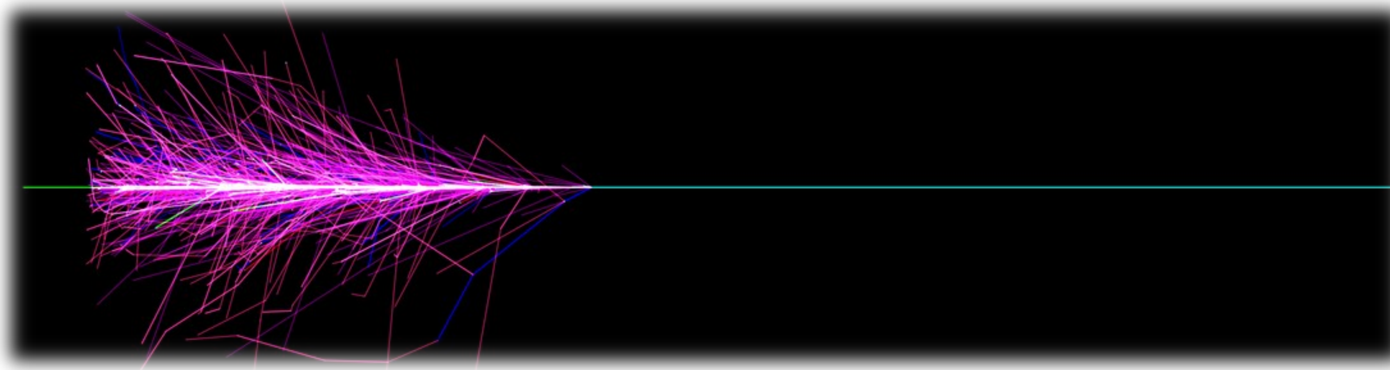
technische universität dortmund

SFB 876 Providing Information by Resource-Constrained Data Analysis

lehrstuhl physik e5



GPU Accelerators for the usage in the cosmic ray simulation CORSIKA

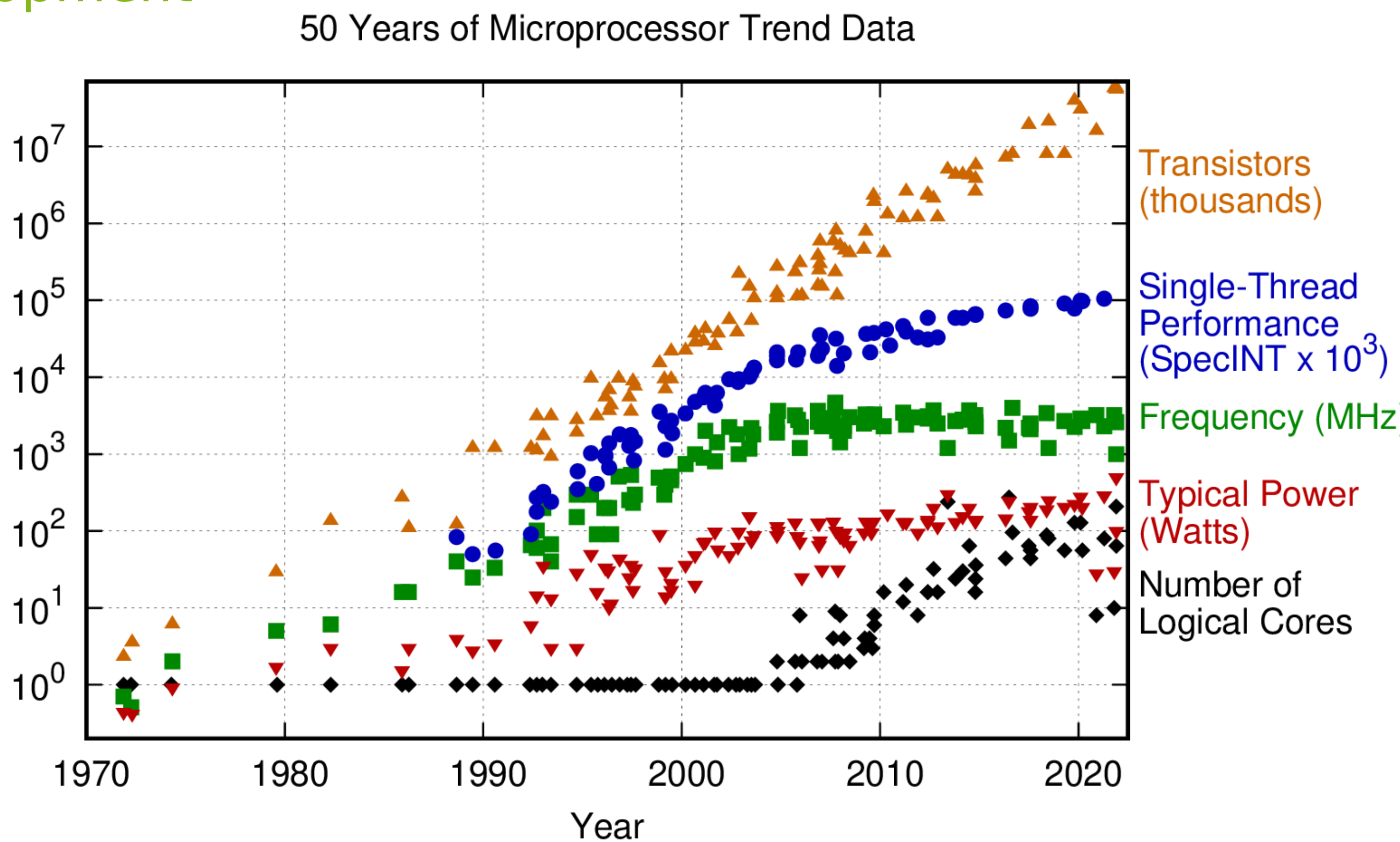
Dominik Baack

Corsika Workshop

Heidelberg - 2022

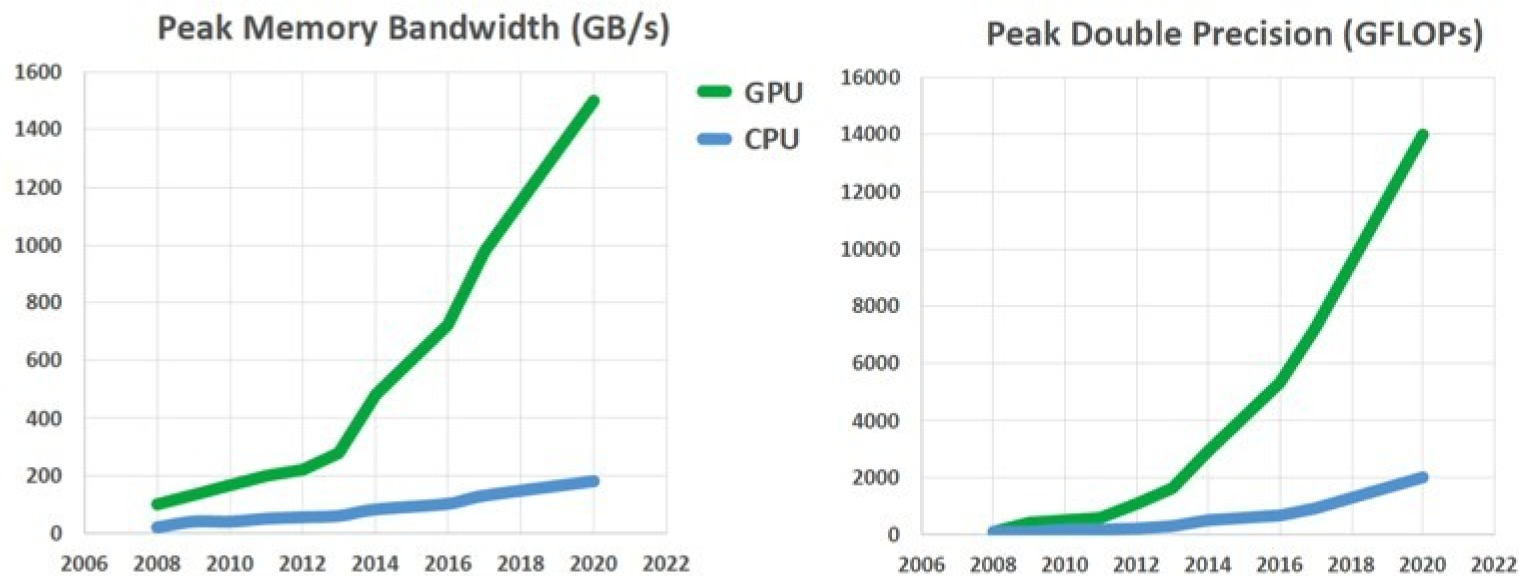
13. 07. 2022

CPU Development



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

GPU Development



https://www.nextplatform.com/2019/07/10/a-decade-of-accelerated-computing-augurs-well-for-gpus/

Acceleration Hardware

-

Formfaktor	H100 SXM	H100 PCIe
FP64	30 teraFLOPS	24 teraFLOPS
FP64-Tensor-Core	60 teraFLOPS	48 teraFLOPS
FP32	60 teraFLOPS	48 teraFLOPS
TF32-Tensor-Core	1.000 teraFLOPS*	800 teraFLOPS*
BFLOAT16-Tensor-Core	2.000 teraFLOPS*	1.600 teraFLOPS*
FP16-Tensor-Core	2.000 teraFLOPS*	1.600 teraFLOPS*
FP8-Tensor-Core	4.000 teraFLOPS*	3.200 teraFLOPS*
INT8-Tensor-Core	4.000 TOPS*	3.200 TOPS*
GPU-Speicher	80 GB	80 GB
GPU-Speicherbandbreite	3 TB/s	2 TB/s
Decoder	7 NVDEC 7 JPEG	7 NVDEC 7 JPEG
Max. Thermal Design Power (TDP)	700 W	350 W
Mehr-Instanzen-Grafikprozessoren	Bis zu 7 MIGs mit je 10 GB	
Formfaktor	SXM	PCIe Zwei Steckplätze mit Luftkühlung
Konnektivität	NVLink: 900 GB/s PCIe Gen5: 128 GB/s	NVLINK: 600 GB/s PCIe Gen5: 128 GB/s
Serveroptionen	NVIDIA HGX™ H100-Partner und NVIDIA-Certified Systems™ mit 4 oder 8 GPUs, NVIDIA DGX™ H100 mit 8 GPUs	Partner und NVIDIA-Certified Systems mit 1–8 GPUs

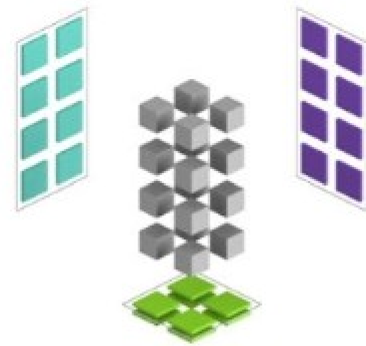
* Mit geringer Dichte dargestellt. Die Spezifikationen sind ohne geringe Dichte um die Hälfte niedriger. Vorläufige Spezifikationen. Änderungen vorbehalten.

GPU Features	NVIDIA Tesla P100	NVIDIA Tesla V100	NVIDIA A100
GPU Codename	GP100	GV100	GA100
GPU Architecture	NVIDIA Pascal	NVIDIA Volta	NVIDIA Ampere
GPU Board Form Factor	SXM	SXM2	SXM4
SMs	56	80	108
TPCs	28	40	54
FP32 Cores / SM	64	64	64
FP32 Cores / GPU	3584	5120	6912
FP64 Cores / SM (excl. Tensor)	32	32	32
FP64 Cores / GPU (excl. Tensor)	1792	2560	3456
INT32 Cores / SM	NA	64	64
INT32 Cores / GPU	NA	5120	6912
Tensor Cores / SM	NA	8	4[2]
Tensor Cores / GPU	NA	640	432
GPU Boost Clock	1480 MHz	1530 MHz	1410 MHz
Peak FP16 Tensor TFLOPS with FP16 Accumulate[1]	NA	125	312/~~624~~[3]
Peak FP16 Tensor TFLOPS with FP32 Accumulate[1]	NA	125	312/~~624~~[3]
Peak BF16 Tensor TFLOPS with FP32 Accumulate[1]	NA	NA	312/~~624~~[3]
Peak TF32 Tensor TFLOPS[1]	NA	NA	156/~~312~~[3]
Peak FP64 Tensor TFLOPS[1]	NA	NA	19.5
Peak INT8 Tensor TOPS[1]	NA	NA	624/1248[3]
Peak INT4 Tensor TOPS[1]	NA	NA	1248/2496[3]
Peak FP16 TFLOPS[1] (non-Tensor)	21.2	31.4	78
Peak BF16 TFLOPS[1] (non-Tensor)	NA	NA	39
Peak FP32 TFLOPS[1] (non-Tensor)	10.6	15.7	19.5
Peak FP64 TFLOPS[1] (non-Tensor)	5.3	7.8	9.7
Peak INT32 TOPS[1,4]	NA	15.7	19.5
Texture Units	224	320	432
Memory Interface	4096-bit HBM2	4096-bit HBM2	5120-bit HBM2
Memory Size	16 GB	32 GB / 16 GB	40 GB
Memory Data Rate	703 MHz DDR	877.5 MHz DDR	1215 MHz DDR

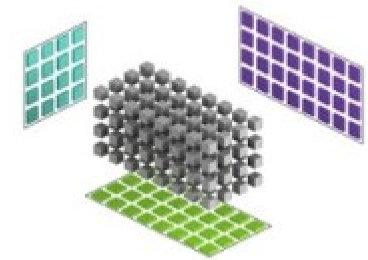
Acceleration Hardware

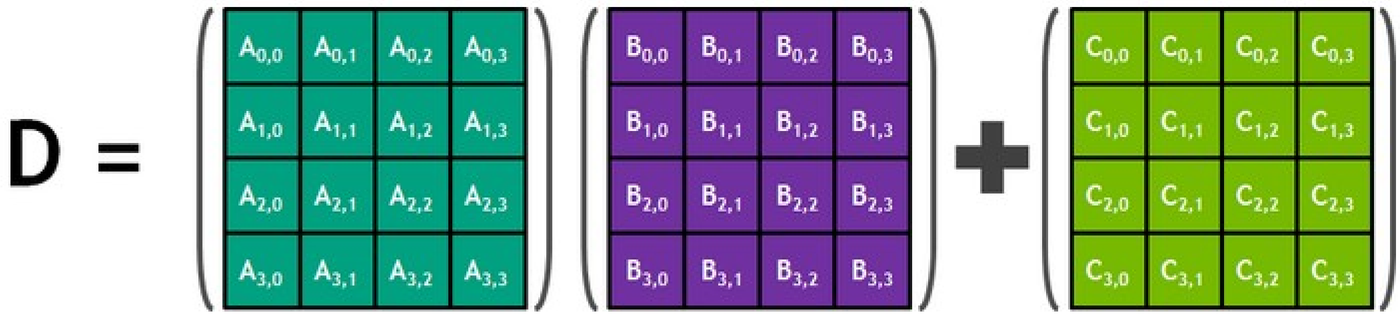
- Tensor Cores A100

FP64

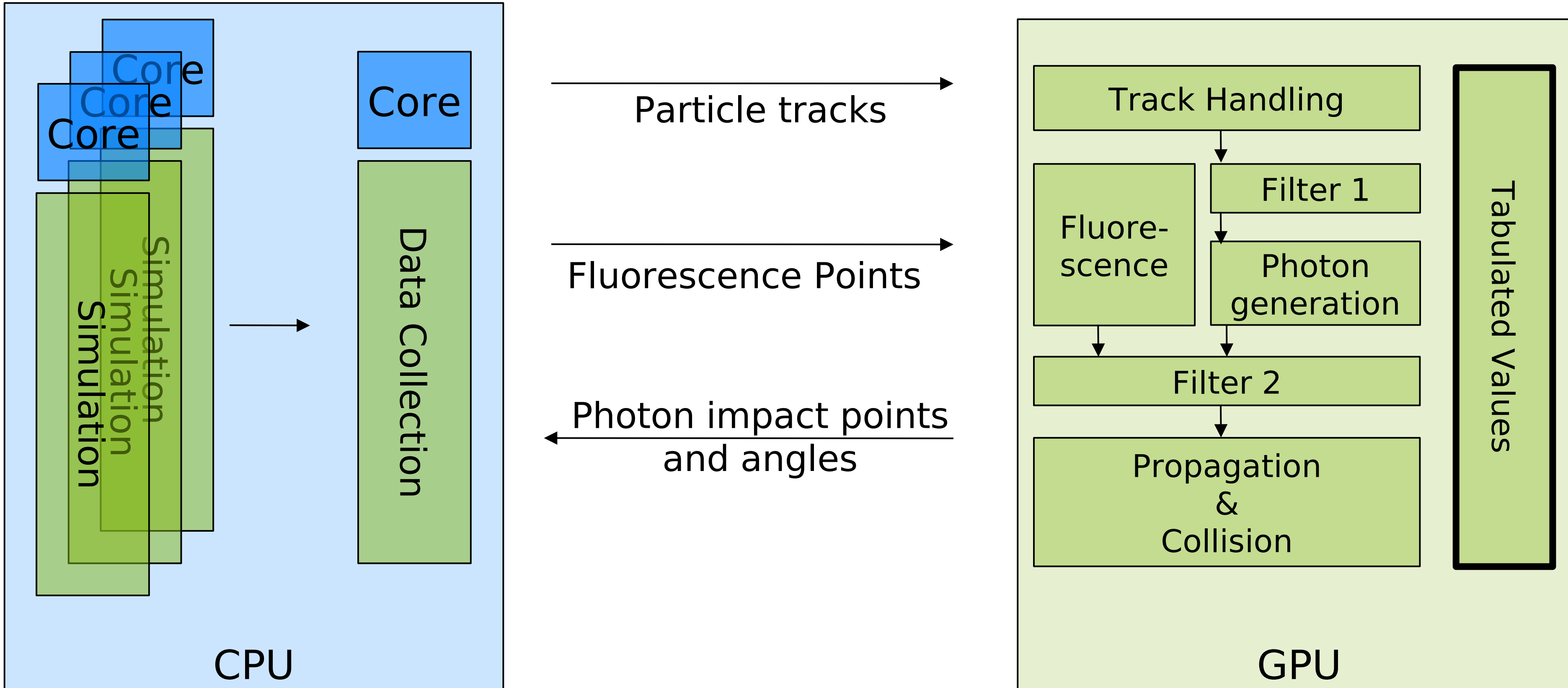


FP32



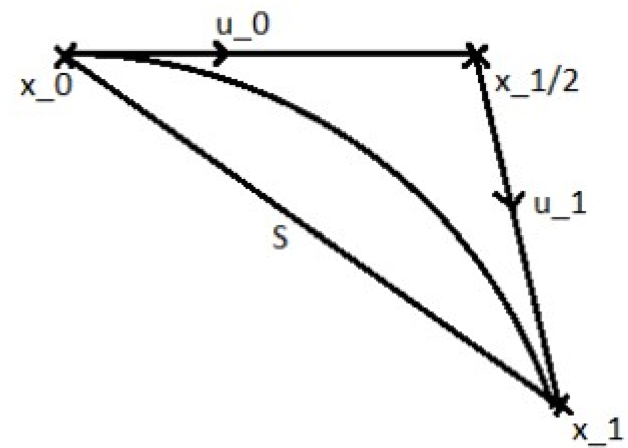


Computing Structure



Prerequisite

- Linear particle tracks
- >>1024 track segments accumulated to reach good performance



Andre Schmidt - KIT

Track Handling

- Photon count:

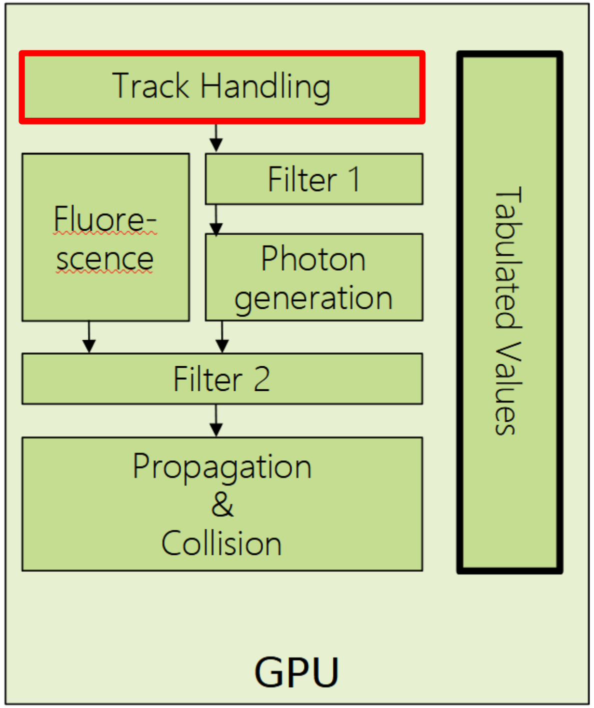
  float64 vs float32

  $$\frac{1}{\beta\ n} \sim \#\ \text{Photon}$$

  – β, n very close to 1

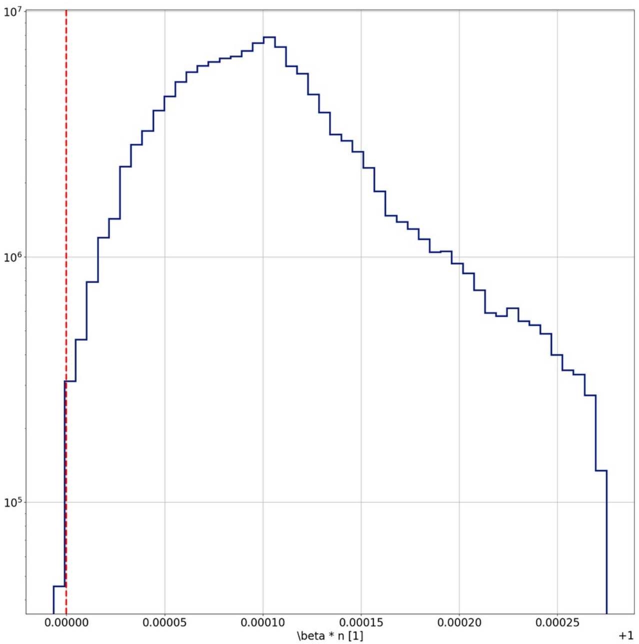
  → 0.03% difference in photon generation

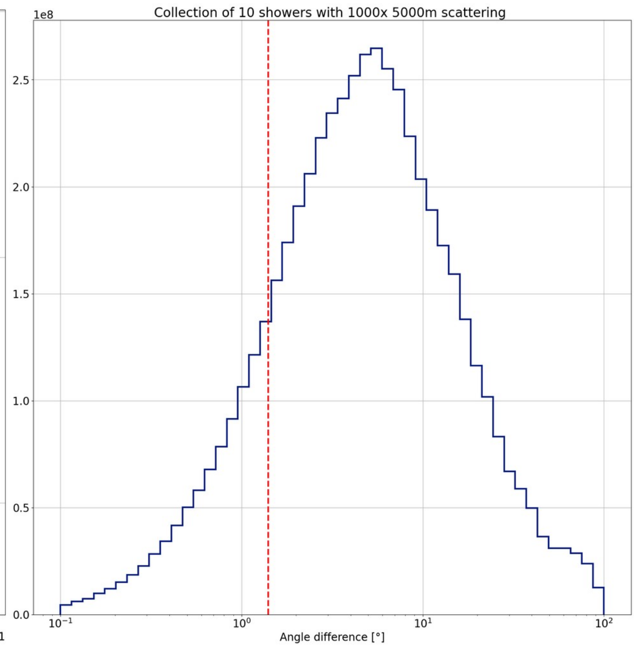
- Mostly multiplications and memory operations

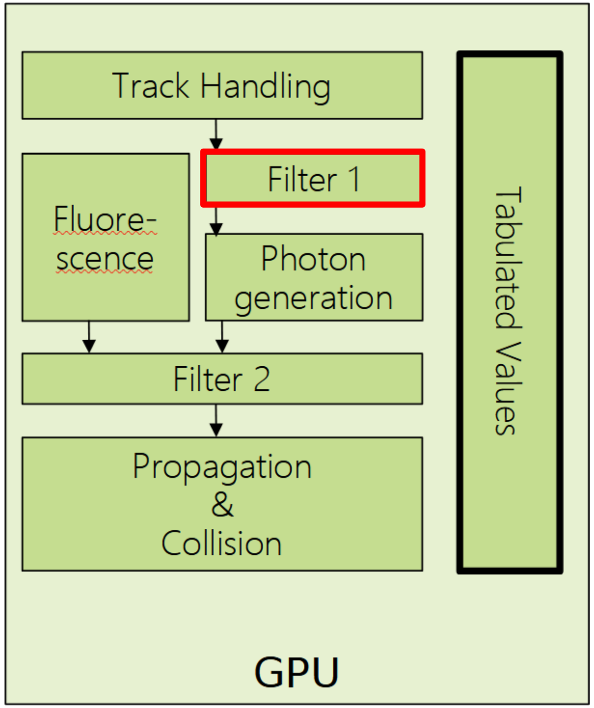


Filter 1

- Angular cuts
- Dot product → Tensor possible

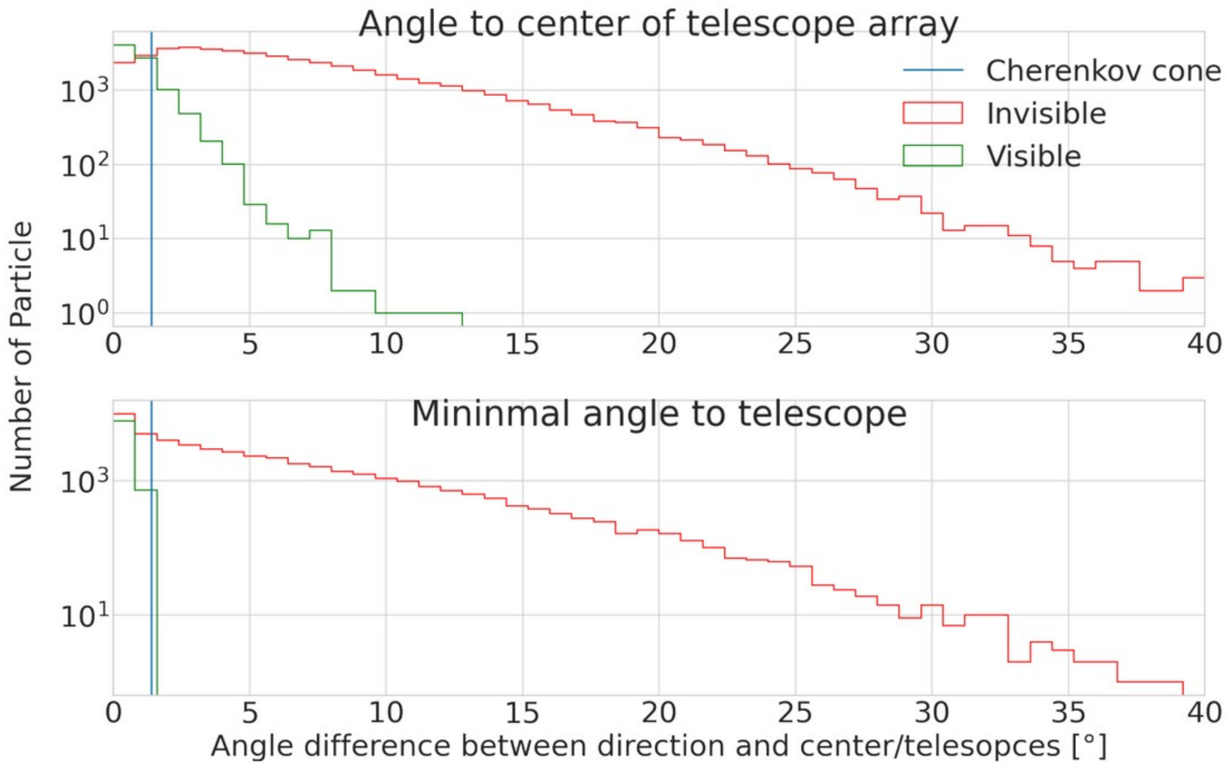


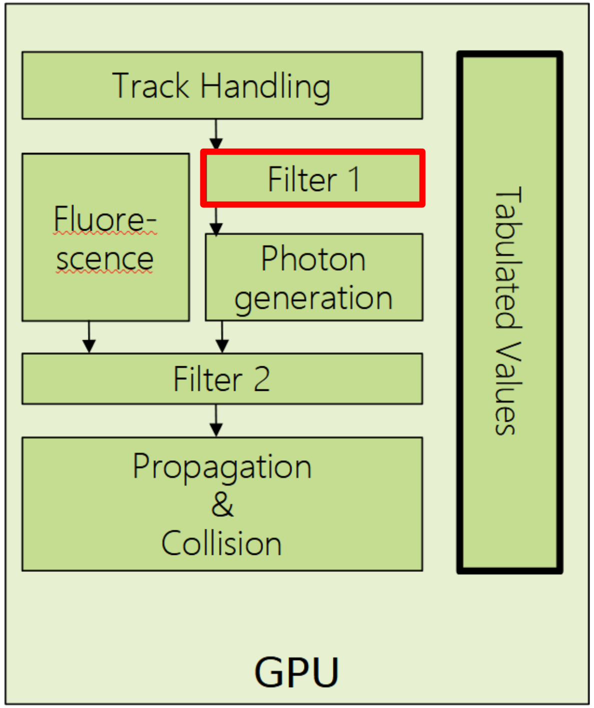




Filter 1

- Angular cuts
- Dot product → Tensor possible





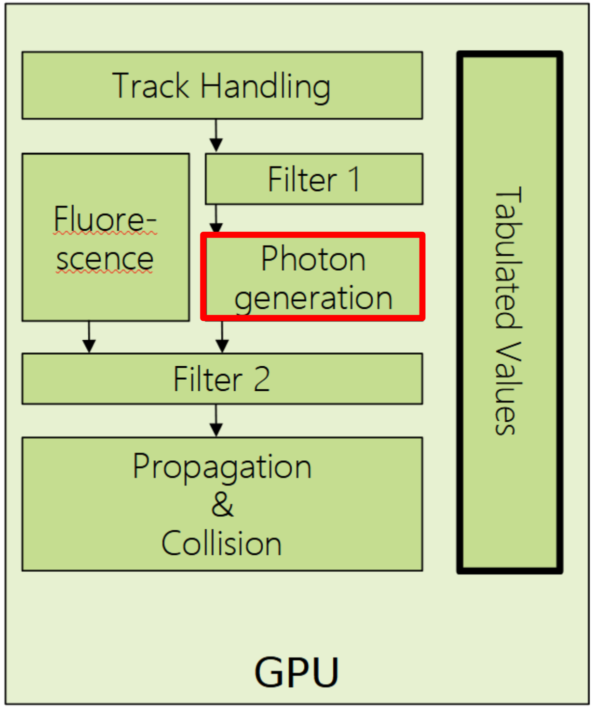
y

Photon Generation

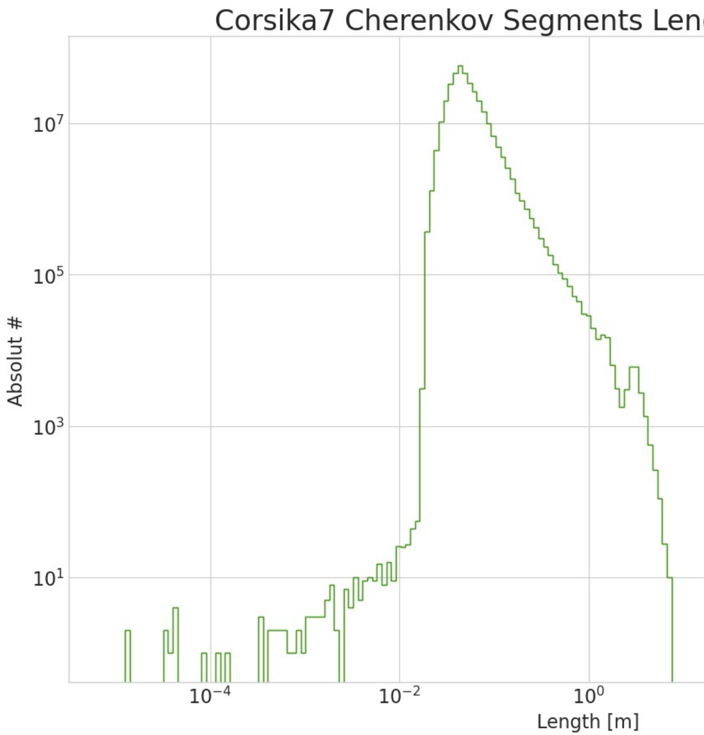
- Generate photons directions → vector
- Rotate and translate into particle system
  Matrix multiplication

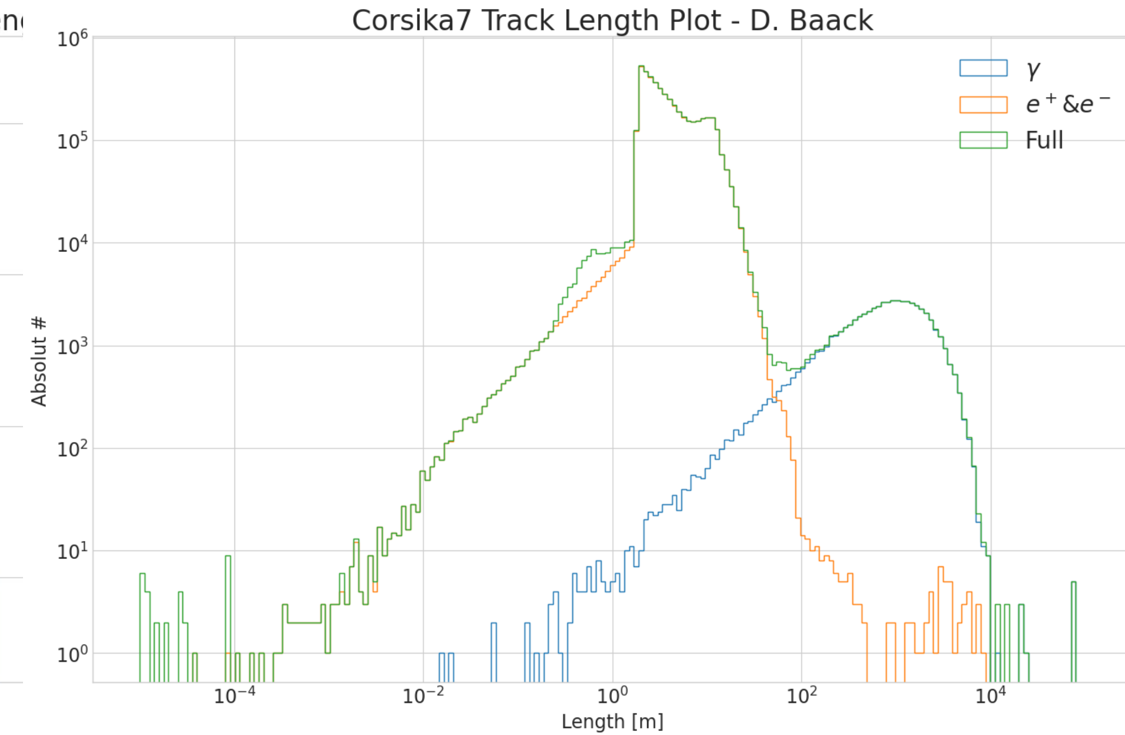
$$N_{\text{cher}} = 2\pi\alpha z^2 L \frac{\lambda_1 - \lambda_2}{\lambda_1 \cdot \lambda_2} \cdot (1 - (n\beta)^{-2})$$

- Rotate and Translate to Emission position



Photon Generation

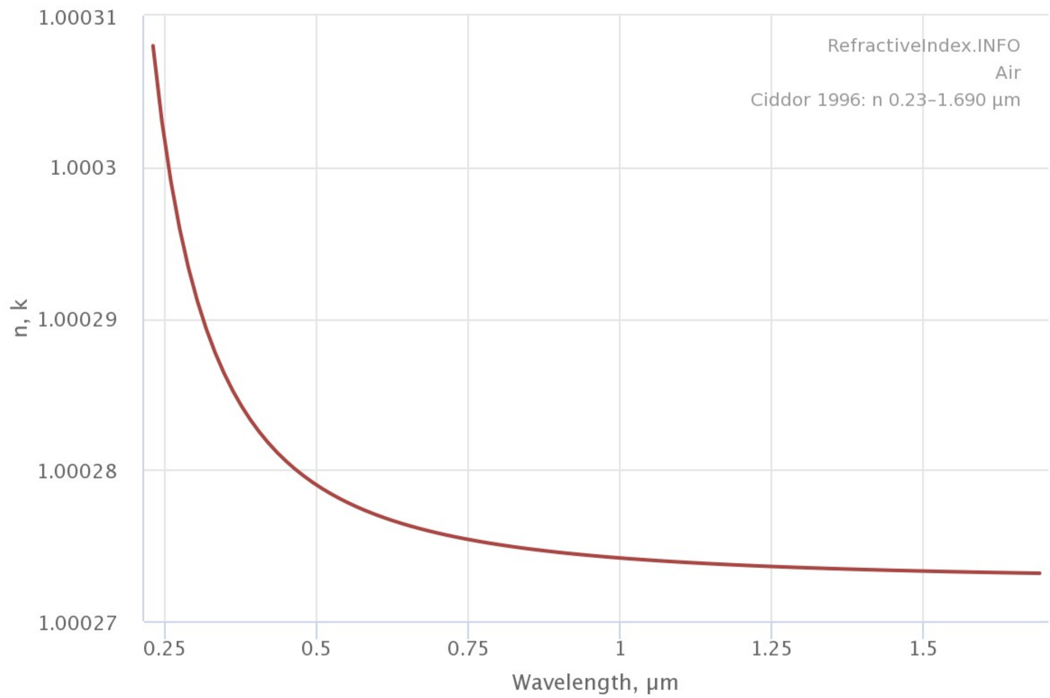




Chromatic Dispersion

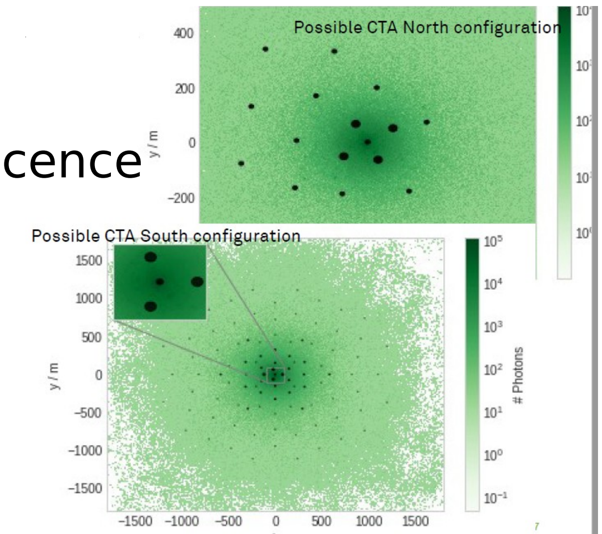
- Emission Wavelength dependent

- Edge cases:
  - Emission only part of specified wavelength possible

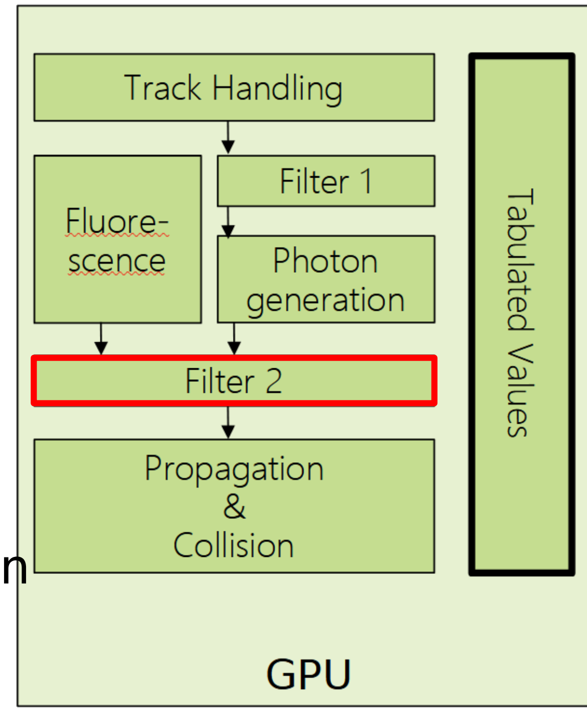


Filter 2

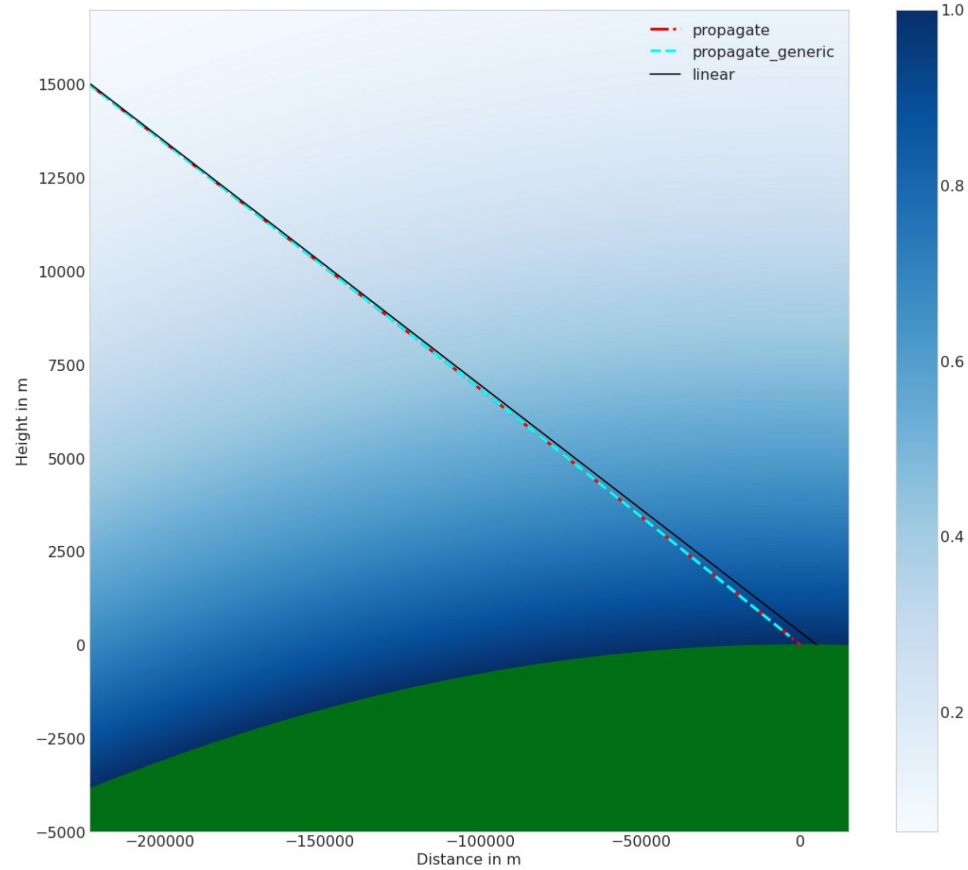
- Merged with fluorescence generation



- Not feasible for default Cherenkov propagation
  - Cuts cost more than simple propagation



Propagation

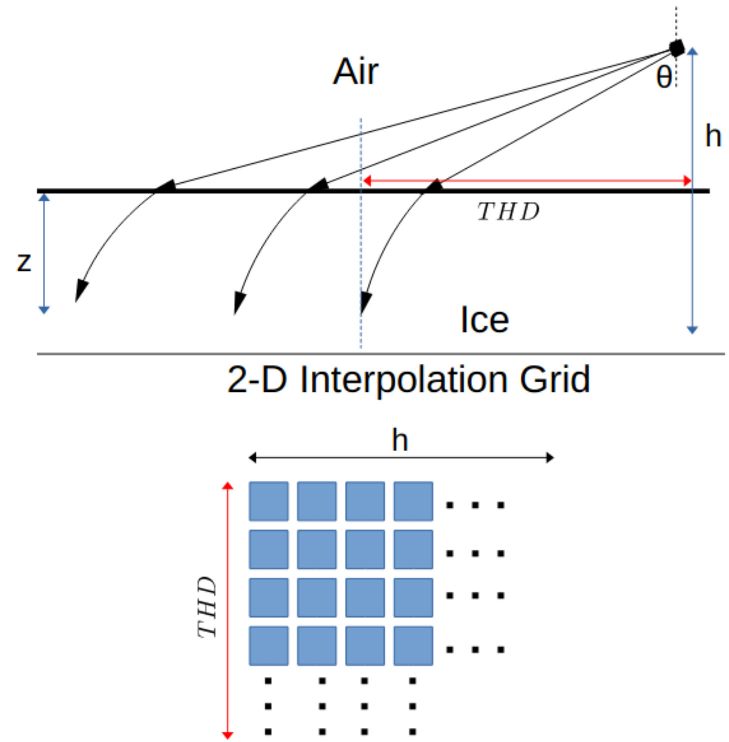


Propagation

of:

- Yesterday:

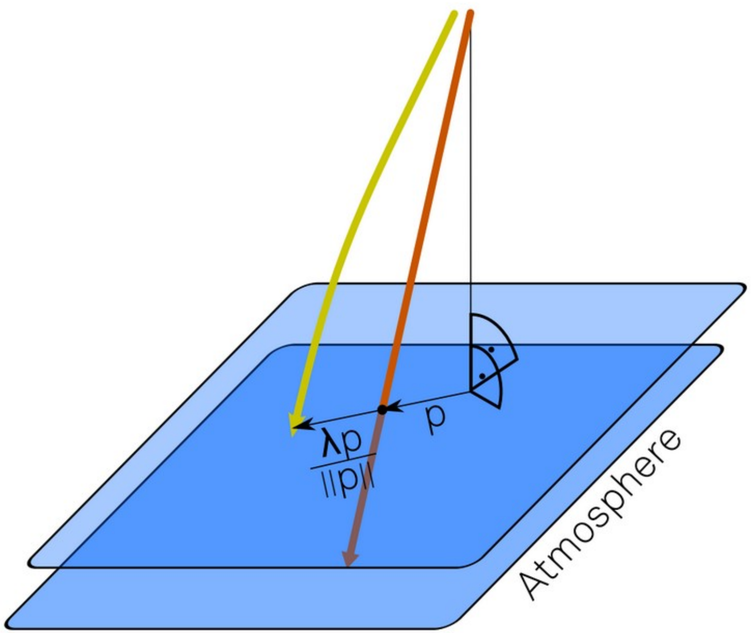
  Propagating Air Shower Radio Signals to In-ice Antennas

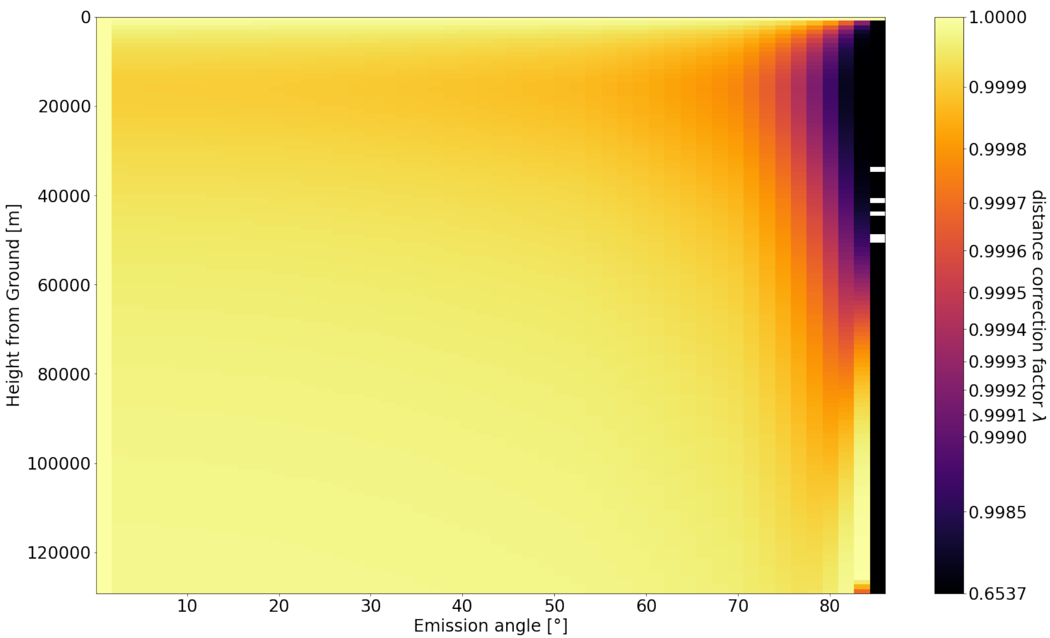


Uzair A. Latif et al 7

Propagation

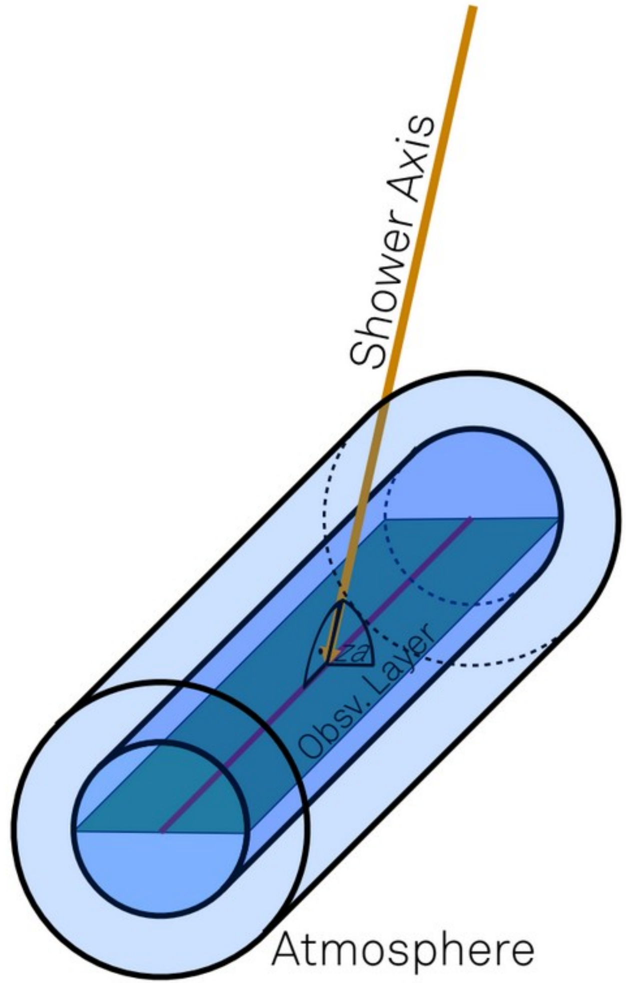
- 2 Parameter: Height and Angle





Propagation

- Offset/Distance, Height, Angle
- Still possible to interpolate in hardware



Conclusion

- Cross-Check with Corsika 6.9, 7.4 and real Magic data started.
- Possible cooperation with Mathieu and LST possible