

Dinosaur (2009 in 2020)

Machine Learning Setup & Choices

Jacob Pyrett

April 2020



Cult of Mac / Why Apple ? / My story

- I've used Apple laptops for about 18 years (~2002) - First one: 1 GHZ titanium Powerbook (TiBook). First job out of college. \$2000 laptop
- Really used to enjoy shopping at Micro Center , Kenwood Apple Store when sold actual software (Aspyr games – mac ports of PC games). EQ for Mac, Doom3 etc. Now on steam.
- Owned 3rd generation iPod, 5th gen iPod Touch, 1st generation iPad, never owned iPhone
- Really enjoyed the journey of OS X. "Cat releases: Lion, Tiger, etc.). About every year was a lot of OS hype.
- Went to San Francisco ! WWDC, developed simple Carbon & Cocoa apps

WWDC 2005 - Moscone Center (Saw Steve Jobs Apple World Wide Developers Conference)



Age: 30 – blurry , did not own digital camera back then. Paid for trip out of own pocket to go as worked at insurance company. Last time the conference ended up going to Apple HQ via bus.

Time for a new computer! for AI / ML learnings

Options


- 1) Desktop , Laptop, or Cloud VM
- 2) Software: deep learning, PyTorch, TensorFlow 2, misc python
- 3) OpenCL or CUDA. (or both?)
- 4) Dedicated GPU. Lots of CPU cores and > 6 GB GPU RAM
- 5) Prefer Linux (Ubuntu) or MacOS. Over Windows 10

TensorBook (lambdalabs.com)



[Servers](#) [Workstations](#) [TensorBook](#) [GPU Cloud](#) [Resources](#)

+1 (866) 711-20

 Due to global supply chain issues caused by COVID-19 we are currently experiencing shipping delays of 4 weeks for Tensorbooks.

TENSORBOOK

Deep Learning Laptop

GPU laptop with RTX 2070 Max-Q or RTX 2080 Max-Q. Ubuntu, TensorFlow, PyTorch, Keras, CUDA, and cuDNN pre-installed.

[Start customizing >](#)

[View top configurations >](#)



Tensorbook Specs : Core i7, 32 GB RAM, Nvidia RTX 2080 (watered down max-q) . Yikes! \$3100

Intel Core i7-9750H
6 Cores, 2.60 GHz

NVIDIA RTX 2080 Max-Q
8 GB VRAM



32 GB
System Memory

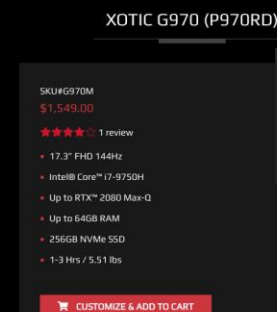
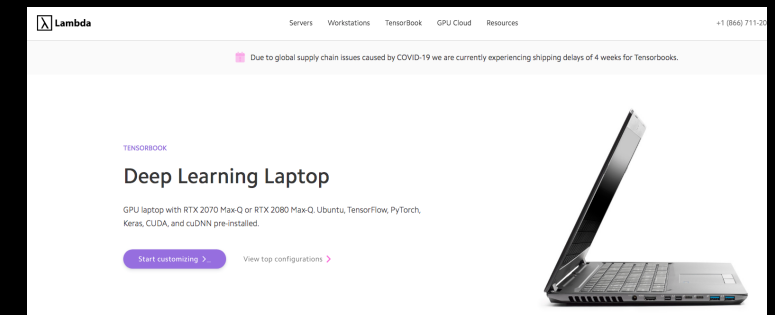
1 TB
NVMe SSD

16.1" Display
Matte, 1920x1080

\$3,100.00


Laptop Puppy Mills! Say hello to Clevo!

These are all the **SAME laptop**: rebranded companies – same Clevo unit
\$2000 - \$2800



What about Apple?

Mac Pro Overview



Customize your Mac Pro.

- 3.5GHz 8-core Intel Xeon W processor, Turbo Boost up to 4.0GHz
- 32GB (4x8GB) of DDR4 ECC memory
- Radeon Pro 580X with 8GB of GDDR5 memory
- 1TB SSD storage
- Stainless steel frame with feet
- Magic Mouse 2
- Magic Keyboard with Numeric Keypad - US English

Processor

Which processor is right for you?

3.5GHz 8-core Intel Xeon W processor, Turbo Boost up to 4.0GHz

3.3GHz 12-core Intel Xeon W processor, Turbo Boost up to 4.0GHz


Can't decide? [We're here to help.](#)

[View gallery](#)

Pickup: Apple Store Pickup is currently unavailable

Order today, delivers: Apr 27 - Apr 29 - Free Delivery options for 45103*

\$6,399.00



Customize your 16-inch MacBook Pro - Space Gray

- 2.3GHz 8-core 9th-generation Intel Core i9 processor, Turbo Boost up to 4.8GHz
- 32GB 2666MHz DDR4 memory
- AMD Radeon Pro 5500M with 8GB of GDDR6 memory
- 1TB SSD storage
- 16-inch Retina display with True Tone
- Four Thunderbolt 3 ports
- Touch Bar and Touch ID
- Backlit Keyboard - US English

Pickup: Apple Store Pickup is currently unavailable

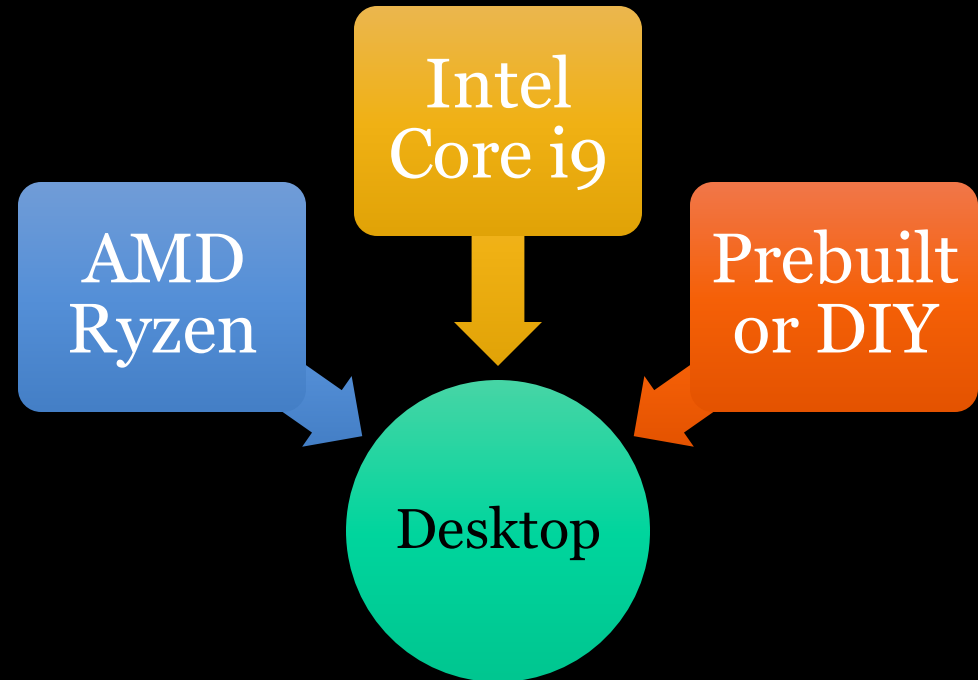
Order today, delivers: Apr 28 - Apr 30 - Fastest Delivery options for 45103*

\$3,299.00

8-core machine , 32 GB for too much money! Price of a new furnace or driveway

Build own PC Desktop

- Nickel-and-dime mentally drives me insane. Just a little more \$ for next level
- Not excited to put it together
- My experience with building PC was more expensive than just buying off the shelf
- Cheaper but still over \$1500. DDR4 is expensive along with a good 8 GB VRAM is expensive, Intel pricey
- State of the art graphics. Desktop graphics rock!





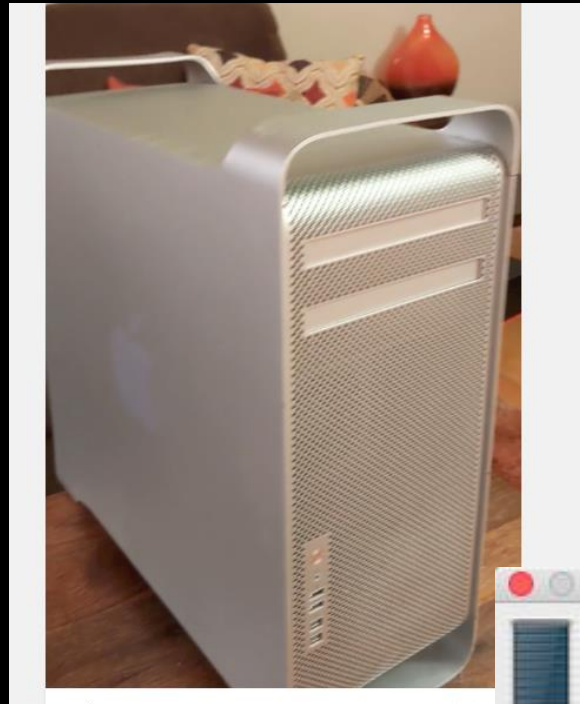
The goal / smart
thing to do

But:



Hello, OLD & used Mac Pro

VIDEO: <https://streamable.com/2p4x3o>



3.5 GHZ cores! -->

WHY ?

Macs must be good for AI – DS team has them

- 12 physical cores : 2 CPU (Intel Xeon) x 6 cores
- 24 virtual cores 3.46 GHZ each
- 128 GB ECC 1333 MHZ DDR3 ram, triple channel – cheap
- Build quality, cooling system
- Support 2 large GPUs, 4 PCI express
- It's a Mac, plays all my favorite games
- I always wanted one of these, were too expensive.
- CHEAP ! \$800 for one with : 64 GB ram, 1 TB SSD



Lessons Learned! - Here Goes

- The CPUs are 10 years old! BUT – they are fast enough.
- Python 3.8 well supported, lots of additional stuff is supported via Homebrew, Jupyter works great. and so far so good!
- Need a GPU. Heard CUDA was more widely used for DL and needed for TensorFlow. So started looking at NVIDIA
- Issue #1: Nvidia stopped providing drivers to Apple – three OS versions back. OK. 13.6 High Sierra is modern enough and still supported.
- Issue #2: Nvidia RTX cards not supported. OK, buy older Nvidia GTX card. Wow, this one looks good.. Ebay Titan X with 12 GB ram and 3000 cuda cores. Wrong choice \$250
- Issue #3: Titan GTX cards come in 4 varieties. Titan (maxwell), Titan X (pascal), Titan Xp, and Titan Z (basically a crappy double pascal). I didn't know and I bought the slowest one – maxwell and should have gotten a pascal based as 2x as fast! RAPIDS (GPU pandas) support missing



Lessons Learned! - Part 2

- Issue #4: Old Xeon CPUs do not have AVX or AVX2 instruction sets!! This is a breaker. Modern python BLAS libraries used for deep learning are based on these. Error: illegal instruction 4
- Issue #5: Prebuilt framework PyTorch only supports GPUs with CUDA capacity over 5. Must build from scratch
- Issue #6: GPU on Mac is not supported by TensorFlow 2
- Issue #7: Build system for PyTorch is broken for missing AVX. Ninja and other systems do not work
- Enjoyed the Mac other wise and really liked the old Titan and bonded with it. Titan is a cool card. A little brainwashed as want to use RAPIDS. (gpu pandas).



INTEL TO THE RESCUE!!! (not)

Intel Software Development Emulator


(<https://software.intel.com/en-us/articles/intel-software-development-emulator>)


Dynamically looks at each instruction and REPLACES missing AVX instructions with SSE instructions while running!

It works, but takes 30x- 50x longer to run for deep learning compared to other work laptop, so not really viable

INTEL TO THE RESCUE #2 (yes!)


[Products](#) [Solutions](#) [Support](#)



USA (English)  [Sign In](#)

PLAIDML

PlaidML is an open source tensor compiler. Combined with Intel's nGraph graph compiler, it gives popular deep learning frameworks performance portability across a wide range of CPU, GPU and other accelerator processor architectures.



Bring Deep Learning to Every Device

PlaidML brings nGraph compatibility to new GPUs and platforms. Included support for OpenCL and Apple Metal across popular GPUs enables the convenience of full-featured deep learning development in the GPUs built into every laptop. Modular hardware backends enable anyone to add hardware support — from embedded SoC to new accelerators.

Automatically Differentiate Machine Learning Frameworks

Deep learning advances routinely require the construction of new neural network operations. PlaidML enables researchers to add operations in hours instead of months through sophisticated code generation algorithms.

Keras Framework : Recompiles base libraries like pandas and pytorch to avoid AVX! using LLVM compiler

INTEL TO THE RESCUE #2 (yes!)



A platform for making deep learning work everywhere.

PlaidML is an advanced and portable tensor compiler for enabling deep learning on laptops, embedded devices, or other devices where the available computing hardware is **not well supported** or the available software stack contains unpalatable license restrictions.

PlaidML sits underneath common machine learning frameworks, enabling users to access any hardware supported by PlaidML. PlaidML supports [Keras](#), [ONNX](#), and [nGraph](#).

As a component within the [nGraph Compiler stack](#), PlaidML further extends the capabilities of specialized deep-learning hardware (especially GPUs,) and makes it both easier and faster to access or make use of subgraph-level optimizations that would otherwise be bounded by the compute limitations of the device.

As a component under [Keras](#), PlaidML can accelerate training workloads with customized or automatically-generated Tile code. It works especially well on GPUs, and it doesn't require use of CUDA/cuDNN on Nvidia hardware, while achieving comparable performance. PlaidML works on all major operating systems: Linux, **macOS**, and Windows.

INTEL TO THE RESCUE #2 (yes!)

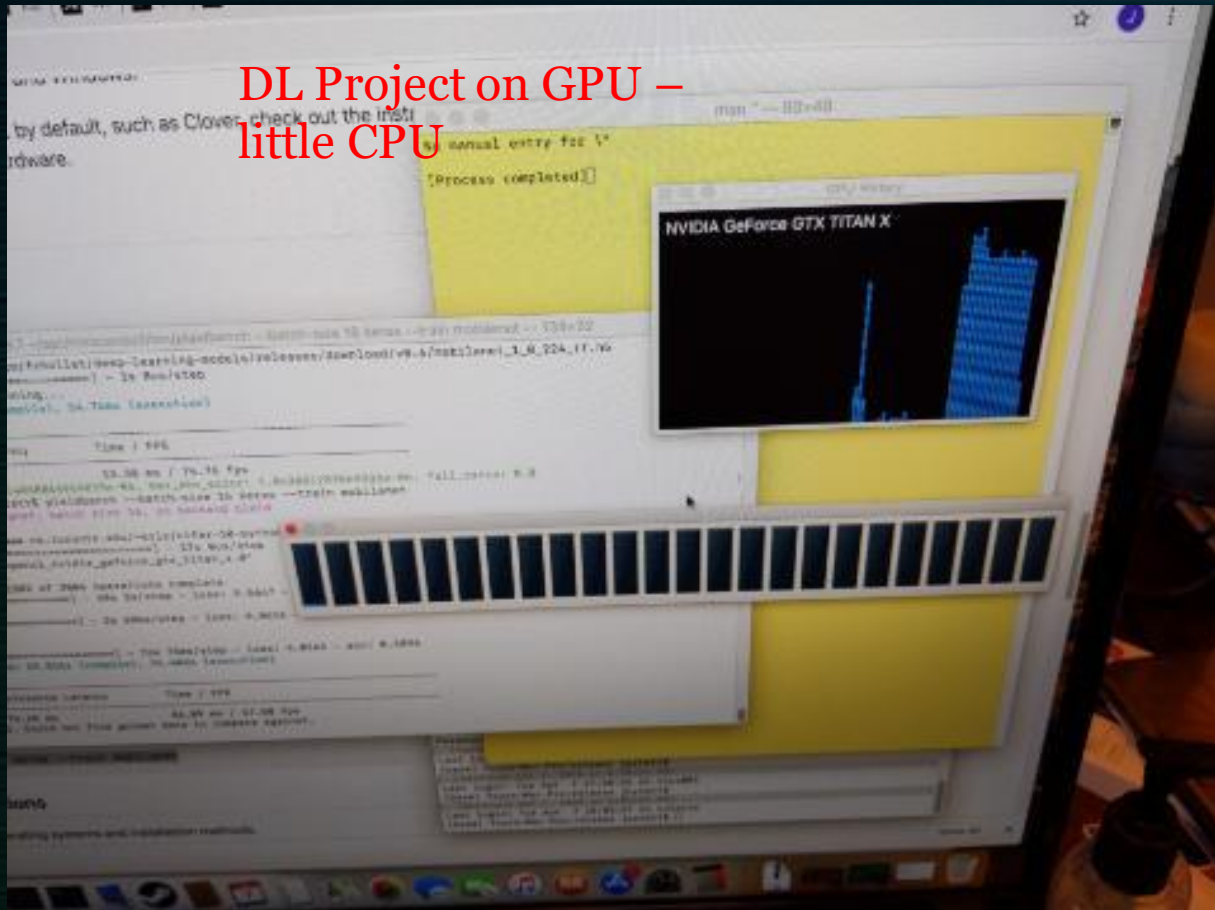


A platform for making deep learning work everywhere.

It did not work immediately OOTB. The rebuild instructions had ambiguous . I eventually figured out how to config it and will be writing a Medium article soon with the details. Possibly contribute to the project. Will see.. But it works!

DL Project on CPU
ONLY – no GPU!

DL Project on GPU –
little CPU



WAIT – What is the solution? For current Mac Books. What about AMD Radeon?



AMD ROCm

Since the ROCm ecosystem is comprised of open technologies: frameworks (**Tensorflow / PyTorch**), libraries (MIOpen / Blas / RCCL), programming model (HIP), inter-connect (OCD) and up streamed Linux® Kernel support – the platform is continually optimized for performance and extensibility.

Tools, guidance and insights are shared freely across the ROCm GitHub community and forums.

AMD ROCm

Applications	HPC & Machine Learning Apps			
Tools	Debugger	Performance Analysis	System Validation	System Management
Frameworks	TensorFlow	PyTorch	Kokkos	RAJA
Math Libraries	MIOpen	FFT, RNG	BLAS, Sparse	Eigen
Communication Libraries	RCCL	UCX, libfabric	MPICH	OpenMPI
Programming Models with ROCm	OpenMP	HIP	OpenCL™	Python
	Fully Open Source ROCm Platform			
Devices	GPU	CPU	APU	DLA