Final project
Big Data Physics

Use linear regression to analyze the number of transistor in processor

Quentin Davillé (12224808)

Index

- Find and process data
- Calculate line of regression
- Moore law
- Conclusion

Find and process data

Data source : https://www.kaggle.com/datasets/michaelbryantds/cpu-and-gpu-product-data?resource=download

	Unnamed: 0	Product	Type	Release Date	Process Size (nm)	TDP (W)	Die Size (mm^2)	Transistors (million)	Freq (MHz)	Foundry	Vendor
0	0	AMD Athlon 64 3500+	CPU	2007-02-20	65.0	45.0	77.0	122.0	2200.0	Unknown	AMD
1	1	AMD Athlon 200GE	CPU	2018-09-06	14.0	35.0	192.0	4800.0	3200.0	Unknown	AMD
2	2	Intel Core i5-1145G7	CPU	2020-09-02	10.0	28.0	NaN	NaN	2600.0	Intel	Intel
3	3	Intel Xeon E5-2603 v2	CPU	2013-09-01	22.0	80.0	160.0	1400.0	1800.0	Intel	Intel
4	4	AMD Phenom II X4 980 BE	CPU	2011-05-03	45.0	125.0	258.0	758.0	3700.0	Unknown	AMD
...	...	...	...	...	...	...	...	...	...	...	...
4849	4849	NVIDIA Quadro 3000M	GPU	2011-02-22	40.0	75.0	332.0	1950.0	450.0	TSMC	NVIDIA
4850	4850	Intel GMA 950	GPU	2005-06-01	90.0	7.0	NaN	NaN	250.0	Intel	Intel
4851	4851	NVIDIA GeForce GT 320M	GPU	2010-03-03	40.0	23.0	100.0	486.0	500.0	TSMC	NVIDIA
4852	4852	NVIDIA GeForce FX 5200	GPU	2003-03-06	150.0	NaN	65.0	29.0	250.0	TSMC	NVIDIA
4853	4853	NVIDIA GeForce 9300 SE	GPU	2008-06-01	65.0	NaN	86.0	210.0	540.0	TSMC	NVIDIA

Find and process data

```
data["Release Date"] = pd.to_datetime(data["Release Date"])
data=data[data["Type"] == "CPU"].reset_index(drop=True) # Only CPU
data = data.dropna(subset=['Release Date']) # Remove CPU without Release date
data = data.dropna(subset=['Transistors (million)']) # Remove CPU without transistor count
data.sort_values(by='Release Date', inplace=True) # Sort the data by Release date
date = data['Release Date']
transistor = data['Transistors (million)']
```

	Unnamed: 0	Product	Type	Release Date	Process Size (nm)	TDP (W)	Die Size (mm^2)	Transistors (million)	Freq (MHz)	Foundry	Vendor
0	0	AMD Athlon 64 3500+	CPU	2007-02-20	65.0	45.0	77.0	122.0	2200.0	Unknown	AMD
1	1	AMD Athlon 200GE	CPU	2018-09-06	14.0	35.0	192.0	4800.0	3200.0	Unknown	AMD
2	2	Intel Core i5-1145G7	CPU	2020-09-02	10.0	28.0	NaN	NaN	2600.0	Intel	Intel
3	3	Intel Xeon E5-2603 v2	CPU	2013-09-01	22.0	80.0	160.0	1400.0	1800.0	Intel	Intel
4	4	AMD Phenom II X4 980 BE	CPU	2011-05-03	45.0	125.0	258.0	758.0	3700.0	Unknown	AMD
...	...	...	...	...	...	...	...	...	...	...	...
4849	4849	NVIDIA Quadro 3000M	GPU	2011-02-22	40.0	75.0	332.0	1950.0	450.0	TSMC	NVIDIA
4850	4850	Intel GMA 950	GPU	2005-06-01	90.0	7.0	NaN	NaN	250.0	Intel	Intel
4851	4851	NVIDIA GeForce GT 320M	GPU	2010-03-03	40.0	23.0	100.0	486.0	500.0	TSMC	NVIDIA
4852	4852	NVIDIA GeForce FX 5200	GPU	2003-03-06	150.0	NaN	65.0	29.0	250.0	TSMC	NVIDIA
4853	4853	NVIDIA GeForce 9300 SE	GPU	2008-06-01	65.0	NaN	86.0	210.0	540.0	TSMC	NVIDIA

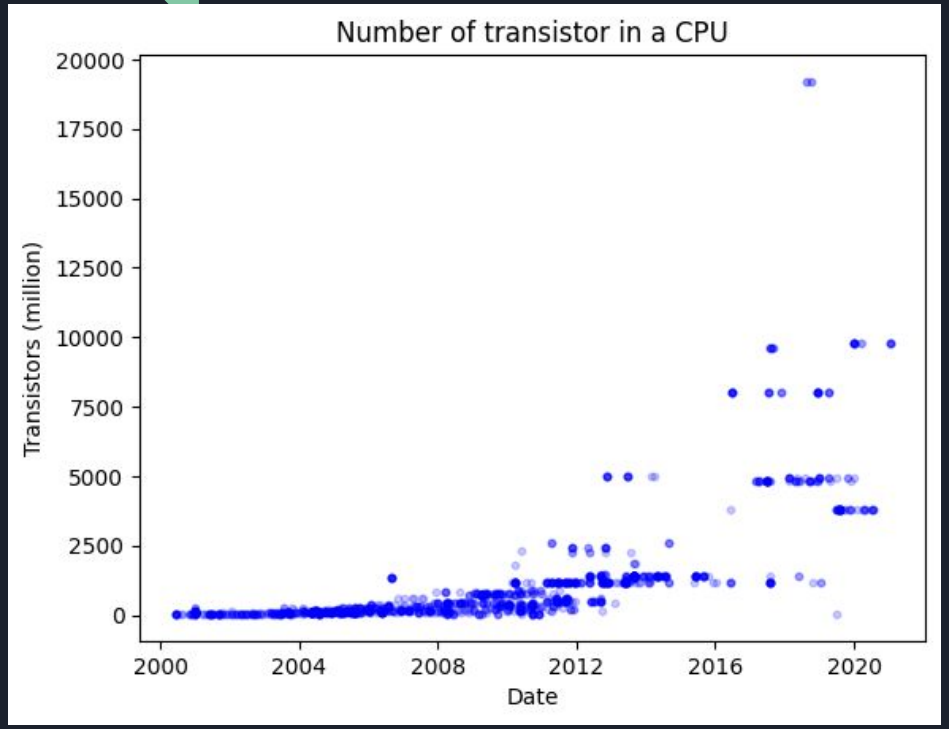
Transistor

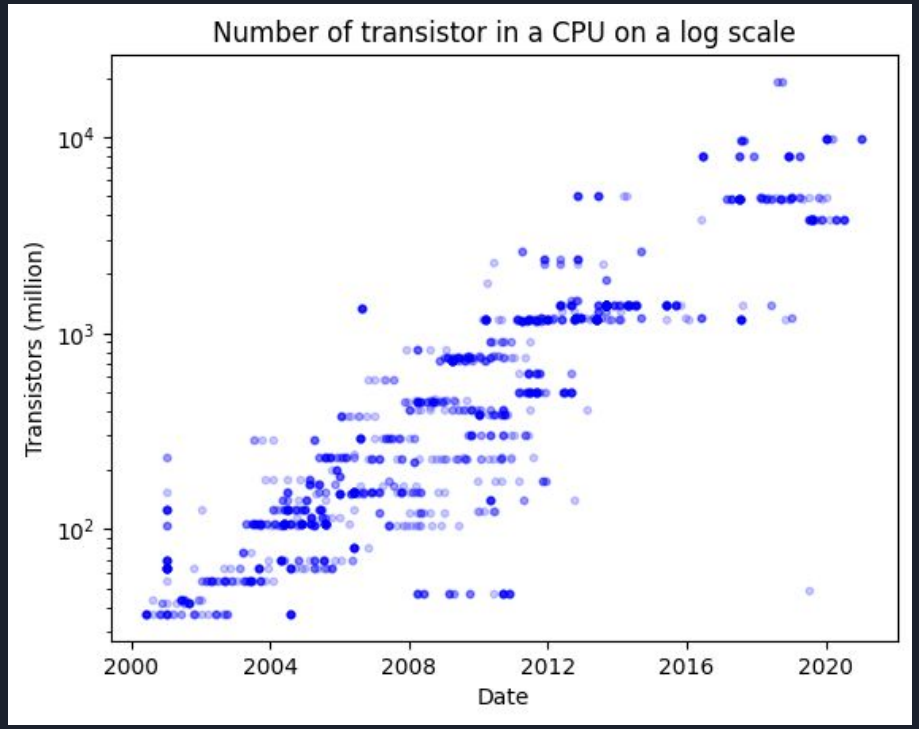
```
0        37.0
1        37.0
2        37.0
3        37.0
4        37.0
        ...
1672    9800.0
1673    9800.0
1670    9800.0
1671    9800.0
1674    9800.0
```

Date

```
0     2000-06-05
1     2000-06-05
2     2000-06-05
3     2000-06-05
4     2000-06-05
         ...
1672  2021-01-12
1673  2021-01-12
1670  2021-01-12
1671  2021-01-12
1674  2021-01-12
```

Find and process data





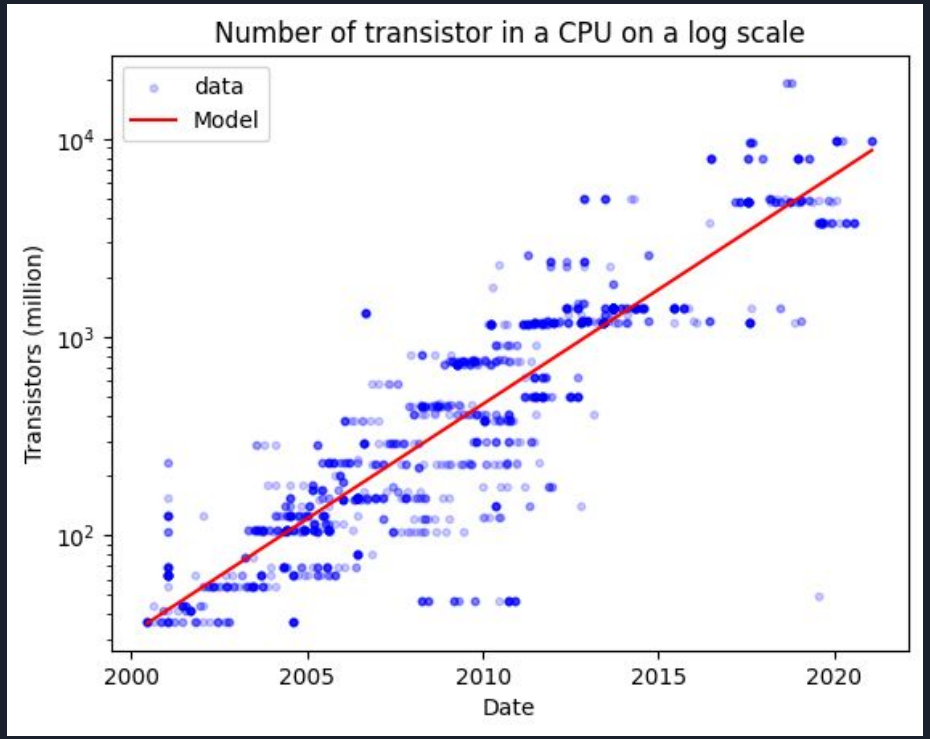
Calculate line of regression

```
oneYear = 365*24*60*60*1000000000 # One year in nanoseconds
x = (pd.to_datetime(date).astype(np.int64) / (oneYear * 2)).astype(np.float64)
y = transistor
res = linregress(x,np.log(y))
slope = res.slope
intercept = res.intercept
print("Slope:", slope)
print("Real slope:", np.exp(slope))
mooreLawSlope = 2
logMooreLawSlope = np.log(mooreLawSlope)
print("Moore law slope:", mooreLawSlope)
print("Moore law slope on a log scale:", logMooreLawSlope)
```

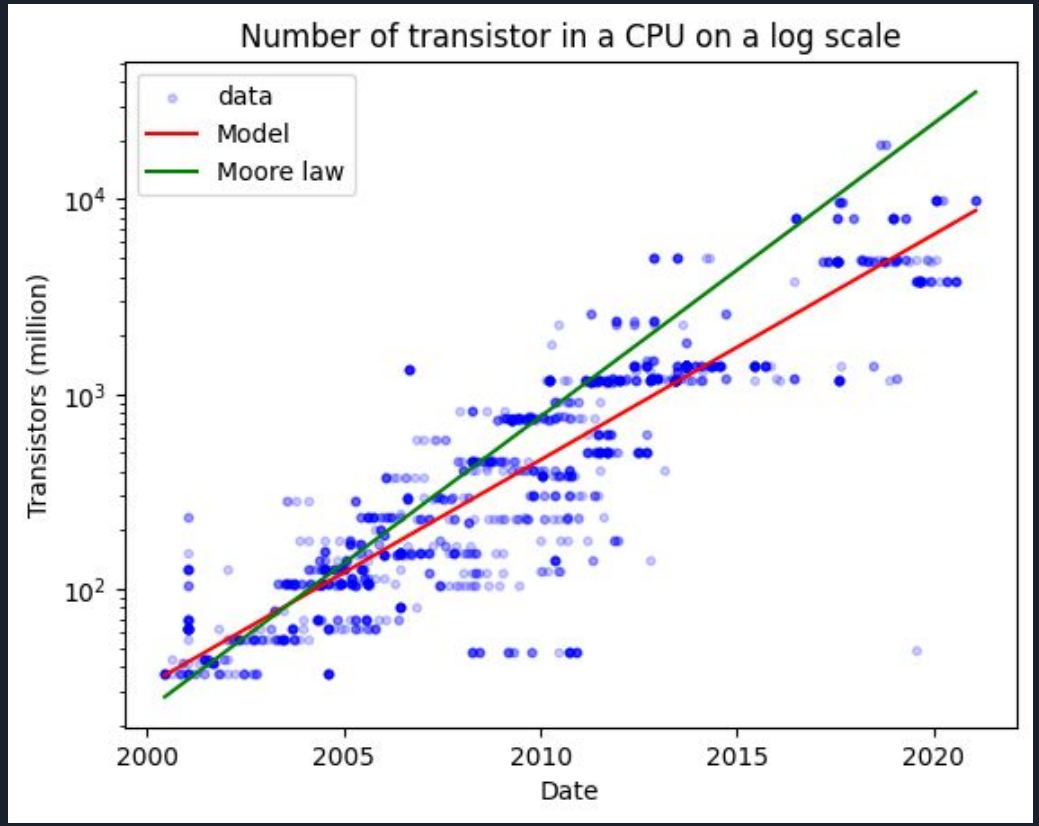
```
Slope: 0.5327481465996401
Real slope: 1.7036076449778932
Moore law slope: 2
Moore law slope on a log scale: 0.6931471805599453
```

Calculate line of regression

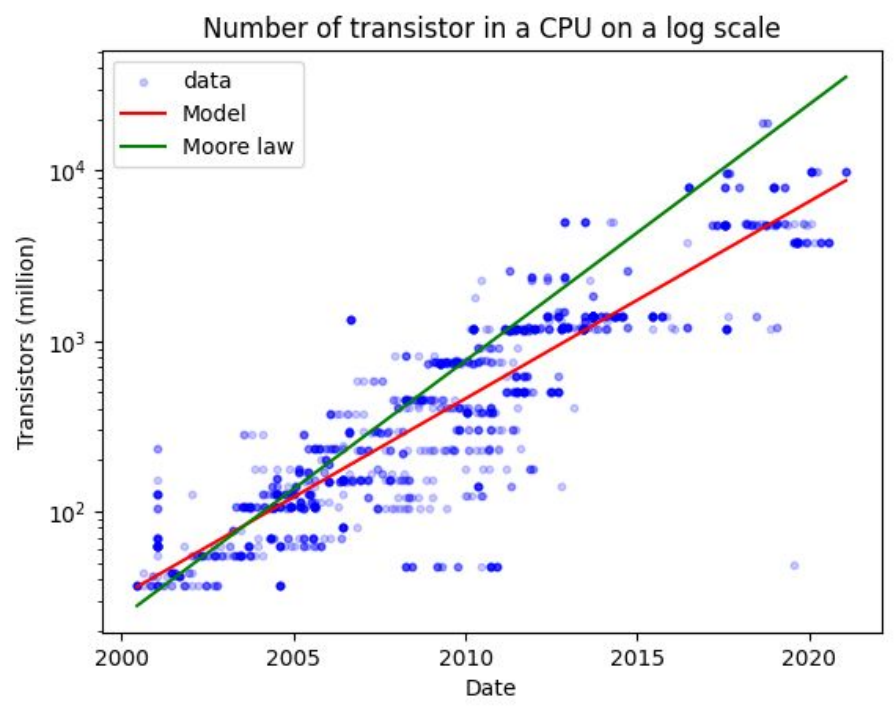
```
model=lambda x: np.exp(x * slope + intercept)
```

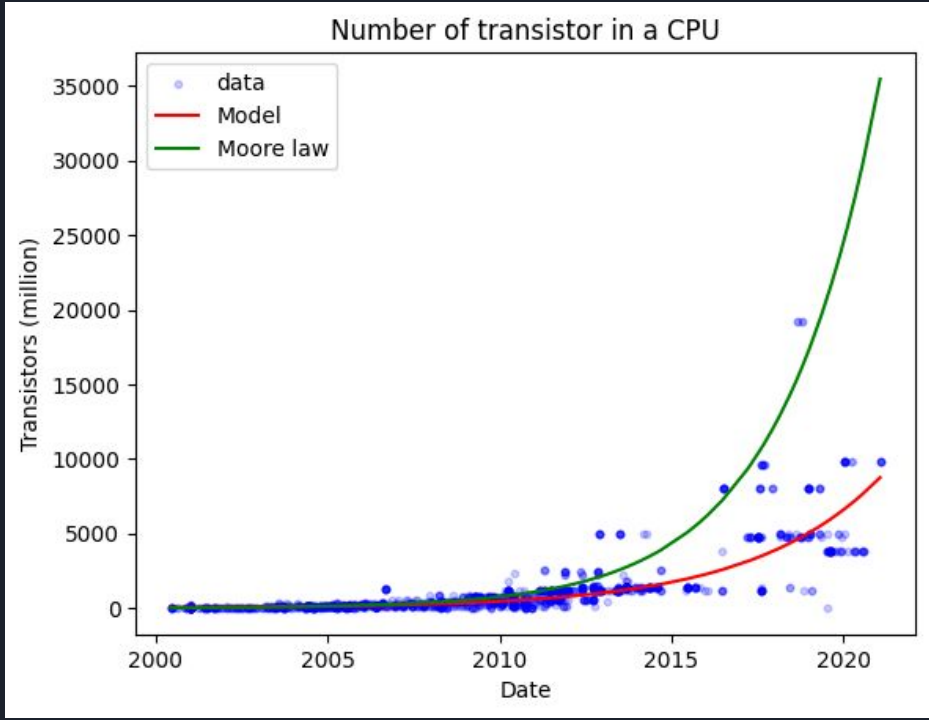


Conclusion


Number of transistor in a CPU on a log scale
data
Model
Moore law
Transistors (million)
10^4
10^3
10^2
2000
2005
2010
2015
2020
Date

Conclusion

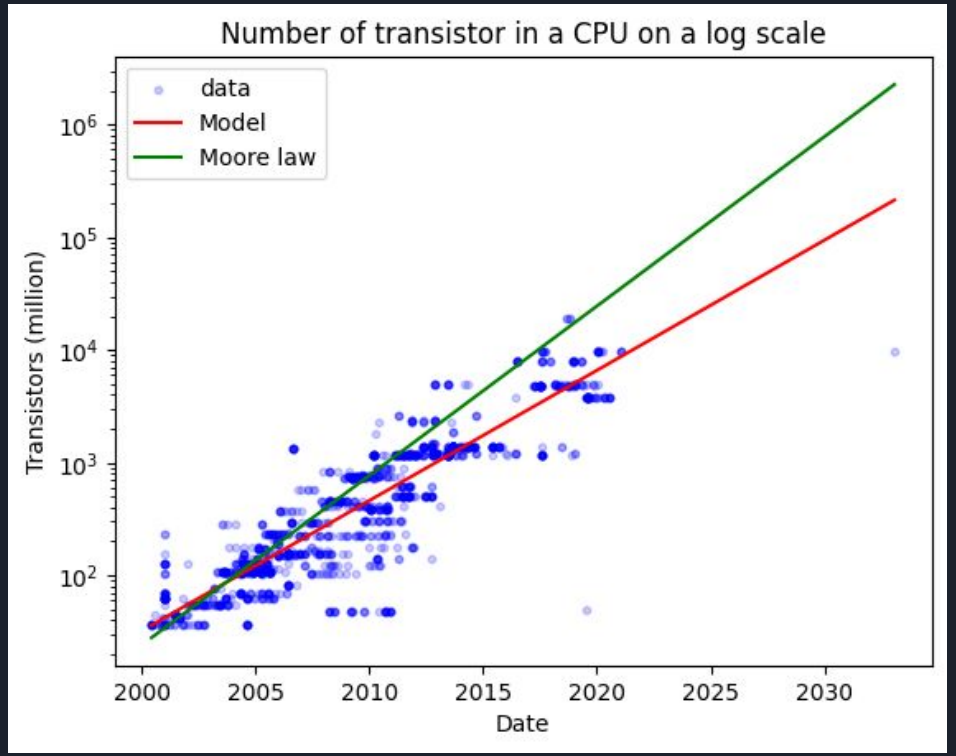

Number of transistor in a CPU on a log scale
data
Model
Moore law
Transistors (million)
10^4
10^3
10^2
2000
2005
2010
2015
2020
Date


Number of transistor in a CPU
data
Model
Moore law
Transistors (million)
35000
30000
25000
20000
15000
10000
5000
0
2000
2005
2010
2015
2020
Date

Conclusion

Average number of transistor in 10 years predicted by moore law: 2269 billions

Average number of transistor in 10 years predicted your model: 213 billions



Thank you

Python source file: https://github.com/D-Quentin/BDF-Final

Dataset source : https://www.kaggle.com/datasets/michaelbryantds/cpu-and-gpu-product-data?resource=download