# Buried Bitline for sub-5nm SRAM Design

R. Mathur*§, M. Bhargava*, S. Salahuddin†, P. Schuddinck†, J. Ryckaert†, S. Annamalai‡, A. Gupta†,
Y. K. Chong*, S. Sinha*, B. Cline*, and J. P. Kulkarni§
*Arm Inc., 5707 Southwest Parkway, Austin, TX, USA. †Imec, Kapeldreef 75, 3000 Leuven, Belgium.
‡Synopsys, 690 E Middlefield Rd, Mountain View, CA, USA. §University of Texas at Austin, TX, USA.
Email: *rahul.mathur@arm.com*

***Abstract*—Buried power rail (BPR), i.e., metal wires below the active transistors, has been proposed for routing power and ground lines to improve the performance and density of standard cells and mitigate the increasing $RC$ parasitics at sub-5nm CMOS technology nodes. In this work, we present the Buried Bit-Line (BBL) SRAM technique which utilizes buried metal interconnects for signal routing instead of power or ground routing to achieve better SRAM performance and lower power consumption while requiring minimal process flow changes to the buried power rail technology. Design technology co-optimization (DTCO) is performed using high-accuracy 3D field solvers for parasitic extraction and industry standard techniques to quantify the effect of BBL technology parameters on SRAM circuit metrics. We show that the proposed BBL SRAM can improve access time by up to 11%, write time by up to 31%, and dynamic power by 4%, effectively equivalent to a technology-node gain improvement.**

## I. Introduction

As transistor dimensional scaling reaches its physical and electrical limits at sub-5nm technology nodes, buried power rail (BPR) technology is a critical scaling booster proposed on the technology roadmap for area scaling as shown in Fig. 1 [1]. Buried power rails are wires that are *buried* in the silicon substrate below the active transistors as shown in Fig. 2. These buried wires can be both wider and/or taller (with demonstrated aspect ratios of up to 7 as shown in Fig. 3) [2], [3] leading to higher capacitance and lower resistance, which is ideal for power and ground routing [4].

Shrinking dimensions while increasing SRAM macro capacity have led to long wires for wordlines (WL) and bitlines (BL), which need to be routed on lower level metal layers. This makes SRAMs particularly sensitive to the rapidly increasing wire resistance in advanced nodes shown in Fig. 4. The BPR approach has been shown to improve SRAM (BPR-SRAM) performance by enabling use of wider and larger spaced BL and WL tracks [5], [6].

In this work we extend the exploration of buried wires for SRAM designs in the form of buried bitlines (BBL-SRAM). We present a detailed design-technology co-optimization (DTCO) analysis to identify the optimum buried wire parameters for BBL-SRAM and evaluate performance and power metrics with the constraint of minimal modification to the buried power rail process, for minimal process cost increase.

## II. Buried Bitline SRAM Design

Both BLs and WLs are regularly placed, long metal interconnects across a large SRAM array and potentially are good candidates for buried signaling. In advanced CMOS FinFET SRAMs employing thin-cell bitcell layout, the power/ground/BLs are routed along the direction of fin orientation while WLs are routed orthogonal to the fin orientation. Hence, BLs are better candidates for buried signaling since they do not require any process change from the BPR approach. On the other hand, buried WL would require either a bitcell layout change or significant modifications to the process steps. Also, the WL signal is unidirectional and digital in nature (with binary values), so it can be strapped with a higher metal layer (e.g., double wordline [7]) to reduce its resistance. In addition, WL slew-rate can be improved by inserting a WL repeater at regular intervals, hence buried WL approach is not considered further in this work. Contrarily, the analog BL signal is driven by the minimum sized (typically 1 or 2 fins) SRAM bitcell transistors during a read operation. As a result, bitlines are not strapped with a higher metal to avoid increased capacitance in small-signal bitline differential development time ($C_{BL} \cdot V_{DD}/I_{read}$) and dynamic power. The BL $RC$ can be mitigated to a some extent by employing Flying Bitline [7] and Double Write Driver [8] circuit techniques. However, these techniques are independent of the process technology and can be used in conjunction with the proposed BBL-SRAM approach. Fig. 4 shows the trend for resistance and capacitance of WL ($R_{WL}$, $C_{WL}$) and BL ($R_{BL}$, $C_{BL}$) scaling from the 14/16nm node to the 5nm process node. $R_{WL}$ trend is better than $R_{BL}$ due to WL strapping and/or circuit techniques.

To quantify the effectiveness of buried signal routing on $R_{BL}$ and $C_{BL}$, accurate parasitic extraction of the BBL to the neighboring metals, devices, and the substrate is necessary. It is obtained by using Synopsys QuickCap® NX [9]. QuickCap NX is a random-walk, high accuracy 3D field solver which is ideally suited for early process exploration of novel concepts such as buried signaling. Fig. 5 shows the 1-1-1 fin SRAM bitcell layout employing the proposed BBL. Based on imec's iN5 layout dimensions mentioned in Table I, 21nm wide BBL can replace the BPR in the STI region. Capacitance associated with buried metal inside the substrate is extracted as junction capacitance. Fig. 6 shows the 2D and 3D cross-section view of the BBL-SRAM bitcell layout as seen in the QuickCap NX.

Fig. 7 shows a comparison of BBL capacitance having default BPR thickness (147nm) with a conventional bitline. Due to high aspect ratio of BPR (AR=7), if the same process is used for BBL, SRAM performance would be severely degraded. Hence, the buried metal thickness is varied across a wide range of values to quantify $R_{BL}$ and $C_{BL}$ sensitivity (Fig. 8). $C_{BL}$ decreases 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, the $R_{BL}$ increases with decreasing BBL thickness. For further analysis, three different points on this curve as bitline dimensions for optimized $C$ ($C_{opt}$), optimized $R$ ($R_{opt}$), and optimized $RC$ ($RC_{opt}$) are chosen as shown in Table II. Fig. 9 and 10 plot the capacitance and resistance of WL and BL for the above configurations. The BPR-SRAM [6] variant, with wider WLs and BLs (resulting in lower $R$, but higher $C$ compared to the baseline) is also included. For the BBL, the buried metal experiences higher capacitance from the substrate, but reduced capacitance from other BEOL metals as it is placed further away from them. BBL $RC_{opt}$ achieves lower $R_{BL}$ and $C_{BL}$ when compared to both the baseline or BPR SRAM. These three BBL configurations can provide options to tradeoff $R$ and $C$ to meet the needs of high performance or low power array designs.

## III. Results and Discussion

The power and performance metrics for different configurations listed in Table II for a representative 34.5kb (256 rows $\times$ 136 columns) SRAM sub-array (Fig. 11), commonly used in L1/L2 caches, is evaluated. The simulation framework and conditions are described in Fig. 12 and Table III respectively.

**SRAM read:** Read margin is quantified as the bitline differential just before the the sense amplifier (SA) trigger. For *iso-performance* comparison, the SA activation is chosen such that the read margin is 150mV for the 111 bitcell in the baseline design. Figure 13 shows that for the 111 bitcell, BBL-$C_{opt}$ has the best read margin (27mV or 18% higher than baseline). For the 122 bitcell (wider cell with more pre-dominant impact of $R_{WL}$), the BPR-SRAM shows the best read margin (21mV or 12% higher). The excessive read margin can be traded off for access-time (performance) improvement by triggering the SA earlier, as shown in Fig. 14. This translates to an *iso-margin access time* improvement of 12% and 8% for the 111 bitcell with BBL-$C_{opt}$ and 122 bitcell with BPR-SRAM respectively.

**SRAM write:** *Static write margin* is quantified as the minimum negative bitline voltage ($V_{NBL}$) required to flip the SRAM bitcell, without any timing constraints. $V_{NBL}$ is a strong function of $R_{BL}$ and therefore, as expected, BBL-Ropt with the lowest $R_{BL}$ amongst the compared configurations, out-performs all others for both the 111 and the 122 bitcells (Fig. 15). For the 111 bitcell, it provides 163mV and 26mV improvements over the baseline and BPR-SRAM, respectively. In fact, BBL-$R_{opt}$ is the only configuration with a positive static write margin (i.e., has the ability to write *without* a NBL write assist). For the 122 bitcell, BBL-$R_{opt}$ shows write margin improvements of 26mV over both the baseline as well as BPR-SRAM. Again, high margin can be traded off for performance by lowering the time to write (WL rise to bit flip). Fig. 16 shows the write time for various configurations for a fixed $V_{NBL}$ = 45mV. As expected, 111 baseline bitcell is not write-able ($V_{NBL}$ required = 161mV). BBL-$R_{opt}$ write time for 111 bitcell is 106ps or 32% better than BPR-SRAM. For the 122 bitcell, and no-assist case (i.e.,$V_{NBL}$ = 0) the BBL-$R_{opt}$ has 62ps or 22% better write time than BPR-SRAM. This write-time improvement can potentially translate into cycle-time improvement if the write operation happens to be the performance limiter for SRAMs in advanced process nodes.

**SRAM dynamic power:** The highly capacitive BL nodes experiencing large voltage swings during SRAM read/write operations can contribute to over 50% of of the total sub-array power. Fig. 17 shows that for both the 111 and 122 bitcells, the best power is provided by the BBL-$C_{opt}$ configuration (as expected, as it has the minimum $C_{BL}$, as seen in Fig. 9). To summarize, (i) for the 111 bitcell, BBL-$C_{opt}$ is the best configuration for the read margin, access time, and dynamic power, while BBL-$R_{opt}$ is best for write-margin and write-time, (ii) for the 122 bitcell, BPR-SRAM option is the best for read margin and access time, BBL-$R_{opt}$ is the best configuration for write-margin and write-time; BBL-$C_{opt}$ is the best option for dynamic power.

**Explorations to minimize $C_{BL}$ of BBL-$R_{opt}$:** The BBL-$R_{opt}$ provides good SRAM write/cycle-time metrics but lags in SRAM read/access time and dynamic power due to its relatively higher $C_{BL}$, which can be mitigated by altering the existing iN5 BPR process slightly. An additional DTCO study is performed by varying STI thickness and BBL depth using QuickCap NX. Fig. 18 shows the reduction in $C_{BL}$ when the STI thickness is increased from 70nm (iN5's BPR default) to 100nm (a new BBL-$R_{opt1}$ configuration) without changing the BBL depth of 15 nm (iN5's BPR default). Furthermore, as the BBL depth increases (from the iN5 BPR default value of 15 nm), the $C_{BL}$ initially decreases and achieves a minima around 25nm BBL depth (Fig. 19), while incurring marginal increase in buried via resistance. This optimized BBL configuration (BBL-$R_{opt2}$) achieves the lowest $R_{BL}$ and nearly the lowest $C_{BL}$ (slightly higher than $C_{BL}$ of BBL-$C_{opt}$). BBL-$R_{opt2}$ shows gains of 6-11% in the access time, 23-31% in the write time, and 1-4% in the dynamic power compared to the baseline and the SRAM BPR configurations (Fig. 20 and Table IV).

## IV. Conclusion

To the best of our knowledge this is the first work which explores the use of buried metal for signal routing. The proposed BBL approach provides significant gains over the baseline as well as BPR-SRAM variants requiring minimal process changes to the already demonstrated BPR technology at a sub-5nm process node. Additionally, further DTCO for BPR parameters is also explored to open the possibilities for future optimizations. The proposed BBL-SRAM offers a generational node gain in SRAM's power and performance and is a promising use-case of leveraging buried metals for $RC$-critical signal routing in all sub-5nm technology nodes.

## References


[1] J. Ryckaert et al., 2019 EDTM, pp. 50-52
[2] A. Gupta et al., 2018 IITC, pp. 4-6.
[3] A. Gupta et al. VLSI 2020
[4] D. Prasad et al., 2019 IEDM pp. 19.1.1-19.1.4.
[5] S. M. Salahuddin et al., 2019 EDL, vol. 40, no. 8, pp. 1261-1264.
[6] S. M. Salahuddin et al., 2020 VLSI.
[7] J. Chang et al.,2017 ISSCC, pp. 206-207.
[8] T. Song et al., 2018 ISSCC, pp. 198-200.
[9] synopsys.com/implementation-and-signoff/signoff/quickcap-nx.html

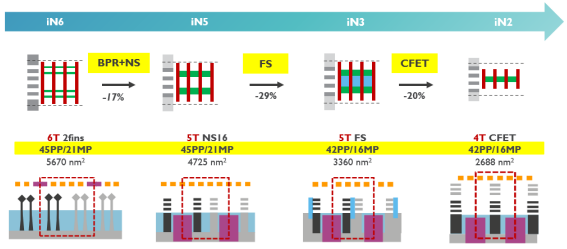


Fig. 1: Technology scaling roadmap from imec showing buried power rails at iN5 (approximately equivalent to foundry 3nm) and beyond.

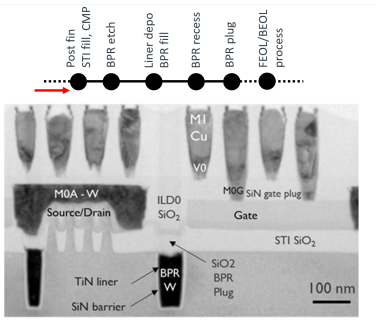


Fig. 2: Cross-section of BPR with transistor front-end and BEOL metal layers [3]. The BPR lines within FEOL have an AR of ∼5 .

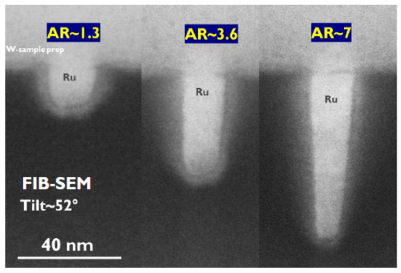


Fig. 3: Focused-ion-beam (FIB) cross-sections of Ruthenium (Ru) lines of different aspect ratios (AR) [2]. Ru lines meet the BPR $R$ target of 50 ohm/um at AR 3 at a width of 33 nm.

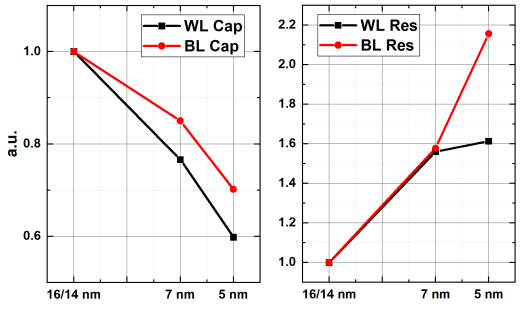


Fig. 4: Resistance and capacitance scaling trends for SRAM wordline and bitline from 16nm to 5nm process node.

| Parameter | Value (nm) |
|---|---|
| Gate pitch | 45 |
| Metal Pitch | 21 |
| Metal (Mx) thickness | 16.5 |
| STI thickness | 70 |
| BPR metal depth | 15 |
| BPR metal width | 21 |
| BPR metal pitch | 84 |
| BPR metal thickness | 147 |

Table I: iN5 default dimensions

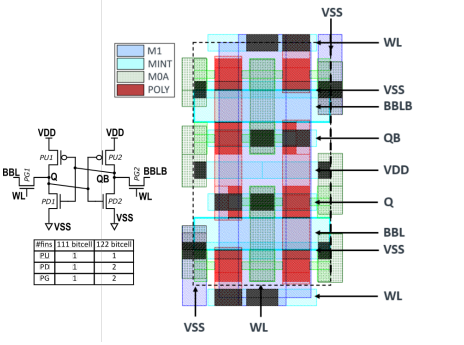


Fig. 5: 6T SRAM bitcell schematics and thin-cell layout view of 111 bitcell with BBL.

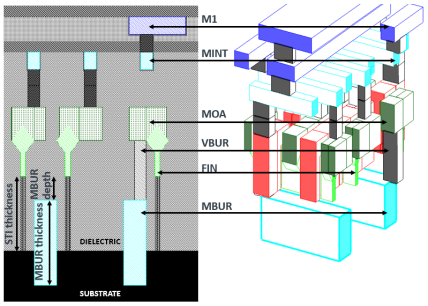


Fig. 6: Cross-sectional and 3D view of 111 bitcell with BBL as seen in QuickCap NX tool.

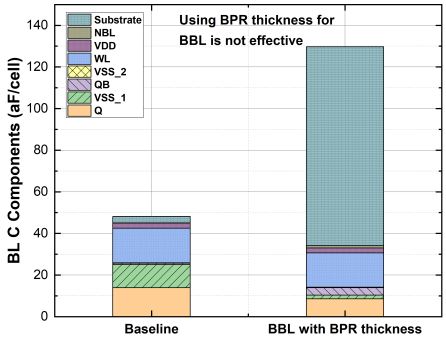


Fig. 7: Comparison of $C_{BL}$ for BBL with default BPR thickness (147nm; AR 7) with baseline SRAM.

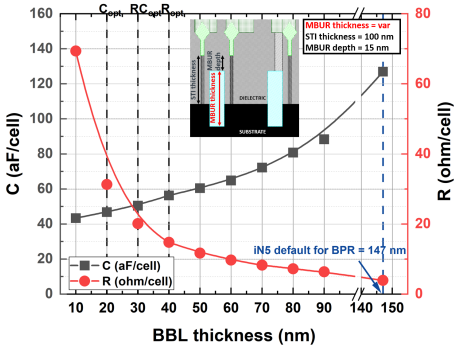


Fig. 8: Impact of thickness of buried metal on the resistance and capacitance of BBL

| Configuration | WL Metal | BL Metal |
|---|---|---|
| Baseline | 31nm M1 + 61nm M2 | 11nm/31nm MINT (111/122) |
| BPR (Buried Power, wider BL/WL [4],[5] | 61nm M1 + 61nm M2 | 31nm/31nm MINT (111/122) |
| BBL Copt | 31nm M1 + 61nm M2 | 21nm MBUR (thickness=20nm) |
| BBL RCopt | 31nm M1 + 61nm M2 | 21nm MBUR (thickness=30nm) |
| BBL Ropt | 31nm M1 + 61nm M2 | 21nm MBUR (thickness=40nm) |

Table II: List of configurations in our experimental setup.

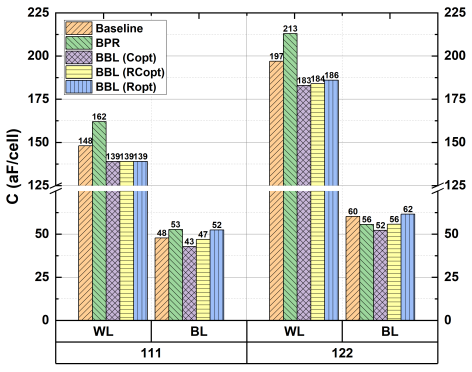


Fig. 9: BL and WL capacitance for different SRAM configurations.

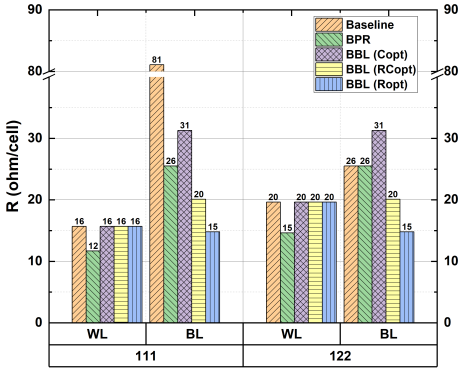


Fig. 10: BL and WL resistance for different SRAM configurations.

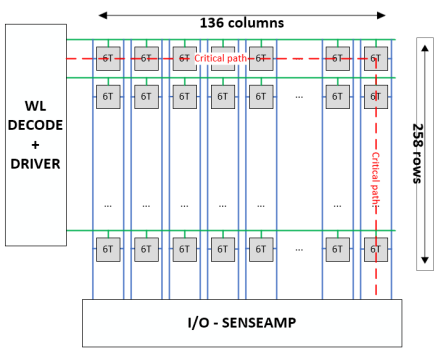


Fig. 11: SRAM sub-array (258 rows x 136 columns) floorplan used in the DTCO analysis.

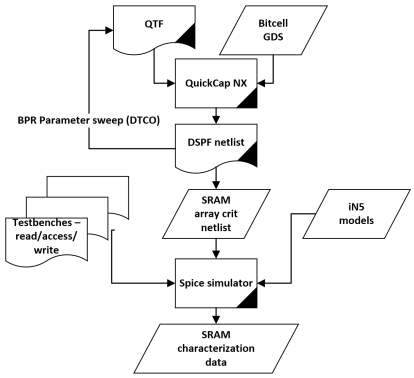


Fig. 12: Extraction and simulation framework. QuickCap Technology File (QTF) supports enhanced geometry description capability and precise silicon profile modeling. DSPF contains detailed network of RC parasitic for every net.

| SRAM Metric | PVT values |
|---|---|
| Read Margin | SS/0.63V/-40C |
| Access Time | SS/0.63V/-40C |
| Write Margin | SF/0.63V/-40C |
| Write Time | SF/0.63V/-40C |
| Dynamic Power | TT/0.7V/25C |

Table III: SRAM metric and the appropriate Process/Voltage/Temperature (PVT) condition used in DTCO analysis.

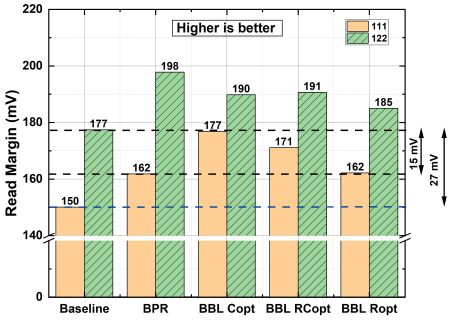


Fig. 13: Comparison of SRAM read margin for different SRAM configurations

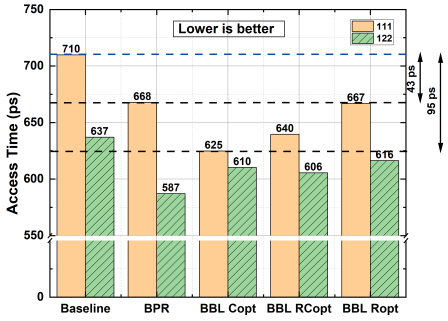


Fig. 14: Comparison of SRAM access time for different SRAM configurations

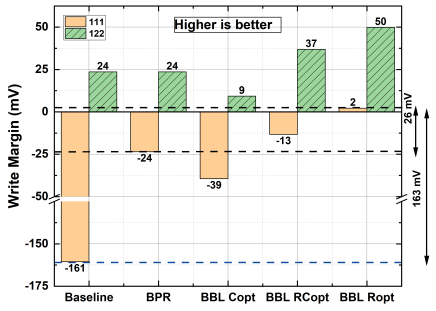


Fig. 15: Comparison of SRAM static write margin for different SRAM configurations.

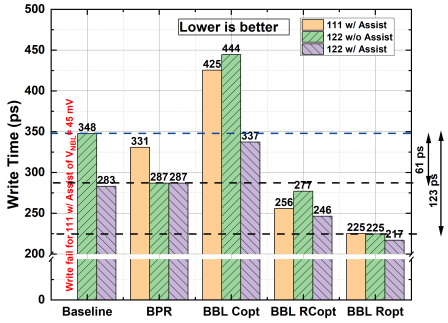


Fig. 16: Comparison of SRAM write time for different SRAM configurations.

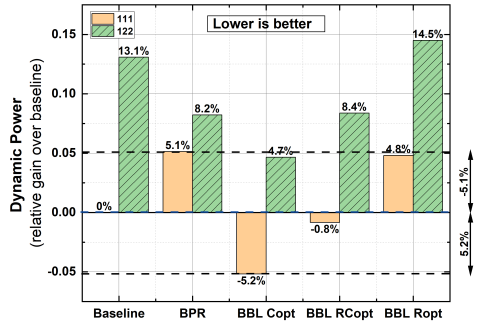


Fig. 17: Comparison of SRAM dynamic power (Read) for different SRAM configurations.

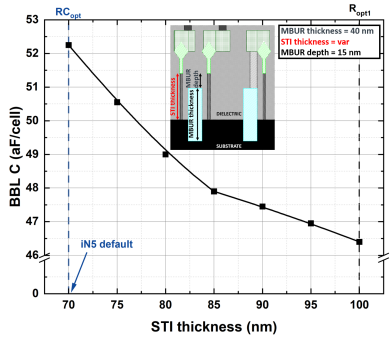


Fig. 18: BBL-Ropt1: $C_{BL}$ improvement with increase in thickness of STI.

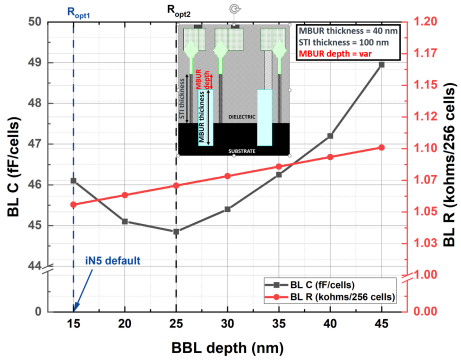


Fig. 19: BBL-Ropt2: $C_{BL}$ improvement by sweeping buried depth and finding optimal point.

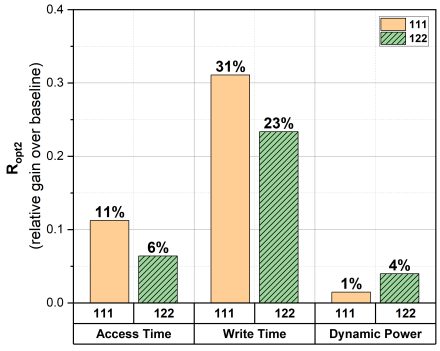


Fig. 20: Relative improvements in SRAM metrics with BPR and selected BBL-Ropt2 over the baseline.

| Metric | Bitcell | BPR | BBL RCopt | BBL Ropt2 |
|---|---|---|---|---|
| Access Time | 111 | 6% | 10% | 11% |
| | 122 | 8% | 5% | 6% |
| Write Time | 111 | -1% | 22% | 31% |
| | 122 | -1% | 13% | 23% |
| Dynamic Power | 111 | -5% | 1% | 1% |
| | 122 | 4% | 4% | 4% |

Table IV: Summary of improvements for BPR and selected BBL configurations relative to baseline SRAM