ISSN 0002-306X. Proc. of the RA NAS and NPUA Ser. of tech. sc. 2017. V. LXX, N 4.

UDC 004.08:621.382.3 RADIOELECTRONICS

V.SH. MELIKYAN, A.V. AVETISYAN, K.H. SAFARYAN

CUT-OFF CIRCUIT FOR DUAL RAIL SRAM PERIPHERY WITH IMPROVED DYNAMIC POWER

New circuit technique to reduce the dynamic mode power of SRAM is proposed. A cut-off-circuit (COC) is proposed to turn off the periphery voltage while keeping memory array voltage as high. This circuit is implemented in a dual-rail-supply SRAM.

Keywords: static random access memory: power; voltage: dynamic; supply; dual-rail.

Introduction. Nowadays, mobile applications are widely used such as smartphone processors, autopilot cars or virtual-reality headsets, and one of the main challenges for those devices is SRAM memories, that can be operated in a low-power state to extend the battery life [1]. One of the methods for power reduction is the power management mode, which affects the design architecture more than clock gating. It increases time delays, as power gated modes must be safely entered and deduced. Architectural trade-offs exist between designing the amount of leakage power saving in low power modes and the energy dissipation to enter and exit the low power modes. An externally switched power supply is a very basic form of power gating to achieve long-term leakage power reduction. To shut off the block for small intervals of time, internal power gating is more suitable, such as power gating, with light sleep mode, which can drop the array to the lowest poseable active power setting, deep sleep, which drops the array voltage to the lowest possible retention power and shut down mode which turns off the array with no data retention, so the array and periphery are power-gated. But the disadvantage of these modes is the area loss, to overcome this problem, a cut-off-circuit (COC) is proposed. Due to this method, the external CUT signal can control the supply switch of the memory periphery, and at the same time keep the memory array content. This paper demonstrates the cut-off circuit, with improved dynamic power. The rest of the paper is organized as follows. In the first section, the schematic design and operation of the conventional power gating schemes are described. The proposed COC scheme for dual-rail memory architecture is explained in detail in the “proposed cut off circuit and operations for dual rail SRAM periphery” section.

Problems and justification of the methodology:

A. Normal Sleep-Circuit.

One of the widely used power gating feature is data-retention gated power [2] scheme, which is provides lower supply across the memory cells, using header and footer devices during the sleep mode (Fig.1). The active current is controlled by the input ENP signal, which activate the PMOS transistor, so internal-VDD starts falling.

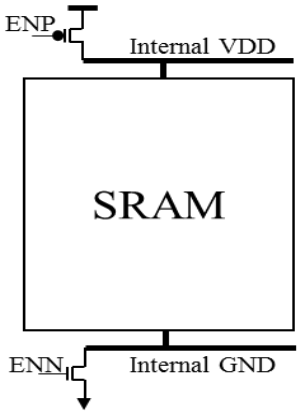

Fig. 1. Normal Sleep Circuit

The normal mode active sink, controlled by the ENN signal, which activate the NMOS transistor and due to that internal-GND plane starts rising.

B. *Sleep-Circuit with P-Diode.*

When the PMOS transistor becomes fast and the NMOS transistor becomes slow, the value of internal GND increases dramatically, despite the low leakage. This is the problem in normal sleep-circuit at cross corners.

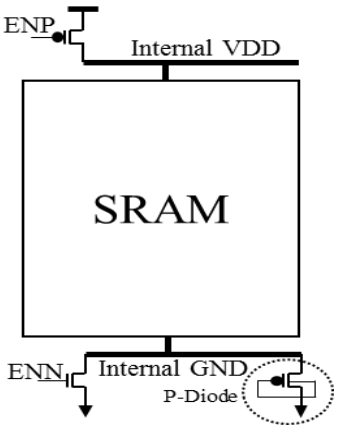

Fig. 2. Sleep Circuit with Footer P-Diode

Due to large process variations, a situation arrives, when at cross corner conditions, rail-to-rail voltage becomes minimum. The solution for these phenomena is to append the P-diode on the internal GND (Fig. 2). To meet the threshold voltage requirement, the p-diode bulk is connected to the ground. Internal-VDD level is reduced because of the increased leakage, which in turn is a feedback to the bulk of P-diode, reducing the threshold voltage of this device.

C. *Sleep-Circuit with Header Feedback P-Diode.*

The above mentioned method, however has a problem when both devices (NMOS and PMOS) are slow. At this point, the internal-VDD has an inordinate value drop. Moreover, to solve this challenge, in addition to footer p-diode, it is decided to use one more p-diode on the header with a feedback (Fig. 3), which will take care of the drop. This feedback reduces the threshold voltage of the header p-diode, thus precludes the internal-VDD from falling to a certain value. The same method can be used at the footer p-diode with an internal-GND. At low temperatures, threshold voltage increases and a stability [3] problem arrives, due to severe reduction in rail-to-rail-voltage, which creates a limitation for the low voltage designs [4]. This is undesirable if such a restriction occurs with a very low leakage condition or corner, where sleep circuit is not required.

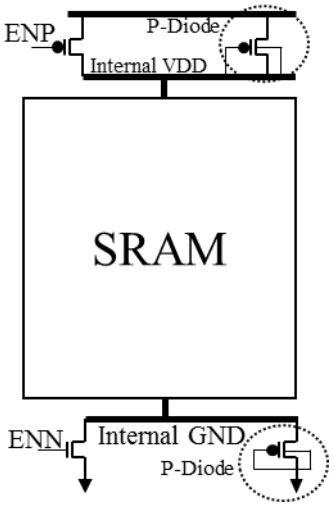

Fig.3. Sleep Circuit with Header Feedback P-Diode

SRAM with dual rail option. The disadvantage of the methods mentioned in the upper section is the area loss when SRAM memory has two physically separated supply voltages for array and periphery, because all the methods are applied on a single SRAM cell, so if the number of words and the number of bits are big, the sleep circuits will dramatically affect the full memory area. Firstly, to

separate the supply for the array and the periphery, the dual rail option is enabled in memory, the periphery power rail and the array power rail are physically separated. Level shifters are used at the boundary of the array, allowing to run the memory array and periphery at different voltages. There are two types of level shifter implementation. If performance is not critical, then periphery voltage can be reduced till the minimum logic functional voltage and, at the same time, keeping the array at the minimum array voltage. In this way, dynamic power is reduced in periphery [5].

1. Level shifters at memory top boundary.
2. Level shifters between control cells and array matrix.

Level shifters at memory top boundary.

Level shifters are added at the memory macro top boundary (Fig. 4), periphery gets the array supply and external supply (VDD) connects to memory input pins. In this case higher internal periphery operating voltage may enable higher speed, but there is no any advantage in dynamic power from the macro when external supply goes low as internal periphery is at higher voltage.

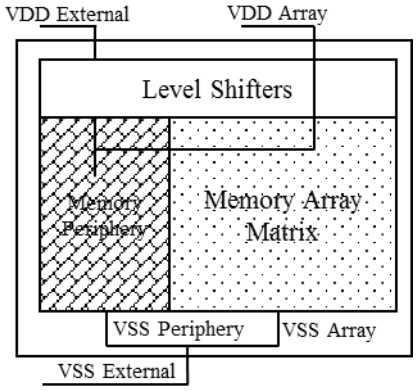

Fig. 4. Dual Rail implementation on top of the memory

Level shifters between control cells and array matrix

Level shifters are added between the periphery and array matrices (Fig. 5). VDD periphery supply is connected to the chip level VDD. Memory macro power is mostly consumed by periphery and it gets power reduction from lower voltage. Some voltage gap between the array and periphery supply can provide lower power.

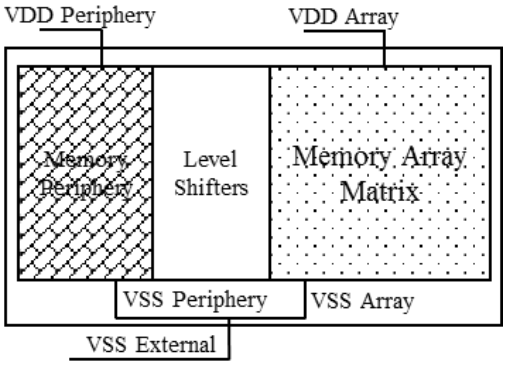

Fig. 5. Dual rail implementation between periphery and array

Proposed cut off circuit and operations for dual rail SRAM periphery. The proposed cut-off-circuit feature is used only when the memory has dual rail option. This circuit can be used to put memory in power down mode, which will allow the periphery supply to turn off completely and keep the memory content. In Fig. 6, the top-level structure of the COC circuit is shown.

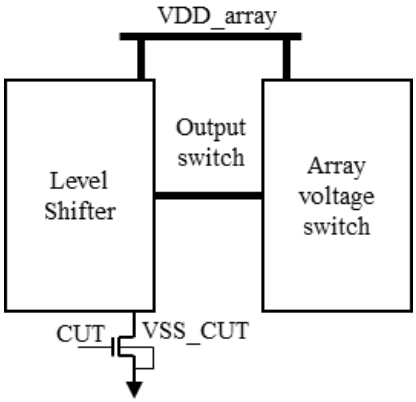

Fig. 6. Top level implementation of COC

This consists of level shifter, array voltage switch and one NMOS transistor. The control signal is CUT, when it is "1" the via NMOS the VSS_CUT discharges to GND, when it is "0" the VSS_CUT is isolated from GND and the source of N2 transistor from level shifter will be floating. This will give an opportunity to control the output switch when CUT signal is disabled and IN and IN inverse signals are respectively "0" and "1". In Fig. 7, the structure of the level shifter is shown. When CUT signal is low, and IN is low the drain of N2 transistor is floating, the P2 is cut off and the current flows from the VDD_array to the output switch. Because the P3 size is small, less current flows along P3. Thus, the

power decreases. The Array voltage switch is presented to control the memory matrix supply, thereby the information in the bit cells can be maintained (Fig. 8).

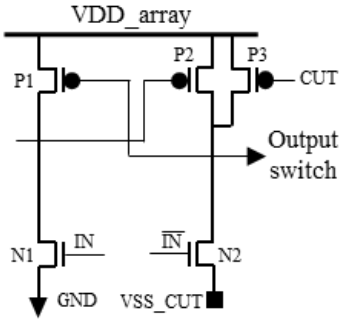

Fig. 7. Structure of level shifter

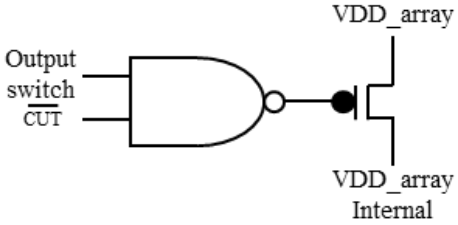

Fig. 8. Structure of the array voltage switch

Measurement results. 512-b SRAM with 6T bitcell and proposed technique has been designed in a 28-NM CMOS technology. The power management is disabled in this memory and the cut-off mode is enabled.

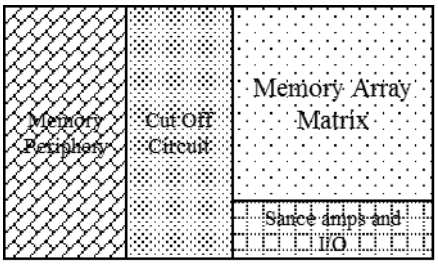

Fig. 9. COC scheme implementation

The memory array is created with 64 physical rows and 8 physical columns. The proposed COC scheme is implemented for the memory array Fig. 9. The peripheral circuit, such as sense amplifiers, control circuit are placed as shown in Fig. 9.

Dynamic power and memory area values for the main PVT corner without using COC scheme is presented in Table 1.

Table 1

Results for regular SRAM

Description	Value
Dynamic Power(uW)	7.77
Area(SM)	300.41

After the COC scheme implementation the dynamic power and the memory area parameters have the following values (Table 2).

Table 2

Results for regular SRAM with COC scheme

Description	Value
Dynamic Power(uW)	7.372
Area(SM)	349.52

As a result, by applying the COC scheme in the SRAM memory, the difference in dynamic power is decreased around 5%, but at the same time, the area is increased by 16%.

Conclusion. 6T SRAM has been presented with an inbuilt cut-off-circuit (COC) scheme to reduce the dynamic power of SRAM. This technique saves dynamic power for the read/write operations. The scheme can be implemented on any SRAM, which has dual rail option. It selectively activates the CUT signal and gives an opportunity to turn off periphery supply and keep memory content then deactivates CUT. As a result, during the read/write cycles, the dynamic power is saved. The dynamic power saving is around 5%, the disadvantage of this method is the memory area increase by 16%.

REFERENCES

1. Embedded SRAM designs for enhancing performance, power and area (PPA) in 16 nm FinFET technology /K. **Nii, Y. Ishii, M. Yabuuchi, et al** //2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT). – September, 2016. – P. 563 – 566.
2. **Jiao H., Kursun V.** Power gated SRAM circuits with data retention capability and high immunity to noise: A comparison for reliability in low leakage sleep mode // 2010 International SoC Design Conference. – April, 2010. – P. 5 – 8.
3. Write-back technique for single-ended 7T SRAM cell /**V. Melikyan, A. Avetisyan, D. Babayan, et al** //2017 IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO). – April, 2017. – P. 112 – 115.

4. A new approach of multi voltage and adaptive voltage scaling techniques for 16 nm FinFET RISC processor / **D. Babayan, E. Babayan, A. Avetisyan, et al** //2017 IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO) – April, 2017. – P. 128 – 131.
5. **Qazi M., Stawiasz K., Chang L., Chandrakasan A.P.** A 512 kb 8 T SRAM macro operating down to 0.57 V with an AC-coupled sense amplifier and embedded data-retention-voltage sensor in 45 nm SOI CMOS // IEEE Journal of Solid-State Circuits. January, 2011. – Vol. 46, No. 1. – P. 85 – 96.

Synopsys Armenia CJSC, Synopsys Armenia Educational Department (SAED). The material is received 04.10.8017.

Վ.Շ. ՄԵԼԻՔՅԱՆ, Ա.Վ. ԱՎԵՏԻՍՅԱՆ, Կ.Հ. ՍԱՖԱՐՅԱՆ

ՍՏԱՏԻԿ ԿԱՄԱՅԱԿԱՆ ԸՆՏՐՈՒԹՅԱՄԲ ՀԻՇԱՍԱՐՔԻ ԴԻՆԱՄԻԿ ՀԶՈՐՈՒԹՅԱՆ ՆՎԱԶԵՑՄԱՆ ՍԽԵՄԱ

Առաջարկվում է նոր սխեմա՝ ստատիկ կամայական ընտրությամբ հիշասարքի դինամիկ ռեժիմի հզորությունը նվազեցնելու նպատակով: Ներկայացվում է մի շղթա, որն անջատում է շրջագծի լարումը՝ պահպանելով հիշողության մատրիցի լարումը: Ի տարբերություն դինամիկ հզորության կրճատման այլ մեթոդների՝ առաջարկվող մեթոդն ապահովում է դինամիկ հզորության նվազեցումը՝ միևնույն ժամանակ ունենալով մակերեսի նվազագույն աճ: Այս սխեման իրականացված է կրկնակի սնուցման դողերով ստատիկ կամայական ընտրությամբ հիշասարքերում:

Առանցքային բառեր. ստատիկ կամայական ընտրությամբ հիշասարք, հզորություն, լարում, դինամիկ, կրկնակի դող:

В.Ш. МЕЛИКЯН, А.В. АВЕТИСЯН, К.Г. САФАРЯН

СХЕМА ДЛЯ СНИЖЕНИЯ ДИНАМИЧЕСКОЙ МОЩНОСТИ В СТАТИЧЕСКОЙ ПАМЯТИ С ПРОИЗВОЛЬНЫМ ДОСТУПОМ

Предложена новая схема для снижения мощности динамического режима статической памяти с произвольным доступом. Предлагается цепь, которая отключает напряжение периферии, при этом сохраняя напряжение матрицы памяти. В отличие от других методов снижения динамической мощности, предложенный метод имеет меньшую разницу в площади памяти и в то же время обеспечивает наибольшее снижение динамической мощности. Эта схема реализована в статической памяти с произвольным доступом с двумя шинами питания.

Ключевые слова: статическая память произвольного доступа, мощность, напряжение, динамический, двойная шина.