

Aaryaman Vasishta

linkedin.com/in/adyaman | aaryaman.vasishta@gmail.com | aaryaman.net | +44 07586572385

INTERESTS

- Real-time Path Tracing • FOSS
- Machine Learning • Neural Rendering
- GPU Kernel Optimization

EDUCATION

UNIVERSITY OF TOKYO

MASTERS IN INFORMATION SCIENCE AND TECHNOLOGY

2019-2021 | Tokyo, JP

Computer Graphics Group

Thesis: MCMC using Neural Networks

Advisor: Toshiya Hachisuka

PUNE UNIVERSITY

BE IN COMPUTER ENGINEERING

2012-2016 | Pune, IN

First Class with Distinction

Pune Institute of Computer Technology

LINKS

Github://jamm

BitBucket://jamm

SKILLS

PROGRAMMING

- C • C++ • Python • MATLAB • Lua
- Bash • x86 Assembly • Kotlin

VERSION CONTROL

Git • Perforce

TOOLS

- vim • Visual Studio • Slurm
- OpenGL • HIP • CUDA
- MySQL • Pytorch
- Cassandra • Kubernetes • Docker

OPEN SOURCE

CONTRIBUTIONS

OSS CONTRIBUTIONS

- ROCm • TheRock • hipBLASLt
- PyTorch • llama.cpp • llamafile
- aotriton • Mitsuba 2 • PBRTv4
- Chromium • WINE • nouveau
- Appleseed • Zandronum • ScummVM

GOOGLE SUMMER OF CODE

- Mentor - 2018 WineHQ, 2020 PCL
- Student in 2015 and 2016 for WineHQ

AWARDS

- 2023 AMD Executive Spotlight - Stable Diffusion Windows enablement
- 2023 AMD Executive Spotlight - SHARK Windows ROCm backend enablement
- 2023 AMD Executive Spotlight - For contributing to AI accelerated demos - RDNA3 launch event.
- 2023 AMD spotlight - RDNA3 launch event - achieved 2.7x AI performance uplift vs. RDNA2
- 2020 Japan Student Services Organization (JASSO) scholarship award
- 2018 Rakuten Tech Division top 1% award with stock bonus

EXPERIENCE

ADVANCED MICRO DEVICES MTS MACHINE LEARNING ENGINEER | TOKYO, JAPAN

June 2025 - Present

- Enabled ROCm and PyTorch on Windows for Radeon GPUs.
- Contributing performance optimizations in Triton
- Future GPU workload optimizations as part of shift-left

STABILITY AI MACHINE LEARNING ENGINEER | LONDON, UK

March 2024 - April 2025

- **Stable Fast 3D** - wrote MPS and CUDA backends for the texture baker, worked on tri-tri BVH optimizations for the uv-unwrapper.
- CUDA extensions for **Stable Point Aware 3D** and inference optimizations for **Stable Virtual Camera**. FP8 Quantizations for **Stable Diffusion 3.5**.
- Dataset generation and preprocessing for **Stable Virtual Camera** and talking-heads AI models.

ADVANCED MICRO DEVICES RESEARCH ENGINEER | TOKYO, JAPAN

April 2021 - February 2024

- Influenced multiple future hardware architectures on neural rendering by developing forward-looking workloads and evaluating them under different hardware constraints.
- Contributed to future GPU (RDNA4) HW enhancements after performing analysis of online AI/ML neural rendering workloads such as instant-ngp, Neural Radiance Caching. HW performance uplift projected up-to 30%.
- Awarded executive spotlight for enabling the Windows ROCm backend for nod.ai SHARK. (later acquired by AMD). Demonstrated the viability of the ROCm backend on Windows as a more flexible alternative to Vulkan backend.
- Awarded executive spotlight for writing fully-fused MLP kernels for current and next-gen RDNA3 GPUs, backing the gen-on-gen AI ops uplift of 2.7x.
- Re-implemented instant-ngp (Instant NeRF) from scratch, which includes fully-fused MLP kernels with WMMA optimizations (tensor cores), grid encoding and occupancy grids for in-house research work with competitive performance for RDNA3 GPUs.
- Ported tiny-cuda-nn and instant-ngp to HIP, supporting RDNA3 and MI GPUs, utilizing the WMMA ops of RDNA3 and matrix cores of MI. As part of this effort, modified CUTLASS to use HIP using rocWMMA backend.
- Enabled HIP support for llamafile and contributed towards RDNA3 compatibility for llama.cpp on Windows.

(CONTD.) ADVANCED MICRO DEVICES RESEARCH ENGINEER | TOKYO, JAPAN

April 2021 - February 2024

- Contributed to **HIP-RT** project infrastructure (bitcode linking, kernel baking, compilation performance).
- Worked on **PBRTv4 port from CUDA to HIP** with HIP-RT backend. Integrated support for interactive mode using HIP-OpenGL interop.
- Ported Radeon ProRender from OpenCL and CUDA to HIP within 48 hours and **improved CPU performance up-to 73%, GPU up-to 25%**.
- Wrote a popular tutorial on RDNA3's WMMA instructions with sample code - https://gpuopen.com/learn/wmma_on_rdna3/.
- Released Orochi - single-binary runtime linking for both HIP and CUDA. Hardened HIP runtime by writing test cases across HIP and CUDA to fix HIP issues using black-box reverse engineering.

RAKUTEN | SOFTWARE ARCHITECT/LEAD ENGINEER | TOKYO, JAPAN

October 2016 – March 2019

- **Awarded for being among the top 1% of the Technology Division employees.**
- **Mentored and onboarded new engineers and interns** from top Canadian universities. Mentored several new engineers and interns from project kickoff and infra setup until deployment in production.
- **Designed, developed and deployed large scale, zero-downtime, cloud-native** core identity services utilizing **Kubernetes** and **Cassandra** serving **billions of worldwide Rakuten logins daily**. Overall **savings of ¥12 million annually**.
- **RIT: Deep learning for Soft-segment background removal** of e-commerce images.

UBISOFT | INTERN GAMEPLAY PROGRAMMER, PUNE STUDIO

January 2016 – May 2016

- **Worked on gameplay and engine layers**, fixing issues while remastering **South Park™: The Stick of Truth™** to PS4 and Xbox One.

PROJECTS

PATH TRACER | PHYSICALLY BASED RENDERER USED FOR RESEARCH

2019 - Present

- **Cross-Platform and Written from scratch** using C++17.
- Current features: Live preview, multi-threaded rendering, Parallel SAH BVH and SBVH. Integrators: Path Tracing with Next-Event-Estimation and MIS, PSS-MLT, PRT using Spherical Harmonics.
- BSDFs supported: Diffuse, Phong, Rough conductor using GGX/Beckmann microfacet model, Dielectric and Metal.

WINEHQ | OPEN-SOURCE COMPATIBILITY LAYER FOR RUNNING WINDOWS PROGRAMS ON POSIX-COMPLIANT OS's

2015 - 2018

- **Re-implemented** Microsoft's Direct3D Retained Mode, a **3D scene graph API** in C resulting in **improved compatibility** across legacy applications and games. Contributions included in Valve's Proton layer for Steam on Linux.
- Wrote tests against Microsoft's undocumented API while adhering to **black-box reverse engineering** methodology.

PUBLICATIONS

- **Stable Virtual Camera: Generative View Synthesis with Diffusion Models**
International Conference on Computer Vision, ICCV'25
Jensen (Jinghao) Zhou, Hang Gao, Vikram Voleti, **Aaryaman Vasishta**, Chun-Han Yao, Mark Boss, Philip Torr, Christian Rupprecht and Varun Jampani.
arXiv preprint.<https://arxiv.org/abs/2503.14489>.
- **SPAR3D: Stable Point-Aware Reconstruction of 3D Objects from Single Images**
Computer Vision and Pattern Recognition, CVPR'25
Mark Boss, Zixuan Huang, **Aaryaman Vasishta**, James M. Rehg, and Varun Jampani.
arXiv preprint.<https://arxiv.org/abs/2501.04689>.
- **SF3D: Stable Fast 3D Mesh Reconstruction with UV-unwrapping and Illumination Disentanglement**
Computer Vision and Pattern Recognition, CVPR'25
Mark Boss, Zixuan Huang, **Aaryaman Vasishta**, and Varun Jampani.
arXiv preprint.<https://arxiv.org/abs/2408.00653>.
- **HIPRT: A Ray Tracing Framework in HIP**
Proceedings of the ACM on Computer Graphics and Interactive Techniques (**High-Performance Graphics**), 2024.
Daniel Meister, Paritosh Kulkarni, **Aaryaman Vasishta**, and Takahiro Harada.
<https://doi.org/10.1145/3675378>.