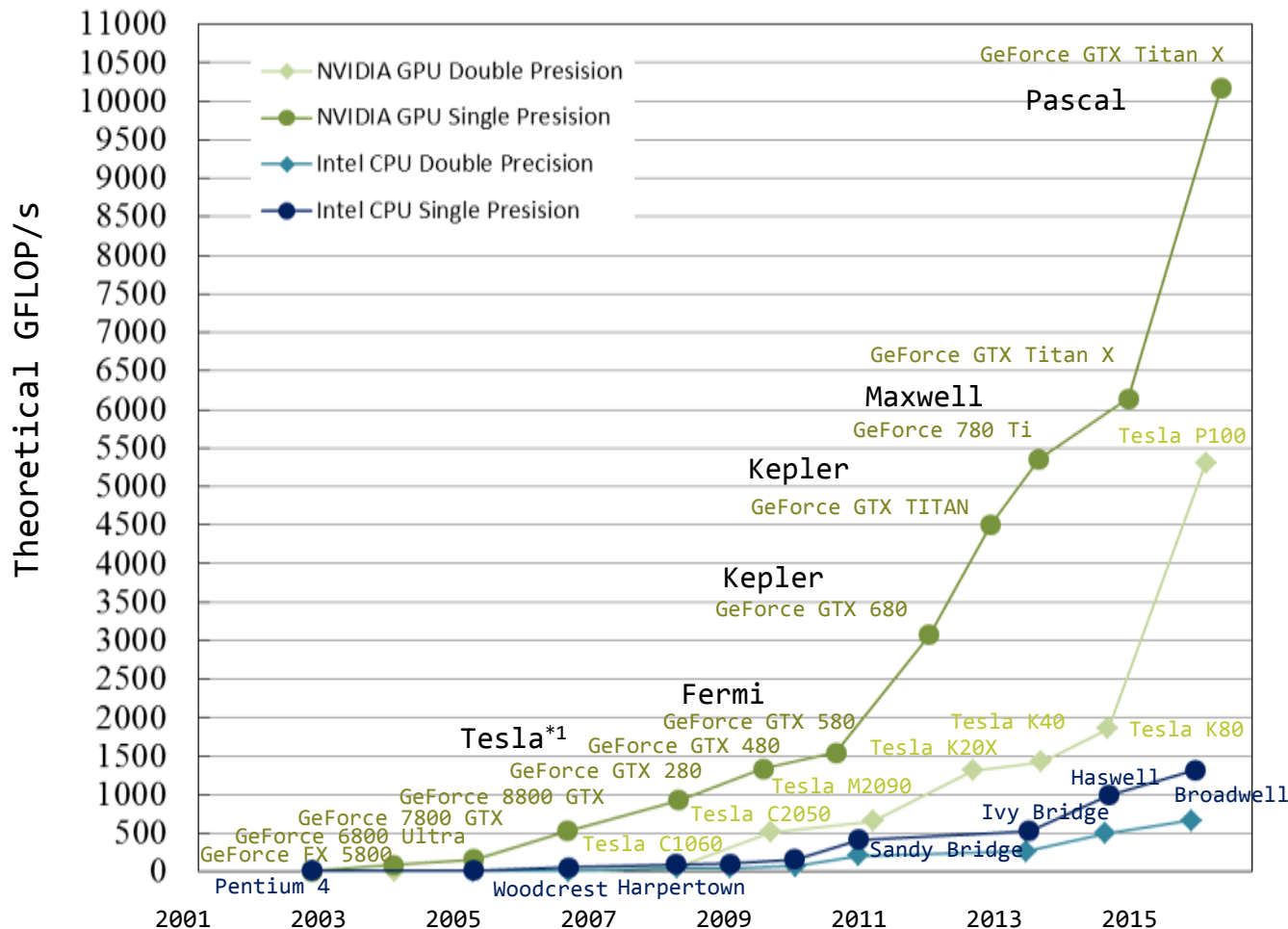


FLOPSとバンド幅の遷移および GPUアーキテクチャの模式図

名古屋大学未来材料・システム研究所 出川智啓

GPUの性能の遷移（理論演算性能）

Excel Sheet



GeForce
ゲーム用

Quadro
CG用

Tesla
GPGPU用

*1コードネーム

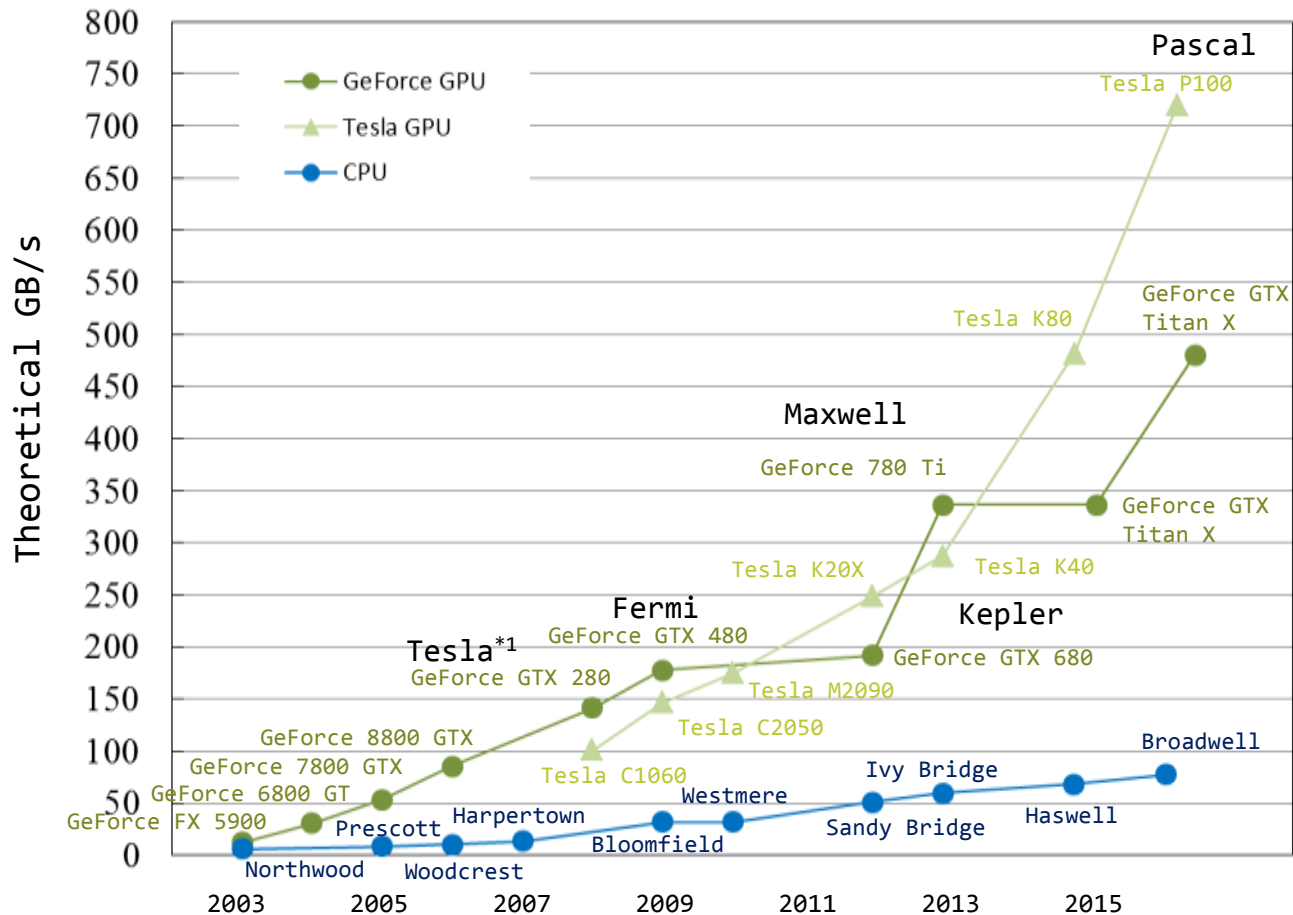
*2製品ファミリ

<http://docs.nvidia.com/cuda/cuda-c-programming-guide/>で公開されている資料を基に作成

2017/4/1

GPUの性能の遷移（理論バンド幅）

Excel Sheet



GeForce
ゲーム用

Quadro
CG用

Tesla
GPGPU用

<http://docs.nvidia.com/cuda/cuda-c-programming-guide/>で公開されている資料を基に作成

Teslaアーキテクチャ

▶ Tesla C1060の仕様

- ▶ SM数 30
- ▶ CUDA Core数 $240 (= 8 \text{ Core/SM} \times 30 \text{ SM})$
- ▶ キャッシュを搭載せず



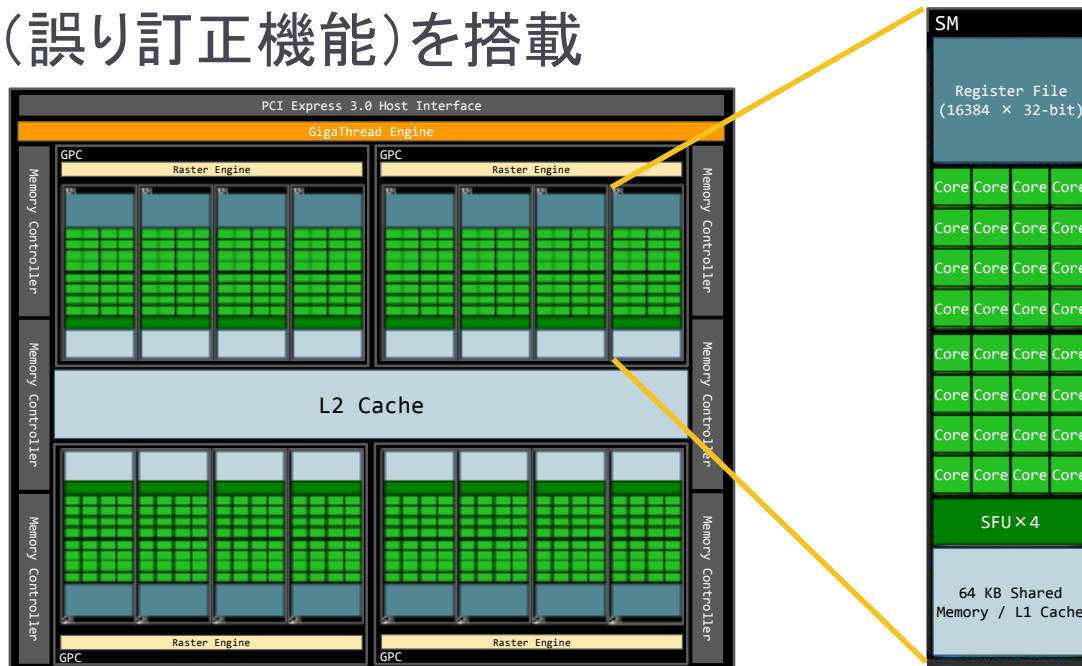
<http://www.anandtech.com/show/2549/2>で公開されている画像を基に作成

Fermiアーキテクチャ

詳細はhttp://www.nvidia.co.jp/docs/IO/81860/NVIDIA_Fermi_Architecture_Whitepaper_FINAL_J.pdfを参照のこと

▶ Tesla M2050の仕様

- ▶ SM数 14
- ▶ CUDA Core数 $448 (= 32 \text{ Core/SM} \times 14 \text{ SM})$
- ▶ L1/L2 キャッシュを搭載
- ▶ ECC(誤り訂正機能)を搭載



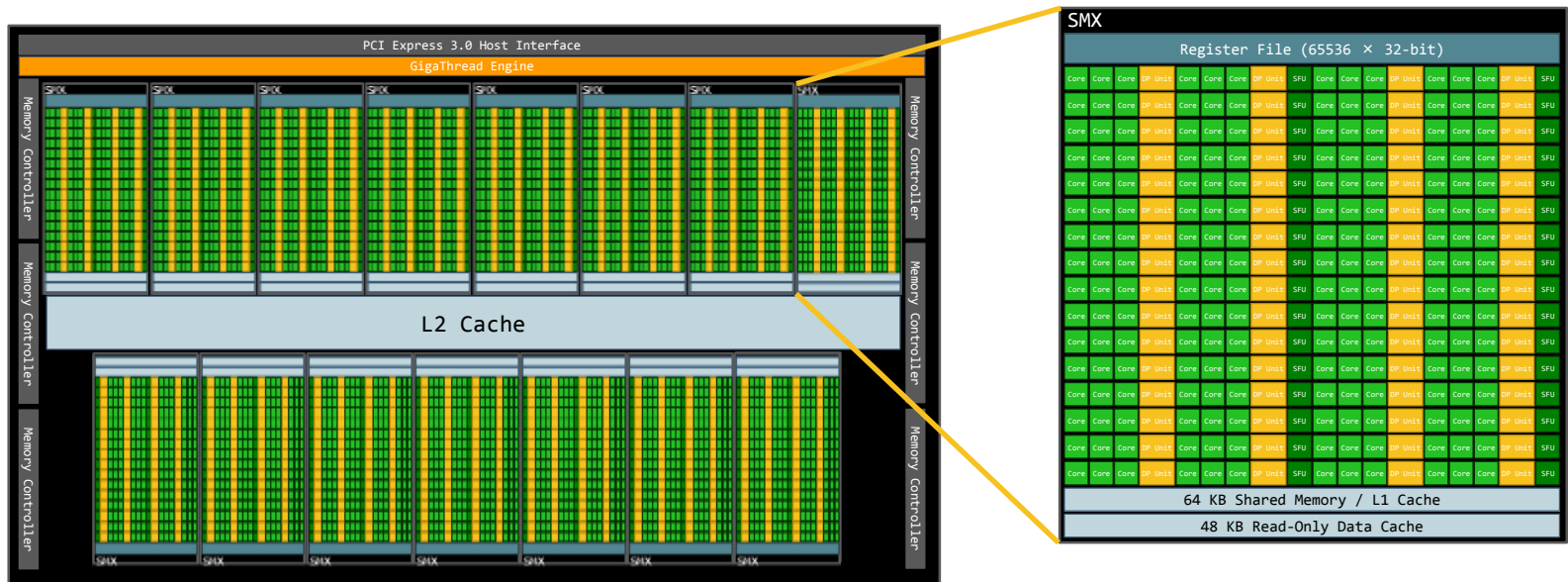
<http://www.anandtech.com/show/2849/3>で公開されている画像を基に作成

Keplerアーキテクチャ

詳細は<https://www.nvidia.co.jp/content/apac/pdf/tesla/nvidia-kepler-gk110-architecture-whitepaper-jp.pdf>を参照のこと

▶ Tesla K20c/mの仕様

- ▶ SMX数 13
- ▶ CUDA Core数 $2,496 (= 192 \text{ Core/SM} \times 13 \text{ SMX})$



https://library.creativecow.net/kaufman_debra/NVIDIA-VGX/1で公開されている画像を基に作成

Maxwellアーキテクチャ

第1世代の詳細は<https://www.nvidia.co.jp/content/product-detail-pages/geforce-gtx-750-ti/geforce-gtx-750ti-whitepaper.pdf>を参照のこと

▶ GeForce GTX TITAN Xの仕様

- ▶ SMM数 24
- ▶ CUDA Core数 $3,072 (=128 \text{ Core/SM} \times 24 \text{ SM})$
 - ▶ 倍精度演算器は搭載していない



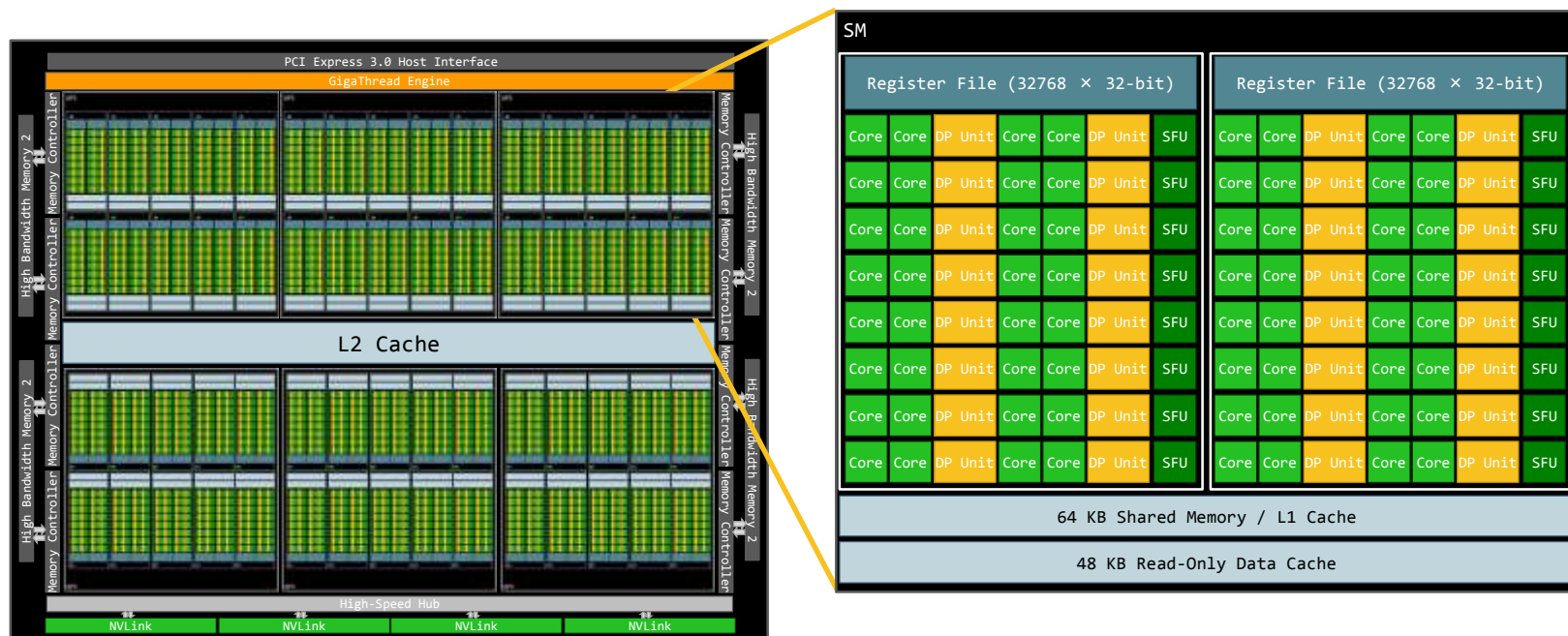
<http://www.itmedia.co.jp/pcuser/articles/1409/19/news051.html>で公開されている画像を基に作成

Pascalアーキテクチャ

詳細は<http://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf>を参照のこと

▶ Tesla P100の仕様

- ▶ SM数 56
- ▶ CUDA Core数 3584 (=64 Core/SM × 56 SM)

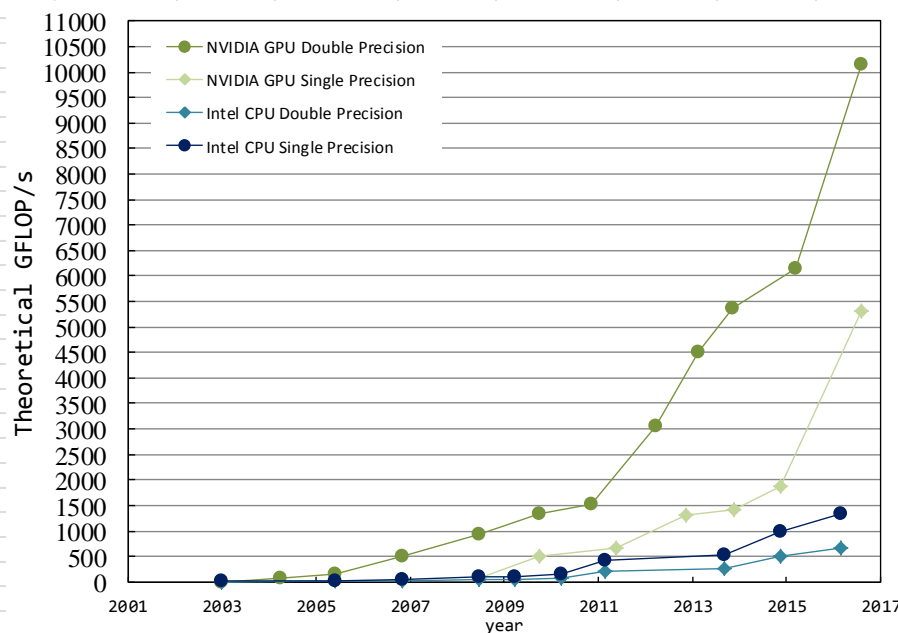


<http://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf>で公開されている画像を基に作成

理論演算性能 (Embedded Excel Sheet)

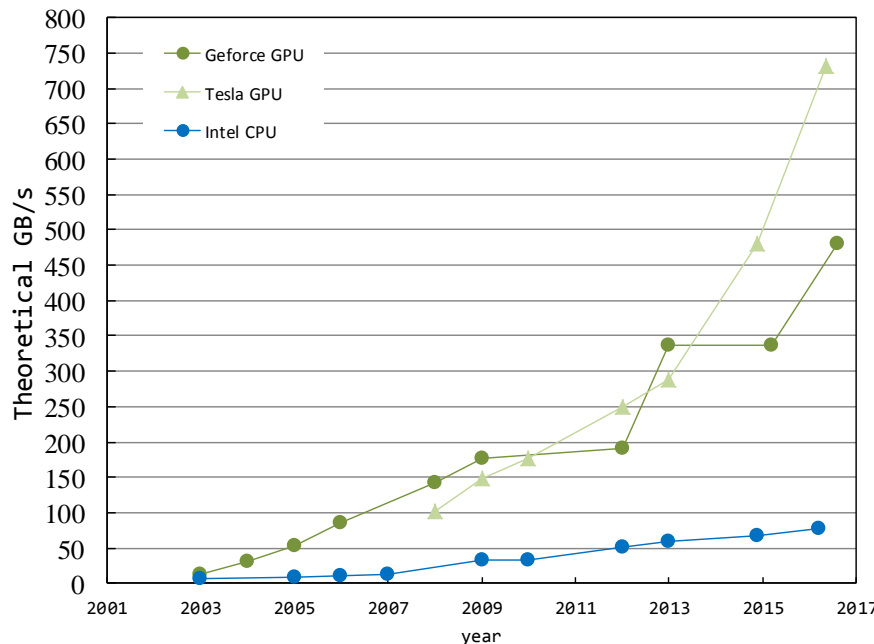
year	NVIDIA GP	year	NVIDIA GP	year	Intel CPU	year	Intel CPU Double Precision
2003.000	0.00E+00	2008.462	7.80E+01	2003.000	7.60E+00	2003.000	3.80E+00
2004.248	7.72E+01	2009.751	5.15E+02	2005.413	2.66E+01	2005.413	1.33E+01
2005.413	1.54E+02	2011.369	6.66E+02	2006.825	5.12E+01	2006.825	2.66E+01
2006.832	5.17E+02	2012.864	1.31E+03	2008.456	9.00E+01	2008.456	4.24E+01
2008.462	9.28E+02	2013.877	1.43E+03	2009.233	1.10E+02	2009.233	5.26E+01
2009.751	1.34E+03	2014.872	1.87E+03	2010.204	1.68E+02	2010.204	6.29E+01
2010.846	1.52E+03	2016.594	5.30E+03	2011.151	4.26E+02	2011.151	2.16E+02
2012.224	3.07E+03			2013.688	5.32E+02	2013.688	2.66E+02
2013.137	4.50E+03			2014.871	9.90E+02	2014.871	4.95E+02
2013.855	5.36E+03			2016.160	1.32E+03	2016.160	6.68E+02
2015.203	6.14E+03						
2016.594	1.02E+04						

- ▶ プログラミングガイドの図からデータを推定
- ▶ GPU倍精度のみ正しい値に修正
- ▶ 残りは近似値



理論バンド幅* (Embedded Excel Sheet)

year	Geforce GPU	year	Tesla GPU	year	Intel CPU
2003.000	1.26E+01	2008.000	1.02E+02	2003.000	6.29E+00
2004.000	3.08E+01	2009.000	1.49E+02	2005.000	8.81E+00
2005.000	5.35E+01	2010.000	1.78E+02	2006.000	1.07E+01
2006.000	8.56E+01	2012.000	2.50E+02	2007.000	1.32E+01
2008.000	1.42E+02	2013.000	2.88E+02	2009.000	3.21E+01
2009.000	1.77E+02	2014.884	4.80E+02	2010.000	3.21E+01
2012.000	1.92E+02	2016.351	7.32E+02	2012.000	5.10E+01
2013.000	3.36E+02			2013.000	5.98E+01
2015.196	3.36E+02			2014.879	6.81E+01
2016.604	4.80E+02			2016.189	7.77E+01



- ▶ プログラミングガイドの図からデータを推定
- ▶ Tesla GPUのみ正しい値に修正
- ▶ 残りは近似値