State retention flip flop architectures with different tradeoffs using crystalline indium gallium zinc oxide transistors implemented in a 32-bit normally-off microprocessor

To cite this article: Niclas Sjökvist et al 2014 Jpn. J. Appl. Phys. 53 04EE10

View the article online for updates and enhancements.

Related content
- A normally-off microcontroller unit with an 85% power overhead reduction based on crystalline indium gallium zinc oxide field effect transistors
  Kazuaki Ohshima, Hidetomo Kobayashi, Tatsuji Nishijima et al.
- Electrical characteristics and short-channel effect of c-axis aligned crystalline indium gallium zinc oxide transistor with short channel length
  Yoshiyuki Kobayashi, Shinpei Matsuda, Daisuke Matsubayashi et al.
- Crystalline In–Ga–Zn–O FET-based configuration memory for multi-context field-programmable gate array realizing fine-grained power gating
  Munehiro Kozuma, Yuki Okamoto, Takashi Nakagawa et al.
State retention flip flop architectures with different tradeoffs using crystalline indium gallium zinc oxide transistors implemented in a 32-bit normally-off microprocessor

Niclas Sjökvist1–3*, Takuro Ohmaru1, Atsuo Isobe1, Naoaki Tsutsui1, Hikaru Tamura1, Wataru Uesugi1, Takahiko Ishizhu1, Tatsuya Onuki1, Kazuaki Ohshima1, Takanori Matsuzaki1, Hidetoshi Mimura1, Atsushi Hirose1, Yasutaka Suzuki1, Yoshinori Ieda1, Tomoki Atsumi1, Yutaka Shionoiri1, Kiyoshi Kato1, Gensuke Goto1, Jun Koyama1, Masahiro Fujita2, and Shunpei Yamazaki1

1Semiconductor Energy Laboratory Co., Ltd., Atsugi, Kanagawa 243-0036, Japan
2VLSI Design and Education Center, University of Tokyo, Bunkyo, Tokyo, 113-8656, Japan
3Student, Linköping University, 581 83, Linköping, Sweden
E-mail: nsi012@sel.co.jp

Received September 23, 2013; accepted December 27, 2013; published online March 13, 2014

As leakage power continues to increase when transistor sizes are downscaled, it becomes increasingly hard to achieve low power consumption in modern chips. Normally-off processors use state-retention and non-volatile circuits to make power gating more efficient with less static power. In this paper, we propose two novel state-retention flip-flop designs based on a parallel and series retention circuit architectures utilizing crystalline indium gallium zinc oxide transistors, which can achieve state retention with zero static power. To demonstrate the application of these different designs, they are implemented in a 32-bit normally-off microprocessor with an energy break-even time of 1.47 µs for the parallel type design and 0.93 µs for the series type design, at a clock frequency of 15 MHz. We show that decreasing the power supply duty cycle to 0.9%, the average current of the processor core can be decreased by over 90% using either type of flip-flop. © 2014 The Japan Society of Applied Physics

1. Introduction

Power consumption has become one of the most important issues in electronic design, and there is an industry trend toward an increased focus on low-power chips and devices.1) At the same time, the semiconductor industry continues to downscale devices with each new generation so as to fit more devices per chip.2,3) A well known side effect from downscaling transistor sizes for silicon (Si) transistors is an increased leakage current, which increases the static power consumption.4–7) This means while it is desirable to keep power consumption as low as possible, it becomes harder to do so as transistor sizes continue to decrease.

To combat this problem, new technologies and architectures are needed. One suggested solution is a design architecture called normally-off (NOFF) computing, which is meant to decrease the static power consumption by combining state-retention and non-volatile devices with power gating, so that the circuit remains in a power-off state when not in use.8–10)

The devices required to realize a NOFF central processing unit (CPU) such as non-volatile static random access memory (RAM) and flip flops (FFs) have already been realized in a number of technologies.10–22) Previous reports show that power gating using magnetoresistive RAM (MRAM) or ferroelectric RAM (FeRAM) as a nonvolatile element are effective in reducing the average power. However, MRAM elements are driven by current and require a high write energy which results in a large power overhead.17) There is a report of a processor with FeRAM produced with a 130 nm technology,22) but future scaling is predicted to be difficult.23)

In this study, we use c-axis aligned crystalline indium gallium zinc oxide (CAAC-IGZO) transistors which is a subset of CAAC oxide semiconductor (CAAC-OS) technology. We have previously reported on the success of CAAC-IGZO transistors in NOFF applications.24–26,39,40) By using CAAC-IGZO elements in nonvolatile circuits, the writing energy can be reduced. Future scaling is possible to 100 nm or less,27) which indicates that it is a promising choice for advanced low-power applications.

We will further demonstrate the feasibility of NOFF computing using a hybrid semiconductor process technology utilizing both Si and CAAC-IGZO transistors. This is done by implementing two different design architectures of state-retention flip flops (SRFFs) utilizing CAAC-IGZO transistors in a 32-bit NOFF CPU. A suitable figure of merit for comparing the effectiveness of each implementation is the energy break-even time (BET),28) which will be measured for each implementation.

2. Crystalline IGZO technology

IGZO material has been subject to widespread research, and there have been several papers about its crystalline properties. Homologous structures have been reported,31–35) as well as a single crystalline structure.36,37)

It has been shown that CAAC-IGZO transistors experience an extremely low off-state current on the order of 10–23 A/µm,29,30) which makes them suitable for creating non-volatile devices. The typical electrical characteristics of an CAAC-IGZO transistor can be seen in Fig. 1.

CAAC-IGZO transistors have been used in many different applications, which include dynamic random access memory,28) NOFF processors39,40) and a field-programmable gate array.41) There have also been papers summarizing the utilization of CAAC-IGZO in displays and LSI.42)

3. Normally off CPU

We have designed and produced a NOFF CPU which utilizes power gating (PG) to decrease the power consumption. The core utilizes state retention registers designed with CAAC-IGZO SRFFs, and the cache has memory cells designed with both CAAC-IGZO and Si transistors to achieve data retention during power off. A table detailing the specifications is
shown in Table I, while a micrograph of the produced NOFF microprocessor is shown in Fig. 2.

The core is designed with a 32-bit “Microprocessor without Interlocked Pipeline Stages version 1” (MIPS-I) instruction set architecture (ISA). Other blocks of interest is the 2 KB non-volatile cache memory and a power management unit (PMU) which controls the power gating sequence. The PMU can exit the power gating sequence either by using an internal counter, or by an external interrupt signal. With the exception of the PMU, all these blocks are powered off during periods of no activity.

4. State retention flip flop functionality

A SRFF has the same functionality as an ordinary FF, with the addition of being able to backup and restore its data to a retention circuit. By utilizing the low off-state current of CAAC-IGZO transistors, it is possible to use a retention circuit which stores electric charge on a retention capacitor, effectively eliminating the static power consumption during power off.

We will present two different SRFFs based on CAAC-IGZO transistors, with different architectures and tradeoffs.

4.1 Parallel type design

The first design is a a master slave FF with a retention circuit in parallel with the master latch. This design has a small performance penalty, at the cost of some area overhead (13.5% compared to an equivalent FF design without the retention circuit) due to its additional Si transistors. The CAAC-IGZO transistor and retention capacitor are stacked upon the Si device layer, and does not contribute to any area overhead. The schematic view of this design can be seen in Fig. 3(a) while the layout can be seen in Fig. 4(a).

The timing chart can be seen in Fig. 5. When the PMU initiates a backup or restore sequence, it will change the control signals of the SRFF as shown in the chart. A backup of the current state is performed in a single step (TB1). During TB1, CLK is set high which locks the current data in the master latch, and OS\textsubscript{G} is kept high which opens the CAAC-IGZO transistor and charges the retention node FN. After this, the circuit can be powered off with the current state isolated on the retention capacitor.

The restore sequence consists of three separate steps (TR1–TR3). In TR1, the power supply is restored and stabilized. During this stage, the CLK and RESET signals are kept low, which will allow the node connecting the master and slave latch together to be precharged. In TR2, CLK is kept high with RESET still low. Depending on the voltage of the retention node FN, the precharged node will either keep its charge or be discharged. Finally in TR3, RESET is set high and the circuit is restored from the previously precharged node.

4.2 Series type design

The other design architecture is a standard master slave FF with the retention circuit in series with the master latch. This design has no extra Si transistors, and both the CAAC-IGZO transistor and retention capacitor are stacked upon the Si device layer which results in zero area overhead. As the retention circuit affects the circuit during normal operation, the maximum frequency will heavily depend on both the speed of the CAAC-IGZO transistor and the size of the retention capacitor. The schematic view of this design can be seen in Fig. 3(b) and the layout in Fig. 4(b).

The timing chart for the backup, power off and restore sequence can be seen in Fig. 6. During normal operation,
OS\_G is set high so that the master latch operates normally. This also means that the retention capacitor is always charged or discharged with the input D. Therefore, in order for this design to retain its state, OS\_G should be set low so the retention node is isolated. This can be done simultaneously as the circuit is powered off, resulting in an immediate backup.

Two steps (TR1–TR2) are required to restore the state. In TR1, the supply voltage is stabilized and CLK kept high which restores the master latch according to the voltage of the retention node FN. In TR2, RESET is set high and the remaining circuit is restored. Normal operation can then be resumed.

5. Flip flop frequency

To determine the operating range and frequency of these new SRFF designs, we have measured their performance during normal operation. The designs have been fabricated using a Si gate length of 0.35 µm, a CAAC-IGZO gate length of 0.8 µm and a retention capacitor of size 0.1 pF. The result is presented in a Shmoo plot indicating whether normal operation passed or not for each combination of supply voltage/frequency. The maximum frequency and operating range for the series design can be seen in Fig. 7, which show a maximum frequency of 45 MHz at 2.5 V.

The parallel design has an operating frequency larger than 134 MHz for a supply voltage equal to or higher than 1.3 V (limited by our test equipment). Simulation results for the parallel design show a maximum frequency of 1.28 GHz at a supply voltage of 2.5 V. Simulation of a reference FF shows a maximum frequency of 1.32 GHz at a supply voltage of 2.5 V.

5.1 Scaling

The parallel design’s retention circuit has a very small influence on the maximum operation frequency during normal operation, since the CAAC-IGZO transistor is closed. Due to this, the parallel design will show similar improvements from scaling as a normal Si FF.

To investigate the scalability of the series design, we have performed simulations using a predictive technology model (PTM) for the Si transistors with a gate length of 32 nm and an a custom model for a CAAC-IGZO transistor gate.
length of 65 nm. As the performance of this circuit is affected by the channel width of the CAAC-IGZO transistor, we display the results as frequency versus channel width of the CAAC-IGZO transistor. The size of the retention capacitor is 0.01 pF. The PTM simulation results are shown in Fig. 8, which shows the maximum frequency at 2.5 V for different CAAC-IGZO channel widths.

6. Normally off CPU performance

Since each design type of SRFF has very different maximum frequency, it is of interest to see how it affects the maximum frequency of the NOFF CPU. We have used the Dhrystone benchmark program\(^{43}\) to stress the NOFF CPU. The measured results for the parallel implementation can be seen in Fig. 9(a), whereas the results from the series implementation can be seen in Fig. 9(b).

Even though the difference between the SRFFs’ frequency is large, the different implementations of the NOFF CPU are not far apart in maximum frequency. This is because the frequency is not limited by the performance of the SRFFs, but by the large physical size of the CPU core which has large propagation delays in our current fabrication process.

7. Break-even time

The number of cycles required for the backup and restore sequence in the CPU is listed in Table II (together with an inherent two clock cycle delay from the PMU). The total number of cycles required for a power gating sequence is then equal to the sum of the delay from the PMU, the length of the backup sequence, the number of power off cycles and the length of the restore sequence. This is expressed as

<table>
<thead>
<tr>
<th>Design</th>
<th>(N_{PMU})</th>
<th>(\Sigma TB)</th>
<th>(\Sigma TR)</th>
<th>(\Sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel</td>
<td>2</td>
<td>9</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>Series</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

Fig. 7. Measured performance and operating range of the series design SRFF.

Fig. 8. Simulated performance of the series design SRFF with the predictive technology model at a supply voltage of 2.5 V for the series design state retention flip flop.

Fig. 9. Measured performance of the NOFF CPU as a function of supply voltage with the parallel implementation shown in (a) and the series implementation in (b).
\[ N_{\text{PG}} = N_{\text{PMU}} + \sum_n TBn + N_{\text{OFF}} + \sum_n TRn. \]  

(1)

The BET is the number of cycles where the average current using power gating is less than or equal to the average current without power gating. We have identified the BET by measuring the current while changing the number of cycles during which the circuit is powered off, ranging from \( N_{\text{OFF}} = 2 \) (which is the smallest possible value possible by the PMU) up until the maximum power-off sequence length. The power-off time is controlled either by an internal 16-bit counter (up to 65536 cycles) in the PMU, or externally by an interrupt signal which can give an indefinitely long power-off time.

To compare the different implementations, we use a test program which performs power gating for a number of cycles, followed by a loop program having ten no operation instructions (NOP) and a jump instruction.

The results for both implementations can be seen in Fig. 10. The parallel design experiences a decreased current with \( N_{\text{OFF}} = 2 \) cycles, which gives a BET of 22 cycles (1.47 µs at 15 MHz). The reason why power-savings are achieved even with such a small number of power-off cycles is the long backup sequence. Since the clock is stopped during the backup sequence, the dynamic power is reduced during this time. It is worth noting that the BET of 22 cycles is limited by the PMU as it is the smallest possible backup and recovery sequence for the parallel implementation.

Break-even energy for the series design is achieved at \( N_{\text{OFF}} = 5 \) cycles, which gives a BET of 14 cycles (0.93 µs at 15 MHz). To see the impact for longer time periods, we can investigate the current consumption when increasing the power-off time of the supply voltage (effectively changing its duty cycle). The results for the NOFF CPU core can be seen in Fig. 11. It is worth noting that the average core current consumption is decreased by over 99% when the supply voltage duty cycle decreases from 100 to 0.9%. This effectively means that the average power consumption of the core can be reduced with at least 99% by adjusting the power-off time.

8. Retention time

The NOFF CPU’s registers will have a retention time limited by the leakage current experienced by the retention node. We have measured the retention time of both implementations at 85 °C by first writing to the registers and then checking their contents after they have been powered off for a period of time. It has been observed that both designs can achieve an retention time of at least 24 h at 85 °C.

9. Conclusions

We have designed and produced two different architectures of SRFFs utilizing both Si and CAAC-IGZO transistors. One has its retention latch in parallel with the standard flip-flop design, and one has its retention latch in series. The parallel design offers high frequency at the cost of 13.5% area overhead, while the series design has zero area overhead at the cost of frequency degradation.

Although the series design SRFF suffers from a large frequency degradation, the design has a very simple control scheme and zero area overhead. The maximum frequency makes it suitable for systems operating at several tens of MHz (e.g., microcontrollers).

These different SRFF designs have then successfully been implemented in a normally off microprocessor, where they could achieve an energy break-even time at 15 MHz of 1.47 µs for the parallel version and 0.93 µs for the series version. However, it is found that the break-even time for the parallel version is limited by the current PMU design, and could be further decreased for future designs.

The efficient implementation of these SRFFs show that CAAC-IGZO transistors are very suitable for implementation in NOFF circuits. This type of NOFF microprocessor is useful in any application with long idle times, achieving very large power savings.

We plan to scale down the process technology for higher performance and are also investigating the use of CAAC-IGZO transistors in peripheral circuits so that a NOFF microcontroller can be realized.