Andres Torrubia

antorbus@gmail.com
andrestorrubia.com
github.com/antorbus
+1 919 918 0912
linkedin

Even if Moore's Law dies, I believe there are still orders of magnitude of performance waiting to be extracted from ingenuous fabrication of the whole tech stack, from silicon to API. I envision myself at the interface where PCB meets software.

EDUCATION

Duke University
Triple Major in

- BSE Electrical and Computer Engineering
- BS Mathematics
- BS Computer Science

2023 - 2027
GPA: 3.89

SKILLS

Programming Languages
C/C++, CUDA, Python, Verilog, MIPS/RISCV assembly

Languages
English (Native), Spanish (Native), Mandarin (C1)

Software
Proficient: PyTorch, NumPy, OpenMP, Git, Linux/Unix, Arduino,
Working Knowledge: KiCad, LTspice, Docker, Xilinx Vivado, GNU/LLVM toolchain

Practical
Prototyping, soldering, data-sheet reading

EXPERIENCE

HummingBird: A Fast IO Aware Attention Kernel For Small Sequences
Research (advised by Prof. Emily Wenger) Fall 2024

- Built a multi-head self-attention kernel for small sequences (<= 48 tokens)
- Achieved 3x wall-clock time speedup compared to state-of-the-art methods such as flash attention
- Optimized specifically for 4090s and 3090s using bare CUDA at FP16

Metamaterial Inverse Design Problem
Research (with Prof. David R. Smith) Ongoing (Started April)

- Currently developing GPU-accelerated Discrete Dipole Approximation algorithms to simulate and control electromagnetic responses for 3d metamaterials

Kaggle Silver Medal, GoogleAI4Code
Deep Learning Competition Summer 2022

- Competed solitary against 1135 teams, ultimately ranking among Kaggle Grandmasters
- Developed and trained LLMs to automatically reconstruct the order of markdown and code cells in Python notebooks
- Approximated the non-differentiable Kendall Tau loss function using softrank to enable gradient-based optimization, later refined the approach to better approximate Spearman's rho
- Learned Rich Sutton's "The Bitter Lesson"

FPGA Raycaster Engine
Software/Hardware Project Fall 2024

- Designed and built a custom 5-stage pipelined, RISC-esque CPU architecture implemented on a FPGA and developed a full raycasting engine in bare-metal assembly, optimized the memory architecture to achieve real-time performance

Recommendation Systems and Image Generation
Intern, Freepik Company Summer 2022 and Summer 2023

- Developed a transformer-based recommendation system trained autoregressively, used Faiss for similarity search and multi-GPU for fast training and inference
- Engineered an image dataset generation pipeline from public domain media using FFmpeg and captioning models. Validated the dataset's efficacy by fine-tuning diffusion models with LoRA

LightLemur, A Lightweight Tensor Differentiation Library
Software Project Ongoing (Started January)

- Leading a team of 3 people to build a reverse mode tensor differentiation library
- Achieving close to PyTorch performance on CPU using OpenMP and SIMD optimizations

AM Radio
Hardware Project Spring 25

- In my quest to learn electronics, I have built an AM radio from scratch, from circuit to PCB, without internet help, only data-sheets and fundamentals