Beyond-Cu Intermediate-Length Interconnect Exploration for SRAM Application

Zhenlin Pei, *Graduate Student Member, IEEE*, Francky Catthoor, *Fellow, IEEE*, Zsolt Tokei, *Member, IEEE*, and Chenyun Pan, *Senior Member, IEEE*

***Abstract*—Promising interconnect materials continue to emerge and are considered as potential replacements for Cu interconnects. In this paper, an intermediate-length interconnect technology exploration framework for SRAM array-level performance based on CACTI 7 is presented to efficiently optimize various emerging interconnect technologies. Three graphene-based interconnect materials are benchmarked against their traditional Cu counterpart for optimal SRAM array-level performance. Furthermore, we investigate the scalability of the emerging interconnect for the SRAM application under five technology nodes based on ASU Predictive Technology Model for FinFET devices and quantify the impact of different cache sizes on key performance metrics, including energy-delay product and energy-delay-area product. Results demonstrate that the cache using thick graphene interconnects has over 19% EDP reduction comparing the Cu counterpart. It is shown that up to 37% of the energy-delay-area product (EDAP) improvement can be achieved by thick graphene interconnects for 2MB cache size at 7nm.**

***Keywords*—Interconnect, graphene, SRAM, mean free path, delay, energy-delay product, energy-delay-area product, design/technology co-optimization**

I. Introduction

The last decades witnessed the booming performance of VLSI systems in terms of frequency, throughput, and energy efficiency. This exponential improvement is fueled by density, integrated, and performance improvements by scaling FETs and back-end-of-line (BEOL) interconnects. However, as the technology node scales beyond 10 nm, the interconnect has become the dominant factor of the chip performance, and technology research and development escalates rapidly more "interconnect-centric" [1, 2]. The delay and energy dissipation associated with interconnects become major concerns in the VLSI systems due to the large resistivity of the traditional Cu interconnects that are limited by thick barriers and size effects [1, 3, 4]. To address the interconnect challenge, significant research efforts have been focused on local short interconnects, and many new materials have been explored, such as graphene, Ruthenium, and Cobalt [5-11]. At the global level, emerging interconnect technologies, such as optical and ballistic interconnects have been studied to further improve the delay and energy efficiency for long-distance communication [12-16]. However, limited work has been performed for intermediate-length interconnect, and there is an increasing need to investigate novel interconnects for intermediate-length that range from tens to hundreds of micrometers.

In the past decade, significant research activities have been performed on graphene in the electronic device community due to its excellent electronic properties, including a long mean free path (MFP) and large current conduction capability [17]. It has been considered as a potential candidate to replace the Cu interconnect to enable power-efficient computing systems [5]. The challenge of graphene interconnect is associated with its large contact resistance, which limits its usage in the local short interconnects. Due to the fabrication limitation, the small number of graphene layers also constrains its usage in the global long interconnects. As a result, graphene interconnects are promising candidates for intermediate-length interconnects that are overlooked in existing research.

To understand the true advantage of novel interconnect technology, it is important to capture circuit-level perspectives instead of a simple repeater-based driver-receiver model. For example, graphene has the benefit of a small interconnect capacitance, which saves energy thanks to the small geometry and quantum capacitance. However, the resistance of the graphene interconnect can be higher than the Cu if only a small number of graphene layers are used [18]. In addition, performance trade-offs exist in the design of interconnect width because wider interconnects help to reduce latency at the cost of the routing area overhead, which is particularly important for circuit- and system-level performance. As a result, many key material and structural parameters can only be optimized at the higher circuit and system levels.

SRAM is one major component in today's processors and occupies a significant portion of the chip area. An SRAM array contains bitlines, wordlines, and H-trees that span a wide range of widths and lengths from local, intermediate, to global levels, making it an ideal circuit for benchmarking intermediate-length interconnects using emerging technologies. Traditionally, the interconnects for bitlines and wordlines locate at lower metal levels for maximum memory density. As the technology node enters sub-10nm, the delay associated with the large local interconnect resistance motivates the shift of interconnects to upper layers such as the 5th or 6th metal layers. A comprehensive interconnect benchmarking for the

Manuscript received in September 2021. This work was supported by IMEC.

Zhenlin Pei and Chenyun Pan are with the Department of Electrical Engineering, The University of Texas at Arlington, Arlington, TX 76010 USA.

Francky Catthoor and Zsolt Tokei are with the IMEC, 3001 Leuven, Belgium.

memory array-level performance becomes critical to realize the true benefits of graphene interconnects for intermediate-length interconnects, such as bitlines and wordlines. Existing work has investigated graphene-based nanomaterials for crossbar-based memory applications [19-22], but no comprehensive studies have been performed on the SRAM using graphene interconnects.

To capture the electrical performance of the SRAM, the most accurate way is to perform a detailed simulation based on the placement and routing EDA design flow. These simulations, however, are very time-consuming. One simulation with a single set of parameters may take hours to complete. Explorations of different emerging interconnect technologies with multiple material and structural properties over a large design space are extremely inefficient, if not impossible. As a result, compact models are critical to allow an efficient interconnect-circuit co-design and realize the full potential of emerging interconnect technologies.

In this paper, we perform a large design space exploration for graphene-based interconnects based on CACTI 7, a well-known and open-source model for optimizing cache [23]. Such a validated cache simulator allows us to efficiently explore various interconnect configurations with reasonable accuracy at the early design stage. Three interconnect configurations will be investigated to understand the true benefit of graphene-based memory, including the thick graphene, graphene-capped Ruthenium, and graphene-capped Cu interconnects. Key trade-offs among different interconnect parameters are investigated, including the width, number of graphene layers, and effective grain size, to maximize SRAM array-level performance. The comparison among different interconnect configurations and their Cu-based counterparts will provide general guidelines to material technologists and system designers to develop and identify suitable graphene-based interconnects for energy-efficient memory systems. The design methodology presented in this work is generic and can be applicable to a variety of devices and interconnect combinations. The major contributions of this work are highlighted below.

- We develop an efficient interconnect/memory co-design method based on CACTI 7 to enable a large design space exploration to maximize the potential advantage of graphene at the memory-level performance.
- Three emerging and promising graphene-based interconnect materials are benchmarked against their Cu counterparts for optimal SRAM performance metrics.
- We perform the scalability analysis for graphene-based SRAM applications under different cache sizes based on ASU Predictive Technology Model for FinFET devices[24].
- Valuable insights are provided to interconnect technologists and circuit designers to mutually understand the material and process requirements and to design more suitable interconnect materials for memory systems.

II. Modeling Approaches

A. Graphene Interconnect Model

For general graphene-based interconnects, the current is given by the Landauer formula [25] :

$$I = \frac{e}{h}\int_{-\infty}^{+\infty} 2\sum_{n}\frac{MFP}{L+MFP}[f(E-\mu_1)-f(E-\mu_2)]\,dE \quad (1)$$

where L is the interconnect length, e is the elementary charge, h is Planck's constant, E is the energy level, f is the Fermi distribution function, μ_1 and μ_2 are electrical potentials on two sides of the graphene, and MFP is the effective mean free path of graphene. The effective MFP depends on several factors, including the property of substrate material and the graphene edge roughness. In this work, the MFP is fitted based on the mobility extracted from experimental data [26, 27] using the semiclassical equation (2) [28]

$$\sigma = en\mu = \frac{2e^2}{h}\left(\sqrt{\pi n}\cdot MFP\right) \quad (2)$$

where σ is the conductivity, n is carrier density, and μ is mobility. MFP is set as $460nm$ at $n \approx 4.85\times10^{10}cm^{-2}$ based on existing work [29]. From the fitted MFP shown in Fig. 1, the MFP scales down with the width due to the edge scattering. With the assumption of side contacts, the graphene contact resistance is obtained as $R_{con}/(WN_{layer})$, where R_{con} is the contact resistance, N_{layer} is the number of graphene layers, and W is the interconnect width. The value of contact resistance is assumed to be 100 Ω·μm based on recent experimental work [30].

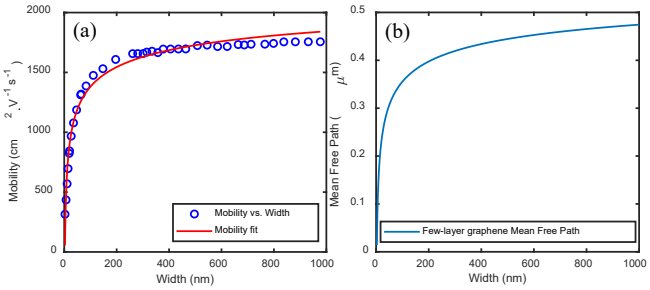

Fig. 1. (a) Few-layer graphene mobility fitting model as the function of width [26], and (b) effective mean free path versus width.

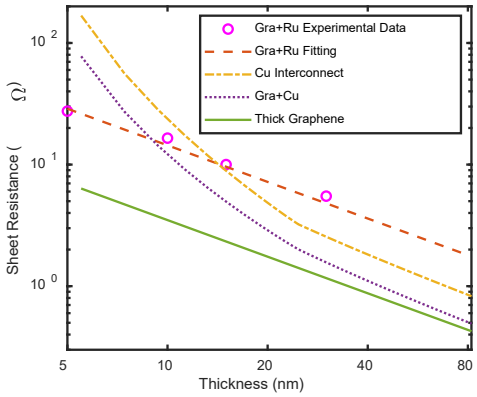

Fig. 2. Comparison of the sheet resistance versus the thickness for four interconnect materials. The inverse function is used to fit the sheet resistance for the graphene-capped Ru based on the experimental data [31]. The aspect ratio is assumed to be 2 for all interconnects, and the effective grain size factor is 3 for graphene-capped Cu interconnect.

In this work, four types of materials are investigated to quantify their impacts on the SRAM array-level performance, including (1) the Cu interconnect as the baseline, (2) the graphene-capped Ruthenium, (3) the graphene-capped Cu, and

(4) the thick graphene interconnects. Here, Ru is considered as one potential option because it has received much research attention, and it has **(i)** competitive resistivity to Cu in fine dimensions and **(ii)** the potential for higher electromigration reliability than Cu [29, 31-33]. The Cu resistivity model follows existing work with a grain boundary reflectivity of 0.5 and a side wall specularity of 0.5 [34]. For graphene-capped Cu interconnect, the electron scatters less frequently inside Cu, and we use 3× of the grain size to capture this effect based on the experimental work [35, 36]. For the graphene-capped Ruthenium, experimental data is used to quantify its performance under different thicknesses [31]. The comparison of sheet resistance and resistance per unit length for four materials is shown in Fig. 2 and Fig. 3.

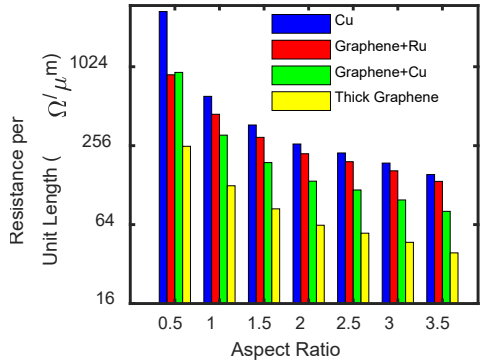

Fig. 3. Resistance per unit length versus aspect ratio for four materials with a interconnect width of 18nm at the 7nm technology node.

The capacitance per unit length of the interconnect follows the model in CACTI, which is based on vertical, horizontal, and fringe capacitance, shown in (3).

$$\begin{aligned} C_{wire/length} &= C_{vertical} + C_{sidewall} + C_{fringe} \\ &= 2 \cdot \varepsilon_V \cdot \frac{W_{wire}}{T_{ild}} + 2 \cdot Miller \cdot \varepsilon_H \\ &\cdot \frac{T_{wire}}{Spacing_{wire}} + C_{fringe} \end{aligned} \tag{3}$$

where ε_V is the vertical dielectric constant of 3.9, ε_H is the horizontal dielectric constant of 2.11~2.18 that scales with the technology node, T_{ild} is the interlayer dielectric thickness which is assumed to be the wire thickness, C_{fringe} is the fringe capacitance of $0.115fF/\mu m$, and $Miller$ is assumed to be 1.5, which is used to capture the Miller effect during the switching activity.

B. Device Models

For the case study, CMOS FinFET high-performance and low-power device models are adopted from ASU Predictive Technology Model for FinFET 7, 10, 14, 16, and 20 nm technology nodes [24]. Device parameters, including drain capacitance, gate capacitance, on current, temperature-dependent leakage current, supply voltage, and threshold voltage are extracted by Synopsys HSPICE simulations [37].

C. Memory Models

The SRAM memory models are adopted based on CACTI, a well-known and open-source simulator for optimizing memory [23, 38]. CACTI sweeps the memory organization parameters to obtain optimal memory parameters based on the target metrics, which can be defined by the user. CACTI considers comparator, Mux driver, output driver, decoder, column Mux, sense amps, wordline, bitline, tag array, and data array in the logical and physical organization of the cache.

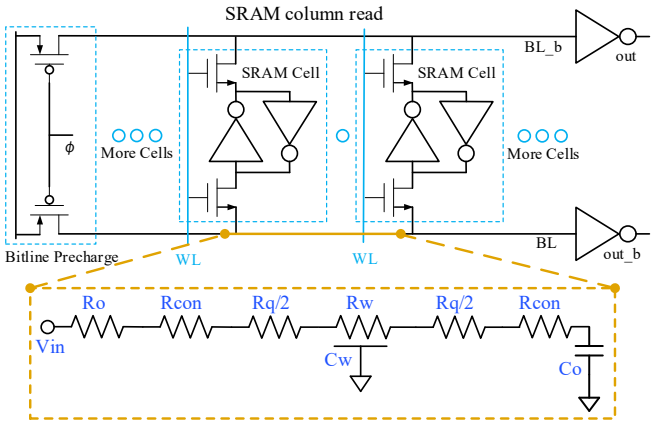

Fig. 4. SRAM cells in a column with equivalent circuit model for a graphene bitline, where C_0 and R_0 are the device input capacitance and output resistance, respectively, C_w and R_w are the interconnect capacitance and resistance, respectively, R_q is the quantum resistance, and R_{con} is the contact resistance.

To model the delay for generic interconnects, an equivalent circuit model for the SRAM bitline is shown in Fig. 4. The critical path for an array access time is mainly divided into three parts: H-tree outside/inside the bank and path inside the mat. CACTI has been validated based on SPICE and published data of commercial cache, including Sun SPARC 90nm L2 cache and Intel 65nm L3 cache [38]. Such a validated cache simulator allows us to efficiently explore various interconnect configurations with reasonable accuracy at the early design stage. Key trade-offs among different interconnect parameters are investigated, including the width, number of graphene layers, aspect ratio, and effective grain size, to maximize array-level SRAM performance. The comparison among different interconnect configurations and their Cu-based counterparts will provide general guidelines to material technologists and system designers to develop and identify suitable graphene-based interconnects for energy-efficient memory systems.

III. SIMULATION RESULTS

Based on the modeling approaches in Section II, the performance analyses are performed at the cache level. Five technology nodes (7, 10, 14, 16, and 20nm), four interconnect materials (Cu, graphene-capped Ru, graphene-capped Cu, and thick graphene interconnects), and three cache sizes (128KB, 2MB, and 32MB) are investigated in the case study.

A. Impact of Interconnect Geometry on Cache Performance

One of the key advantages of graphene-based interconnect is its small resistance thanks to the large MFP. Since the interconnect resistance is highly dependent on its geometry, we analyze the impact of aspect ratio as well as the width of interconnect on SRAM array-level performance. Under different width assumptions, the energy and access time breakdown bar charts for different aspect ratios are shown in Fig. 5. Note that due to the large span in different energy and access time components, the y axis is shown in the log scale. The total energy of the cache is mainly consumed by four parts,

namely the decoder, the wordline, the bitline, and the H-tree, where H-tree dominates the overall energy due to the large interconnect length. In Fig. 5, the access time decreases with the aspect ratio due to the increasing cross-section area and decreasing resistance. Meanwhile, the total energy increases because the energy associated with the interconnect capacitance increases. As a result, a clear trade-off between energy and access time can be observed in Fig. 6 (a). In short, the cache performance is either **(i)** limited by the energy dissipation if the aspect ratio is too large or **(ii)** limited by the access time if the aspect is too small. From Fig. 6 (b), an optimal aspect ratio can be observed to minimize the energy-delay product (EDP) for a given width. The EDP versus the aspect ratio for each width scaling factor is shown in Fig. 6 (b), where an optimal aspect ratio exists to minimize the overall EDP.

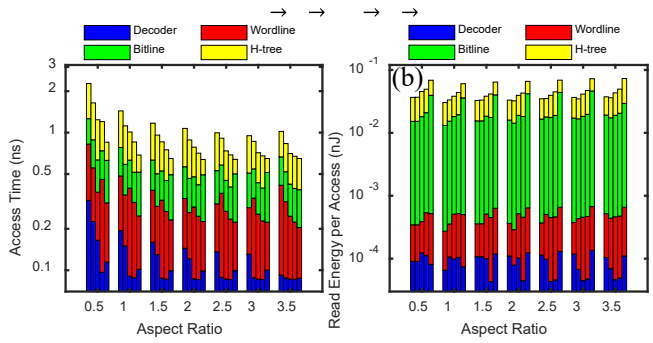

Fig. 5. (a) Access time and (b) read energy per access breakdown bar chart versus aspect ratio of four interconnect materials for a cache size of 2MB at the 7nm technology node.

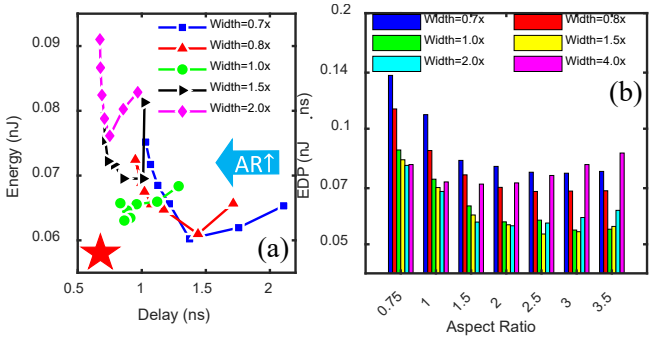

Fig. 6. (a) Energy versus delay and (b) EDP versus aspect ratio for a cache size of 2MB using thick graphene interconnect under width scaling factor of 0.7, 0.8, 1, 1.5, 2, and 4 at the 7nm technology node. The red star shows the preferred corner.

In terms of the width scaling factor, a larger width helps to further reduce the interconnect resistance, leading to a better access time. However, increasing the width leads to a larger cell area and hence increases the interconnect length, leading to a diminishing return in the access time improvement, shown in Fig. 5. Meanwhile, the energy keeps increasing due to the longer interconnects. As a result, within each aspect ratio, an optimal width also can be observed to minimize EDP, as shown in Fig. 6 (b). Since cell density is one important metric of the SRAM application, we investigate the impact of interconnect geometry on energy-delay-area product, as shown in Fig. 7. With the consideration of area, optimal aspect ratio and width also exist, and the cache prefers to use a smaller width compared to the design for EDP minimization.

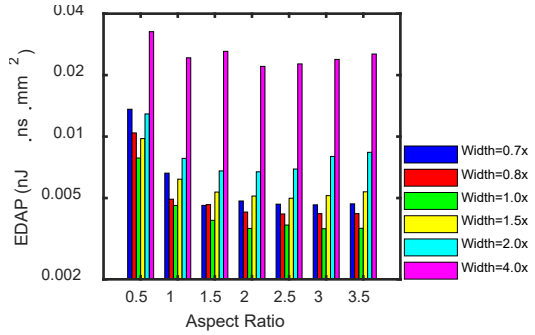

Fig. 7. EDAP versus aspect ratio for SRAM using thick graphene interconnects under the width scaling factor of 0.7, 0.8, 1, 1.5, 2, and 4 versus aspect ratio for a cache size of 2MB at the 7nm technology node.

B. Impact of Material on Cache Performance

To quantify the impact of interconnect materials on the cache-level performance, the access time and energy per access time as a function of the width scaling factor for four types of materials are shown in Fig. 8 (a) and (b), respectively. Each bar in the figure is based on the optimal aspect ratio. Thick graphene is the best material for the access time due to the small resistance as shown in Fig. 2. Note that the y-axis is in log scale due to the large span in access time and energy.

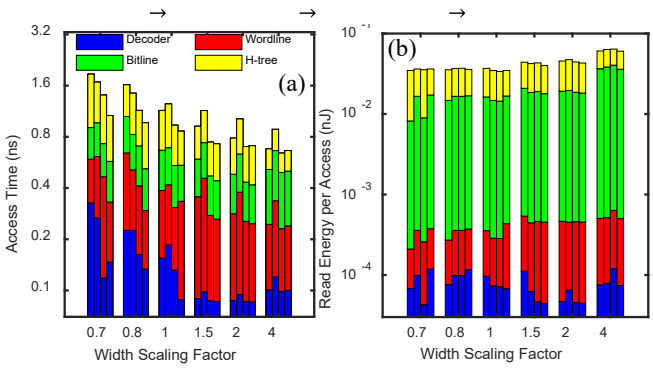

Fig. 8. (a) Access time and (b) read energy per access breakdown bar chart versus the width scaling factor with the optimal aspect ratio for minimizing EDP for a cache size of 2MB at the 7nm technology node.

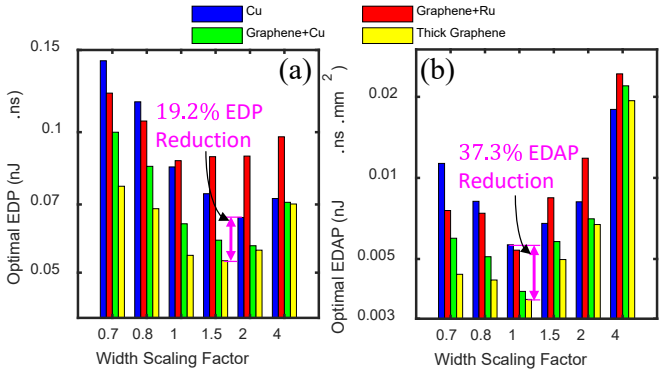

Fig. 9. Optimal (a) EDP and (b) EDAP of cache using four interconnect materials for a cache size of 2MB under width scaling factor of 0.7, 0.8, 1, 1.5, 2, and 4 at the 7nm technology node.

Comparing to the Cu counterpart, up to 19% and 37% reduction can be observed for the cache using thick graphene interconnects in terms of EDP and EDAP, respectively, as shown in Fig. 9. The performance improvement benefits mainly come from the superior conductivity of graphene, especially for minimizing EDAP, where a relatively small width is used.

C. *Impact of Technology Node on Cache Performance*

To quantify the impact of technology node on the performance of cache using thick graphene, the optimal access time breakdown bar chart under five technology nodes is demonstrated in Fig. 10. Each point in the figure is obtained based on the optimal aspect ratio for minimizing EDP as discussed in Section III A.

As the technology node scales down, the access time improves due to the small footprint area and short interconnects, such as bitline, wordline, and H-tree. Compared to the Cu-based cache, the one using thick graphene can provide over 40% reduction in access time under the nominal width at the 7nm technology node. A large reduction of the access time can be observed at a small technology node, as illustrated in Fig. 10 (b). In addition, the access time improvement using graphene increases as the width scales down. This is because Cu interconnect suffers from a significant size effect due to the electron scattering and a thick barrier, which increase the interconnect resistance substantially at a small dimension.

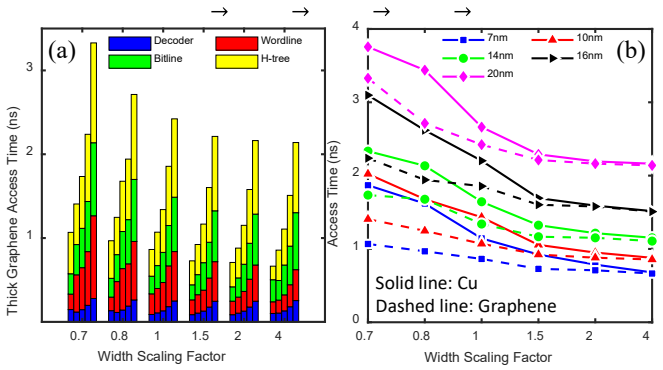

Fig. 10. (a) Access time breakdown bar chart of an SRAM using thick graphene interconnects and (b) comparison of the access time for Cu and thick graphene interconnects versus width scaling factor for a cache size of 2MB at 7, 10, 14, 16, and 20nm technology nodes.

In Fig. 11 (a), the read energy breakdown bar chart is illustrated as a function of width for five technology nodes. As the technology scales down, the short interconnects and small devices lead to a reduction in capacitance, hence small read energy. In addition, the read energy increases with the interconnect width because of the larger area and longer interconnects that need to be switched. By sweeping the width scaling factor, a clear trade-off between energy per access and access time can be observed in Fig. 11 (b). In short, the cache performance is either **(i)** limited by the energy dissipation if the width is too large or **(ii)** limited by the access time if the width is too small. Compared to the Cu-based SRAM, up to 19% EDP reduction can be observed for the one using thick graphene interconnects at the 7nm technology node, as shown in Fig. 11 (b).

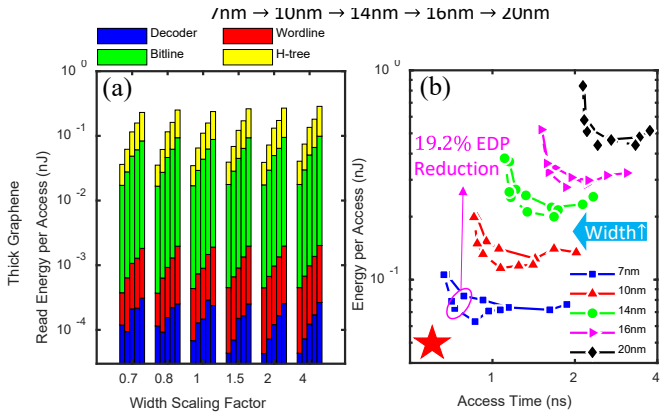

Fig. 11. (a) Read energy per access breakdown bar chart of an SRAM using thick graphene interconnects and (b) comparison of energy versus access time of cache using Cu and thick graphene interconnects for a cache size of 2MB at 7, 10, 14, 16, and 20nm technology nodes. Solid and dashed lines are for SRAMs using Cu and thick graphene interconnects, respectively. The red star shows the preferred corner.

D. *Impact of Cache Size on Improvement*

The simulations performed in previous subsections are based on a cache size of 2MB. To capture the impact of cache size at the SRAM array-level performance, Fig. 12 shows the EDP and EDAP reduction for the SRAM using thick graphene interconnects comparing to the Cu counterpart under three cache sizes. In general, a larger cache gains more benefits by using graphene interconnects, especially at a small technology node. This is because the performance of a large cache is more dominant by longer interconnects due to the large area. The narrow interconnect also takes more advantage of the small resistance of graphene-based interconnect compared to its Cu counterpart. For a small cache size of 128KB, the EDP reduction saturates for small technology nodes. This is mainly due to the fact that (i) the interconnects are relatively short and limited by the contact resistance of graphene and (ii) the device resistance dominates the delay at a small cache size. For example, the interconnect resistance-related delay (e.g. R_wC_w and R_wC_o) are ~25% of the total delay at the 7nm and 10nm technology nodes, and the rest of the delay is determined by the device resistance. As a result, the difference of EDP reduction between 7nm and 10nm is small.

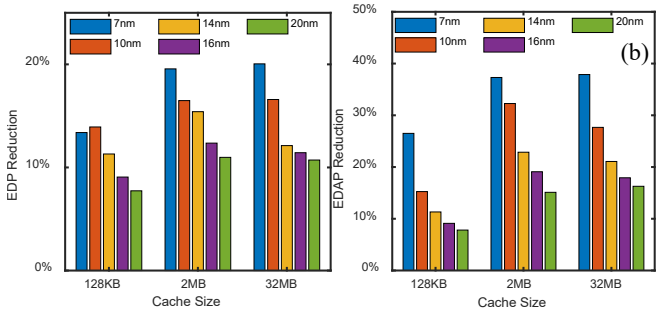

Fig. 12. (a) EDP and (b) EDAP reduction of the SRAM using thick graphene interconnects compared to Cu-based SRAM versus cache size at 7, 10, 14, 16, and 20nm technology nodes.

Compared to the EDP improvement, the EDAP reduction is more substantial, as shown in Fig. 12 (b). This is because when the area is taken into the consideration, the cache prefers to use a small interconnect width, which better takes advantage of the small resistance of graphene.

IV. Conclusion

In this paper, we quantify the benefits of three graphene-based interconnect configurations for SRAM array-level performance, including graphene-capped Ru, graphene-copped Cu, and thick graphene, based on the CACTI simulator. Key interconnect parameters, such as width, aspect ratio, and effective grain size, are investigated to minimize various performance targets, including EDP and EDAP. Under the optimal width and aspect ratio, thick graphene provides the best performance thanks to its small resistance compared to its counterparts with up to 19% and 37% reduction in EDP and EDAP, respectively, compared to the Cu counterparts. Compared to the optimal width to minimize EDP, the optimal width for minimizing EDAP decreases due to the area overhead. Graphene-capped Cu also performs better compared to graphene-capped Ru and Cu thanks to its larger effective grain size based on the experimental observation. Graphene-capped Ru underperforms its counterparts mainly due to the large resistance that is observed in the experiment. When it comes to the impact of cache size on performance, a larger cache leads to more EDP and EDAP reduction for sub-10nm due to the longer interconnects, such as H-tree and bitlines.

References

[1] G. Bonilla, N. Lanzillo, C.-K. Hu, C. Penny, and A. Kumar, "Interconnect scaling challenges, and opportunities to enable system-level performance beyond 30 nm pitch," in *2020 IEEE International Electron Devices Meeting (IEDM)*, 2020, pp. 20.4. 1-20.4. 4. DOI: 10.1109/IEDM13553.2020.9372093

[2] J. D. Meindl, "Beyond Moore's law: The interconnect era," *Computing in Science & Engineering,* vol. 5, pp. 20-24, 2003. DOI: 10.1109/MCISE.2003.1166548

[3] R. Brain, "Interconnect scaling: Challenges and opportunities," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 9.3. 1-9.3. 4. DOI: 10.1109/IEDM.2016.7838381

[4] D. Prasad, A. Ceyhan, C. Pan, and A. Naeemi, "Adapting interconnect technology to multigate transistors for optimum performance," *IEEE Transactions on Electron Devices,* vol. 62, pp. 3938-3944, 2015. DOI: 10.1109/TED.2015.2487888

[5] R. Chen, J. Liang, J. Lee, V. P. Georgiev, R. Ramos, H. Okuno, D. Kalita, Y. Cheng, L. Zhang, and R. R. Pandey, "Variability study of MWCNT local interconnects considering defects and contact resistances—Part I: pristine MWCNT," *IEEE Transactions on Electron Devices,* vol. 65, pp. 4955-4962, 2018. DOI: 10.1109/TED.2018.2868421

[6] J. Jiang, J. H. Chu, and K. Banerjee, "CMOS-compatible doped-multilayer-graphene interconnects for next-generation VLSI," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, pp. 34.5. 1-34.5. 4. DOI: 10.1109/IEDM.2018.8614535

[7] F. Griggio, J. Palmer, F. Pan, N. Toledo, A. Schmitz, I. Tsameret, R. Kasim, G. Leatherman, J. Hicks, and A. Madhavan, "Reliability of dual-damascene local interconnects featuring cobalt on 10 nm logic technology," in *2018 IEEE International Reliability Physics Symposium (IRPS)*, 2018, pp. 6E. 3-1-6E. 3-5. DOI: 10.1109/IRPS.2018.8353641

[8] V. Huang, D. Shim, H. Simka, and A. Naeemi, "From Interconnect Materials and Processes to Chip Level Performance: Modeling and Design for Conventional and Exploratory Concepts," in *2020 IEEE International Electron Devices Meeting (IEDM)*, 2020, pp. 32.6. 1-32.6. 4. DOI: 10.1109/IEDM13553.2020.9371945

[9] S. Dutta, S. Kundu, A. Gupta, G. Jamieson, J. F. G. Granados, J. Bömmels, C. J. Wilson, Z. Tőkei, and C. Adelmann, "Highly scaled ruthenium interconnects," *IEEE Electron Device Letters,* vol. 38, pp. 949-951, 2017. DOI: 10.1109/LED.2017.2709248

[10] O. V. Pedreira, K. Croes, A. Leśniewska, C. Wu, M. Van Der Veen, J. De Messemaeker, K. Vandersmissen, N. Jourdan, L. Wen, and C. Adelmann, "Reliability study on cobalt and ruthenium as alternative metals for advanced interconnects," in *2017 IEEE International Reliability Physics Symposium (IRPS)*, 2017, pp. 6B-2.1-6B-2.8. DOI: 10.1109/IRPS.2017.7936340

[11] M. H. van der Veen, N. Heyler, O. V. Pedreira, I. Ciofi, S. Decoster, V. V. Gonzalez, N. Jourdan, H. Struyf, K. Croes, C. Wilson, and Z. Tokei, "Damascene benchmark of Ru, Co and Cu in scaled dimensions," in *2018 IEEE International Interconnect Technology Conference (IITC)*, 2018, pp. 172-174. DOI: 10.1109/IITC.2018.8430407

[12] L. Yang, T. Zhou, H. Jia, S. Yang, J. Ding, X. Fu, and L. Zhang, "General architectures for on-chip optical space and mode switching," *Optica,* vol. 5, pp. 180-187, 2018. DOI: 10.1364/OPTICA.5.000180

[13] H. Gu, K. Chen, Y. Yang, Z. Chen, and B. Zhang, "MRONoC: A low latency and energy efficient on chip optical interconnect architecture," *IEEE Photonics Journal,* vol. 9, pp. 1-12, 2017. DOI: 10.1109/JPHOT.2017.2651586

[14] J. Bashir, E. Peter, and S. R. Sarangi, "BigBus: A scalable optical interconnect," *ACM Journal on Emerging Technologies in Computing Systems (JETC),* vol. 15, pp. 1-24, 2019. DOI: 10.1145/3289391

[15] D. Choi and K. Barmak, "On the potential of tungsten as next-generation semiconductor interconnects," *Electronic Materials Letters,* vol. 13, pp. 449-456, 2017. DOI: 10.1007/s13391-017-1610-5

[16] H. Sharma and K. S. Sandha, "Thermally aware modeling and performance analysis of MLGNR as on-chip VLSI interconnect material," *Journal of Electronic Materials,* vol. 48, pp. 4902-4912, 2019. DOI: 10.1007/s11664-019-07281-9

[17] S. Hu, J. Chen, N. Yang, and B. Li, "Thermal transport in graphene with defect and doping: phonon modes analysis," *Carbon,* vol. 116, pp. 139-144, 2017. DOI: 10.1016/j.carbon.2017.01.089

[18] C. Pan, P. Raghavan, A. Ceyhan, F. Catthoor, Z. Tokei, and A. Naeemi, "Technology/circuit/system co-optimization and benchmarking for multilayer graphene interconnects at sub-10-nm technology node," *IEEE Transactions on Electron Devices,* vol. 62, pp. 1530-1536, 2015.

[19] B. Alimkhanuly, J. Sohn, I.-J. Chang, and S. Lee, "Graphene-based 3D XNOR-VRRAM with ternary precision for neuromorphic computing," *npj 2D Materials and Applications,* vol. 5, pp. 1-10, 2021. DOI: https://doi.org/10.1038/s41699-021-00236-x

[20] J. Liang, S. Yeh, S. S. Wong, and H.-S. P. Wong, "Effect of wordline/bitline scaling on the performance, energy consumption, and reliability of cross-point memory array," *ACM Journal on Emerging Technologies in Computing Systems (JETC),* vol. 9, pp. 1-14, 2013. DOI: https://doi.org/10.1145/2422094.2422103

[21] J. Liang, S. Yeh, S. S. Wong, and H.-S. P. Wong, "Scaling challenges for the cross-point resistive memory array to sub-10nm node-an interconnect perspective," in *2012 4th IEEE International Memory Workshop*, 2012, pp. 1-4. DOI: 10.1109/IMW.2012.6213650

[22] S. Sun and D. Jiao, "Multiphysics Modeling and Simulation of 3-D Cu–Graphene Hybrid Nanointerconnects," *IEEE Transactions on Microwave Theory and Techniques,* vol. 68, pp. 490-500, 2020. DOI: 10.1109/TMTT.2019.2955123

[23] R. Balasubramonian, A. B. Kahng, N. Muralimanohar, A. Shafiee, and V. Srinivas, "CACTI 7: New tools for interconnect exploration in innovative off-chip memories," *ACM Transactions on Architecture and Code Optimization (TACO),* vol. 14, pp. 1-25, 2017. DOI: https://doi.org/10.1145/3085572

[24] Predictive Technology Model (PTM), available online at http://ptm.asu.edu, 2012.

[25] S. Datta, *Quantum transport: atom to transistor*: Cambridge University Press, 2005.

[26] A. Contino, I. Ciofi, R. Baert, X. Wu, I. Asselberghs, U. Celano, C. J. Wilson, Z. Tokei, G. Groeseneken, and B. Soree, "Circuit Delay and Power Benchmark of Graphene against Cu Interconnects," presented at the IEEE International Interconnect Technology Conference (IITC), Brussels, Belgium, 2019.

[27] K. I. Bolotin, K. J. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. e. Hone, P. Kim, and H. Stormer, "Ultrahigh electron mobility in suspended graphene," *Solid state communications,* vol. 146, pp. 351-355, 2008. DOI: https://doi.org/10.1016/j.ssc.2008.02.024

[28] K. I. Bolotin, K. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, and H. Stormer, "Ultrahigh electron mobility in suspended graphene," *Solid State Communications,* vol. 146, pp. 351-355, 2008. DOI: 10.1016/j.ssc.2008.02.024

[29] S. Achra, X. Wu, V. Trepalin, T. Nuytten, J. Ludwig, S. Brems, V. Afanas'ev, C. Huyghebaert, B. Soree, M. Heyns, I. Asselberghs, and Z. Tőkei, "Characterization of Interface Interactions between Graphene and Ruthenium," presented at the IEEE International Interconnect Technology Conference (IITC), San Jose, California, USA, 2020. DOI: 10.1109/IITC47697.2020.9515595

[30] W. S. Leong, H. Gong, and J. T. Thong, "Low-contact-resistance graphene devices with nickel-etched-graphene contacts," *ACS nano,* vol. 8, pp. 994-1001, 2014. DOI: https://doi.org/10.1021/nn405834b

[31] S. Achra, I. Asselberghs, X. Wu, S. Brems, C. Huyghebaert, B. Sorée, M. Heyns, and Z. Tokei, "Graphene-Ruthenium hybrid interconnects," presented at the IEEE International Interconnect Technology Conference (IITC), Brussels, Belgium, 2019.

[32] T. Nogami, "Overview of interconnect technology for 7nm node and beyond-New materials and technologies to extend Cu and to enable alternative conductors," in *2019 Electron Devices Technology and Manufacturing Conference (EDTM)*, 2019, pp. 38-40. DOI: 10.1109/EDTM.2019.8731225

[33] X. Zhang, H. Huang, R. Patlolla, W. Wang, F. W. Mont, J. Li, C.-K. Hu, E. G. Liniger, P. S. McLaughlin, and C. Labelle, "Ruthenium interconnect resistivity and reliability at 48 nm pitch," in *2016 IEEE International Interconnect Technology Conference/Advanced Metallization Conference (IITC/AMC)*, 2016, pp. 31-33. DOI: 10.1109/IITC-AMC.2016.7507650

[34] C. Pan and A. Naeemi, "A Proposal for a Novel Hybrid Interconnect Technology for the End of Roadmap," *Electron Device Letters, IEEE,* vol. 35, pp. 250-252, 2014. DOI: 10.1109/LED.2013.2291783

[35] H. C. Lee, M. Jo, H. Lim, M. S. Yoo, E. Lee, N. N. Nguyen, S. Y. Han, and K. Cho, "Toward near-bulk resistivity of Cu for next-generation nano-interconnects: Graphene-coated Cu," *Carbon,* vol. 149, pp. 656-663, 2019. DOI: https://doi.org/10.1016/j.carbon.2019.04.101

[36] S. Achra, X. Wu, V. Trepalin, T. Nuytten, J. Ludwig, V. Afanas' ev, S. Brems, B. Sorée, Z. Tokei, and M. Heyns, "Metal induced charge transfer doping in graphene-ruthenium hybrid interconnects," *Carbon,* vol. 183, pp. 999-1011, 2021. DOI: https://doi.org/10.1016/j.carbon.2021.07.070

[37] HSPICE, *Synopsys, Mountain View, CA, USA,* 2021.

[38] S. Thoziyoor, N. Muralimanohar, J. H. Ahn, and N. P. Jouppi, "CACTI 5.1," HP Laboratories, Palo Alto. 2008.