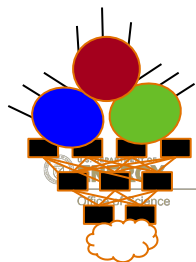




GELATO Update 2/23/24

Liam Sherman



Model Synthesis

- Model synthesized at fixed<16,10> precision
- 25ns clock with part 'xcvu9pflgb2104-2L' from topo board

Performance Estimates

Timing

Summary

Clock	Target	Estimated	Uncertainty
ap_clk	25.00 ns	19.613 ns	3.12 ns

Latency

Summary

Latency (cycles)		Latency (absolute)		Interval (cycles)		
min	max	min	max	min	max	Type
0	0	0 ns	0 ns	1	1	function

Detail

Instance

Loop

Summary

Name	BRAM_18K	DSP48E	FF	LUT	URAM
DSP	-	-	-	-	-
Expression	-	-	0	2	-
FIFO	-	-	-	-	-
Instance	-	504	0	78663	-
Memory	-	-	-	-	-
Multiplexer	-	-	-	36	-
Register	-	-	914	-	-
Total	0	504	914	78701	0
Available	4320	6840	2364480	1182240	960
Available SLR	1440	2280	788160	394080	320
Utilization (%)	0	7	~0	6	0
Utilization SLR (%)	0	22	~0	19	0

Model Effectiveness

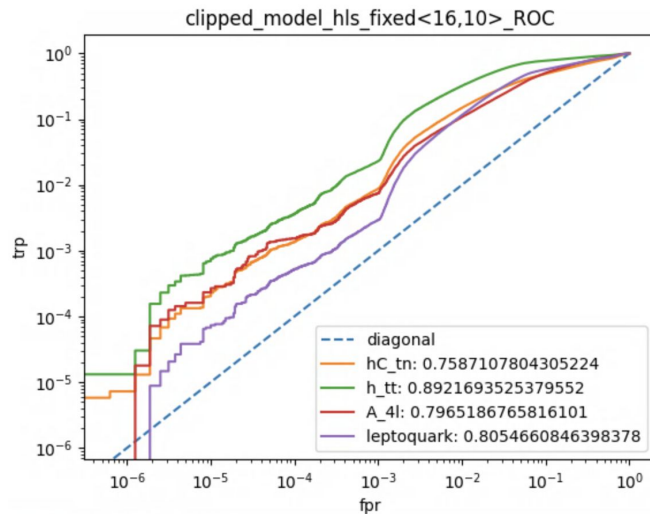
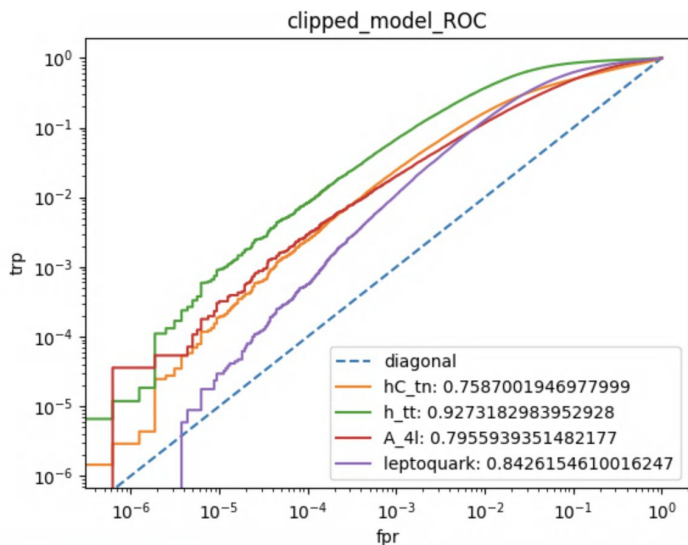
SLAC

ap_fixed<width bits, integer bits>

0101.1011101010



- The keras model has similar effectiveness on all signals to the hls4ml model



Signal Efficiency Before and After Synthesis

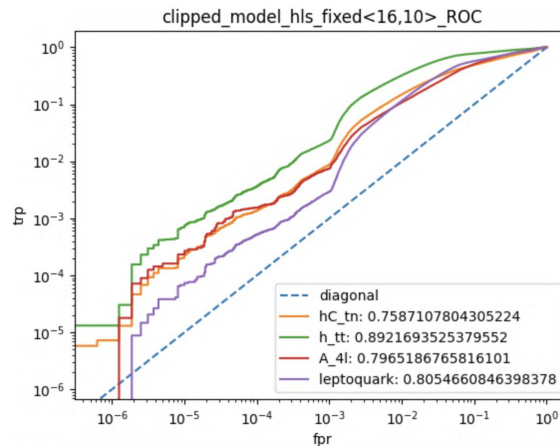
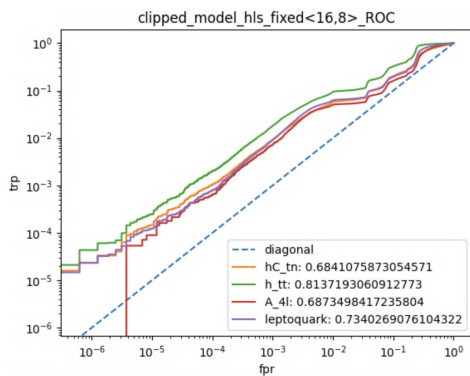
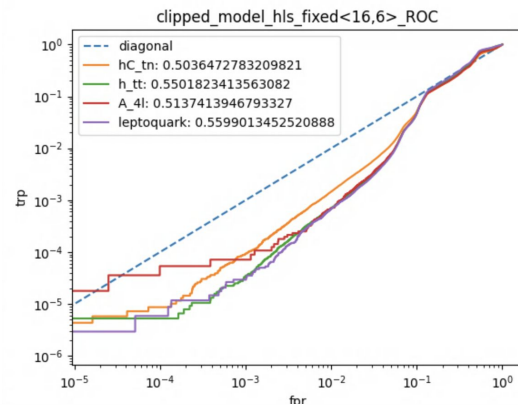
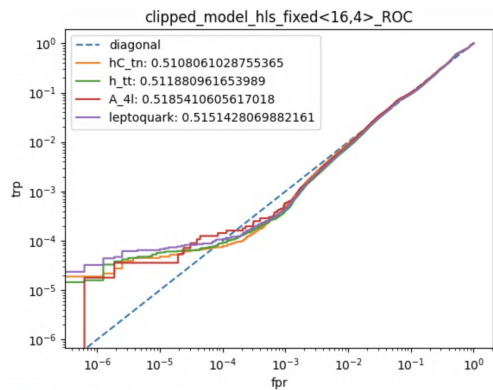
SLAC

Signal Acceptance at 100KHz

Model	hC_tn	h_tt	A_4l	leptoquark
Keras	0.05509	0.14201	0.041272	0.030739
Hls4ml <16,10>	0.046826	0.1200833	0.0351623	0.02542

Strong Dependence on distribution of int vs decimal

SLAC

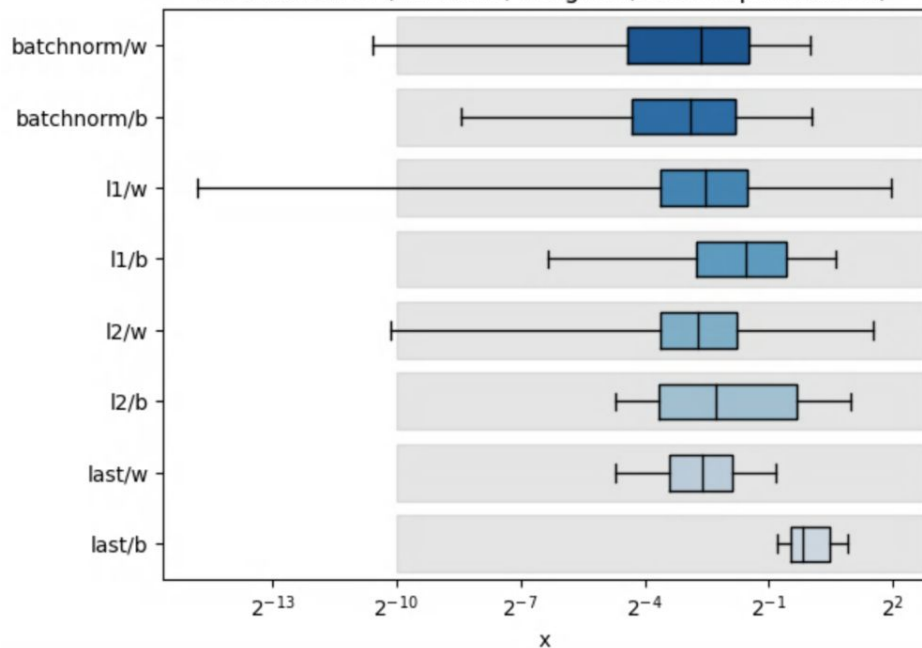


Weight value distribution

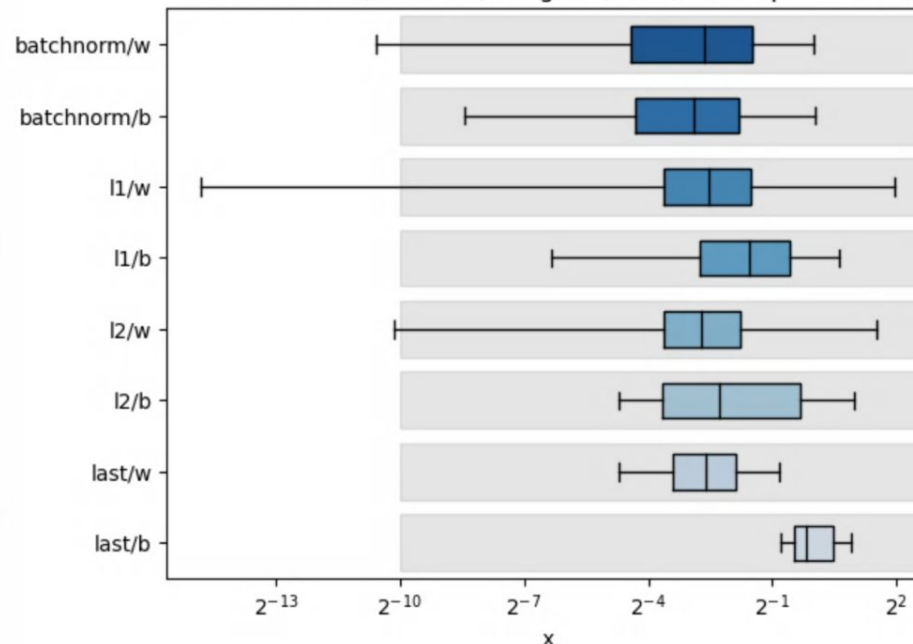
SLAC

<16,6>

Distribution of (non-zero) weights (before optimization)



Distribution of (non-zero) weights (final / after optimization)

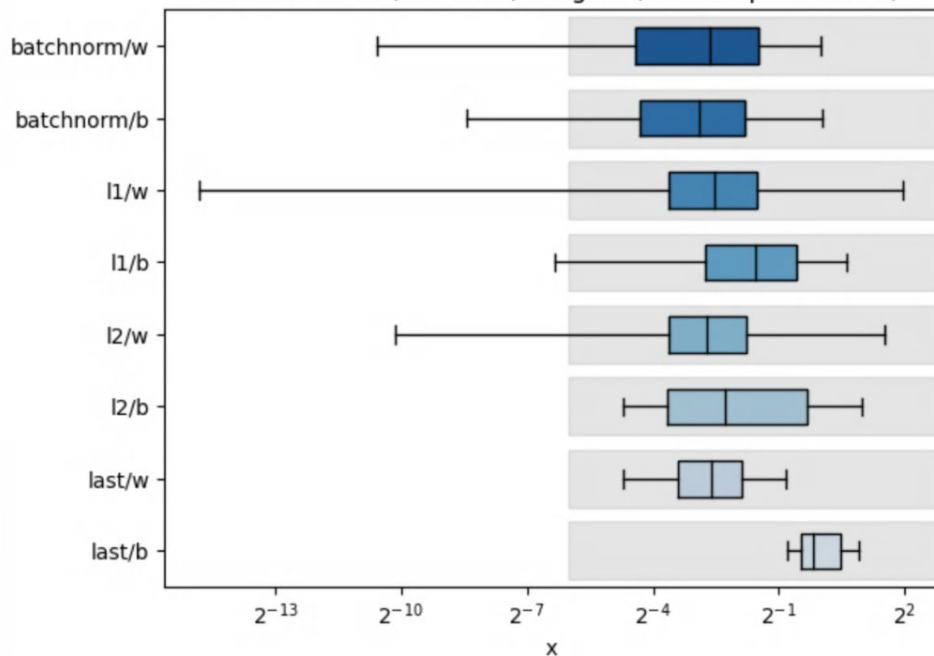


Weight value distribution

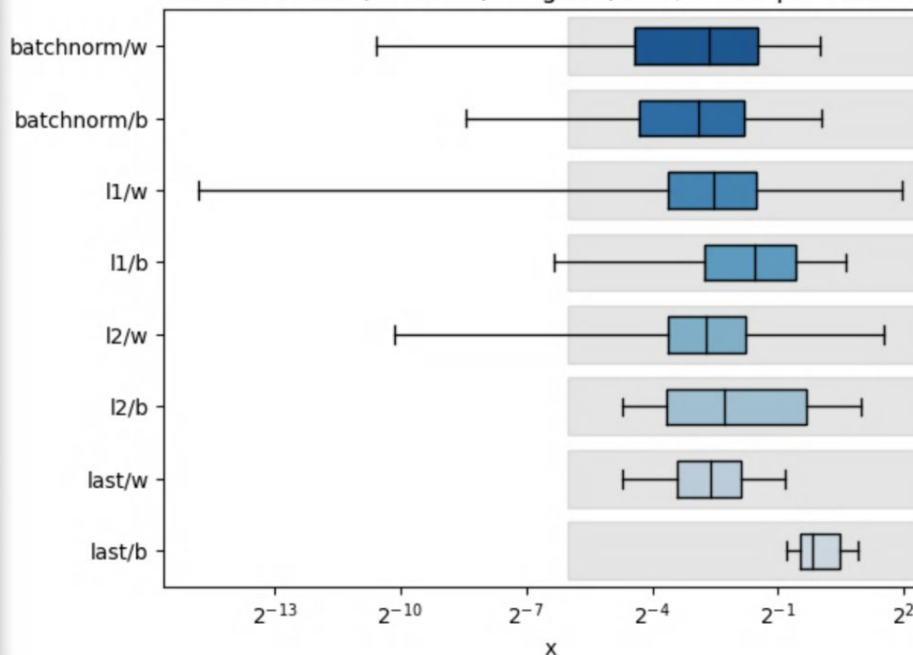
SLAC

<16,10>

Distribution of (non-zero) weights (before optimization)

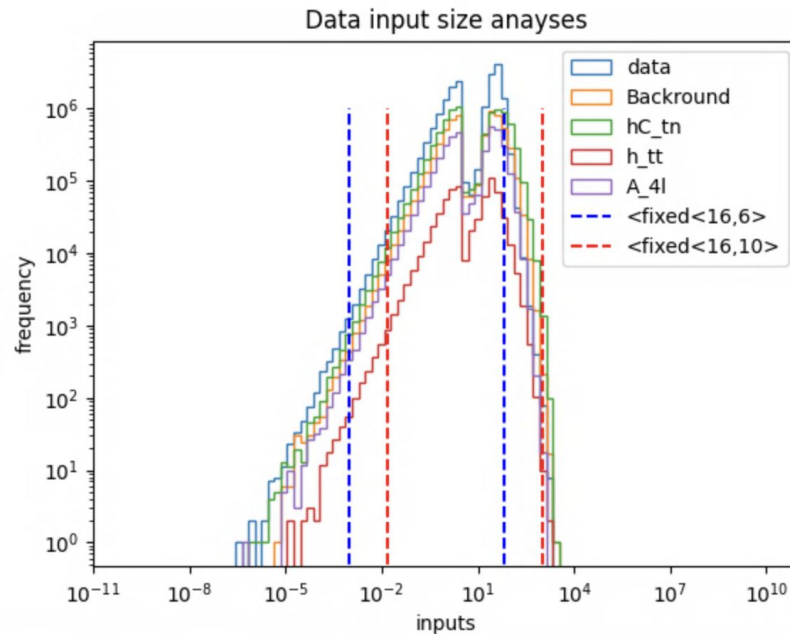


Distribution of (non-zero) weights (final / after optimization)



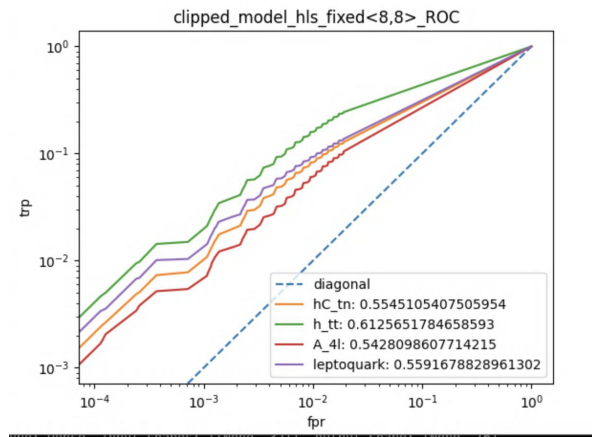
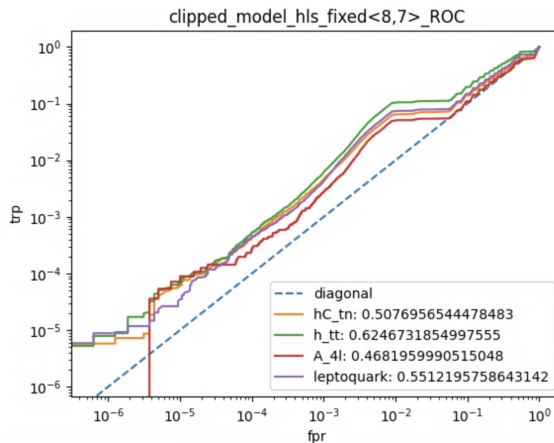
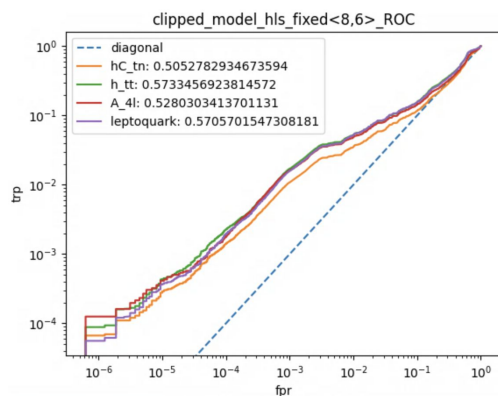
Input Value Distribution

- I think this is the cause of the performance drop between $\langle 16, 10 \rangle$ and $\langle 16, 6 \rangle$ precision



8 bit performance

SLAC



- Quantization Aware Training likely needed to be effective at 8 bit precision