



GPU Native FP Throughput Rates (Relative To FP32)						
Name	Chip	FP16	FP64	INT8	INT4	
Radeon RX Vega		2x	1/16	4x	N/A	
A100 SXM4	GA100	2x	1/2			<a href="https://developer.nvidia.com/cuda-gpus">https://developer.nvidia.com/cuda-gpus</a>
Titan RTX	TU102	2x	1/32	4x	8x	<a href="https://en.wikipedia.org/wiki/AMD_RX_Vega_series">https://en.wikipedia.org/wiki/AMD_RX_Vega_series</a> <a href="https://instinct.radeon.com/en/exploring-amd-vega-deep-learning/">https://instinct.radeon.com/en/exploring-amd-vega-deep-learning/</a>
Tesla T4	TU104	2x	1/32	4x	8x	
Quadro RTX 8000	TU102	2x	1/32	4x	8x	
Quadro RTX 6000	TU102	2x	1/32	4x	8x	
Quadro RTX 5000	TU104	2x	1/32	4x	8x	
Quadro RTX 4000	TU104	2x	1/32	4x	8x	
Quadro GV100	GV100	2x	1/2	4x	N/A	
Quadro GP100	GP100	2x	1/2	4x	N/A	
RTX 2080 Ti	TU102	2x	1/32	4x	8x	
RTX 2080	TU104	2x	1/32	4x	8x	
RTX 2070	TU106	2x	1/32	4x	8x	
Tesla V100	GV100	2x (w/o Tensor Cores)	1/2	4x	N/A	
Titan V	GV100	2x (w/o Tensor Cores)	1/2	4x	N/A	<a href="https://www.anandtech.com/show/12170/nvidia-titan-v-preview-rtiomachy">https://www.anandtech.com/show/12170/nvidia-titan-v-preview-rtiomachy</a>
Titan Xp	GP102	1/64	1/32	4x	N/A	<a href="https://www.anandtech.com/show/11263/nvidia-announces-nvidia-titan-xp-full-gp102-for-1200">https://www.anandtech.com/show/11263/nvidia-announces-nvidia-titan-xp-full-gp102-for-1200</a>
Titan X (Pascal)	GP102	1/64	1/32	4x	N/A	
GTX 1080 Ti	GP102	1/64	1/32	4x	N/A	<a href="https://www.anandtech.com/show/11172/nvidia-unveils-geforce-gtx-1080-ti-next-week-699">https://www.anandtech.com/show/11172/nvidia-unveils-geforce-gtx-1080-ti-next-week-699</a>
GTX 1080	GP104	1/64	1/32	4x	N/A	<a href="https://www.anandtech.com/show/10325/the-nvidia-geforce-gtx-1080-and-1070-founders-edition-review5">https://www.anandtech.com/show/10325/the-nvidia-geforce-gtx-1080-and-1070-founders-edition-review5</a>
GTX 1070 Ti	GP104	1/64	1/32	4x	N/A	<a href="https://www.anandtech.com/show/11987/the-nvidia-geforce-gtx-1070-ti-founders-edition-review">https://www.anandtech.com/show/11987/the-nvidia-geforce-gtx-1070-ti-founders-edition-review</a>
GTX 1070	GP104	1/64	1/32	4x	N/A	
GTX 1060	GP106	1/64	1/32	4x	N/A	<a href="https://www.anandtech.com/show/10580/nvidia-releases-geforce-gtx-1060-3gb">https://www.anandtech.com/show/10580/nvidia-releases-geforce-gtx-1060-3gb</a>
GTX 1050 Ti	GP107	1/64	1/32	4x	N/A	<a href="https://www.anandtech.com/show/10980/nvidia-launches-geforce-gtx-1050-ti-gtx-1050-for-laptops">https://www.anandtech.com/show/10980/nvidia-launches-geforce-gtx-1050-ti-gtx-1050-for-laptops</a>
Tesla P100	GP100	2x	1/2	N/A	N/A	<a href="https://www.anandtech.com/show/">https://www.anandtech.com/show/</a> <a href="https://www.anandtech.com/show/">https://www.anandtech.com/show/</a> <a href="https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf">https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf</a>
Tesla P40	GP102	N/A	1/2	4x	N/A	<a href="https://www.anandtech.com/show/10675/nvidia-announces-tesla-p40-tesla-p4">https://www.anandtech.com/show/10675/nvidia-announces-tesla-p40-tesla-p4</a>
Tesla P4	GP104	N/A	1/2	4x	N/A	<a href="http://www.nvidia.com/object/accelerate-inference.html">http://www.nvidia.com/object/accelerate-inference.html</a>
Jetson TX2	Tegra X2	2x	N/A ?	N/A	N/A	<a href="https://devtalk.nvidia.com/default/topic/1026771/data-process-about-tensorrt-int8-and-fp16-inference-engine/">https://devtalk.nvidia.com/default/topic/1026771/data-process-about-tensorrt-int8-and-fp16-inference-engine/</a>
Jetson TX1	Tegra X1	2x	N/A ?	N/A	N/A	
Tesla M40	GM200	N/A (Promoted to FP32)	1/32	N/A	N/A	<a href="https://www.anandtech.com/show/9778/nvidia-announces-tesla-m40-m4-server-card-data-center-machine-learning">https://www.anandtech.com/show/9778/nvidia-announces-tesla-m40-m4-server-card-data-center-machine-learning</a>
Tesla M60	GM204	N/A (Promoted to FP32)	1/32	N/A	N/A	
Tesla M4	GM206	N/A (Promoted to FP32)	1/32	N/A	N/A	
Titan X (Maxwell)	GM200	N/A (Promoted to FP32)	1/32	N/A	N/A	
GTX Titan	GK110	N/A (Promoted to FP32)	1/3	N/A	N/A	<a href="https://www.anandtech.com/show/12170/nvidia-titan-v-preview-rtiomachy">https://www.anandtech.com/show/12170/nvidia-titan-v-preview-rtiomachy</a>
GTX 680	GK104	N/A (Promoted to FP32)	1/24	N/A	N/A	
Tesla K40	GK110	N/A (Promoted to FP32)	1/3	N/A	N/A	

Year	YearMonth	GLOPS_F32	GLOPS_F16	GLOPS_TENSOR	TOPS_INT	Annotation	Core	TDP	F32-16 multiplier	F32-INT multiplier	Price
2020	2020.03.01	2958	5915			RTX 2080	RTX 2080	300	2		
2020	2020.03.01	1784	3568			RTX 2070	RTX 2070	220	2		
2020	2020.05.14	312000	624000	624000	1248	A100 SXM4	A100 SXM4	400	2	4	
2020	2020.05.14	9500	19000	312000	624	Arpeggio A100	A100 SXM4	400	2	4	
2019	2019.04.03	2881	5762			RTX 1600	RTX 1600	75	2		
2019	2019.23.09	10138	20275	89200	178.4	RTX 2080 Super	RTX 2080 Super	250	2	4	
2019	2019.07.01	6118	12236	12236	2447.6	RTX 2070 Super	RTX 2070 Super	215	2	4	
2019	2019.07.09	6123	12246	37420	14.8	RTX 2080 Super	RTX 2080 Super	175	2	4	
2019	2019.07.09	438	876			RTX 1600	RTX 1600	75	2		
2019	2019.02.22	4608	9216			RTX 1600 Ti	RTX 1600 Ti	100	2		
2019	2019.01.15	5242	10484	51800	103.2	RTX 2080	RTX 2080	180	2	4	
2019	2019.12.18	10312	20624	150000	300.2	RTX 2080 Ti	RTX 2080 Ti	250	2	4	
2019	2019.11.13	7100	14200			Quadro RTX 4000	Quadro RTX 4000	180	2	4	
2019	2019.08.13	10000	20000	130000	260.0	Quadro RTX 8000	Quadro RTX 8000	250	2	4	
2019	2019.08.13	16300	32600	261	0.000	Quadro RTX 8000	Quadro RTX 8000	295	2	4	
2019	2019.03.27	14600	29200	118000	236.0	Quadro DV100	Quadro DV100	250	2	4	
2019	2019.02.27	15444	30888	118000	237.7	RTX 2080 Ti FE	RTX 2080 Ti FE	250	2	4	
2019	2019.02.27	11700	23400	107800	215.2	RTX 2080 Ti	RTX 2080 Ti	200	2	4	
2019	2019.08.13	12000	24000	80000	179.4	Quadro RTX 5000	Quadro RTX 5000	285	2	4	
2019	2019.09.20	10000	20000	84800	169.6	RTX 2080 FE	RTX 2080 FE	225	2	4	
2019	2019.09.20	8000	17440	80000	161.1	RTX 2080	RTX 2080	215	2	4	
2019	2019.09.12	8100	16200	60000	120	TESLA T4	TESLA T4	75	2	4	
2019	2019.10.17	7900	15800	60000	120	RTX 2070 FE	RTX 2070 FE	185	2	4	
2019	2019.10.17	6487	12974	60700	121.4	RTX 2070	RTX 2070	175	2	4	
2017	2017.01.03	8812	17624			RTX Titan X (Maxwell)	RTX Titan X (Maxwell)	375	2	4	
2017	2017.06.01	14028	28056	130000	56.12	RTX Titan Xp	RTX Titan Xp	250	2	4	
2017	2017.12.07	13000	26000	120000	56.12	RTX Titan V	RTX Titan V	250	2	4	
2017	2017.08.14	11016	22032	46.072	RTX Titan Xp	RTX Titan Xp	345	2	4		
2017	2017.04.06	10790	21580	43.16	RTX Titan Xp	RTX Titan Xp	250	2	4		
2017	2017.03.10	10600	21200	42.400	RTX Titan V	RTX Titan V	250	2	4		
2017	2017.02.08	10300	20600	40.600	RTX Titan V	RTX Titan V	230	2	4		
2017	2017.08.14	10215	20430	40.80	RTX Titan Xp	RTX Titan Xp	285	2	4		
2017	2017.07.26	8384	16768	33.144	RTX Titan Xp	RTX Titan Xp	210	2	4		
2017	2017.11.02	7816	15632	31.264	RTX 1070 Ti	RTX 1070 Ti	180	2	4		
2017	2017.09.21	6884	13768	5.18	RTX Titan Xp	RTX Titan Xp	290	2	4		
2017	2017.09.25	2000	4000	0-19000	0-19000	RTX Titan Xp	RTX Titan Xp	165	2	4	
2017	2017.09.17	1402	2804	3.690	RTX 1070	RTX 1070	150	2	4		
2017	2017.05.02	441.6	883.2	0.000	RTX Titan Xp	RTX Titan Xp	15	2	4		
2017	2017.12.18	0	0	0	0	RTX Titan Xp	RTX Titan Xp	300	0	4	
2016	2016.02.13	11700	23400	47.03	RTX Titan Xp	RTX Titan Xp	200	2	4		
2016	2016.08.02	10157	20314	40.628	RTX Titan X (Maxwell)	RTX Titan X (Maxwell)	300	2	4		
2016	2016.08.20	9340	18680	36.840	RTX Titan Xp	RTX Titan Xp	300	2	4		
2016	2016.02.27	6238	12476	33.912	RTX Titan Xp	RTX Titan Xp	180	2	4		
2016	2016.08.20	6912	0	0	0	RTX Titan Xp	RTX Titan Xp	300	0	4	
2016	2016.08.10	5932	0	0	0	RTX Titan Xp	RTX Titan Xp	180	0	4	
2016	2016.08.13	5443	0	0	0	RTX Titan Xp	RTX Titan Xp	75	0	4	
2016	2016.08.15	3470	0	0	0	RTX Titan Xp	RTX Titan Xp	120	0	4	
2016	2016.10.25	1881	0	0	0	RTX Titan Xp	RTX Titan Xp	75	0	4	
2016	2016.10.25	1733	0	0	0	RTX Titan Xp	RTX Titan Xp	75	0	4	
2016	2016.08.22	700	0	0	0	RTX Titan Xp	RTX Titan Xp	15	0	4	
2016	2016.03.30	680	0	0	0	RTX Titan Xp	RTX Titan Xp	140	0	4	
2015	2015.03.30	9500	0	0	0	RTX Titan Xp	RTX Titan Xp	200	0	4	
2015	2015.11.10	6844	0	0	0	RTX Titan Xp	RTX Titan Xp	200	0	4	
2015	2015.03.31	6144	0	0	0	RTX Titan Xp (Maxwell)	RTX Titan Xp (Maxwell)	280	0	4	
2015	2015.08.01	5382	0	0	0	RTX 580 Ti	RTX 580 Ti	200	0	4	
2015	2015.08.20	3229	0	0	0	RTX 580	RTX 580	100	0	4	
2015	2015.01.27	2368	0	0	0	RTX 580	RTX 580	100	0	4	
2015	2015.11.10	2195	0	0	0	RTX 584	RTX 584	75	0	4	
2015	2015.08.20	1072	0	0	0	RTX 680	RTX 680	60	0	4	
2015	2015.01.04	512	0	0	0	RTX 680	RTX 680	15	0	4	
2014	2014.11.07	6278	0	0	0	RTX 680	RTX 680	375	0	4	
2014	2014.03.25	6122	0	0	0	RTX Titan Z	RTX Titan Z	375	0	4	
2014	2014.02.18	5721	0	0	0	RTX Titan Black	RTX Titan Black	200	0	4	
2014	2014.08.18	4812	0	0	0	RTX 680	RTX 680	185	0	4	
2014	2014.08.18	3484	0	0	0	RTX 570	RTX 570	145	0	4	
2014	2014.02.18	1386	0	0	0	RTX 780 Ti	RTX 780 Ti	60	0	4	
2014	2014.02.18	1044	0	0	0	RTX 780	RTX 780	55	0	4	
2014	2014.02.29	780	0	0	0	RTX 780	RTX 780	64	0	4	
2014	2014.08.18	683	0	0	0	RTX 750	RTX 750	61	0	4	
2014	2014.03.27	306	0	0	0	RTX 750	RTX 750	200	0	4	
2013	2013.11.07	5046	0	0	0	RTX 780 Ti	RTX 780 Ti	200	0	4	
2013	2013.10.08	5040	0	0	0	RTX 740	RTX 740	235	0	4	
2013	2013.02.21	4500	0	0	0	RTX Titan	RTX Titan	250	0	4	
2013	2013.02.23	3977	0	0	0	RTX 780	RTX 780	200	0	4	
2013	2013.09.30	3513	0	0	0	RTX 770	RTX 770	220	0	4	
2013	2013.09.30	2460	0	0	0	RTX 760	RTX 760	170	0	4	
2013	2013.08.25	2238	0	0	0	RTX 680	RTX 680	170	0	4	
2012	2012.04.09	2021.78	0	0	0	RTX 680	RTX 680	200	0	4	
2012	2012.05.01	4577	0	0	0	RTX 680	RTX 680	185	0	4	
2012	2012.05.10	2460	0	0	0	RTX 670	RTX 670	170	0	4	
2012	2012.08.13	1881.6	0	0	0	RTX 680	RTX 680	140	0	4	
2012	2012.08.13	1428.8	0	0	0	RTX 680 Ti	RTX 680 Ti	110	0	4	
2012	2012.08.13	812.3	0	0	0	RTX 650	RTX 650	64	0	4	
2012	2012.04.24	728.8	0	0	0	RTX 650	RTX 650	75	0	4	
2012	2012.05.15	311	0	0	0	RTX 600	RTX 600	65	0	4	
2012	2012.05.15	298.8	0	0	0	RTX 600	RTX 600	69	0	4	
2012	2012.05.15	183.5	0	0	0	RTX 610	RTX 610	59	0	4	
2011	2011.03.24	2483.3	0	0	0	RTX 580	RTX 580	365	0	4	
2011	2011.01.26	1283.4	0	0	0	RTX 580 Ti	RTX 580 Ti	170	0	4	
2011	2011.05.17	1075	0	0	0	RTX 580	RTX 580	150	0	4	
2011	2011.03.15	691.2	0	0	0	RTX 580 Ti	RTX 580 Ti	116	0	4	
2011	2011.02.21	311	0	0	0	RTX 460	RTX 460	50	0	4	
2011	2011.05.14	298.8	0	0	0	RTX 550	RTX 550	50	0	4	
2011	2011.04.12	152.5	0	0	0	RTX 450	RTX 450	29	0	4	
2010	2010.11.09	1581.1	0	0	0	RTX 580	RTX 580	244	0	4	
2010	2010.12.07	1405.4	0	0	0	RTX 570	RTX 570	230	0	4	
2010	2010.03.25	1345	0	0	0	RTX 480	RTX 480	200	0	4	
2010	2010.03.25	1088.6	0	0	0	RTX 470	RTX 470	215	0	4	
2010	2010.07.12	907.2	0	0	0	RTX 480	RTX 480	180	0	4	
2010	2010.08.13	601.3	0	0	0	RTX 450	RTX 450	106	0	4	
2010	2010.10.11	298.8	0	0	0	RTX 450	RTX 450	89	0	4	
2010	2010.02.15	257.3	0	0	0	RTX 340	RTX 340	89	0	4	
2010	2010.02.15	240	0	0	0	RTX 320	RTX 320	75	0	4	
2010	2010.02.15	187.5	0	0	0	RTX 320	RTX 320	43	0	4	
2010	2010.09.03	134.4	0	0	0	RTX 420	RTX 420	25	0	4	
2009	2009.01.08	1192.3	0	0	0	RTX 285	RTX 285	289	0	4	
2009	2009.01.15	708.48	0	0	0	RTX 285	RTX 285	204	0	4	
2009	2009.03.03	470	0	0	0	RTX 280	RTX 280	150	0	4	
2009	2009.03.10	470	0	0	0	RTX 130	RTX 130	141	0	4	
2009	2009.03.10	208	0	0	0	RTX 140	RTX 140	183	0	4	
2009	2009.10.12	130.6	0	0	0	RTX 220	RTX 220	58	0	4	
2009	2009.03.10	120	0	0	0	RTX 130	RTX 130	75	0	4	
2009	2009.03.10	89.6	0	0	0	RTX 120	RTX 120	60	0	4	
2009	2009.10.12	44.8	0	0	0	RTX 110	RTX 110	30.5	0	4	
2009	2009.11.27	44.8	0	0	0	RTX 110	RTX 110	30.5	0	4	
2009	2009.11.25	22.4	0	0	0	RTX 100	RTX 100	25	0	4	
2009	2009.03.10	22.4	0	0	0	RTX 100	RTX 100	25	0		















Name	Release date	Release price
GTX 1080 Ti	03/10/2017	\$699
GTX 1070 Ti	11/02/2017	\$399
GTX 1060	07/19/2016	\$299
GTX 1050 Ti	10/25/2016	\$139
GTX 1050	10/25/2016	\$109
GT 1030	05/17/2017	\$79
GTX 1080	05/27/2016	\$599
GTX 1070	06/10/2016	\$379
GTX 1060	07/19/2016	\$299
GTX 1660 Super	10/29/2019	\$229
GTX 1660 Ti	02/22/2019	\$279
GTX 1660	03/14/2019	\$219
GTX 1650 Super	11/22/2019	\$159
GTX 1650	04/23/2019	\$149