FLOPs : floating-point operations

FLOP/s : floating point operations per second

FLOP/s :

$$C = 2 \times N_c \times R \times D \times E$$

N_c : number of connections

between neurons in an unrolled NN

$$R = \frac{\text{Operations per backward pass}}{\text{Operations per forward pass}} \approx 2$$

(backward-forward FLOP ratio)

D : number of training examples

E : number of training epochs

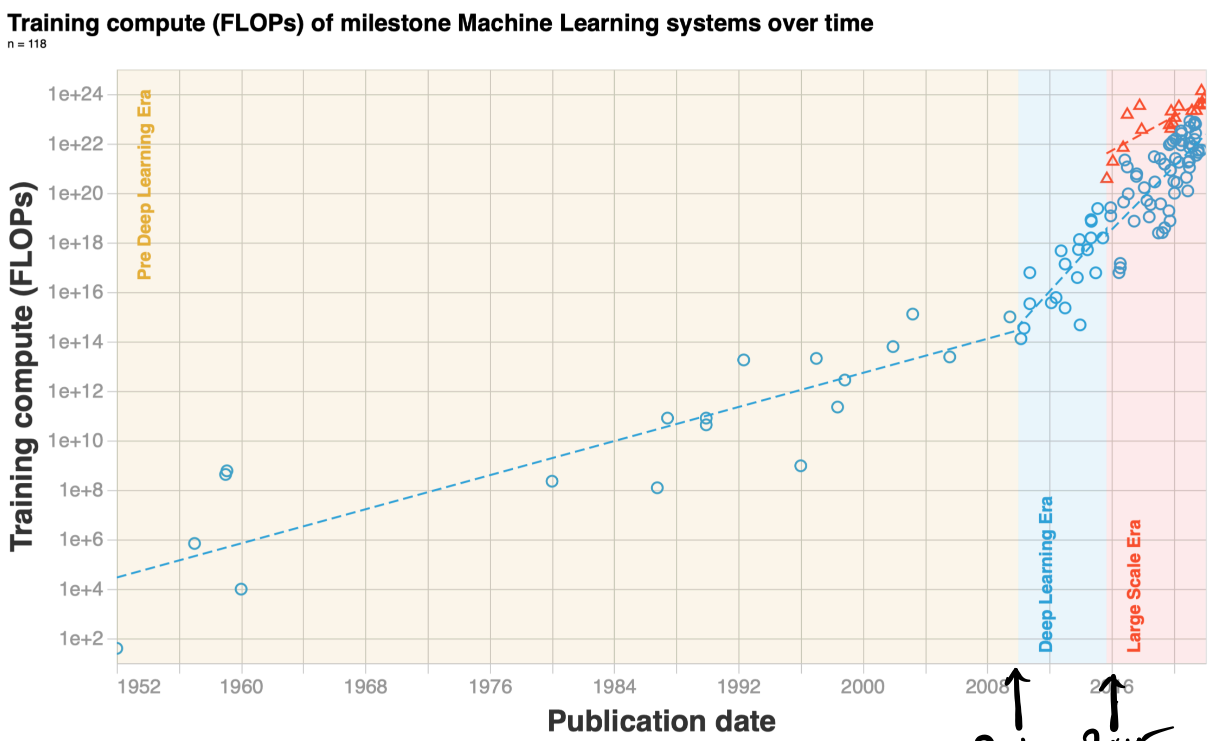
Moore's Law:

transistor density doubles roughly every two years

Period	Data	Scale (start to end)	Slope	Doubling time
1952 to 2010 Pre Deep Learning Era	No low outliers ($n = 19$)	3e+04 to 2e+14 FLOPs	0.2 OOMs/year [0.1; 0.2; 0.2]	21.3 months [17.0; 21.2; 29.3]
2010 to 2022 Deep Learning Era	No outliers ($n = 80$)	7e+14 to 2e+18 FLOPs	0.6 OOMs/year [0.4; 0.7; 0.9]	5.7 months [4.3; 5.6; 9.0]
September 2015 to 2022 Large-Scale Era	High outliers ($n = 19$)	4e+21 to 8e+23 FLOPs	0.4 OOMs/year [0.2; 0.4; 0.5]	9.9 months [7.7; 10.1; 17.1]

TABLE I: Summary of our main results. In 2010 the trend accelerated along the with the popularity of Deep Learning, and in late 2015 a new trend of large-scale models emerged.

Log-linear trends



2010 2015

Fig. 1: Trends in $n = 118$ milestone ML models between 1952 and 2022. We distinguish three eras. Notice the change of slope circa 2010, matching the advent of Deep Learning; and the emergence of a new large-scale trend in late 2015.

AlexNet 2012 → 2012 ImageNet Competition

Period	Outliers	Scale (FLOPs)	Slope	Doubling time	R^2
1952-2009	All models ($n = 19$)	3e+04 / 2e+14	0.2 OOMs/year [0.1; 0.2; 0.2]	21.3 months [16.2; 21.3; 31.3]	0.77
1952-2011	All models ($n = 26$)	1e+04 / 3e+15	0.2 OOMs/year [0.1; 0.2; 0.2]	19.6 months [15.6; 19.4; 25.0]	0.83
2010-2022	All models ($n = 98$)	1e+15 / 6e+22	0.7 OOMs/year [0.6; 0.7; 0.7]	5.6 months [5.0; 5.6; 6.2]	0.70
	Regular-scale ($n = 77$)	4e+14 / 2e+22	0.7 OOMs/year [0.6; 0.7; 0.7]	5.6 months [5.1; 5.6; 6.2]	0.78
2012-2022	All models ($n = 91$)	1e+17 / 6e+22	0.6 OOMs/year [0.5; 0.6; 0.7]	5.7 months [4.9; 5.7; 6.7]	0.58
	Regular-scale ($n = 80$)	4e+16 / 2e+22	0.6 OOMs/year [0.5; 0.6; 0.7]	5.7 months [4.9; 5.7; 6.7]	0.69

TABLE II: Log-linear regression results for ML models from 1952 to 2022.

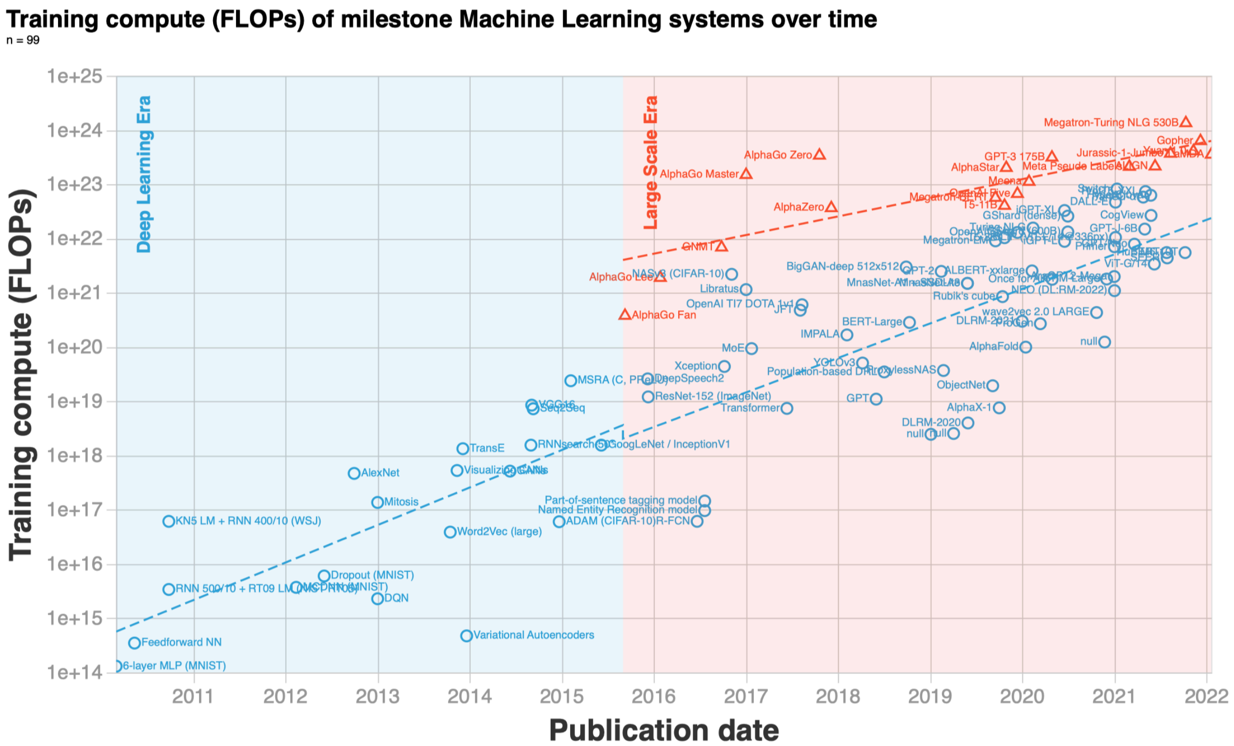


Fig. 2: Trends in training compute of $n = 99$ milestone ML systems between 2010 and 2022. Notice the emergence of a possible new trend of large-scale models around 2016. The trend in the remaining models stays the same before and after 2016.

2009 - 2012 Speech Recognition

2010