Integrating high frequency capacitance measurement for monitoring process variation of equivalent oxide thickness of ultra-thin gate dielectrics

Yuegang Zhao
Keithley Instruments, Inc.

Introduction

As CMOS transistors have gotten smaller and smaller, so has the thickness of their gate dielectrics. This presents a great challenge to traditional capacitance measurement used to monitor dielectric thickness for process variation. First, the relationship between the capacitance value in the inversion or accumulation region of the capacitance-voltage (C-V) curve to the gate oxide thickness is no longer simple. It’s necessary to apply new models, including quantum mechanics and polysilicon depletion effects, to determine oxide thickness accurately from the C-V curve [1, 2]. Second, gate leakage increases exponentially as thickness decreases due to tunneling of carriers through the ultra-thin gate [3]. The gate capacitor becomes very lossy due to high leakage, and the gate capacitance measurement shows roll-off effects in both the inversion and accumulation regions of the C-V curve [4]. These roll-off effects make it impossible for engineers to extract $C_{OX}$ directly and use it to monitor thickness variations in production. The roll-off behavior is also dependant on the DC leakage of the gate. Therefore, even for two gate dielectrics with the same physical thickness and area, the lower quality one with higher gate leakage will show the greater roll-off in the C-V curve, which makes it more difficult to monitor thickness variations.

Some roll-off effects in the C-V curve are device related [5, 6]. At high frequency the two main factors are channel resistance and contact resistance. These effects could be modeled by a different equivalent circuit model and could be reduced by a new device layout. On the other hand, some of the roll-offs in C-V measurement are related to non-optimized setups, including cabling, connectors, and probe station setup [7]. The first part of the paper provides a comprehensive overview of difficulties and precautions on C-V measurement on ultra-thin gate dielectrics using LCR meters at high frequencies (1–100MHz). The second part of the paper explores C-V measurement at radio frequency (RF) as one of the approaches to solving the
high leakage induced measurement problem. In general, the crossover to RFCV occurs for gate oxide equivalent oxide thickness (EOT) in the range of 1.7nm to 1.0nm.

**Error analysis**

Most of the challenges of using the LCR meters currently available for monitoring EOT variation come from getting correct capacitance measurements on very leaky gate materials. The effect of gate leakage in capacitance measurement can be represented by the dissipation factor (D) or quality factor (Q), where

\[ D = \frac{G}{\omega C}, \quad \text{and} \]

\[ Q = \frac{1}{D}. \]

\( G \) and \( C \) are the conductance and capacitance of the gate dielectrics respectively, and \( \omega = 2\pi f \), with \( f \) being the frequency of the AC stimulus. An ideal capacitor without any parasitics has an infinite \( Q \) or zero \( D \), while an ideal resistor has an infinite \( D \). As gate oxide thickness decreases to less than 2nm, the effect of higher \( D \) starts to show up in the capacitance measurement. It is not unusual to see gate capacitance have \( D \) larger than 10 or even 100 at 1MHz. The direct result of large \( D \) is a roll-off of the C-V curve in the inversion or accumulation region. Sometimes the measured capacitance value is negative [8].

Let’s quickly examine how measurement error is related to \( D \). For the simple parallel circuit in *Figure 1a*, the capacitance error can be simplified as follows:

\[ \frac{\Delta C}{C} = E_0 + \Delta \theta \cdot D. \]

Here \( D \) can be expressed as

\[ D + \frac{G}{\omega C} = \cot(\theta) \]

and \( E_0 \) denotes basic measurement error on a perfect capacitor. The definition of \( \theta \) (phase angle) is shown in *Figure 1b*. 
Figure 1. Schematic diagram of equivalent circuit and AC impedance measurements: (a) Parallel equivalent circuit model (b) Definitions of phase and amplitude of AC impedance measurements.

Eq. 1 is very important in determining errors in capacitance measurement at high D. First of all, it suggests that the measurement error is linearly dependent on D. In addition, it suggests that phase error, amplified by D, becomes the dominant source of error as D increases. There are two ways to reduce the capacitance measurement error: to increase the frequency, thereby reducing the D, or increase phase measurement accuracy, thereby reducing the phase error. Phase error comes from imperfections in the test system and measurement conditions, including cables and connectors, probes, and chuck.

High frequency C-V measurement with LCR meters

It’s not uncommon to question the minimum gate oxide thickness that current LCR meters can measure. In fact, what’s important is not how thin the gate oxide is, but how leakage the gate is. As the quality of the gate oxide differs, material differs, and technology differs, gate oxides with similar equivalent oxide thicknesses may have leakage currents that differ by several orders of magnitude. One important factor to characterize the quality of the gate oxide, as described above, is D (dissipation factor) at a certain frequency (because D is inversely proportional to frequency). Since D is directly related to measured phase (θ, as shown in Eq. 2), the principal limitation of currently available LCR meters is their inability to resolve small phase angles due to high dissipation factor. This is mainly because the test frequency is not high enough to reduce the D factor, and there is no calibration method on those LCR meters to measure the small phase angle accurately. This sets a theoretical limit on how well those LCR meters perform on thin gate oxide measurement. On the other hand, even when using a current LCR meter, the way in which the LCR meter is set up in the measurement system also affects the quality of the C-V measurement dramatically. We will briefly review some of the improvements that can be made when setting up an LCR meter, then assess the theoretical limitation of thin oxide measurement with existing LCR meters.

Cabling is very important in C-V measurements. The overall cable length in the system must be kept as close as possible to the calibration length of the LCR meter. Any deviation in physical cable length from the calibration cable length introduces phase error.
Figure 2 shows an example of the effects of different calibration cable lengths on C-V measurements on 1.3nm gate oxide. In the measurement setup, the physical cable length is close to two meters. Different cable length values are used as inputs for cable calibration and C-V curves are measured accordingly. In Figure 2, we see that capacitance measurements with small D (around 0V, since hardly any DC current flows at small DC bias) are not affected very much by variations in cable length; when D is small, the overall measurement error is dominated by $E_0$, according to Eq. 1. Phase error does not play a leading role here. On the other hand, when D is large, such as in the inversion region, where large DC current flows due to tunneling, the cable-induced phase error effect becomes significant. When the calibration length is close to the actual physical length, the inversion region is nice and flat. However, when the calibration length is shorter than the physical length, the curve starts rolling up. When the calibration length is longer than the physical length, the curve rolls off.

![Figure 2. The effect of different cable calibration lengths on C-V measurement on a 1.3nm gate oxide transistor.](image)

Proper shield jumper location is another factor in ensuring C-V measurement accuracy. Proper operation of an LCR meter requires that the shields of the coaxial cables be properly connected as close to the DUT as possible. These shields provide a current return path that compensates for parasitic inductance from the cabling (Figure 3a). If the cable shields are not tied properly, close to the DUT, there will be some parts of the cables (close to the contact point to the DUT) not returning current in the shield. This results in extra phase errors due to the inductance effect. Figure 3b shows the effect of proper shield connections on capacitance measurement, especially when D is high. As can be seen from Figure 3b, the measurement without the shield jumper shows large roll-offs in both the accumulation and
inversion regions, while with the shield jumper close to the DUT, there is a significant reduction in roll-offs.

**Figure 3.** The effect of shield jumpers on C-V measurements: (a) Correct cabling setup for C-V measurement. The red line between the shields of the Voltage Force and Voltage Sense terminals (which are labeled HV and LV, respectively) is the shield jumper that ties the shields of the cables together close to the DUT. (b) The effect of the shield jumper on the C-V measurement of a 1.3nm gate oxide. The C-V curve shows less roll-off in the inversion region with a properly located shield jumper.

For C-V measurements in production environments, probe cards for ultra-thin oxide characterization are specially designed to reduce the parasitic capacitance and inductance of wiring and probe needles. To achieve the best results, we recommend using a special probe card with minimized parasitic capacitance reserved just for thin-gate C-V measurements. When measuring a four-terminal transistor, the source, drain, and substrate are tied together in the probe card level to achieve the shortest cables possible to those terminals. Shield jumpers, which as mentioned previously are critical to capacitance measurement on leaky capacitors, are used on the probe card so that the signal path shields are tied as close to the DUT as possible.

Cable calibration is critical to successful C-V measurement on high D capacitors. Cable calibration includes open, short, and load calibrations. While both short and load calibrations require a proper test structure on the wafer (e.g., short and 50Ω load), open calibration does not. It has been found that the quality of short calibration determines the overall measurement quality. When calibrating on a short structure, there are inevitably some contact resistances. Short calibration with a high contact resistance results in noisy measurements and roll-offs in the C-V curve. Therefore, it is crucial to reduce contact resistance as much as possible during calibration. In a production environment, it is crucial to have an auto-calibration procedure. The system can be set up so that it performs calibration automatically when certain calibration criteria are met. Those criteria include the duration of
the previous calibration, whether the previous calibration failed or not, or whether it is the first time to calibrate. To be consistent, calibration is only performed on the first wafer of every cassette. The user can change calibration criteria according to specific needs.

It is well known that some roll-offs in C-V curves are due to the channel resistance of the MOSFET [6]. Figure 4 shows an example of capacitance measurements on three transistors with different gate lengths with area normalized capacitance value. Those transistors are on the same site on a wafer and very close to each other. It clearly shows that channel resistance-induced roll-off effect. Even though this effect can be compensated for by proper device modeling, it’s undesirable in a production environment. We recommend using a short channel transistor to minimize channel resistance-induced roll-off. The trade-off of using a small area transistor is that the fringing capacitance is relatively large. Fortunately the fringing capacitance can be subtracted by measuring capacitances with two different gate areas. The area of the gate should be designed so that the capacitance value to be measured is around 1–2pF. Higher capacitance values result in measurement range overload due to high leakage, while lower capacitance values result in noisy measurements due to resolution limitations of the LCR meter.

Figure 4. The effect of channel resistance on C-V measurement on 1.3nm gate oxide. Three curves represent C-V measurements of 1µm x 10µm, 5µm x 10µm, and 10µm x 10µm transistors respectively. Measurement on shorter channel length shows less roll-off in the inversion region.

Besides measurement accuracy, noise is another important factor with thin gate oxide C-V measurement. Again, based on Eq. 1, measurement noise is directly proportional to D factor. As D factor increases dramatically as gate thickness decreases, measurement noise, or
repeatability of the measurement, which was not a problem before, becomes an issue. The most common source of noise is the chuck, especially a thermal chuck. If the device under test is not isolated from the chuck, as is typically the case with an NMOS transistor in a CMOS process, chuck noise can couple into measurement and becomes obvious when D is high. Figure 5 shows the effect of measurement noise coupled with noise from a thermal chuck. There are several solutions to this problem:

- Use a better isolated, low noise chuck.
- Use higher frequencies so that D is reduced.
- Turn off the power to the thermal chuck when it’s not in use.
- Design the test structure so that the chuck is isolated from the body of the transistor.

Figure 5. Example of chuck noise coupled into C-V measurement noise at high D. The insert in the graph shows scope plots of the chuck noise.

One common misunderstanding is that the higher the test frequency used, the better the capacitance measurement on thin gate oxide will be. In principle, this is true, because D is smaller at higher frequencies (this is usually true for frequencies less than 100MHz, where the DUT can be represented by a parallel equivalent circuit). However, to implement this principle with the LCR meters currently on the market, one other important factor must be considered, which is their basic measurement accuracy at higher frequencies. This is related to the $E_0$ term in Eq. 1. $E_0$ represents the measurement accuracy on an ideal capacitor ($D = 0$). $E_0$ is a function of frequency. By reviewing the published specifications for currently available LCR meters [10], one can easily learn that the measurement accuracy degrades as frequency moves toward the high end of the frequency range (1–100MHz).
Another way to look at the problem is to combine dissipation factor, frequency, and the measurement specification for the LCR meter and draw a plot of measurement accuracy as a function of frequency. Figure 6 shows that on a gate dielectric with a $D = 10$ at 1MHz, the least error occurs at frequencies of around 1MHz. Therefore, a LCR meter with 1MHz capability should be sufficient for measuring gates with $D = 10$. For example, as shown in Figure 7 for a 1MHz C-V measurement on a 1.3nm oxide—the largest $D$ at that frequency is close to 20.

**Figure 6.** Example of specification error of an LCR meter for $D = 10$ (1MHz) across frequency.

**Figure 7.** Example of C-V measurement at 1MHz on 1.3nm gate oxide.
Following a similar approach, if a gate dielectric has a $D = 100$ at 1MHz, the sweet spot moves to a higher frequency (100MHz). However, at that frequency, the lowest measurement error rises to around 10%, which may be unacceptably high. To go even further, for a gate with a $D = 1000$ at 1MHz, which is equivalent to $D = 10$ at 100MHz, the minimum measurement error for frequencies up to 100MHz is around 30%, which makes this LCR meter completely unsuitable for this type of measurement. This exercise sets a theoretical limit on the maximum leakiness of an oxide that can be measured using the currently available LCR meters that operate at frequencies from 1 to 100MHz. Depending on the $D$ of the material being measured, this exercise may be useful in determining the level of leakage at which a specific LCR can no longer measure a particular oxide. Alternatively, it’s also possible, for a given $D$, to determine the optimal measurement frequency.

**RF capacitance measurement**

With current LCR meters, the product of phase measurement accuracy and $D$, as in Eq. 1, is limited, which limits the capacitance measurement accuracy. At 110MHz, the infrastructure, including cabling, probe card, and calibration, is similar to that of an RF measurement. It requires a full calibration set, including phase, open, short, and load calibration. At the same time, performance is again limited by the product of phase measurement accuracy and $D$. With an alternative approach, using the RF technique, one can measure capacitance at much higher frequency, in particular at operating frequency, such as 2.4GHz. Usually the measurement is done at a frequency greater than 1GHz, the point of maximum $Q$. At such frequencies, $D$ will remain relatively small for the foreseeable future (according to the International Technology Roadmap for Semiconductors) [11]. Conductance due to leakage ceases to be an issue.

The RF capacitance of the DUT is derived from the complex conductance ($Y$),

$$C = \frac{|Y|^2}{2\pi f \text{Im}(Y)}$$

which is calculated from s-parameters measured on a two-port network (Figure 8). A vector network analyzer is used to measure the RF scattering parameters. The characteristic impedance of the overall transmission line of the system is optimized for 50Ω, with a 20GHz bandwidth. RF signals from the VNA are passed to the RF probe card through a dedicated pathway. The DC bias and RF signal are mixed in the Keithley S600 Series testhead in very close proximity to the DUT. The component of complex impedance ($Z$) of the DUT can be calculated from reflection parameters $S_{11}$ and $S_{22}$,

$$S_{11} = \frac{Z_{11} - Z_0}{Z_{11} + Z_0} \text{ and } S_{22} = \frac{Z_{22} - Z_0}{Z_{22} + Z_0}$$

with $Z_0 = 50\Omega$. The frequency of measurement is selected so that the AC impedance of the DUT is close to 50Ω, because the measurement accuracy of the vector network analyzer is...
optimized around 50Ω impedance. For example, a 1pF capacitor has 50Ω impedance at around 3GHz.

**Figure 8. Schematic of a two-port network for s-parameter measurement**

The parasitics embedded in the measurement system and the DUT are part of the technical difficulties involved in using RF measurement. Significant work has been done in measurement methodology and device layout to de-embed the parasitics. *Figure 9* shows a simplified circuit model of a real transistor. The goal of RF capacitance measurement is to get $C_{OX}$. However, $C_{OX}$ is surrounded by imperfections in the physical device. Those imperfections include overlap capacitance between the gate contact and the source/drain well, gate resistance (due to poly silicon), lead inductance (from DUT to contact pads), contact resistance (between probe needle and contact pads), and channel resistance (mentioned previously). Some of the imperfections can be extracted by the de-embedding technique, especially the effects of contact resistance, lead inductance, and parasitic capacitance. DUT measurement results are corrected by subtracting those measured on de-embedding structures. Typical de-embedding structures include open, short, and thru (*Figure 10*). For short de-embedding,

$$Z = Z_{\text{meas}} - Z_{\text{short}},$$

for open de-embedding,

$$Y = Y_{\text{meas}} - Y_{\text{open}},$$

and for thru de-embedding,

$$\frac{1}{Y} = \frac{1}{Y_{\text{meas}}} - \frac{1}{Y_{\text{through}}}.$$

Combinations of two or more of the de-embeddings can also be used. It is common for open and short de-embeddings to be used together to correct both parasitic capacitance and contact resistance. Sometimes open and thru are used together when the series inductance of the DUT is not optimized.
One basic assumption of RF C-V measurement is that the characteristic impedance of the system’s transmission line is 50Ω. The closer the impedance of the transmission line to 50Ω, the better the measurement result will be. The device layout on the wafer should be adjusted to match the transmission line impedance. A ground-signal-ground structure is required for RF measurement. As mentioned earlier, channel resistance effects will show up in C-V measurements on long-channel devices, especially at higher frequencies. It is recommended that small transistors be used for RF capacitance measurements. To achieve a better signal-to-noise ratio, many small area transistors can be connected in parallel to make a large device. More details on the design of test structures for RF C-V measurement are available [9].

Contact resistance variations between consecutive probe contacts can limit the repeatability of RF C-V measurements. For example, if the DUT has a characteristic impedance of 100kΩ at 1MHz, a 1Ω variation in contact resistance will cause only a 0.01%
error. However, the characteristic impedance of the same device drops to 100Ω at 1GHz, so that same 1Ω variation in contact resistance will induce a 1% error. Most of the variation in contact resistance is due to buildup of aluminum oxide on the tip of the probe needle. These variations have been engineered out of the S600 Series through the use of automated probe cleaning. A Gage R&R study shows less than 5% variation in most cases (30% variation in Gage R&R is considered “good”). *Figure 11* shows an example of repeated RF C-V measurements.

![Figure 11. Overlay of two RF C-V measurements at 2.4GHz.](image)

Calibrations down to the probe tips are required to make accurate measurements. A full calibration set includes open, short, thru, and load calibrations. Calibration will compensate for imperfections in the transmission line, including parasitic capacitance and lead inductance on the probe card and connectors. However, calibration cannot compensate for contact resistance, because the contact resistance between the probe needle and the gold contacts on the calibration substrate is not the same as that between the probe needle and the aluminum pad on actual wafer under tests. These subtle differences are also compensated for in the S600 Series automation.

A full suite of RF capacitance measurements involves:

- Loading calibration wafer and performing calibration. This is required only if the probe card is changed or if more than 72 hours has elapsed since the last calibration. The S600 Series accomplishes this in a way that is compliant with all 300mm automation requirements.
- Loading wafers from cassette.
- Performing s-parameter measurements on de-embedding structure (once per lot).
• Performing s-parameter measurements on actual DUT. The resolution of the Keithley system is sufficient to measure a single 100fF DUT, extract gate and fringing capacitance, and correct for poly depletion.

• Outputting de-embedded results.

*Figure 12* demonstrates the correlation of RF C-V measurement results on a 13Å gate oxide at 2.4GHz with a 20MHz C-V measurement using an LCR meter. It shows excellent agreement between the two methods.

![Figure 12. RF C-V measurement on a 1.3nm gate oxide.](image)

**Comparison of high frequency and RF C-V techniques**

Comparing high frequency C-V (HFCV) and radio frequency C-V (RFCV) results only makes sense when both techniques can yield valid EOT measurements. In general, this applies to gate oxides with EOTs typically ranging from 1.7nm to 1.0nm. The actual range varies, depending on the specific process. For thick oxide, HFCV has a clear advantage because of cost and ease of use. For ultra-thin oxide, HFCV is no longer capable because of the reasons stated previously in this paper. When both HFCV and RFCV are valid options, the following comparison may help users determine the best time to migrate to RFCV based on cost, technology roadmap, and other factors.
1. At frequency >1MHz, special care has to be made for cabling and connection (such as a dedicated signal pathway), an RF-like calibration suite, such as open, short, load, has to be deployed to calibrate down to the prober tip.
2. See [12].
3. See [8].
4. DC bias is provided by a source measure unit through a bias Tee, while RF measurement is AC coupled through the bias Tee using a Vector Network Analyzer.

**Conclusion**

Common C-V measurement errors with currently available LCR meters are discussed, as well as their limitations in making C-V measurements on ultra-thin gate oxides. Techniques to enhance an LCR meter’s performance to near its theoretical limits, such as cabling, probe card, and connectors, are discussed. The RF C-V technique is discussed and deployed in production environment to monitor EOT variations for ultra-thin gate oxide. New options available make the S600 Series the ultimate tool for monitoring EOT variation in production environments for the current technology node, as well as for several future technology nodes.
References


[10] Agilent 4294 LCR meter specification, document #5968-3809E.


About the Author

Yuegang Zhao is a Senior Applications Engineer with the Semiconductor Business Group of Keithley Instruments in Cleveland. He received his Masters Degree in Semiconductor Physics from the University of Wisconsin, and his B.S. in Physics from Peking University, Beijing, China. He has six years of experience in semiconductor physics and device characterization.