Transient Modeling of Ferroelectric Capacitors for Nonvolatile Memories
Ali Sheikholeslami and P. Glenn Gulak, Member, IEEE

Abstract—Present ferroelectric (FE) capacitor models mostly rely on continuous hysteresis loop characteristics of FE materials. Our experimental results show that this approach overestimates the remanent and saturation polarizations available for nonvolatile semiconductor memories by more than 50%. A behavioral transient model based on pulse measurement results is proposed and implemented as an HSPICE macro-model. The model mainly consists of two nonlinear capacitors, corresponding to the two different polarization states of an FE capacitor.

I. INTRODUCTION

FERROELECTRIC (FE) capacitors have long been recognized as suitable storage elements for nonvolatile memory cells [1], [2]. Low-density ferroelectric random access memories (FRAM's) (such as 512 × 8 b and 4 k × 8 b) are now commercially available and compete with both electrically erasable programmable read-only memories (EEPROM's) and battery-backed static random access memories (SRAM's) [3]. Higher density FRAM's (such as 32 k × 8 b, 100 ns access time) have been successfully designed and tested [4], and the search continues to provide higher density FRAM's with faster read/write time and higher reliability. Meanwhile, FE circuit simulation and optimization are hindered by a lack of accurate CAD circuit models. As a result, current design methods are primitive, and FE circuits are usually designed and margined independently of the actual transient characteristics of the FE capacitor. In fact, the best modeling techniques used presently in memory circuit simulation for this technology only provide for static dc modeling; few transient models have yet been proposed [5].

The complex behavior of an FE capacitor in response to arbitrary voltage patterns is perhaps the main difficulty towards universal modeling of this element. From the circuit point of view, the FE capacitor is a nonlinear element with memory and, hence, has the complexity of both. A nonlinear but memoryless capacitor can be modeled behaviorally by a piecewise linear curve in the Q–V plane (charge versus applied voltage). In any voltage interval bounded by two consecutive breakpoints, the nonlinear capacitor would be equivalent to a series combination of a linear capacitor and a voltage source. The accuracy of such model depends on the number of breakpoints chosen in approximating the nonlinear curve, that is, the more breakpoints, the more accurate the model. A nonlinear with memory capacitor, on the other hand, is modeled by several piecewise linear curves. Each curve represents a nonlinear capacitor corresponding to a state of the capacitor memory. It is evident, therefore, that the larger the memory depth (the number of memory states), the more complex the model. In fact, the number of measurements required for behavioral modeling of a nonlinear device is equal to the number of breakpoints chosen to approximate a nonlinear characteristic multiplied by the number of states.

For an FE capacitor, the number of states can be very large if a general model is intended. However, only a few states need to be considered if modeling is employed for conventional memory applications such as a 1T–1C FRAM cell. Fig. 1 shows the circuit diagram of a 1T–1C memory cell activated through wordline, and written or read through bitline and driveline. In writing a binary digit 0 to the cell, a positive voltage (normally the full power supply, $V_{DD}$) is applied to bitline while the driveline is grounded and the wordline is asserted. The voltage waveform that appears across the capacitor is a positive square pulse with amplitude $V_{DD} - V_T$, where $V_T$ is the threshold voltage of the access transistor. In writing a binary digit 1, a positive voltage ($V_{DD}$) is applied to the driveline while the bitline is grounded and the wordline is asserted. The voltage waveform across the capacitor in this case is a negative pulse with amplitude $V_{DD}$. The voltage waveforms for the read operations are also pulses with similar

![Fig. 1. A 1T–1C FRAM cell.](image-url)
rise and fall times, but with possibly different amplitudes [6]. For example, in [6] it is explained that the capacitor can also experience $2/3 V_{DD}$ depending on the read–write sequence. It is this limited variety of pulse amplitudes that helps simplify the behavioral modeling of FE capacitors for memory applications.

In this paper, we illustrate a modeling procedure based on the above approach. For simplicity, only two distinct polarization states are considered in these examples. However, the procedure is easily extensible to modeling with a larger number of states. Fig. 2 depicts the two polarization states on a hysteresis loop along with its corresponding state transition diagram. According to this diagram, the next state of the capacitor is determined by its current state and the polarity of the applied pulse. The capacitor remains in the positive polarization state if the input pulse has a positive polarity, or changes to a negative state if the input pulse has a negative polarity. Similarly, the capacitor remains in the negative polarization state if the input pulse has a negative polarity, or changes to a positive state if the input pulse has a positive polarity.

A two-state transition diagram forms the basis of our proposed model. Accordingly, it consists of two distinct parts corresponding to two distinct states, and an algorithm that dictates switching from one state to another depending on the input pulse polarity. In Sections II and III of this paper, we discuss the details of an experimental procedure that leads to the construction of each part. Next, in Section IV, we propose the model by incorporating the experimental results into the state transition diagram. As will be seen in the same section, such a model easily lends itself to an HSPICE macro-model implementation. Finally, some simulation results generated by this model will be presented.

II. EXPERIMENTAL PROCEDURE

There are two types of circuits that are used for behavioral characterization of an FE capacitor: the Sawyer–Tower circuit and the pulse measurement circuit [1], [2]. The Sawyer–Tower circuit is normally used to record the hysteresis loop of an FE capacitor, that is, the polarization variation in response to a low-frequency sinusoidal voltage waveform. The hysteresis loop illustrates a periodic switching of polarization in response to an applied periodic voltage waveform. On the other hand, the pulse measurement circuit is used to record the transient current of an FE capacitor in response to nonperiodic one-shot pulses with short durations (on the order of the circuit time constant). As such, the pulse measurement results convey much more information with regard to FE memory design where the capacitor experiences one-shot read/write pulses with short durations. At the same time, the pulse measurement procedure is more involved and requires special considerations. We present some of these considerations through the description of our measurement setup.

Our hysteresis loop measurement setup is a Sawyer–Tower-based circuit with a 10-nF integrating capacitor. A continuous 1-KHz sinusoidal waveform is applied to the circuit, and the voltage across the integrating capacitor is monitored by a 300-MHz 1-Gsamples/s digitizing oscilloscope. The charge on the 10-nF capacitor at the peak value of the input waveform is divided by the FE capacitor area to find the polarization parameter $P_f$. Similarly, the charge corresponding to the zero value of the input waveform is divided by the FE capacitor area to find $P_r$.

The pulse measurement circuit consists of a series combination of an FE capacitor and a small resistor. As shown in Fig. 3, the circuit is driven by a programmable pulse generator through a coaxial cable. All coaxial cables used in this setup have 50-Ω intrinsic impedances and are terminated at one end by a 50-Ω resistor. This eliminates signal reflections from the terminal nodes (i.e., the scope channels and the pulse generator terminal). This consideration is particularly important for transient measurements, as the delay produced by the cable is comparable with the signal rise time. The signal delay per unit length (one meter) of the cable is 5 ns while the signal rise time is only 1.8 ns. Therefore, the reflected signal would be in the period of interest for the transient analysis. For a fully matched setup, there are no reflections from the scope terminals. The signals captured by the scope are only delayed versions of signals at nodes $X$ and $Y$. These delays can be made equal by using the same length for both cables. On the other hand, the signal reflections from the capacitor nodes are unavoidable due to both capacitive and nonlinear characteristics of the FE capacitors. In fact, even if the FE capacitor were replaced by a perfectly linear resistor, the signal reflections from node $X$ would still occur.

We have used FE capacitor samples (PZT: $Zr/Ti = 40/60$) from two different vendors, which we refer to as capacitor $C_A$ and capacitor $C_B$ throughout this paper. Capacitor $C_A$ is $80 \times 80 \mu m^2$ while capacitor $C_B$ is $100 \times 100 \mu m^2$ in size. Since the RC time constants of the measurement circuit in these
two cases are unequal, two different sets of pulse durations are chosen for the small and the large size capacitors. The pulse durations are not critical in our measurements provided they are long enough to ensure any transients have died down within these time periods. The voltage pattern is shown in Fig. 4. The initial negative pulse brings the capacitor to its negative state (refer to Fig. 2). After a few seconds, which is the time required for the computer to initialize the pulse generator, two positive pulses are applied. The first positive pulse stimulates both the switching and the nonswitching current to flow through the capacitor. The second positive pulse only stimulates the nonswitching current [7].

The experiment is repeated 20 times, using 20 different pulse amplitudes uniformly distributed between zero volts and the full supply voltage. To limit the maximum field applied to our capacitor samples of 2500-Å thickness, 10 V is chosen as the maximum pulse amplitude voltage.

The measurement results are expected to be repeatable if both the dielectric thickness and the voltage amplitude are scaled down by the same factor. In fact, the electric field remains unchanged by these scalings. Meanwhile, note that the film thickness cannot be reduced beyond a limit imposed by the process yield.

III. RESULTS AND DISCUSSIONS

Three parameters of interest to FE nonvolatile memory design are the remanent polarization, the saturation polarization, and the switching time of the FE capacitor. The first two parameters can be derived from both hysteresis loop and pulse measurements whereas the last parameter can only be derived from pulse measurements. It will be shown in this section that the hysteresis loop-based polarization can overestimate these parameters by more than 50%. The switching time is argued to be negligible in comparison with the RC time constant of a minimum size FE memory cell.

Two sets of experimental results are shown in Fig. 5. Fig. 5(a) shows the transient current of capacitor \( C_A \) corresponding to the first and second positive pulses as captured by channel B of the scope. Fig. 5(b) shows similar results for capacitor \( C_B \). Comparing Fig. 5(a) and (b), the transient current of capacitor \( C_A \) overshoots while the other one decays smoothly to zero. The main reason for this difference lies in the relative sizes of the capacitors. The measurement circuit is an effective resistor inductor capacitor (RLC) circuit due to the inductance of the probe tips. Such a circuit can cause three types of behavior, namely, underdamped behavior (oscillatory response), critically damped behavior, and overdamped behavior (nonoscillatory response). An oscillatory response is caused by capacitances less than a critical value, determined by the series resistance and inductance in the measurement circuit. Likewise, a nonoscillatory response is caused by capacitances greater than the critical value. In our measurement circuit, the capacitance value of capacitor \( C_A (C_B) \) happens to be below (above) the critical value. Therefore, capacitor \( C_A (C_B) \) shows oscillatory (nonoscillatory) transient behavior. It is expected, however, to observe nonoscillatory transient currents for capacitor \( C_A \) with larger areas. It is also expected to observe nonoscillatory results if less inductive probe tips are used in the measurement.

Despite the transient current dependence on the inductance value in a series RLC circuit, the final charge stored on the capacitor is independent of the inductance value and can be calculated by integrating the transient current over the pulse interval. Referring to Fig. 5(a), the integral of the first pulse current over the interval (50 ns, 150 ns) represents the sum of the switching and the nonswitching charge of \( C_A \), while the integral of the second pulse current over the same interval represents only the nonswitching charge of \( C_A \). We denote by \( P_1 \) and \( P_0 \), respectively, the integral of the first and second pulse currents over the mentioned period divided by the FE capacitor area. The integrals are calculated numerically using the trapezoidal method. Similarly, \( P_1 \) and \( P_0 \) are calculated for \( C_B \) with (100 ns, 400 ns) as the time interval [refer to Fig. 5(b)]. Note that \( P_1 \) (the available polarization increment for the digital 1 state) corresponds to \( P_s + P_r \) while \( P_0 \) (the available polarization increment for the digital 0 state) corresponds to \( P_s - P_r \) of the hysteresis loop (refer to Fig. 2). \( P_s \) and \( P_r \) can be expressed in terms of \( P_1 \) and \( P_0 \) by the following:

\[
P_s = \frac{1}{2}(P_1 + P_0) \quad \text{and} \quad P_r = \frac{1}{2}(P_1 - P_0).
\]

In short, the capacitor size is not of critical importance as far as the remanent and the saturation polarizations are concerned. The switching time measurement, on the other hand, is quite sensitive to the inductance value, especially when the FE capacitance is comparable to its critical value. In this case, it is important to choose smaller capacitors and ac probe tips with less inductance.
Based on the above arguments, the polarization parameters, \( P_1 \) and \( P_0 \), are the only valid data that can be extracted from our pulse measurement results. These parameters are extracted as functions of the pulse amplitude and compared with their corresponding values derived from hysteresis loop measurement in Fig. 6. Fig. 6(a) compares \( P_1 \) and \( P_0 \) for capacitor \( C_A \) as derived from pulse measurements (dashed curves) with their counterparts derived from hysteresis loop measurements (solid curves). Fig. 6(b) illustrates similar parameters for capacitor \( C_B \).

First, note that the polarization values for capacitor \( C_B \) are at least twice as high as the polarization values for capacitor \( C_A \). At 8-V pulse amplitude, for example, \( P_1 \) and \( P_0 \) for capacitor \( C_B \) are 2.5 and 2.6 times their corresponding values for capacitor \( C_A \). Also note that at the same input voltage, \( P_r \) is four times larger for capacitor \( C_B \). These are two significant advantages of capacitor \( C_B \).

Second, note that the polarization parameters for capacitor \( C_B \) saturate at around 10 V while capacitor \( C_A \) behaves almost linearly in the same range. This indicates that higher polarization values are achievable by capacitor \( C_A \) at higher voltages. Higher polarizations can also be obtained by reducing the thickness in capacitor \( C_A \).

Finally, note that the hysteresis loop characteristic overestimates \( P_1 \) by 85% and 104% for capacitors \( C_A \) and \( C_B \), respectively. This overestimation is caused by the longer period of the applied signal (1 ms instead of 100 ns), as well as its periodicity.

A. Switching Time

The switching speed of an FE capacitor from one polarization state to another is a figure of merit for the capacitor in nonvolatile memory applications. This speed is normally characterized by the switching time of the capacitor. In this section, we investigate practical limitations in the measurement of this parameter.

Switching time is defined as the time required for a constant applied voltage to switch the polarization state of the capacitor. In practice, switching time is difficult to measure since it is much smaller than both the RC time constant of the measurement circuit and the rise time of the applied pulse. In fact, by the time the voltage across the capacitor is constant, the switching has already been completed. Therefore, it remains ambiguous as to what fraction of this time is due to the intrinsic switching time and what fraction is due to the pulse rise time and the RC time constant. A study by Larsen et al. on PZT capacitors [8] demonstrates the switching time as a function of the capacitor area. As the area of the capacitor decreases (equivalent to reducing the RC time constant of the circuit), the measured switching time decreases. If this function is extrapolated to where the capacitor area is zero, a switching time of 1.8 ns can be predicted. This value is exactly equal to the rise time of the applied pulse. Therefore, the switching time measurement is restricted by the pulse rise time, which is an instrument limitation. The inherent switching time (i.e., the switching time when not restricted by the circuit RC time constant and the instrument limitation) is expected to be around 100 ps [9]. This time is at least one order of magnitude less than the typical RC time constant of a one-transistor memory cell (refer to Fig. 1). In 0.8-\( \mu \)m CMOS technology, for example, such time constants are around 10 ns, assuming a 1-pF bitline capacitance and 2:1 ratio for the bitline capacitance to the cell capacitance [6].

Based on the above arguments, we neglected the switching time in comparison with the RC time constant of the cell. The following calculation shows that this approximation underestimates the memory access time by less than 10% if the number of cells per bitline is greater than 10. Each cell typically contributes 4 fF to the bitline capacitance, including the diffusion and the metal parasitic capacitances. For 10 cells, the total capacitance on the bitline will be 40 fF. The 2:1 ratio [6] results in 20 fF for the FE capacitor and, hence, 300 ps for the RC time constant. Assuming 15 k\( \Omega \) as the access transistor ON resistance and a minimum of three time constants for the signal development time, the access time exceeds 900 ps. This must be compared with 1000 ps if the switching time is included. This error reduces to 0.78% in a 128 cells per bitline memory and 0.39% in a 256 cells per bitline memory.

Assuming zero switching time, on the other hand, greatly simplifies the transient analysis of the FE capacitor, and also reduces the measurement complexity that otherwise is required. We present this analysis in the next section where we will introduce our proposed model.

IV. ZERO SWITCHING-TIME TRANSIENT (ZSTT) MODEL

In the last section, we mentioned that the switching time of the FE capacitor can be neglected in comparison with the RC time constant of an FE memory cell. The transient model introduced in this section is based on this assumption and, therefore, called the zero switching-time transient (ZSTT) model. Zero switching-time implies that the charge increment on the FE capacitor will reach its final value instantaneously.\(^1\) Therefore, the charge increment is only a function of the applied voltage and the initial state of the capacitor, not a

\(^1\) Although an instantaneous charge increment implies an infinite current, this never occurs in a practical circuit. This is because the current is always limited by an external resistance such as the ON resistance of the access transistor.
Fig. 7. The characteristics of the two nonlinear capacitors representing capacitors (a) $C_A$ and (b) $C_B$ in the ZSTT model.

Fig. 8. An implementation of the ZSTT model.

The function of time. In other words

$$\Delta Q = \Delta Q(V, Q_{\text{init}})$$

where $V$ and $Q_{\text{init}}$ represent the applied voltage and the initial polarization charge on the capacitor. Since there are only two initial states that are important to a memory cell (i.e., digital 0 and 1), the above equation can be broken into two parts, each corresponding to one initial state. In other words

$$\Delta Q = \begin{cases} 
\Delta Q_0(V) & \text{for digital 0 state} \\
\Delta Q_1(V) & \text{for digital 1 state.}
\end{cases}$$

Each part of (3) represents, in general, a nonlinear function of voltage that must be determined. Let us first determine $\Delta Q_0$. A digital zero state is defined to be the positive polarization state of the capacitor. A positive voltage, in this case, results in a polarization increment of $P_0$ (or, equivalently, $P_a - P_r$) whereas a negative voltage results in a charge decrement of $P_1$ (or, equivalently, $P_a + P_r$). Recall from the last section that both of these parameters are functions of the applied voltage and can be determined experimentally. Also note that a polarization increment is equivalent to a charge increment if multiplied by the capacitor area. Therefore, $\Delta Q_0$ can be expressed as a function of $P_a$, $P_r$, and the capacitor area by the following:

$$\Delta Q_0(V) = A\{P_a(V) - P_r(V)\}u(V) + A\{P_a(V) + P_r(V)\}u(-V)$$

where $A$ represents the capacitor area and $u(V)$ represents the unit step function. In this equation, we have assumed both $P_a$ and $P_r$ are odd functions of the applied voltage. Cases in which this assumption is not valid can be treated by minor modifications. Equation (3) can be further simplified as

$$\Delta Q_0(V) = A\{P_a(V) - \text{sgn}(V)P_r(V)\}$$

where $\text{sgn}(V)$ represents the signum function that is equal to $-1$ for negative values and $+1$ for positive values of $V$. A similar argument for the digital 1 state results in

$$\Delta Q_1(V) = A\{P_a(V) + \text{sgn}(V)P_r(V)\}.$$
five break points in the full domain of $\Delta Q_0$ and $\Delta Q_1$. These functions are shown in Fig. 7(b).

The piecewise linear approximation of $\Delta Q_0(V)$ and $\Delta Q_1(V)$ results in two piecewise linear capacitors that can be implemented in most circuit simulators. We present an HSPICE implementation of the model in the next section.

A. Model Implementation

A circuit representation of the ZTT model is shown in Fig. 8. $C_0(V_0)$ and $C_1(V_1)$ represent the two nonlinear capacitors corresponding to the two binary states of the FE capacitor (discussed in the last section). If the binary state of the FE capacitor is 0, switches $S_{01}$ and $S_{12}$ are closed while switches $S_{11}$ and $S_{02}$ are open. In this case, the equivalent capacitance looking into the input terminals of $C_0(V_0)$. Meanwhile, a voltage-controlled voltage source, which is equal to $V_1$, is connected to $C_1(V_1)$ to initialize this capacitor for the opposite binary state. For binary state 1, the states of the switches are the reverse: switches $S_{11}$ and $S_{02}$ are closed while switches $S_{01}$ and $S_{12}$ are open. Therefore, the capacitance looking into the input terminals is $C_1(V_1)$, and the controlled source is connected to $C_0(V_0)$ for the similar reason mentioned above.

If the initial binary state of an FE capacitor is known, its subsequent binary states can be determined by the applied pulse pattern. A binary state 0 remains unchanged in response to a positive pulse as well as the leading edge (falling edge) of a negative pulse. It only changes on the trailing edge (rising edge) of the negative pulse. A binary state 1 remains unchanged in response to a negative pulse as well as the rising edge of a positive pulse. One example of an input voltage pattern and its corresponding control signals for the switches are illustrated in Fig. 9. The initial binary state of the capacitor, in this example, is considered to be 0. Therefore, $S_{01}$ and $S_{12}$ are initially closed, and the equivalent capacitance looking into the input terminals is $C_0(V_0)$. On the rising edge of the first negative pulse, switches $S_{11}$ and $S_{02}$ are closed and switches $S_{01}$ and $S_{12}$ are open. In this case, the capacitance looking into the input terminals is $C_1(V_1)$. This situation will be reversed again on the falling edge of the first positive pulse.

The ZTT model can be easily implemented as an HSPICE macro-model if the two nonlinear capacitors in Fig. 8 are replaced by their piecewise linear approximations. The switches can be replaced by voltage-controlled resistors (VCR’s) that exhibit high and low resistances for the open and closed states of the switches, respectively. The controlling voltages for these switches can be derived by the method described above.

V. SIMULATION RESULTS

The ZTT transient model can be used to simulate the electrical behavior of an integrated FE capacitor in any of the standard memory configurations. In this section, we present an example of such a simulation. In particular, we use this model in conjunction with the IT-1C FE memory cell (Fig. 1) to simulate an FRAM read operation.

A typical decision that must be made during the course of any FRAM design is the selection of an appropriate capacitor area given certain design parameters. A transient simulation of a read operation is useful in making this selection. We wish to determine (for capacitors $C_A$ and $C_B$) the area that results in an identical 100-mV differential signal on a 1-pF bitline.

As mentioned earlier, capacitors $C_A$ and $C_B$ were fabricated for test purposes. The dielectric thickness, in both cases, is optimized for 10-V supply operation. However, it is expected that the capacitors would behave similarly if the thickness was decreased for 5-V supply operation. Therefore, we assume that the horizontal axes in Fig. 6 can be scaled down to 5 V without changing the polarization parameters.

As shown in Fig. 10, the wordline (WL) and driveline (DL) in both designs are driven by identical pulses, having 1-ns rise time. The capacitor sizes are then optimized, an incremental process requiring several simulation runs, to obtain the same differential voltage (100 mV) on the bitline for the digital 1 and 0 stored data. The scaled sizes are predicted to be 11 $\mu$m$^2$ and 5 $\mu$m$^2$ for capacitor $C_A$ and $C_B$, respectively. Therefore, an area saving of 54% is expected by using vendor B capacitors.

Fig. 10 illustrates the final simulation results.
VI. CONCLUSION

In this paper, we described the experimental procedure for device characterization of FE capacitors. Our experimental results show that the continuous hysteresis loop characteristic overestimates the remanent and saturation polarizations available for a nonvolatile memory by more than 50%.

The switching time of the FE capacitor can be neglected in comparison with the RC time constant of a memory cell. Based on this argument and pulse measurement results, a new transient model (the ZSTT model) was proposed and implemented as an HSPICE macro-model. The model was used to optimize the capacitor sizes for a read operation in a 1T-1C FRAM design.

ACKNOWLEDGMENT

The authors would like to thank D. Somppi, S. Wood, V. K. Chivukula, and P. Leung for their cooperation in the experimental part of this research.

REFERENCES


Ali Sheikholeslami was born in Mashhad, Iran, in 1966. He received the B.Eng. degree from Shiraz University, Shiraz, Iran, in 1990, and the M.A.Sc. degree from the University of Toronto, Canada, in 1994, both in electrical engineering. Currently, he is a Ph.D. candidate in the Department of Electrical and Computer Engineering, University of Toronto, where his research is focused on ferroelectric memory design and modeling, very large scale integration (VLSI) memory design, content-addressable memories, and multiple-valued logic.

P. Glenn Gulak (S'82-M'83) received the Ph.D. from the University of Manitoba, Canada, while holding a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship. From 1985 to 1988 he was a Research Associate in the Information Systems Laboratory and the Computer Systems Laboratory at Stanford University, Stanford. Currently, he is a Professor in the Department of Electrical and Computer Engineering at the University of Toronto, Canada. His research interests are in the areas of circuits, algorithms, and VLSI architectures for digital communications and signal processing applications.

Dr. Gulak has received five teaching awards for undergraduate courses taught in both the Department of Computer Science and the Department of Electrical and Computer Engineering at the University of Toronto. He has served on the ISSCC Signal Processing Technical Subcommittee since 1990 and currently serves as Program Secretary for ISSCC.