Exploring Trends in Microprocessor Design: Does Moore's Law have a cousin?

Prepared by Jason Glass

Northwestern University
FINAL REPORT EECS 452

List of Figures

List of Tables

Abstract:

This paper follows the analysis of 20 years of microprocessor design and performs regression to determine exponential and linear design trends. This paper discusses the existing laws and their limits. This paper applies linear and exponential fits to several measured and derived computational parameters such as number of cores, power usage and clock frequency.

This paper determines that at least 5 parameters exist with an R^2 over 0.8 and 4 over 0.9. These fits are then used to make predictions about the state of computation in the coming years.

*Lastly this paper combines several parameters to introduce the scaling factors of (Max Cores * Average Frequency) and (Max Cores * Average Frequency)/TDP to model theoretical computational power and computational efficiency of CMP systems in the coming years. These factors have R^2 respectively of 0.9533 and 0. 9362.*

Introduction:

Moore's Law is considered one of the fundamental exponential laws of computing. It states that every 18 months (the exact timeframe varies) the number of transistors per microprocessor will double and thus computing power will roughly double. This number has varied from 24 months to 18 months to 13 months in the modern iterations of the law. Moore's Law however is not the only exponential, or proposed, exponential law of computing. The answer to the titular question is in fact yes.

Other exponential laws such as Dennard Scaling stated that performance per watt scales at roughly the same rate as Moore's Law (1). This law however broke down in 2006 due to arrival at the power gate. The breakdown of Dennard scaling directly contributed to the increase in CMP systems versus the traditional SMP systems that dominated computing for decades.

Koomey's Law states that computations per joule of energy has been doubling approximately every 1.57 years, or roughly in accord with Moore's Law (2). Interestingly in 2048 a predicted breakdown of Koomey's Law has been noted due to the arrival of the Thermodynamic-Computational concept of the Landauer's limit. This limit specifies a minimum amount of energy needed for computation by entropy law and as computational efficiency increases eventually irreversible computing will no longer be sustainable.

This paper explores if any other computing laws may be derived from the analysis of over 1000 intel microprocessors from 2004 to the modern day. My focus was particularly on how the trends CMP design can be observed graphically and statistically. I also address new scaling parameters to compete with Moore's Law and Koomey's Law for CMP systems versus SMP systems.

Related Work

At the time of inception for this paper I was unaware of previous work in the field of data analysis on microprocessor design trends. However, there is one notable paper I found which covers the same work, *Moore's Law* by Domaille et al from Plymouth University (3). This student group conducted almost identical research to my own but on a smaller and more validation scale.

They determined the R^2 value of Moore's law in 2014 using the top 500 supercomputer list referenced in their paper. They ultimately found an experimental R^2 value of over 0.97 for an R_{peak} and R_{max} value which roughly corresponds to computing power. They conclude that computing power

doubles roughly ever 13 months and total number of cores ever 22 months. I will discuss my own results compared to these later.

Methodology

My methodology for this paper was simple. Intel provides a database for all their chips ever released which is conveniently located on Wikipedia.com (4)(5)(6)(7)(8)(9). The bulk of my data preparation was data scrubbing and choose of parameters for analysis.

One parameter that did not make it into the final analysis was notably die size (mm^2). I wanted to include this parameter as an important metric of computation density and other physical constrained parameters but die size information was neither widely nor accurately available for most of Intel's builds.

The data set consists of 1080 data entries of Xeon, i3, i5, i7 and i9 processors from 2004 to 2018 release dates. An original 1500 entries were pruned due to incomplete data such as no price. There are 590 Xeon microprocessor entries, 112 i3 cores, 180 i5 cores, 190 i7 cores, and 5 i9 cores in the data set.

The following parameters were gathered from the data:

- Thermal Dynamic Power (TDP) – Roughly power usage as defined by Intel
- Cores
- Threads
- Release Price
- Frequency

The following parameters were derived from the data:

1. Frequency/Threads
2. Watts/Cycle
3. (Release Price)/TDP (USD/Watts)
4. Threads/ (Release Price)
5. Frequency/ (Release Price) (Ghz/USD)
6. Threads/TDP

Note all derived parameters were also evaluated at the Log10 scale, when log10 scale provided higher coefficient of determination then log was used. All averages were calculated on a yearly basis. For a few parameters adjusted year calculations are shown as the first year, 2004, and the last year, 2018, of data are both partial years and did not always capture trend data.

Parameter 1 highlights the trend of Clock rate compared to number of threads, an analysis of CMP system efficiency. Parameter 2 is simply how many watts are used per Cycle, energy efficiency. Parameter 3 identifies if consumers are paying per processing power over time. Parameter 4 identifies how much consumers are paying per thread. Parameter 5 is like 4 but Frequency per price. Lastly is the Watts used per Thread parameter 6.

After all parameters had been calculated linear regression was performed. For each data set I applied the either an Average, Min, or Max function to the data by year. For example, Max Cores by year would show what the chip with the most number of cores had per year. For each parameter I identified the data transformation that yielded the greatest R^2 value for a linear fit.

After the first step of regression I also applied either an exponential or logarithmic fit to the data depending on the slope of the tonicity.

The results section highlights the most interesting and tangible conclusions from the data analysis. I provide both quantitative and qualitative analysis of results.

Data:

The following sections contains the visualizations of each raw data set for both the measured and derived parameters.

1. Cores
2. Threads
3. Frequency (Ghz)
4. Threads
5. Thermal Dynamic Power – TDP (Watts)
6. Release Price (USD)
7. TDP/Cycle (nW/Cycle)
8. (Release Price)/TDP (USD/W)
9. Frequency/Threads
10. TDP/Threads
11. Threads/ (Release Price)
12. Frequency/ (Release Price) (GHz/USD)
13. Threads/TDP

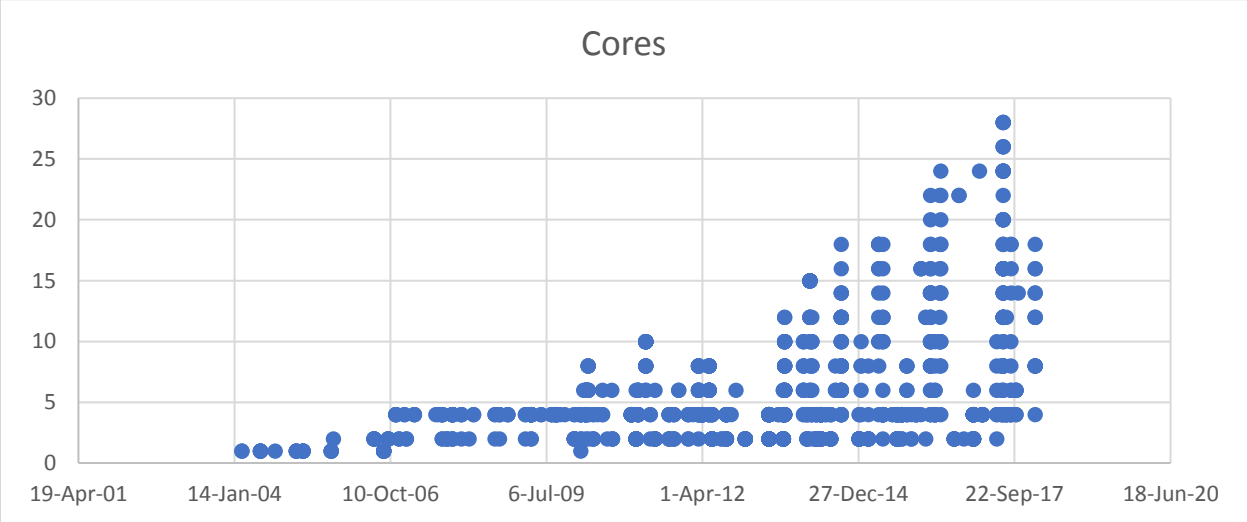

Figure 1 - Cores per Chip

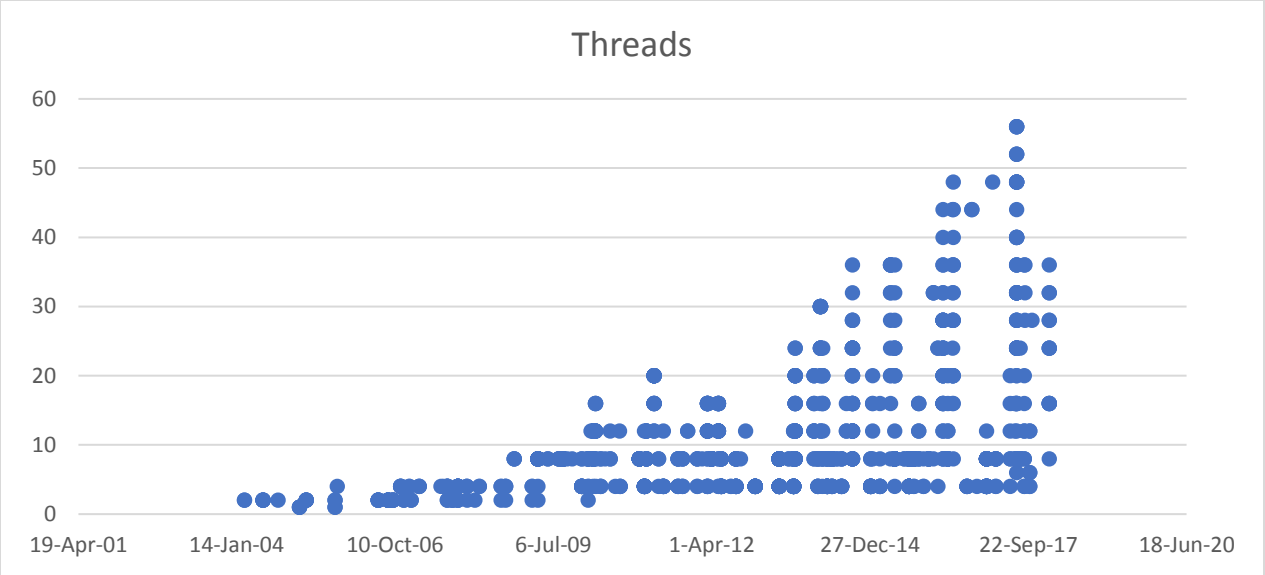

Figure 2 - Threads per Chip

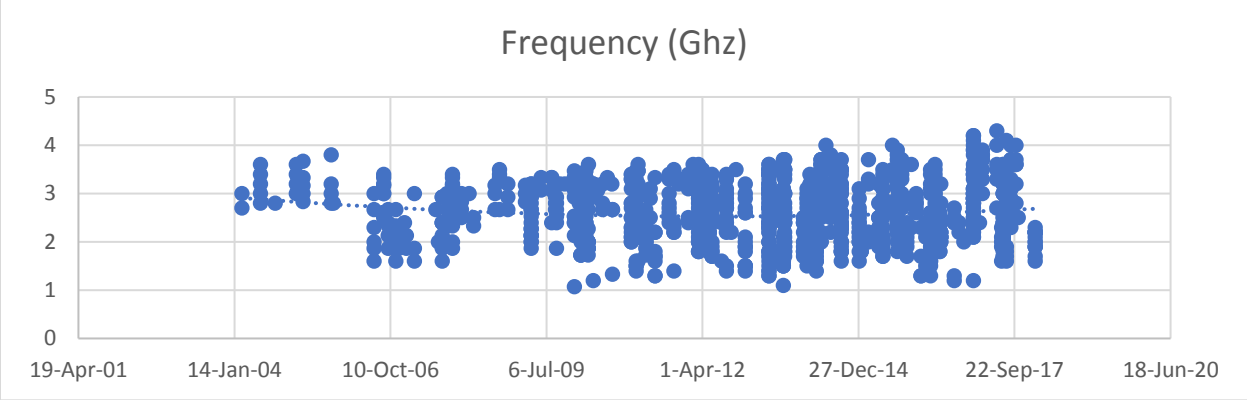

Figure 3 - Clock Frequency of Chip (Ghz)

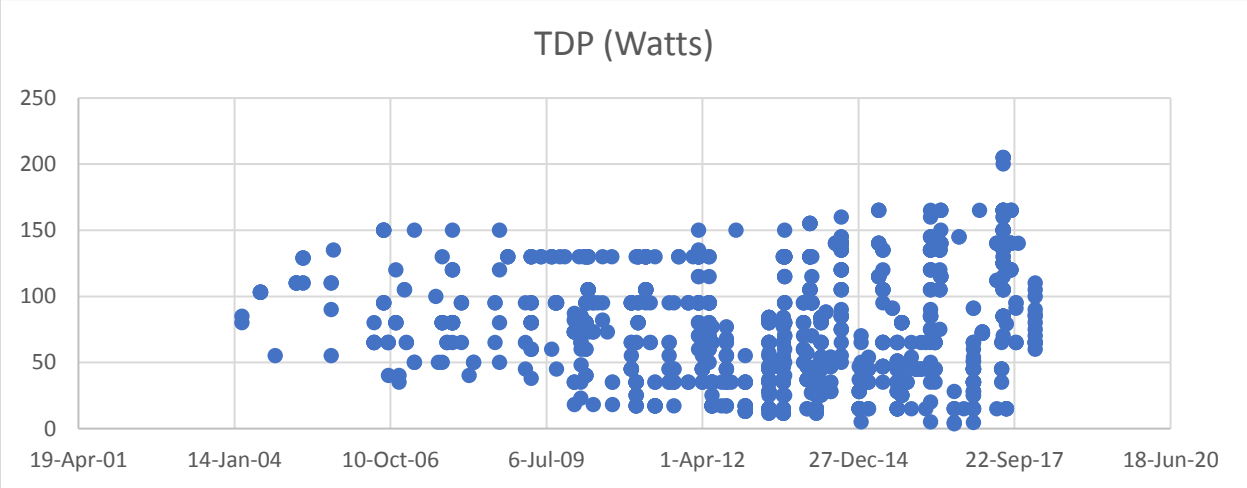

Figure 4 - TDP of Chip (Watts)

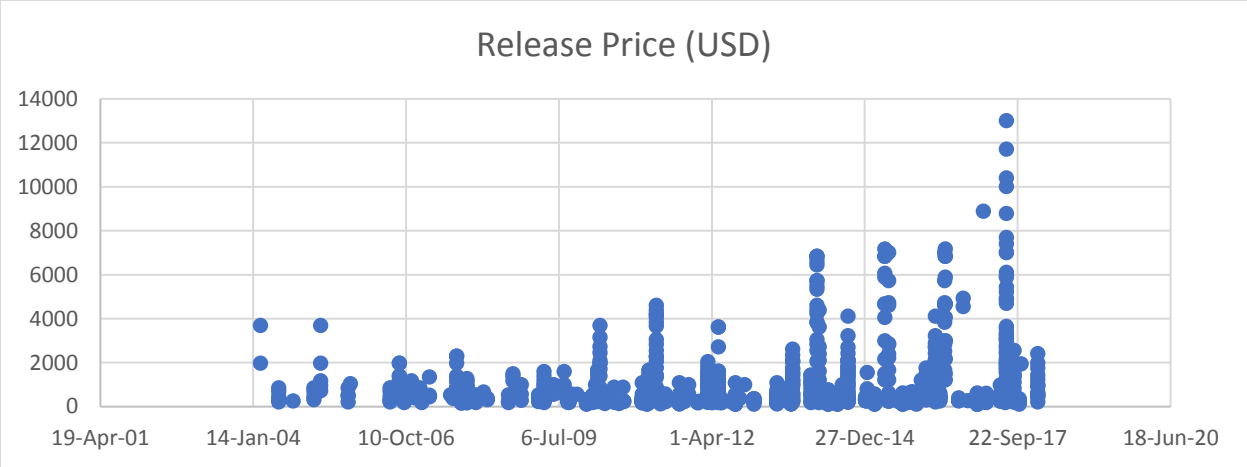

Figure 5 - Release Price of Chip (USD)

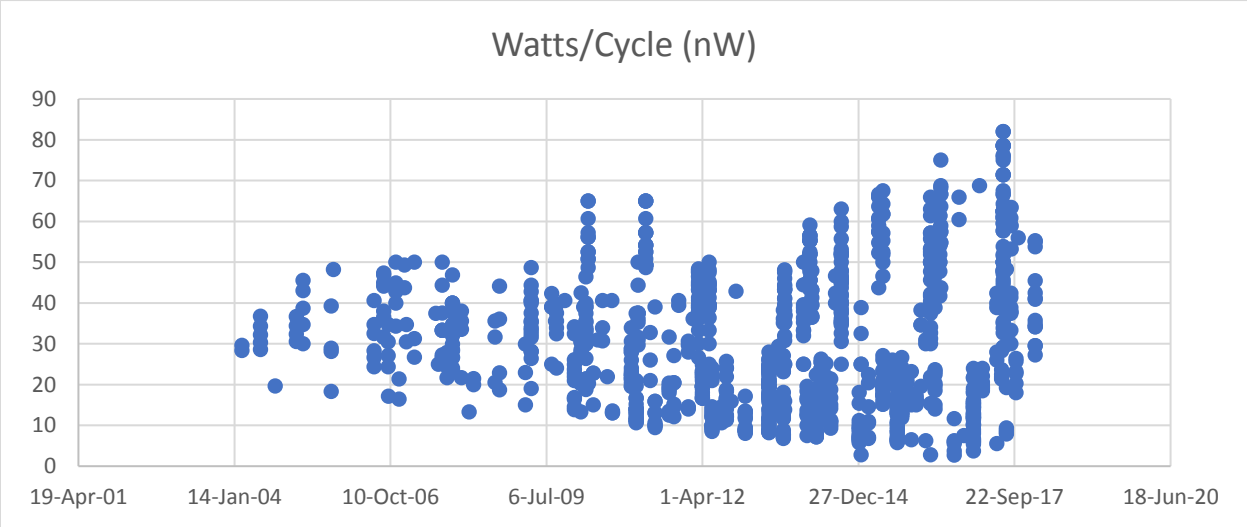

Figure 6 - Watts/Cycle

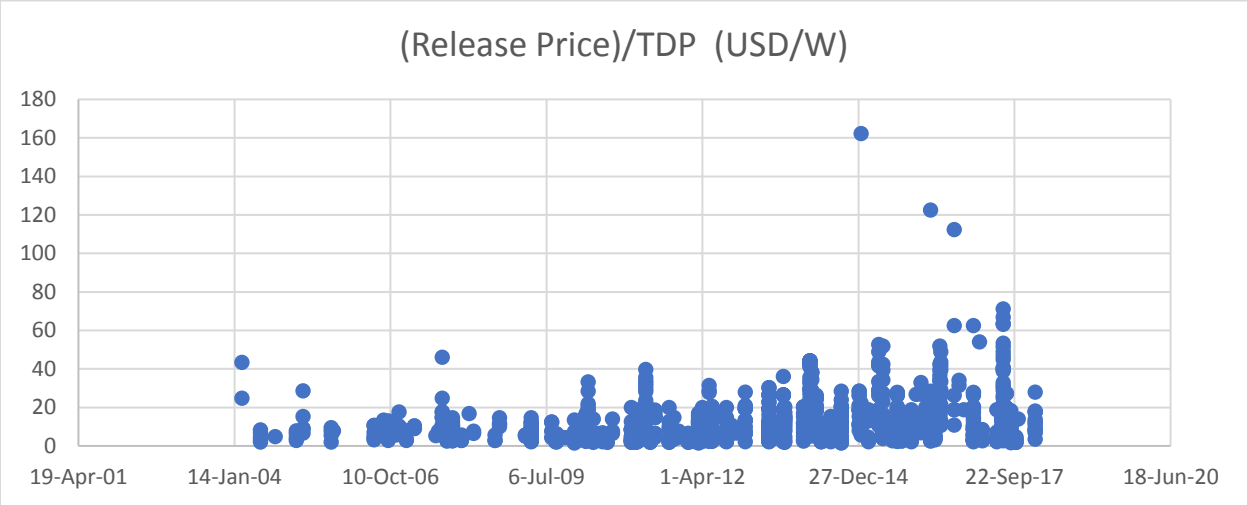

Figure 7 – (Release Price)/TDP (USD/W)

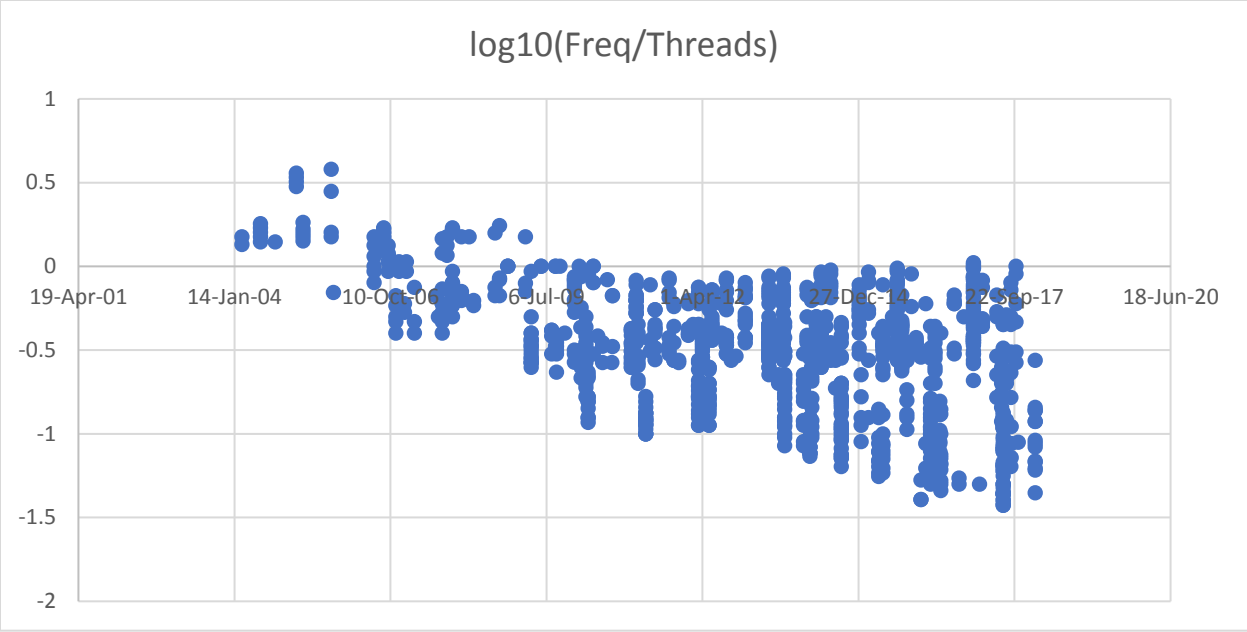

Figure 8 - Freq/Threads (GHz)

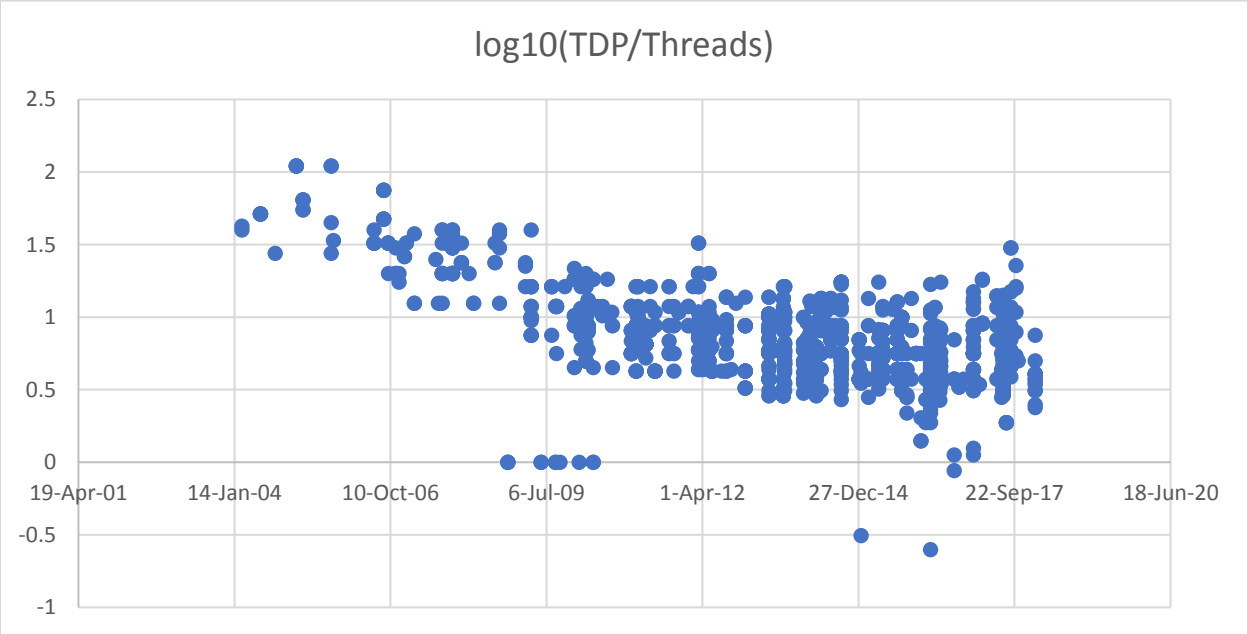

Figure 9 - TDP/Threads (Watts)

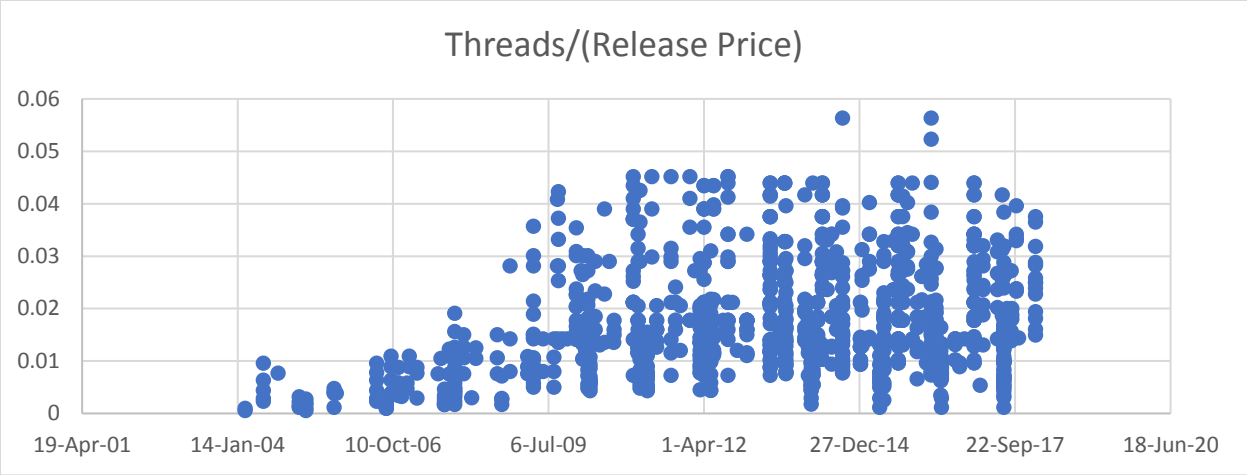

Figure 10 - Threads/ (Release Price)

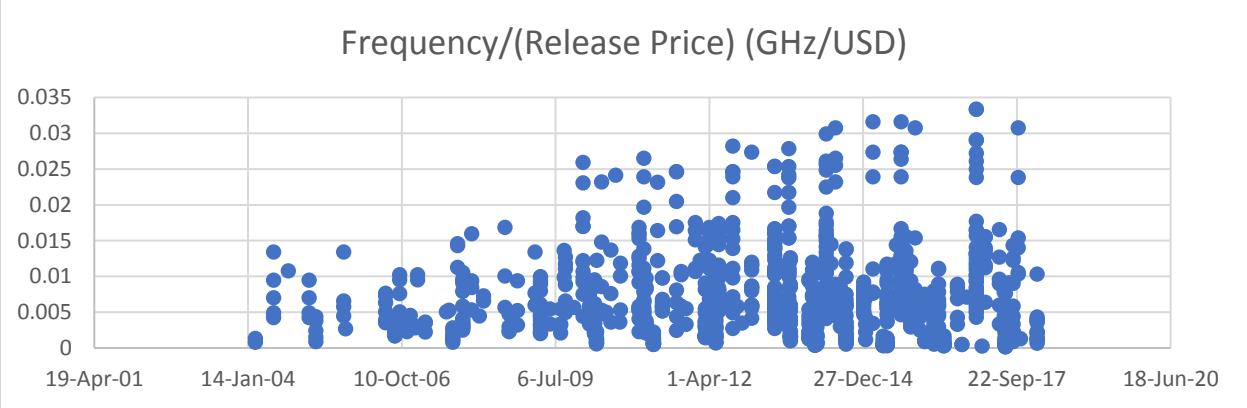

Figure 11 - Frequency/ (Release Price) (GHz/USD)

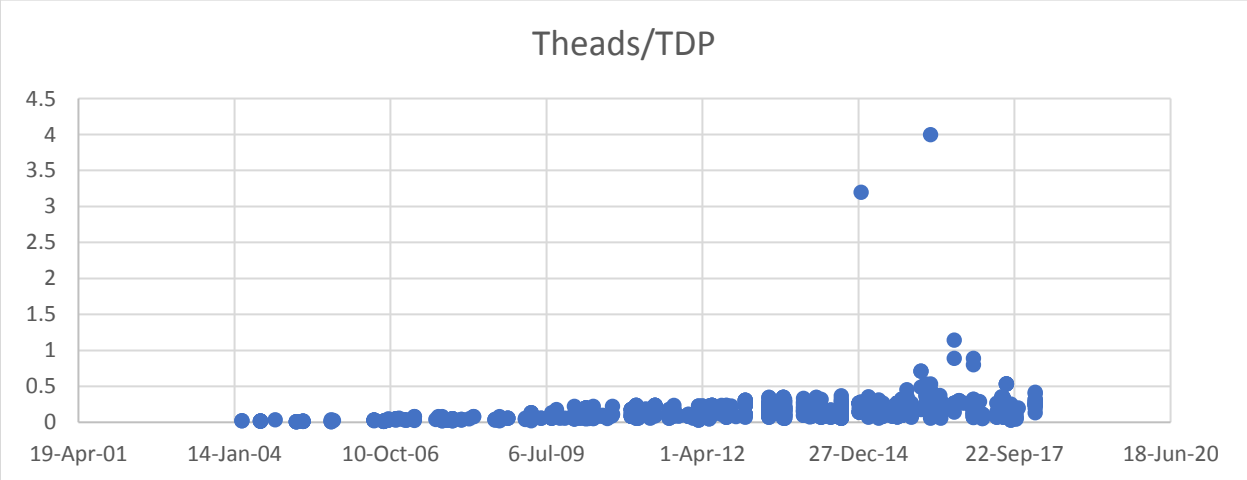

Figure 12 - Threads/TDP

Results:

In this section I will highlight the five parameters I found to be most interesting based on their R^2 value. I will also discuss how the variance of several parameters changed over time. Lastly, I will discuss two exponential scaling laws that I believe provide long term predictive ability in the trends of microprocessor design. I also highlight several exponential fits on parameters that suit the fit as opposed to a linear fit.

Table 1 is a summary of the brute force regression approach I applied to this data set. The 3 transformations I applied on the granularity of *per annum* are Average, Max, and Min. For the Cores and Threads parameters it makes sense to take the Max as the increasing number of cores in CMP systems is of great interest. Meanwhile for more overall trends like Threads/TDP it makes sense to take the average of all chips manufactured that year.

Highlighted in green in each column is the transformation which attainted the highest R^2.

Table 1 - Regression Results - 5 Highest R^2 Parameters

Fit	Cores	Threads	log(Freq/Threads)	Threads/TDP	Threads/Price
Avg	y = 0.631x - 0.1305	y = 1.376x - 1.7962	y = -0.0774x + 0.2388	y = 0.0196x - 0.0262	y = 0.0014x + 0.0032
Avg R^2	0.8281	0.8625	0.8787	0.9157	0.7886
Max	y = 1.7464x - 3.1048	y = 3.6214x - 7.7714	y = -0.0475x + 0.4292	y = 0.1472x - 0.485	y = 0.003x + 0.0108
Max R^2	0.8525	0.8631	0.6867	0.296	0.6682
Min	y = 0.1179x + 0.9238	y = 0.3214x + 0.6952	y = -0.1052x + 0.0081	y = 0.0052x + 0.0024	y = 0.0004x + 0.0003
Min R^2	0.5029	0.7067	0.9433	0.5626	0.2087
R^2	0.8525	0.8631	0.9433	0.9157	0.7886
Exponential or Log Fit	$y = 1.4541e^{0.2063x}$	$y = 2.2185e^{0.2285x}$	y = -0.604ln(x) + 0.2893	$y = 0.0192e^{0.199x}$	y = 0.0084ln(x) - 0.0008
Exp or Log R^2	0.9103	0.9322	0.9498	0.9137	0.8169

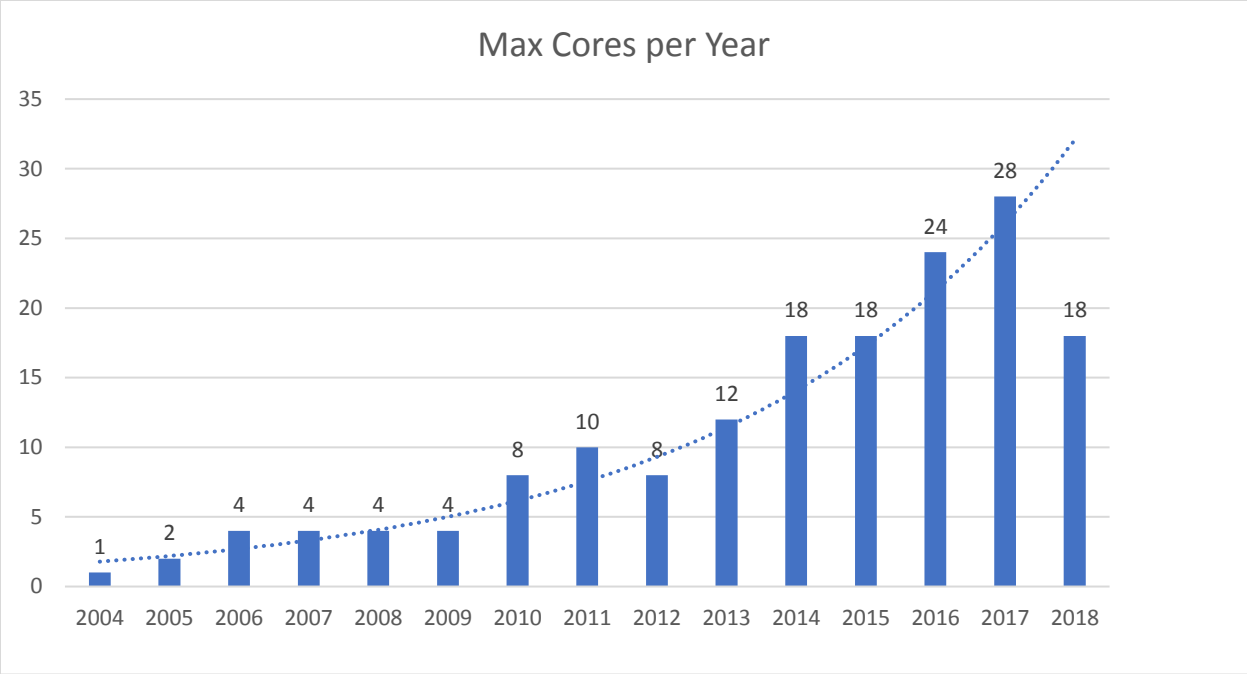

Figure 13 - Max Cores per Year

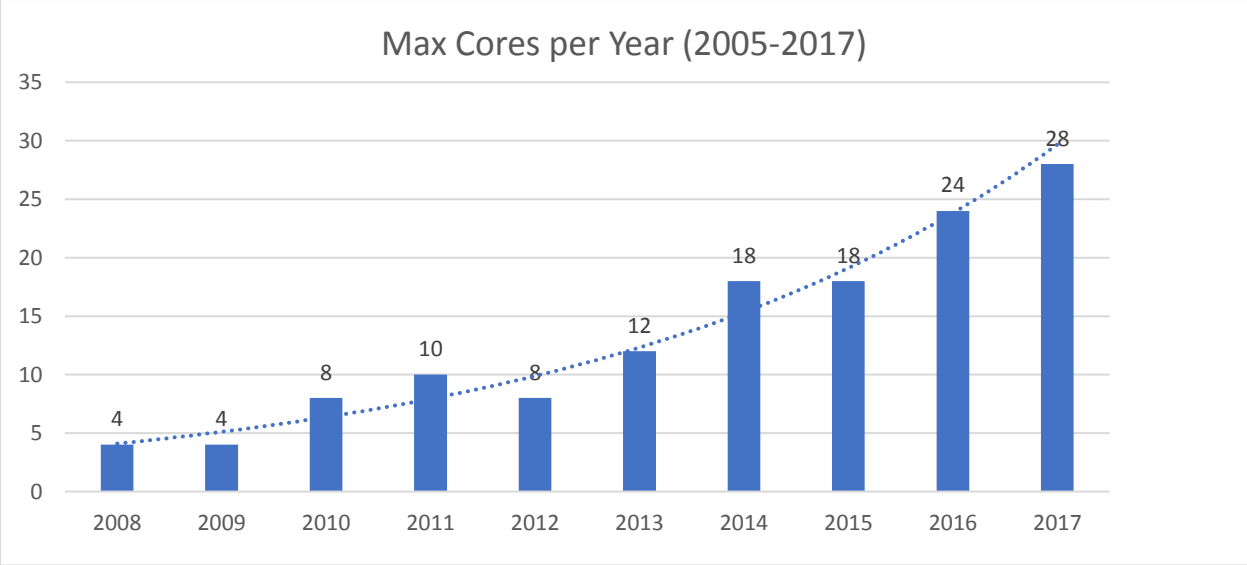

Figure 14 - Max Cores per Year (2005-2017)

After removing the data from 2004 and 2018 the adjusted regression statistics are y = 2.6303x - 1.0667 and R^2 = 0.926. This is an increase of 0.0735 in R^2. This figure makes it clear that the number of cores per system at maximum is roughly doubling every 3 years, this will be discussed further in the **Predictions** section of this report.

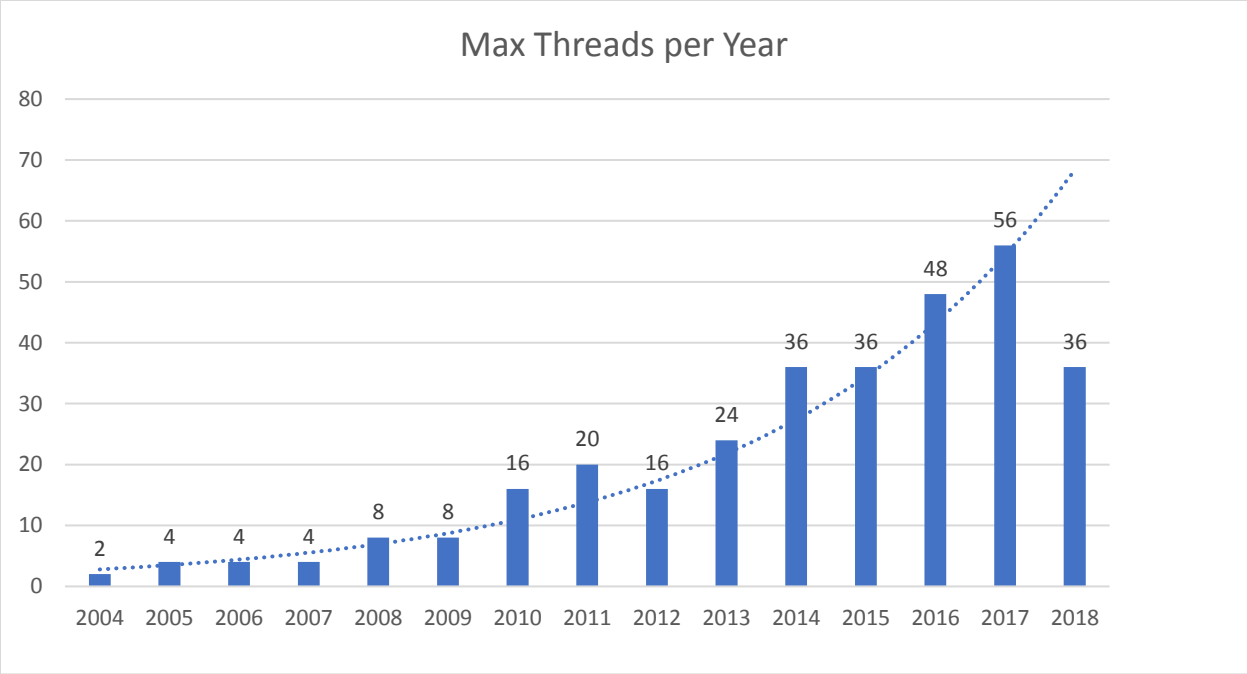

Figure 15 - Max Threads per Year

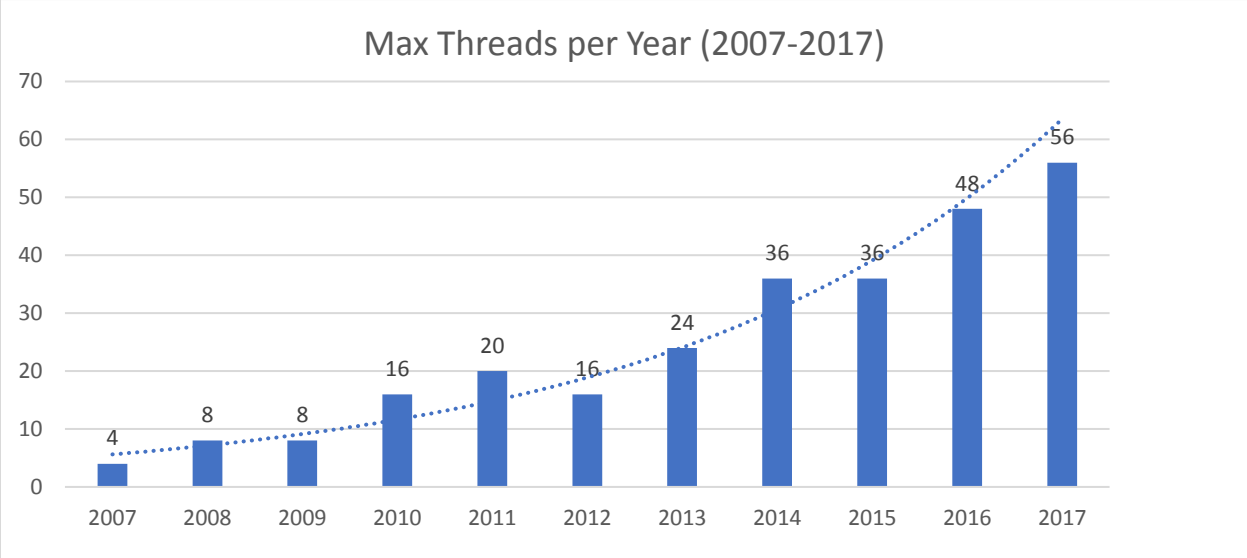

Figure 16 - Max Threads per Year (2007-2017)

After removing the data for 2004,2005,2006 and 2018 the new fit statistics are y = 4.9818x - 5.1636 and R^2 = 0.9292. This is an increase of 0.661. This data matches that of the Max Cores per year as expected. The only difference in data is the minimal number of Intel processors which do not have 2-way issue, aka hyperthreading as defined by Intel.

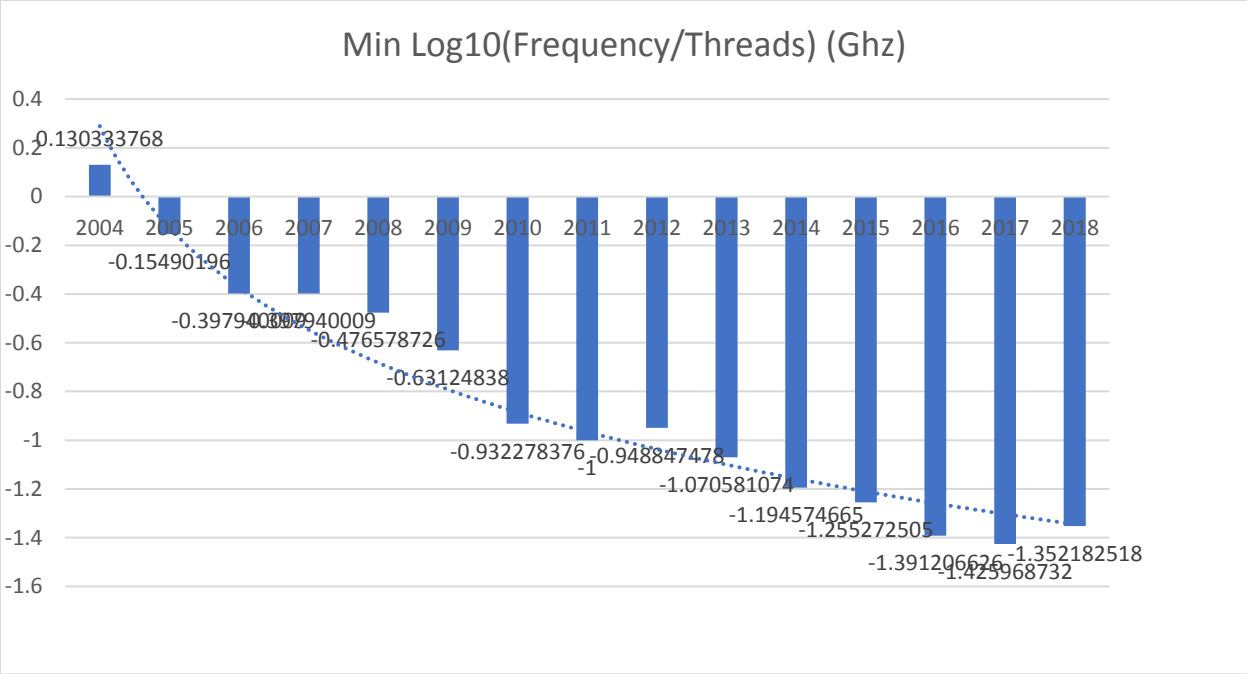

Figure 17 – Min Log10(Frequency/Threads) by Year(Ghz)

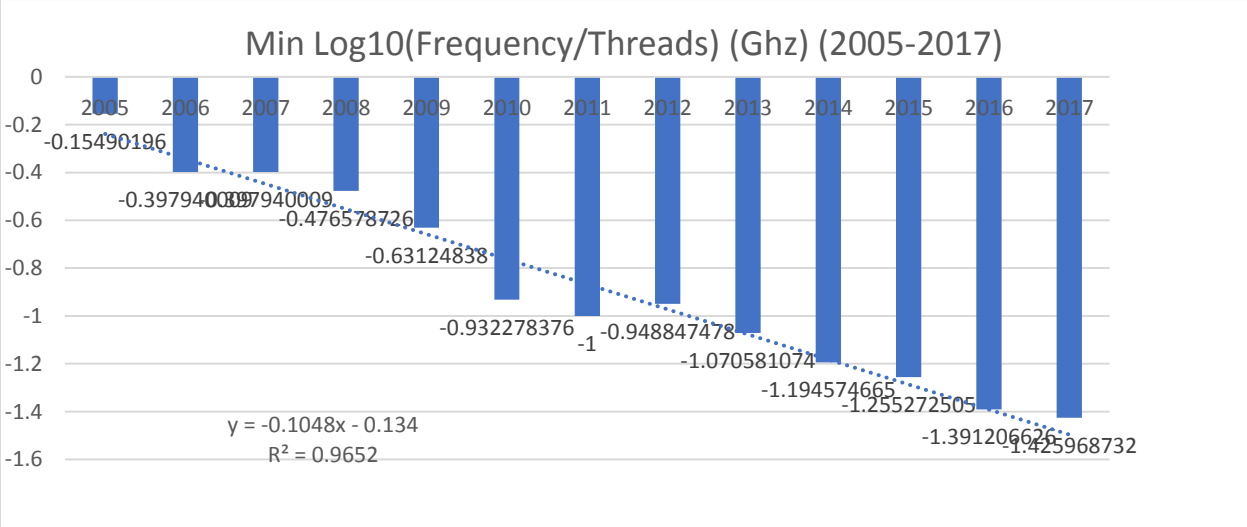

Figure 18 -Min Log10(Frequency/Threads) by Year (Ghz) (2005-2017)

After removing the data for 2004 and 2018 the new fit statistics are y = -0.1048x - 0.134 and R^2 = 0.9652. This is an increase of 0.221 over the original linear fit. The fist can also be expressed exponentially as $y = 0.7346e^{-0.241x}$. This fit shows a clear trend in ratio of Processing Power to the clock frequency. Efficiency is increasing. The current ratio is log10 = -1.4 or there is a roughly 1:25 ratio of Clock to Threads.

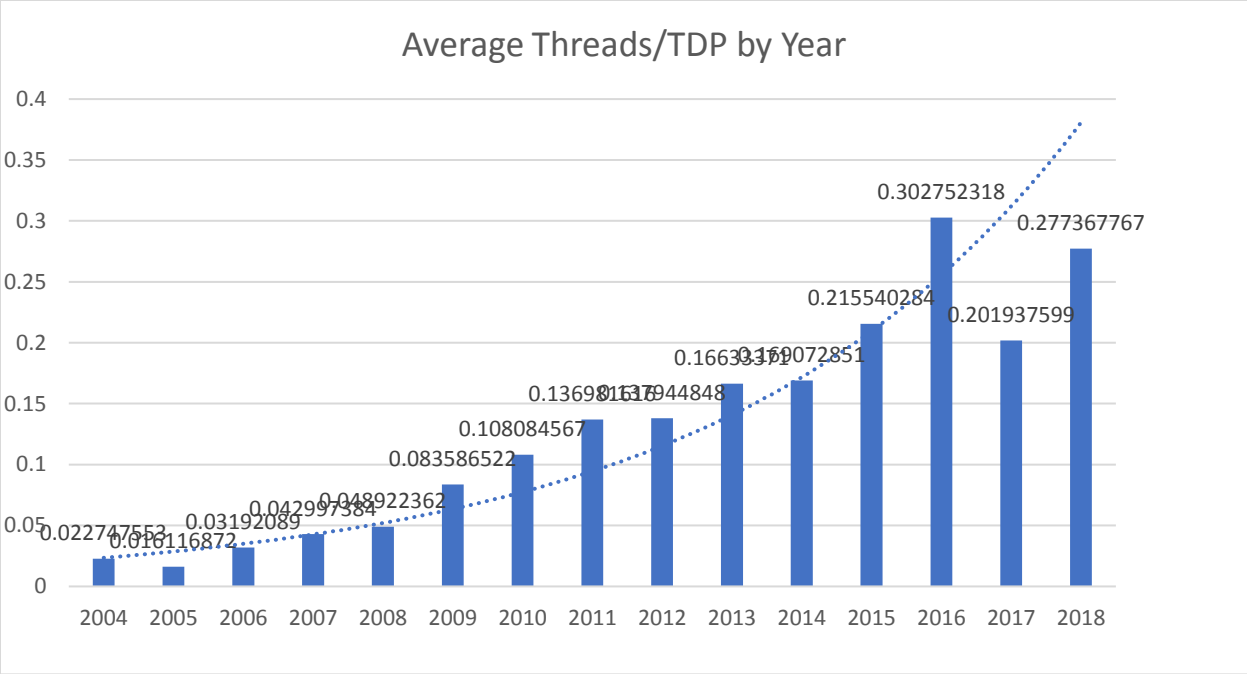

Figure 19 - Average Threads/TDP by Year

This figure demonstrates that on average each thread is using more power each year. This proves that CMP are becoming more efficient per watt.

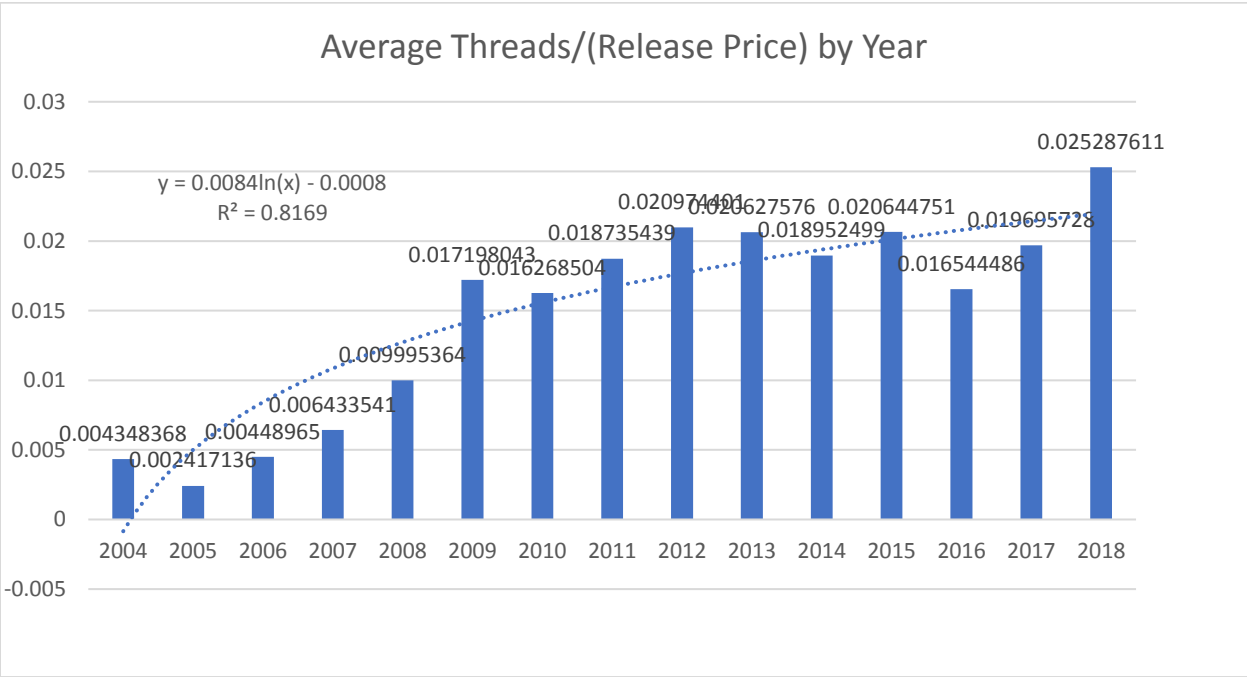

Figure 20 - Average Threads/ (Release Price) by Year (USD)

This parameter shows that while on average consumers are paying more per thread each year, this model has an R^2 of over 0.8 for a logarithmic fit. So, it is fair to conclude that an asymptomatic

relationship for Threads/Price does exist and consumers will slowly stop paying more per thread if this relationship holds.

The following sections displays the visualization of variance by year of the following parameters:

- (Release Price)/TDP by Year
- Frequency/Threads by Year
- Cores by Year
- TDP by Year
- Watts/Cycle by Year
- Threads/TDP by Year

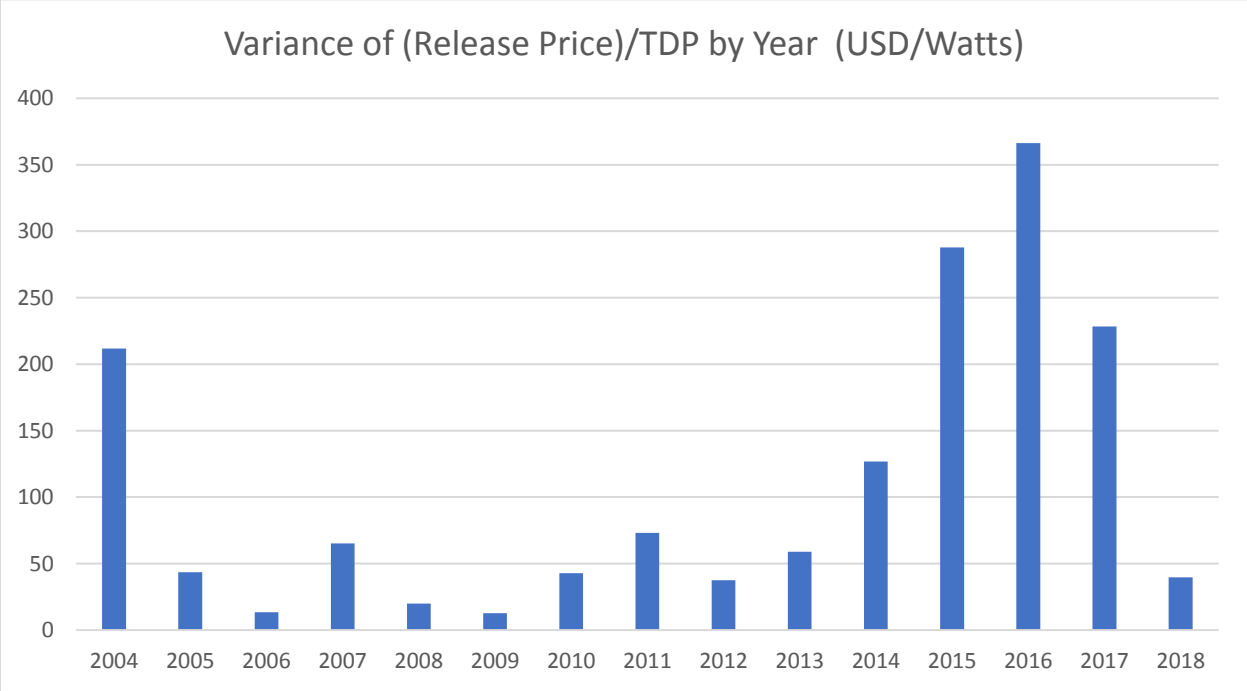

Figure 21 - Variance of (Release Price)/TDP by Year (USD/Watts)

A brief anecdotal and interesting conclusion is that for almost 10 years the difference in Price/Power did not change but in the last few years there is far more variety in specialization and usage of cores. I speculate the increased number of low-power embedded mobile processors along with high-performance Xeon server chips in the last 3 years is the reason for a sharp increase in variance.

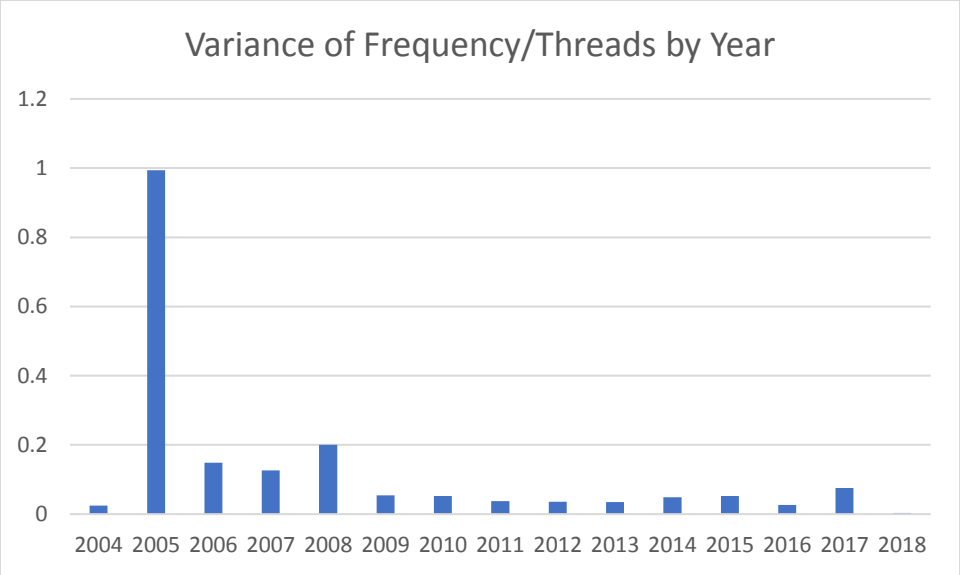

Figure 22 - Variance of Frequency/Threads by Year

Variance of Frequency/Thread has not changed for over 10 years. This matches with the focus on CMP designs that emerged in 2006 when Dennard scaling began to fail.

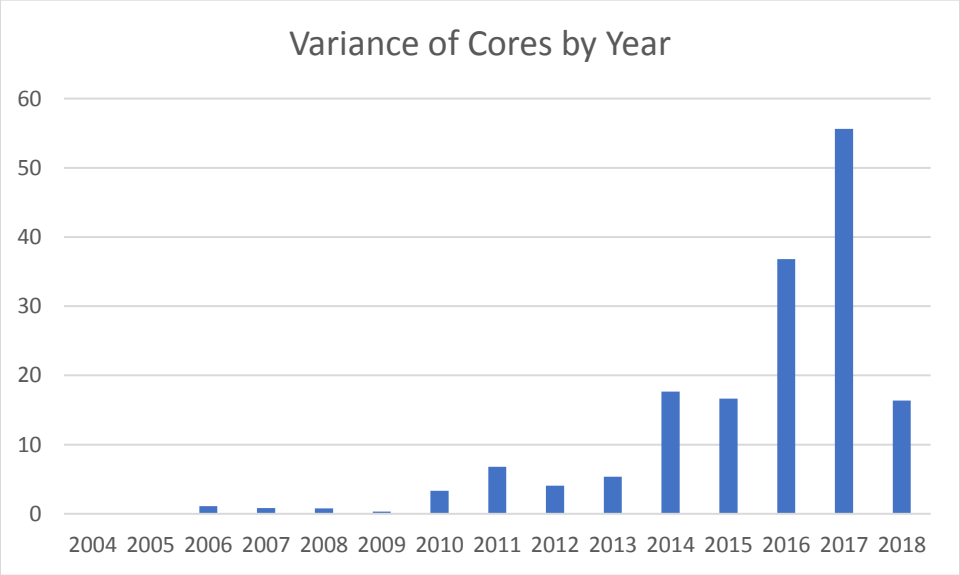

Figure 23 - Variance of Cores by Year

There has been a noted increase in the various number of types of processor in the last few years. From powerful low-power designs for embedded use, with 1 or 2 cores. To large 20+ core systems for servers, many of the latest Xeon chips have 32+ cores.

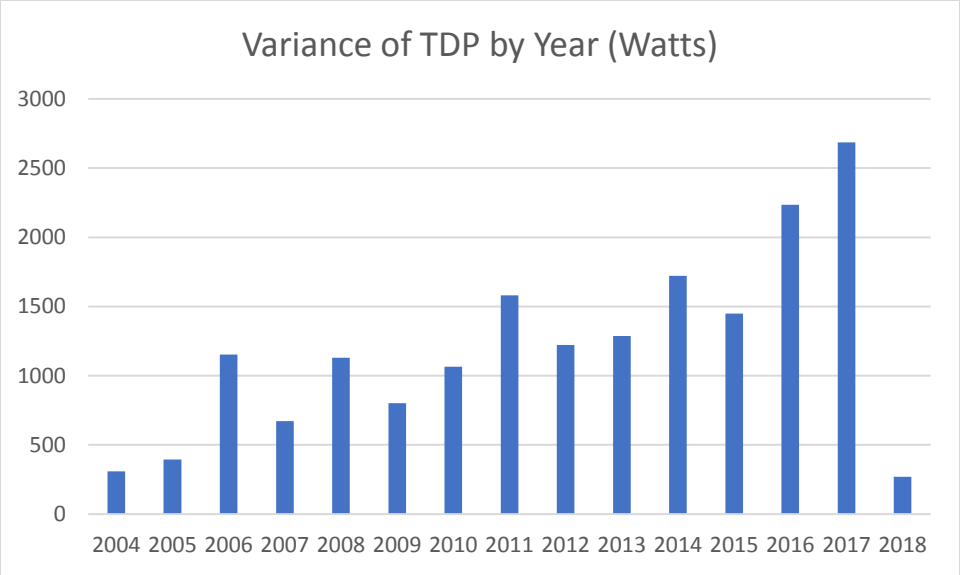

Figure 24 - Variance of TDP by Year (Watts)

There has always been a significant various in the TDP of all designs, specialization has always existed, but in matching with previous conclusions, the range of specialization has vastly increased in the last several years. Variance roughly doubled from 2012 to 2017.

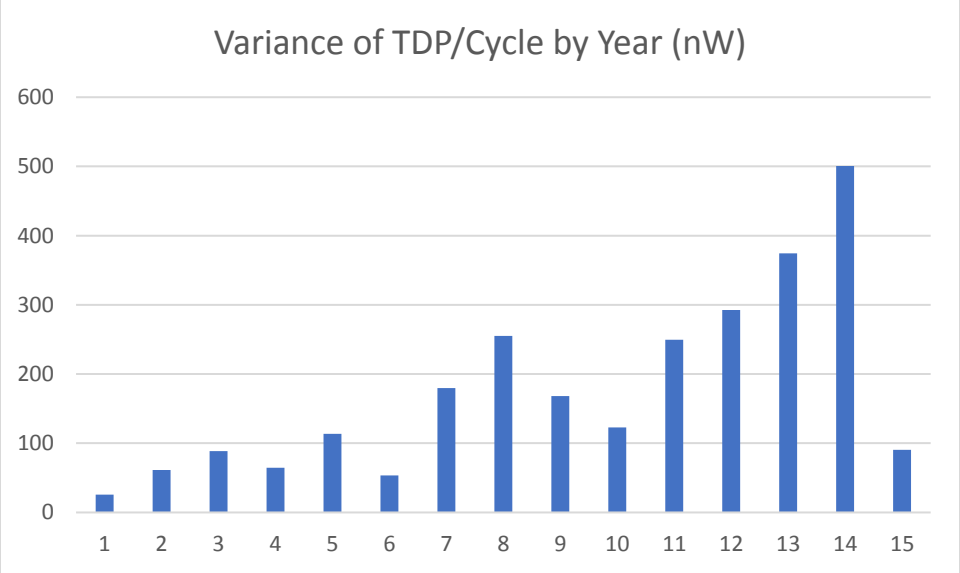

Figure 25 - Variance of TDP/Cycle by Year (nW)

A similar conclusion to TDP by year, but when average frequency is factored in the relationship still holds.

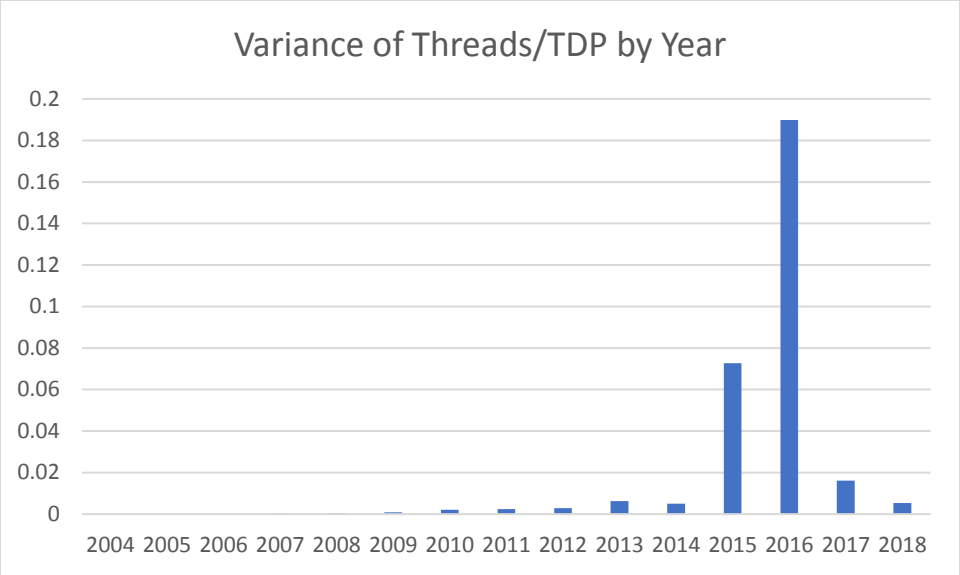

Figure 26- Variance of Threads/TDP by Year

There has only been a large variance in Threads/TDP for 3 years, although 2017 has a low variance as well. It would be interesting to do a further market analysis on 2015 and 2016 to see if there was in fact a large amount of design variation that year and then a singularity in 2017.

The following section contains two scaling parameters I believe add value to the analysis of microprocessor manufacturing trends:

- Max Cores * Average Frequency by Year
- (Max Cores * Average Frequency)/ (Average TDP) by Year

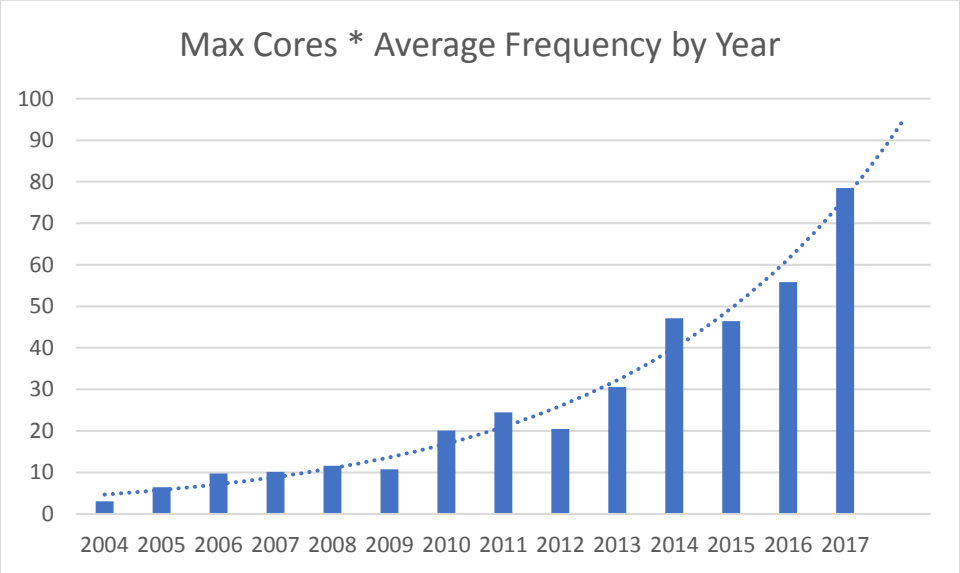

*Figure 27 - Max Cores * Average Frequency by Year*

My first proposed scaling factor is Max Cores * Average Frequency. This is a rough estimate of total theoretical computational power of a chip. The exponential fit statistics are $y = 7.5235e^{0.2148x}$ and $R^2 = 0.9533$. This is a very strong fit. This equation states that rough computing power will double every 3 years.

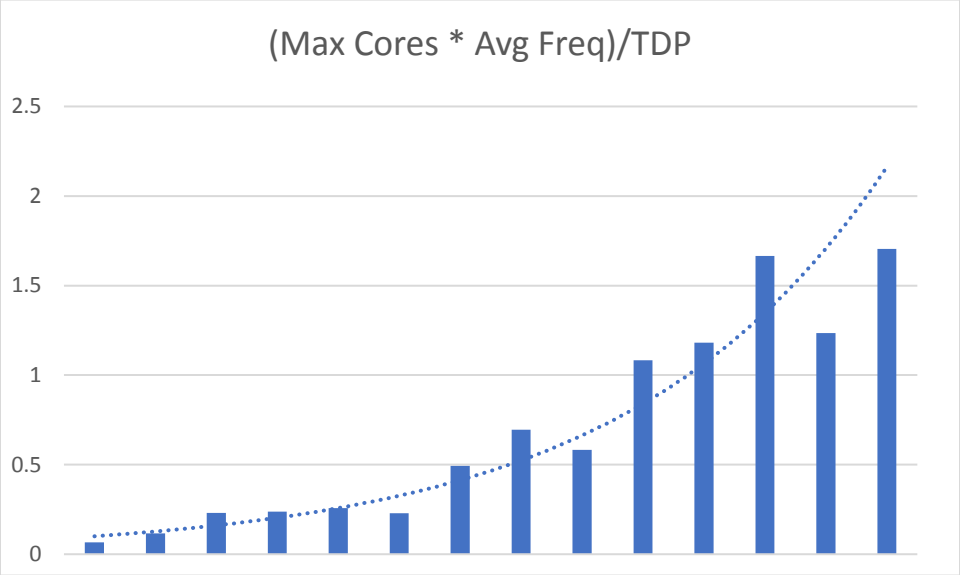

*Figure 28 - (Max Cores * Average Frequency)/TDP*

My second proposed scaling is simply the first divided by TDP to incorporate power into the fit. The exponential fit statists are $y = 0.0791e^{0.2363x}$ $R^2 = 0.9362$. This equation states that the efficiency of computation will double roughly ever 3 years.

I believe that these two parameters avoid the pitfalls of transistor-based scaling factors, as we know the power wall will obscure the theoretical vs actualized performance of chips. However, if we upscale transistors simply to number of cores and multiply that by the clock frequency we get a much more accurate estimate of the computational power of a CMP system. Moore's Law was based off the trends of SMP systems where for a long time more transistors directly meant an equal increase in performance.

Predictions

Just for fun based on my top 5 parameters I will make the following predictions about computing:

1. Cores - $y = 1.4541e^{0.2063x}$

   The max number of cores per chip will equal 128 in the year 2023

2. Threads - $y = 2.2185e^{0.2285x}$

The max number of threads will be 256 in the year 2023.

3. Log(Frequency/Threads) - $y = -0.604\ln(x) + 0.2893$

   The frequency/thread will be a ratio of 1:100 in the year 2041

4. Threads/TDP - $y = 0.0192e^{0.199x}$

   We will achieve a value of 1 thread per watt in the year 2027

5. Threads/Price - $y = 0.0084\ln(x) - 0.0008$

   We will get 0.1 threads per dollar in the year 152004…..hmmm this prediction might take a while

Conclusion

In this paper I conducted an analysis on the last 20 years of microprocessor design to determine if any other exponential laws, or linear for that matter, could be derived. My analysis ended up yielding 5 parameters with R^2 over 0.8 and 4 over 0.9, and one over 0.95 when and adjusted year range was used.

Those are 4 parameters that with a limited data set and data fitting have very promising values of determination. The initial question of "Are there more exponential trends in computing to find?" seems to be a definitive yes.

The ultimate determination of my data was the creation of two closely related scaling factors that roughly show the computational power of a CMP system. These factors are (Max Cores * Average Frequency) and the same value divided by TDP. One parameter is strictly theoretical computational power and the second is simply computational efficiency per Watt. Two related but slightly different parameters.

The two parameters had R^{2+} over 0.93 and I confidently conclude they can adequately model theoretical computational power for CMP chips designs in the coming years.

References

(1) “Dennard Scaling.” *Wikipedia*, Wikimedia Foundation, 6 Mar. 2018, en.wikipedia.org/wiki/Dennard_scaling.

(2) Domaille, Susannah, et al. “Moore's Law.” *The Plymouth Student Scientist*, vol. 2104, no. 7, ser. 2, pp. 185–194. *2*.

(3) “Koomey's Law.” *Wikipedia*, Wikimedia Foundation, 3 Mar. 2018, en.wikipedia.org/wiki/Koomey%27s_law.

(4) “List of Intel Core i3 Microprocessors.” Wikipedia, Wikimedia Foundation, 27 Feb. 2018, en.wikipedia.org/wiki/List_of_Intel_Core_i3_microprocessors.

(5) “List of Intel Core i5 Microprocessors.” Wikipedia, Wikimedia Foundation, 27 Feb. 2018, en.wikipedia.org/wiki/List_of_Intel_Core_i5_microprocessors.

(6) “List of Intel Core i7 Microprocessors.” Wikipedia, Wikimedia Foundation, 27 Feb. 2018, en.wikipedia.org/wiki/List_of_Intel_Core_i7_microprocessors.

(7) “List of Intel Core i9 Microprocessors.” Wikipedia, Wikimedia Foundation, 27 Feb. 2018, en.wikipedia.org/wiki/List_of_Intel_Core_i9_microprocessors.

(9) “List of Intel Xeon Microprocessors.” Wikipedia, Wikimedia Foundation, 27 Feb. 2018, en.wikipedia.org/wiki/List_of_Intel_Core_Xeon_microprocessors.