Chapter 3
Variation-Tolerant SRAM Write and Read Assist Techniques

3.1 Introduction

There are stringent requirements to lower the power consumption and achieve higher speed in today's SoCs and microprocessors. Voltage scaling combined with technology scaling has been effective in achieving both requirements. However, the large increase in random variations in advanced CMOS technology nodes has created several challenges for SRAM design. This is exacerbated by the high demand for low voltage and high density memories for SoC [1]. Dealing with SRAM cell stability at lower supply voltages is currently one of the biggest challenges in SRAM design [2–4].

3.2 SRAM Stability Metrics

Several metrics have been proposed to analyze the SRAM stability margins. These metrics are beneficial in studying how different design options affect SRAM stability, analyzing the effectiveness of assist techniques, predicting yield, optimizing the memory V_{min}, as well as determining the process window at low voltage, as shown in Fig. 3.1. At low voltage, the write and read margins reduce, which diminishes the operation window. In the next section, we study different metrics used for write and read margins.

Static metrics are typically easier to simulate and measure since they assume that bitcell stability is time-independent. Static metrics have been used extensively in measurement and simulation of bitcell stability [3, 4, 6–11].

Dynamic stability metrics are time-dependent, and have been proposed to improve the accuracy of determining SRAM stability. However, dynamic metrics are more difficult to simulate and analyze since they require transient simulation, while static

M. H. Abu-Rahma and M. Anis, *Nanometer Variation-Tolerant SRAM*,
DOI: 10.1007/978-1-4614-1749-1_3,

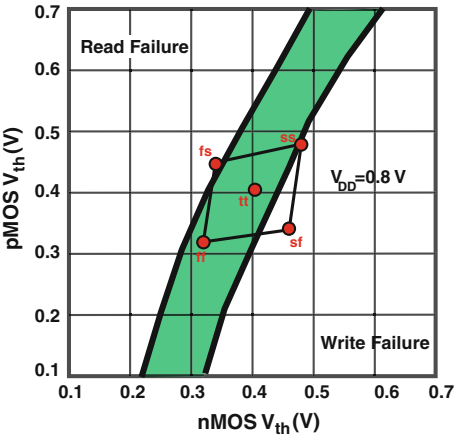

Fig. 3.1 Write and read stability metrics used to determine the V_{th} window of operation [5]

metrics may be determined using DC simulation. In the next sections, we review the commonly used static and dynamic metrics for SRAM write and read stability [3, 4, 6–11].

3.2.1 Static Write Margin

In write operation, BLB is pulled to zero using write driver, while WL is enabled, as shown in Fig. 3.2. Therefore, the NMOS PG2 is turned ON, which results in a voltage drop in the storage node QB holding data 1. When this voltage falls below $V_{\text{DD}} - V_{\text{th}}$ for the PU1, PU1 starts the feedback action. For stable write operation, PG2 should be stronger than PU2.

To quantify write stability, one of the most widely used metrics is the write static noise margin (WSNM), which uses a similar concept to the read SNM described in Sect. 3.2.3. In this approach, the voltage transfer characteristics (VTC) of the two sides of the bitcell are obtained from DC simulation, while the bitlines (BL and BLB) are driven to the write operation condition (i.e., BL connected to V_{DD} and BLB to zero, as shown in Fig. 3.3). For a successful write operation, there should only be one cross-point between the VTCs of the bitcell inverters, which implies that there is only one solution (the cell is monostable), as shown in Fig. 3.3a. The separation between the two VTCs indicates the bitcell write margin. If the separation is reduced, the bitcell write ability worsens until the separation reaches zero, which means that there are two or three cross points, so the bitcell changes from monostable to bistable or metastable. Therefore, to measure write stability, the WSNM is defined as the width of the minimum square that can be enclosed between the two VTCs. Using this definition, the bitcell failure condition is defined as WSNM $=0$ [12]. Improvements in the definition of WSNM have been proposed [6].

Another write stability metric, the bitline write margin (WVBL), is derived from the BL voltage, as shown in Fig. 3.3b. In this method, the bitcell is configured in write operation with one bitline (BL) connected to V_{DD}, and the other bitline (BLB)

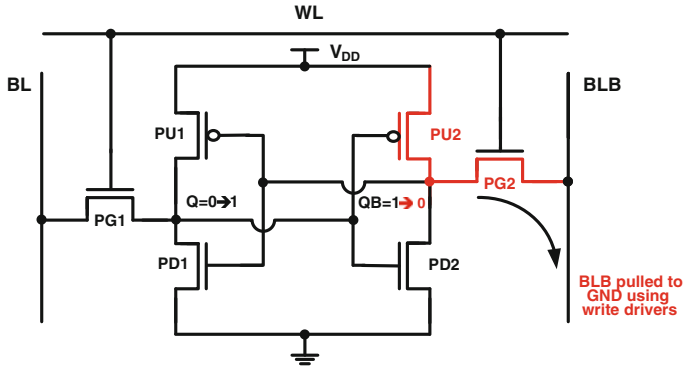

Fig. 3.2 Bitcell in write operation

swept from V_{DD} to zero. As the BLB voltage is reduced, the voltage at which the internal nodes (Q and QB) flip is defined as the bitline write margin (WVBL). The lower the bitline, the more difficult it is to write the bitcell. WVBL can also be defined as the maximum bitline voltage at which the bitcell can be written. The write failure criteria in this case is defined as WVBL $=0$ [12].

Similar to the WVBL write margin definition, the WL voltage can be used to assess write ability. In this case, the WL voltage is swept from zero to V_{DD} to find the minimum voltage at which the bitcell is written (Q and QB flip) and is called $WVWL$, as shown in Fig. 3.3c. The lower the value of WVWL, the easier it is to write on the bitcell. The write failure criteria is defined as WVWL $\geq V_{DD}$ [12].

Other DC write margins have been proposed, such as the N-curve method, as shown in Fig. 3.3d, which uses current information instead of voltage [4, 12].

It is important to note that due to device variations, the bitcell margin needs to be evaluated on both sides of the bitcell (i.e., in the case of writing zero via BL or via BLB), since variations will cause the margin to be different for each side of the cell. Hence, the write margin becomes the worst of the two sides margins.

3.2.2 Dynamic Write Margin

The main limitation of DC metrics is that they ignore the nonlinear dynamics of write operation by neglecting the bitcell time-dependence. For example, static metrics assume that the WL pulse width is infinite and only account for the voltage amplitude. In reality, the write pulse width is proportional to the memory cycle time. In addition, static margins assume that one bitline is actively pulled to zero while the other bitline stays at V_{DD}, which is not realistic in dynamic operation.

Figure 3.4 shows the distribution of bitcell write time, defined as the time required for the internal storage node to flip after the WL is enabled. The distribution is

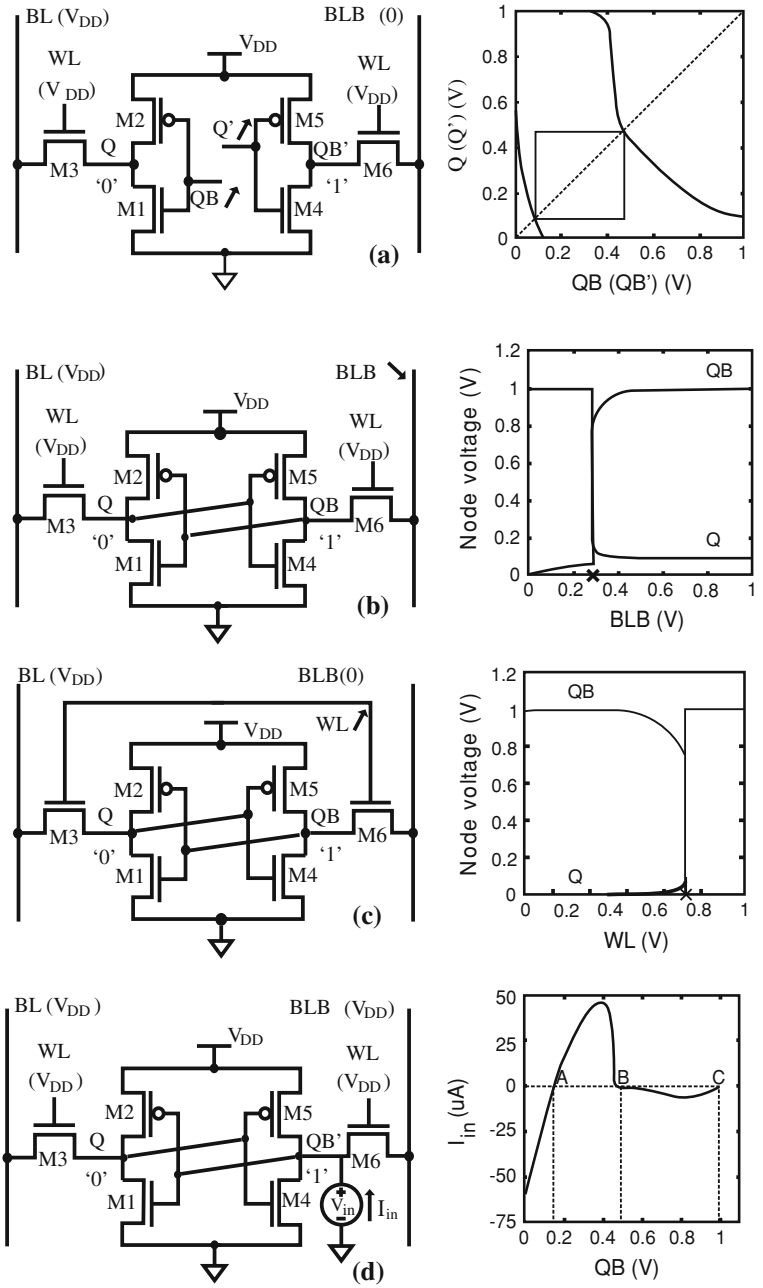

Fig. 3.3 Different DC definitions for bitcell write margin including **a** write SNM, **b** WVBL, **c** WVWL, and **d** *N*-curve [12]

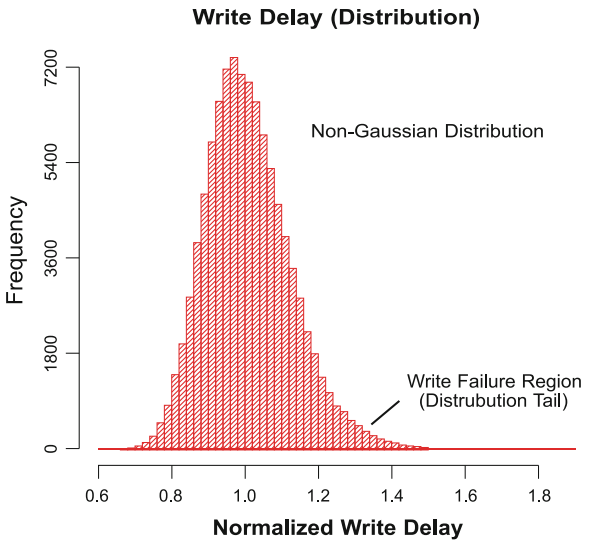

Fig. 3.4 Distribution of bitcell write time which shows that the distribution is non-Gaussian with a long tail toward higher write time [13]

not a Gaussian distribution, and shows a long tail toward higher bitcell write time. Dynamic write failure can be defined as the probability that the bitcell write time is larger than the WL pulse width [13–15]. Other definitions use the differential voltage between bitcells storage nodes at the end of write operation, as shown in Fig. 3.5. The impact of WL pulse width on write dynamics is shown in Fig. 3.6. As the WL pulse width increases, the failure probability (or measured fail bit count) decreases and approaches the DC value.

Comparisons between static and dynamic metrics have shown that static metrics are optimistic when predicting write failure probability [10, 13, 16], in some cases by 3 orders of magnitude [10]. This underestimation of failure results from static metrics' assumption that the WL pulse width is infinite; therefore, failures due to insufficient WL pulse width are not captured in the static write margin. This causes dynamic write V_{min} to be typically larger than static V_{min} [15]. Moreover, the fact that the distribution of dynamic write margin is not Gaussian complicates the statistical yield estimation [12].

One of the interesting findings from dynamic write stability is the relationship between PMOS pull-up strength and write fails. Adjusting the PMOS pull-up to be weaker improves static write margin; however, it may reduce dynamic stability since the bitcell write time (time to flip) increases due to slower rising time of the bitcell storage node [6, 19]. As shown in Fig. 3.7, this type of dynamic failure causes frequency-dependent failure since the bitcell internal node stays near the metastable point and may take several cycles to reach V_{DD}. Hence, read operation fails if it is performed immediately after write operation [6, 19].

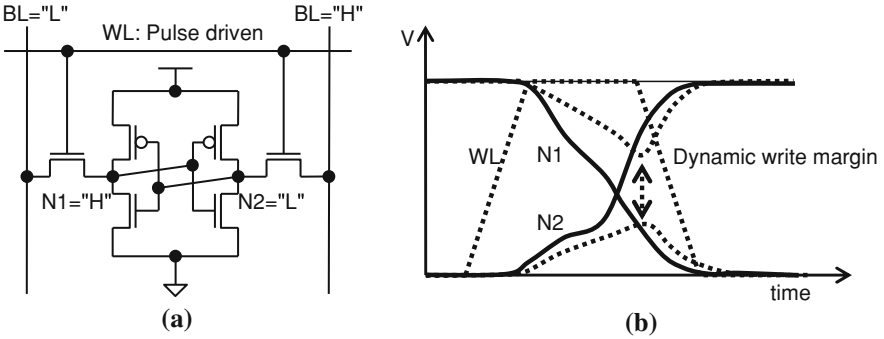

Fig. 3.5 Dynamic write margin definition based on transient analysis. **a** Memory cell state during dynamic (transient) analysis **b** Waveform during dynamic analysis [17]

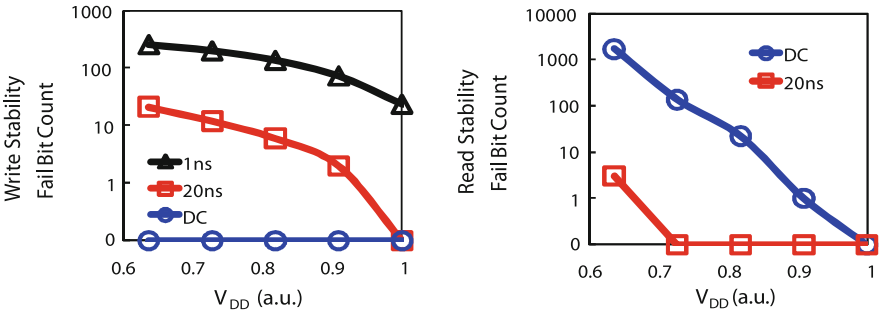

Fig. 3.6 Measured fail bit count dependence on V_{DD} and wordline pulse width for write and read stability [18]

3.2.3 Static Read Stability

In read operation, read disturb may occur after the wordline is enabled. The voltage at the internal storage node storing a zero (Q) slightly rises due to the voltage divider between the pass-gate transistor (PG1) and the pull-down (PD1), as shown in Fig. 3.8. If the voltage at Q rises above the threshold voltage of PD2, the cell may flip its state. In this case, stable read operation requires that PD1 should be stronger than PG1. Read stability failure increases with process variations, which affect all the transistors in the bitcell [3, 13, 20].

To quantify the bitcell's robustness against this type of failure, static noise margin (SNM) is one of the most commonly used metrics [21]. SNM is defined as the maximum amount of voltage noise that a cell can tolerate [21]. SNM is calculated by finding the largest square which fits inside the VTCs (butterfly curves), as shown in Fig. 3.9. A larger SNM implies higher robustness for the bitcell. However, due to WID variations, each transistor in the bitcell experiences different magnitude of variation, hence, the symmetry of the bitcell is lost. This causes large spread in SNM

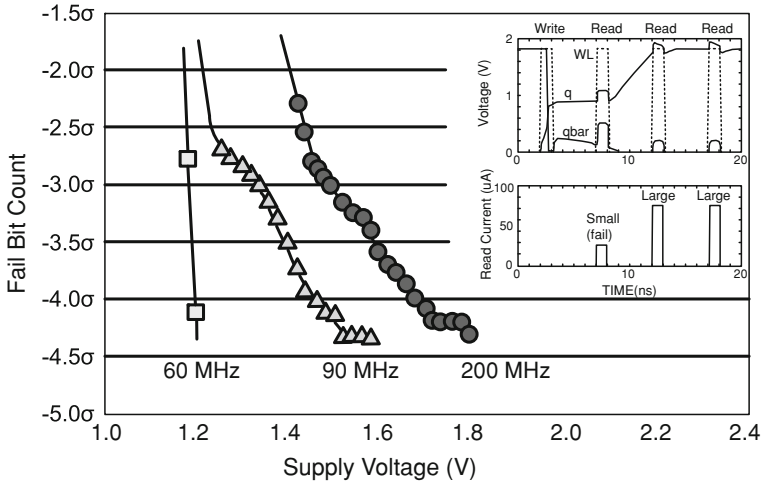

Fig. 3.7 Measured frequency-dependent failure attributed to dynamic write fail. The inset shows simulation of the write failure where the bitcell internal node (V_{n1}) does not fully reach V_{DD} after write operation in one cycle. The failure appears as a frequency-dependent failure since at lower frequency V_{n1} has sufficient time to reach V_{DD} before the next cycle [19]

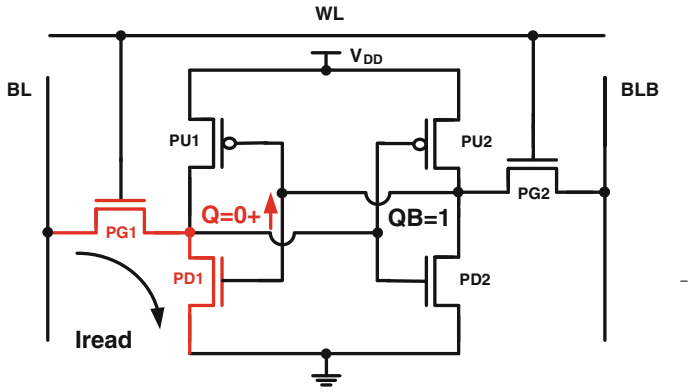

Fig. 3.8 Bitcell in read operation

as shown in measured SNM butterfly curves in Fig. 2.29. Read stability failure occurs if SNM reaches zero [13, 20, 21].

Other techniques for static read stability margins use the bitcell current such as the N-curves method [4, 23]. Because SNM is estimated using DC simulation, this method of measuring read stability does not account for the dynamic nature of read operation. In the next section, we present metrics for dynamic read stability.

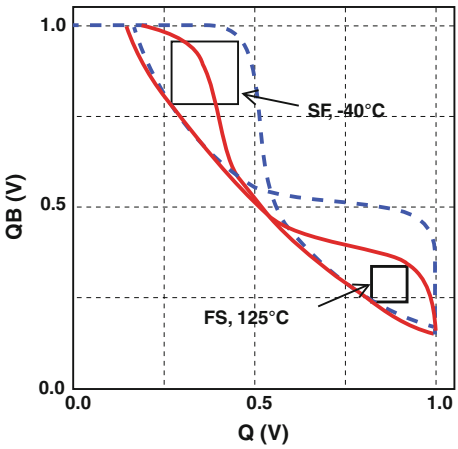

Fig. 3.9 SNM butterfly curves for a 45 nm SRAM bitcell simulated at skewed process corners, FS: fast NMOS, slow PMOS and SF: slow NMOS, fast PMOS. In this simulation, WID variations are not included, so the curves are symmetrical [22]

3.2.4 Dynamic Read Stability

Conventional SNM read stability metric cannot capture the dynamic behavior of the WL, BL, and internal storage nodes of the bitcell. Several works have recently investigated how the dynamic operation of the bitcell affects read stability. DC SNM gives pessimistic results compared to dynamic stability because DC SNM assumes the WL pulse width is infinite and that the BL is actively pulled to V_{DD} in read operation. In reality, the WL pulse width is typically a percentage of the cycle time, hence, the bitcell disturb may not have enough time to flip within the WL pulse width. Also, BLs are typically not pulled high in read operation, but are precharged before the WL is enabled, therefore, the BL capacitance has strong impact on the dynamic stability of the bitcell. Conventional static SNM can overestimate the probability of read flip failure by 6 orders of magnitude [10]. Moreover, static metrics cannot account for dynamic issues such as the impact of supply noise on stability [10].

Several dynamic read margin definitions have been proposed [10, 17, 24, 25]. Figure 3.10 shows the definition of dynamic read stability based on transient simulation [17]. The margin is defined as the minimum differential voltage between the storage nodes when the WL is high. The dynamic read margin strongly depends on the WL pulse width and the bitline capacitance [17, 26].

Several models have been developed to study the dynamic stability. Rigorous analysis using nonlinear system theory has shown that two conditions are required to cause a bitcell flip: the noise should, 1- exceed the static SNM, and 2- be sustained longer than a minimum critical duration [9, 25, 27, 28].

The impact of WL pulse width on dynamic read stability is shown in Fig. 3.6. As the WL pulse width decreases, the read flip failure probability (or measured fail bit count) decreases drastically. The impact of bitline capacitance on read failure can be estimated by simulating dynamic read stability, and shows that reducing the BL capacitance results into significant read stability improvement. Figure 3.11 shows

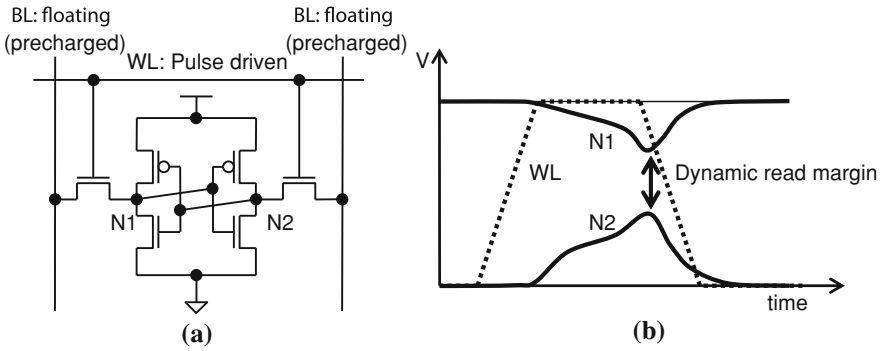

Fig. 3.10 Dynamic read margin definition based on transient analysis with the bitlines floating [17] **a** Memory cell state during dynamic(transient) analysis **b** Waveform during analysis

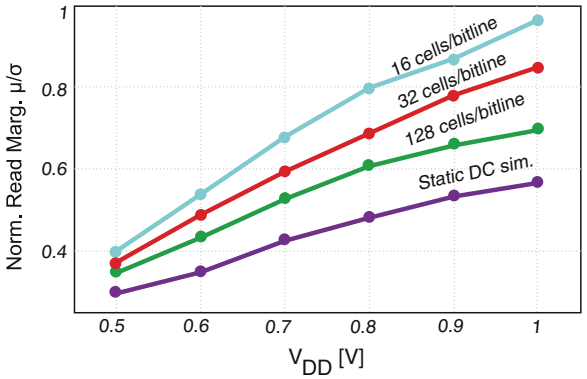

Fig. 3.11 Simulation of static and dynamic read stability margins for a 6T bitcell in 28 nm technology. Dynamic margins show much lower $V_{\min}$ compared to static [26]

simulation results for static and dynamic read margins for 6T cell in 28 nm technology. For the same read margin (constant μ/σ), $V_{\min}$ calculated by dynamic read margin is much lower $V_{\min}$ than that predicted by static simulation (0.7 V for dynamic while 0.9 V for static). Moreover, reducing the bitline capacitance by reducing the number of bitcells per bitline, can be used to further reduce by 100 mV when using 16 cells/bitline instead of 128 cells/bitline. This allows the bitcell to operate at lower voltage at the expense of larger area overheard, since additional read circuitry is required when using shorter bitlines (larger number of memory banks) [10, 24, 26, 29]. The dynamic read stability concept has wide applications in the area of read assist techniques, as will be discussed in the next sections.

Table 3.1 Impact of bitcell size increase on V_{min} reduction [30]

Bitcell area (normalized)	V_{min} (V)
1	1.1
1.11	1.0
1.22	0.8
1.44	0.7

3.3 Bitcell Design for Low Voltage Operation

The bitcell V_{min} depends strongly on random variations. To reduce SRAM V_{min}, the simplest technique is to increase the 6T bitcell area. Since the V_{th} variations are inversely proportional to the square root of device area, increasing the device area reduces the variations, which helps improve the bitcell stability, as shown in Table 3.1. However, increasing bitcell size comes with large area overhead, which increases the product cost significantly and limits the usage of large area 6T bitcells to small density memories.

Due to the conflicting requirements of read and write stability, it is challenging to reduce the V_{min} for 6T bitcell without a large increase in bitcell area. Instead of 6T, an 8T bitcell has been proposed to lower V_{min} by decoupling the read and write operations. Figure 3.12 shows an 8T bitcell, which has isolated read and write ports. The read operation is single-ended, using the read bitline (RBL) and read wordline (RWL), which eliminates read disturb failure since there is no contention between the pass-gate and the pull-down. The write operation is performed by differential write using WBL, $WBLB$, and the wordline (WL) which is similar to that in 6T bitcell. Since the WWL, WBL, and $WBLB$ are not used in read operation, bitcell devices they can be optimized solely for write operation where PG/PU current ratio can be increased without considering the PG/PD read constraint in 6T bitcell. Since the 8T bitcell has a dedicated write port, it does not suffer from read disturb and the limit for low voltage read stability is retaining the data (retention failure). Figure 3.13 compares the SNM for 6T and 8T bitcells in 32 nm technology node. The SNM for the 8T bitcell is 76 % higher than the 6T bitcell which allows the 8T bitcell to have lower V_{min} than the 6T.

While the 8T bitcell provides lower V_{min}, it adds 30–40 % area penalty [26, 31], as shown in Fig. 3.14. Also, 8T read operation is single-ended, which requires large signal sensing and limits the maximum number of cells on the bitline. The other option for read operation is to use a sense amplifier with one input connected to a reference voltage, which is typically slower and more susceptible to supply noise compared to differential sensing. Another difference between 8T and 6T bitcells is that the 8T bitcell does not allow column muxing, which is also called column interleaving. Therefore, the periphery circuits for 8T bitcells are larger than the 6T since each column requires dedicated read and write circuitry and cannot be shared for multiple columns. Moreover, without column interleaving, 8T bitcells become

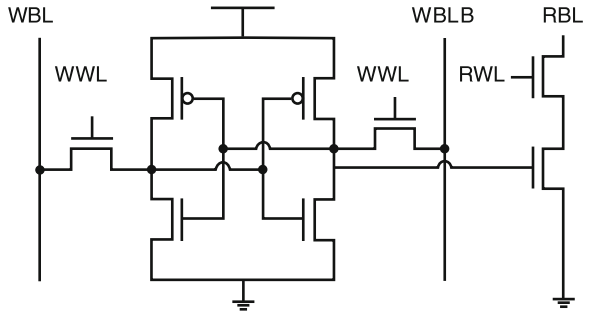

Fig. 3.12 8T bitcell with separate read and write ports which provides lower V_{min} compared to conventional 6T bitcell [31]

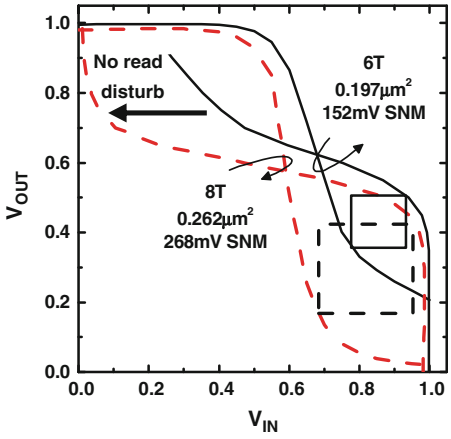

Fig. 3.13 8T bitcell provides higher SNM compared to 6T due to elimination of read disturbs [31]

more sensitive to soft errors, where multi-cell upsets (MCUs) cause multiple fails in the same word and may not be corrected using ECC, as discussed in Chap. 2.

In efforts to achieve the benefits of an 8T cell while reducing the area overhead, a 7T bitcell has been proposed as shown in Fig. 3.15. An NMOS transistor (N5) is added between the storage node (V2) and the pull-down device (N2) [32]. The WWL is used for write operation while WL is used for both read and write operations. The protection device N5 is controlled using the complement of WL ($/WL$). In read operation, WL is high and $/WL$ is low, hence, $N5$ prevents $V2$ from decreasing and the cell cannot flip even if V1 increases. In write operation, both WL and WWL are high, and one of the bitlines is pulled to zero, similar to 6T bitcell write operation. Read operation is single-ended using only BL.

A major disadvantage of 7T bitcells is that if $V2$ is storing a zero, then the node is floating, which requires a dynamic retention condition. If the WL activation period is longer than that the data retention time of $V2 = 0$, the stored data will be lost due

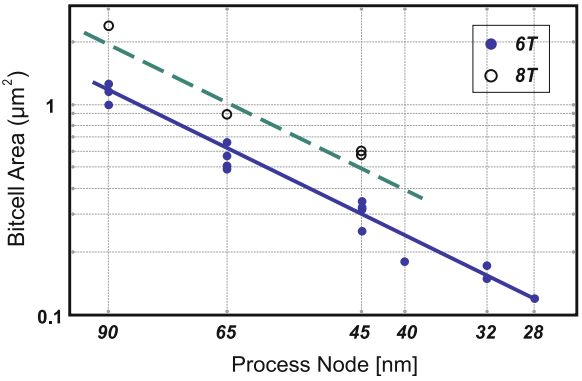

Fig. 3.14 Bitcell area scaling for 6T and 8T bitcells [26]

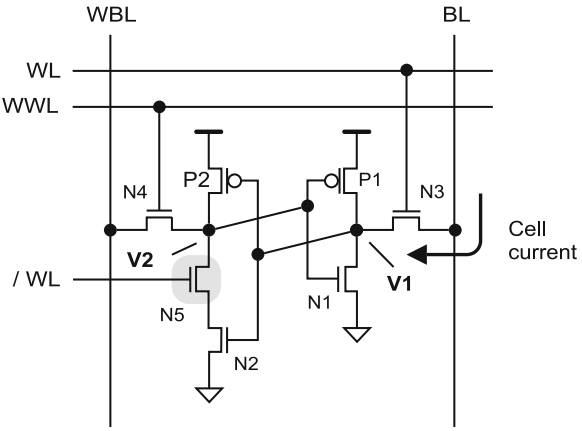

Fig. 3.15 7T bitcell with protection device to prevent read disturb [32]

to leakage current. In order to achieve stable read operation, the WL pulse width should be lower than the retention-time-to-fail due to leakage.

To push the SRAM operation down to subthreshold region, 10T bitcells have been proposed [33, 34]. Figure 3.16 shows a 10T subthreshold bitcell with fully differential read that can support column interleaving. In read mode, the WL is enabled and $VGND$ is pulled to ground while WWL is disabled. Therefore, the storage nodes Q and QB are isolated from the bitlines in read mode, and the read SNM is the same as the hold SNM of conventional 6T SRAM. In write operation, both WL and WWL are enabled to transfer the write data to the cell node. To improve write ability of 10T bitcells at low voltage, WWL boosting has been proposed.[1]

Recently, a 10T bitcell that allows contention-free writes and improves V_{min} has been proposed [34], as shown in Fig. 3.17. In this bitcell, complementary WL

[1] WL boosting is one of the write assist techniques that will be described in the next sections.

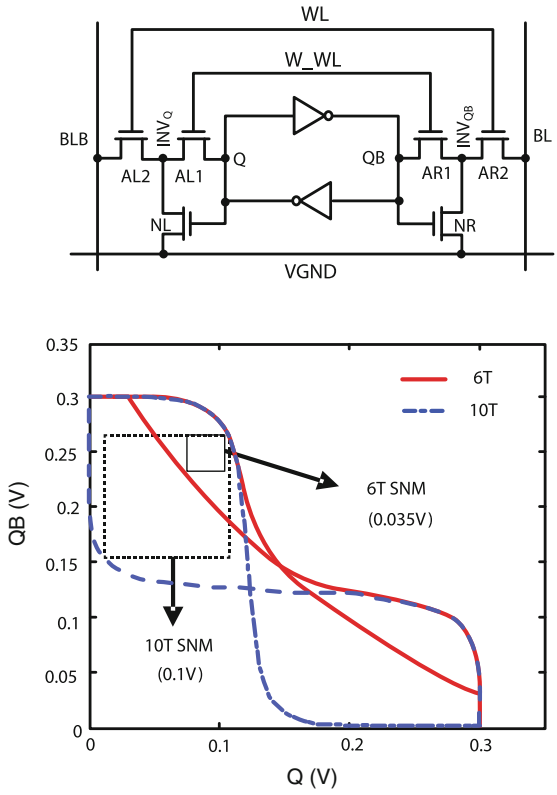

Fig. 3.16 10T bitcell that allows SRAM operation down to subthreshold region [33]

($WRWL$ and $\overline{WRWL}$) are used to access the bitcell in write operation using a transmission gate consisting of NMOS and PMOS transistors. Also, the write wordlines are used to disable the bitcell internal pull-up and pull-down devices, allowing the write operation to be contention-free when the $WRBL$ is used to write the bitcell. The read operation uses a separate read port, similar to 8T bitcells [34]. In addition to the bitcells mentioned above, other types of bitcells have been proposed to improve V_{min}, such as asymmetrical 6T bitcell [24].

3.4 Write and Read Assist Circuits

As discussed earlier, one of the biggest challenges in designing SRAM bitcells is to achieve balanced write and read margins. This is because of the conflicting requirements on the bitcell design, which are difficult to balance by conventional sizing and V_{th} adjustments [35]. To address this challenge, extensive research has been done in

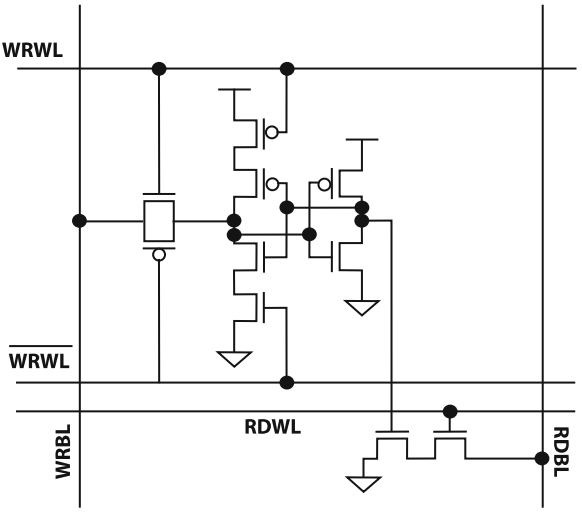

Fig. 3.17 10T bitcell that allows contention-free writes [34]

the area of write and read assist techniques to reduce the conflict between write and read operation and allow the memories to operate at lower supply voltage.

Read assist techniques include using higher supply voltage for the bitcell, wordline under-drive (WLUD), lower bitline capacitance, lower bitline voltage, modulating device characteristics using body bias, and read and write back [36, 37]. Since write operation in 6T bitcells has conflicting requirements with read operation, write assist techniques tend to perform the opposite operation compared to read assist [36, 37]. Write assist methods include lowering bitcell supply voltage, wordline boosting, negative bitline write, and body bias to improve the strength of NMOS pass-gate versus the PMOS pull-up. Assist techniques can be broadly classified into two categories: single supply and dual supply techniques. In the next sections, we discuss the implementation details of different state of the art write and read assist techniques [36–42].

3.5 Dual Supply Assist Techniques

In dual supply approaches, one power supply is used for the bitcell array, and a different supply voltage is used for the periphery circuit [30, 42–44]. In this way, the SRAM's supply voltage can be kept at a higher voltage compared to that of logic. The logic's supply voltage can be scaled to reduce power consumption, while the SRAM's supply is kept constant. This ensures that the SRAM read stability is sufficient, since SRAM failures can be eliminated at higher supply voltage (at higher V_{DD}, the impact of variations decreases). Many varieties of dual supply memories have been proposed [38, 39, 41, 42, 45–48].

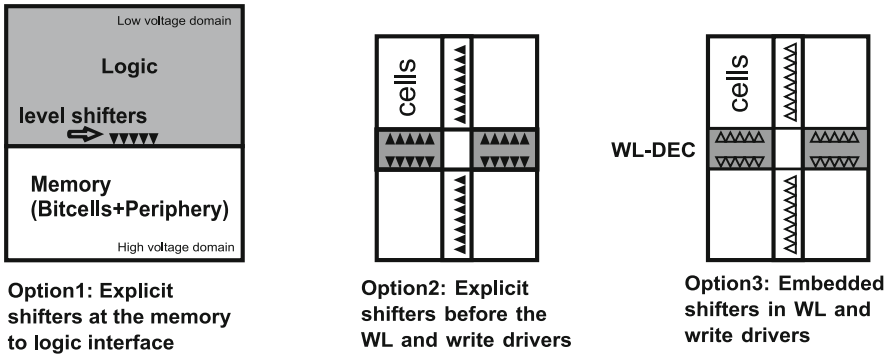

Fig. 3.18 Different options for partitioning the voltage domains in dual supply read and write assist techniques [30]

An important design decision related to dual supply assist is how to partition the voltage domains. Different types of partition have been proposed that cover different points in the tradeoff between memory area, power, and bitcell stability, as shown in Fig. 3.18. In the simplest approach, the whole memory (bitcell+periphery) lays on the high voltage domain, hence, level shifters are only needed in the interface between memory and logic [30]. The second option is to place the level shifters internal to the memory at the boundary between the bitcell and periphery, either using explicit level shifters or embedded level shifters, as shown in Fig. 3.20. The first option provides the smallest area since it involves the lowest number of shifters; however, the power consumption is the highest because the memory periphery is at the high voltage domain and cannot be scaled. Using level shifters inside the memory, the power consumption can be reduced by 20–30 %, and the use of embedded level shifter further reduces the area overhead [30]. Figures 3.19 and 3.20 shows the implementation of level shifters in the BL and the WL drivers, respectively [30]. These types of embedded level shifters help to reduce the memory area compared to using explicit level shifters, and prevent leakage paths in the boundary between the low and high voltage domains.

In concept, designing the memory so that the read and write operations use different supply voltages significantly improves cell stability. In read operation, WL voltage is lower than the array voltage, which increases SNM (pass-gate drive capability decreases). In write operation, the WL voltage is higher than the array voltage, which improves the bitcell write margin [48]. The challenge is that in high density memories, column interleaving is often used to improve area efficiency and reduce soft error rate. Hence, in both read and write operations, bitcells that are on the same wordline will experience a dummy read operation (half-selected bitcells). Therefore, the simple implementation of static dual supply along the wordline direction cannot address read and write stability for column interleaved memories.

To address the issue of read and write for column interleaved memories, column-based dual-power supply has been proposed [35, 45, 49]. The bitcell's dynamic dual

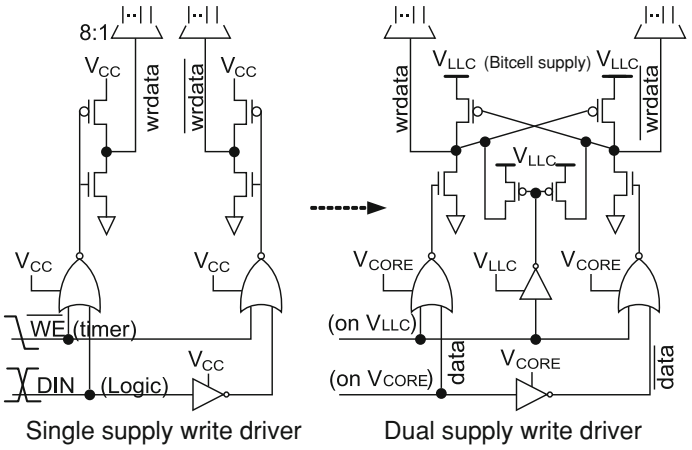

Fig. 3.19 Embedded BL level shifter implementation in dual supply memories [30]

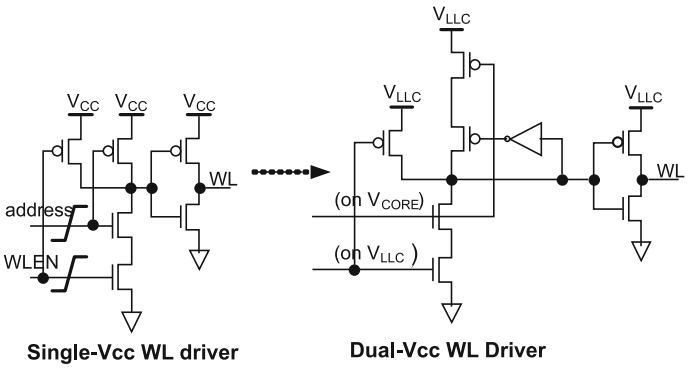

Fig. 3.20 Embedded WL level shifter implementation in dual supply memories [30]

supply is switched during read and write operations at a column level as shown in Fig. 3.21. To improve read stability, a higher VCC is connected to the memory cells in read operation, which increases the cell SNM. In write operation, a lower voltage is connected to the memory cell which makes the cell easier to flip. The change from high to low voltage is accomplished using a column-based power multiplexer (mux) as shown in Fig. 3.21. Therefore, only the column that is being written experiences low voltage $V_{\text{CC\_lo}}$, while half-selected bitcells that are on the wordline stay connected to $V_{\text{CC\_hi}}$, which ensures that the read stability for those cells is not reduced. The low supply voltage needs to be higher than the retention V_{min} to prevent bitcells on the column being written from losing the stored data. Another key point in this approach is that the local VCC needs to be stabilized before the WL is turned on to prevent any stability issues, which requires that the VCC select signals are generated early enough.

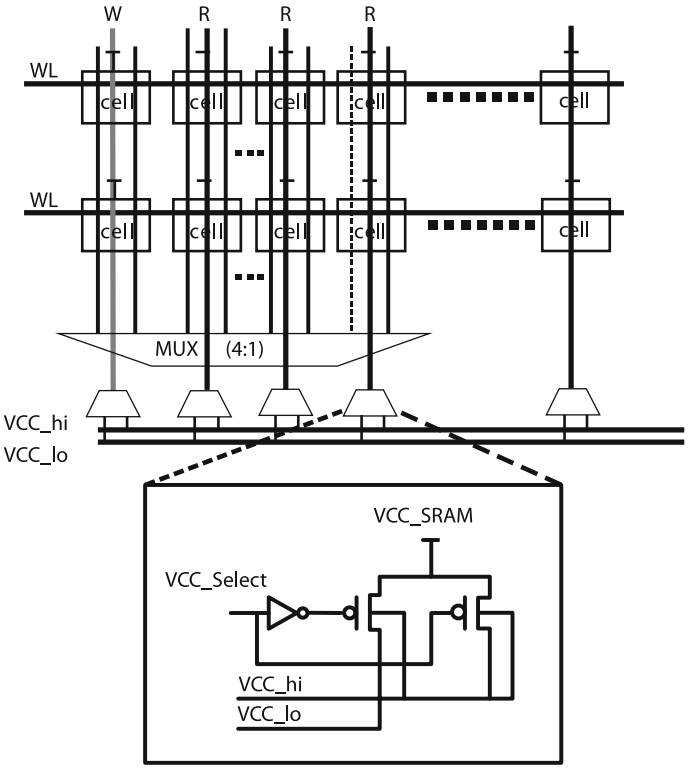

Fig. 3.21 Write assist using column-based dual supply [43]

In dual supply read assist techniques, the bitline can be precharged to the high [30, 39] or low voltage [42, 50]. Figure 3.22 shows the impact of lowering the bitline voltage on read SNM. As the bitline voltage decreases, the strength of the NMOS pass-gate decreases, hence, the SNM initially improves. However, as the bitline voltage decreases further, SNM starts decreasing after a certain point since the bitcell experiences a dummy write operation. This limits the minimum precharge voltage of the bitlines, and the power reduction achieved by lowering the logic voltage [50]. Figure 3.23 shows an implementation of dual supply read assist with the WL and bitcell supply at high voltage while the bitlines are precharged to the low supply voltage. One way to extend the range of the low supply voltage is to have the cell voltage track the logic voltage, while still maintaining higher voltage than the bitlines [50].

To improve the effectiveness of dual power supply SRAM, programmable WL voltage level control has been proposed [45]. In this technique, both the low and high supply voltages are provided to the WL drivers, which are level programmable using a digital code, as shown in Fig. 3.24. The WL voltage level is controlled adaptively depending on the process corner. So that the WL control balances between read disturb and write failure, as shown in Fig. 3.25

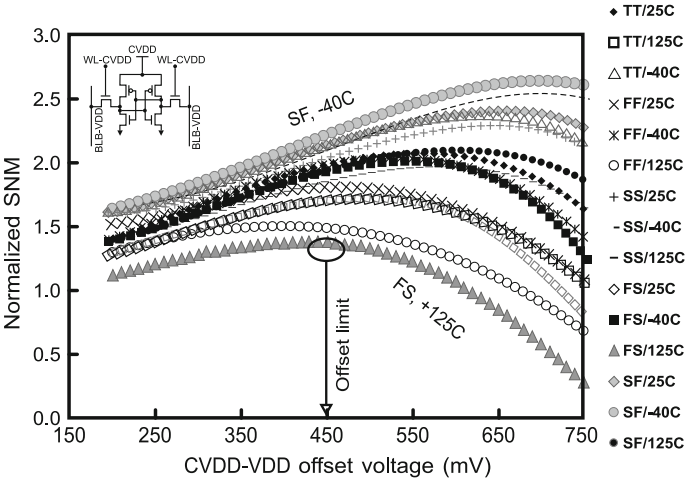

Fig. 3.22 Impact of lowering the bitline voltage on read SNM for different process corners and temperatures in 45 nm technology [50]

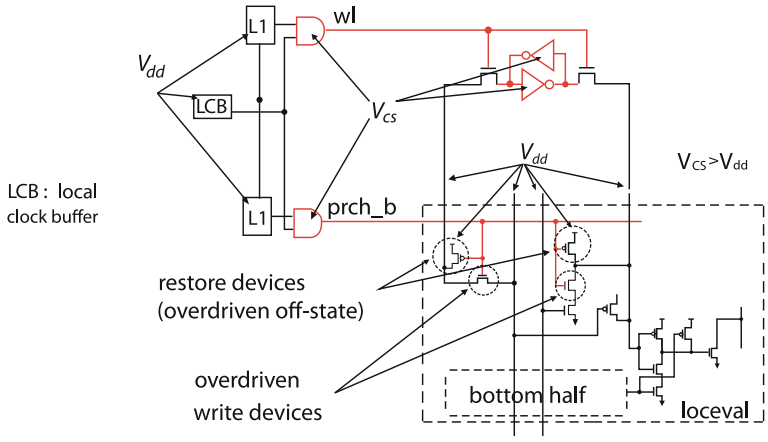

Fig. 3.23 Implementation of dual supply read assist with the WL and bitcell supply at high voltage while the bitlines are precharged to the low supply voltage [42]

Another technique that utilizes dual supplies use two voltage levels to control the body bias [44]. Using the body bias for the NMOS and PMOS separately can improves the SRAM margins, since SRAM margins are sensitive to the ratio of NMOS to PMOS drive. In column-based body bias, the PMOS pull-up is forward biased in read operation to improve read stability. In write operation, for the columns selected for write, the body bias is reversed, which weakens the pull-up device and improves write ability. Half-selected bitcells stay at the forward-bias conditions, which is the same as in the read operation.

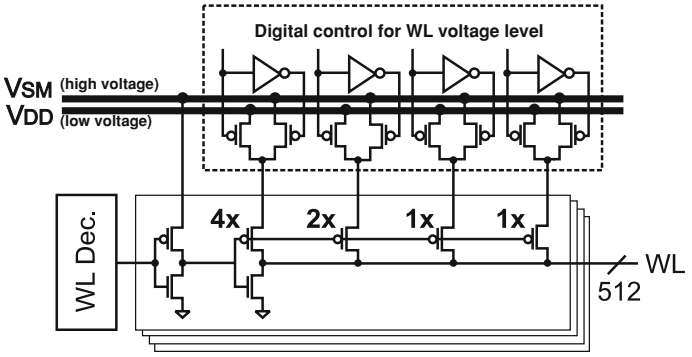

Fig. 3.24 Dual supply read assist using adaptive WL voltage level control [45]

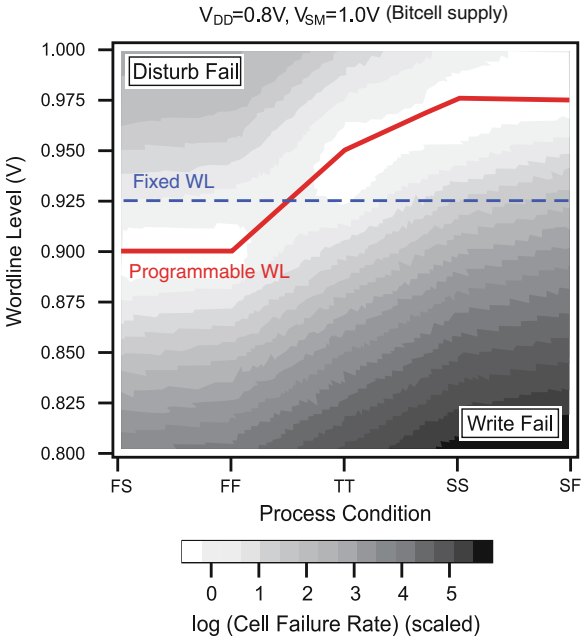

Fig. 3.25 Programmable WL used to track the process corner and apply the optimal WL voltage to balance between read and write failures [45]

3.6 Single Supply Write and Read Assist Techniques

Although the dual V_{DD} concept seems relatively simple, it introduces significant challenges. First, voltage level shifters are required at the interface between the bit-cell array voltage and the logic voltage. These level shifters tend to consume large

area which lowers the memory's area efficiency. In addition, level shifters introduce additional delay in the memory critical path and hence, cause speed penalty. Moreover, the power grid design at the chip level becomes challenging since a dedicated power grid is required for memories, which adds cost due to routing and extra metal resources required to distribute the additional power supply.

The degree of difficulty in implementing a dual supply power grid depends on the chip architecture. For example, microprocessor designs use relatively few kinds of SRAM architectures having large capacity, such as caches, which can be physically placed in a close proximity on a chip [22]. This simplifies the design of a dual power grid since all the memories are physically located near each other. However, in SoC design, there are typically hundreds of SRAM architectures and they are not necessarily placed in close proximity on a chip. This makes it difficult to have a dedicated power grid for all the memories. Therefore, for an SoC, it is always desirable to use a single power supply for the SRAM [5, 22, 51, 52].

In single supply assist techniques, additional circuitry is added to assist write and read operations and provide adequate margins. In the next section we present an overview of single supply assist techniques.

3.6.1 Supply Collapse Write Assist

As mentioned in the previous sections, lowering the bitcell supply voltage improves write margin, since the PMOS pull-up is weaker. Several implementations have been proposed to reduce or collapse the bitcell supply voltage in write operation. These techniques rely on lowering the voltage of the bitcell supply during write operation [45, 53–56]

One way to implement supply collapse is to float the bitcell supply power [53]. The V_{dd} lines are separated per column ($Vddm[n]$) as shown in Fig. 3.26. $Vddm[n]$ is controlled using the power switch $MSW[n]$, which is turned off in write operation, so that $Vddm[n]$ floats. The bitcell write current (from $Vddm$ to the bitline pulled low) discharges the $Vddm$ voltage, which improves the write margin [53]. The reduction of $Vddm$ voltage may cause the bitcells on the same column to lose data due to retention failures. Therefore, careful design of this technique is needed to prevent bitcells from failing due to retention. In addition to floating bitcell supply, actively pulling down the supply has been proposed [55, 57, 58] as well as using charge sharing to reduce the supply voltage [51].

Recently, the concept of collapsing the supply voltage had been extended to the bitcell supply voltage below the static data retention voltage [55, 59]. To prevent the resulting data retention failure of unselected cells, the duration of voltage collapse must be controlled, requiring a dynamic data retention approach. The maximum time within which the data is retained in the bitcell after fully collapsing the supply voltage Td, max has been measured as shown in Fig. 3.27. The measured data shows that bitcells can retain data even if the supply voltage is fully collapsed as long as the collapse time is less 3 ns and 10 ns for the SP and LP cases, respectively.

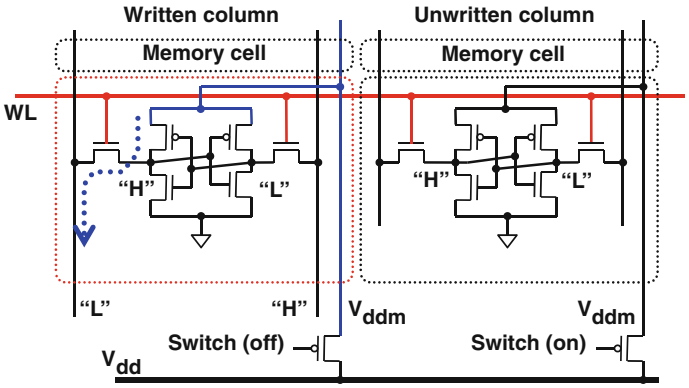

Fig. 3.26 Single supply write assist by floating the bitcell supply in write operation [53]

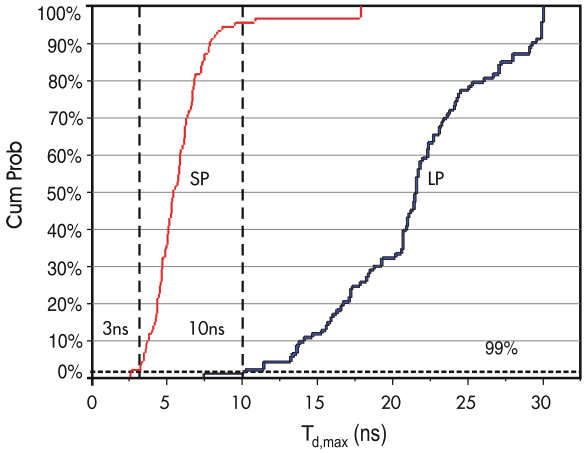

Fig. 3.27 Distribution of the maximum time a bitcell can sustain a collapsed voltage without flipping ($T_{d,\max}$) in 32 nm. SP and LP are two process flavors; LP being the low power option. Measured $T_{d,\max}$ show that bitcells can sustain 3 ns and 10 ns of supply collapse at the 99th percentile for SP and LP options, respectively [59]

To take advantage of dynamic data retention, transient voltage collapse write assist (TVC) has been proposed [55]. In this approach, the write assist lowers the bitcell supply below the retention voltage during write operation, hence, eliminating the contention between the pass-gate and the pull-up. Figure 3.28 shows the implementation of the TVC technique. In write operation, the header device ($PVCS$) is disabled and the cell supply (VCS) is actively discharged via the NMOS pull-down (NWR). The slew rate and minimum voltage of VCS can be controlled by varying the pulse width $COLPULSE$ and the parallel PMOS clamp devices PB. This technique requires careful control of timing and the lowest VCS voltage to ensure that the collapse time is sufficient for write without causing retention failures on unselected cells [55].

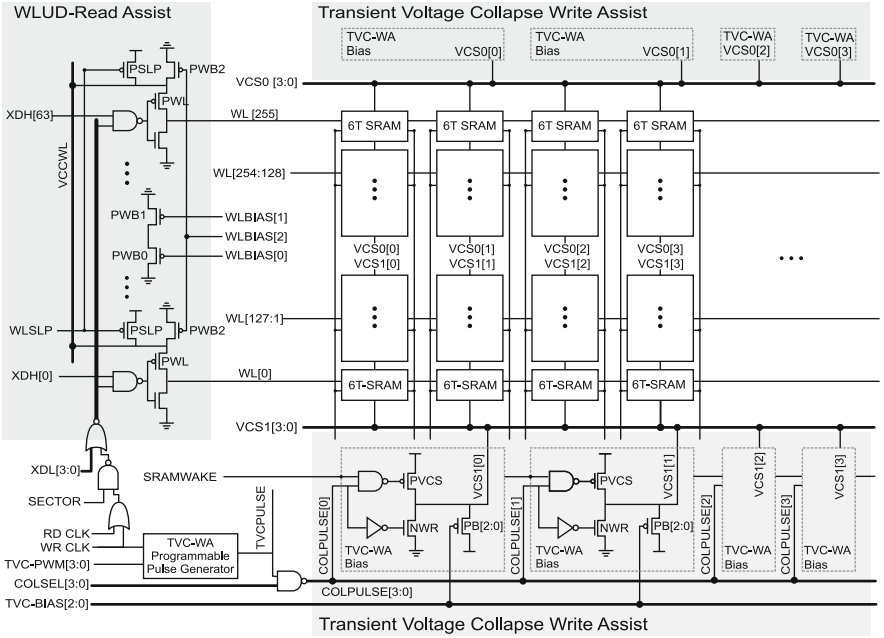

Fig. 3.28 Transient voltage collapse write assist and wordline under drive read assist [55]

Other supply collapse techniques include charge sharing to reduce the supply voltage. Also self-induced supply collapse was proposed to improve the write margin for 8T bitcell [54]. The scheme pre-discharges the bitlines prior to write operation: the virtual bitcell supply slightly droops before the start of the write operation. A similar idea to lowering the supply is to boost the virtual ground [2], but it requires larger series devices in the Vss path due to higher current demands [55].

3.6.2 Negative Bitline Write Assist

Another write assist technique relies on improving the strength of the NMOS pass-gate by applying a small negative voltage [60–64]. Using a small negative bitline voltage (∼200 mV) the gate-source bias of the pass-gate increases, which increases the write margin, as shown Fig. 3.29. The negative voltage is typically generated using a boosting capacitor connected to the bitlines and enabled at the end of the conventional write operation (after the bitline is pulled down to ground). Large improvements in write margin, in the range of 2–3 orders of magnitude reduction in failure probability can be achieved using this technique as shown in Fig. 3.29.

Figure 3.30 shows one implementation of the negative bitline technique [63]. In this approach, after the bitline is pulled low, the bitline is floated by disabling

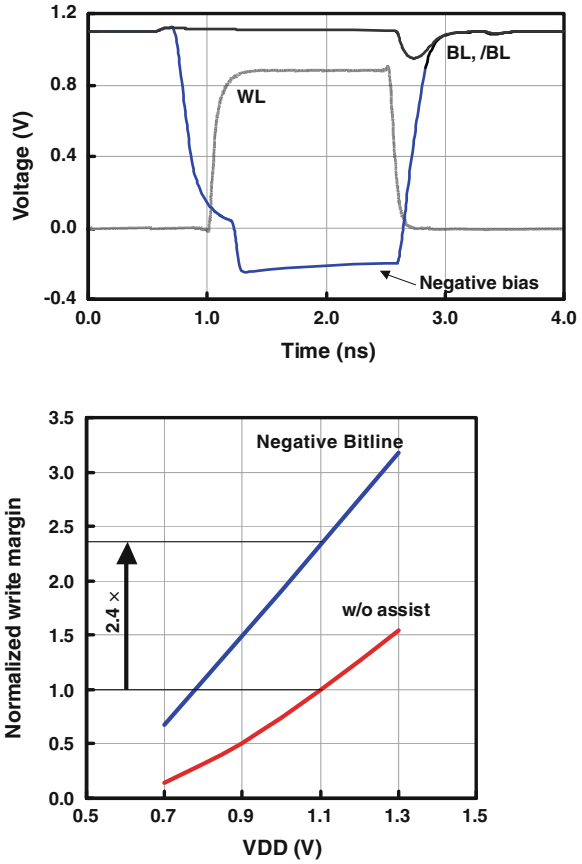

Fig. 3.29 Concept of negative bitline to improve write margin [60]

the column select devices. By using a high to low transition on the other side of the coupling capacitor C_{boost}, the bitlines are temporarily transitioned to negative bitline. The ratio of C_{boost} to total bitline capacitance determines the coupling coefficient and the magnitude of the negative voltage. The value of the negative bias should not be excessive to prevent flipping bitcells on the same column. This may occur because the pass-gate of non-selected bitcells may turn on if the bitline negative voltage exceeds the threshold voltage of the pass-gate, causing a retention failure. To allow memory configurations that have different number for cells per bitline, improvements in this technique use automatic BL negative level adjustment [61]. These improvements limit the value of the negative voltage, allowing this technique to be suitable for compilable memories.

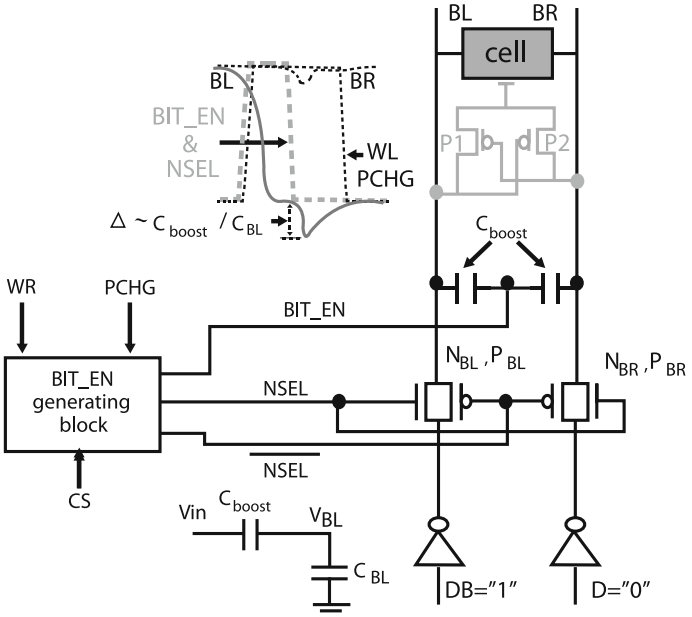

Fig. 3.30 Circuit implementation of negative bitline write assist [63]

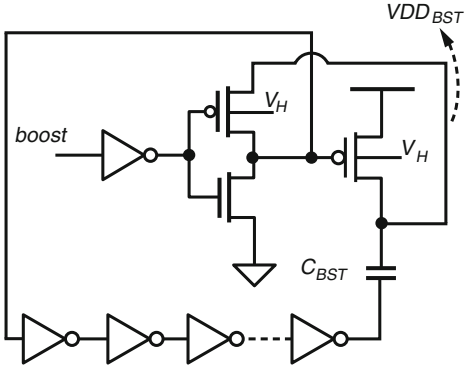

Fig. 3.31 Write assist using WL boost to improve write ability in 8T bitcells [26]

3.6.3 Wordline Boosting Write Assist

Wordline boosting technique is another approach to improve bitcell write ability. This technique is typically used with 8T bitcells that do not support column interleaving, and therefore, do not suffer from the half-select problem [26]. Boosting WL effectively reduces the write $V_{\min}$ for the 8T bitcells. One way to boost the write

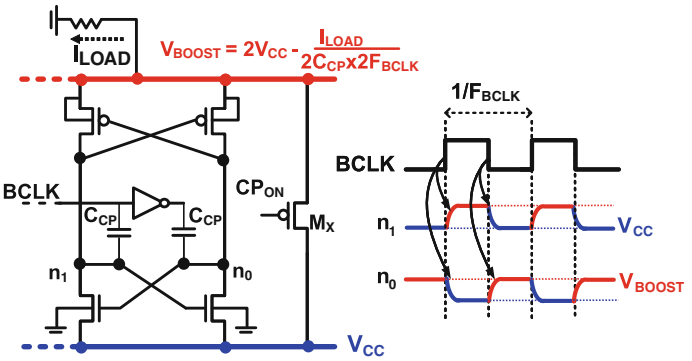

Fig. 3.32 Wordline boosting write assist technique employing a charge pump to boost wordline voltage in write operation [52]

wordline WWL is to use a boosting capacitor, as shown in Fig. 3.31. Boosted voltage ($V_{DD,BST}$) powers the WL drivers, the write drivers, and the column select devices during write operation [26].

Another approach to boost the WL is to use a charge pump [52] as shown in Fig. 3.31. The pump boosting ratio is determined by the load current, which comprises all the active and inactive level-shifters connected to the WL. In this implementation, if the boosted voltage exceeds the V_{max} of the device, the charge pump is turned off to maintain the gate oxide and junction reliability of the devices connected to the V_{boost} rail. Since the boosting ratio depends on the pump frequency, it is important to track the process and temperature condition to provide the optimal boost. On-chip write ability sensors that dynamically track the temperature and aging has been used to provide the optimal boosting ratio [52]. Boosting techniques using charge pumps consume large power consumption due to the dynamic power of the charge pump and the need to increase the pump frequency (F_{BLK}) to improve the boosting ratio [52] (Fig. 3.32).

Recently, a novel WL boosting approach has been proposed which uses the intrinsic coupling capacitance from the device and interconnect to the WWL to create a boosting capacitor, as shown in Fig. 3.33. The technique uses two types of intrinsic coupling capacitance. The first coupling capacitance is found in the WWL interface of the WWL driver ($C1$), while the second coupling capacitance is at the WWL interface to the bitcell NMOS write pass-gate ($C2$ and $C3$). To enable the use of the coupling capacitance in the WWL driver, the input of the driver is transitioned to low to create the rising transition on the WWL. After the WWL reaches full V_{DD}, the WWL is floated, and the coupling capacitance is enabled by switching the $BOOST$ signal from low to high. Similarly, the second coupling is enabled by pre-discharging the WBL and $WBLX$ signals, and depending on the data polarity, one bitline is brought high after the WWL has been floated. This technique can achieve a total boost of 20 % of the supply voltage, with the pass-gate coupling contributing 17 % while the WWL driver contributes 3 %. This boosting technique is scalable

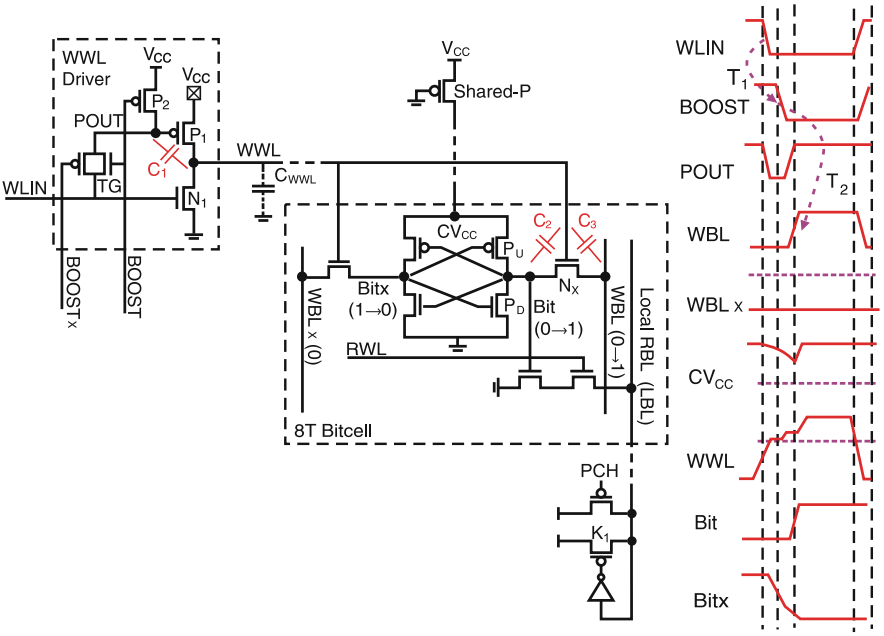

Fig. 3.33 Write assist employing wordline boosting using intrinsic capacitive coupling from (1) the bitlines to the WL and (2) WL driver input to the WL output [54]

to any number of bits per WWL since the intrinsic bitline capacitance scales in the same ratio.

3.6.4 Wordline Under-Drive Read Assist

Lowering the wordline voltage in read operation decreases the pass-gate strength and improves the read stability. This technique is one of the most commonly used read assist techniques [22, 45, 51, 58, 65, 66]. Figure 3.28 shows an implementation of WLUD, where the WL voltage is adjusted using a voltage divider between the PMOS PWL device and the programmable pull-down $PWB1, 2, 3$.

The wordline under drive technique is enabled in both read and write operation so that half-selected bitcells experience lower WL voltage. However, lowering the WL voltage degrades the write margin, so a write assist technique is usually implemented with the WLUD approach. In addition, the WL level should be controlled depending on the process and temperature condition, so that the lowest WL voltage occurs in the worst global corner for read stability (fast NMOS and slow PMOS), while the WL voltage stays high in other corners which are write-limited. To address this requirement, adaptive control of the wordline level depending on the process corner

has been proposed [22, 45]. In [22], WL voltage varies depending on D2D and temperature variations. Replica transistors are used to control the WL voltage level, which improves process tracking and increases SNM. A bitcell-based sensor has been proposed to dynamically optimize the level of the WL for each die [65]. The on-die sensor determines if the die is read or write-limited, and a programmable control applies the optimal WLUD. Since read and write stability strongly depend on the temperature, a die can shift from being write-limited at low temperature to read-limited at higher temperature. The sensor dynamically tracks the temperature and process variation and applies the optimal WLUD level [65].

3.6.5 Lower Bitline Read Assist

As discussed in Sect. 3.5, lowering the bitline voltage level improves read stability since the pass-gate strength decreases [2, 11, 29, 64, 67]. In one such approach, the bitline precharge voltage is reduced before the bitcell is accessed, increasing SNM. The bitline voltage can be reduced by slightly discharging the bitlines before the start of read operation [29]. In another implementation, the bitlines are precharged using an NMOS transistor, so the precharge voltage is one V_{th} lower than V_{DD}. Another technique precharges different columns to V_{DD} and GND and uses charge sharing to reduce the bitline voltage [67]. More details about this approach will be discussed in the case study presented in Sect. 3.7.

3.6.6 Short Bitline Read Assist

Dynamic read stability analysis revealed that reducing the bitline capacitance reduces read disturb failures. Lowering the number of cells per bitline from 256 to 64 cells per bitline reduces failure[2] rate by 50X, as shown in Fig. 3.34. Reducing the number of bitcells per columns is therefore an effective approach to improve read stability and has been explored to reduce SRAM V_{min} [17, 26].

3.6.7 Read and Write Back Assist

Several read assist techniques that rely on improving the dynamic stability of the bitcell have been proposed. The duration of the WL pulse width has a strong impact on read and write stability. If the WL pulse width is very short, the dynamic read

[2] The DC failure rate is 500X higher than the rate compared to dynamic failure rate for the 64 cells per bitline case. Also, DC failure is 140 mV higher than dynamic failure, which shows the importance of dynamic read stability.

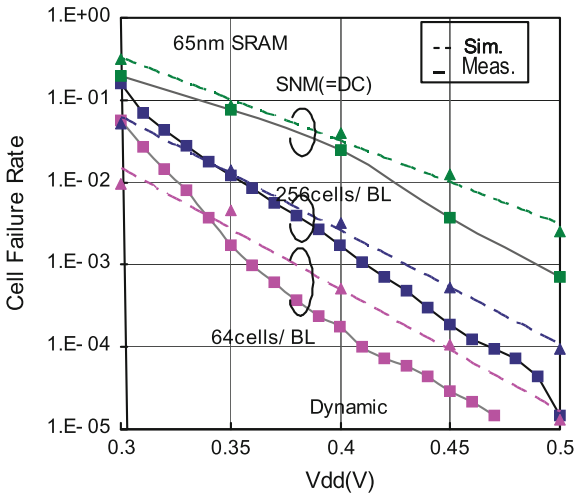

Fig. 3.34 Measured static and dynamic read stability versus supply voltage. Reducing the number of bitcells per columns reduces the probability of read disturb [24]

stability improves since the bitcell is not given enough time to flip. However, the write margin worsens for the same reason [29]. Therefore, an optimal WL pulse width balances read and write stability. To decouple the opposite requirements of WL pulse width for read and write operations, the read and write back approach has been proposed (also known as read modify write) [29, 49]. In this technique, each column has a dedicated sense amplifier (column multiplexer=1). As the name implies, the write operation starts with a read operation followed by a write operation. In the read operation, a narrow WL pulse width is used, to increase read stability. The read data is stored on a per column sense amplifier, which acts as the write driver. Next, the WL is enabled again for a longer pulse width that is sufficient for write operation. The same concept can be applied in read operation only, where the sense amplifier is used to read the data from the bitcell and write it back at the end of the read operation [49]. Hence, it provides data recovery by writing back the original data. Read and write back technique increases power consumption since every column undergoes full signal amplification. Moreover, it has a large area overhead (8 % area increase) since a SA is integrated per column and cannot be shared between several columns as typically is the case with high density memories [29]. In addition, mismatch in the sense amplifier may cause the bitcell to write incorrect data, corrupting the stored one.

3.7 Case Study: Selective Precharge Read Assist Technique

In SRAM design, the bitlines are typically precharged to V_{DD} before accessing the bitcell. However, as discussed earlier, reducing the bitline voltage V_{BL} before accessing a bitcell improves read stability, as shown in Fig. 3.35. This increase in

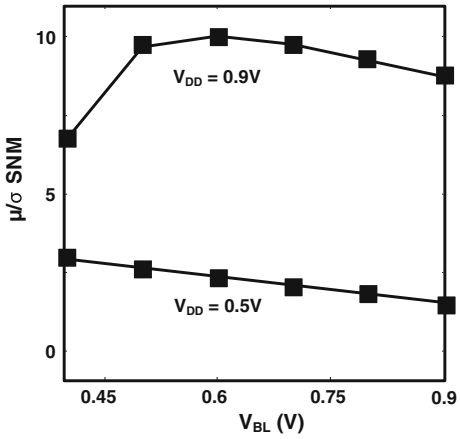

Fig. 3.35 $\frac{\mu}{\sigma}$ SNM versus bitline precharge voltage. Lower bitline voltage during a read access improves bitcell read stability (SNM) [11]

stability results from the decrease in pass-gate (access transistor) strength reduces as the bitline voltage decreases. This effectively increases the bitcell α ratio, defined as the ratio of pull-down to pass-gate strength. Hence, SNM improves as ΔV_{BL} increases where ΔV_{BL} is defined here as $V_{\mathrm{DD}} - V_{\mathrm{BL}}$ just before the wordline (WL) is asserted. Note that as ΔV_{BL} increases,[3] SNM reaches a maximum, and further increases in ΔV_{BL} cause significant SNM reduction. Therefore, accurate control of ΔV_{BL} is important to prevent ΔV_{BL} from exceeding the maximum SNM point.

From a circuit point of view, the relation between ΔV_{BL} and SNM has been exploited into increase bitcell stability [2, 29] . A pulsed bitline approach has been used to control the duration of an NMOS pull-down [29]. This pull-down device discharges the bitline, which increases ΔV_{BL} just before the WL is enabled. However, this technique is sensitive to PVT variations since ΔV_{BL} is a strong function of the pulse duration, which will vary with PVT variations. Therefore, a complex timing scheme may be required to control ΔV_{BL} at different PVT conditions. Alternatively, an NMOS device has been used to precharge the bitlines, by having one V_{th} drop [2]. Due to the strong sensitivity of V_{th} to PVT variations, the effectiveness of this technique decreases in different PVT corners. Moreover, low V_{th} devices are required to ensure that ΔV_{BL} does not cause read disturbs in the worst-case conditions, which adds additional processing cost, especially for low cost SoCs.

In this section, we examine implementation details of a read assist technique that uses lower BL voltage to improve read stability [67]. This technique uses a single supply voltage approach to overcome the obstacles in implementing dual supply read assist techniques. Instead of precharging the bitlines to V_{DD}, different parts of the bitlines are precharged to V_{DD} or (predischarged) to GND. Using charge sharing, the final required bitline voltage can be precisely controlled using the capacitance ratio. Therefore, this technique is highly immune against process variations (both

[3] ΔV_{BL} in this chapter should not be confused with the bitline differential voltage ΔV_{bl}. Here, ΔV_{BL} is the reduction in bitline precharge level before accessing the bitcell.

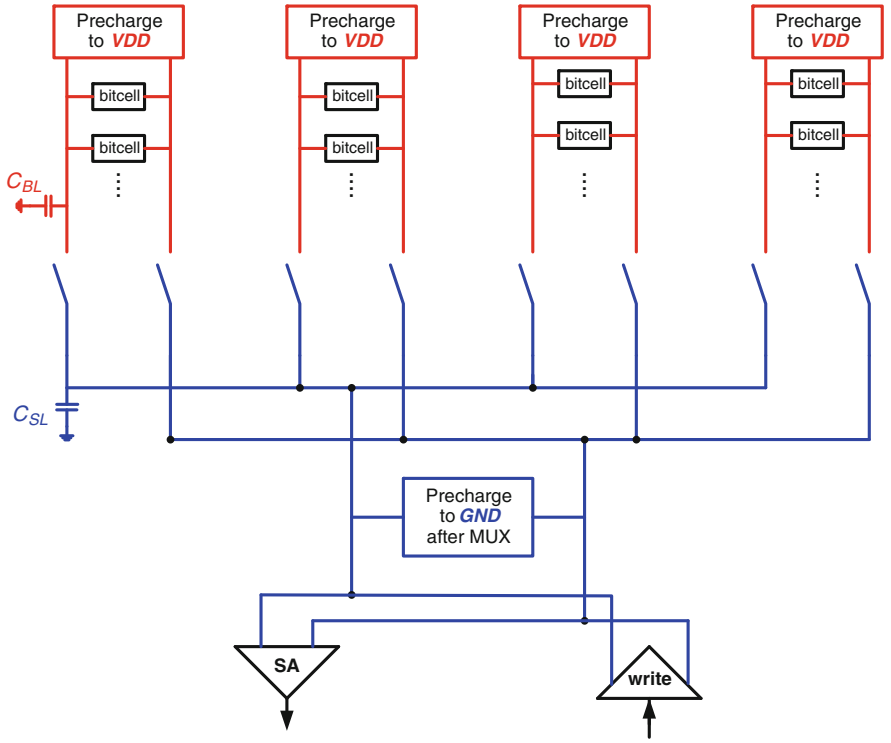

Fig. 3.36 Selective Precharge operation. Step 1: Precharge to V_{DD} and GND

front-end and back-end) since the capacitance ratio is weakly dependent on PVT corners.

3.7.1 Circuit Operation

Figure 3.36 shows a simple schematic for the selective precharge technique with four bitline columns connected to the read and write circuitry (sense amplifier and write drivers). Bitlines, BL/BLB, refer to the upper part of the bitlines connected directly to the bitcells (before the column select). Sense/Write lines, SL/SLB, refer to the lower part of the bitline connected to the sense amplifier and write drivers (after the column select).

Selective precharge operation can be divided into three main steps. First, BL and BLB are precharged to V_{DD} as in conventional approaches, while SL and SLB are pre-discharged to GND, as shown in Fig. 3.36. In the second step, the column select devices (MUX) on each bitline column are enabled as shown in Fig. 3.37. Hence, charge sharing occurs between the upper and lower bitlines for $BL0 - BL3$. The

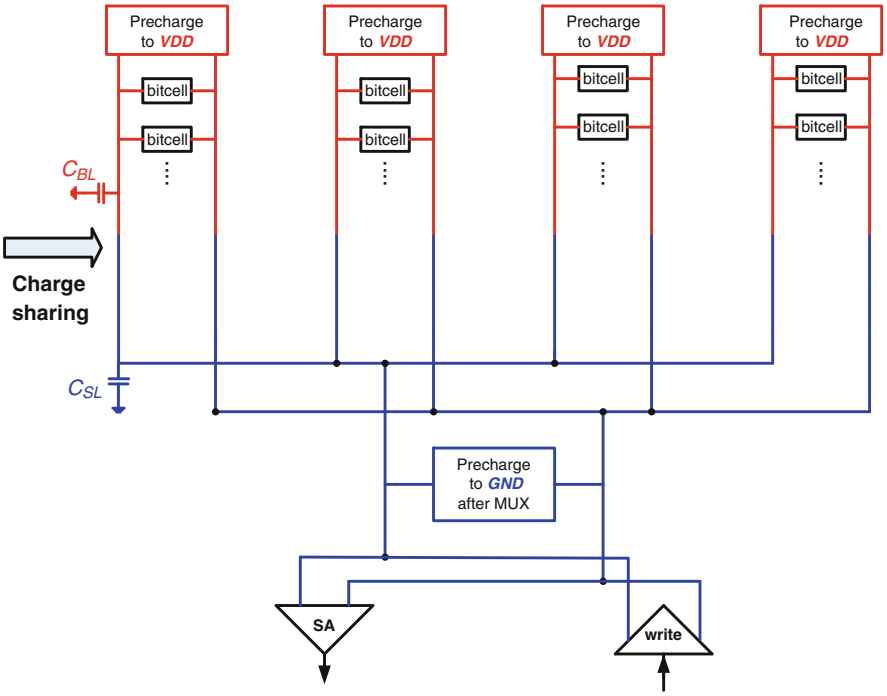

Fig. 3.37 Selective Precharge operation. Step 2: Charge sharing

final bitline voltage after charge sharing is determined by the capacitance ratio of upper to lower bitlines (BL and SL). Charge sharing reduces the bitline voltage so, SNM improves as discussed earlier. In the third step, the MUX devices for all unselected columns are disabled, while the selected column MUX stays on as shown in Fig. 3.38. In this case, the selected column provides access to the required bitcell, while SNM of half-selected bitcells is also improved since their bitline voltages have also been reduced.

Figure 3.39 shows the implementation of the selective precharge technique. A NOR gate is added for each bitline column to control the column select. Figure 3.40 shows the precharge circuits for both V_{DD} and GND. Figure 3.41 shows the timing diagram for selective precharge operation. ch_sh is activated using the rising edge of the precharge disable (for PMOS pull-up). When ch_sh is high, the PMOS devices in the column select MUX are on. Hence, charge sharing occurs between all bitlines sharing the same read/write circuitry and SL/SLB line. Therefore, BL/BLB voltage decreases while SL/SLB voltage increases. ch_sh is disabled using mux_state, which is a dummy column select signal. Therefore, bitcells see reduced bitline voltage when the bitcell is accessed. At the end of operation, BL/BLB are precharged back to V_{DD} while SL/SLB are precharged to GND, as shown in Fig. 3.41.

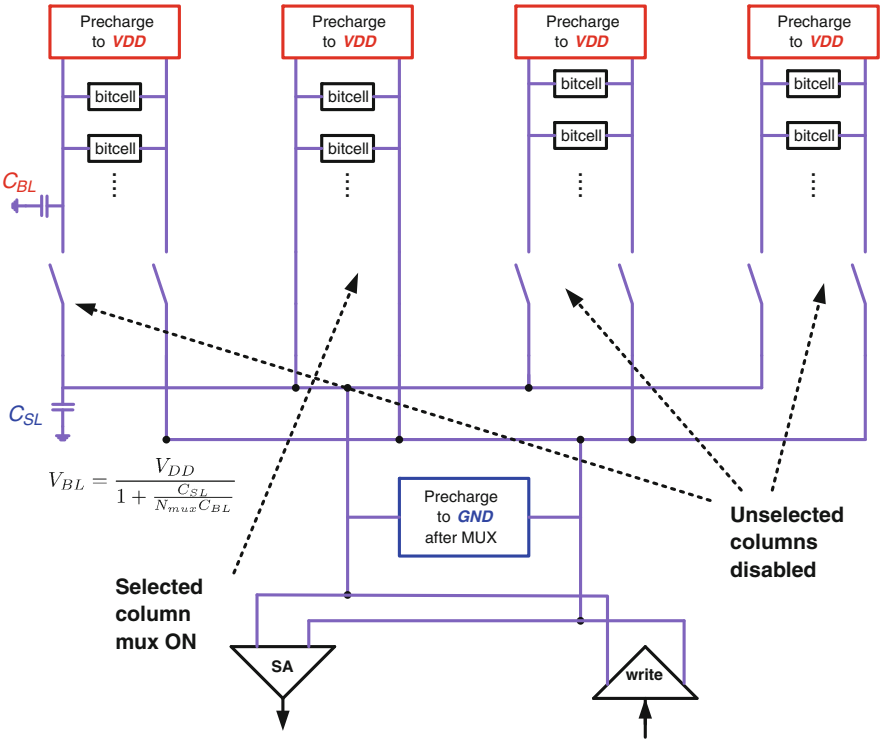

Fig. 3.38 Selective Precharge operation. Step 3: Unselected columns disabled

Selecting the location of precharge to V_{DD} or GND sets the required value of ΔV_{BL}. For example, if a larger ΔV_{BL} is required, one or more of the bitlines can be precharged to GND instead of V_{DD}, as shown in Fig. 3.42. Therefore, the proposed technique allows changing ΔV_{BL} by selecting in which points should be precharged to V_{DD} or GND. Note that in this technique, no additional supply voltages are required to generate the desired bitline voltage, which reduces the design complexity. In addition, since the final ΔV_{BL} voltage depends solely on capacitance ratio, its value is not influenced by process variations.

3.7.2 Access Time Improvement

In SRAM, the read operation determines the access time of the memory. One of the limiting factors determining the access time is the delay from the clock to the WL enable, which is part of the memory critical path. The proposed technique introduces another signal, ch_sh, which should be enabled before WL is asserted. To accommodate the ch_sh signal shown in Fig. 3.41, the WL enable path may

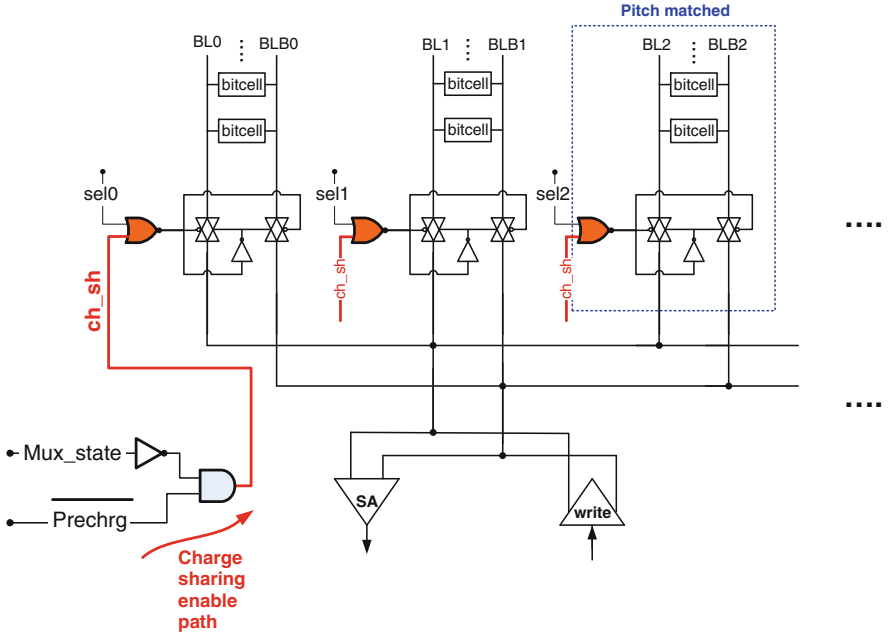

Fig. 3.39 Selective precharge schematic

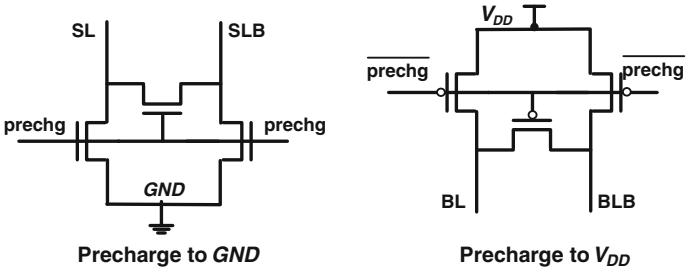

Fig. 3.40 Precharge to V_{DD} and GND circuits, including equalize transistors

be delayed. This delay will therefore increase the memory access time. Hence, a technique to reduce (or recover) access time is required.

In addition to the clock to WL delay, another contributor to the memory's access time is the WL pulse width T_{WL}. T_{WL} is the time required for the bitcell to discharge the bitlines and generate sufficient input differential for the sense amplifier to allow correct read operation. T_{WL} typically contributes to approximately 30 % of the memory access time [44]. To reduce this delay component, we exploit the relation between ΔV_{BL} and the sense amplifier (SA) input offset.

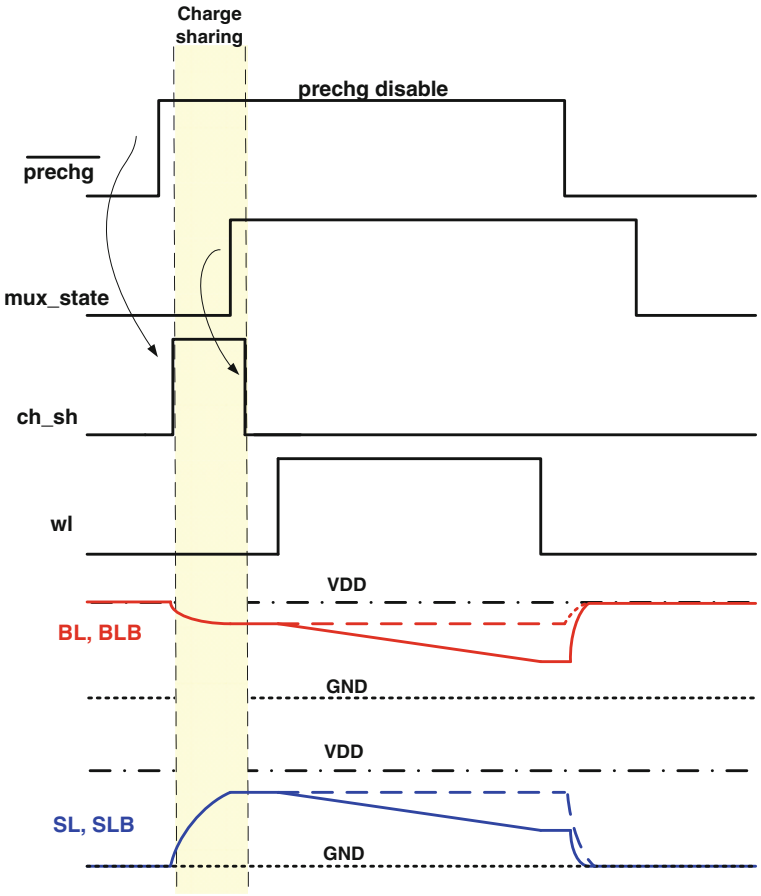

Fig. 3.41 Selective precharge timing diagram

There are many types of sense amplifiers used in SRAM design. However, current latch sense amplifier (CLSA) is one of the most widely used types due to its high speed and isolation as discussed in Sect. 5.4.2 (CLSA shown in Fig. 5.11). Moreover, it has been shown that reducing bitline voltage (common mode) improves the SA offset [68]. This characteristic of CLSA makes it very attractive in the proposed selective precharge technique. By reducing the bitline voltage (increasing ΔV_{BL}), the SA offset ($\sigma_{\mathrm{SA,offset}}$) decreases, allowing T_{WL} to be reduced for a given failure probability. The reduction in T_{WL} can therefore compensate for the increase in clock to WL delay, as will be shown in Sect. 3.8.

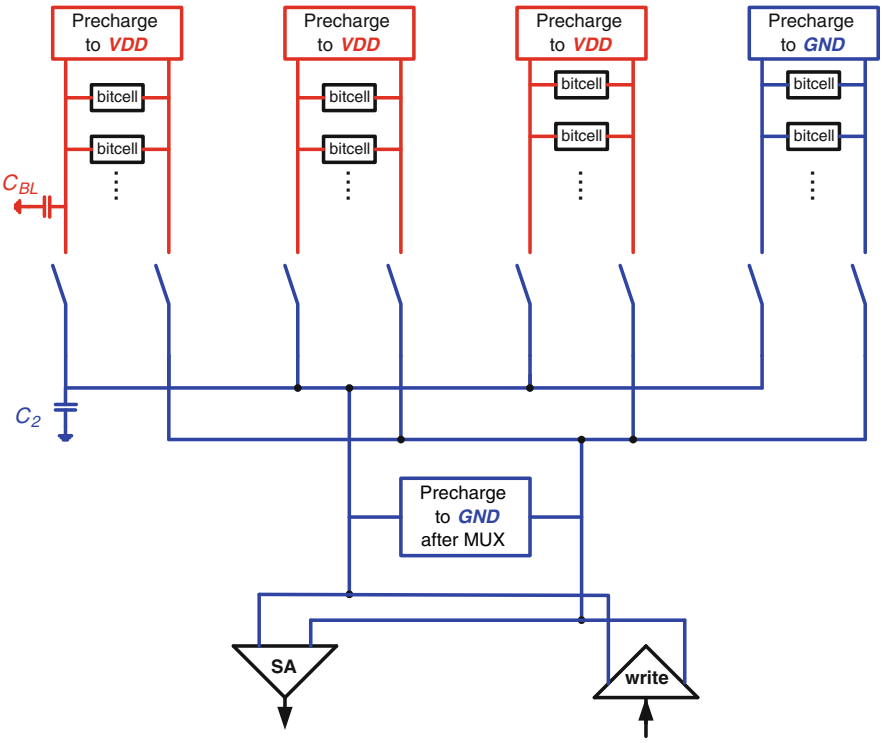

Fig. 3.42 Achieving larger range of ΔV_{BL} by precharging one of the bitlines to GND

3.8 Results and Discussion

To test the proposed read assist technique, a full-custom 512 kb SRAM was designed and implemented in an industrial 45 nm technology as shown in Fig 3.43. Table 3.2 provides details on the memory architecture. In this section, we present the post-layout simulation results for the proposed technique.

In Fig. 3.44, the read operation is shown for a bitcell on the first column (enabled using MUX0). In the beginning of the operation, $BL0$ and $BLB0$ are set to V_{DD} while SL and SLB are set to zero. Charge sharing operation is activated using ch_sh, which activates all the MUX transistors. Therefore, the $BL0/BLB0$ voltages decrease while SL/SLB increase as shown in Fig. 3.44, and they settle to a value determined by the capacitance ratio. Note that charge sharing happens quickly and that the results voltage is not sensitive to the ch_sh pulse width (a wider pulse does not affect the settling voltage after charge sharing). After charge sharing is completed, the MUX devices (PMOS) for all unselected columns are disabled (MUX1), while the selected column stays selected (MUX0). Therefore, when WL is asserted, the accessed and half-selected bitcells see a reduced bitline voltage, which increases the bitcell's SNM.

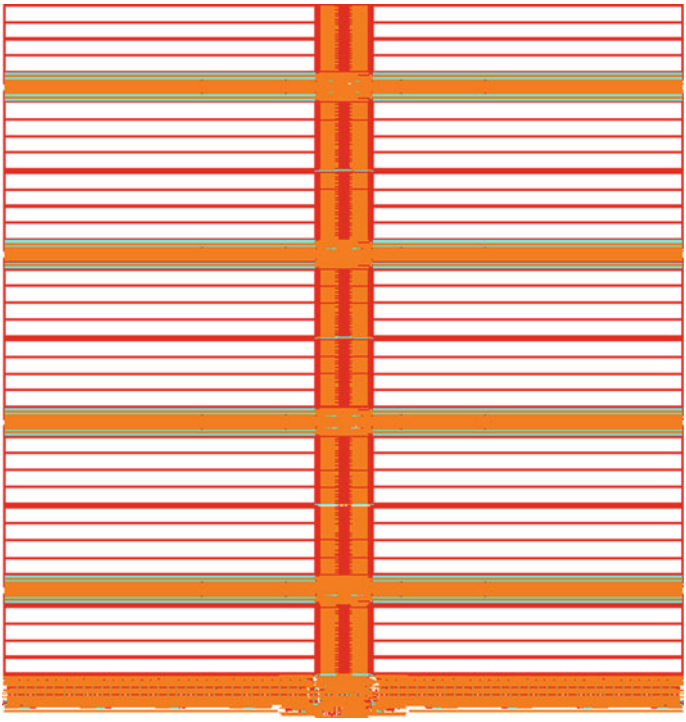

Fig. 3.43 Layout of the designed 512 kb memory in 45 nm technology

At the end of the read operation, the bitlines are precharged to V_{DD} while the sense lines are precharged to GND.

The impact of bitline voltage on the CLSA speed and input offset is shown in Fig. 3.45. Monte Carlo transient simulations were used to measure the SA's offset. As ΔV_{BL} increases, the SA delay slightly decreases until it reaches a minimum. Beyond that point, the SA delay increases. In the meantime, the SA input offset (σ_{offset}) decreases monotonically with the increase in ΔV_{BL}. The reduction in σ_{offset} improves the robustness of the SA and decreases the probability of read access failures. Therefore, the WL pulse width can be reduced accordingly based on the following:

$$\frac{T_{WL2}}{T_{WL1}} = \frac{\sigma_{SA,offset_2}}{\sigma_{SA,offset_1}}, \tag{3.1}$$

where T_{WL} is the time allowed for the bitcell to generate the bitline differential before enabling the SA. This large reduction in σ_{offset} reduces the access time of the memory. Because T_{WL} is about 30 % of memory access time, and as shown in Fig. 3.45, SA offset can be reduced by up to 25 %, and access time improves by 7 %. In reality, to accommodate the ch_sh pulse, the WL enabled path may be slightly delayed, so this

Table 3.2 512 kb memory design information

Technology	45 nm low power (LP) CMOS
Density	512 kb
Memory width (word size)	64 bits
Memory depth (number of words)	8,192 words
Banks	16 (32 kb each)
Rows/bank	128
Columns/bank	256

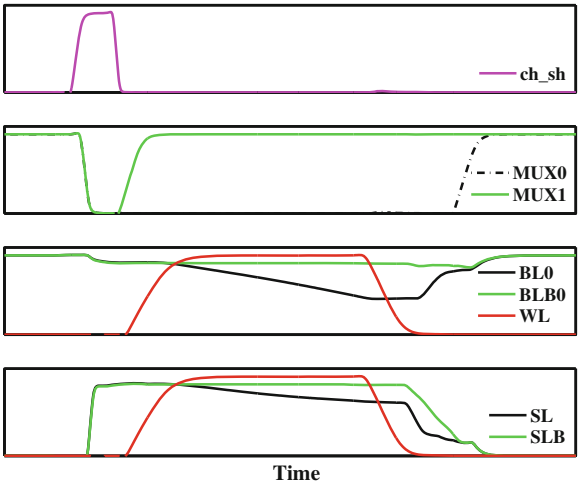

Fig. 3.44 Simulation waveforms for selective precharge read operation. MUX0/1 is the gate voltage for the PMOS device in the column select for column 0 and 1, respectively

improvement in speed would be smaller. Nevertheless, since charge sharing requires a very short time, the impact on access time improvement is negligible.

Charge sharing operation enabled when a bitcell is accessed for write operation, as shown in Fig. 3.46. In that case, half-selected bitcells experience reduced BL voltage to improve read stability (bitlines $BL1$ and $BLB1$). However, write margin degrades due to the lower drive capability of the pass-gate which may cause a write failure [39]. To improve the write ability of the selected bitcell, we use a CMOS write driver, as shown in Fig. 3.47. Therefore, the BL voltage lost in charge sharing is recovered using the pull-up device in the write driver. Hence, the CMOS write driver protects the write margin in write operation.

To estimate the bitcell read stability, we simulated the butterfly curve as shown in Fig. 3.48, which shows the bitcell SNM for a nominal bitcell which does not include WID variations. However, as discussed in Chap. 2, WID variations will cause

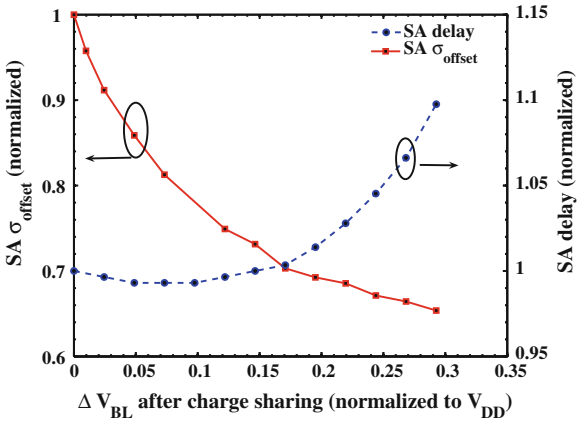

Fig. 3.45 Sense amplifier delay and input offset versus ΔV_{BL} after charge sharing

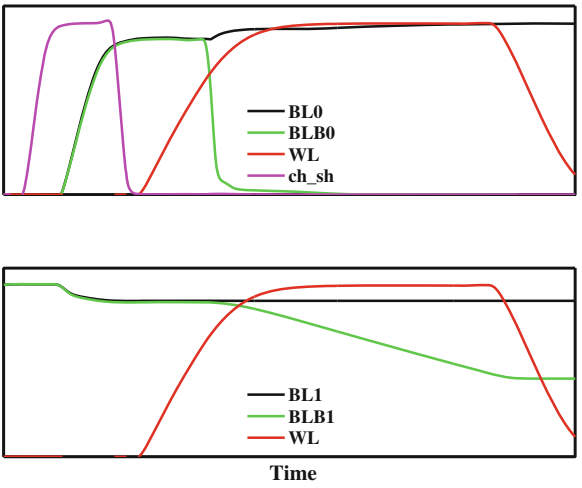

Fig. 3.46 Simulation waveforms for selective precharge write operation. $BL0/BLB0$ are accessed for write operation while $BL1/BLB1$ are half-selected bitlines

each transistor in the bitcell to have different V_{th}, which will cause the bitcell be asymmetric. This is shown by the Monte Carlo simulation results in Fig. 3.49, which shows a large spread in the VTC characteristics of the bitcell. This spread translates to large variation in SNM.

The improvement in SNM using the proposed technique is shown in Fig. 3.50. Monte Carlo simulations are used to measure the impact of local variation on the bitcell's SNM. To ensure high yield for the embedded memories, 6σ of SNM local variation is included. As ΔV_{BL} increases, SNM increases linearly until it reaches

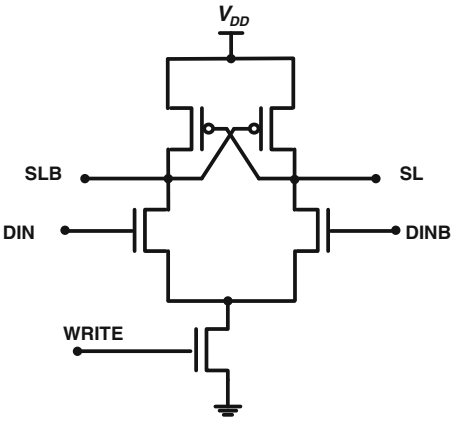

Fig. 3.47 Full-swing CMOS write driver used to improve write margin

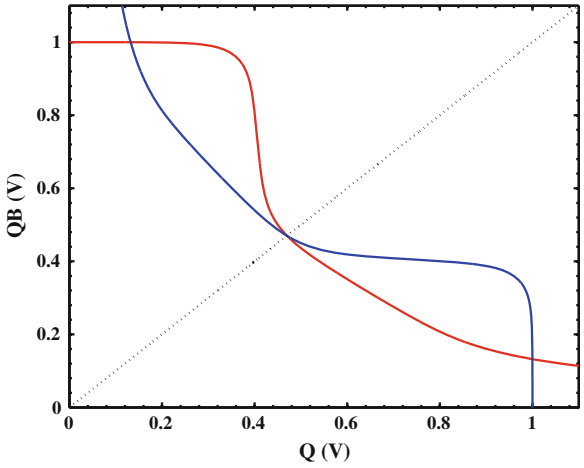

Fig. 3.48 SNM simulation results for nominal devices without WID variations

a maximum. Any further increase in ΔV_{BL} causes SNM to decrease significantly, which deteriorates cell read stability, agreeing with previous results [29, 39].

To evaluate the robustness of the proposed scheme in precisely controlling ΔV_{BL}, different process corners and post-layout RC extraction options were simulated, as shown in Table 3.3. ΔV_{BL} does not change significantly across different conditions (9–12 %), demonstrating the robustness of the proposed technique against PVT variations.

c

Figure 3.51 shows the process window curves (V_{th}), which are used to determine the operating limit of the memory accounting for 6σ of local variation coverage. In

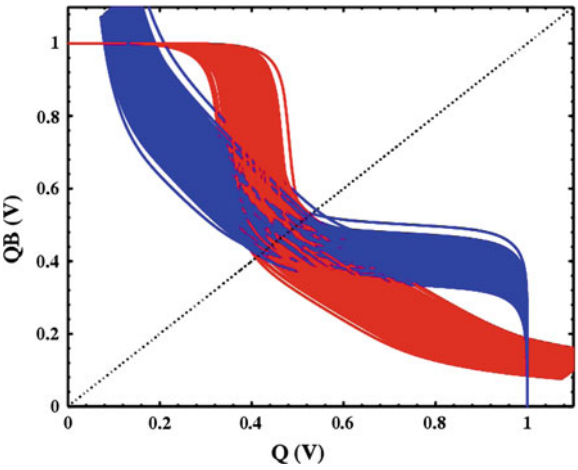

Fig. 3.49 SNM simulation results using Monte Carlo simulation for 1,000 MC runs

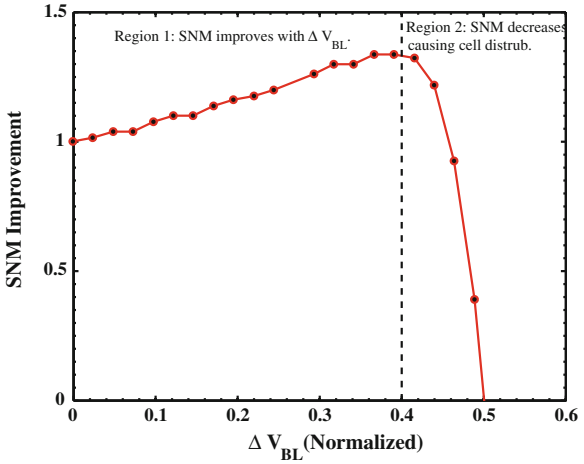

Fig. 3.50 SNM improvement versus ΔV_{BL} after charge sharing using the proposed technique

this simulation, the D2D variations are swept for NMOS and PMOS V_{th}. For each point of D2D variation, Monte Carlo simulation using WID variations is used to find the mean and sigma for SNM. We define the failure region as that where SNM reaches zero. Using the selective precharge technique (solid line), the operating window is expanded relative to the conventional approach. This increase in operation window reduces the failure probability by more than 100X.

To validate the improvements in cell stability using the proposed selective precharge technique, the designed 512 kb memory was fabricated in 45 nm tech-

Table 3.3 ΔV_{BL} for different conditions

Process	Slow	Slow	Nominal	Fast
Temp.	–40	125	25	–40
Parasitic C	Max	Max	Nominal	Min
ΔV_{BLA}	9.8 %	9.4 %	11.1 %	12 %

[A] Normalized to V_{DD}.

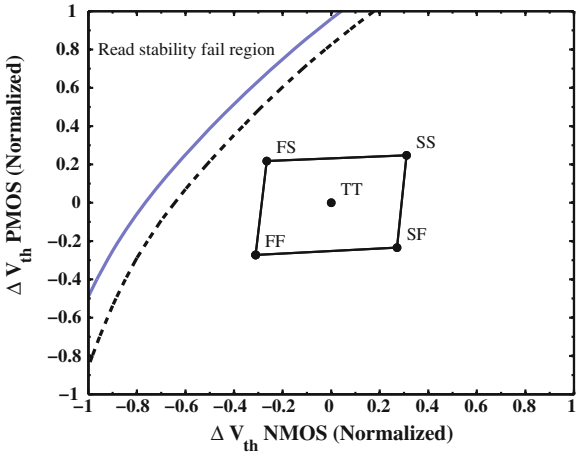

Fig. 3.51 Read stability operating window for selective precharge (*solid line*) compared to the conventional approach (*dotted line*). Simulation accounts for 6σ of local variations

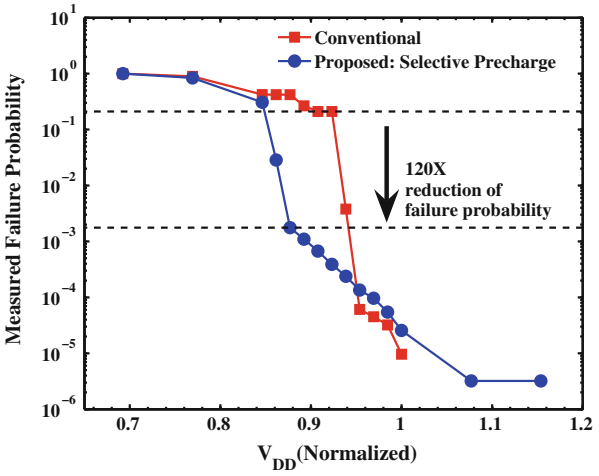

Fig. 3.52 Measured failure probability for the fabricated 512 kb memory for the proposed technique (selective precharge) and the conventional approach

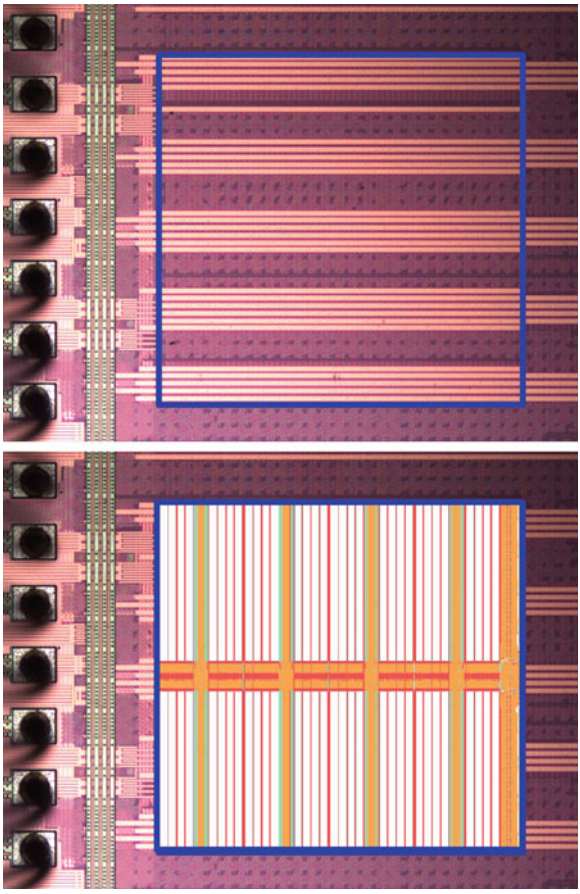

Fig. 3.53 Chip micrograph for the fabricated 512 kb memory in 45 nm technology. Upper figure shows the location of the memory and the lower overlays the memory layout

nology. Figure 3.53 shows the fabricated test chip micrograph. Measured results for cell failure probability using the conventional and the proposed technique are shown in Fig. 3.52, which shows that the proposed technique reduces the failure probability by more than 120X, validating the improvement in bitcell read stability.

The proposed technique has a small area overhead (<2 %) and shows strong robustness against process variations. In addition, it requires only one supply voltage, eliminating the need of level-shifters that cause large area and speed penalty. Moreover, the timing generation is simple since it re-uses timing signals available in SRAM design. Finally, the technique improves the memory speed, bitcell stability, and operating window, demonstrating its effectiveness.

3.9 Summary

The increase in local variations in nanometer technologies strongly affects SRAM cell stability. Various metrics used to analyze write and read stability were discussed. Detailed overview of the state-of-the art assist techniques and their impact on conventional SRAM design approaches was presented. As a case study, the implementation details of a single supply read assist technique were discussed. The proposed technique, selective precharge, allows precharging different parts of the bitlines to V_{DD} and GND and uses charge sharing to precisely control the bitline voltage which increases the bitcell stability. In addition to improving SNM, the proposed technique also improves memory access time. A 512 kb memory was designed to demonstrate the effectiveness of this technique in an industrial 45 nm technology. The technique significantly improves read stability, and provides high robustness against process variations.

References

1. K. Itoh, M. Horiguchi, M. Yamaoka, Low-voltage limitations of memory-rich nano-scale CMOS LSIs, in *33rd European Solid State Circuits Conference, 2007. ESSCIRC*, pp. 68–75, 11–13 Sept. 2007
2. A. Bhavnagarwala, S. Kosonocky, Y. Chan, K. Stawiasz, U. Srinivasan, S. Kowalczyk, M. Ziegler, A sub-600 mv, fluctuation tolerant 65 nm CMOS SRAM array with dynamic cell biasing, in *Proceedings of IEEE Symposium on VLSI Circuits*, pp. 78–79, 2007
3. S. Mukhopadhyay, H. Mahmoodi, K. Roy, Statistical design and optimization of SRAM cell for yield enhancement, in *Proceedings of International conference on Computer Aided Design*, pp. 10–13, 2004
4. E. Grossar, M. Stucchi, K. Maex, W. Dehaene, Read stability and write-ability analysis of SRAM cells for nanometer technologies. IEEE J. Solid-State Circ. **41**(11), 2577–2588 (2006)
5. M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada, K. Yanagisawa, T. Kawahara, Low-power embedded SRAM modules with expanded margins for writing, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, vol. 1, pp. 480–611, 2005
6. K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura, H. Kobatake, Redefinition of write margin for next-generation SRAM and write-margin monitoring circuit, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, pp. 2602–2611, 2006
7. S. Mukhopadhyay, H. Mahmoodi, K. Roy, Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled CMOS. Comput. Aided Des. Integr. Circ. Syst. IEEE Trans. **24**(12) PP. 1859–1880, Dec. 2005
8. W. Dong, P. Li, G. Huang, SRAM dynamic stability: Theory, variability and analysis, in *IEEE/ACM International Conference on Computer-Aided Design, ICCAD 2008*, pp. 378–385, Nov. 2008
9. Y. Zhang, P. Li, G.M. Huang, Separatrices in high-dimensional state space: system-theoretical tangent computation and application to SRAM dynamic stability analysis, in *Proceedings of the 47th Design Automation Conference*, ser. DAC '10. (ACM, USA, 2010), pp. 567–572
10. M. Khellah, D. Khalil, D. Somasekhar, Y. Ismail, T. Karnik, V. De, Effect of power supply noise on SRAM dynamic stability, in *Proceedings of IEEE Symposium on VLSI Circuits*, pp. 76–77, June 2007

11. A. Bhavnagarwala, S. Kosonocky, C. Radens, K. Stawiasz, R. Mann, Q. Ye, K. Chin, Fluctuation limits and scaling opportunities for CMOS SRAM cells, in *Proceedings of the International Electron Devices Meeting (IEDM)*, pp. 659–662, 2005
12. J. Wang, S. Nalam, B. Calhoun, Analyzing static and dynamic write margin for nanometer srams, in *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 129–134, Aug 2008
13. K. Agarwal, S. Nassif, Statistical analysis of SRAM cell stability, in *DAC '06: Proceedings of the 43rd Annual Conference on Design Automation*, pp. 57–62, 2006
14. G. Huang, W. Dong, Y. Ho, P. Li, Tracing SRAM separatrix for dynamic noise margin analysis under device mismatch, in *Behavioral Modeling and Simulation Workshop, BMAS 2007. IEEE International*, pp. 6–10, Sept. 2007
15. S. Nalam, V. Chandra, R. Aitken, B. Calhoun, Dynamic write limited minimum operating voltage for nanoscale SRAMs, in *Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 1–6, March 2011
16. R. Joshi, R. Kanj, S. Nassif, D. Plass, Y. Chan, C.-T. Chuang, Statistical exploration of the dual supply voltage space of a 65nm PD/SOI CMOS SRAM cell, in *Solid-State Device Research Conference, 2006. ESSDERC 2006. Proceeding of the 36th European*, pp. 315–318, Sept. 2006
17. M. Yamaoka, K. Osada, T. Kawahara, A cell-activation-time controlled SRAM for low-voltage operation in DVFS SoCs using dynamic stability analysis, *ESSCIRC: 34th European Solid State Circuits Conference*, pp. 286–289, Sept. 2008
18. S.O. Toh, Z. Guo, B. Nikolić, Dynamic SRAM stability characterization in 45nm CMOS, in *IEEE Symposium on VLSI Circuits (VLSIC)*, pp. 35–36, June 2010
19. S. Ikeda, Y. Yoshida, K. Ishibashi, Y. Mitsui, Failure analysis of 6T SRAM on low-voltage and high-frequency operation. IEEE Trans. Electron Devices **50**, 1270–1276 (2003)
20. R. Heald, P. Wang, Variability in sub-100 nm SRAM designs, in *Proceedings of International conference on Computer Aided Design*, pp. 347–352, 2004
21. E. Seevinck, F. List, J. Lohstroh, Static-noise margin analysis of MOS SRAM cells. IEEE J. Solid-State Circ. **22**(5), 748–754 (1987)
22. K. Nii, M. Yabuuchi, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, H. Makino, Y. Yamagami, S. Ishikura, T. Terano, T. Oashi, K. Hashimoto, A. Sebe, S. Okazaki, K. Satomi, H. Akamatsu, H. Shinohara, A 45-nm bulk CMOS embedded SRAM with improved immunity against process and temperature variations. IEEE J. Solid-State Circ. **43**(1), 180–191 (2008)
23. C. Wann, R. Wong, D. Frank, R. Mann, S.-B. Ko, P. Croce, D. Lea, D. Hoyniak, Y.-M. Lee, J. Toomey, M. Weybright, J. Sudijono, SRAM cell design for stability methodology, in *IEEE VLSI-TSA International Symposium on VLSI Technology (VLSI-TSA-Tech)*, pp. 21–22, April 2005
24. A. Kawasumi, T. Yabe, Y. Takeyama, O. Hirabayashi, K. Kushida, A. Tohata, T. Sasaki, A. Katayama, G. Fukano, Y. Fujimura, N. Otsuka, A single-power-supply 0.7 V 1 GHz 45 nm SRAM with an asymmetrical unit-β-ratio memory cell, in *Solid-State Circuits Conference, ISSCC 2008. Digest of Technical Papers. IEEE. International*, pp. 382–622, Feb. 2008
25. M. Sharifkhani, M. Sachdev, SRAM cell stability: A dynamic perspective. IEEE J. Solid-State Circ. **44**(2), 609–619 (2009)
26. M. Sinangil, H. Mair, A. Chandrakasan, A 28 nm high-density 6T SRAM with optimized peripheral-assist circuits for operation down to 0.6 V, in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), IEEE. International*, pp. 260–262, Feb. 2011
27. B. Zhang, A. Arapostathis, S. Nassif, M. Orshansky, Analytical modeling of SRAM dynamic stability, in *IEEE/ACM International Conference on Computer-Aided Design, ICCAD '06*, pp. 315–322, Nov. 2006
28. M. Wieckowski, D. Sylvester, D. Blaauw, V. Chandra, S. Idgunji, C. Pietrzyk, R. Aitken, A black box method for stability analysis of arbitrary SRAM cell structures, in *Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 795–800, March 2010
29. M. Khellah, Y. Ye, N. Kim, D. Somasekhar, G. Pandya, A. Farhang, K. Zhang, C. Webb, V. De, Wordline and bitline pulsing schemes for improving SRAM cell stability in low Vcc 65 nm CMOS designs, in *Proceedings of IEEE Symposium on VLSI Circuits*, pp. 9–10, 2006

30. M. Khellah, D. Somasekhar, Y. Ye, N.S. Kim, J. Howard, G. Ruhl, M. Sunna, J. Tschanz, N. Borkar, F. Hamzaoglu, G. Pandya, A. Farhang, K. Zhang, V. De, A 256-kb dual-V_{CC} SRAM building block in 65-nm CMOS process with actively clamped sleep transistor. IEEE J. Solid-State Circ. **42**(1), 233–242 (2007)
31. L. Chang, D. Fried, J. Hergenrother, J. Sleight, R. Dennard, R. Montoye, L. Sekaric, S. McNab, A. Topol, C. Adams, K. Guarini, W. Haensch, Stable SRAM cell design for the 32 nm node and beyond, in *Symposium on VLSI Technology, 2005. Digest of Technical Papers*, pp. 128–129, June 2005
32. K. Takeda, Y. Hagihara, Y. Aimoto, M. Nomura, Y. Nakazawa, T. Ishii, H. Kobatake, A read-static-noise-margin-free SRAM cell for low-vdd and high-speed applications. IEEE J. Solid-State Circ. **41**(1), 113–121 (2006)
33. I.J. Chang, J.-J. Kim, S. Park, K. Roy, A 32 kb 10 T sub-threshold SRAM array with bit-interleaving and differential read scheme in 90 nm CMOS. IEEE J. Solid-State Circ. **44**(2), 650–658 (2009)
34. S. Jain, S. Khare, S. Yada, V. Ambili, P. Salihundam, S. Ramani, S. Muthukumar, M. Srinivasan, A. Kumar, S.K. Gb, R. Ramanarayanan, V. Erraguntla, J. Howard, S. Vangal, S. Dighe, G. Ruhl, P. Aseron, H. Wilson, N. Borkar, V. De, S. Borkar, A 280 mV-to-1.2 V wide-operating-range IA-32 processor in 32 nm CMOS, in *IEEE. International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 66–68, Feb. 2012
35. K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, M. Bohr, A 3-GHz 70-Mb SRAM in 65-nm cmos technology with integrated column-based dynamic power supply. IEEE J. Solid-State Circ. **41**(1), 146–151 (2006)
36. R.W. Mann, J. Wang, S. Nalam, S. Khanna, G. Braceras, H. Pilo, B.H. Calhoun, Impact of circuit assist methods on margin and performance in 6T SRAM. Solid-State Electron. **54**(11), 1398–1407 (2010)
37. H. Yamauchi, A discussion on SRAM circuit design trend in deeper nanometer-scale technologies. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **18**(5), 763–774 (2010)
38. M. Sinangil, N. Verma, A. Chandrakasan, A reconfigurable 65 nm SRAM achieving voltage scalability from 0.25 to 1.2 V and performance scalability from 20kHz to 200 MHz, in *Solid-State Circuits Conference, 2008. ESSCIRC 2008. 34th European*, pp. 282–285, Sept. 2008
39. B. Campbell, J. Burnette, N. Javarappa, V. von Kaenel, Power-efficient dual-supply 64 kB L1 caches in a 65 nm CMOS technology, in *Proceedings of IEEE Custom Integrated Circuits Conference*, pp. 729–732, 2007
40. T. Suzuki, H. Yamauchi, K. Satomi, H. Akamatsu, A stable SRAM mitigating cell-margin asymmetricity with a disturb-free biasing scheme, in *Proceedings of IEEE Custom Integrated Circuits conference*, pp. 233–236, 2007
41. R. Joshi, R. Houle, D. Rodko, P. Patel, W. Huott, R. Franch, Y. Chan, D. Plass, S. Wilson, S. Wu, and R. Kanj, A high performance 2.4 Mb L1 and L2 cache compatible 45 nm SRAM with yield improvement capabilities, in *Proceedings of IEEE Symposium on VLSI Circuits*, pp. 208–209, June 2008
42. J. Pille, C. Adams, T. Christensen, S. Cottier, S. Ehrenreich, T. Kono, D. Nelson, O. Takahashi, S. Tokito, O. Torreiter, O. Wagner, D. Wendel, Implementation of the cell broadband engine in a 65 nm SOI technology featuring dual-supply SRAM arrays supporting 6 GHz at 1.3 V, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, pp. 322–606, 11–15 Feb. 2007
43. K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, M. Bohr, A 3-GHz 70 MB SRAM in 65 nm CMOS technology with integrated column-based dynamic power supply, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, vol. 1, pp. 474–611, 10–10 Feb. 2005
44. M. Yamaoka, T. Kawahara, Operating-margin-improved SRAM with column-at-a-time body-bias control technique, in *33rd European Solid State Circuits Conference, 2007. ESSCIRC*, pp. 396–399, 11–13 Sept. 2007
45. O. Hirabayashi, A. Kawasumi, A. Suzuki, Y. Takeyama, K. Kushida, T. Sasaki, A. Katayama, G. Fukano, Y. Fujimura, T. Nakazato, Y. Shizuki, N. Kushiyama, T. Yabe, A process-

variation-tolerant dual-power-supply SRAM with 0.179 um2 cell in 40 nm CMOS using level-programmable wordline driver, in *IEEE International on Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009*, pp. 458–459,459a, Feb. 2009

46. F. shi Lai, C.-F. Lee, On-chip voltage down converter to improve SRAM read/write margin and static power for sub-nano CMOS technology. IEEE J. Solid-State Circ. **42**(9), 2061–2070 (2007)
47. Y. Hirano, M. Tsujiuchi, K. Ishikawa, H. Shinohara, T. Terada, Y. Maki, T. Iwamatsu, K. Eikyu, T. Uchida, S. Obayashi, K. Nii, Y. Tsukamoto, M. Yabuuchi, T. Ipposhi, H. Oda, Y. Inoue, A robust SOI SRAM architecture by using advanced ABC technology for 32 nm node and beyond LSTP devices, in *Proceedings of IEEE Symposium on VLSI Technology*, pp. 78–79, June 2007
48. Y. Morita, H. Fujiwara, H. Noguchi, K. Kawakami, J. Miyakoshi, S. Mikami, K. Nii, H. Kawaguchi, and M. Yoshimoto, A Vth-Variation-Tolerant SRAM with 0.3-V minimum operation voltage for memory-rich SoC under DVS environment, in *Proceedings of IEEE Symposium on VLSI Circuits*, pp. 13–14, 2006
49. H. Pilo, C. Barwin, G. Braceras, C. Browning, S. Lamphier, F. Towler, An SRAM design in 65-nm technology node featuring read and write-assist circuits to expand operating voltage. IEEE J. Solid-State Circ. **42**(4), 813–819 (2007)
50. Y.H. Chen, G. Chan, S.Y. Chou, H.-Y. Pan, J.-J. Wu, R. Lee, H. Liao, H. Yamauchi, A 0.6 V dual-rail compiler SRAM design on 45 nm CMOS technology with adaptive SRAM power for lower VDD_{min} VLSIs. IEEE J. Solid-State Circ. **44**(4), 1209–1215 (2009)
51. S. Ohbayashi, M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Imaoka, Y. Oda, T. Yoshihara, M. Igarashi, M. Takeuchi, H. Kawashima, Y. Yamaguchi, K. Tsukamoto, M. Inuishi, H. Makino, K. Ishibashi, H. Shinohara, A 65-nm SoC embedded 6T-SRAM designed for manufacturability with read and write operation stabilizing circuits. IEEE J. Solid-State Circ. **42**(4), 820–829 (2007)
52. A. Raychowdhury, B. Geuskens, J. Kulkarni, J. Tschanz, K. Bowman, T. Karnik, S.-L. Lu, V. De, M. Khellah, PVT-and-aging adaptive wordline boosting for 8T SRAM power reduction, in *IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 352–353, Feb. 2010
53. M. Yamaoka, N. Maeda, Y. Shinozaki, Y. Shimazaki, K. Nii, S. Shimada, K. Yanagisawa, T. Kawahara, 90-nm process-variation adaptive embedded SRAM modules with power-line-floating write technique. IEEE J. Solid-State Circ. **41**(3), 705–711 (2006)
54. J. Kulkarni, B. Geuskens, T. Karnik, M. Khellah, J. Tschanz, V. De, Capacitive-coupling wordline boosting with self-induced VCC collapse for write VMIN reduction in 22-nm 8T SRAM, in *IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 234–236, Feb. 2012
55. E. Karl, Y. Wang, Y.-G. Ng, Z. Guo, F. Hamzaoglu, U. Bhattacharya, K. Zhang, K. Mistry, M. Bohr, A 4.6 GHz 162 Mb SRAM design in 22 nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry, in *IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 230–232, Feb. 2012
56. S. Damaraju, V. George, S. Jahagirdar, T. Khondker, R. Milstrey, S. Sarkar, S. Siers, I. Stolero, A. Subbiah, A 22 nm IA multi-CPU and GPU system-on-chip, in *IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 56–57, Feb. 2012
57. M. Khellah, N.S. Kim, Y. Ye, D. Somasekhar, T. Karnik, N. Borkar, F. Hamzaoglu, T. Coan, Y. Wang, K. Zhang, C. Webb, V. De, PVT-variations and supply-noise tolerant 45 nm dense cache arrays with diffusion-notch-free (DNF) 6T SRAM cells and dynamic multi-vcc circuits, in *2008 IEEE Symposium on VLSI Circuits*, pp. 48–49, June 2008
58. B. Mohammad, M. Saint-Laurent, P. Bassett, J. Abraham, Cache design for low power and high yield, in *9th International Symposium on Quality Electronic Design, 2008. ISQED 2008*, pp. 103–107, March 2008
59. Y. Wang, E. Karl, M. Meterelliyoz, F. Hamzaoglu, Y.-G. Ng, S. Ghosh, L. Wei, U. Bhattacharya, K. Zhang, Dynamic behavior of SRAM data retention and a novel transient voltage collapse technique for 0.6 V 32 nm LP SRAM, in *IEEE International on Electron Devices Meeting (IEDM)*, pp. 32.1.1–32.1.4, Dec. 2011

60. K. Nii, M. Yabuuchi, Y. Tsukamoto, S. Ohbayashi, Y. Oda, K. Usui, T. Kawamura, N. Tsuboi, T. Iwasaki, K. Hashimoto, H. Makino, H. Shinohara, A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment, in *IEEE Symposium on VLSI Circuits*, pp. 212–213, June 2008
61. Y. Fujimura, O. Hirabayashi, T. Sasaki, A. Suzuki, A. Kawasumi, Y. Takeyama, K. Kushida, G. Fukano, A. Katayama, Y. Niki, T. Yabe, A configurable SRAM with constant-negative-level write buffer for low-voltage operation with 0.149um2 cell in 32 nm high-k metal-gate CMOS, in *IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 348–349, Feb. 2010
62. N. Shibata, H. Kiya, S. Kurita, H. Okamoto, M. Tan'no, T. Douseki, A 0.5-V 25-MHz 1-mW 256-kb MTCMOS/SOI SRAM for solar-power-operated portable personal digital equipment - sure write operation by using step-down negatively overdriven bitline scheme. IEEE J. Solid-State Circ. **41**(3), pp. 728–742, March 2006
63. S. Mukhopadhyay, R. Rao, J.-J. Kim, C.-T. Chuang, SRAM write-ability improvement with transient negative bit-line voltage. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **19**(1), 24–32 (2011)
64. H. Pilo, I. Arsovski, K. Batson, G. Braceras, J. Gabric, R. Houle, S. Lamphier, F. Pavlik, A. Seferagic, L.-Y. Chen, S.-B. Ko, C. Radens, A 64 Mb SRAM in 32 nm High-k metal-gate SOI technology with 0.7 V operation enabled by stability, write-ability and read-ability enhancements, in *IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 254–256, Feb. 2011
65. H. Nho, P. Kolar, F. Hamzaoglu, Y. Wang, E. Karl, Y.-G. Ng, U. Bhattacharya, K. Zhang, A 32 nm High-k metal gate SRAM with adaptive dynamic stability enhancement for low-voltage operation, in *IEEE International on Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 346–347, Feb. 2010
66. M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, S. Imaoka, H. Makino, Y. Yamagami, S. lshikura, T. Terano, T. Oashi, K. Hashimoto, A. Sebe, G. Okazaki, K. Satomi, H. Akamatsu, H. Shinohara, A 45 nm low-standby-power embedded SRAM with improved immunity against process and temperature variations, in *Proceedings of the International Solid-State Circuits Conference ISSCC*, pp. 326–606, 11–15 Feb. 2007
67. M. H. Abu-Rahma, M. Anis, S.S. Yoon, A robust single supply voltage SRAM read assist technique using selective precharge, in *Proceedings of the 34th European Solid State Circuits Conference ESSCIRC*, pp. 234–237, 2008
68. B. Wicht, T. Nirschl, D. Schmitt-Landsiedel, Yield and speed optimization of a latch-type voltage sense amplifier. IEEE J. Solid-State Circ **39**(7), 1148–1158 (2004)