors), and then simply add them together. If determining the spectrum of the output

voltage, rather than of the phase, is the final objective, perform an additional (nonlinear) phase modulation step, as shown in the right-most block of Fig. 4.

Aside from its simplicity, the LTV model informs design in important ways. Among other insights, the LTV theory explicitly identifies the role that symmetry can play in suppressing the effects of 1/f noise. In principle, total suppression is possible, allowing excellent close-in performance despite the acknowledged inferiority of the devices' 1/f noise properties. Although perfect suppression is unsustainable in practice, usefully large reductions are readily obtained (see Fig. 5).

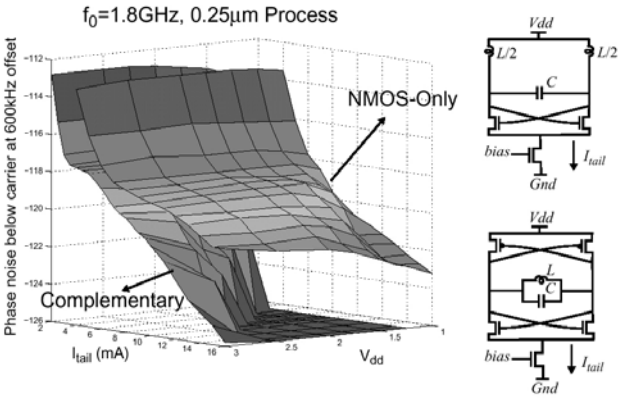


Fig. 5. Illustration of symmetry's effect on suppressing 1/f noise upconversion.

The phase noise performance of CMOS oscillators is now fully competitive with many bipolar-based oscillators, and generally superior to those realized in III-V technologies. None but the most optimistic imagined this possibility a decade ago.

## V. Some Practical Matters: ESD

Demonstrating the ability of CMOS to provide low-noise amplifiers and oscillators at GHz frequencies in an academic setting is one thing. Providing adequate ESD protection without impairing these hard-won achievements is quite another. Many cellphone manufacturers are beginning to require some front-end circuits to evince ESD withstands of 10kV (HBM). Needless to say, this requirement presents a significant challenge. Conventional ESD structures (e.g. snap-back clamps) are hard-pressed to provide adequate protection without degrading performance at GHz frequencies. Using diplexers to decouple signal paths from ESD paths can ease the problem [12], but to date this approach has only demonstrated 3kV withstands at 5GHz [13].

To provide adequate protection forces the use of a structure with substantial capacitance. The diplexer approach resonates this capacitance at signal frequencies to minimize the impact of the ESD structure on the signal path. The effectiveness of this strategy is limited by factors such as the attainable Q values for the resonating elements. An approach that considerably relaxes the design constraints, while simultaneously enabling broadband (not just high carrier-frequency) operation, is to distribute the total ESD capacitance among several segments that collectively behave as a lumped transmission line (see Fig. 6) [14]. At the relatively low frequencies comprising an ESD pulse, the segments effectively merge and act as one large ESD protection structure. At high frequencies, the distributed nature becomes evident, and the overall structure behaves as a simple delay line. A demonstration prototype shows >12kV HBM withstand over a bandwidth exceeding 30GHz.

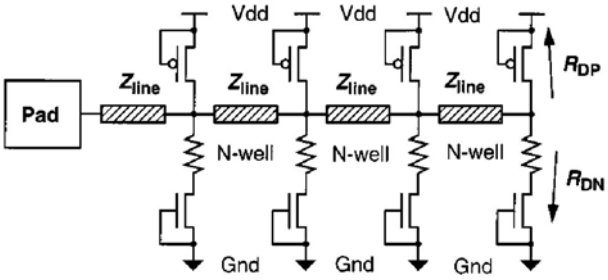


Fig. 6. Distributed ESD protection for broadband circuits.

A drawback of this approach is the relatively large area consumed. However, it remains the highest-performance ESD strategy yet demonstrated.

## VI. Possible Futures

The prevailing belief is that CMOS scaling will continue and transistors will keep getting faster for about another decade or so (as in the extrapolations on the right-hand part of Fig. 1). This trend nicely complements the widespread natural impulse to push bandwidths and datarates ever upward. A reasonable question is whether there are compelling applications in these new frequency ranges for, unlike digital circuits, one must additionally consider the propagation properties of the environment. As frequencies increase into the millimeter-wave bands, absorption and diffraction generally increase dramatically. For some short-range, high datarate applications (e.g., home HD theater interconnection), transmission at frequencies corresponding to strong absorption peaks, such as that due to oxygen around 60GHz, is actually desirable, to reduce co-channel interference. Covert inter-satellite communication uses bands with this same property to reduce the probability of intercept by ground-based eavesdroppers.

As frequencies increase even further, various resonances with the chemical bonds of a great many substances of interest enable spectroscopic analysis with

THz signals. Environmental and security applications abound for this type of technology.

Deep submillimeter (near-THz) waves are also capable of imaging through opaque objects. Although their penetrating and resolving power are not as great as that of x-rays, they also don't damage biological tissues because of their non-ionizing nature. Early detection of subcutaneous protocancers and dental caries represents just a couple of the many medical applications that would be enabled by the deployment of inexpensive near-THz systems. CMOS, perhaps working in tandem with some devices built in other technologies, seems well positioned as a medium for building these future THz systems [15].

The "better, faster" vector is a familiar one, and the field will certainly move along it. To identify other important futures, perhaps it is useful to look at the broad sweep of wireless history in brief.

Marconi's century-old contribution to wireless may be characterized as the commercialization of station-to-station wireless telegraphy. The second age of wireless, broadcasting, upended Marconi's worldview with a fundamentally asymmetric station-to-persons communications model. Today we are enjoying the prominence of the third age of wireless, in which symmetrical person-to-person communications takes place at will.

Given that all possible permutations of *stations* and *persons* have thus been covered, one might reasonably wonder if the basic outline of wireless history is now complete (aside from the aforementioned obsession with increased data rates). Of course, the answer is no.

What is unfolding now is a fourth age of wireless in which *things* communicate with other *things*. There are already classes of devices that have this characteristic, of course. RFID tags, Bluetooth audio headsets, keyless entry remotes and the like are all examples of Fourth Age devices. Indeed, one may argue that CMOS has made the Fourth Age practical. As an example, consider a representative passive RF ID tag:

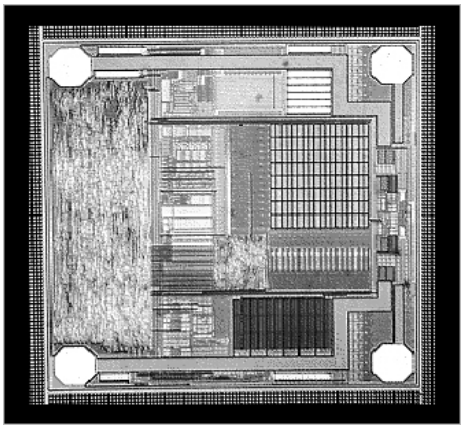

Fig. 7. Passive RFID tag (Impinj)

At less than about 1mm on a side, this mixed-signal device contains a substantial number of digital gates. It is hard to imagine any technology other than CMOS having the ability to realize such a chip at costs consistent with the demands of the target markets.

As another example, consider the trends toward "strong digital, weak analog" and multiband operation as typified by the software-defined radio and its cousins. The combination of RF and analog blocks with heavy digital processing units is most practical when realized in CMOS form. The "DRP" from Texas Instruments is a representative of this class of systems (Fig. 8) [16]. It is evident from the figure that most of the die is occupied by digital elements. As with the passive RFID tag example, realizing this system in other than CMOS is all but unthinkable.

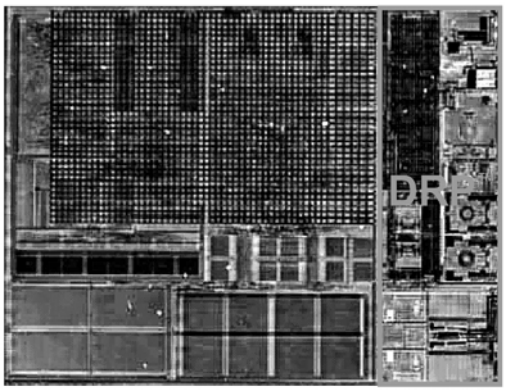

Fig. 8. Digital RF Processor (Texas Instruments)

Each Age of Wireless grew to a size and importance that dwarfed the Age that preceded it. If history is any guide to the future, we may expect the Fourth Age of wireless to be the largest yet, all thanks to CMOS.

## VII. Conclusion

The prominence of RF CMOS was not predicted by most experts, and its success has generally exceeded the expectations of even its proponents by a good margin. The combination of scaling and global research activity aimed at moving the field forward has produced important contributions that benefit more than CMOS technology alone, and assure a future of ubiquitous wireless communication.

## Acknowledgments

The author gratefully acknowledges the contributions of his students and colleagues. The author also wishes to thank the support of the RFIT organizers, particularly for their faith that a manuscript would be submitted in time for publication, despite a complete lack of supporting evidence.

## REFERENCES


[1] Thomas H. Lee, *A Fully Integrated, Inductorless FM Receiver*, Massachusetts Institute of Technology Doctoral Dissertation, December 1989.

[2] C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Brodersen, "Millimeter-wave CMOS design," *IEEE J. Solid-State Circuits*, vol. 40, pp. 144-155, Jan. 2005.

[3] Van Der Ziel, A., "Gate noise in field effect transistors at moderately high frequencies," *Proceedings of the IEEE*, vol.51, no.3, pp. 461-467, March 1963.

[4] Derek K. Shaeffer and Thomas H. Lee, "A 1.5V, 1.5GHz CMOS low noise amplifier," *VLSI Circuits Symposium Digest of Technical Papers*, June 1996, pp. 32-33.

[5] Derek K. Shaeffer and Thomas H. Lee, "A 1.5V, 1.5GHz CMOS low noise amplifier," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 5, May 1997, pp. 745-759.

[6] We note here only that the authors come from a well-respected university, and should therefore know better.

[7] Jindal, R.P., "Hot-electron effects on channel thermal noise in fine-line NMOS field-effect transistors," *Electron Devices, IEEE Transactions on*, vol.33, no.9, pp. 1395-1397, Sep 1986.

[8] Abidi, A.A., "High-frequency noise measurements on FET's with small dimensions," *Electron Devices, IEEE Transactions on*, vol.33, no.11, pp. 1801-1805, Nov 1986.

[9] Scholten, A.J.; Tiemeijer, L.F.; van Langevelde, R.; Havens, R.J.; Zegers-van Duijnhoven, A.T.A.; Venezia, V.C., "Noise modeling for RF CMOS circuit simulation," *Electron Devices, IEEE Transactions on*, vol.50, no.3, pp. 618-632, March 2003.

[10] Hajimiri, A.; Lee, T.H., "A general theory of phase noise in electrical oscillators," *Solid-State Circuits, IEEE Journal of*, vol.33, no.2, pp.179-194, Feb 1998.

[11] Hull, C.D.; Meyer, R.G., "A systematic approach to the analysis of noise in mixers," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on [see also Circuits and Systems I: Regular Papers, IEEE Transactions on]*, vol.40, no.12, pp.909-919, Dec 1993.

[12] Vickram, Vathulya; Alok, Govil; Tirdad, Sowlati., "ESD protection circuit for use in RF CMOS IC design," U.S. patent 6894567, filed 4 Dec. 2001.

[13] Leroux, P.; Steyaert, M., "A 5 GHz CMOS low-noise amplifier with inductive ESD protection exceeding 3 kV HBM," *Solid-State Circuits Conference, 2004. ESSCIRC 2004. Proceeding of the 30th European* , vol., no., pp. 295-298, 21-23 Sept. 2004.

[14] Kleveland, B.; Maloney, T.J.; Morgan, I.; Madden, L.; Lee, T.H.; Wong, S.S., "Distributed ESD protection for high-speed integrated circuits," *Electron Device Letters, IEEE*, vol.21, no.8, pp.390-392, Aug 2000.

[15] Thomas H. Lee, "Candidate THz sources: the history and future (?) of velocity-modulated devices," *Proceedings of SPIE* -- Volume 6120, *Terahertz and Gigahertz Electronics and Photonics V*, R. Jennifer Hwu, Kurt J. Linden, Editors, Mar. 7, 2006.

[16] Staszewski, R.B.; Muhammad, K.; Eliezer, O., "Digital RF Processor (DRP) for Mobile Phones," *Radio Frequency Integrated Circuits (RFIC) Symposium, 2007 IEEE*, vol., no., pp.181-184, 3-5 June 2007.