A Review of Recent Research Works on Negative Capacitance Field Effect Transistor

Muhibul Haque Bhuyan*

Department of Electrical and Electronic Engineering
Southeast University, Tejgaon I/A, Dhaka 1208, Bangladesh

Abstract
The size of CMOS technology is being continuously scaled down in the nano scale regime. But the Boltzmann’s electron distribution creates a major obstacle to it and causes to reach a limit, i.e., the sub-threshold slope value of 60mV/decade attainable at normal room temperature. Reduction of power consumption has become a critical challenge for computational circuits and thus restricts the processing speed of data rate. Negative capacitance transistor proposes to break this limit of SS further down. A Negative Capacitance Field Effect Transistor engages a ferroelectric material in the gate region of the device structure and thus provides an effect of negative capacitance. As such internal voltage amplification occurs and thereby reduces the sub-threshold slope. In this paper, construction and the physics based working theory behind the NCFET structures, motivation towards the research works on NCFET and their comparative attainment etc. will be discussed and reported elaborately.

Keywords: NCFET, MOSFET, CMOS, Computational Speed, Negative Capacitance Transistor.

I. Introduction
As per aviation report, the IT industry emits 2% of global CO₂ emissions, which is equivalent to about 7% of the total electricity used in the world. But this figure may be doubled by 2030. We all know that conventional FETs require a change of 60 mV of channel potential to conduct a current of 10 orders of magnitude at room temperature. This means that minimum channel potential can’t be lowered than this thermionic limit, which is commonly known as sub-threshold tyranny. Negative Capacitance Field Effect Transistor (NCFET), as an evolving FET architecture, is being considered to alleviate this thermionic limit. To manufacture the ultra-low-power consumer electronics products, NCFET is thought as a good candidate.

Boltzmann electron distribution limit is considered unavoidable for the conventional Metal Oxide Semiconductor Field Effect Transistor (MOSFET). Already this has set the lowest boundary of the sub-threshold slope of 60 mV/decade (Y. Taur and T. H. Ning, 1998; A. P. Chandrakasan and R. W. Brodersen, 1995). Thus further reduction of the power consumption by the device has become an impossible task and hence making the energy efficient devices as well as the huge demanding ultra-low-power applications in the areas of Internet-of-Things (IoT) and wearable computing devices. NCFET is a promising replacement for CMOS device to eradicate the problem of the lowest value of sub-threshold slope of 60 mV/decade at room temperature by using the ferroelectric materials and as such increasing the gate voltage drop internally (S. Salahuddin and S. Datta, 2008; V. V. Zhirkov and R. K. Cavin, 2008; M. Kobayashi and T. Hiramoto, 2015). Two-dimensional (2D) transition metal dichalcogenides (TMDs) have shown superior resistance to short channel due to their thinness at atomic level (B. Radisavljevic et al., 2011; S. B. Desai, et al., 2016; L. Liu et al., 2013; M. Chhowalla et al., 2016) and this has demonstrated steeper sub-threshold slope over a wide voltage range applied to the gate of NCFET structure (H. Ota et al., 2017; A. Sharma and K. Roy, 2017; F. A. McGuire et al., 2016). NCFETs with TMDs as channel material and ferroelectric hafnium zirconium oxide (HZO) (M. Si et al., 2018; F. A. McGuire et al., 2017; M. Si et al., 2018-2) or polymer (F. A. McGuire et al., 2016; X. Wang et al., 2017) as ferroelectric gate have been found in many literatures so far. Layered ferroelectric materials with flat surface at atomic level can provide pronounced performance and elevated reliability of NCFETs over the bulk ferroelectric materials by optimizing the dangling
bonds and interface traps that come from charged impurity atoms. A few types of 2-D layered materials were hypothetically demonstrated or experimentally determined as the ferroelectric materials (F. A. McGuire et al., 2016). Of these, CuInP2S6 (CIPS) was found with the switchable polarization property down to 4 nm of thickness at room temperature (F. Liu et al., 2016).

Prof. Sayeef Salahuddin and Prof. Supriyo Datta of Purdue University, USA hypothesized about NCFET in 2007 (S. Salahuddin and S. Datta, 2008). Continuous downscaling of silicon transistors were being made so that more transistors could be jam-packed onto a smaller chip. But at one stage, silicon transistors approached to its fundamental physical limits. As a result, major semiconductor industries became concerned over it and tried to explore the new methods to enhance chip design without changing basic structure of the transistor but modifying the existing design a little bit. NCFETs came to rescue silicon transistor’s race and persist the enhancement of microprocessors till 2030s. Hence improvement in clock speed is also observed. This has come as a blessing for those engineers who work on finite element analysis and computational fluid dynamics for more processing speed.

In the subsequent section of this paper, we will discuss about the problems of the conventional transistors, define the sub-threshold slope, explain ferroelectricity and negative capacitance effect, describe construction and physics based working principles of NCFETs, provide a comparative summary of NCFETs with their attainment level of sub-60 mV/decade sub-threshold slope (SS).

II. Problems of Conventional Transistors

With the conventional transistors, the chip manufacturers are facing a lot of problems. Leakage current is one of them. Due to the shrinking of transistor’s size, off-state current (I_{off}) that flows through the channel of the transistor when the device remains idle and no voltage is applied on the gate, has increased steadily. Thus power is wasted as heat energy causing the CPU to become hot requiring extra power for the cooling fan to dispose extra heat from the system and hence needs further power.

Though multicore processors can diminish the leakage current by switching off the idle cores but leakage current starts to flow if these idle cores are switched on again.

Another problem of the conventional transistor is the voltage reduction with chip size reduction. We know that threshold voltage (V_{th}) is required at the gate to switch the transistor on. In each generation of CPUs, this V_{th} has declined because of the reduction of the transistor size. But the applied voltage hasn’t been scaled down proportionately with the size of the transistor. This has created a difficult situation as the power used by the chip is directly proportional to the square of the applied voltage. That means, more transistors are being switched ON/OFF in a comparatively smaller area. Thus the power density of the chips have increased manifold, and thus keeping the chips cool became challenging than before and hence increasing the complexity of the modern CPU’s cooling systems.

Short Channel Effects (SCE) arises due to the threshold voltage reduction with the shortening of channel length of the FET (M. H. Bhuyan, 2011). In the short channel regime of the FET, size of the depletion layer becomes similar to its depletion width and as such, quantum mechanical phenomena, such as, quantum tunneling effect plays a dominating role. Therefore, shrinking of transistor size is not an only option to get higher speed, higher packing density and lower power dissipation. The alternative way is to find new architectures of the transistor. Using an additional layer of ferroelectric material for getting the negative capacitive effect might provide a better solution to eradicate this problem.

III. Sub-threshold Slope

For the conventional MOS transistors, it is assumed that the drain to source current becomes zero when gate voltage is less than the threshold voltage (V_{th}) of the transistor. But in practice it doesn’t happen rather when the gate voltage is smaller than the threshold voltage, the drain to source current in the channel decreases exponentially to zero that yields off state current or sub-threshold current. In the drain to source current (I_{DS}) vs. gate to source voltage (V_{GS}) characteristic curve (Fig. 1), this region is known as sub-threshold region and the slope of this regime is known as the sub-threshold slope (SS).
The sub-threshold slope (SS) of the transistor can be defined as $\frac{d(V_{GS})}{d(\log I_{DS})}$, i.e., the amount of gate voltage necessary to change the drain to source current by one decade. This slope depends on the channel depletion capacitance. It has been proved that the sub-threshold slope can be written as-

$$SS = \frac{dV_{GS}}{dI_{DS}} = \left(\frac{kT}{q}\ln 10\right)\left(1 + \frac{C_D}{C_i}\right)$$  \hspace{1cm} (1)

where $C_D$ is the depletion capacitance of the channel and $C_i$ is the insulator capacitance of the gate. At room temperature ($T = 300$ K), we get-

$$SS = 60 \times 10^{-5} \left(1 + \frac{C_D}{C_i}\right)$$  \hspace{1cm} (2)

But we know that $C_D << C_i$, thus theoretical limit of $SS$ is 60 mV/decade (G. McFardland and M. Flynn, 1995).

If we can reduce the ratio of current at ON and OFF states (i.e., $I_{ON}/I_{OFF}$) then we can reduce the leakage current and with the reduction of leakage current the power dissipation can also be reduced. Thus the current ratio needs to be as low as possible, i.e. the sub-threshold slope needs to be steeper. Different types of techniques have been applied to alleviate this problem. For example, impact ionization, tunneling and positive feedback mainly by amplifying the gate voltage (C. W. Yeung et al., 2012).

We know that MOS transistor is designed in such a way that when we apply gate voltage above the threshold voltage, it becomes turned ON, i.e., current starts to flow through the device. When the gate voltage level is well below the threshold voltage of the device then ideally the device current should remain zero. However, in practice, this is not zero. A small leakage current flows in the device and this current is called sub-threshold current. However, if the supply voltage is not scaled down with the gate oxide thickness, electric field in the oxide region increases. This increased electric field in the gate oxide region degrades the oxide layer gradually and thereby causing dielectric breakdown ultimately (K. Yamabe, 1985). However, to enhance reliability and life-time, gate oxide thickness cannot be made extremely thin. Therefore, we need improvisation in the gate oxide region so that sub-threshold slope can be reduced further so that the power dissipation and improved reliability and performance as well as longer life time of the MOS transistor can be ensured.

**IV. Fabrication Method of NCFET**

The key fabrication steps of Ge NCFETs are described in this section from the literature J. Li et al. (2019) in Fig. 2.

![Figure 2: Schematic diagram of an NCFET (J. Li et al., 2019)](image_url)

Four-inch n-type Ge wafers on (001) plane with a resistivity of 0.088–0.14 Ω-cm have been used as the preliminary substrate. Then pre-gate cleaning is done. After that Ge wafer is placed into an ultra-high vacuum chamber for two rounds of surface passivation using Si2H6 for the time duration of 40 and 60 minutes respectively to improve the channel current. After the completion of this process, TaN/HZO/TaN/HfO2 stack is deposited with the thicknesses of 4.35 and 4.5 nm for the HfO2 dielectric layer and HZO ferroelectric layer respectively. Then gate patterning and etching steps are done and source/drain (S/D) regions are implanted using boron ions (B+) at 30 keV of implantation energy with ion dose of $1 \times 10^{15}$ cm$^{-2}$. It is to be noted that as the substrate is n-type Ge, therefore p-type
dopant (boron) is used for both source and drain regions. Besides, metal layer with Ni is made using a lift-off process over the source/drain (S/D) regions. As a final task, rapid thermal annealing process is carried out for the duration of 30s at 450°C of temperature. Thus the NCFET structure is fabricated as shown in the schematic diagram of Fig. 2. The internal metal gate (TaN) in between HZO and HfO2 in the fabricated NCFET structure counterbalances the channel potential at the surface of the device.

V. Ferroelectricity and Negative Capacitance Effects

There are two types of non-linear dielectric materials, one is called paraelectric and the other is called ferroelectric. In paraelectric material, there is no polarization when electric field is removed. But in ferroelectric material, there are two possible states of polarization when electric field is removed. To get ferroelectricity, we need ferroelectric materials. This type of material has nonlinear polarization (i.e., separation between positive and negative charges) variation with voltage. That is, a small variation in supply voltage causes a large variation in polarization. This nonlinearity means that a ferroelectric material can polarize at a specific voltage spontaneously. This is known as the coercive voltage. By exploiting this property, it is possible to have negative capacitance. Thus when voltage applied to the material is increased, the charge in the material is decreased and vice versa. This happens due to the movement of the atoms within the lattice of the material. However, this effect is observed only within a small voltage range though stabilization of negative capacitance is possible by reducing the base voltage required to turn on the device and hence changing its many other characteristics.

A ferroelectric layer is usually added between the gate and substrate material as shown in Fig. 1 though there are many ways to add this material into the transistor. Adding this materials means gate voltage is increased and threshold voltage is decreased and thus power consumption is reduced.

The ferroelectric material can increase the efficiency of a transistor by strengthening the gate voltage. This diminishes the $V_{th}$ needed to turn on the transistor and therefore drops off the power usage by the transistor.

In conventional FET, when voltage ($V_G$) is applied to the gate, voltage drops across the substrate ($\psi_s$) and the insulating layer ($V_i$). Applying KVL, we can-

$$V_G = V_i + \psi_s$$  (3)

When the insulating layer is replaced by a ferroelectric material then the voltage drop across the silicon substrate is increased as the negative capacitance of the ferroelectric material reduces the voltage drop across this layer. Thus if we want to maintain the same voltage drop across the semiconductor layer then we can reduce the applied gate voltage and hence we can reduce the power consumption by the device or in other words, we can increase the on state current.

One important parameter of the transistor is its sub-threshold slope. It is defined as the change of current in the transistor’s channel with the voltage change made at the gate terminal. The inverse of the sub-threshold slope is called sub-threshold swing ($S$), which is defined as the change of voltage needed at the gate terminal to raise current flow from the drain to the source terminal by ten times. In a conventional FET, the lowest value of $S$ is 60 mV/decade, but practical values lie in the range of 80-120 mV/decade.

When the sub-threshold slope is reduced in a negative FET, threshold voltage is also reduced, which means that the less amount of gate voltage is required to turn on the device. Therefore, less amount of voltage is required to have maximum current flow in the channel due to the reduction of saturation voltage ($V_{sat}$) and hence we can run the transistor at maximum operating frequency with a reduced voltage. This advantage comes with the fact that the speed of operation of CPUs with this kind of transistor is faster. Since power dissipation is proportional to the square of the applied voltage, hence a slight decrement in voltage yields a massive power saving and hence a reduced amount of heat dissipation.

Further advantage of NCFET is the less current flow during the off state of the transistor, which signifies reduced transistor leakage current and hence less power loss, which yields longer battery life and smaller energy expenditures.
VI. Construction and Characteristics of NCFET Structure

A negative capacitance FET uses a thin layer of ferroelectric material in between the gate metal and the gate dielectric material as used in the conventional FET structure as shown in Fig 3. Other things of the FET remain same as in the conventional FET. However, the processing steps are needed to be changed. These steps are already discussed in section 2 of this paper.

Due to this little change in the structure, the ferroelectric hysteresis, which is found in Fe-FET structure, is disappeared as shown in Fig. 4. The internal voltage drop at the ferroelectric-dielectric interface is larger than the gate voltage in such a way that the sub-threshold slope goes down under the Boltzmann thermionic limit of 60 mV/decade as demonstrated in Fig. 5. This value is usually measured at room temperature. Due to this phenomenon, the on-state current ($I_{DS}$) is reduced at a lower supply voltage ($V_{DD}$), and this yields a significant reduction of the power consumption by the transistor. Moreover, the threshold voltage ($V_{th}$) would escalate as $V_{DD}$ is raised as shown in Fig. 6.

VII. Theory of Operation of NCFET

In this section, the physics behind the operation of a NCFET is described briefly from the article of S. Salahuddin et al., 2008.

According to this article, a ferroelectric material preserves energy from the phase transition and during that action, it gives itself to be biased at a condition where its capacitance becomes negative (S. Salahuddin et al., 2008). Already, there is a capacitance due to depletion region of the device. Since negative capacitance is a part of the insulating layer’s capacitance hence it is added in series to the gate capacitance. Therefore, the equivalent capacitance is increased and hence the sub-threshold swing of the transistor is reduced well below the sub-threshold swing’s minimum limit 60 mV/decade by not changing the transport dynamics of the transistor. That means the on-state current of such type of device can be as high as the supply voltage that may be reduced considerably. Consequently, the effect of negative capacitance has the possibility of leading to an excessively low voltage device yet providing the excellent switching properties of the transistor.
Figure 7 (a) shows the energy landscape of a ferroelectric material, which exhibits two energy minima at non-zero polarization without an applied electric field to its gate terminal— one for positive charge and another is for negative charge. At the curvature of the energy landscape around \( Q = 0 \), one needs to have \( C < 0 \) to match the curvature. Because, we know that the energy stored by a conventional capacitor is given by \( Q^2/2C \). So, if energy is positive then \( C \) is positive but if energy is negative then it is obvious that as \( Q^2 > 0 \) then the capacitance should be negative even when charge is negative. Thus, a ferroelectric material can exhibit the property of the negative capacitance effect around the state \( Q = 0 \).

In Fig. 7 (b), two capacitors— one is the positive capacitance \( C_d \) and the other is the negative capacitance \( C_i \), are connected in series. Since the positive capacitance depends on the applied voltage therefore this must have zero charge when the applied voltage is zero. This condition confirms that the ferroelectric material should be biased at \( Q = 0 \) for this structure at zero gate bias voltage, i.e. with the capacitance being negative. The series equivalent capacitance may be expressed by equation (4).

\[
C_{eq} = \frac{|C_i|C_d}{|C_i| - C_d}
\]  

Figure 7: (a) Energy landscape of a ferroelectric material and (b) series connected depletion and insulating layer capacitors (S. Salahuddin et al., 2008)

This shows that for \( |C_i| > |C_d| \), which is usually the case, \( C_{eq} \) will be greater than both \( |C_i| \) and \( |C_d| \). But we know in conventional case, when capacitors are connected in series then the equivalent capacitance is always lower than even the smallest capacitance of the series connected capacitors. This is a unique property of negative capacitance with respect to the conventional ordinary capacitance.

This series combination of capacitor has a one-to-one mapping with a FET considering the fact that these two capacitors, \( |C_i| \) and \( |C_d| \) can be regarded as the gate insulator capacitance and semiconductor’s depletion layer capacitance respectively. Since the equivalent capacitance is increased rather than decreased when \( C_i \) is negative, it is possible to obtain an extra quantity of charge at the similar gate voltage or alternatively we can say that the applied gate voltage may be reduced for the similar quantity of charge in the channel. Thus the supply voltage is reduced in an NCFET. By applying the same reasoning, we can infer that the sub-threshold slope of such type of devices can be reduced well below the minimum value 60 mV/decade at room temperature for the conventional FETs.

VIII. Comparative Study of NCFET

A wide variety of NCFET structure is found in various literatures on NCFET. In this paper, a comparative summary of various NCFET structures is reported in Table 1. In this table, proposed NCFET structure, ferroelectric dielectric materials used in the gate stack, minimum sub-threshold slope in mV/decade, and typical application areas based on the respective NCFET structure are shown. It has been observed from this table that the lowest value of minimum sub-threshold slope (\( SS_{min} \)) is 11.7 mV/decade at 1 \( \mu \)m and 14.4 mV/decade at 0.5 \( \mu \)m of gate lengths. Besides, from the structure column, it is seen that the basic construction remains the same. The only difference is the insertion of the ferroelectric material in the gate stack. Mostly used ferroelectric material is Hafnium Zirconium Oxide (HZO) along with Molybdenum disulphide (MoS\(_2\)). But some other inorganic and organic materials have also been used. The main application areas are in the devices where ultra-low and high gain is of paramount importance, such as, sensors, mobile communication devices, flexible electronics etc. It is also to be noted that the most of the research works have been conducted in the first and second decades of the 21st century.
Table 1 Comparative summary report of various NCFET structures proposed and tested by the researchers

<table>
<thead>
<tr>
<th>NCFET Structure</th>
<th>Ferroelectric Dielectric Material</th>
<th>$SS_{\text{min}}$ (mV/dec)</th>
<th>Typical Application based on this NCFET Structure</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CuInP$_2$S$_6$ (CIPS Flake) and MoS$_2$</td>
<td>Hf$<em>{0.5}$Zr$</em>{0.5}$O$_2$ (HZO) and WSe$_2$</td>
<td>28</td>
<td>High-gain inverter</td>
<td>X. Wang et al., 2019</td>
</tr>
<tr>
<td>Hf$<em>{0.5}$Zr$</em>{0.5}$O$_2$ (HZO) and MoS$_2$</td>
<td>Hf$<em>{0.5}$Zr$</em>{0.5}$O$_2$ (HZO) and WSe$_2$</td>
<td>14.4</td>
<td>Steep-slope NCFET</td>
<td>M. Si et al., 2018</td>
</tr>
<tr>
<td>Hf$<em>{0.5}$Zr$</em>{0.5}$O$_2$ (HZO) and MoS$_2$</td>
<td>Hf$<em>{0.5}$Zr$</em>{0.5}$O$_2$ (HZO) and WSe$_2$</td>
<td>&lt;60</td>
<td>Mobile phones, distributed sensors, and emerging components for IoT</td>
<td>M. Si et al., 2018-2</td>
</tr>
<tr>
<td>Ferroelectric (Fe) polymer, poly (vinylidene difluoride-trifluoroethylene). 2-D semiconductor, e.g., MoS$_2$ and MoSe$_2$ as the transistor’s channel</td>
<td>Ferroelectric (Fe) polymer, poly (vinylidene difluoride-trifluoroethylene). 2-D semiconductor, e.g., MoS$_2$ and MoSe$_2$ as the transistor’s channel</td>
<td>24.2</td>
<td>Future flexible electronics</td>
<td>X. Wang et al., 2017</td>
</tr>
<tr>
<td>HfZrO$_x$</td>
<td>HfZrO$_x$</td>
<td>60</td>
<td>Bulk MOSFET, fully depleted SOI-FETs</td>
<td>A. Sharma and K. Roy, 2019</td>
</tr>
<tr>
<td>Ferroelectric polymer, poly (vinylidene difluoride-trifluoroethylene) with MoS$_2$</td>
<td>Ferroelectric polymer, poly (vinylidene difluoride-trifluoroethylene) with MoS$_2$</td>
<td>11.7 and 14.4 at 1μm and 0.5μm channel length</td>
<td>Extremely low voltage NCFETs</td>
<td>F. A. McGuire et al., 2016</td>
</tr>
<tr>
<td>Ferroelectric is deposited over a metal/high k dielectric gate stack</td>
<td>Ferroelectric is deposited over a metal/high k dielectric gate stack</td>
<td>28.3</td>
<td>Ultra-low power devices</td>
<td>C. W. Yeung et al., 2012</td>
</tr>
</tbody>
</table>
IX. Conclusions and Future Scopes

The main advantage of NCFET is that we don’t need to redesign the basic architecture of the device. We only need to insert a ferroelectric material with the gate stack. Therefore, we don’t need to change the manufacturing machine, just need to change few processing steps that are being used for the conventional FETs as well.

If the semiconductor manufacturing industries can harness the negative capacitance regime of ferroelectric materials then it is possible to change the total scenario of the microprocessor architectures. This would yield several benefits like smaller transistor leakage current, tiny voltage drop in the channel, reduced power consumption and dissipation etc. Therefore, this would improve the ratio of on-state to off-state current, increase devices’ life cycle, save power used by the data centers, reduce amount of heat generation, reduce energy requirement for cooling mechanisms. In this way, electricity expenses would be possible to minimize. This indicates that it has cumulative positive impacts on all aspects of electronics and electricity and hence the costs.

Researchers are still trying their level best to improve the performance parameters of NCFET. They are also trying to investigate and explore with the new ferroelectric materials (both organic and inorganic types) having stability over a wide range of variation of the operation parameters and compatibility of various technology nodes to reduce the minimum sub-threshold slope of the device further and thus harnessing more benefits. Therefore, this device seems to be very promising for the upcoming days for the researchers as well as semiconductor device manufacturing industries.

References


