Optical Physical Foundation Models: Motivation, Opportunities, and Challenges

P L McMahon[1], L G Wright[2], T Wang[3], and T Onodera[1]

[1]*Applied and Engineering Physics, Cornell University, Ithaca NY, USA*
[2]*Department of Applied Physics, Yale University, New Haven CT, USA*
[3]*Dept. of ECE/ENG, Boston University, Boston MA, USA*
Contact Email: pmcmahon@cornell.edu

Foundation models are deep neural networks (such as GPT-5, Gemini 3, and Opus 4) trained on large datasets that can perform diverse downstream tasks – text and code generation, question answering, summarization, image classification, and so on. The philosophy of foundation models is to put effort into a single, large ($\sim 10^{12}$-parameter) general-purpose model that can be adapted to many downstream tasks with no or minimal additional training. We argue that the rise of foundation models presents an opportunity for optical-neural-network development: in contrast to when different models were used for different tasks, it now makes sense to build special-purpose, fixed hardware implementations of neural networks, manufactured and released at the roughly 1-year cadence of major new foundation-model versions.

Beyond conventional digital-electronic inference hardware with read-only weight memory, we advocate a more radical re-thinking: hardware in which the neural network is realized directly at the level of the physical design and operates via the hardware's natural physical dynamics – Physical Foundation Models (PFMs). PFMs could enable orders-of-magnitude advantages in energy efficiency, speed, and parameter density. For $\sim 10^{12}$-parameter models, this would both reduce the high energy burden of AI in datacenters and enable AI in edge devices that today are power-constrained to far smaller models. PFMs could also enable inference hardware for models much larger than current ones: 10^{15}- or even 10^{18}-parameter PFMs seem plausible by some measures. Back-of-the-envelope calculations based on an example of a 3D nanostructured glass medium suggest the possibility of favorable scaling. The talk will conclude with mention of some of the major research challenges that must be resolved for trillion-parameter PFMs and beyond to become reality.

This will be a perspective talk based on arXiv:2604.27911