

About me

- Sasank Chilamkurthy
- Co-author of PyTorch
- Former CTO of Qure.ai
- Working on a new Edge AI startup
- <https://www.linkedin.com/in/sasank-chilamkurthy/>
- <https://twitter.com/sasank51>



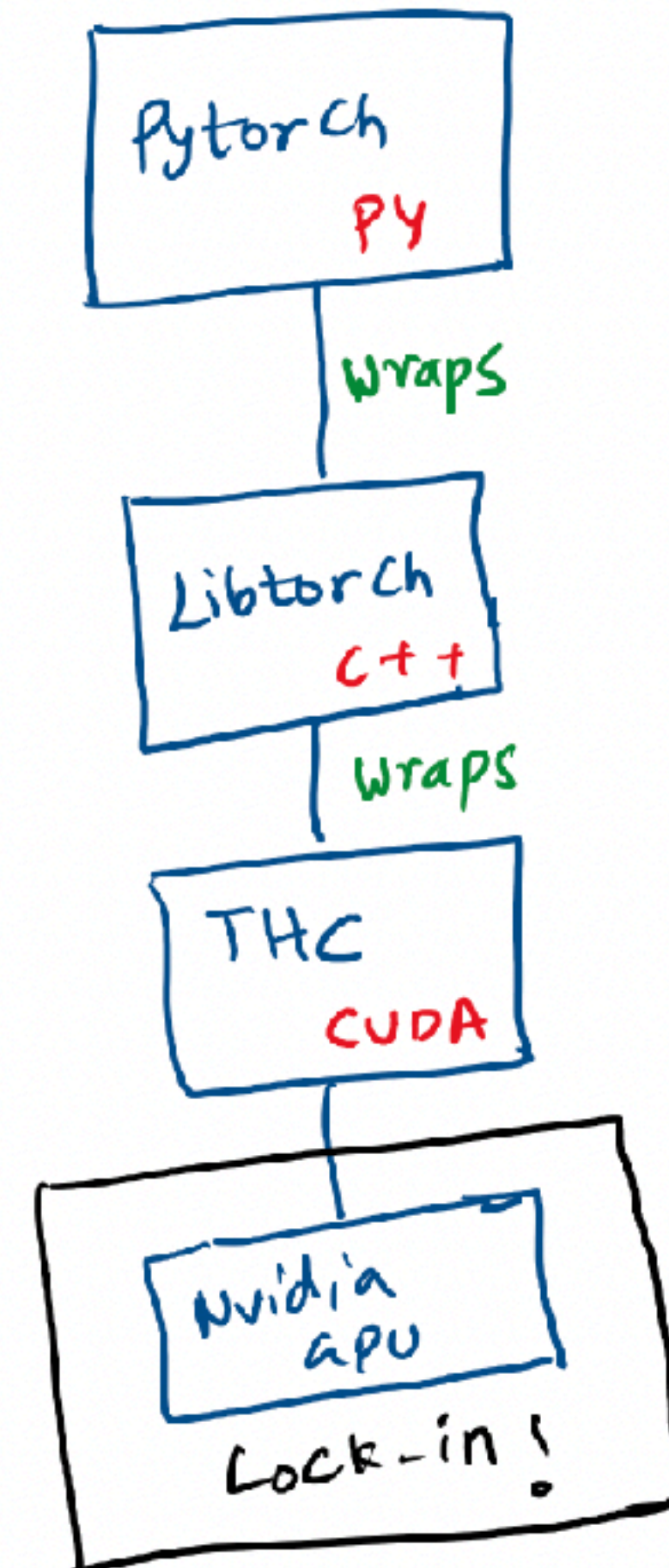
Virtualization for GPUs

- AI apps are often limited to Nvidia devices because they use CUDA.
- We're building a virtualisation runtime that lets AI devs use PyTorch for apps on any device.
- It's user-friendly, allowing devs to switch devices without extra steps.
- Helps devs save by using cheaper GPUs in the cloud and makes AI deployment on edge systems easier.

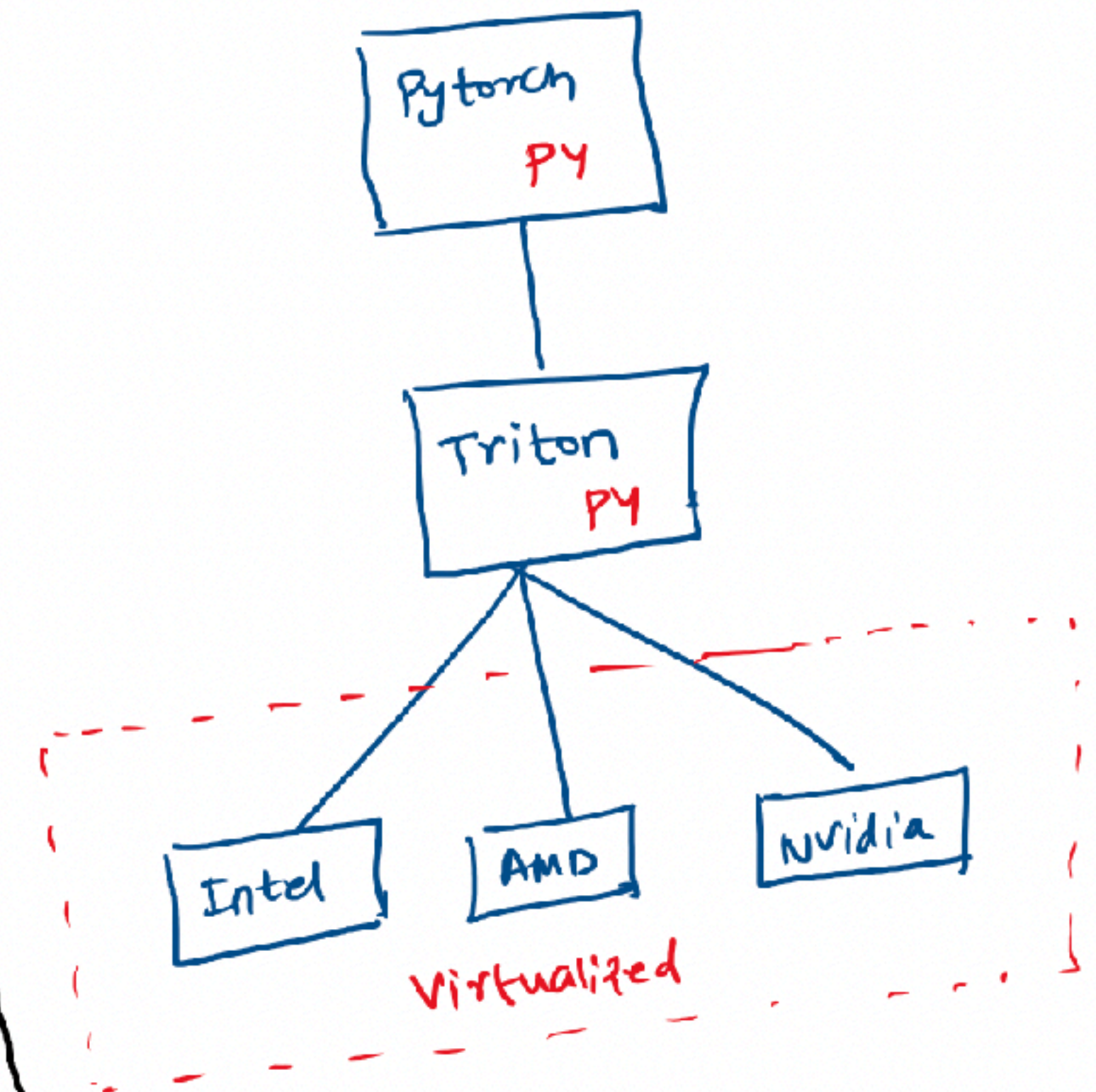
How it Works

- Using open-source tools like PyTorch, OpenAI Triton, llama.cpp, and SYCL
- Making Triton DSL compile easily to DPC++, RoCM, and more.
- PyTorch code runs on any GPU, no changes or add-ons needed

Before



After



Business Model

- Vertical integration, like Apple, for a seamless hardware-software mix.
 - Offering affordable AI developer PCs equipped with our virtualization runtime for app development
 - Selling specialized hardware for edge computing, compatible with our runtime
 - Providing cloud services using cost-effective GPUs
- Focus on developer market