# A New Visualization Framework for NVMExplorer to Compare the Performance of Memory Technologies

## Abstract

*This document presents an interactive visualization framework based in Jupyter Notebook and made with python to for guided data analysis on the performance of embedded non-volatile memory technologies, leveraging data generated by the NVMExplorer framework.*

## 1. Introduction

With exponential growth in the computer science landscape, data-intensive such as DNNs (deep neural networks) or graph processing have become to much of a burden on the memory technologies we use. Prior work suggests that CMOS-compatible eNVMs (embedded non-volatile memories) can be specialized as effective memory resources to optimize storage efficiency for data-intensive workloads. Architectures leveraging eNVMs can obviate the need for constant accesses to off-chip DRAM, saving otherwise overwhelming energy cost, and offering better power efficiency than SRAM for on-chip memory resources.

Modern research focuses on designing these state-of-the-art technologies towards specific workload patterns and application use cases based on their properties, as using one technology solution over another may create immense trade-offs. Prior simulation tools aim to empower users to compare performance and potential of these memory technologies for distinct traffic patterns and quantify such trade-offs; `NVMExplorer` is an open source design space exploration (DSE) framework which, based on circuit and device parameters, system constraints, and application level behavior, provides statistics on the different memory cell configurations' performance [2]. While the scope and capabilities of these tools are valuable, existing data visualizations allow comparison across many metrics, but without guidance or available tutorial materials for a new user to contextualize or refine results.

For this reason, this work develops and presents a guided experimentation and data visualization tool to confidently compare eNVM potential across use cases in a user-focused way. The presented work, which we call "NVMExplorer: Story Mode" makes it easier for the user to digest the results of their cross-computing-stack design studies, including guided reading of data across key metrics and filtering data according to user's optimization priorities. Using `NVMExplorer: Story Mode`, a system designer considering different eNVM proposals would be able to determine which one is preferable when running their application based on speed and energy efficacy.

## 2. NVMExplorer: Story Mode

### 2.1. Background Tools

The new visualization tool, `NVMExplorer: Story Mode`, presents the essential data reported by `NVMExplorer` design studies in a clear and coherent manner. This platform utilizes pandas, numpy, ipywidgets and matplotlib libraries for Python and Jupyter Notebook, and uses both guiding text and visualizations to explain to new users what the possible variables affecting performance are, and illustrate their manifestation for each memory technology and design configuration. To make it accessible to the user, the entire jupyter notebook and example data for a guided tutorial will be made available open-source, with options for users to add file names to load their own data.

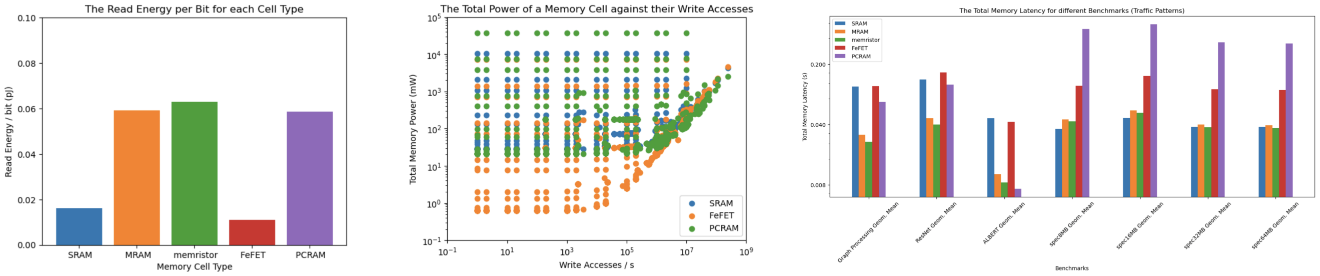
### 2.2. Custom, Interactive Filtering

Additionally, our application uses interactive toggles to make each graph adaptable to what the user desires to see, including toggles for all of the memory cell types and other configuration parameters, so that the user can choose when to see results corresponding to each cell dynamically. Similarly, there are toggles for the graph scales; some statistics are clearly understood by analyzing their orders of magnitude instead, like the latency in seconds or the power in picojoules, so using a turning on the log scale might shed light to new data interpretations. Finally, the scatter plots also include sliders to adjust the range of either axis. Again, this allows the user to customize by zooming and filtering the exact data that they want to investigate.

### 2.3. A Cross-Stack Narrative

A major accomplishment of this framework is the guided narration order that it follows. With `NVMExplorer: Story Mode`, we start our investigation by analyzing eNVM characteristics at the device level. The provided materials guide a user through comparing the characteristics of the study's cell types. It explores statistics such as the cell area, area efficiency (proportion of cell area in the memory array, rather than periphery circuitry) [1], the read and write energy per bit, and potential endurance. Looking at these architecture specific statistics helps paint the picture on each cell's nature, and may provide reasons on their array-level or application performance.

Second, the user delves into more array level statistics and generic traffic. This refers to when the whole memory array works together on any type of workload running on a computer. This means that for each memory configuration,



(a) Chapter 1: Read Energy per Bit against Memory Cell Type

(b) Chapter 2: Total Memory Power Against Write Accesses

(c) Chapter 3: Mean Memory Latency for each cell under the simulated traffic of different applications.

Figure 1: Visualizations from the cross-stack levels comprising each chapter produced by `NVMExplorer: Story Mode`

there are unique power and latency results for different traffic patterns, which are defined by a number of read and write accesses per second of operation. Hence, this becomes a dense design space cluttered with a range of potential memory performance, and comparing in a methodical, customizable way with provided toggles incredibly aids the user to narrow in on a range of traffic or limited budget for allowable power or latency.

The exploration concludes with looking at how results change based on the behavior and traffic patterns of specific software applications running on a chosen architecture. First, the tool presents how read and writes vary per benchmark type to give an idea of what running a certain application means in terms of data movement.`NVMExplorer` supports certain application use cases natively: DNNs, graph processing, and SPEC. As a driving example to demonstrate the improved user experience and accelerated data analysis of our proposed tool, we consider image classification tasks using ResNet, and natural language processing tasks using ALBERT. An additional critical metric to consider at this application-driven stage is the potentially limited life expectancy of the cells, which uses endurance data to extrapolate how long the eNVM array could functionally serve the system before degrading.

Although this organization is encouraged for comprehension, each user is also free to research in their own fashion. Skipping the first part could be helpful for a person knowledgeable in eNVMs for example. In addition, if they want to look at a statistic that is not hard-coded, the notebook makes it easy for you to replace the variable name once from the code, and a similar graph will show with the same tailoring settings.

## 3. Evaluation

The `NVMExplorer: Story Mode` visualization tool allows for a thorough analysis of memory technologies. The real time graph options allow the viewer to lend focus to specific data points or ranges, while the cross-stack organization contributes to an evolving understanding of the memory cell performance at run time; both of these properties permit a plethora of research questions to be answered. Some questions could regard the latency that your program will concur. For example, you can investigate which cell configuration or combination will make ResNet50 have less latency when computing thousands of inferences. On another hand, this framework can answer which cell would allow you to finish accessing, processing and visualizing large amounts of data from Facebook, given that the program needs to run for hours. Another area open to research would be the environmental impact of running certain algorithms. How does the energy consumed per second change per cell when running scientific style computations? Is there a way to optimize this?

While this research tool allows for the `NVMExplorer` data to be easily comprehensible, it comes with downsides. At times, loading the bar charts from the first chapter incurs light latency, and so does customizing the visualization options. Moreover, it brings the user to a deep exploration, for which they must spend some time learning, and unless you already understand how to use the tool, it is not as efficient for quick questions.

## 4. Conclusion

In essence, this research commends the use of `NVMExplorer: Story Mode` to fully comprehend and create conclusions about your `NVMExplorer` data. As commented, the features and implementation facilitate a learning experience for personalized research questions and can help software developers optimize the performance of their projects, posing a benefit to the computer science community.

## References

[1] X. Dong, C. Xu, Y. Xie, and N. P. Jouppi, “Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 7, pp. 994–1007, 2012.

[2] L. Pentecost, A. Hankin, M. Donato, M. Hempstead, G.-Y. Wei, and D. Brooks, “Nvmexplorer: A framework for cross-stack comparisons of embedded non-volatile memories,” in *2022 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 2022, pp. 938–956.