Investigating Harmonic Distortion Dips in High-Resistivity Silicon Substrates

Vandroogenbroek Arthur
UCLouvain
Louvain-la-Neuve, Belgium
arthur.vandroogenbroek@uclouvain.be

Rack Martin
UCLouvain
Louvain-la-Neuve, Belgium
martin.rack@uclouvain.be

Allibert Frédéric
Soitec
Grenoble, France
frederic.allibert@soitec.com

Raskin Jean-Pierre
UCLouvain
Louvain-la-Neuve, Belgium
jean-pierre.raskin@uclouvain.be

***Abstract*— Harmonic distortion (HD) measurements are widely used to assess substrate linearity, typically exhibiting a quadratic relationship between the second harmonic (H_2) and the fundamental component (H_1). However, measurements on high-resistivity (HR) silicon substrates can present unexpected dips in H_2 vs. H_1 curves. While previously observed, the underlying physical mechanisms remain insufficiently understood. This paper investigates these dips through TCAD simulations and RF measurements, demonstrating that they result from the interplay between variations in equivalent capacitance and conductance. The phase difference between these two harmonic sources under certain bias conditions leads to destructive interference, enabled by carrier-lag phenomena in the silicon substrate, explaining several abnormal features in the H_2 vs. H_1 curves.**

***Keywords*— Harmonic distortion, high-resistivity Si substrate, large signal analysis, non-linear characterization, dip.**

I. Introduction

Harmonic Distortion (HD) measurements are a common method for quantifying and characterizing device nonlinearity. This process involves injecting RF power at a specific frequency f_0 and measuring the output power at harmonic frequencies, which are integer multiples of f_0 as depicted in Fig. 1. In this paper, the power at the n-th harmonic frequency $n \cdot f_0$ is denoted as H_n. To describe the substrate linearity, H_2 values measured on a specific coplanar waveguide (CPW) are often reported for H_1 = 15 dBm [1]. We can derive from mathematical analysis, that H_2 is proportional to P_{in}^2, while H_1 is proportional to P_{in} [2]. As a result, a plot of H_2versus H_1 on a logarithmic scale is expected to yield a straight line with a slope of 2.

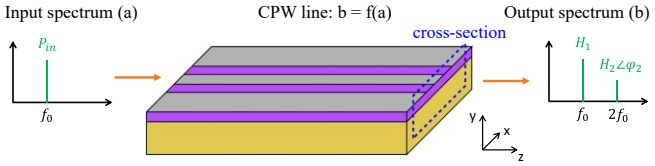


Fig. 1. HD measurement principle for characterization of substrate linearity.

Despite this expectation, HD measurements on high-resistivity substrates can present deviations, such as variations in the slope and the appearance of a local minimum, referred to as a "dip," in the H_2-vs-H_1 curve. While this phenomenon has been previously reported in [3], existing explanations remain incomplete. This paper provides a more comprehensive investigation into the physical origins of these dips through measurements and TCAD simulations.

II. Experimental Observation of Dips

Figure 2 shows an example of H_2-vs-H_1 measurements where a dip is observed. The device is an 8 mm-long CPW line fabricated on a 5 kΩ·cm p-type silicon substrate with a 4.24 µm oxide layer provided by CEA-Leti. In these experiments, the RF input signal at 900 MHz was swept from -5 to 24 dBm, while DC bias ranging from -20 V to 20 V was applied.

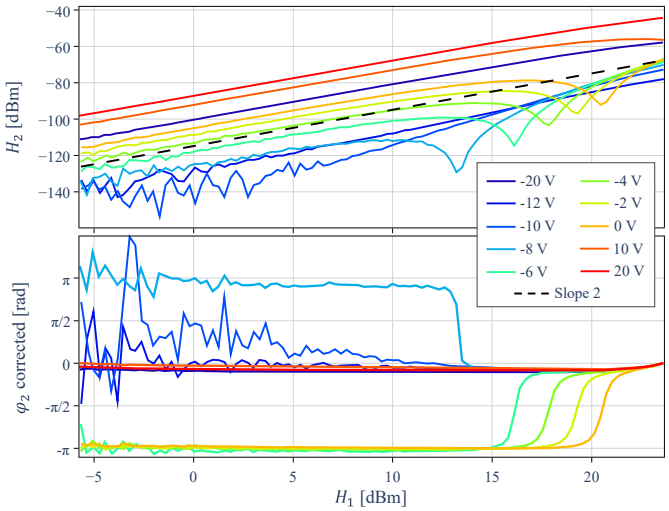


Fig. 2. Measurements of power (top) and corrected phase (bottom) of H_2 as a function of H_1 power, for bias conditions from -20 to 20 V and RF input power levels ranging from -5 to 24 dBm at a fundamental frequency of 900 MHz.

In Fig. 2, H_2 power and phase are plotted as function of H_1 power. The experimental setup based on an PNA-X analyzer from Keysight inherently provides phase information as part of the harmonic analysis. The measured phase represents the difference between the phase of the detected signal and that of PNA-X's local oscillator. The phase values plotted in Fig. 2 are corrected for measurement-induced time delays and normalized to $\varphi = 0$ at P_{in} = 24 dBm.

Fig. 2 shows that for low input power levels, the expected slope of 2 is always observed, except at a bias of -10 V. For biases above -10 V, dips are present and shift to higher H_1 values when increasing the DC bias. Each time a dip is present it is accompanied by an abrupt phase shift of approximately 180° (π). Conversely, for biases below -10 V, dips are either absent or require input power levels exceeding 24 dBm to be noticeable.

III. Physical Interpretation of Dips Based on TCAD

A. Dips in TCAD Simulations

This study employed the same TCAD modelling approach detailed in [4]-[5]. To focus on understanding the underlying mechanisms behind dips, instead of quantitively reproducing the results shown in Fig. 2, a simplified structure was analyzed. Interface traps were omitted to streamline the analysis. Accounting for their absence, the interfacial fixed oxide charge density Q_{ox} was set to $5 \cdot 10^9$ C/cm^2.

Fig. 3 presents the H_2-vs-H_1 curves as a function of bias at 900 MHz, derived from TCAD simulations of the structure. These curves exhibit similarities to those in Fig. 2, such as a slope of 2 in the low-power region and a similar bias-dependent position of the dips. Furthermore, the simulated curves exhibit saturation after the dip (in the 25-38 dBm of H_1 range), also observed for some of the measurement curves, even though H_1 was limited in the experimental setup to 24 dBm.

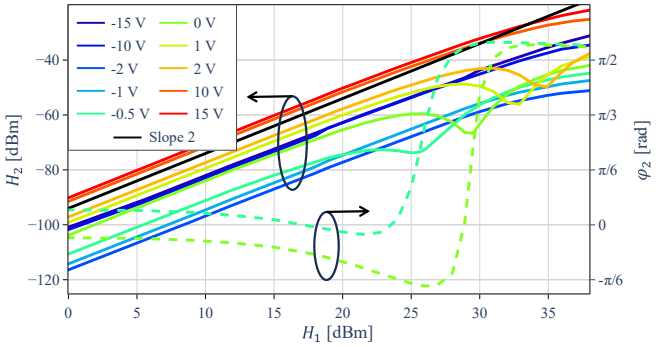


Fig. 3. TCAD simulation of H_2 vs H_1 at 900 MHz for various DC biases. The phase ϕ_2 is also plotted for the harmonic data corresponding to biases of -0.5 V and 0 V.

The simulations used to produce the results in Fig. 3 enable accurate extraction of the phase of H_2, with phase information plotted alongside power for the curves at –0.5 V and 0 V in Fig. 3. As in Fig. 2, dips are accompanied by abrupt phase changes. Though these simulations do not quantitatively reproduce the measurements in Fig. 2, they provide a good qualitative reproduction of the observed behaviors and they can bring physical insight into the origin of dips.

B. Substrate Equivalent Circuit

In [4], it was demonstrated that harmonic generation originates from the variable impedance beneath the oxide layer, which arises due to fluctuations in electron and hole concentrations in silicon substrate. The substrate impedance can be modelled as an infinite network of local elements, each comprising a conductance in parallel with a capacitance, as illustrated in Fig. 4(a). While the local capacitance of each cell is proportional to the permittivity and remains constant, the local conductance is proportional to the conductivity, which depends on the electron and hole concentrations according to:

$$\sigma = \sigma_n + \sigma_p = q\mu_n n + q\mu_p p \tag{1}$$

As a result, only local conductances are represented as variable non-linear elements in Fig. 4(a). However, when considering the cumulative effect of all elements in the network, the resulting complex impedance between the signal line and the ground lines exhibits variations in both its real and imaginary parts. Consequently, the equivalent impedance is effectively characterized by a variable conductance in parallel with a variable capacitance as shown in Fig. 4(b) [6].

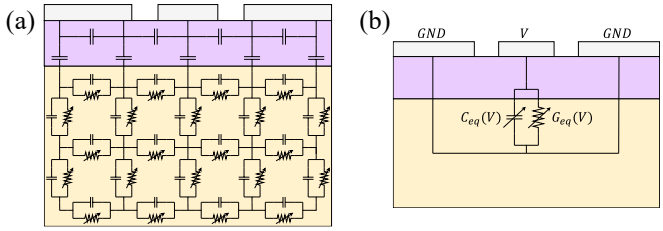


Fig. 4. (a) Representation of the non-linear impedance network in the cross section of the CPW structure. (b) Equivalent impedance representation.

C. Physical Explanation of the Dips

Both variations in equivalent conductance and capacitance contribute to harmonic generation. As a first-order analysis, the frequency components in the substrate current I can be examined by exciting the CPW cross-section equivalent circuit in Fig. 4(b) with a single-tone voltage of amplitude A at angular frequency ω_0:

$$V = A \cdot \sin(\omega_0 t) \tag{2}$$

The substrate current response is made up of two components: (i) I_G due to the real part of the equivalent substrate admittance G_{eq}, and (ii) I_C due to the imaginary part C_{eq}:

$$I_{tot} = I_G + I_C = G_{eq}(V) \cdot V + C_{eq}(V) \cdot \frac{\partial V}{\partial t} \tag{3}$$

At low power levels, the non-linear functions $G_{eq}(V)$ and $C_{eq}(V)$ can be approximated using Taylor-series expansions:

$$G_{eq}(V(t)) = G_0 + G_1 V + \cdots = G_0 + G_1 A \cdot \sin(\omega_0 t) + \cdots \tag{4}$$

$$C_{eq}(V(t)) = C_0 + C_1 V + \cdots = C_0 + C_1 A \cdot \sin(\omega_0 t) + \cdots \tag{5}$$

Substituting (2), (4) and (5) into (3) yields:

$$I_G = A\left[G_0 \cdot sin(\omega_0 t) + A \cdot G_1 \frac{1-cos(2\omega_0 t)}{2} + \cdots\right] \tag{6a}$$

$$I_C = A\left[\omega_0 C_0 \cdot cos(\omega_0 t) + A \cdot \omega_0 C_1 \frac{sin(2\omega_0 t)}{2} + \cdots\right] \tag{6b}$$

Analyzing the current responsible for the second harmonic generation at frequency $2\omega_0$ $I_{tot@2\omega_0}$, we can write:

$$I_{tot@2\omega_0} = I_{G@2\omega_0} + I_{C@2\omega_0} \tag{7}$$

$$I_{tot@2\omega_0} = A \cdot A_G \cdot \cos(2\omega_0 t) + A \cdot A_C \cdot \sin(2\omega_0 t) \tag{8}$$

Where A_G and A_C represent the amplitude of variation in G_{eq} and in C_{eq} admittance responsible for generating harmonics at $2f_0$. As long as the low-power Taylor-series expansion holds, we have:

$$A_G = -AG_1/2 \tag{9}$$

$$A_C = A\omega_0 C_1/2 \tag{10}$$

As $I_{G@2\omega_0}$ and $I_{C@2\omega_0}$ are in quadrature (i.e., phase-shifted by $\pi/2$), the total current magnitude is given by:

$$\left|I_{tot@2\omega_0}\right| = A^2/2\sqrt{G_1^2 + \omega_0^2 C_1^2} \quad (11)$$

Thus, increasing the fundamental power amplitude A results in an A^2 increase in $I_{tot@2\omega_0}$, giving the expected 2:1 slope in the H_2 -vs- H_1 power curves. However, the above mathematical description cannot justify a decrease in $I_{tot@2\omega_0}$ amplitude for an increase in A (i.e. observing a dip), indicating that some assumptions break down at higher power levels.

In particular, the presence of a dip requires that A_G and/or A_C are not strictly linear functions of A. Moreover, a phase difference between $I_{G@2\omega_0}$ and $I_{C@2\omega_0}$ between $\pi/2$ and $3\pi/2$ is required to produce destructive interference and cause a dip when A_G approaches A_C.

The phase difference between equivalent conductance and capacitance can be attributed to the spatial distribution of local elements influencing their variations, with G_{eq} and C_{eq} potentially affected by different sets of elements. Indeed, elements located further from the signal line experience a larger phase delay compared with those closer due to the finite carrier propagation delay. At GHz frequencies, this delay τ is no longer negligible, as carrier inertia prevents an instantaneous response to RF signal [4]. This can be modelled by modifying equations (4)-(5) and (7)-(8) as follows:

$$G_{eq}(t) = G_0 + G_1 A \cdot \sin\left(\omega_0(t - \boldsymbol{\tau_G})\right) + \cdots \quad (12)$$

$$C_{eq}(t) = C_0 + C_1 A \cdot \sin\left(\omega_0(t - \boldsymbol{\tau_C})\right) + \cdots \quad (13)$$

$$I_{G@2\omega_0} = A \cdot A_G \cdot \cos(2\omega_0 t + \boldsymbol{\phi_G}) \quad (14)$$

$$I_{C@2\omega_0} = A \cdot A_C \cdot \sin(2\omega_0 t + \boldsymbol{\phi_C}) \quad (15)$$

Since the elements affecting G_{eq} and C_{eq} could differ, ϕ_G and ϕ_G may not be equal. Consequently, the phase difference between $I_{G@2\omega_0}$ and $I_{C@2\omega_0}$ can be expressed as:

$$\Delta\varphi = \pi/2 + \phi_G - \phi_C \quad (16)$$

The closer $\Delta\varphi$ is to π, the more both current sources cancel each other when their amplitude becomes equal, resulting in a more pronounced dip.

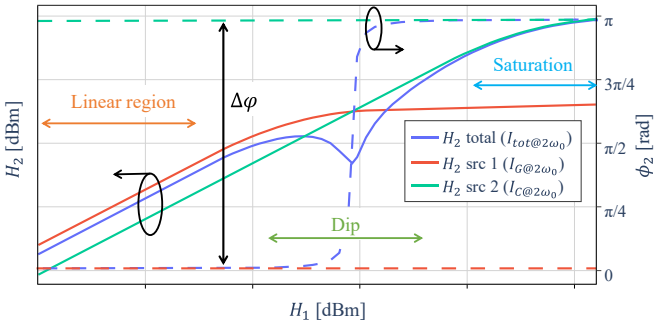


Fig. 5. Power (solid, left axis) and phase (dashed, right axis) of two theoretical harmonic sources and their combined responses.

If considering that both sources can only saturate with power due to the saturation of A_G and A_C, the significant phase shift observed at the dip's position in both measurements and simulations can only occur if the stronger of the two sources ($I_{G@2\omega_0}$ for example) saturates before the weaker source ($I_{C@2\omega_0}$). This is illustrated in Fig. 5 that shows the magnitude and phase of two theoretical sources with a $\Delta\varphi$ close to π, along with their combined responses. At low power, the first source dominates and $I_{tot@2\omega_0}$ follows the behavior described by equations (9)-(10) and (14)-(15). As a result, A^2 can be isolated in $I_{tot@2\omega_0}$, leading to the observed 2:1 slope in Fig. 5.

At a certain power level, however, the nonlinear conductance response saturates and A_G is no longer proportional to A. Consequently, the red curve in Fig. 5 saturates. As power increases, while the first source saturates, the second continues to grow, increasing destructive interference between their 180° out-of-phase components until they reach equal power, at which point a dip appears in the total response.

D. Intrinsic Behavior at -0.5 V Bias

Measurements and simulations revealed the presence of a threshold bias voltage above which dips begin to appear. This behavior is explained by the fact that, below this threshold, holes accumulate beneath the signal line, causing σ_p to dominate in equation (1) and $I_{G@2\omega_0}$ to be primarily driven by hole modulation. Conversely, above the threshold, electrons become the main contributors to both equation (1) and $I_{G@2\omega_0}$. In SiO_2 MOS structures, this threshold is known to be negative because of the positive interface charge (Q_{ox}) at the Si/SiO_2 interface, which attracts free electrons near the interface [7].

In this paper, we will focus on the behavior above that threshold voltage and more particularly to the -0.5 V bias curve in Fig. 3. Table 1 provides 2D cross-sectional views of the simulated CPW line showing the electron and hole concentrations extracted from TCAD at various H_1 power levels. These cross-sections are represented with a logarithmic scale applied to the y axis (see Fig. 1). The "Min" column depicts the carrier concentration when the RF signal reaches its most negative value, repelling electrons and attracting holes beneath the signal line. Conversely, the "Max" column shows the carrier concentration when the RF signal reaches its most positive value, attracting electrons and repelling holes.

Table 1. Cross-sections of the CPW line showing electron and hole concentrations at different H_1 power levels (9 to 32 dBm) for a bias of -0.5 V. The "Min" column corresponds to the concentration when the signal reaches its most negative voltage, while the "Max" column represents the concentration at the most positive voltage.

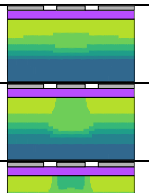
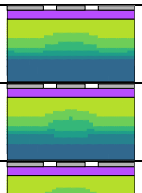
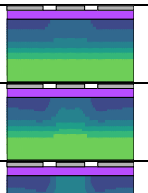
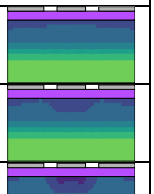
H_1 (dBm)	Electron concentration		Hole concentration	
	Min	Max	Min	Max
9				
20				
25				
29				
32				

Table 1 provides cross-sectional data for the full substrate, though our analysis focuses on the shallow regions beneath the oxide and between both planar ground lines. These areas experience stronger electric fields and greater carrier modulation, making them the primary contributors to harmonic generation [6]. Fig. 6 conceptually depicts how local impedance close to Si/SiO₂ interface beneath the signal line and how G_{eq} and C_{eq} vary at different power levels, providing a simplified explanation of the simulation results. When the local conductance G becomes smaller than the local capacitive admittance ωC, G_{eq} decreases and C_{eq} increases. Note that this conceptual representation considers that G_{eq} and C_{eq} are only driven by the local elements at the Si/SiO₂ interface beneath the signal line. Therefore, there is no delay τ_G and τ_C and $I_{G@2\omega_0}$ and $I_{C@2\omega_0}$ are in quadrature.

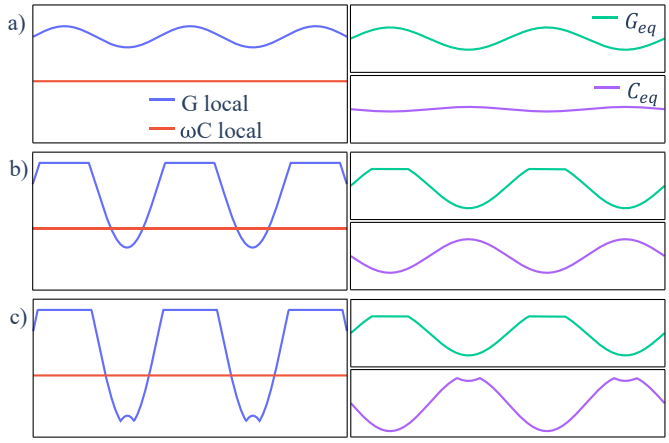


Fig. 6. Conceptual representation of the variation of local impedance below the signal line and equivalent impedance (a) in linear region, (b) in dip region and (c) in saturation region.

To better understand H2's behavior, the response is divided into three distinct regions, as illustrated in Fig. 5: the linear regions where there is a slope 2:1 in the H_2-vs-H_1 curve, the dip region, where deviations from this trend occur, and the saturation region.

1) *Linear Region*

At low power levels, the electron concentration in the region of interest consistently exceeds the hole concentration. As a result, the term $q\mu_n n$ dominates in (1), allowing the analysis to consider only the electron contribution. Since the bias applied is significantly higher than the threshold voltage, a highly conductive layer of electrons is present beneath the signal line. In this region, the admittance of local capacitances is considerably lower than that of local conductances, regardless of whether the signal is at its "Max" or "Min" value (Fig. 6a). Therefore, $I_{G@2\omega_0}$ dominates in harmonic generation and as long as A_G does not saturate, the power curve will exhibit a slope 2:1.

2) *Dip Region*

As power increases, two key effects emerge in the dip region. First, electron concentration modulation begins to saturate, along with G_{eq}, as shown in Fig. 6(b). Second, when the signal reaches its minimum, electron repulsion creates a depletion zone, oscillating between depletion and inversion. This transition induces capacitance variations, which increase with power, while conductance variations remain saturated. As a result, $I_{G@2\omega_0}$ saturates while $I_{C@2\omega_0}$ continues to scale quadratically with power, maintaining a 2:1 slope as shown in Fig. 5. Once $I_{C@2\omega_0}$ exceeds $I_{G@2\omega_0}$ in amplitude, the phase of the total response $I_{tot@2\omega_0}$ shifts from the phase of $I_{G@2\omega_0}$ to that of $I_{C@2\omega_0}$. Due to the phase difference of the two components, this crossover produces destructive interference, resulting in the abrupt phase shift and dip observed in Fig. 3.

3) *Saturation Region*

For power levels above 25 dBm, the assumption that electrons are the sole contributors in (1) becomes invalid. When the signal reaches its "Min" value, hole concentration surpasses electron concentration in certain regions, leading to the increase in the local conductance minimum as shown in Fig. 6(c). As this minimum rises with increasing power, the influence of local conductance variations on C_{eq} diminishes. Consequently, A_C begins to saturate, resulting in the flattening of the response observed in the high-power region of Fig. 3.

IV. Conclusion

This paper investigates the physical mechanisms responsible for the dips observed in harmonic distortion measurements on high-resistivity silicon substrates. Through TCAD simulations, it is demonstrated that these dips arise from the phase difference between variations in equivalent conductance and capacitance within the substrate. This phase difference's physical origin stems from finite carrier inertia, that respond in a finite time that can be a significant fraction of an excitation's time-period at RF frequencies. Additionally, the saturation behavior at high power is attributed to the increasing influence of hole concentration, which becomes non-negligible at elevated input power levels.

References

[1] F. Allibert *et al.*, "Engineering SOI substrates for RF to mmWave frontends," *Microw. J.,* vol. 63, p. 72, Oct. 2020.

[2] J. C. Pedro and N. B. Carvalho, Intermodulation Distortion in Microwave and Wireless Circuits. Norwood, MA: Artech House, 2003.

[3] M. Rack and J.-P. Raskin, "RF harmonic distortion modeling in silicon-based substrates including non-equilibrium carrier dynamics," *2017 IEEE MTT-S International Microwave Symposium (IMS)*, Honolulu.

[4] M. Rack, F. Allibert and J.-P. Raskin, "Modeling of Semiconductor Substrates for RF Applications: Part I—Static and Dynamic Physics of Carriers and Traps," in *IEEE Transactions on Electron Devices*, vol. 68, no. 9, pp. 4598-4605, Sept. 2021, doi: 10.1109/TED.2021.3096777.

[5] M. Rack, F. Allibert and J.-P. Raskin, "Modeling of Semiconductor Substrates for RF Applications: Part II—Parameter Impact on Harmonic Distortion," in *IEEE Transactions on Electron Devices*, vol. 68, no. 9, pp. 4606-4613, Sept. 2021, doi: 10.1109/TED.2021.3096781.

[6] C. Roda Neve and J.-P. Raskin, "RF Harmonic Distortion of CPW Lines on HR-Si and Trap-Rich HR-Si Substrates," in *IEEE Transactions on Electron Devices*, vol. 59, no. 4, pp. 924-932, April 2012.

[7] E. H. Nicollian and J. R. Brews, MOS Physics and Technology. Hoboken, NJ: Wiley, 1982.