The Great Processor Shortage: History and Projection of an Emerging Trend

1st Dang Vo
Faculty of IT
University of Science
Ho Chi Minh, Vietnam
vndang21@vp.fitus.edu.vn

2nd Hau Nguyen
Faculty of IT
University of Science
Ho Chi Minh, Vietnam
ntthau21@vp.fitus.edu.vn

3rd Vinh Thai
Faculty of IT
University of Science
Ho Chi Minh, Vietnam
tvvinh21@vp.fitus.edu.vn

4th Thuc Doan
Faculty of IT
University of Science
Ho Chi Minh, Vietnam
dnthuc21@vp.fitus.edu.vn

***Abstract*—The recent chip shortage has caused tremendous disruption in various industries, notably the chip development industry itself. Naturally, two questions arise: First, when will the shortage subside and end? Second, how will the chip industry change from the shortage? This research seeks to answer these questions by modeling the relationship between a processor's metrics and its release date using a simple polynomial regression model. This is coupled with the least-squares method to fit said model to the acquired data points. Results show that the chip shortage had already ended in Q3 2022, and the shortage did not seem to have much of an effect on the processor development industry. However, the shortage might reemerge should there be another global crisis similar to those in recent years.**

***Index Terms*—CPU, Correlation, Economics, Electronics industry, GPU, Polynomial regression, Semiconductor shortage.**

I. Introduction

Imagine it's the simpler times around the start of 2019. You, a hard-working corporate employee, haven't a care in the world about your dusty old desktop as your company already provides you with high-end ones at work. As 2020 hits, so does the COVID-19 pandemic, and now you must work from home. Fast forward to 2021, you have just gotten accustomed to remote working on your sub-par desktop. And then disaster strikes, your graphics card fails on you. At first, you are utterly enraged. But then, you start to sympathize with your 7-year-old card: "Maybe it's time for an upgrade" – you thought. You start scouring the internet for a replacement only to be hard-pressed to find that the average graphics card price has skyrocketed from $267 to a whopping $1077 in just under 2 years [1]!

But how did all this happen? Indeed, The COVID-19 pandemic has been a great detriment to our societies for almost half a decade. It has not only disrupted our daily lives but also advancements in various fields. One such field is the field of processing unit development, i.e., the field concerning the development of CPUs and GPUs. Coupled with the pandemic, over the past few years, this field has been relentlessly attacked by various advents such as the China-United States trade war, cryptocurrency mining, the Russia-Ukraine war... [2] to the point of stagnation. At the supply's end, due to the shutdown of virtually all supply lines of intermediate products, processor manufacturers have been struggling to acquire enough materials to keep up with the rising demands. While at the demand's end, long periods of lockdowns imply a surge in remote work creating an increase in demand for computers in which processors reside. These factors caused prices to shoot up to unprecedented figures which led to the great processor shortage in the second half of 2020.

The processor shortage has caused great disruptions in various industries, largely those that utilize processors in their production procedures. However, though mostly detrimental, the shortage gave rise to many significant innovations in processor architectures in order to reduce manufacturing costs, thus lowering the barrier of entry to processors. Naturally, two questions arise: First, when will the shortage, along with all its demerits, subside and end? And second, how will the processor industry change from the shortage? But the insight into the future lies within the past and thus, we must first take a step back into the past and investigate the history of the great processor shortage. Hence the title (and purpose) of the research.

Although there have been a few efforts in tackling the impact of the computer chip shortage, as stated in section II, these efforts are made mostly to address the shortcomings of the computer chip supply chains rather than the actual developments of the chip itself. Furthermore, there does not appear to be any rigorous attempt at predicting when the shortage will end. All in all, to the best of our knowledge, an investigation into when the shortage will end and how it will impact the development of processors has not yet been **formally** studied.

Answering these questions will provide us with an insight into why the shortage happened and how to prevent such a catastrophe in the future. Furthermore, knowledge of when the shortage ends will tell us if the shortage is long-term or short-term. This will aid manufacturers to be more flexible in timely adjusting the level of production to suit the ever-changing demands. All in all, the research will help get the state of the economy back on track and ensure its stability for the near future.

The research will leverage datasets on common metrics of processors to analyze the evolution of the processor shortage. Armed with that knowledge, the paper will attempt to project when the shortage ends and what is the next big trend for the processor industry. To achieve this, we will employ the help of a polynomial regression model coupled with the least-squares method – A reliable tool for short-term predictions. By feeding the model historical data on processors' metrics, it can learn the hidden patterns and be able to predict when the shortage will end along with the future of the development of processing units.

It can be said that there are many factors which contribute to the processor shortage. However, the research will only address the impact of past advents on the great processing unit shortage and not potential future events. Moreover, the research will only dissect in detail the trend in performance, efficiency, and value in CPUs and GPUs from the first quarter of 2017 to the first quarter of 2023. Specifically:

- performance refers to speed,
- efficiency refers to power consumption, and
- value refers to the amount of performance per dollar, i.e., how much bang you are getting for your bucks.

Overall, the objectives of this paper are two folds:

1) To investigate the history of the great processor shortage and draw a conclusion on when it will end.
2) To project an emerging trend in the development of CPUs and GPUs in terms of performance, efficiency, and value.

II. Related Works

In recent years, the number of research raising concerns about the chip shortage implies a shared interest in analyzing it.

Most of these studies are aimed at revealing a vulnerability in the global **chip supply chain** and how to remedy it: [3], [4], and [5] try to identify the cause of the shortage by investigating the said vulnerability and suggest remedies to the problem; [6] attempts to lessen the impact of the shortage by suggesting the recycling chip testing method in order to increase the production yield.

Others see the crisis as an opportunity to learn about **supply chain** management: [7] attempts to provide a new framework for strategically managing global supply chains based on lessons learned from the shortage; [8] explores how an industry-wise systemic disruption starts and propagates using the shortage as an example; Lastly, [9] and [10] tell an engaging story on how the chip shortage supply chain impacted the world.

However, the aforementioned research differs from this paper in that they concern the evolution of the **chip supply chains** rather than that of **the actual chips** themselves. Some studies do address the development of chips such as [11]. But [11] only investigates how the price will change with a large emphasis on the chip market rather than the technological advancements in the chip manufacturing industry.

To the best of our knowledge, an investigation into when the shortage will end and how it will impact the development of processors has not yet been formally studied.

But fortunately, data analysis using regression has been done before. To achieve our goal of forecasting the development of processors, we leverage the use of a polynomial regression model. The model is well known for its accuracy in predicting the short-term growth of a population and has already been applied to solve various problems such as predicting rainfall [12] or the infamous COVID-19 pandemic's death toll [13].

III. Methodology

A. Data selection

But first, what is a "shortage"? In the economics lingo, a more common jargon for "shortage" is "scarcity". Thus, to analyze the processor shortage is to analyze the scarcity of processors. And as scarcity is closely related to a commodity's price, investigating the history of the shortage and predicting when it will end is equivalent to dissecting and projecting the average **price of processors over time**. Moreover, the research also aims to forecast an emerging trend in the development of processors in terms of **performance, efficiency, and value**. Therefore, to conduct this research, we must first acquire data on the performance, efficiency, release date, release price, and price over time of processors.

While datasets on the former 3 attributes are abundant, those on processors' prices are surprisingly scarce. In the end, we decided to consult PassMark for their CPU dataset [14] and GPU dataset [15]. Conveniently, their datasets already contain metrics dedicated to evaluating performance, efficiency, and value.

Moreover, the metrics on prices can be found by navigating to the respective processor's page in which there is a pricing history graph. Thanks to this graph, we could not only extract the release date and release price of a processor but also its price history.

Additionally, we also kept trivia information such as the names or categories of processors for easy data inspection.

B. Data overview

To summarize, we divide our data into 2 datasets:

1) The CPU dataset, of 1110 entries, comprises of:

- Trivia data:
  - Name
  - Category
- Performance measuring data:
  - Mark[1] (Performance rating)
- Efficiency measuring data:
  - TDP (W) (The highest heat a processor can handle)
  - Power performance (Performance per TDP watt)
- Value measuring data:
  - Release date
  - Release price
  - Value (Performance per Release price dollar)
- Pricing history (List of price points in the pricing history)

[1]Formulas for marks can be found at https://forums.passmark.com/performancetest/4599-formula-cpu-mark-memory-mark-and-disk-mark.

2) The GPU dataset, of 237 entries, comprises of:

- Trivia data:
  - Name
  - Category
- Performance measuring data:
  - G3D mark (3D graphics performance rating)
- Efficiency measuring data:
  - TDP (W)
  - Power performance (3D Performance per TDP watt)
- Value measuring data:
  - Release date
  - Release price
  - Value (3D Performance per Release price dollar)
- Pricing history

C. Data acquisition

Acquiring data on performance and efficiency is as simple as copying the table HTML from the page itself and parsing it to a CSV file. To obtain the data on the release date, release price, and price over time, we crawled the link to each of the processor's page extracted from the aforementioned table HTML.

Both tasks were accomplished using Python's Requests and BeautifulSoup4 module.

D. Data cleaning

The data-cleaning process is as follows:

1) Drop irrelevant or derived columns
2) Drop rows without release dates
3) Process datatypes of each column
4) Process pricing history:
   a) Remove rows without any price data
   b) Adjust prices for inflation
   c) Sort price points by date and remove duplicates
5) Infer release date and price from the pricing history
6) Fill null cells with their columns' mean values
7) Remove irrelevant rows, notably those from prior to 2017 as they are not within the scope of the research
8) Recalculate derived columns such as values and power performances

The process was performed using Python's Pandas module.

E. Data analyzing

As data analysis is ultimately the step that will lead us to our answer, it is important to revise the questions the research seeks to answer:

1) When will the shortage end based on its historical data?: As shortage is closely related to price, predicting when the shortage will end is equivalent to pinpointing when the average processor price is stabilized. Moreover, as the question is purely one of economics, we will only rely on the pricing history data to make the said prediction.

To achieve this, we will first look at the past trends in the average price of every quarter from 2017 onwards. And with each rise and fall during the shortage, we will attempt to associate it with an event that occurred at that time. This will lend us some insights into how future events can pan out. Furthermore, as the shortage began around the second half of 2020, the price data from prior to that period can be considered stable or normal. By comparing today's average price with the pre-shortage price (after adjusting for inflation of course!), we can make an educated guess on when the shortage will end.

Second, we will leverage the use of a simple polynomial regression model to predict the end of the shortage. A simple polynomial regression model is an analytic tool used for modeling the expected value of a dependent variable y in terms of the value of an independent variable x as an nth degree polynomial of the form:

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + ... + \alpha_n x^n$$

In other words, it "fits" a polynomial graph between data points such that the sum of the residuals (the differences between these data points and the polynomial) is minimum. To attain this minimum, the parameters of the polynomial in question are adjusted using the least-squares method. Instead of minimizing the sum of the residuals, the method seeks to minimize the sum of the residuals squared. This minimum can be easily found by setting the gradient of the polynomial to zero and solving for the parameters. However, achieving the minimum does not mean that our model fits the datapoints perfectly. It just means that that is the best that our model could do, not that our model is the best. To measure how accurate a model is at fitting/predicting, we can use the coefficient of determination, denoted as R^2.

This tool is perfect for our problem as we only have 1 independent variable (date) and 1 dependent variable (price). Additionally, the relationship between them is obviously not linear. What's more, this tool can be conveniently used by importing Python's famous NumPy module.

2) What will be the next big thing in the development of processors in terms of performance, efficiency, and value?: The question asks us to **predict** the upcoming **trend** in the development of processors. Sounds familiar, doesn't it? Indeed, we could just perform a simple polynomial regression analysis on the dataset where the independent variable is the release date, and the dependent variables are the metrics related to performance, efficiency, and value. However, if the relationship is linear, the polynomial regression method is prone to overfitting. Overfitting is when a regression model fits too closely to a given dataset that it loses its generality, rendering the model unsuitable for prediction. To combat overfitting, we must first determine whether release date has a strong linear relationship with any other metrics. Enters Pearson correlation coefficient.

Pearson correlation coefficient (PCC) is a measurement of how linear the correlation (relationship) between 2 metrics is. PCC values range from -1 to 1 with values close to 0 meaning no correlation, values approaching -1 indicate negative correlation, and values approaching 1 represent positive correlation. The strength of the linear correlation is measured by the absolute value of the PCC value.

Table I
CORRELATIONS BETWEEN RELEASE DATE AND OTHER METRICS

CPU correlations	
	Release date
Mark	0.375
TDP (W)	0.115
Release Price	-0.006
Power Perf.	**0.508**
Value	0.205

GPU correlations	
	Release date
G3D Mark	**0.150**
TDP (W)	0.129
Release Price	-0.083
Power Perf.	0.062
Value	0.075

As table I demonstrates, none of the correlations are above 0.6. Thus, the relationships are likely not to be linear. However, in the chip industry, most metrics grow exponentially rather than linearly. A great example of this is Moore's law which states that the number of transistors in an integrated circuit will double every other year. Knowing if the relationship is exponential is beneficial as it will lead to a better prediction model. Fortunately, we can also achieve this insight using PCC after applying log transformation – A type of transformation that linearizes an exponential model by replacing every dependent variable y with $\log y$ – on the data:

Table II
CORRELATIONS AFTER APPLYING LOG TRANSFORMATION

CPU correlations	
	Release date
Mark	0.388
TDP (W)	0.043
Release Price	0.046
Power Perf.	**0.461**
Value	0.444

GPU correlations	
	Release date
G3D Mark	0.031
TDP (W)	**0.141**
Release Price	0.008
Power Perf.	-0.020
Value	0.033

By table II, it doesn't seem that any of the correlations are exponential as the highest PCC value is a measly 0.488.

Another way to check for the type of model we are dealing with is by simply looking at each metrics' scatter plots: Figure 1 shows us that the patterns, or lack thereof, are seemingly random. The only exception to this is the power performance plot in figure 1 which is somewhat linear. This is also prevalent in table I where we already explained that its PCC value is still not strong enough to be considered linear.

Thus, we must apply a polynomial regression model to solve this problem. On top of that, as the patterns are quite chaotic, we will apply regression on the average price of every quarter rather than individual data points. This will help us ensure an accurate regression model at the cost of losing information on individual processors. However, the advantages far outweigh the disadvantages as we do not care about **every** processor, merely the trend in **all** processors, within the scope of this research. Moreover, in the case that polynomial regression

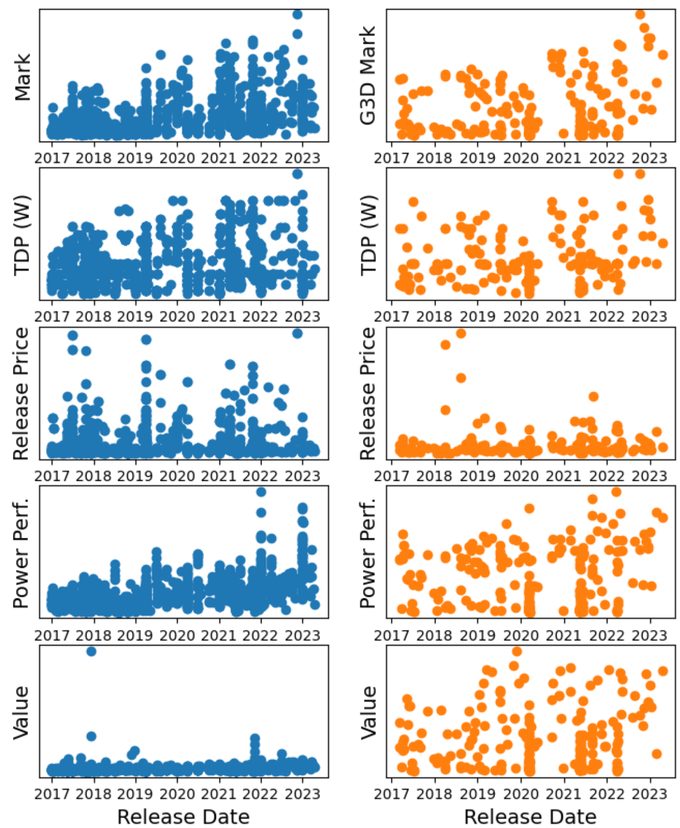

Figure 1. CPU (left) and GPU (right) metrics versus release date

fails to predict the dependent variable due to the pattern being too chaotic, we will resort to linear regression instead. Linear regression is a special case of polynomial regression where the polynomial's degree is 1. Although not as robust as polynomial regression, linear regression can help us identify the overall trend of the data more easily by looking at its slope.

IV. RESULTS

A. When will the shortage end based on its historical data?

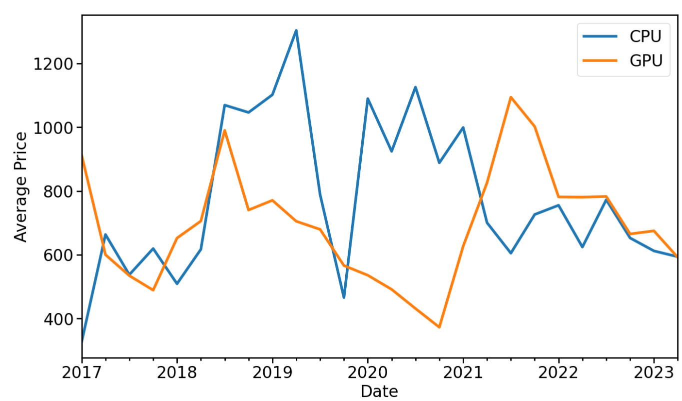

Figure 2. Inflated average pricing history of CPUs and GPUs

1) An educated guess: First, we will summarize the history of the shortage using the average prices shown in figure 2.

We will start our analysis in Q4 2019 as that is when the shortage started to take effect [2]. As we can see, CPU and GPU prices were on the decline prior to the pandemic, sitting at around \$450 and \$550 respectively. But as soon as Q1 2020 hit, CPU price spiked to about \$1050, more than double its price at the end of 2019! This is due to the drastic shift to remote working and learning in various countries coupled with the shutdown of many chip supply chains in favor of practicing social distancing. This was not the case for GPUs however, presumably because most remote jobs require CPUs more than GPUs, whose price continued to decline. CPU price remained stagnated while that of GPU keep decreasing until Q4 2020 when the China-United States trade war started which caused both prices to increase. In 2021, as CPU production picked up, its price significantly dropped. However, GPU prices kept growing due to the upsurge in demands from the cryptocurrency mining craze. After which, in Q3 2021, GPU prices started to drop while CPU prices oscillated around \$700. Today, the average CPU and GPU prices are around \$600.

Comparing these prices with those from the pre-shortage era (2017 – 2019), we can deduce that CPU prices have already recovered from the shortage since about Q2 2021. On the other hand, although GPU prices have only stabilized since Q3 2022, they are still on the decline today. Thus, we can conclude that the processor shortage had already ended in Q3 2022.

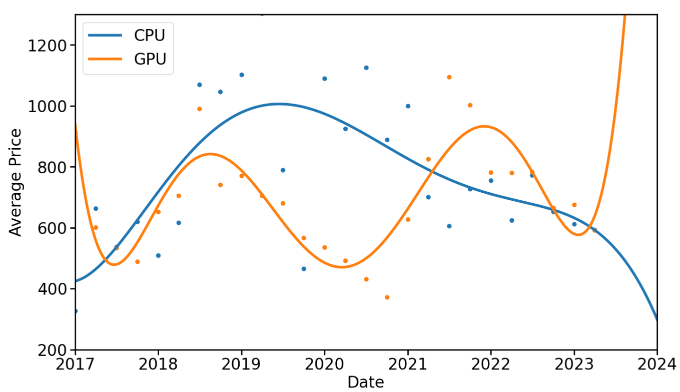

Figure 3. Average pricing history with polynomial regression curves

2) Prediction using polynomial regression: The polynomial regression curves shown in figure 3 are both of degree 6 as that is the optimal degree suggested by NumPy to ensure that we won't overfit. The CPU curve, sporting an R^2 value of 0.49, suggests that CPU prices will continue to drop, down to \$300 in 2024. On the other hand, the GPU curve, with an R^2 value of 0.74, forecasts that GPU prices will skyrocket to unprecedented figures.

B. What will be the next big thing in the development of processors in terms of performance, efficiency, and value?

For this question, we will apply both polynomial and linear regression on data points of average processors' metrics over each quarter, represented by a scatter plot, to see the current trend in processors. All polynomial regressions are of the optimal degree of 7 as NumPy suggested and are represented by dotted lines. Linear regressions are represented by continuous lines.

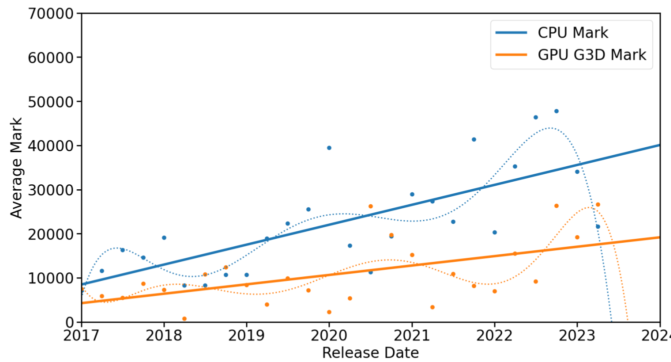

Figure 4. Average CPU mark and GPU G3D mark versus release date

1) Performance: As shown in figure 4, the average CPU mark and GPU G3D mark seem to have an upward trend with CPU marks being both higher and rising faster than those of GPU as suggested by the linear regressions. According to the polynomial regressions, however, the average mark for both will decrease in the upcoming year. The R^2 values for the polynomial regression of CPU and GPU are 0.72 and 0.52 respectively.

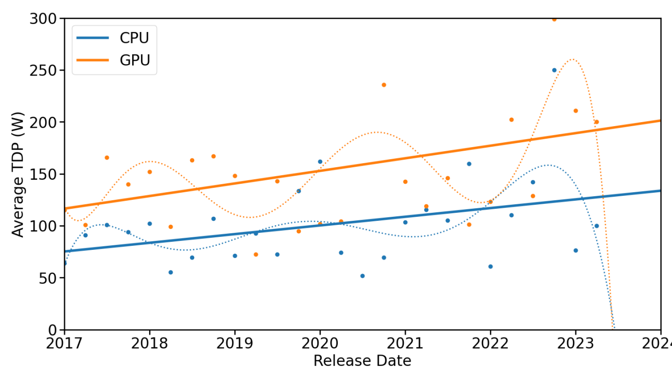

Figure 5. Average TDP (W) versus release date

2) Efficiency: In figure 5, the average TDP seems to be rising at the same rate for both CPU and GPU with the latter being about $50W$ higher than the former. Again, the polynomial regression suggests that the average TDP will fall in the middle of 2023. The R^2 values for the polynomial regression of CPU and GPU are 0.30 and 0.38 respectively.

Figure 6 also tells a similar story: The average power performance seems to rise for both with the CPU's rate of change being significantly higher. However, the polynomial regressions propose the opposite: CPU's power performance will decrease while that of GPU will increase dramatically in Q2 2023. The R^2 values for the polynomial regression of CPU and GPU are 0.64 and 0.54 respectively.

3) Value: As demonstrated in figure 7, the average release price (already adjusted for inflation) will drop gradually according to the linear regression. On the other hand, the

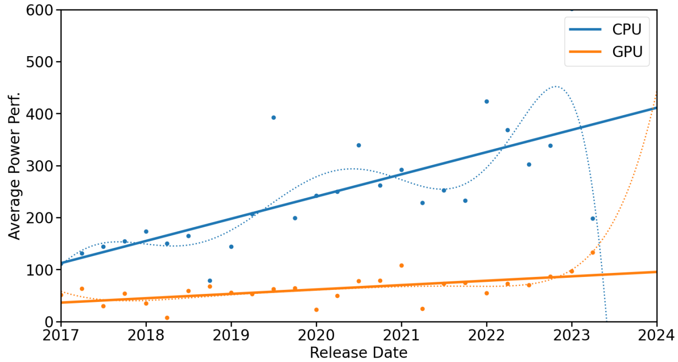

Figure 6. Average power performance versus release date

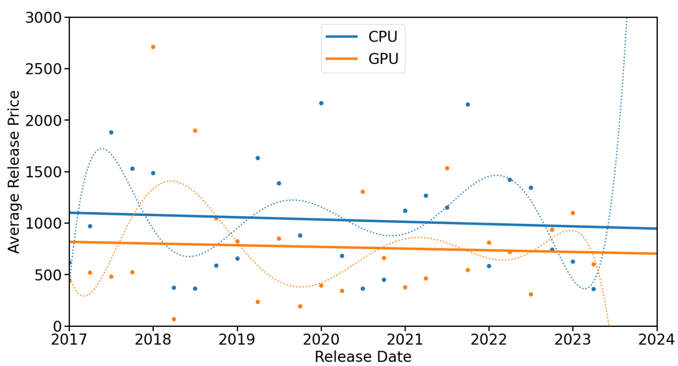

Figure 7. Average release price versus release date

polynomial regressions suggest that CPU's release price will increase while GPU's decrease. The R^2 values for the polynomial regression of CPU and GPU are 0.37 and 0.24 respectively.

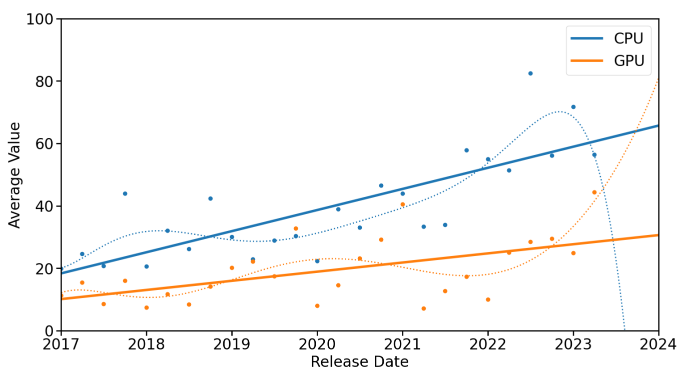

Figure 8. Average value versus release date

As the release price drops and the mark increases, the average value should increase, all else being equal. This is indeed the case as demonstrated by the linear regressions in figure 8. The polynomial regressions contradictorily present that CPU values will decrease while GPU value's rate of change is even higher than it seems. The R^2 values for the polynomial regression of CPU and GPU are 0.76 and 0.48 respectively.

V. Discussion

A. When will the shortage end?

Our findings show that we can make an educated guess that the shortage had already ended in Q3 2022. However, applying polynomial regression gives us a different result, i.e., the price of CPU will continue to drop rather than stabilize while that of GPU will skyrocket.

Unfortunately, as the pricing history is rather chaotic with no real pattern, our regression models are unreliable at predicting even the nearest future regardless of their R^2 values. Thus, it is evidently better to rely on our educated guess instead which concludes that the shortage ended in Q3 2022.

However, as our historical analysis of the shortage might suggest, the processor shortage was not caused by a single event, but rather a series of them. If this series of events continues, the chip shortage will likely reemerge. Ultimately, this result should be taken with a grain of salt.

Thus, even if it is highly likely that the shortage has come to an end, chip manufacturers should still err on the side of caution. In some senses, the chip shortage was caused by the unreadiness of manufacturers in adapting to global crises. As such, measures should be taken now to ensure such a catastrophe will not happen in the future.

B. What is the next big trend in the development of chips?

Once again, as the patterns in the metrics are very random, polynomial regression did not yield any reasonably good results despite however high its R^2 score is. Therefore, we can only rely on our results from the linear regressions to draw a conclusion.

Generally, despite the shortage, every metric of processors, CPU and GPU alike, evolved in a positive way from before the shortage. We found no evidence that supports the fact that the shortage had changed the processor industry in any way or form.

This could imply that the chip manufacturers did not combat the shortage by adapting the chips to the consumer's needs but rather by some other methods.

C. Limitations

One of the biggest limitations of this study is simply the lack of datasets of processors that are up to date, let alone ones with pricing history. Due to the lack of data, we were not able to analyze the data categorically which could be the reason why the patterns in the data seemed so random.

Another limitation is that polynomial regression did not seem to be very effective at predicting the everchanging price in the processor market or even the metrics of processors in the short term.

D. Future research

Future research on this topic should try to acquire a broader dataset before going any further. Moreover, a better regression model should be used for prediction as a polynomial model does not seem to be a good match for this type of problem.

VI. CONCLUSION

Our research found that even though it is highly likely that the shortage has come to an end, chip manufacturers should still be cautious in order to avoid such a catastrophe in the future.

What is more, despite the shortage, every single metric of processors evolved in a positive way from before the shortage. This could imply that the chip manufacturers did not combat the shortage by adapting the chips to the consumer's needs but rather by some other methods.

Future research on this topic should try to acquire a broader dataset. Furthermore, a better regression model should be used for prediction.

VII. ACKNOWLEDGMENTS

I would like to express my greatest gratitude to my supervisor, Doctor Nguyen Ngoc Thao, for not only providing crucial feedback at the start of the research but also guiding us throughout our endeavor, ultimately shaping this research up to what it is today.

I am also grateful to my teammates who greatly supported me during this journey. Without their running me errands and their occasional burst of good ideas, I would not have been able to finish this research within the confine of the deadline. Though, I would appreciate it more had they provided me with more moral support.

Lastly, I would also like to mention some of my classmates whose ideas I ~~stole~~ borrowed for this research.

REFERENCES

[1] Dave James. Average graphics card prices have halved since the start of 2022, 2022. Last accessed 20 April 2023.

[2] Wikipedia. 2020–present global chip shortage, 2023. Last accessed 20 April 2023.

[3] Hannah Casper, Autumn Rexford, David Riegel, Amanda Robinson, Emily Martin, and Mohamed Awwad. The impact of the computer chip supply shortage. In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Bangalore, India, 2021.

[4] Xiling Wu, Caihua Zhang, and Wei Du. An analysis on the crisis of ”chips shortage” in automobile industry. *Journal of Physics: Conference Series*, 1971, 2021.

[5] Wassen Mohammad, Adel Elomri, and Laoucine Kerbache. The global semiconductor chip shortage: Causes, implications, and potential remedies. *IFAC-PapersOnLine*, 55(10):476–483, 2022.

[6] Chung-Huang Yeh and Jwu E Chen. Recycling test methods to improve test capacity and increase chip shipments. *IEEE Design and Test*, pages 1–1, 2022.

[7] Jennifer Chandler. When the supply chain breaks: strategies for the chip shortage. *Journal of Business Strategy*, 2022.

[8] Vinay Ramani, Debabrata Ghosh, and ManMohan Sodhi. Understanding systemic disruption from the covid-19-induced semiconductor shortage for the auto industry. *Omega*, 113(10):102720, 2022.

[9] Duncan Stewart, Dan Hamling, Ariane Bucaille, and Gillian Crossan. My kingdom for a chip: The semiconductor shortage extends into 2022. *Technology, Media, and Telecommunications Predictions*, 2022.

[10] Kristin Dziczek. Why the automotive chip crisis isn't over (yet), 2022. Last accessed 20 April 2023.

[11] Stephen M. Rothrock. Coronavirus, chip boom, and supply shortage: The new normal for global semiconductor manufacturing. *International Symposium on Microelectronics*, 2021(1):26–30, 2021.

[12] Jany George, Letha Janaki, and Jairaj Gomathy. Statistical downscaling using local polynomial regression for rainfall predictions – a case study. *Water Resources Management*, 30:183—-193, 2016.

[13] Debanjan Parbat and Monisha Chakraborty. A python based support vector regression model for prediction of covid19 cases in india. *Chaos, Solitons and Fractals*, 138(10):109942, 2020.

[14] PassMark SOFTWARE. Cpu mega list, 2023. Last accessed 25 April 2023.

[15] PassMark SOFTWARE. Video card list, 2023. Last accessed 25 April 2023.