Buried Interconnects for Sub-5 nm SRAM Design

R. Mathur, Senior Member, IEEE, M. Bhargava, Member, IEEE, B. Cline, Member, IEEE, S. Salahuddin, A. Gupta, P. Schuddinck, J. Ryckaert, and J. P. Kulkarni, Senior Member, IEEE

Abstract — The metal resistance increase due to interconnect pitch scaling was traditionally offset by interconnect length scaling. This is no longer the case for deeply scaled nodes with narrow critical dimensions (CDs) of metals. Static random access memories (SRAMs) route long critical signals such as bitlines and wordlines in lower metal layers of the back-end-of-line (BEOL) and are particularly affected by this metal resistance increase. Buried power rail (BPR) has been proposed in sub-5-nm nodes for routing power and ground lines to improve the performance and density of standard cells and mitigate voltage IR drop issues. We extend the work by exploring the use of buried interconnect for both signal and power routing in SRAMs with minimal process flow changes to the already proposed BPR technology. A high-accuracy 3-D field solver is used for accurate parasitic extraction of SRAM bitcells with buried interconnects. Industry-standard methods are used to evaluate the SRAM macro-level power–performance metrics. We show that the buried interconnects can improve SRAM access time by up to 11%, write time by up to 28%, and dynamic power by 4%, effectively equivalent to one full technology-node gain improvement.

Index Terms — Bitline (BL), buried power rail (BPR), Design Technology Co-Optimization (DTCO), interconnect, resistance, static random access memories (SRAM), wordline (WL).

I. INTRODUCTION

INTERCONNECTS play a critical role in enabling high-performance logic and memory in advanced CMOS technology. Copper (Cu) is widely used in modern interconnects because of its low bulk resistivity and high electrical reliability [1]. However, at nanoscale dimensions, its resistivity increases rapidly due to increased electron scattering at the surface and grain-boundary interfaces [2]. Furthermore, Cu interconnects require a barrier to prevent metal diffusion into the dielectric and a liner to promote metal fill. The scaling of the barrier and liner is nonideal, causing an increasingly lower proportion of the wire cross section available for Cu that is responsible for electrical conduction [3], [4]. Furthermore, the impact from line edge roughness to interconnect resistance increases when scaling to narrower wires [5]. As a result, interconnect resistance has entered an exponential increase regime, becoming a key system performance limiter in sub-5-nm technology nodes [6]–[8]. Therefore, there is a need to explore alternative materials and approaches to reduce interconnect resistance.

Ruthenium (Ru) is one of the alternative materials that can augment or replace Cu in modern interconnects [12], [13]. It offers a lower resistance in narrow wires [14] and a higher metal to barrier/liner volume ratio compared to Cu [15], [16]. Ru can also withstand the front-end-of-line (FEOL) process thermal budget [17]. Buried interconnects in FEOL oxide and substrate between the transistor fins have been proposed for power routing (also called buried power rail or BPR) [18], [19]. The aspect ratio of BPR can be made taller to achieve a low IR drop in power distribution network (PDN). A study at 3-nm process node on an Arm CPU showed that BPR could reduce the worst case voltage IR drop by 1.7×–7× [20]. A part of the improvement in IR drop can also be traded off with a more relaxed strapping distance to the global power grid, alleviating congestion in lower metal layers and possibly resulting in smaller design footprints [17].

Static random access memories (SRAMs) are particularly sensitive to the rapidly increasing metal resistance in advanced nodes. Critical signals, such as wordlines (WLs) and bitlines (BLs), are typically routed in lower level metal layers of back-end-of-line (BEOL) over long lengths to support a large subarray design and achieve a high SRAM area density. Fig. 1 shows the scaling trend for resistance of WL ($R_{WL}$) and BL ($R_{BL}$) in advanced process nodes. It is observed that the resistance of these metals has increased by 2–3.5× from 14-/16- to 3-nm process node.
BPR is considered as one of the important scaling boosters in the sub-5-nm process technology nodes [19]. This work assumes that the BPR technology may be available for logic design in future technology offerings and therefore available for SRAM design for use in the best possible way. However, BPR may not lead to straightforward scaling of the SRAM bitcell as the bitcell dimensions may not be metal limited. This work, instead, explores how it can be utilized to benefit SRAM design by mitigating the high resistance of critical signals and improving the SRAM power performance at the macro level. The BPR approach has been shown to improve SRAM (BPR-SRAM) performance by creating space for wider BL and WL tracks [9], [10]. However, wider wires increase the metal capacitance, limiting signal delay gain and degrading dynamic power. We previously reported the direct use of buried interconnects for signal routing in the form of buried BLs (BBLs) in SRAM [11]. This article is an enhanced and more complete version of it while making the following new contributions:

1) a new bitcell configuration of BBL-buried VSS (BVSS) SRAM featuring both buried signal and power routing.

2) a simulation analysis to identify thickness (height) of BVSS metal for minimal voltage IR drop on the bitcell ground.

3) comparison of different buried interconnects bitcell variants for SRAM macro-level power and performance.

II. SRAM Design With Buried Metals

Both BLs and WLs are regularly placed, long metal interconnects across a large SRAM array, and are good candidates for buried signaling. In the advanced FinFET SRAM layout, BLs are routed along the direction of fin orientation, whereas WLs are routed orthogonally to the fin orientation. Buried WL would require significant modifications to the BPR process flow in which etching and fill for buried interconnects are done post fin patterning [19]. Contrarily, BBL does not require any significant change to the process flow with BPR.

Besides the processing challenges, there are electrical reasons for preferring BBL over buried WL. A relatively large CMOS gate drives the WL signal, and the WL metal can be strapped with another higher metal layer of BEOL (e.g., double WL [21]) to reduce its resistance at the expense of some capacitance increase, but an overall improvement in $RC$ delay. Furthermore, the WL slew rate can be improved by inserting a WL repeater after a specific interval [22]. On the other hand, the analog BL signal is driven by high-$V_t$, minimum-sized (typically 1 or 2 fins) SRAM bitcell transistors during a read operation. Consequently, BL is not strapped with a higher metal to avoid increased capacitance, which would slow the small-signal BL differential development ($C_{BL} \cdot V_{DD}/I_{read}$) during the SRAM read operation. Furthermore, SRAM dynamic power is a much stronger function of the capacitance of BLs than the capacitance of WLs, and higher BL capacitance will lead to a significant increase in SRAM dynamic power. For these reasons, BBL became the primary choice for design exploration, and the buried WL option is not considered further in this work. Note that the BL $RC$ can be mitigated to some extent by employing flying BL [21] and double write driver [23] circuit techniques. However, these techniques are independent of the process technology and can be used in conjunction with the various buried interconnect SRAM approaches as described next.

A. Baseline SRAM

The baseline design [Fig. 2(a)] is a typical thin-cell layout of a 1-1-1 [pull-up transistor (PU)-pass-gate transistor...
Fig. 3. Layout, cross-sectional, and 3-D view of (a) 1-1-1 fin HD bitcell and (b) 1-2-2 fin HC bitcell with BBL-BVSS as seen in QuickCap NX tool. The bitcell dimension of HD cell is 168 nm × 90 nm and the HC cell is 210 nm × 90 nm.

(PG)-pull-down transistor (PD)] fin high-density (HD) SRAM bitcell drawn using the imec’s N5 process design rules (Table I). The baseline bitcell is 168 nm in height and 90 nm in width, and these dimensions are maintained for all subsequent bitcell variants with buried interconnect. The power [Voltage Drain Drain (VDD)], ground [Voltage Source Source (VSS)], BL [BL/negative bit line (NBL)], and bitcell storage nodes (Q/QB) nets utilize the horizontal MINT (lowermost local metal interconnect) layer. Due to the crowding of nets in the MINT layer, the width of the BL wire is only 11 nm. The WL and VSS utilize the vertical M1 layer. The WL is also strapped with a wide wire in the M2 layer (not shown in the figure) for lower resistance (common to all configurations). The N5 process node allows M1 and M2 to be routed parallel over the SRAM array. However, for a different process node where M1 and M2 are strictly orthogonal, WL could be routed in M1 and M3.

B. BPR SRAM

The BPR SRAM bitcell [Fig. 2(b)] routes the VDD and VSS nets in buried metal layer (MBUR) [9], [10]. This alleviates congestion in the MINT and M1 layers to widen BL and WL wires. Wider wires reduce the resistance significantly but at the expense of some increase in capacitance.

C. BBL SRAM

Recognizing the fact that BLs are routed in the direction of the transistor fin, the BBL SRAM [Fig. 2(c)] routes the BLs in the MBUR layer in the FEOL oxide between the fins [11]. The BBL signal experiences a larger coupling capacitance from the substrate but smaller coupling capacitance from other BEOL metals. BBL can achieve lower resistance of BL depending on the thickness (height) of the buried metal.

D. BBL-BVSS SRAM

BBL-BVSS is a unique bitcell where both power (VSS) and signal (BL) nets are proposed to be buried [Fig. 2(d)]. This design improves upon the state-of-the-art BBL bitcell [11] by additionally creating space in the M1 layer for wider WL wire through BVSS. The VDD net is not buried to reduce coupling capacitance to BBL. VSS routed below the device and VDD above the device are done only for the array section, which is usually custom done for SRAMs. Fig. 3 shows the top-level layout and 2-D and 3-D cross-section view of the BBL-BVSS bitcell as seen in the QuickCap NX [24].

Due to the high aspect ratio of BPR (AR = 7), if the same process is used for BBL, SRAM performance would be severely degraded due to high BL capacitance. Hence, the buried metal thickness is varied across a wide range of values to quantify $R_{BL}$ and $C_{BL}$ sensitivity (Fig. 4). $C_{BL}$ decreases almost linearly with decreasing thickness since the capacitance to the substrate reduces as the buried metal is confined inside the shallow trench isolation (STI) dielectric region. As expected, $R_{BL}$ increases with decreasing BBL thickness. The buried metal parameters for BBL (width = 21 nm, thickness = 30 nm, and AR ~1.5) for optimal BL signal $RC$ delay for the N5 process node from Design Technology Co-Optimization (DTCO) in [11] are maintained for BBL-BVSS SRAM. However, keeping the same thickness for BVSS as BBL may lead to a suboptimal power grid.
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4
IEEE TRANSACTIONS ON ELECTRON DEVICES

Fig. 4. Impact of thickness of buried metal on the resistance and capacitance of BBL.

Fig. 5. Impact of thickness of BVSS metal on the bitcell VSS resistance and the bitcell read current simulated at SS/0.63 V/−40 °C.

with high-voltage IR drop on the bitcell ground (Fig. 5). Consequently, the bitcell read current may degrade by as much as 12% (at SS/0.63 V/−40 °C), assuming 256 rows per BL SRAM subarray design and tap connections to BEOL VSS grid available only at the two edges of the subarray. A buried metal with a width of 21 nm and a thickness of 63 nm or larger (AR > = 3) can be utilized to limit the voltage IR drop on BVSS within 5 mV and bitcell read current degradation to 4% or less.

The VDD net (routed in MINT) could also incur voltage IR drop depending on the subarray size and metal resistance due to the lack of connection to a higher metal grid. Note that there is no active current drawn from the VDD terminal for the read operation; therefore, the IR drop on the VDD net is negligible. For write, the VDD net delivers current initially during preflip contention and then later for flip completion. The IR drop in VDD due to the flip completion current may increase the write completion time, but the IR drop due to the preflip contention current usually helps the write operation. The IR drop during a write can impact the stability of bitcells in the unselected columns (in a design with column mux > 1). If this effect is significant, then the VDD nets per column may be decoupled at the edge of the array, similar to how they are configured in transient voltage collapse (TVC) write-assist implementation [25]. Another alternative is introducing gap or break cells in the SRAM array to connect the VDD to a more robust grid in higher metal. The bitcell layout of Fig. 2(d) could also be modified by routing VDD in MBUR, which has a much lower resistance due to its tall aspect ratio. This method comes at the slight expense of increasing coupling capacitance between buried VDD and BBL. However, in the extraction results, the increase in BBL cap was less than 2%.

Table II summarizes the WL and BL metal width and thickness for the different SRAM configurations. The metal width values are chosen to optimize WL and BL signal RC for the N5 process node, and they could be different for other technology nodes.

### III. Parasitic Extraction of Buried Interconnect

Accurate parasitic extraction of the buried metal to the neighboring metals, device structures, and the substrate is necessary to gain insights into the extent of coupling of BL and WL with other nets and substrate. $C_{WL}$ and $C_{BL}$ splits of different SRAM configurations are plotted in Fig. 6 for HD bitcell. The BPR SRAM [10] exhibits a smaller contribution to $C_{WL}$ and $C_{BL}$ from BVSS when compared to the baseline SRAM where VSS is routed adjacent to BL (in MINT) and the WL (in M1 and M2). However, the reduction in $C_{WL}$ and $C_{BL}$ is offset by an increase in coupling between wider WL and BL metals to other neighboring metals. Besides,
in the absence of VSS in M1, the WL experiences coupling from neighboring cell metals (WL_L and WL_R). The BBL SRAM [11] experiences a larger contribution to \( C_{BL} \) from the substrate due to its proximity with BBL. However, the increase in \( C_{BL} \) is balanced by reduced coupling from other BEOL nets such as Q and VSS. For BBL-BVSS, the coupling capacitance contribution from VSS to \( C_{WL} \) is smaller, but the coupling capacitance contribution from VSS to \( C_{BL} \) is higher as both VSS and BL are buried. Overall, both BBL and BBL-BVSS SRAM designs maintain similar \( C_{BL} \) compared to the baseline despite doubling the width and thickness of the BL metal (MBUR). Furthermore, BBL-BVSS SRAM achieves lower \( C_{BL} \) compared to the baseline despite doubling the width of the WL metal (M1).

Fig. 7 extends the comparison of WL and BL by including the resistance plot and adding data for 1-2-2 fin high current (HC) bitcell. While the BPR SRAM [10] results in lower \( R \) with wider WLs and BLs, it suffers from increased \( C \) compared to the baseline. BBL SRAM [11] achieves lower \( R_{WL} \) when compared to both the baseline or BPR SRAM and lower \( C_{BL} \) when compared to BPR SRAM. However, BBL SRAM does not provide any direct optimizations for the WL and suffers from worse \( R_{BL} \) compared to BPR SRAM. BBL-BVSS SRAM improves BBL SRAM by preserving its lower \( R_{BL} \) and additionally lowering \( R_{WL} \).

The extracted netlist of SRAM bitcell is obtained by using Synopsys QuickCap NX. This high-accuracy 3-D field solver is ideally suited for early process exploration of novel concepts such as buried interconnects. The Graphic Design System (2nd version) (GDSII) layout views of different SRAM bitcell configurations along with a QuickCap Technology File (QTF) from the imec’s N5 process node are provided as input to the tool. The QTF accurately describes the process geometries and allows modifying process parameters such as buried metal thickness and depth for DTCO. The extracted netlist of the bitcell is stitched together to create an SRAM subarray (Fig. 8) critical path netlist. The selected subarray size (258 rows × 136 columns) is commonly used in memory macros for L1/L2 caches [26]. Macro-level simulations using the SPICE models for imec’s 5-nm process node are performed for evaluating SRAM read, write, and dynamic power. The complete extraction and simulation framework is shown in Fig. 9.
IV. RESULTS AND DISCUSSION

This section presents the results from the macro-level simulations under the conditions in Table III. The power performance of SRAMs with different buried metal interconnect configurations is compared.

A. SRAM Read

The SRAM read margin is a measure of the voltage differential (SAdiff) developed at the BL pair when the sense amplifier (SA) is triggered. For iso-performance comparison (read access time of 710 ps), the SA activation (WL to SA trigger time) is chosen such that the SAdiff is 150 mV for the HD bitcell in the baseline design. Fig. 10(a) shows that BBL-BVSS SRAM improves the SAdiff by 23 mV (15%) for the HD bitcell and 27 mV (15%) for the HC bitcell compared to the baseline design. The improvement in read margin can be attributed to the improvement in WL and BL signal slews (Table IV). The excess read margin can be traded off for read access-time (CLK to Q) improvement by triggering the SA earlier, as shown in Fig. 10(b). This translates to an iso-margin read access-time improvement of 77 ps (11%) and 65 ps (10%) for the HD and HC bitcell, respectively.

B. SRAM Write

Static write margin is quantified as the minimum BL voltage (VNBL) required to flip the SRAM bitcell without any timing constraints. It can be considered as a theoretical limit case where the signal has enough time to settle to its final steady-state value. A negative static write margin means that BL is driven below VSS (or negative voltages) and indicates the need for write-assist features. VNBL is a strong function of RBBL, and therefore, both BBL and BBL-BVSS with the lowest RBBL among the compared configurations outperform all others for both the HD and HC bitcells [Fig. 10(c)].

Authorized licensed use limited to: University of Texas at Austin. Downloaded on January 30,2022 at 15:36:57 UTC from IEEE Xplore. Restrictions apply.
BPR have the same BL width and thickness for the HC bitcells leading to the same \( R_{BL} \), and write margin. This is also the same case for BBL and BBL-BVSS SRAM for the HD and HC bitcells. The dynamic aspect of the write operation, which includes timing constraints such as WL and BL RC parasitic effects and signal slews, is captured in the write time (WL rise to bit flip) results. A high write margin can be traded off for performance by lowering the write time. Fig. 10(d) shows the write time for various configurations for a fixed \( V_{NBL} = -45 \text{ mV} \). As expected, the HD baseline bitcell is not write-able (\( V_{NBL} \) required = \( -161 \text{ mV} \)). BBL-BVSS provides 28% better write time compared to BPR-SRAM for the HD bitcell and 19% better write time compared to the baseline for the HC bitcell. This write time gain can potentially translate into cycle-time (frequency) improvement if the write operation happens to be the performance limiter for SRAMs in advanced process nodes.

C. SRAM Dynamic Power

Multiple highly capacitive BL nets experiencing voltage swings during SRAM read/write operations can contribute to over 50% of the total SRAM power [27]. Unlike the BPR configuration, both the BBL and BBL-BVSS achieve low \( R_{BL} \), while maintaining or achieving modest gains in \( C_{BL} \) compared to the baseline design (Fig. 7). The dynamic power improvement with BBL and BBL-BVSS SRAM relative to the baseline is summarized in Table V.

V. CONCLUSION

Deepest scaled nodes with a narrow critical dimension of metals suffer from increasingly higher interconnect resistance, which becomes a system performance bottleneck. Therefore, exploring methods for lower interconnect resistance is essential. This work proposes using a buried interconnect for signal routing in SRAM design to reduce the metal resistance of critical signals. High-accuracy 3-D field solver-based parasitic extraction is performed to understand the coupling between the buried metal and other FEOL and BEOL structures. A new BBL-BVSS SRAM bitcell design is proposed, which provides significant gains over the baseline, BPR, and BBL SRAM variants requiring minimal process changes to the already demonstrated BPR technology. Accurate parasitic extraction of SRAM bitcells with buried interconnects is performed, and industry-standard methods are used to evaluate the SRAM macro-level power–performance metrics. Our findings suggest that buried interconnects, if carefully sized, can help mitigate the interconnect resistance issue in SRAMs and augment the technology scaling roadmap for sub-5-nm process nodes.

ACKNOWLEDGMENT

The 5-nm technology files and simulation models were provided by imec. Tool support for extraction was provided by Synopsys.

REFERENCES