

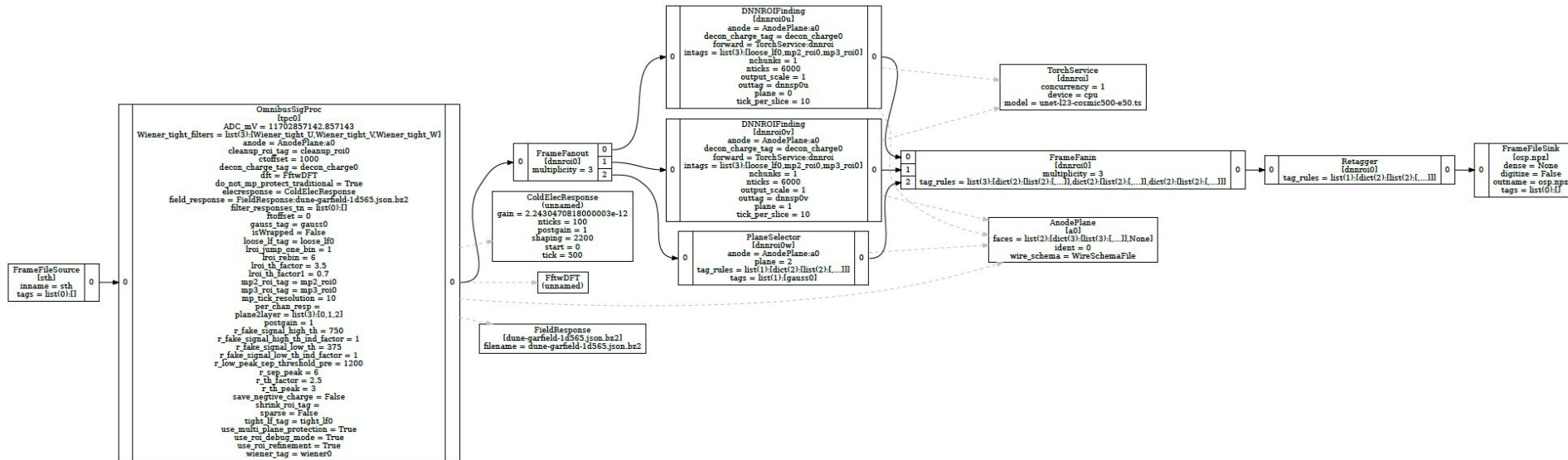
Performance Comparison between OSP and SPNG (DUNE)

Context

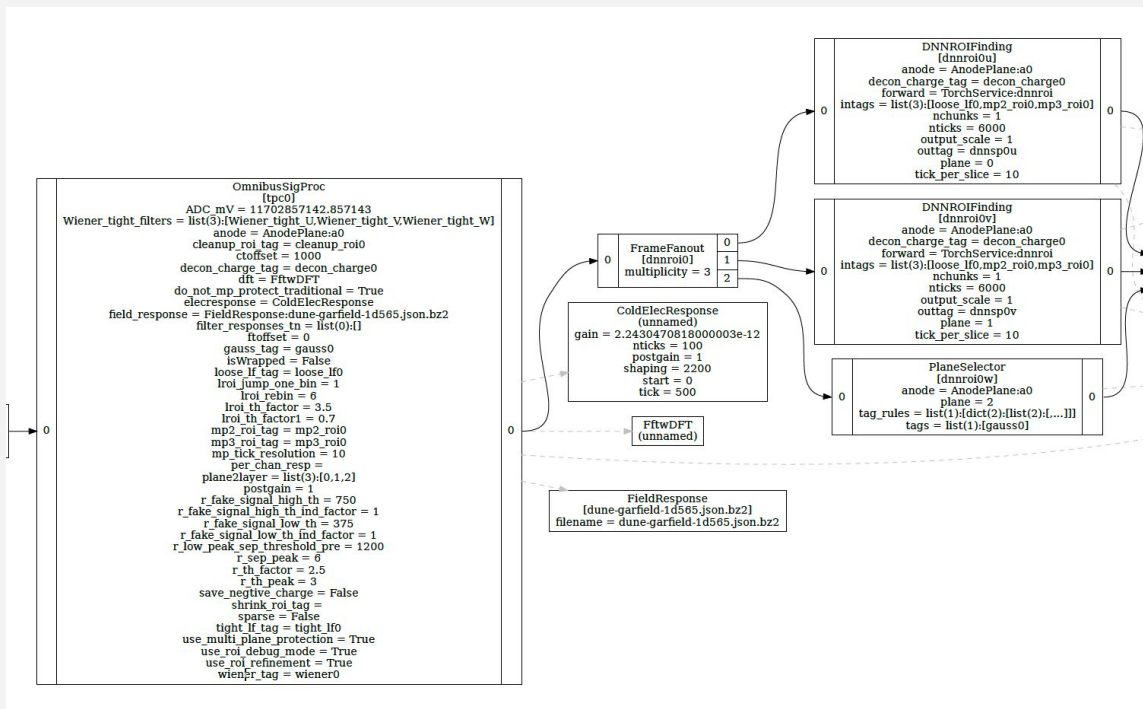
- This comparison of OSP and SPNG is done after the previous issue of slow down of SPNG nodes in GPU compared to CPU was addressed.
- A thorough analysis of the issue and the subsequent mitigation/solution is discussed in this github [issue](#)

OSP : Original Signal Processing

- Current State of the art signal processing used by wire-cell → Original Signal Processing (OSP)



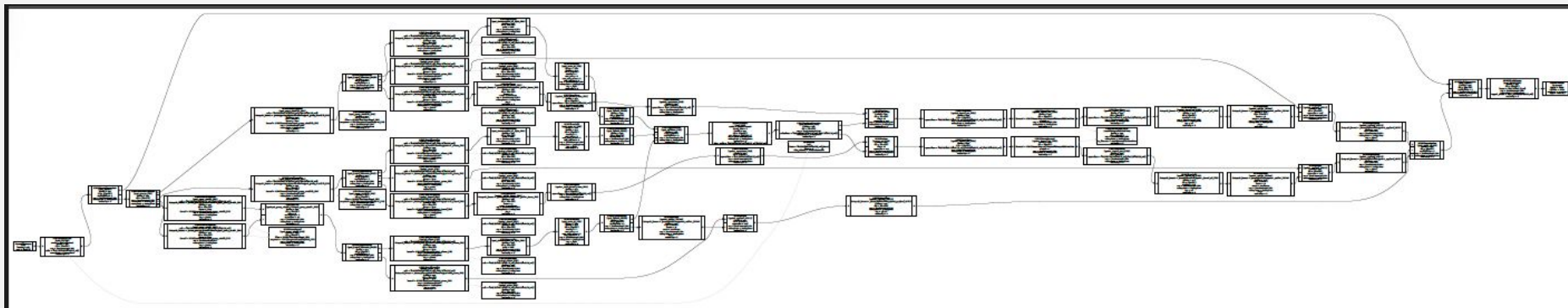
Original Signal Processing



CPU Heavy

Inference done via GPU in
DNNROIFinding Node.

SPNG



Profiling SPNG and OSP

- 10 “events”
- SPNG: `adc-to-spng.jsonnet`
- OSP: `adc-to-osp.jsonnet`
- Hostname: `wcgpu0.phy.bnl.gov`
- CPU: AMD Ryzen Threadripper 7970X 32 cores
- Architecture: `x86_64`
- GPU: NVIDIA GeForce RTX 4090
- NVIDIA Driver: `555.42.06`
- CUDA Driver: `12050`
- Profiling was done end to end for both SPNG and OSP workflows

Metric	ADC-to-SPNG (wall)	ADC-to-OSP (wall)
Full Profile Span	13.410 sec	46.332 sec
CUDA Kernels launched	326,703	3194
CUDA kernel Time	2975.235 ms	286.106 ms
CUDA runtime/API calls	386824	6055
CUDA runtime/API time	3774.143 ms	350.662 ms
Synchronization events	1127	504
Synchronization time	2506.186 ms	164.713 ms
Memory copy operations	56572	578

Breaking down the Malloc operations

Copy Kind	SPNG Count	SPNG Bytes (MB)	SPNG Time (ms)	OSP Count	OSP Bytes (MB)	OSP Time (ms)	Time Ratio (SPNG/OSP)
Device-to-Device	55,425	30,906.804	92.650	74	4,300.800	7.519	12.32×
Host-to-Device	351	806.821	43.082	150	168.813	6.637	6.49×
Device-to-Host	796	768.106	43.257	354	38.403	1.776	24.36×

OSP and SPNG Running in CPU and GPU

SP type	CPU	GPU	Speed up
SPNG	92.8 sec	10.9 sec	8.5x
OSP	51.47	46.58	1.10x

OSP and SPNG Running in CPU and GPU (Using Single CPU Core)

SP type	CPU (sec)	GPU (sec)	Speed up
SPNG	313	11	28.4x
OSP	136	50	2.7x

Setting `OMP_NUM_THREAD=1` so that torch is forced to use single CPU core

Breakdown (Single OMP Threads)

