The Thermal Scaling: from Transistor to Array

Tianbing Chen, Tzung-Yin Lee, Justin Allum, and Mike McPartlin

Wafer Foundry Engineering, Skyworks Solutions, Inc.
20 Sylvan Rd, Woburn, MA 01801, tianbing.chen@skyworksinc.com

***Abstract* — The electrical and thermal performance scaling from transistor to array was studied in this paper to help improve the predictive modeling of the electrical-thermal behavior of bipolar arrays. The dc and ac performance scales well at low bias but not to the medium and high bias because thermal resistance does not scale to the total emitter area. It is demonstrated that after the correction of the R_{TH} in array models, the simulation can predict the electrical-thermal behavior of power arrays.**

***Index Terms* — SiGe HBT, scaling, self-heating, thermal resistance, array, power amplifier.**

I. Introduction

In the past decade the cellular power amplifier market has been a constant battle ground over the product performance, ruggedness, and cost for semiconductor processing technologies such as GaAs, CMOS, and SiGe BiCMOS etc. With the optimal combination of performance, cost, functional integration, and manufacturability, the SiGe BiCMOS technology has emerged recently as one of the leading candidates for cellular, WiMAX, and Wi-Fi RF front-end systems [1]. In order to compete against GaAs technology, which has very small parasitics, one of the key development efforts for SiGe BiCMOS process was to improve the performance through parasitic reductions [2]. In a SiGe PA array design, a multiple-finger HBT unit cell is typically preferred over a single-finger counterpart for higher power added efficiency and also for higher power gain due to its lower parasitic capacitances. The use of the multiple-finger device has inevitably led to the higher power density, and so a healthy balance between circuit performance and transistor self-heating is required. During the ruggedness test the PA can operate at much higher voltage or power density for high mismatch conditions. This puts further constrains on the PA thermal design.

Due to the PA's need for a very large output power (Pout) the total device area for the power cell becomes very large. The power cell has to be composed of an array of multiple-finger cells. The mutual heating and associated parasitic capacitance in between unit cells within an array are layout dependent and a non-optimized design can cause the PA power performance degradation. There has been a lot of research regarding the self-heating on the single transistor [3][4][5]. However, there has been very limited work on the design of transistor arrays that were used in the PA design. The foundry process design kit normally covers the device performance well for a single transistor, but not for the array-level performance. For example, mutual heating depends on the transistor spacing in the layout and also the size of the power cell, and it is rarely modeled. Without proper physical understanding of the array heating, the design is empirical and it relies heavily on the designer's experience. The performance scaling from a single transistor to an array is critical for successful predictive modeling of a PA design in this context.

This work is organized as follows: the subsequent section presents the details for the device technology and experiment design, followed by the dc and ac performance scaling study from single transistor to arrays. Thermal resistance is then analyzed for arrays with various layouts and sizes, and a model simulation validation on data with justified thermal resistance parameters is presented in the end.

II. Device Technology and Experiment

The SiGe HBTs used in this investigation are from a commercial SiGe HBT BiCMOS technology that features 6.0-V BV_{CEO}, 35-GHz f_T SiGe HBTs and 5-V Si CMOS devices. The through-wafer vias (TWVs) are also included in this process to enable the single-ended amplifier architecture. The devices under test (DUTs) were laid out in GSG pad format for testing. To study the scaling effect, DUTs with 1, 2, 4, and 7 transistors were designed in this work. A simplified layout top view is shown in Fig. 1 for DUTs that contain different numbers of transistors. The spacing between two neighboring NPN devices was kept at a constant of 3.5 µm from p+ to p+ substrate rings. The unit SiGe HBT device used in this design is a 2 finger power transistor with emitter width and length of 0.9 µm and 30µm, respectively.

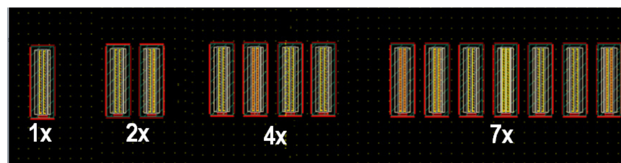

Fig.1. Simplified top view of the layout for NPN cells with different number of transistors.

III. DC & RF PERFORMANCE SCALING

Fig. 2 shows the forward Gummel characteristics on the left for all four DUTs with different numbers of transistors. At low- and medium V_{BE} the collector and base current scales well with the total emitter area as expected. The current density at large V_{BE} (>0.8V) is different for the different DUTs because at large current the IR drop on the BE junction does not scale to the total emitter area. The β vs. J_C plot shown on the right of Fig.2 suggests that β is the same for all DUTs, and is independent of the total number of transistors, other than some die-to-die variations.

The measured output curves at constant low V_{BE} for SiGe HBTs with different number of transistors is shown in Fig. 3. At V_{BE} of 0.72, 0.74, and 0.76V the transistor current density is virtually the same for all DUTs. However, the current slope starts to diverge for the largest DUT with seven transistors at V_{BE} of 0.76V, where Jc is only as low as 10 μA/μm^2. The mismatch of the early voltage is likely due to the self-heating difference for DUTs with a different number of transistors.

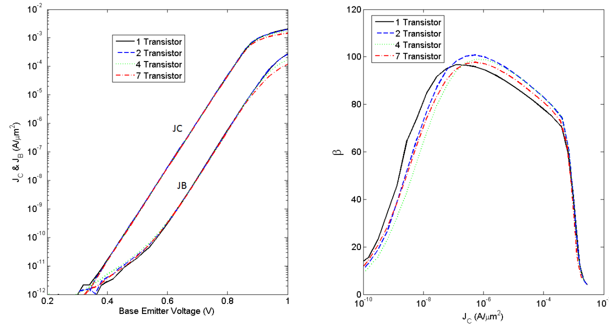

Fig. 2. Forward Gummel characteristics for DUTs with different numbers of transistors: a) J_B and J_C versus V_{BE}; b) β versus J_C

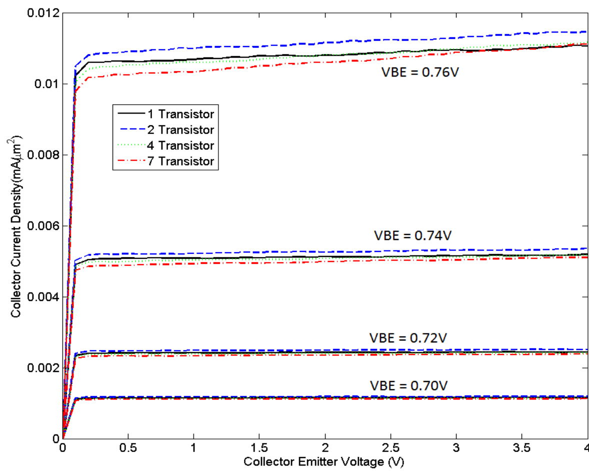

Fig. 3.Constant low V_{BE} output characteristics for SiGe HBTs with a varying number of transistors.

To examine the self-heating difference, the constant V_{BE} output characteristics were measured under higher V_{BE} for all DUTs with different numbers of transistors. As we can see from Fig. 4, the IV mismatch increases with V_{BE}. Some previous work has demonstrated that the SiGe HBT stability in the bias range of 0.1mA/μm^2 is defined by a combined effect of thermal instability and impact ionization [4]. Base current reversal at constant V_{BE} drive was tested to gauge the impact ionization of all DUTs. The results shown in Fig. 5 suggest that BV_{CEO} is 6.2V for all DUTs. The calculated avalanche factor shown to the right in Fig. 5 also suggests that there is no impact ionization difference. So we can safely conclude that the IV mismatch from Fig. 4 is merely caused by the self-heating difference. As the number of transistors in a DUT increases, there is more junction temperature rise due to the increased mutual heating. Constant base current drive tests were also performed to analyze the transistor junction temperature rise with the increase of power dissipation. As we can see from Fig.6a, little collector current density difference exists for different DUTs at constant base current drive. However, the V_{BE} from Fig.6b, which is an electrical parameter that tracks junction temperature, is dramatically different among all DUTs. As the power dissipation increases proportional to V_{CE} to the first order, the transistor junction temperature will increase with V_{CE}. With the increase of the junction temperature, the V_{BE} decreases to maintain a fixed base current. The ΔV_{BE} relative to its peak value in Fig. 6b represents the junction temperature rise at a given bias condition. It can be clearly seen from Fig. 6b that the junction temperature rises more as the number of transistor increases for the DUTs, consistent with data analysis in Fig. 4. It is common practice now to operate bipolar transistors at bias levels higher than BV_{CEO} but lower than BV_{CBO}. The trick is to make the input resistance seen from the base small, which is close to the constant V_{BE} test condition shown in Fig. 4. If the mutual heating is not accounted for in simulation, the simulated J_C for Fig. 4 will be the same family curves independent of the number of transistors. This type of over-optimism in simulation can lead to under ballasting conditions in the actual circuit design.

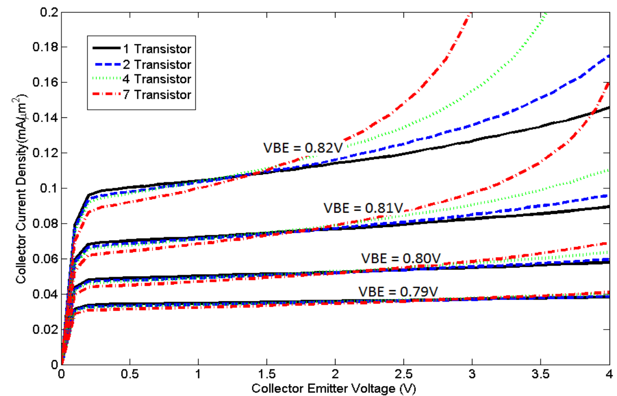

Fig. 4. Constant high V_{BE} output characteristics for SiGe HBTs with a varying number of transistors.

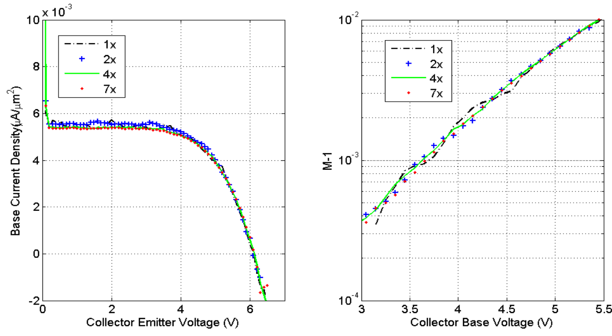

Fig. 5. Base current reversal measurement for DUTs with a different number of transistors: a)J_B versus V_{CE} with V_{BE} = 0.68V; b) Calculated avalanche factor for the same test.

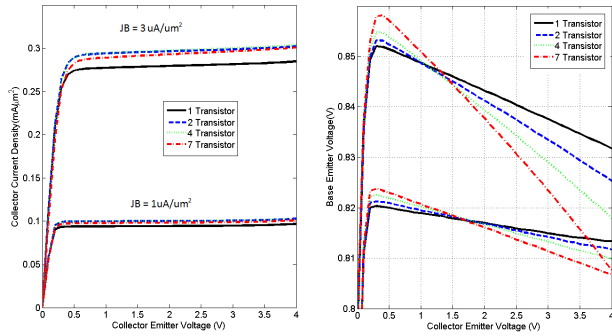

Fig. 6. Constant IB drive tests for DUTs with a varying number of transistors: a)J_C versus V_{CE}; b)V_{BE} versus V_{CE} .

In addition to the *dc* performance, the *ac* performance scaling was examined for all DUTs. As can be seen from Fig.4 there are thermal stability issues for large DUTs when biased in conventional constant V_{BE} conditions for Jc>0.1mA/μm^2 and V_{CE}>2V bias conditions. The constant base current drive was used in the s-parameter characterization to test all DUTs under full bias range. Typical G_{max}/I_C and f_T/I_C curves are plotted in Fig. 7. The extracted G_{max} and f_T are similar for all DUTs when biased at the same collector current density.

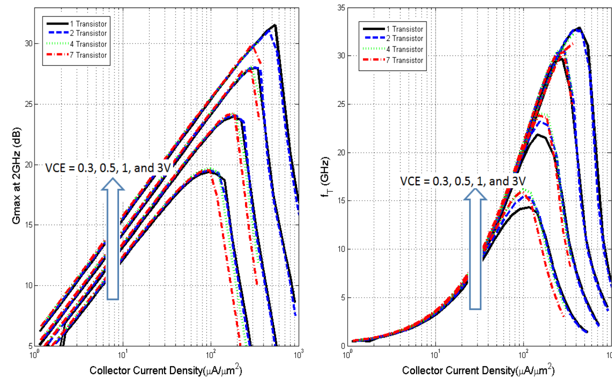

Fig. 7. a) The measured maximum transducer power gain (G_{max}) *vs.* collector current density (J_C) for DUTs with varying number of SiGe HBTs; b) Measured cut-off frequency (f_T) *vs.* collector current density (J_C) for DUTs with varying number of SiGe HBTs.

IV. THERMAL RESISTANCE

To quantify the level of transistor heating, the thermal resistance (R_{TH}) was extracted for all DUTs by using V_{BE} to monitor the junction temperature [6]. For DUTs with multiple transistors collected in parallel, the measured V_{BE} represents the electrical average of the junction temperature of all transistors. So the extracted R_{TH} stands for the average heating of DUTs by accounting for both the individual transistor self-heating and the mutual heating between transistors. The measured R_{TH} vs. number of transistors was plotted in Fig. 8 together with its R_{TH} parameter from the default HiCUM model. We use R_{TH0} to stand for the thermal resistance for a single transistor, and for an n transistor array, R_{TH} is simulated as $R_{TH} = R_{TH0}/n$. Due to the exclusion of the mutual heating factor in the default transistor array model, the simulated R_{TH} is smaller than the measured R_{TH}. To visualize the gap between model and data, a new parameter M_{TH} was plotted in Fig. 8. M_{TH} is defined as the R_{TH} multiplication factor for a transistor array:

$$M_{TH} = \frac{R_{TH}_DATA}{R_{TH}_SIMU} \quad (1)$$

With the increase of the array size there is more mutual heating, which is not accounted for in a typical npn model. As a result, the M_{TH} increases together with the number of transistors. Consider a simple current mirror application as shown in Fig. 9 I_{REF} is biased on a single transistor, Q1, while the output is from the collector of a large power array, Q2. The V_{BE} of Q2 is fixed at a given I_{REF}, and I_{OUT} increases with V_{OUT} due to the increased power dissipation. The simulated I_{OUT} will not be accurate if mutual heating is not accounted for in the Q2 model.

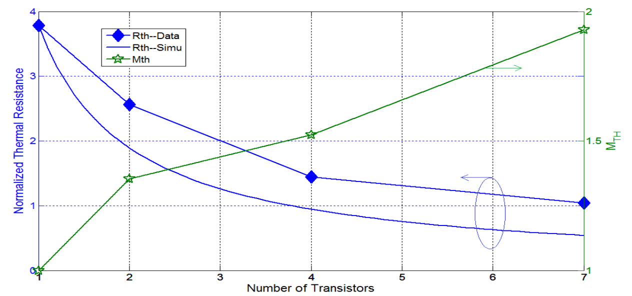

Fig. 8. The measured and the simulated thermal resistance for all DUTs with a varying number of transistors.

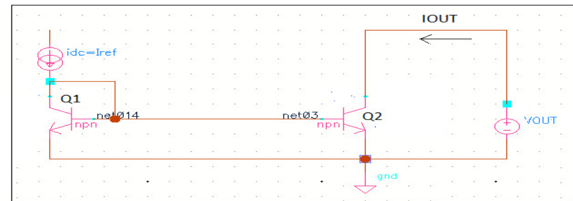

Fig. 9. Simple two-transistor current mirror

Other than the size of the array, the npn mutual heating is also a strong function of device spacing in the layout.

Thermal resistances were extracted for 7x arrays with varying npn spacing and plotted versus the total array width in Fig. 10. With the increase of npn spacing and also the total array width, the mutual heating is reduced and so is the thermal resistance.

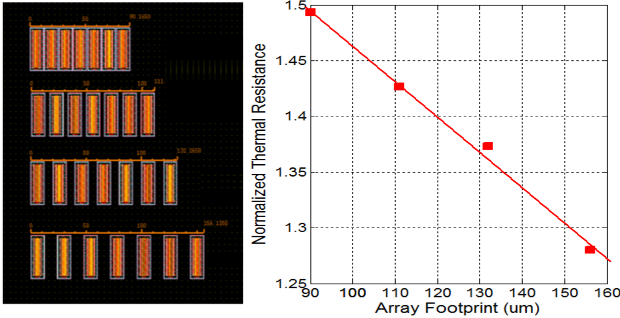

Fig. 10. a) The layout for 7x array with varying npn spacing; b) Measured thermal resistance for 7x arrays.

A previous study on an SOI BiCMOS technology had shown that the transistor thermal resistance is a strong function of device footprint [5]. Thermal resistances were extracted for 30+ transistors and arrays on two die. The results are plotted as a function of device footprint in Fig. 11. In reference [3] the heat removal from a planar device on bulk substrate is considered radial in all directions from the active area and R_{TH} can be approximated as

$$R_{TH} = \frac{1}{4\kappa Si\sqrt{LW}} \quad (2)$$

Where L and W define the box area of the device footprint, and KSi is the silicon thermal conductivity. Indeed in Fig. 11 we can see that all the R_{TH} data can be well fitted into a generic function of device footprint.

$$R_{TH} = \frac{A}{\sqrt{Footprint}} + B \quad (3)$$

This has made predictive thermal modeling possible.

V. Model Validation

Today's commercial bipolar models such as HiCUM, Mextram, and VBiC are fairly accurate in terms of simulating the electrical thermal behavior of a single transistor. However, the mutual heating is rarely included in an array application. By default, if one plots the simulated family curves at a constant VBE drive for DUTs in Fig.1, the J_C vs. V_{BE} curves will be the same independent of the number of transistors. Through data analysis, we know that the J_C mismatch in Fig.4 is caused mainly by the R_{TH} difference. The calculated M_{TH} factor in Fig.8 is implemented into the models for all DUTs and the model playback is shown in Fig. 12. The simulated IV curves with adjusted R_{TH} in models fit the measurement data very well. We also demonstrated in an actual PA design that the ICQ simulation-vs.-Si gap was brought down from 14% to less than 2% after the implementation of the correct Rth model for the array device.

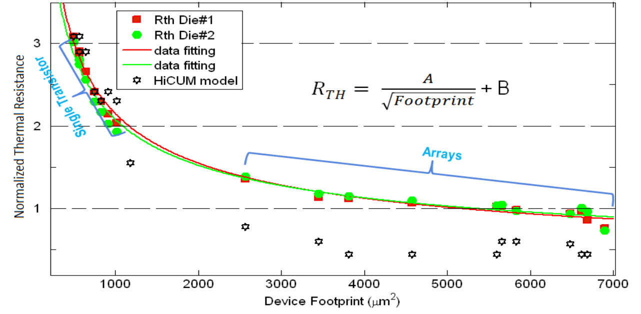

Fig. 11. Measured thermal resistance versus device footprint for various single transistors and arrays.

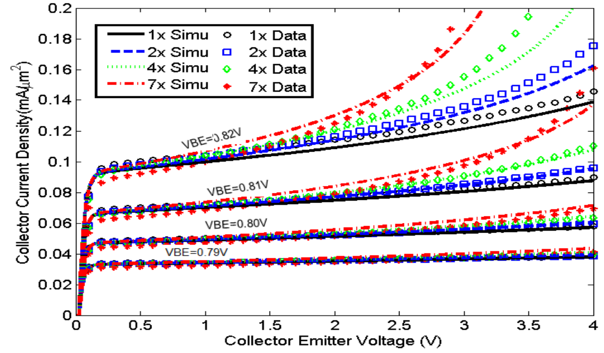

Fig. 12. Model validation on constant high V_{BE} output characteristics for DUTs with varying number of transistors.

VI. Conclusion

This paper investigates the *dc*, ac, and thermal scaling from a single transistor to an array. The dc and ac performance scale well at relatively low bias, but mismatch occurs at medium to high bias due to the non-scaling nature of the thermal resistance to the emitter area. After the R_{TH} in the models is adjusted to account for the mutual heating in arrays, the simulation is found to fit the IV curves of arrays at a much higher power density.

References

[1] J. Brewer and P. Gammel, "Silicon BiCMOS: The Optimal Technology for RF Power," Microwave Journal, Sep. 12, 2010.

[2] R. Jos, "Future developments and technology options in Cellular Phone Power Amplifiers: from power amplifier to integrated RF frond-end module," IEEE BCTM, Proceedings of 2000, pp. 118-125.

[3] A. Pacelli *et al.*, "Compact Modeling of thermal resistance in bipolar transistors on bulk and SOI substrate", IEEE Trans. Elec. Dev. Vol.49, No.6, 2002, pp.1027-1033.

[4] J. Kim *et al.*, "Safe operating area from self-heating, impact ionization, and hot carrier reliability for a SiGe HBT on SOI", IEEE BCTM, Proceedings of 2007, pp. 230-233 .

[5] T. Chen *et al.*, "Footprint design optimization in SiGe BiCMOS SOI technology", IEEE BCTM, Proceedings of 2008, pp. 230-233.

[6] J.-S. Rieh *et al.*, "Measurement and modeling of thermal resistance of high speed SiGe heterojunction bipolar transistors, " Proc. Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems, 2001, pp. 98-101.