SAGE: Self-Tuning Approximation for Graphics Engines

Mehrzad Samadi\textsuperscript{1}, Janghaeng Lee\textsuperscript{1}, D. Anoushe Jamshidi\textsuperscript{1}, Amir Hormati\textsuperscript{2}, and Scott Mahlke\textsuperscript{1}

University of Michigan\textsuperscript{1}, Google Inc. \textsuperscript{2}
GPU Specific Approximation

Goal: Hardware-aware approximation

Input Program → Static Compiler

Approximation Methods
GPU Specific Approximation

**Goal:** Hardware-aware approximation

Input Program ➔ Static Compiler ➔ Approximation Methods ➔ Approximate Kernels ➔ Tuning Parameters
GPU Specific Approximation

**Goal:** Hardware-aware approximation

**Input Program** → **Static Compiler** → **Approximation Methods** → **Approximate Kernels** → **Runtime system** → **Tuning Parameters**
We Can Control Output Quality.
We Can Control Output Quality.

Across 10 applications:

- 2.5x speedup with 90% output quality
- 2.0x speedup with 95% output quality