Linux Plex Media Server Cheat Sheet (26-April-2019) ashley@elpamsoft.com

No warranty. Use at own risk. Results may vary. Some information from 3rd party sources. No copyrights implied. See your doctor if problems persist.

GPU Transcoding using nVidia NVENC and NVENC

Graphics Card Model	Chipset Family	Chip Model	# of Chips	Session Limit	NVENC FPS				NVDEC FPS			VRAM	Streams for VRAM*	VRAM Streams Card Load * (Less than 100% Best)	Recommended Plex Streams	VDPAU	Notes
					Chips	h264 720p	h264 1080p	HEVC	Chips	H264	HEVC						
GeForce GT 630 - 640 GeForce GT 710 - 730	Kepler	GK208	1	2	1	157-490	78-220	-	1	161	-	1GB	4	74%	4	D	NVENC: No H264 4:4:4 or Lossless on Kepler K620 Can be GK107 or GM107
GeForce GT 630 - 640 GeForce GTX 650, GT 740		GK107	1	2	1	157-490	78-220	-	1	161	-	1-2GB	4-8	74% - 149%	4-5		
Quadro K420, K600			1	2	1	157-490	78-220	-	1	161	-	1GB	4	74%	4		
Quadro K620			1	2	1	157-490	78-220	-	1	161	-	1-2GB	4-8	74% - 149%	4-5		
Quadro K2000			1	U	1	157-490	78-220	-	1	161	-	2GB	8	149%	5		
GeForce GTX 645 -660 Ti Boost GeForce GT 740		GK106	1	2	1	157-490	78-220	-	1	161	-	1-2GB	4-8	74% - 149%	4-5		
Quadro K4000			1	U	1	157-490	78-220	-	1	161	-	3GB	12	224%	5		
GeForce GTX 660 - 690 GeForce GTX 760 - 770		GK104	1	2	1	157-490	78-220	-	1	161	-	2-8GB	8-32	149% - 596%	5		
Quadro K4200-K5000			1	U	1	157-490	78-220	-	1	161	-	4-8GB	16-32	298% - 596%	5		
GeForce GT 780 - 780 Ti	Kepler (2nd Gen)	GK110	1	2	1	157-490	78-220	-	1	161	-	3-6GB	12-25	224% - 466%	5		Kepler 2nd Gen
GeForce GTX Titan / Titan Black			1	2	1	157-490	78-220	-	1	161	-	6GB	25	466%	5		
GeForce GTX Titan Z			2	2	2	157-490	156-440	-	2	322	-	6GB	25	233%	10		Double NVENC/NVDEC Chips
Quadro K6000		GK110B	1	U	1	157-490	78-220	-	1	161	-	12GB	49	913%	5		
GeForce GTX 745 - 750 Ti	Maxwell (1st Gen)	GM107	1	2	1	502-826	211-345	-	1	376	-	2-4GB	8-16	70%	8-12	E	PureVideo hardware and software running on the shader array to decode HEVC (H.265) as partial/hybrid hardware video decoding.
Quadro K1200			1	2	1	502-826	211-345	-	1	376	-	4GB	16	139%	12		
Quadro K2200			1	U	1	502-826	211-345	-	1	376	-	4GB	16	139%	12		
Telsa M10			1	U	1	502-826	211-345	-	1	376	-	8GB	32	287%	12		
GeFroce 830M-940MX, MX110-130		GM108	1	0	0	-	-	-	0	-	-	1-2GB	0	-	-		No NVENC/NVDEC Chips
Quadro K620M			1	0	0	-	-	-	0	-	-	2GB	0	-	-		
GeForce GTX 750, 950 - 960	Maxwell (2nd Gen)	GM206	1	2	1	641-1111	261-432	142-200	1	376	408	1-3GB	4-12	32% - 96%	4-12	F	Dedicated HEVC Main (8-bit) & Main 10 (10-bit) and VP9 hardware decoding video decoding up to 4K
Quadro M2000			1	U	1	641-1111	261-432	142-200	1	376	408	4GB	16	128%	12	F	
GeForce GTX 960 Ti - 980		GM204	1	2	2		522-864	284-400	1	376	?	2-8GB	8-33	64%-263%	8-12	E	PureVideo hardware and software running on the shader array to decode HEVC (H.265) as partial/hybrid hardware video decoding. Expect Slow HEVC using CPU
Quadro M4000-M5000 Series			1	U	2		522-864	284-400	1	376	?	8GB	33	263%	12		
GeForce GTX 980 Ti		GM200	1	2	2		522-864	284-400	1	376	?	6GB	25	199%	12		
GeForce GTX Titan X			1	2	2		522-864	284-400	1	376	?	12GB	50	391%	12		
Quadro M6000 Series			1	U	2		522-864	284-400	1	376	?	12GB	50	391%	12		
GeForce GT 1030	Pascal	GP108	1	0	0		-	-	1	658	720	2GB	0	-	-	H	No NVENC Support

GeForce GTX 1050 / 1050 Ti	Pascal	GP107	1	2	1		388-631	259-395	1	658	720	4GB	14	67%	14		Pascal adds 8K NVDEC Support and 8K NVENC on some cards??
Quadro P400-P620			1	2	1		388-631	259-395	1	658	720	2GB	7	33%	7		
Quadro P1000			1	2	1		388-631	259-395	1	658	720	4GB	14	67%	14		
GeForce GTX 1050 / 1050 Ti		GP106	1	2	1		388-631	259-395	1	658	720	2-4GB	7-14	33% - 67%	7-14		
GeForce GTX 1060			1	2	1		388-631	259-395	1	658	720	3-6GB	10-20	48% - 95%	10-20		
Quadro P2000			1	U	1		388-631	259-395	1	658	720	5GB	17	81%	17		
GeForce GTX 1060		GP104	1	2	1		388-631	259-395	1	658	720	3-6GB	10-20	48% - 95%	10-20		
GeForce GTX 1070 - 1080			1	2	2		776-1262	518-790	1	658	720	8GB	27	123%	22		Dual NVENC Cards gives double FPS, except P4000
Quadro P4000			1	U	1		388-631	259-395	1	658	720	8GB	27	128%	21		
Quadro P5000			1	U	2		776-1262	518-790	1	658	720	16GB	55	314%	22		
GeForce GTX 1080 Ti		GP102	1	2	2		776-1262	518-790	1	658	720	11GB	38	205%	22		
GeForce GTX Titan X Titan Xp			1	2	2		776-1262	518-790	1	658	720	12GB	49	223%	22		
Quadro P6000			1	U	2		776-1262	518-790	1	658	720	24GB	82	447%	22		
Titan V	Volta	GV100	1	2	3		1164-1893	595-908	1	658	720	12GB	41	187%	22	I	3x NVENC, Limited by 1x NVDEC
Titan RTX	Turing	TU102	1	2	1		446-725	595-908	1	1316	1440	24GB	82	552%	14	J	HEVC 8K encoding at 30FPS, HEVC B-Frames support and up to 25% bitrate savings for HEVC and up to 15% bitrate savings for H.264. Estimated FPS from 3rd party info
GeForce RTX 2080 Ti			1	2	1		446-725	595-908	1	1316	1440	11GB	38	276%	14		
GeForce RTX 2080		TU104	1	2	1		446-725	595-908	1	1316	1440	8GB	27	182%	14		
GeForce RTX 2060 / 2070		TU106	1	2	1		446-725	595-908	1	1316	1440	8GB	27	182%	14		
GeForce GTX 1660 Ti / 1660		TU116	1	2	1		446-725	595-908	1	1316	1440	6GB	20	135%	14		
GeForce GTX 1650	Volta	TU117	1	U	1		388-631	595-908	1	658	720	4GB	14	108%	13		
Intel 2xxx, G440-G870 (except non GPU Models)	Sandy Bridge	HD 2000, HD 3000	1	U	1		120	-	1	120	-	-			4		h.264 only. No transcoders on Pentium or Celeron Series
Intel 32xx, 33xx, 34xx, 35xx, 37xx, G16x, G2xx	Ivy Bridge	HD 2500, HD 4000	1	U	1		206		1	206		-			4		Poor image quality encoding
Intel 41xx, 46xx, 47xx, G3xx, G18x	Haswell	HD 4200-4600, HD 5000	1	U	1		166		1	166		-			4		Performance regression. H.264/MPEG-4 AVC, VC-1 and H.262/MPEG-2 Part 2
Intel 57xx, 56xx, 68xxK, 69xxK+X (except non GPU Models)	Broadwell	HD 5300-6300P	1	U	1				1			-			4		VP8 hardware decoding
Intel 6xxx Only	Skylake	HD510, 530, 580	1	U	1			150	1		150	-			4		h.265 8bit Support
Intel 7xxx-8xxx (except 79xx) (not 7640X, E3-1xx0)	Kaby, Coffee, Wiskey Lake	HD610-HD640	1	U	1				1			-			4		h.265/HEVC Main10/10-bit VP9 8-bit and 10-bit
Intel (except non GPU Models)	Ice Lake	Gen11	1	U	1				1			-			4		VP9 8/10-bit HDR10 Tone Mapping
					Chips		**H264**	**HEVC**	**Chips**	**H264**	**HEVC**	**VRAM**					
Graphics Card Model	**Chipset Family**	**Chip Model**	**# of Chips**	**Session Limit**			**NVENC FPS**		**NVDEC FPS**				Streams for VRAM*	VRAM Streams Card Load * (Less than 100% Best)	**Recommended Plex Streams**	**VDPAU**	

*h.264 FPS Based on 1080p@20MBPS YUV 4:2:0 8-bit. HEVC (h.265) FPS based on 1080p@20MBPS 8-bit
*VRAM listed are common sizes.
*NVENC FPS based on Single Pass quality profile. (High Quality – High Performance)
*VDPAU nVidia PureVideo Information [https://en.wikipedia.org/wiki/Nvidia_PureVideo#Nvidia_VDPAU_Feature_Sets]

VRAM Bandwidth

Preliminary testing sees a single NVDEC job on a 128Bit GTX 1050 Ti 4GB (Pascal) unable to use more than 30% (112 FPS) of the NVDEC. Two streams hold about 50%, Three about 80% and more than four streams to reach 100% NVDEC saturation. A 256Bit GTX 970 4GB (Maxwell 2nd Gen) can hit 100% NVDEC saturation (376FPS) with a single stream.

The difference between the 128Bit, 192Bit and 256Bit Memory bandwidth needs further testing. It looks like a 128Bit memory bus will not cause performance issues with multiple transcodes but will see Plex offline "Sync" jobs only able to use 30% of the NVDEC chip.

VRAM Stream Card Load

Card load is calculated from the smallest of the "NVENC FPS" and "NVDEC FPS" then divided by "Streams for VRAM" combined FPS to provide card. For example; Quadro K2200 4GB model can NVDEC 1080p@376 FPS and NVENC 720p@502 FPS and fit a maximum 16 transcodes in its 4GB VRAM. NVDEC being the smaller FPS we take 16x h.264.1080p@30 FPS equals 480 FPS, more than the NVDEC can process. This card will be under 128% Load to deliver 16 streams causing buffering but could deliver 12 streams at 96% load. See "*Recommended Plex Streams*" for more information.

Exceeding VRAM

Exceeding the VRAM usage will cause new transcodes to buffer indefinitely, even once VRAM has dropped below the maximum. The Plex Client will need to stop the play request and request it again once VRAM usage has dropped. Choose a card with enough VRAM to avoid this.

Streams for VRAM

"Streams for VRAM" is how many Plex Streams will fit in VRAM at any one time, this figure is based on 1080p@15MBPS to 720p@4MBPS per stream.
See "*Recommended Plex Streams*" and "*Exceeding VRAM*" for stream buffering issues.

TRANSCODE PROFILE	KEPLER	MAXWELL 1ST GEN	MAXWELL 2ND GEN	PASCAL & VOLTA
720P@3MBIT TO SD 2MBIT		102MB	207MB	190MB
720P@3MBPS TO 720P 2MBPS		104MB	190MB	195MB
720P@6MBPS TO 720P@2MBPS		98MB	218MB	170MB
720P@6MBPS TO 720P@4MBPS		108MB	229MB	190MB
1080P@15MBPS TO 720P@2MBPS		122MB	231MB	280MB
1080P@15MBPS TO 720P@4MBPS		133MB	248MB	300MB
1080P@15MBPS TO 1080P@8MBPS		142MB	250MB	320MB
4K@68MBPS TO 720P@4MBPS	-	-		1220MB
4K@68MBPS TO 1080P@8MBPS	-	-		1290MB

Recommended Plex Streams

Maximum 1080p@15Mbps streams to maintain live streams without buffering pauses based on FPS of chipset and VRAM available. This calculation is based on 1080p@15MBPS to 720p@4MBPS per stream. See "*Streams for VRAM*" for more precise usage.

Multiple NVENC Chips

Some cards have either multiple NVENC chips or multiple Graphics chips each containing an NVENC chip.

Taking into account that you need to decode the same FPS as you encode let's look at the GTX 1070; It supports NVENC h.264@1262 FPS but only NVDEC h.264@658 FPS, this would mean 658 FPS@1080p is the maximum throughput with a "single pass" encoding profile, not the full 1262 FPS.

Using a "high quality" profile the GTX 1070 supports NVENC h.264@776 FPS, you will see improved image quality and smaller encoded streams while still reaching the 658 FPS@1080p throughput. Input streams lower than 1080p or that do not use NVDEC (eg. Unsupported CODEC) can allow this card to NVENC over the NVDEC limit. HEVC NVDEC on this card supports 720FPS meaning a HEVC to h.264 workload would yield a transcode of 720FPS, 60FPS better than h.264 to h.264.

GPU Power Usage

Without an attached monitor, a GTX 970 with a max draw of 145w will draw about 80w with 100% NVDEC to NVENC transcoding.

Microsoft Windows

Session Limits can be overridden in Windows [https://github.com/keylase/nvidia-patch/tree/master/win]. This can be a tricky setup, a more sturdy option choose a Quadro Card with no session limits.

Linux

Session Limits can be overridden on GeForce and Quadro Cards [https://github.com/keylase/nvidia-patch] Please make sure the card has enough VRAM for this.

NVDEC for Linux needs v1.15.1.791 or newer and a patch. [https://github.com/revr3nd/plex-nvdec] Please make sure the card has enough VRAM for this.

Linux NVENC / NVDEC Monitoring

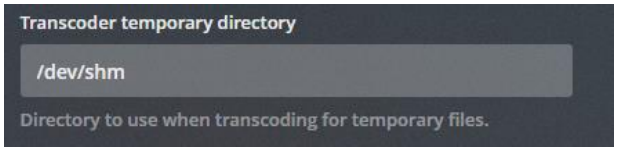
Live usage stats:
watch -n 2 'nvidia-smi -q -d UTILIZATION'

Usage Report:
nvidia-smi

Plex Media Server Optimisation

- Run Plex Library from an SSD (NVMe SSD for best performance) by making a symlink in Linux
  sudo ln -s "/SSD/Plex Media Server" "/var/lib/plexmediaserver/Library/Application Support/Plex Media Server"

- Move the "Transcoder temporary directory" to a RAM disk using the Linux default: "/dev/shm"
  /dev/shm allows by default 50% of RAM to be used as cache
  You will need about 1GB per 1080p stream for a 900 sec "Transcoder default throttle buffer"



VDPAU Reference

VP4	C	Mar-11	Supports complete acceleration for MPEG-1, MPEG-2, MPEG-4 Part 2 (a.k.a. MPEG-4 ASP), VC-1/WMV9 and H.264. Global motion compensation and Data Partitioning are not supported for MPEG-4 Part 2.
VP5	D	Apr-11	Similar to feature set C but added support for decoding H.264 with a resolution of up to 4032×4080 and MPEG-1/MPEG-2 with a resolution of up to 4032×4048 pixels.

VP6	E	Feb-14	Similar to feature set D but added support for decoding H.264 with a resolution of up to 4096×4096 and MPEG-1/MPEG-2 with a resolution of up to 4080×4080 pixels. GPUs with VDPAU feature set E support an enhanced error concealment mode which provides more robust error handling when decoding corrupted video streams. Cards with this feature set use a combination of the PureVideo hardware and software running on the shader array to decode HEVC (H.265) as partial/hybrid hardware video decoding.
VP7	F	Jan-15	Introduced dedicated HEVC Main (8-bit) & Main 10 (10-bit) and VP9 hardware decoding video decoding up to 4096 × 2304 pixels resolution.
	G		Introduced dedicated hardware video decoding of HEVC Main 12 (12-bit) up to 4096 × 2304 pixels resolution.
VP8	H	May-16	Feature Set H are capable of hardware-accelerated decoding of 8192x8192 (8k resolution) H.265/HEVC video streams.
VP9	I	Nov-17	
VP10	J	Apr-19	

Sources

https://en.wikipedia.org/wiki/Nvidia_PureVideo#Nvidia_VDPAU_Feature_Set
https://developer.nvidia.com/video-encode-decode-gpu-support-matrix
https://developer.nvidia.com/nvidia-video-codec-sdk#NVENCFeatures
http://developer.download.nvidia.com/assets/cuda/files/NVENC_DA-06209-001_v07.pdf (720p FPS Kepler to Maxwell 2nd Gen)
https://github.com/MarkRepo/NvencEncoder/blob/master/doc/NVDEC_DA-06209-001_v08.pdf (NVENC_DA-06209-001_v08 1080p FPS Kepler to Pascal)
https://github.com/MarkRepo/NvencEncoder/blob/master/doc/NVENC_DA-06209-001_v08.pdf (NVDEC_DA-06209-001_v08 1080p FPS Kepler to Pascal)

Samples

```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 418.56       Driver Version: 418.56       CUDA Version: 10.1     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  GeForce GTX 970     Off  | 00000000:01:00.0 Off |                  N/A |
| 60%   75C    P2    79W / 148W |   2758MiB /  4041MiB |     27%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                       GPU Memory |
|  GPU       PID   Type   Process name                             Usage      |
|=============================================================================|
|    0       860      C   /usr/lib/plexmediaserver/Plex Transcoder2    217MiB |
|    0      4689      C   /usr/lib/plexmediaserver/Plex Transcoder2    248MiB |
|    0     10363      C   /usr/lib/plexmediaserver/Plex Transcoder2    229MiB |
|    0     12134      C   /usr/lib/plexmediaserver/Plex Transcoder2    219MiB |
|    0     12150      C   /usr/lib/plexmediaserver/Plex Transcoder2    194MiB |
|    0     14154      C   /usr/lib/plexmediaserver/Plex Transcoder2    252MiB |
|    0     15169      C   /usr/lib/plexmediaserver/Plex Transcoder2    203MiB |
|    0     19964      C   /usr/lib/plexmediaserver/Plex Transcoder2    269MiB |
|    0     21399      C   /usr/lib/plexmediaserver/Plex Transcoder2    207MiB |
|    0     27496      C   /usr/lib/plexmediaserver/Plex Transcoder2    207MiB |
|    0     29197      C   /usr/lib/plexmediaserver/Plex Transcoder2    199MiB |
|    0     29311      C   /usr/lib/plexmediaserver/Plex Transcoder2    290MiB |
+-----------------------------------------------------------------------------+

==============NVSMI LOG==============
```

```
Timestamp                           : Sat Apr 13 20:54:10 2019
Driver Version                      : 418.56
CUDA Version                        : 10.1

Attached GPUs                       : 1
GPU 00000000:01:00.0
    Utilization
        Gpu                         : 32 %
        Memory                      : 19 %
        Encoder                     : 39 %
        Decoder                     : 84 %
    GPU Utilization Samples
        Duration                    : 18446744073709.22 sec
        Number of Samples           : 99
        Max                         : 33 %
        Min                         : 10 %
        Avg                         : 0 %
    Memory Utilization Samples
        Duration                    : 18446744073709.22 sec
        Number of Samples           : 99
        Max                         : 20 %
        Min                         : 2 %
        Avg                         : 0 %
    ENC Utilization Samples
        Duration                    : 18446744073709.22 sec
        Number of Samples           : 99
        Max                         : 51 %
        Min                         : 35 %
        Avg                         : 0 %
    DEC Utilization Samples
        Duration                    : 18446744073709.22 sec
        Number of Samples           : 99
        Max                         : 99 %
        Min                         : 81 %
        Avg                         : 0 %
```