Appendix: Artifact Description/Artifact Evaluation

Artifact Description (AD)

I. Overview of Contributions and Artifacts

A. Paper's Main Contributions

Provide a list of all main contributions of the paper.

C_1 HDL implementation of M^{3}XU.
C_2 Performance emulation framwork for M^{3}XU.
C_3 Applications using M^{3}XU.

B. Computational Artifacts

List the computational artifacts related to this paper along with their respective DOIs. Note that all computational artifacts may be archived under a single DOI.

A_1 https://doi.org/YY.YYYY/zenodo.0XXXXX

Provide a table with the relevant computational artifacts, highlight their relation to the contributions (from above) and point to the elements in the paper that are reproducible by each artifact, e.g., which figures or tables were generated with the artifact.

Artifact ID	Contributions Supported	Related Paper Elements
A_1	C_1	Table 3 Figure 5 (a), 5 (b)
A_1	C_2	Table 2, Table 4 Figure 4, Figure 5(c), 5(d)
A_1	C_3	Figures 6-9

II. Artifact Identification

Provide the following six subsections for each computational artifact A_i.

A. Computational Artifact A_1

Relation To Contributions

Provided the source code for:

- HDL for hardware synthesize result.
- Setting up performance emulation framwork for M^{3}XU.
- Performance emulation of case study application using M^{3}XU.

Briefly explain the relationship between the artifact and contributions.

Expected Results

- Area energy consumption of M^{3}XU.
- Raw perfromance data including throughput and latency of M^{3}XU.
- End to end latency of application M^{3}XU.

Provide a higher level description of what outcome to expect from the corresponding experiments. Provide an explanation of how the results substantiate the main contributions.

Algorithm A should be faster than Algorithms C and B in all GPU scenarios.

Expected Reproduction Time (in Minutes)

- 5 hrs for HDL synthesization.
- 120 hrs for M^{3}XU microbenchmark on all input sizes.
- 24 hrs for case study applications.

Estimate the time required to reproduce the artifact, providing separate estimates for the individual steps: Artifact Setup, Artifact Execution, and Artifact Analysis.

The expected computational time of this artifact on GPU X is 20 min.

Artifact Setup (incl. Inputs)

Hardware: Nvidia A100 DGX Station. Or equivlent system installed with Nvidia A100 PICE GPUs.
Software:

- Linux kernel version 5.4.0-81-generic
- CUDA 11.4, driver 470.57.0
- Synopsis design compiler with the 45nm FreePDK45 library
- Nvidia cutlass

Datasets: Input data are synthetic by framework or provided.
Installation and Deploymen: Configuration and build commands are provided.

Hardware: Specify the hardware requirements and dependencies (e.g., a specific interconnect or GPU type is required).

Software: Introduce all required software packages, including the computational artifact. For each software package, specify the version and provide the URL.

Datasets / Inputs: Describe the datasets required by the artifact. Indicate whether the datasets can be generated, including instructions, or if they are available for download, providing the corresponding URL.

Installation and Deployment: Detail the requirements for compiling, deploying, and executing the experiments, including necessary compilers and their versions.

Artifact Execution

- follow ReadME.md for cutlass installation and M^{3}XU integration.
- Configure licenses for Synopsis design compiler.
- Makefiles are provided for each case study applications.

Provide an abstract description of the experiment workflow of the artifact. It is important to identify the main tasks (processes) and how they depend on each other.

A workflow may consist of three tasks: T_1, T_2, and T_3. The task T_1 may generate a specific dataset. This dataset is then used as input by a computational task T_2, and the output of T_2 is processed by another task T_3, which produces the final results (e.g., plots, tables, etc.). State the individual tasks T_i and provide their dependencies, e.g., $T_1 \rightarrow T_2 \rightarrow T_3$.

Provide details on the experimental parameters. How and why were parameters set to a specific value (if relevant for

the reproduction of an artifact), e.g., size of dataset, number of data points, input sizes, etc. Additionally, include details on statistical parameters, like the number of repetitions.