2012 International Conference on Solid-State and Integrated Circuit (ICSIC 2012)
IPCSIT vol. 32 (2011) © (2011) IACSIT Press, Singapore

New Circuit Design Architecture for a 300-MHz 40nm 1Mb Embedded STT-MRAM with Great Immunity to PVT Variation

Hung-Chang Yu [+], Kai-Chun Lin , Ku-Feng Lin , Chin-Yi Huang , Yu-Der Chih , Tong-Chern Ong , Luan C. Tran and Fu-Lung Hsueh

Taiwan Semiconductor Manufacturing Company
Hsin-Chu 300, Taiwan, ROC

Abstract. A 1Mb STT-MRAM implemented in 2-cell per bit architecture is demonstrated in 40nm technology. By using newly developed sense amplifier and word-line driver as well, 300MHz access speed can be acquired and, compared to one cell per bit using conventional sensing scheme, huge immunity to PVT variation can be achieved. This paper shows an effective architecture to overcome the large deviation in TMR and MTJ resistance due to hard control on smoothness of MgO thin film and MTJ uniformity in 40nm or 28nm technology.

Keywords: MTJ, MgO oxide thickness, Rap, Rp, 2T2MTJ, 2-cell per bit.

1. Introduction

Many circuit techniques have been proposed for read operation on STT-MRAM. Self-reference is proposed to minimize the effect of resistance variation but access time overhead is too much [1]. Conventional sensing schemes, like negative resistance read scheme [2] and sense amplifier with source side degeneration circuit [3], were proposed on read operation but still neither of above can overcome the issue of read failure due to MTJ resistance variation. Fig.1 shows that the conventional sensing scheme uses two cells as reference cells which are stored Rp (MTJ low resistance) and Rap (MTJ high resistance) and generate reference resistance Rref . Rref is used to generate reference current or voltage. When reading a cell, sense amplifier detects that cell current or voltage and compares it with reference current or voltage in order to distinguish the resistance stored in the read cell. As long as the condition Rp < Rref < Rap is not violated and Rp, Rref and Rap are well separated as well, the successful read operation can be done. However as process technology is scaling down, MgO oxide thickness ($\tau < 1$nm), the roughness of MgO thin film and MTJ cross-sectional area (A) have prominent effect on large stochastic spread of variation in MTJ resistance [5] so that Rp, Rref and Rap are easily not well separated. Once resistance variation is large but TMR is not sufficiently large, Rref can overlap with Rp and Rap. See Fig. 2. Under such situation, Rp may be larger than Rref and Rap may be smaller than Rref. Thus read operation will fail. Even if merge reference scheme mentioned in [4] is used, the problem of insufficient sensing margin between Rp (or Rap) and Rref still cannot be avoided

An alternative solution is proposed in this paper. The proposed solution skips using Rref but uses two cells as one data bit's storage. One cell stores a bit value and the other cell stores the complement of that bit. The proposed scheme doesn't need to compare Rp or Rap with a reference resistance. The sensing scheme used in 2-cell per bit just compares a pair of cells with opposite states written into these 2 cells . A well design differential sense amplifier can detect the voltage or current difference resulted from cells' resistance difference which can be as small as just 500Ω, Hence even if TMR is as small as 30%, according to the Monte Carlo simulation on the proposed design the read operation still can be 100% successfully. Recently

[+] Corresponding author. Tel.: + (886939828162); fax: +(88635768812).
E-mail address: (hcyum@tsmc.com).

there have been some discussion on non-volatile SRAM by combining 6-T SRAM cell with MTJ cell, like [6]. That kind of architecture in fact is not practical because it ignores the risk of MTJ disturbance and MgO thin film breakdown due to un-controlled voltage across MTJ when performing voltage equalization in 6-T cell. When reading the cell, MTJ will be easily disturbed too due to the same root cause. Besides, in 40nm 2T2MTJ cell size is still much smaller than 6-T SRAM cell size. Therefore it's better to implement non-volatile SRAM-competitive memory by using 2T2MTJ instead of using 6-T SRAM cell combined with 2 MTJ. To implement high speed 2-cell per bit STT-MRAM, new sense amplifier, word line driver and write scheme are proposed in this paper. Part II shows the design of high speed sensing scheme using current mode differential sense amplifier. Part III shows novel word line drivers to drive word lines up to different voltage levels. Part IV shows the write scheme.

2. High speed current mode sense amplifier

Recently proposed current mode sense amplifier [7] claimed to achieve high-speed operation. But there is no bit-line clamp architecture used in sense amp of [7] and hence there are two problems resulted from using that sense amp: one problem is cell disturbance and the other is MgO thin film damage. From [7] it can be seen that during pre-charge phase, bit-line voltage can be much higher than 1.0 volt across PVT corners. To avoid cell disturbance, voltage across MTJ should be far less than 0.5v. Besides, high voltage stress (> 1.0v) across MTJ will easily degrade reliability of MgO thin film and even further cause MgO thin film breakdown. Hence well defined bit-line voltage is important to avoid MTJ being disturbed or even damaged. In this paper the proposed circuit of high-speed current mode sense amplifier is shown in Fig. 4. In this circuit M1 M2 M3 M4 constitute cross-coupled positive feedback loop which is commonly used in high speed sense amplifier. The source terminals of M1 and M2 are low impedance input terminals used for input current. Constant tail current flowing through M7 is designed to maintain fixed common mode output level. M8 and M9 are clamp devices used to clamp bit line voltage smaller than 0.3v in order to avoid disturbing the states stored in MTJ cells. When sense amplifier is enabled and equalized, the current flowing through M1 and M2 are Id1 and Id2. Output nodes SAO_L and SAO_R are initially equalized at Veq and then will be released after EQ is off. Right after EQ is off, the initial output difference voltage $\Delta Vout$ of sense amplifier can be calculated as follows.

Since $Id1 = gm1 \times Vgs1$, $Id2 = gm2 \times Vgs2$; $Vgs1 = Veq - V1$ *and* $Vgs2 = Veq - V2$

$\Rightarrow \Delta Vgs = Vgs1 - Vgs2 = V2 - V1 = \Delta V$; $\Delta I = Id1 - Id2 = gm1 \times Vgs1 - gm2 \times Vgs2$

Assume M1 and M2 are identical, hence $gm1=gm2=gm \Rightarrow \Delta I = gm \times (V2-V1) = gm \times \Delta V$;

$\Delta Vout = \Delta I \times Rout = gm \times \Delta V \times [(gm + gmb)ro1rv1 + ro1 + rv1] || ro3$, *Rout is the output impedance; ro1, ro3 are the drain-source impedances of M1and M3; rv1 is the impedance looked into the node V1*

Once the differential voltage is introduced into the output nodes, the positive feedback loop will quickly amplify the difference voltage and then the sense amplifier is driven into the latched state. Simulation shows that 2ns sensing time can be achieved. However if there is mismatch in M1 - M6, sensing speed may be effected. Thus symmetrical layout of M1 - M6 is crucial to minimize the probability of mismatch.

Various simulations have been done to compare the performance among proposed sense amplifier, conventional current mode sense amplifier in Fig.1 and differential sense amplifier in [7] as well. All these three amplifiers are designed and optimized in thick gate-oxide devices and typical power supply is 2.5volt. RC load on each bit-line is considered based on the array of 512 word-line by 512 pairs of bit-line. Fig. 5 shows that for the proposed sense amplifier it only takes 2ns from enable signal EN asserted to sense amplifier output data generated. If measuring the period from equalization signal EQ to data output, it's even faster and only takes 1 ns. Fig. 5 shows the comparison of access times v.s. MTJ resistance variation among 2 types of sense amplifiers. It demonstrates that current-mode differential sense amplifier is generally immune to the variation of MTJ resistance but conventional current mode sense amplifier is greatly effected by resistance variation due to large variation in MTJ cell current. In Fig. 6 Monte Carlo simulation results with 1000 samples show that the comparison of read yield with TMR variation. When 1 $\sigma \leq 6\%$ in MTJ resistance, 100% read success can be achieved even if TMR is as low as 40%. However for 1-cell per bit architecture using conventional sensing scheme, it's hard to achieve 100% read success even if TMR=1.0

when σ = 6%. In Fig. 7 Monte Carlo simulation shows that by using proposed sense amplifier in 2-cell per bit architecture the read yield is hardly effected by the variation of MTJ resistance but by using conventional sense amplifier in 1-cell per bit architecture, the read yield drops rapidly when Rp is decreased. That is because when Rp is reduced too much, the parasitical resistance in the read path will gradually dominate the whole resistance of read path and hence conventional current mode sense amplifier cannot precisely detect current change from ∣Rp - Rref ∣or ∣Rap – Rref∣.

3. New charge sharing scheme for word-line driver

Word line voltage is different during read and write cycle. It is raised to 1.5v or above during write operation but during read operation word-line voltage is maintained at 1.1v. To avoid device damage, in conventional word line driver the devices used to drive word-line are all thick gate-oxide devices and during read operation thick gate-oxide device is under driven at 1.1v. See Fig. 8. Since thick gate-oxide device M3 in Fig. 8 is under driven at 1.1v, Idsat will be smaller and word line rising time becomes much longer than spec requirement. In fact, longer word-line rising time becomes the limitation to achieving high speed operation. Hence it's important to figure out a way to reduce the word-line rising time and meanwhile still prevent device from being overstressed. A charge sharing scheme for fast word line switching under different power supplies is proposed and shown in Fig. 9(a). During write operation, M2 is on and raises the selected word line voltage to Vwrt which is 1.5v or above. At this period, M3 and M4 are off but M5 is on so that capacitor C1 is charged to Vwrt. After write operation is complete, M2 is off and M1 is on in order to discharge the word line to ground level. When entering into read operation, M1 and M2 are off and M3 and M4 are turned on. Meanwhile M5 are off too during a read cycle. Due to stored charge shared by word line parasitical capacitance, word line voltage is quickly coupled up to Vcore level (= 1.1v) and maintained at that level by M3. After one cycle of read operation is complete, M3 and M4 are off again and M1 is turned on to discharge the word line to ground level. Meanwhile before next cycle of read operation starts, M5 is turned on again to re-charge the capacitor C1 to Vwrt. Since C1 can be re-charged in between serial read cycles, the voltage of each word line can be pulled up to Vcore quickly and hence the word line voltage's rising speed will not be degraded no matter how many serial read cycles are performed. The simulation results shown in Fig. 9(b) demonstrate that by using this scheme, word line rising speed in read operation can be improved at least 770ps in worst case and at most 15.38ns in best case. Besides, no matter how much variation existing in Vcore, word line rising time only has very little change in read operation. The RC- load on the word line is considered based on the configuration of 512 word-line by 512 pairs of bit-line.

4. Write scheme

Since two cells constitute one data bit, one cell stores 0 (or 1) and the other cell stores the opposite state 1 (or 0), just like 6-T SRAM cell stores a state at one output node and stores the opposite state at the other output node. Therefore when an input data bits are written into cells through write buffer, the write buffer will generate a pair of binary values which are complement to each other on BL and BLB. Fig. 10 shows the write scheme for this 2 cells per bit architecture. One cell is written into Rp state and the write current is flowing from free layer to pinned layer, the other cell is written into Rap state and the write current is flowing from pinned layer to free layer.

5. Chip architecture

The cell array of this 1 Mb STT-MRAM is divided into 4 banks and each bank has 512 word-line by 512 pairs of bit-line. Dual power supplies of 2.5V and 1.1V are used for this chip.

6. References

[1] Tae Yun Kim; Fuminori Kimura; Yusuke Matsui; Tsutomu Yoshihara; Tsukasa Ooishi; Yuji Kihara; Masahiro Hatanaka; "A 75MHz MRAM with Pipe-Lined Self-Reference Read Scheme for Mobile/Robotics Memory System" pp 117-120, ASSCC, 2005

[2] Halupka, D.; Huda, S.; Song, W.; Sheikholeslami, A.; Tsunoda, K.; Yoshida, C.; Aoki, M.; "Negative-resistance read and write schemes for STT-MRAM in 0.13μm CMOS", ISSCC pp 256 – 257, 2010

[3] Kim, J.; Ryu, K.; Kang, S. H.; Jung, S.-O.; "A Novel Sensing Circuit for Deep Submicron Spin Transfer Torque MRAM (STT-MRAM)" IEEE Transactions, pp 1 – 5, 2010

[4] Kim, J.P.; Taehyun Kim; Wuyang Hao; Rao, H.M.; Kangho Lee; Xiaochun Zhu; Xia Li; Wah Hsu; Kang, S.H.; Matt, N.; Yu, N.; "A 45nm 1Mb embedded STT-MRAM with design techniques to minimize read-disturbance", VLSI Symposium, pp 296 – 297, 2011

[5] Shen, Weifeng; Mazumdar, Dipanjan; Zou, Xiaojing; Liu, Xiaoyong; Schrag, B. D.; Xiao, Gang; "Effect of film roughness in MgO-based magnetic tunnel junctions", Appl. Physics Letters, vol. 88, issue 18, pp 182508 - 182508-3, 2006

[6] Weisheng Zhao; Belhaire, E.; Chappert, C.; Mazoyer, P.; "Spintronic Device Based Non-volatile Low Standby Power SRAM" Symposium on VLSI, pp 40 – 45, 2008

[7] Chia-Tsung Cheng; Yu-Chang Tsai; Kuo-Hsing Cheng; "A high-speed current mode sense amplifier for Spin-Torque Transfer Magnetic Random Access Memory", IEEE MWSCAS, pp 181 – 184, 2010

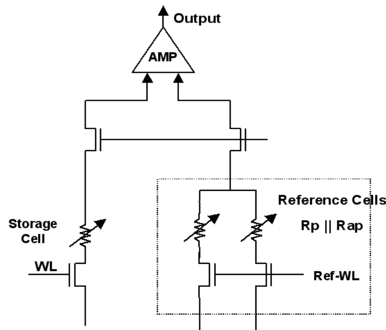


Fig.1: Conventional sense amplifier (Simplified) comparing storage cell with reference cells to determine high/low state of output

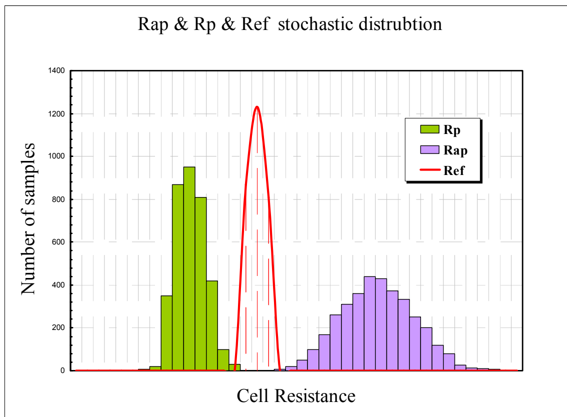


Fig.2 : Stochastically distributed MTJ resistance shows overlap among Rp, Rap and Rref. Situation that Rap < Rref and Rp > Rref can happen

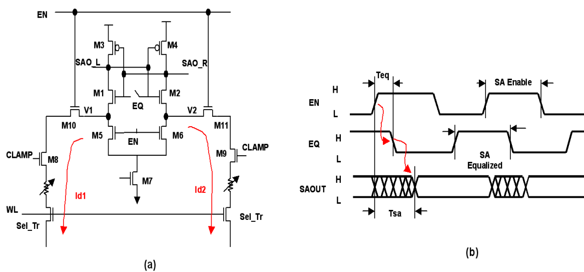


Fig. 3: (a) Circuit of current-mode differential sense amplifier. (b) Operation waveform of proposed sense amplifier

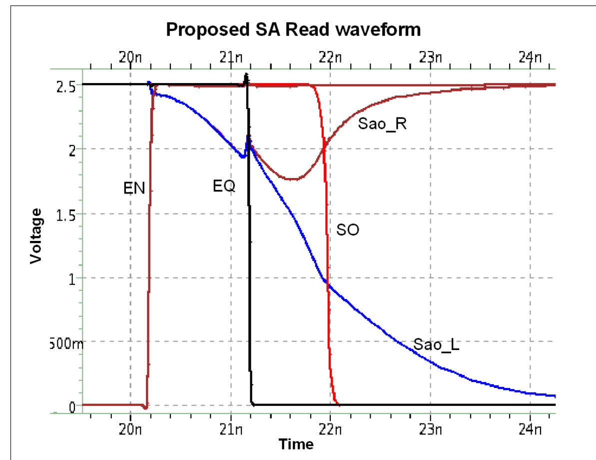


Fig. 4: Simulation shows proposed sense amplifier can achieve 2ns sensing speed.

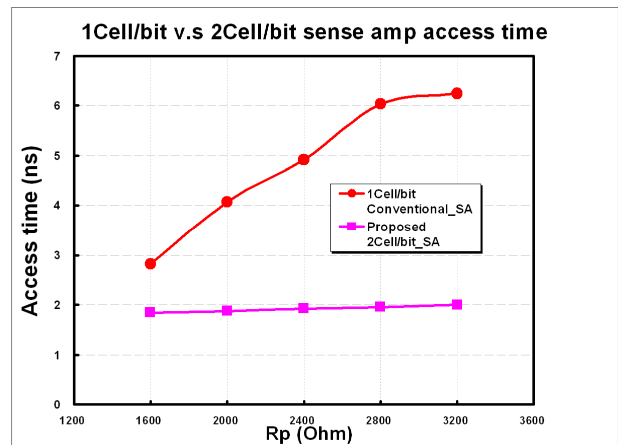


Fig. 5: Comparison of access time v.s. MTJ resistance variation between 2 types of sense amplifiers. Differential current mode sense amplifier keeps fast sensing speed across various MTJ resistance.

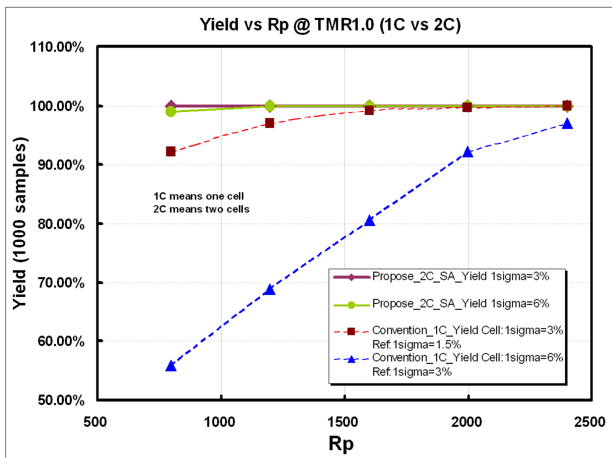


Fig. 6: Monte Carlo simulation shows 2-cell per bit architecture can have 100% read success as long as TMR is 40% or larger when 1 $\sigma \leq 6\%$ in MTJ resistance deviation.

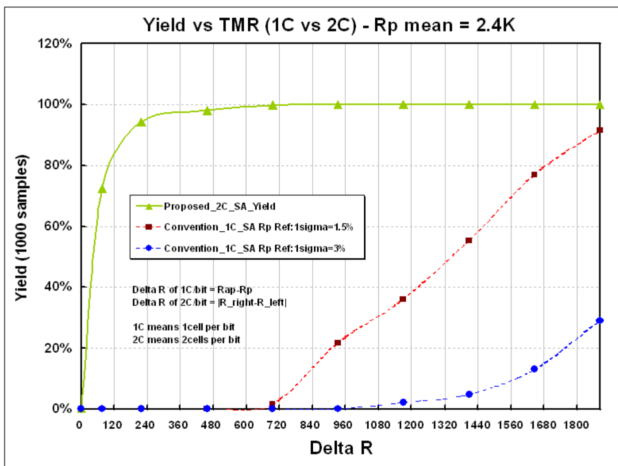


Fig. 7: Monte Carlo simulation shows the read yield of 2-cell per bit architecture is hardly effected by the variation of MTJ resistance. But 1-cell per bit is easily effected by Rp/Rap variation.

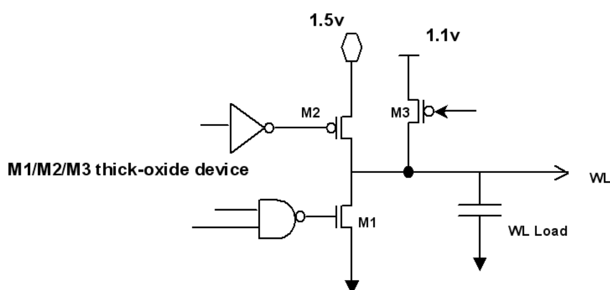
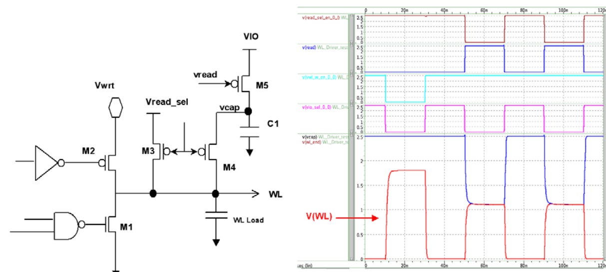
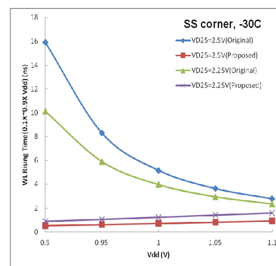


Fig. 8: Using conventional word-line driver to drive word-line to 1.5V or above during write and to 1.1V during read. M3 is under-driven at 1.1V.



(a)



VD25	Vdd	Original (ns)	Proposed (ns)	ΔT (ns)	Improved (%)
2.5	1.1	2.77	0.914	1.856	67
	1.05	3.63	0.805	2.825	78
	1	5.15	0.7	4.45	86
	0.95	8.28	0.607	7.673	93
	0.9	15.9	0.524	15.376	97
2.25	1.1	2.35	1.58	0.77	33
	1.05	2.95	1.4	1.55	53
	1	3.97	1.22	2.75	69
	0.95	5.9	1.04	4.86	82
	0.9	10.14	0.877	9.263	91

*Compared with adjusted WL driver size

(b)

Fig. 9: (a) Word-line driver with effective charge sharing scheme and waveform. (b) Performance comparison shows that word-line rising time can be as small as 500ps by using the proposed scheme.

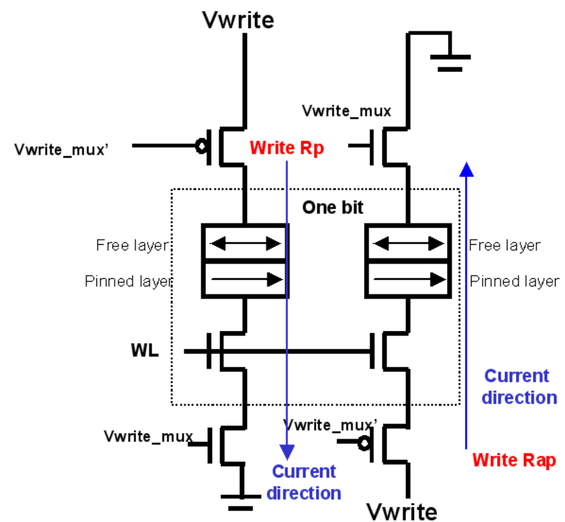


Fig. 10: Write scheme for 2-cell per bit