

GPU性能仕様比較表

製品名	AMD Instinct MI300X	AMD Instinct MI300A APU	AMD Instinct MI250X	AMD Instinct MI250	AMD Instinct MI210	NVIDIA H200 for SXM5	NVIDIA H100 for SXM5	NVIDIA H100 NVL	NVIDIA H100 for PCIe	NVIDIA L40S	NVIDIA L40	NVIDIA L4	NVIDIA RTX 6000 Ada	NVIDIA RTX 5000 Ada	NVIDIA RTX 4500 Ada	NVIDIA RTX 4000 Ada	NVIDIA RTX 4000 SFF Ada	GeForce RTX4090	GeForce RTX4080	GeForce RTX4070 Ti	GeForce RTX4070	NVIDIA A800 40GB Active	NVIDIA A100 for PCIe 80GB	NVIDIA A100 for PCIe 40GB	NVIDIA A40	NVIDIA A30		
製品画像																												
GPUアーキテクチャ	CDNA3	CDNA3	CDNA2	CDNA2	CDNA2	Hopper	Hopper	Hopper	Hopper	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ada Lovelace	Ampere	Ampere	Ampere	Ampere	Ampere		
製造プロセス (nm)	TSMC 5/6nm	TSMC 5/6nm	TSMC 6nm	TSMC 6nm	TSMC 6nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 5nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 5nm	TSMC 5nm	TSMC 5nm	TSMC 5nm	TSMC 7nm	TSMC 7nm	TSMC 7nm	Samsung 8nm	TSMC 7nm
トランジスタ数 (億)	153	146	58.2	58.2	58.2	80	80	80	80	76.3	76.3	35.8	76.3	76.3	76.3	35.8	35.8	76.3	45.9	35.8	35.8	54.2	54.2	54.2	28.3	54.2		
チップのコードネーム	Aqua Vanjaram	Aqua Vanjaram	Aldebaran	Aldebaran	Aldebaran	GH100	GH100	GH100	GH100	AD102	AD102	AD104	AD102	AD102	AD102	AD102	AD104	AD102	AD103	AD104	AD104	GA100	GA100	GA100	GA102	GA100		
Compute Capability						9.0	9.0	9.0	9.0	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.0	8.0	8.0	8.6	8.0	
ベースクロック(MHz)	1216	912	1000	1000	1000	1,665	1,665	1,095	1,095	1065	735	795	915	1,155			720	2,235	2,205	2,310	1,920	765	1,065	765	1,305	930		
GPU Boost クロック(MHz)	2100	2100	1700	1700	1700	1,830 1,980	1,830 1,980	1,830 1,980	1,620	2520	2,490	2,040	2,505	2,550	2,580	2,175	1,560	2,520	2,505	2,610	2,475	1,410	1,410	1,410	1,740	1,440		
CUDAコア数	38,912	29,184	14,080	13,312	6,656	16,896	16,896	16,896	14,592	18,176	18,176	7,424	18,176	12,800	7,680	6,144	6,144	16,384	9,728	7,680	5,888	6,912	6,912	6,912	10,752	3,584		
Tensorコア数	1,216	912	880	832	416	528	528	528	456	568	568	232	568	400	240	192	192	512	304	240	184	432	432	432	336	224		
RTコア数	0	0	0	0	0	0	0	0	0	142	142	58	142	100	60	48	48	128	76	60	50	0	0	0	84	0		
SM(CU)数	304	228	220	208	104	132	132	132	114	142	142	58	142	100	60	48	48	128	76	60	46	108	108	108	84	56		
L1キャッシュ容量 (SM毎)	32KB	32KB	16KB	16KB	16KB	256KB	256KB	256KB	256KB	142	142	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	192KB	192KB	192KB	128KB	192KB	
L2キャッシュ容量	32MB(+256MB)	24MB(+256MB)	16MB	16MB	16MB	50MB	50MB	50MB	50MB	96MB	96MB	48MB	96MB	72MB	72MB	48MB	48MB	96MB	64MB	48MB	36MB	40MB	40MB	40MB	6MB	24MB		
メモリ容量	192GB HBM3	128GB HBM3	128GB HBM2e	128GB HBM2e	64GB HBM2e	141GB HBM3e	80GB HBM3	94GB HBM2e	80GB HBM2e	48GB GDDR6	48GB GDDR6	24GB GDDR6X	48GB GDDR6	32GB GDDR6	24GB GDDR6	20GB GDDR6	20GB GDDR6	24GB GDDR6X	16GB GDDR6X	12GB GDDR6X	12GB GDDR6X	40GB HBM2	80GB HBM2	40GB HBM2	48GB GDDR6	24GB HBM2		
メモリバス(bit)	8,192	8,192	8,192	8,192	4,096	5,120	5,120	6,144	5,120	384	384	192	384	256	192	160	160	384	256	192	192	5,120	5,120	5,120	384	3,072		
メモリ帯域幅(GB/s)	5,325	5,325	3,277	3,277	1,638	4,800	3,352	3,917	2,048	864	864	300	960	576	432	360	280	1,008	717	504	504	1,555	1,935	1,555	696	933		
ECC機能サポート	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	非対応	非対応	非対応	非対応	対応	対応	対応	対応	対応		
バスインターフェース	PCI-Express 5.0 x16	PCI-Express 5.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 5.0 x16	PCI-Express 5.0 x16	PCI-Express 5.0 x16	PCI-Express 5.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16		
TDP(W)	750	550 (Peak 760)	560	500	300	700	700	400	300~350	350	300	72	300	250	210	130	70	450	320	285	200	240	300	250	300	165		
放熱機構	Passive	-	Passive	Passive	Passive	-	-	Passive	Passive	Passive	Passive	Passive	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Passive	Passive	Passive	Passive		
NVLink	対応	対応	対応	対応	対応	対応	対応	対応	対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	対応	対応	対応	対応		
補助電源	不要	不要	不要	不要	CPU(EPS) 8pin	不要	不要	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	不要	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	不要	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	16-pin (12VHPWR)	CPU(EPS) 8pin	CPU(EPS) 8pin	CPU(EPS) 8pin	CPU(EPS) 8pin	
SLOT数	OAM	CPU	OAM	OAM	2	SXM	SXM	2	2	2	2	1	2	2	2	1	2	3	3	2	2	2	2	2	2	2		
サイズ(高さ,長さ)	OAM	CPU	OAM	OAM	FH10.5"L	SMX	SMX	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	LP	FH10.5"L	FH10.5"L	FH10.5"L	FH9.5"L	LP	137H304L	140H336L	FH11.2"L	FH9.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L		
FP64(TFLOPS)	81.7	61.3	47.9	45.3	22.6	33.5	33.5	33.5	25.6	1.4	1.4	0.473	1.423	1.020	0.619	0.418	0.300	1.29	0.762	0.626	0.455	9.7	9.7	9.7	0.585	5.2		
FP64 Tensor Core(TFLOPS)	163	123	95.7	90.5	45.3	66.9	66.9	66.9	51.2	-	-	-	-	-	-	-	-	-	-	-	-	19.5	19.5	19.5	-	10.3		
FP32 (TFLOPS)	163	123	47.9	45.3	22.6	66.9	66.9	66.9	51.2	91.6	90.5	30.3	91.1	65.3	39.6	26.7	19.2	82.6	48.7	40.1	29.1	19.5	19.5	19.5	37.4	10.3		
TF32 Tensor Core(TFLOPS)	654	490	-	-	-	495	495	495	378	183	90.5	61	182	131	79	53	38	82.6	48.7	40.1	29.1	156	156	156	74.8	82.6		
TF32 Tensor Core(TFLOPS) ※	1307	981	-	-	-	989	989	989	756	366	181	121	364	261	159	107	77	165	97.5	80.2	58.3	312	312	312	150	165		
FP16 (TFLOPS)						134	134	134	102	91.6	90.52	30.3	91.1	65.3	39.6	26.7	19.2	82.6	48.7	40.09	29.1	78.0	78.0	78.0	37.4	41.3		
FP16 Tensor Core(TFLOPS) (FP16集積)	1307	981	383	362	181	989	989	989	756	366	362	121	364	261	159	107	77	330	195	160	117	312	312	312	150	165		
FP16 Tensor Core(TFLOPS) ※ (FP16集積)	2614	1962	-	-	-	1979	1979	1979	1513	733	724	242	728	522	317	214	153	661	390	321	233	624	624	624	299	330		
FP16 Tensor Core(TFLOPS) ※ (FP32集積)	1307	981	383	362	181	989	989	989	756	366	181	121	364	261	159	107	77	165	97.5	80.2	58	312	312	312	150	165		
FP16 Tensor Core(TFLOPS) ※ (FP32集積)	2614	1962	-	-	-	1979	1979	1979	1513	733	362	242	728	522	317	214	153	330	195	160	117	624	624	624	299	330		
BFLOAT16 (TFLOPS)	-	-	-	-	-	134	134	134	102	91.6	90.5	30.3	91.1	65.3	39.6	26.7	19.2	82.6	48.7	40.1	29.1	39	39	39	37.4	20.6		
BFLOAT16 Tensor Core(TFLOPS) (FP32集積)	1307	981	383	362	181	989	989	989	756	366	181	121	364	261	159	107	77	165	97.5	80	58	312	312	312	150	165		
BFLOAT16 Tensor Core(TFLOPS) ※ (FP32集積)	2614	1962	-	-	-	1979	1979	1979	1513	733	362	242	728	522	317	214	153	330	195	160	117	624	624	624	299	330		
FP8 Tensor Core(TFLOPS)	2615	1961	-	-	-	1979	1979	1979	1513	733	362	242	728	522	317	214	153	661	390	321	233	-	-	-	-	-		
FP8 Tensor Core(TFLOPS) ※	5230	3922	-	-	-	3958	3958	3958	3026	1466	724	485	1457	1044	634	428	307	1321	780	641	466	-	-	-	-	-		
INT8 Tensor Core(TOPS)	2615	1961	383	362	181	1979	1979	1979	1513	733	362	242	728	522	317	214	153	661	390	321	233	624	624	624	299	330		
INT8 Tensor Core(TOPS) ※			-	-	-	3958	3958	3958	3026	1466	724	485	1457	1044	634	428	307	1321	780	641	466	1247	1247	1247	599	661		
INT4 Tensor Cor(TOPS)			383	362	181	-	-	-	-	733	724	485	1457	1044	634	428	307	1321	780	641	466	1247	1247	1247	599	661		
INT4 Tensor Core(TOPS) ※			-	-	-	-	-	-	-	1466	1448	969	2914	2089	1268	855	613	2642	1560	1283	933	2495	2495	2495	1197	1321		

GPU性能仕様比較表

製品名	NVIDIA A10	NVIDIA V100S	NVIDIA V100	NVIDIA RTX A6000	NVIDIA RTX A5500	NVIDIA RTX A5000	NVIDIA RTX A4500	NVIDIA RTX A4000	NVIDIA RTX A2000 12GB	NVIDIA RTX A2000	GeForce RTX3090 Ti	GeForce RTX3080	GeForce RTX3070	GeForce RTX3060 Ti	GeForce RTX3060	GeForce RTX2080 Ti	GeForce RTX2080	GeForce GTX 1080 Ti	GeForce GTX TITAN X	NVIDIA T1000 8GB	NVIDIA T1000	NVIDIA T600	NVIDIA T400	NVIDIA T400 4GB	Quadro RTX 8000	Quadro RTX 6000	
製品画像																											
GPUアーキテクチャ	Ampere	Volta	Volta	Ampere	Ampere	Ampere	Ampere	Ampere	Ampere	Ampere	Ampere	Ampere	Ampere	Ampere	Ampere	Turing	Turing	Pascal	maxwell	Turing	Turing	Turing	Turing	Turing	Turing	Turing	
製造プロセス (nm)	Samsung 8nm	TSMC 12nm	TSMC 12nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	Samsung 8nm	TSMC 12nm	TSMC 12nm	TSMC 16nm	TSMC 28nm	TSMC 12nm	TSMC 12nm	TSMC 12nm	TSMC 12nm	TSMC 12nm	TSMC 12nm	TSMC 12nm	
トランジスタ数 (億)	28.3	21.1	21.1	28.3	28.3	28.3	28.3	17.4	12.0	12.0	28.3	28.3	17.4	17.4	12.0	18.6	13.6	11.8	8.0	4.7	4.7	4.7	4.7	4.7	18.6	18.6	
チップのコードネーム	GA102	GV100	GV100	GA102	GA102	GA102	GA102	GA104	GA106	GA106	GA102	GA102	GA104	GA104	GA106	TU102	TU104	GP102	GM200	TU117	TU117	TU117	TU117	TU117	TU102	TU102	
Compute Capability	8.6	7.0	7.0	8.6	8.6	8.6	8.6	8.6	8.6	8.6	8.6	8.6	8.6	8.6	8.6	7.5	7.5	6.1	5.2	7.5	7.5	7.5	7.5	7.5	7.5	7.5	
ベースクロック(MHz)	885	1,245	1,230	1,410	1,170	1,170	1,065	735	562	562	1,440	1,440	1,500	1,320	1,410	1,350	1,515	1480	1000	1065	1065	735	420	420	1,395	1,440	
GPU Boost クロック(MHz)	1,695	1,597	1,380	1,800	1,695	1,695	1,650	1,560	1,200	1,200	1,860	1,710	1,725	1,777	1,665	1,635	1,800	1582	1075	1395	1395	1335	1425	1425	1,770	1,770	
CUDAコア数	9,216	5,120	5,120	10,752	10,240	8,192	7,168	6,144	3,328	3,328	10,752	8,704	5,888	4,864	3,584	4,352	2,944	3,584	1,536	896	896	640	384	384	4,608	4,608	
Tensorコア数	288	640	640	336	320	256	224	192	104	104	336	272	184	152	112	544	368	0	0	0	0	0	0	0	576	576	
RTコア数	72	0	0	84	80	64	56	48	26	26	84	68	46	38	28	68	46	0	0	0	0	0	0	0	72	72	
SM(CU)数	72	80	80	84	80	64	56	48	26	26	84	68	46	38	28	68	46	28	24	14	14	10	6	6	72	72	
L1キャッシュ容量 (SM毎)	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	128KB	64KB	64KB	48KB	48KB	64KB	64KB	64KB	64KB	64KB	64KB	64KB	64KB
L2キャッシュ容量	6MB	6MB	6MB	6MB	6MB	6MB	6MB	4MB	3MB	3MB	6MB	5MB	4MB	4MB	3MB	5MB	5MB	2.75MB	3MB	1MB	1MB	1MB	1MB	1MB	6MB	6MB	
メモリ容量	24GB GDDR6	32GB HBM2	32GB HBM2	48GB GDDR6	24GB GDDR6	24GB GDDR6	20GB GDDR6	16GB GDDR6	12GB GDDR6	6GB GDDR6	24GB GDDR6X	10GB GDDR6X	8GB GDDR6	8GB GDDR6	12GB GDDR6	11GB GDDR6X	8GB GDDR6X	11GB GDDR5X	12GB GDDR5	8GB GDDR6	4GB GDDR6	4GB GDDR6	2GB GDDR6	4GB GDDR6	48GB GDDR6	24GB GDDR6	
メモリバス(bit)	384	4,096	4,096	384	384	384	320	256	192	192	384	320	256	256	192	352	256	352	384	128	128	128	64	64	384	384	
メモリ帯域幅(GB/s)	600	1,133	897	768	768	768	640	448	288	288	1,008	760	448	448	360	616	448	484	336.5	160	160	160	80	80	672	672	
ECC機能サポート	対応	対応	対応	対応	対応	対応	対応	対応	対応	対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	対応	対応
バスインターフェース	PCI-Express 4.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 4.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	
TDP(W)	150	250	250	300	230	230	200	140	70	70	450	320	220	200	170	260	215	250	250	50	50	40	30	30	295	295	
放熱機構	Passive	Passive	Passive	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active+水冷	Active	Active	Active	Active	Active	Active	Active	Active
NVLink	非対応	非対応	非対応	対応	対応	対応	対応	非対応	非対応	非対応	対応	非対応	非対応	非対応	非対応	非対応	非対応	非対応	対応	非対応	非対応	非対応	非対応	非対応	非対応	対応	対応
補助電源	PCIe 8pin	2x PCIe 8pin	2x PCIe 8pin	CPU(EPS) 8pin	PCIe 8pin	PCIe 8pin	PCIe 8pin	PCIe 6pin	不要	不要	2x PCIe 8pin	2x PCIe 8pin	2x PCIe 8pin	PCIe 8pin	PCIe 8pin	PCIe 8pin + PCIe 6pin	PCIe 8pin + PCIe 6pin	PCIe 8pin + PCIe 6pin	2x PCIe 8pin	不要	不要	不要	不要	不要	PCIe 8pin + PCIe 6pin	PCIe 8pin + PCIe 6pin	
SLOT数	1	2	2	2	2	2	2	2	2	2	3	3	2	2	2	2	2	2	2	1	1	1	1	1	2	2	
サイズ(高さ、長さ)	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH9.5"L	LP	LP	140H336L	FH11.2"L	FH9.5"L	FH9.5"L	FH9.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	LP	LP	LP	LP	LP	FH10.5"L	FH10.5"L	
FP64(TFLOPS)	0.488	8.2	7.1	0.605	0.542	0.434	0.370	0.300	0.125	0.125	0.625	0.465	0.317	0.270	0.186	0.445	0.331	0.354	0.103	0.078	0.078	0.053	0.034	0.034	0.510	0.510	
FP64 Tensor Core(TFLOPS)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
FP32 (TFLOPS)	31.2	16.4	14.1	38.7	34.7	27.8	23.7	19.2	7.99	7.99	40.0	29.8	20.3	17.3	11.9	14.2	10.6	11.3	3.3	2.5	2.5	1.7	1.1	1.1	16.3	16.3	
TF32 Tensor Core(TFLOPS)	62.5	-	-	77.4	69.4	56	47.3	38	16.0	16	40.0	29.8	20.3	17.3	11.9	-	-	-	-	-	-	-	-	-	-	-	
TF32 Tensor Core(TFLOPS) ※	125	-	-	155	139	111	94.6	77	31.9	32	80.0	59.5	40.6	34.6	23.9	-	-	-	-	-	-	-	-	-	-	-	
FP16 (TFLOPS)	31.2	32.7	28.3	38.7	34.7	27.8	23.7	19.2	7.99	7.99	40.0	29.8	20.3	17.3	11.9	28.5	21.2	0.177	-	5.0	5.0	3.42	2.19	2.19	32.6	32.6	
FP16 Tensor Core(TFLOPS) (FP16集積)	125	131	113	155	139	111	95	77	32	32	160	119	81.3	69.1	47.7	114	85	-	-	-	-	-	-	-	130	130	
FP16 Tensor Core(TFLOPS) ※ (FP16集積)	250	-	-	310	278	222	189	153	64	64	320	238	163	138	95	-	-	-	-	-	-	-	-	-	-	-	
FP16 Tensor Core(TFLOPS) (FP32集積)	125	131	113	155	139	111	94.6	77	31.9	32	80	59.5	40.6	35	24	57	42	-	-	-	-	-	-	-	130	130	
FP16 Tensor Core(TFLOPS) ※ (FP32集積)	250	-	-	310	278	222	189	153	63.9	64	160	119	81.3	69.1	47.7	-	-	-	-	-	-	-	-	-	-	-	
BFLOAT16 (TFLOPS)	31.2	-	-	38.7	34.7	27.8	23.7	19.2	7.99	7.99	40.0	29.8	20.3	17.3	11.9	-	-	-	-	-	-	-	-	-	-	-	
BFLOAT16 Tensor Core(TFLOPS) (FP32集積)	125	-	-	155	139	111	95	77	32	32	80	59.5	40.6	34.6	23.9	-	-	-	-	-	-	-	-	-	-	-	
BFLOAT16 Tensor Core(TFLOPS) ※ (FP32集積)	250	-	-	310	278	222	189	153	64	64	160	119	81.3	69.1	47.7	-	-	-	-	-	-	-	-	-	-	-	
FP8 Tensor Core(TFLOPS)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
FP8 Tensor Core(TFLOPS) ※	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
INT8 Tensor Core(TOPS)	250	-	-	310	278	222	189	153	64	64	320	238	163	138	95	228	170	-	-	-	-	-	-	-	261	261	
INT8 Tensor Core(TOPS) ※	500	-	-	619	555	444	378	307	128	128	640	476	325	277	191	-	-	-	-	-	-	-	-	-	-	-	
INT4 Tensor Cor(TOPS)	500	-	-	619	555	444	378	307	128	128	640	476	325	277	191	455	339	-	-	-	-	-	-	-	522	522	
INT4 Tensor Core(TOPS) ※	1000	-	-	1239	1111	889	757	613	256	256	1280	953	650	553	382	-	-	-	-	-	-	-	-	-	-	-	

※表中の(※)はスパース性高速化有効時の数値になります。※数値は理論性能値です。使用するデバイスや環境、その他の要因によって実性能値は変動します。

GPU性能仕様比較表

製品名	Quadro RTX 5000	Quadro RTX 4000	Quadro GV100	Quadro P6000	Quadro P5000	Quadro GP100	Quadro P2000	Quadro P1000	Quadro P600	Quadro P400	Tesla T4	Tesla P100 SXM2	Tesla P100 PCIe	Tesla P40	Tesla P4	TITAN RTX	TITAN V	TITAN Xp	NVIDIA Titan X
製品画像																			
GPUアーキテクチャ	Turing	Turing	Volta	Pascal	Pascal	Pascal	Pascal	Pascal	Pascal	Pascal	Turing	Pascal	Pascal	Pascal	Pascal	Turing	Volta	Pascal	Pascal
製造プロセス (nm)	TSMC 12nm	TSMC 12nm	TSMC 12nm	TSMC 16nm	TSMC 16nm	TSMC 16nm	TSMC 16nm	Samsung 14nm	Samsung 14nm	Samsung 14nm	TSMC 12nm	TSMC 16nm	TSMC 16nm	TSMC 16nm	TSMC 16nm	TSMC 12nm	TSMC 12nm	TSMC 16nm	TSMC 16nm
トランジスタ数 (億)	13.6	13.6	21.1	11.8	7.2	15.3	4.4	3.3	3.3	3.3	13.6	15.3	15.3	11.8	7.2	18.6	21.1	11.8	11.8
チップのコードネーム	TU104	TU104	GV100	GP102	GP104	GP100	GP106	GP107	GP107	GP107	TU104	GP100	GP100	GP102	GP104	TU102	GV100	GP102	GP102
Compute Capability	7.5	7.5	7.0	6.1	6.1	6.0	6.1	6.1	6.1	6.1	7.5	6.0	6.0	6.1	6.1	7.5	7.0	6.1	6.1
ベースクロック(MHz)	1,620	1,005	1,132	1506	1607	1,304	1,076	1,354	1,329	1,228	660	1328	1189	1303	810	1350	1200	1426	1417
GPU Boost クロック(MHz)	1,815	1,545	1,450	1645	1733	1,442	1,480	1,392	1,557	1,252	1,590	1480	1328	1531	1063	1770	1455	1582	1531
CUDAコア数	3,072	2,304	5,120	3,840	2,560	3,584	1,024	640	384	256	2,560	3,584	3,584	3,840	2,560	4,608	5,120	3,584	3,584
Tensorコア数	384	288	640	0	0	0	0	0	0	0	320	0	0	0	0	576	640	0	0
RTコア数	48	36	0	0	0	0	0	0	0	0	40	0	0	0	0	72	0	0	0
SM(CU)数	48	36	80	30	20	56	8	5	3	2	40	56	56	30	20	72	80	28	28
L1キャッシュ容量 (SM毎)	64KB	64KB	128KB	48KB	48KB	24KB	48KB	48KB	48KB	48KB	64KB	24KB	24KB	48KB	48KB	64KB	96KB	48KB	48KB
L2キャッシュ容量	4MB	4MB	6MB	3MB	2MB	4MB	128KB	1MB	1MB	512KB	4MB	4MB	4MB	3MB	2MB	6MB	4.5MB	3MB	3MB
メモリ容量	16GB GDDR6	8GB GDDR6	32GB HBM2	24GB GDDR5X	16GB GDDR5X	16GB HBM2	5GB GDDR5	4GB GDDR5	2GB GDDR5	2GB GDDR5	16GB DDR6	16GB HBM2	16GB HBM2	24GB GDDR5	8GB GDDR5	24GB GDDR6	12GB HBM2	12GB GDDR5X	12GB GDDR5X
メモリバス(bit)	256	256	4,096	384	256	4,096	160	128	128	64	256	4096	4096	384	256	384	3072	384	384
メモリ帯域幅(GB/s)	448	416	868	433	288	732	140	96	64	32	320	732	732	694	192	672	652.8	480	480
ECC機能サポート	対応	非対応	対応	対応	対応	対応	非対応	非対応	非対応	非対応	対応	対応	対応	対応	対応	非対応	非対応	非対応	非対応
バスインターフェース	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16	PCI-Express 3.0 x16
TDP(W)	265	160	250	250	180	235	75	47	40	30	70	300	250	250	75	280	250	250	250
放熱機構	Active	Active	Active	Active	Active	Active	Active	Active	Active	Active	Passive	Passive	Passive	Passive	Passive	Active	Active	Active	Active
NVLink	対応	非対応	対応	非対応	非対応	対応	非対応	非対応	非対応	非対応	非対応	対応	非対応	非対応	非対応	対応	非対応	非対応	非対応
補助電源	PCIe 8pin + PCIe 6pin	PCIe 8pin	PCIe 8pin	PCIe 8pin + PCIe 6pin	PCIe 8pin	PCIe 8pin	不要	不要	不要	不要	不要	不要	CPU(EPS) 8pin	CPU(EPS) 8pin	不要	2x PCIe 8pin	PCIe 8pin + PCIe 6pin	PCIe 8pin + PCIe 6pin	PCIe 8pin + PCIe 6pin
SLOT数	2	1	2	2	2	2	1	1	1	1	1	SXM	2	2	1	2	2	2	2
サイズ(高さ,長さ)	FH10.5"L	FH9.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L	FH7.7"L	LP	LP	LP	LP	SMX	FH10.5"L	FH10.5"L	LP	FH10.5"L	FH10.5"L	FH10.5"L	FH10.5"L
FP64(TFLOPS)	0.348	0.222	7.4	0.395	0.277	5.2	0.095	0.056	0.037	0.020	0.254	5.3	4.76	0.367	0.170	0.510	7.45	0.354	0.343
FP64 Tensor Core(TFLOPS)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FP32 (TFLOPS)	11.2	7.1	14.8	12.6	8.9	10.3	3.0	1.8	1.2	0.64	8.1	10.6	9.5	11.8	5.44	16.3	14.9	11.3	11.0
TF32 Tensor Core(TFLOPS)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TF32 Tensor Core(TFLOPS) ※	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FP16 (TFLOPS)	22.3	14.2	29.7	0.197	0.139	20.7	6.1	3.6	2.4	1.3	16.3	21.2	19.0	0.184	0.085	32.6	29.8	0.177	0.171
FP16 Tensor Core(TFLOPS) (FP16集積)	89	57	119	-	-	-	-	-	-	-	65	-	-	-	-	130	119	-	-
FP16 Tensor Core(TFLOPS) ※ (FP16集積)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FP16 Tensor Core(TFLOPS) (FP32集積)	89	57	119	-	-	-	-	-	-	-	65	-	-	-	-	130	119	-	-
FP16 Tensor Core(TFLOPS) ※ (FP32集積)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BFLOAT16 (TFLOPS)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BFLOAT16 Tensor Core(TFLOPS) (FP32集積)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BFLOAT16 Tensor Core(TFLOPS) ※ (FP32集積)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FP8 Tensor Core(TFLOPS)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FP8 Tensor Core(TFLOPS) ※	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
INT8 Tensor Core(TOPS)	178	114	-	-	-	-	-	-	-	-	130	-	-	-	-	261	-	-	-
INT8 Tensor Core(TOPS) ※	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
INT4 Tensor Cor(TOPS)	357	228	-	-	-	-	-	-	-	-	261	-	-	-	-	522	-	-	-
INT4 Tensor Core(TOPS) ※	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

※表中の(※)はスパース性高速化有効時の数値になります。※数値は理論性能値です。使用するデバイスや環境、その他の要因によって実性能値は変動します。