Groq - $10M Series A

Groq is building a next-gen processor optimized specifically for machine learning, initially TensorFlow. The long term measurement of success is TBD.

Fundamental Human Need: Understanding and Creation
25% of the population: Yes but indirectly
$100B by 2045: Possibly
Special Person: Yes

Overview

- Computer science is in the midst of an evolution from traditional software to machine learned insights and models. As this happens, code no longer needs to account for outcomes but rather be able to work through problems and adapt their response based on what they've seen.
- This simple reframing isn't broadly possible today as the compute resources and training models don't exist to allow software to make this jump across the board. But as companies, industries, areas collect more and more data, the training models are becoming more robust and can now, increasingly, be tested in the wild (called inference). In order to accelerate this likely movement, a broad swath of compute resources, technologies and toolchains need to be built to enable an ecosystem to develop.
- As software specifically makes this transition, products can become increasingly responsive, personalized and accurate at lower levels of cost and complexity in delivering value to their end users/customers:
    - A next-gen music application can learn more accurately what you like and may like and deliver a more delightful experience of music discovery. [predictive analytics]
    - A next-gen photos application auto categorizes, sorts and extracts key information so that photos are as much about the data about and in the photo as the visual photo itself. [image recognition + deep learning + classifier]
    - A next-gen car understands the tradeoffs of getting from point A to B and can both dynamically route and adapt to driving conditions to optimize for speed, reliability and safety. [image recognition + ai + machine learning]
    - A next-gen retail experience will allow you to walk into a store, pick items off a shelf and walk out. The cameras in the store can locate you, identify both you and the items you've picked and then automatically charge your credit card. [machine learning + image recognition]
- In order for any of this to happen in a scalable way, however, we need to start at an atomic unit and reimagine the underlying architectural framework we use to process information and execute commands. In other words, in the same way we are shifting from CPU to GPU, there is an opportunity to move from the GPU to whatever needs to come next. The reason is that the amount of information needed and the processing power needed is orders of magnitude greater than what a traditional CPU can accomplish.
- The first order improvement has historically been to shift the burden to GPUs but limitations exist with GPUs wrt cost, power and iops depending on the task/application. Further, GPUs and implementation models constrain where machine learning can currently be applied today (works reasonably well in a data center, doesn't work nearly as well in a moving car).
- There is an opportunity to build a new stack that allows a broad ecosystem of machine learning to power most software applications. In order to start, we will first redesign the core computing substrate and optimize it for machine learning tasks, frameworks and pipelines.

Intro to Machine Learning

- To bring yourself up to speed, I recommend starting as follows: overview of machine learning → overview of tensor flow → a useful example of things that are now possible
    - https://en.wikipedia.org/wiki/Machine_learning
    - https://www.tensorflow.org/

    - https://www.youtube.com/watch?v=iBs59GIXhIA
- Some more large scale examples of machine learning in the wild:
    - Google Photos App
    - Voice recognition on Google Nexus phones
    - Facebook NewsFeed
    - Google Search Ranking

Team

- Doug Wightman Co-founder and CEO: https://www.linkedin.com/in/dougwightman
    - 33yrs old
    - Engineer on Google Loon
- Jonathan Ross Co-founder and CTO: https://www.linkedin.com/in/jonathan-ross-12a95156
    - 34 years old
    - Founded TPU project as a 20% project at Google
    - Designed first 3 versions of TPU, implemented on FPGAs
    - Improved perf of the FPGA to match that of ASIC using some clever techniques endemic to software engineering
- Stellar feedback about Jonathan and Doug, especially Jonathan
    - “100x engineer”
    - “Father of TPU”
    - “Insights and horsepower were second to none”

Financing

- $25M pre
- $10M raised - all from Social Capital (we’d own 28.57%)
- 15% pool
- 3 board seats
    - Doug, Jonathan, Chamath
- Competitive Dynamics:
    - Eclipse.vc passed due to conflict with Cerebras (see below)
    - NEA
    - Greylock
    - Accel
    - KP

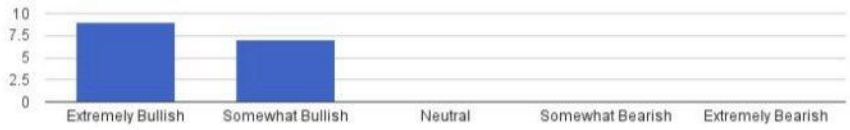
Competitors

- The category is still early but there are a few competitors already:
    - Google
    - Cerebras (raised $33M Series A; pre-product)
    - Nervana (raised $24M; acquired by Intel for $400M)
- We are betting that the original inventor of the TPU can:
    - Recruit a stellar technical team
    - Come up with incremental innovations and insights into the design of a next-gen TPU

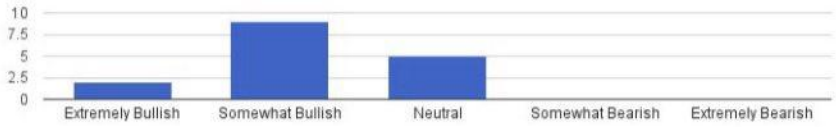
Team Discussion:

LINK To Voting Form

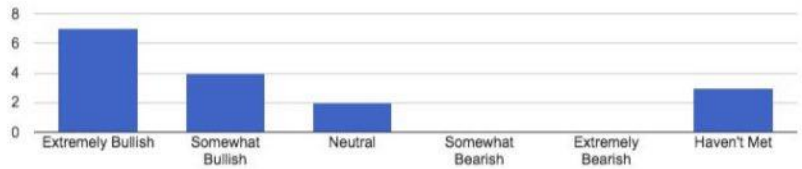
Sentiment on MARKET:



Sentiment on COMPANY:



Sentiment on CEO/Team:



Should we do deal?:

Low	Median / Average	High
$19	$575 / $800	$3,200

Representative quotes from responses:

- Teams is impressive and have done this before, hopefully they can do this outside GOOG
- Talent is the scarcest resource in ML right now, need for ML specific chops will explode - this is one of best teams to build it
- High variability outcome but should back this team for this market...it's a team bet; worst case it gets bought
- Unclear who will win beyond GPUs (or whether new processors too narrow and lose vs GPUs) - GPUs keep getting faster
- Worried about competition from well-funded behemoths - they aren't sitting on their hands
- Unclear how ML will evolve beyond computer vision & NLP
- Is this team too confident? Brilliant on ML dimension but potential blind spots around building hardware/systems, will need to be paired with a business/market person that is adept in this
- Don't think we have a prepared mind or any special skills that gives us an unfair advantage here.
- We should have seen in detail what other players were up to

Non-participants:

Decision	Summary of Team Discussion / Decision Rationale
[What was the decision? Pass / invest]	Struck a deal on $10 on $25; 15% pool