Questions:

* What does your team do?
* How much of it involves machine learning?
* What are some top things that you look for in your engineers?
* In your opinion, what is the top reason an engineer would get pipped. What can someone do to make sure they’re performing well?
* Be honest, be focused, collaboration (don’t work in a silo, share knowledge, be vocal and collaborative)
* What are some key things that I should learn to succeed on this team?

Shift Detection:

* Issue on the product chip that we found that could’ve been flagged by a device parameter earlier in the line. One of our senior engineers brough tup the question – “why didn’t we notice it soon?” the answer was that no one was monitoring this parameters because it was so obscure and we had over 9000 parameters.
* I recognized that we can do some kind of automated shift detection or change point detection analysis. So I did a lot of research and tried many types of algorithms and found the PELT (Pruned exact linear time) algorithm worked best with our data.
* PELT looks at a cost function as well as a penalty value to protect against overfitting. It also prunes the data into segments when looking for the change points. So it iterates through the data and Calculates the minimization at each data point and stores the optimal segmentation at each point. Then it goes back and offers the locations of the changepoints under the assumption that as we increase data, there will be a linearly increasing number of change points. This is controllable by inputs into the algorithm.
* I first build the whole system on Jupyter Notebook to make sure it works. Afterwards, I moved the code up to our server which makes the data pull a lot more efficient. I also used the multithreading on python to speed up the process even more. I set up the a cron job to run this algorithm on all our parameters every 2 weeks and had it set up pub-sub system to inform certain parameter owners of shifts. I also created a dashboard that would graphically show these parameter shifts and well as provide details like median pre shift, median post shift, lot\_id and date of shift, and group the parameters by type.

Clustering Project

* During my down time I wanted to work on a wafer clustering tool based on chip x/y data. I was inspired by a paper written in ieee utilizing this method called Simultaneous Orthogonal Matching Pursuit specifically with wafer regionality data clustering so I wanted to attempt to build it.
* I organized a design thinking session for my team to see what potential customers would want to see out of this project. The paper was missing a lot of background information and equations so I went above and beyond to contact the author of the paper on linked in and set up a series of phone calls with him to help me understand the algorithm.

AWS experience

* I have used AWS S3 and EC2 at Berkeley, also Google Cloud Platform (Big Query, cloud storage)
* I learned about AWS lambda but have never used it.
* I have experience with Map Reduce and docker and Kubernetes and Kafka queues
* IBM Cloud virtual server and Watson X tools

Tell me about yourself – short and crisp

* I have a bachelors and masters degree in Electrical Engineering from UIUC and Georgia Tech both of which are top 5 schools for EE. After I graduated from Georgia Tech, I was hired by IBM in their Semiconductor development team. It’s a highly technical job, but I realized really early on that a huge chunk of the job involves driving value out of the data that we collect in order to maximize yield, performance and reliability… but the tools we were using were super basic and no one was using any state of the art data science techniques.
* So I asked my boss about 1 year into my career at IBM if I can do a second Masters but in data science and Machine Learning at UC Berkeley. He was pretty enthusiastic about it and convinced IBM to pay for my entire tuition. So I started doing full time work and full time school.
* I’d say his belief in me paid off, because a few months after I started at Berkeley, I built my first defect classification model on python. Then I started pulling data directly into python, I automated a lot of my work, and I started using machine learning as well. This jump started what we call the “transformation” protocol in my team and next thing I know, I was teaching my coworkers how to pull data and use python and some basics in data science.
* I’ve been coding in Python specifically in the realm of data science and machine learning since 2017 (7 years) but overall I’ve been coding since 2010 with my Electrical and Computer Engineering background. My strongest object oriented programming language by far is in Python and I am very experienced with SQL queries. But since I’ve been in tech for so long, I have experience with other languages such as C and R and Spark through work and school projects.
* So I have been working at IBM for 8 years now in hardware development, but my role had been pivoting strongly towards data science for the past 6 years. I think the next step in my career is to get a job as an official data scientist and machine learning engineer. I’ve learned a lot at Berkeley, and a lot on my own through my work at IBM, and I’m hoping to learn even more by working at Capitol One.

Strengths

* Constantly learning and trying new things
* Really enjoy being innovative
* I am very easy to work with, not a of conflicts with my team, I can find pretty amicable ways to deal with disagreements while not compromising what I believe is correct

Weakness

* I can be a little shy when it comes to public speaking. I overcome it by practicing a lot and really know my material. And usually when the adrenaline starts going I can get locked in. But I can get pretty nervous before I have to do a presentation to a large group.