Single Event Effects in BIB Rejection NNs

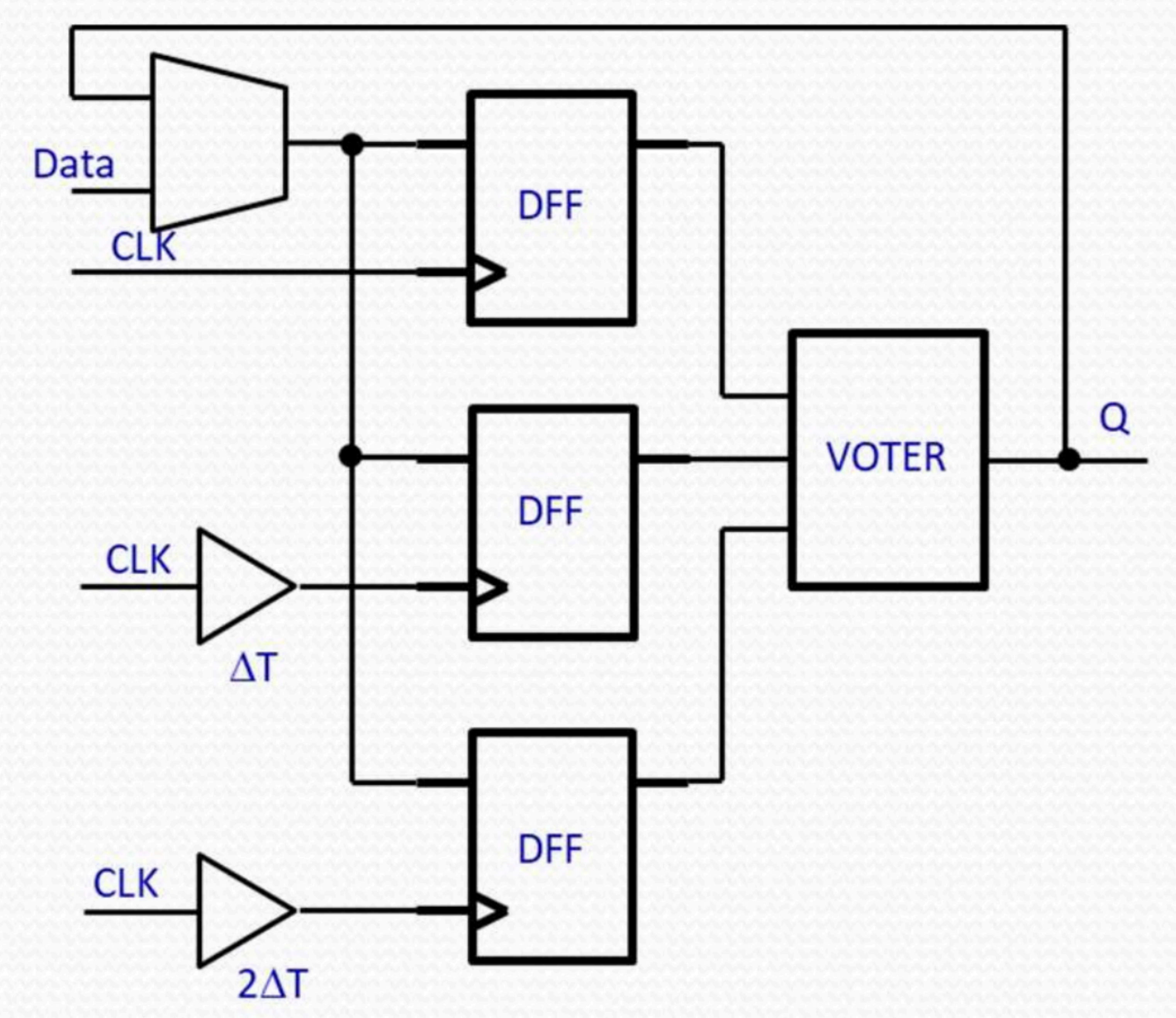
Daniel Abadjiev, Karri DiPetrillo, Eliza Howard, Ryan Michaud, Ryan Roberts, Eric You

Constraints of Deploying ML on Detector

- Forced to use ASICs by constraints of power, material budget, latency, radiation hardness, etc.
- Data altered during readout — no reprocessing to correct issues
  - Choices made in ASIC design are final for ~10 years of detector operation
  - Choices made in model training are final until the weights can be replaced
  - Robustness of chip design and ML architecture is vital
- **How well does a neural network operate in a radiation environment?**
  - Is it better to triplicate or use a larger network?

Single Event Effects in Pixel Detectors

- Single Event Effects (SEEs) include several categories
  - **Single Event Upset (SEU) - Random bit flip in stored memory**
  - Single Event Transient (SET) - Transient signal induced in circuit
  - Single Event Latch-Up (SEL) - Transistor stuck open with high current draw, permanent damage
- Standard mitigation is triple modular redundancy (TMR). 200% overhead.
  - Phase II Pixel detectors only have some logic triplicated due to space constraints.
  - Design target 99% hit efficiency to have no impact on tracking performance.



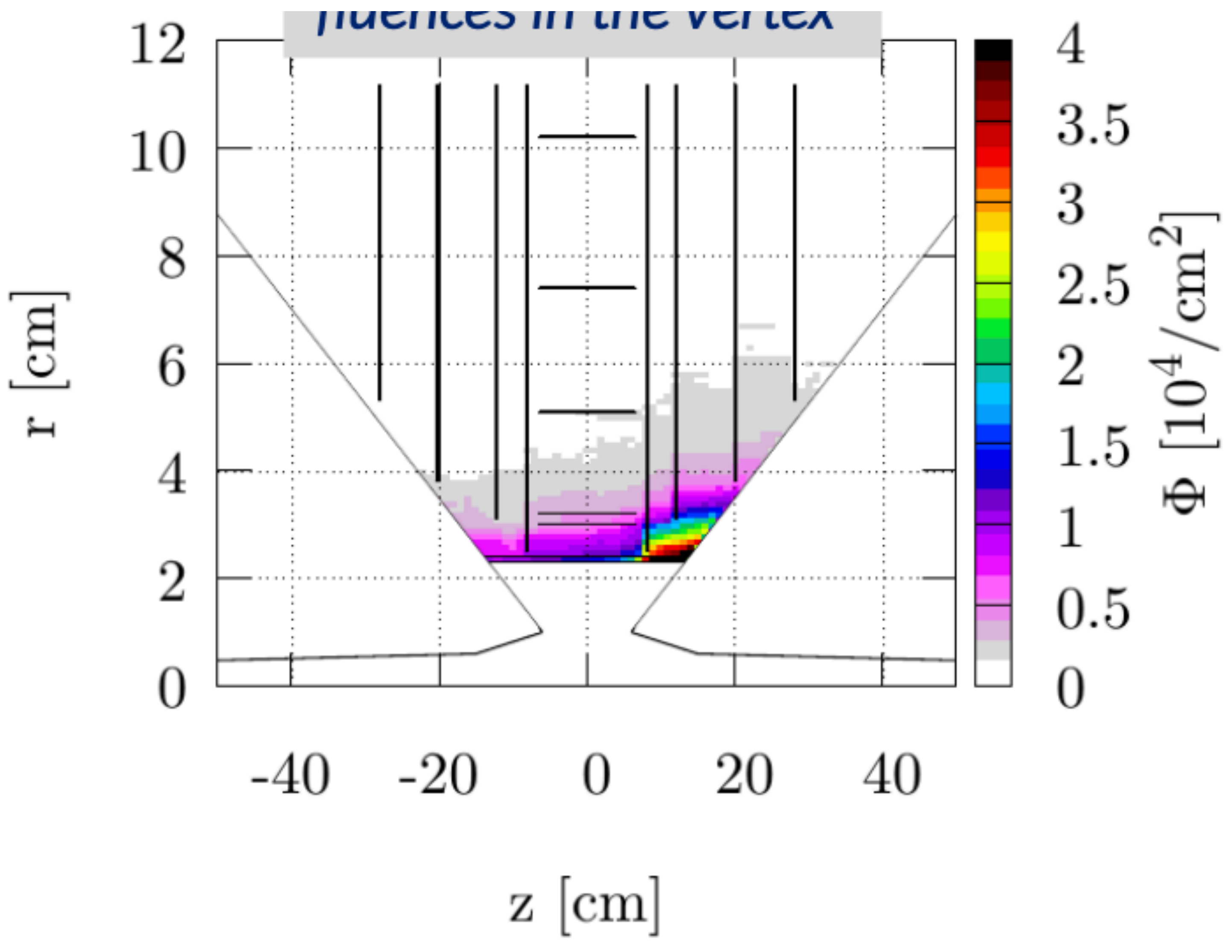
Mohsine Menouni

Estimating SEE Impact on in-ASIC NNs

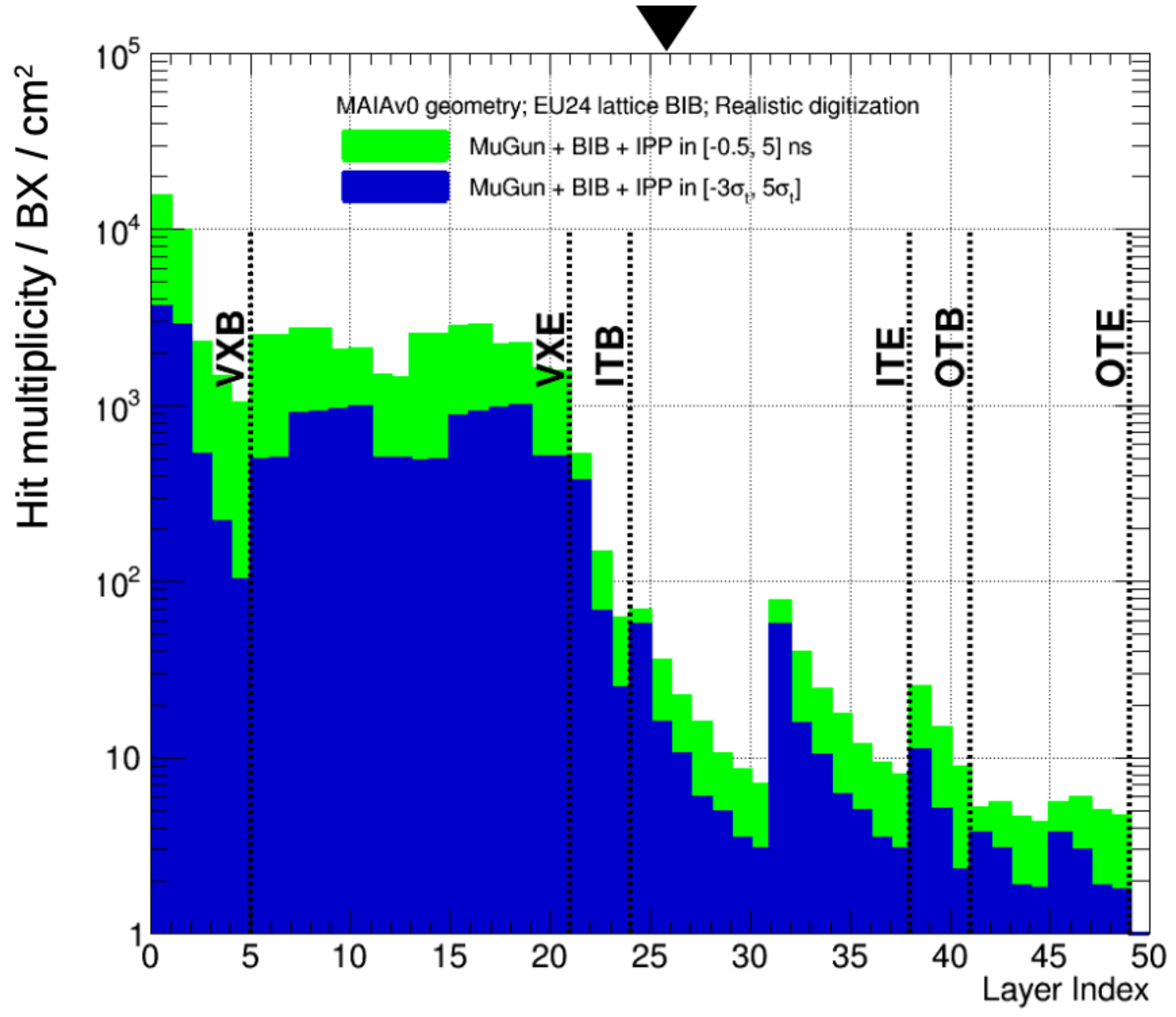
- Estimate impact of **SEUs in the stored weights** of the NN.
  - Expect SEUs to accumulate in unprotected bits until configuration is reset.
  - SEUs in other areas affect a single BX.
  - Rarer effects require simulation of chip design to study.
- **Inject bit flips** into the quantized weights and **re-run inference** to determine the impact.
  - Focus on **changes to signal efficiency and background rejection** for a fixed cut chosen based on radiation-free performance.

BIB Rejection Reminders

- 1st pixel layer sees occupancy > 10%
- Full triggerless readout at a collision rate 20-30 kHz → ~73.6 Gbps per chip
    - Must either **reject hits from BIB** in the readout chip **or operate with a trigger**
    - BIB particles differ from signal in direction, timing, PID, and momentum
- Simple cut on cluster length rejects 20-30% of clusters: 50% reduction in bandwidth (Angira Rastogi)
    - Signal losses ≲ 5%, data rate remains 35x too large with loose timing cut



Flux per BC from in-flight decays to electrons: Daniele Calzolari



Hit multiplicity by tracker layer: Angira Rastogi

NN Performance

- **Simple Model — Cluster Position and Size** 64% background rejection @ 99% signal efficiency

- **Intermediate — Cluster Shape** 81% background rejection @ 99% signal efficiency

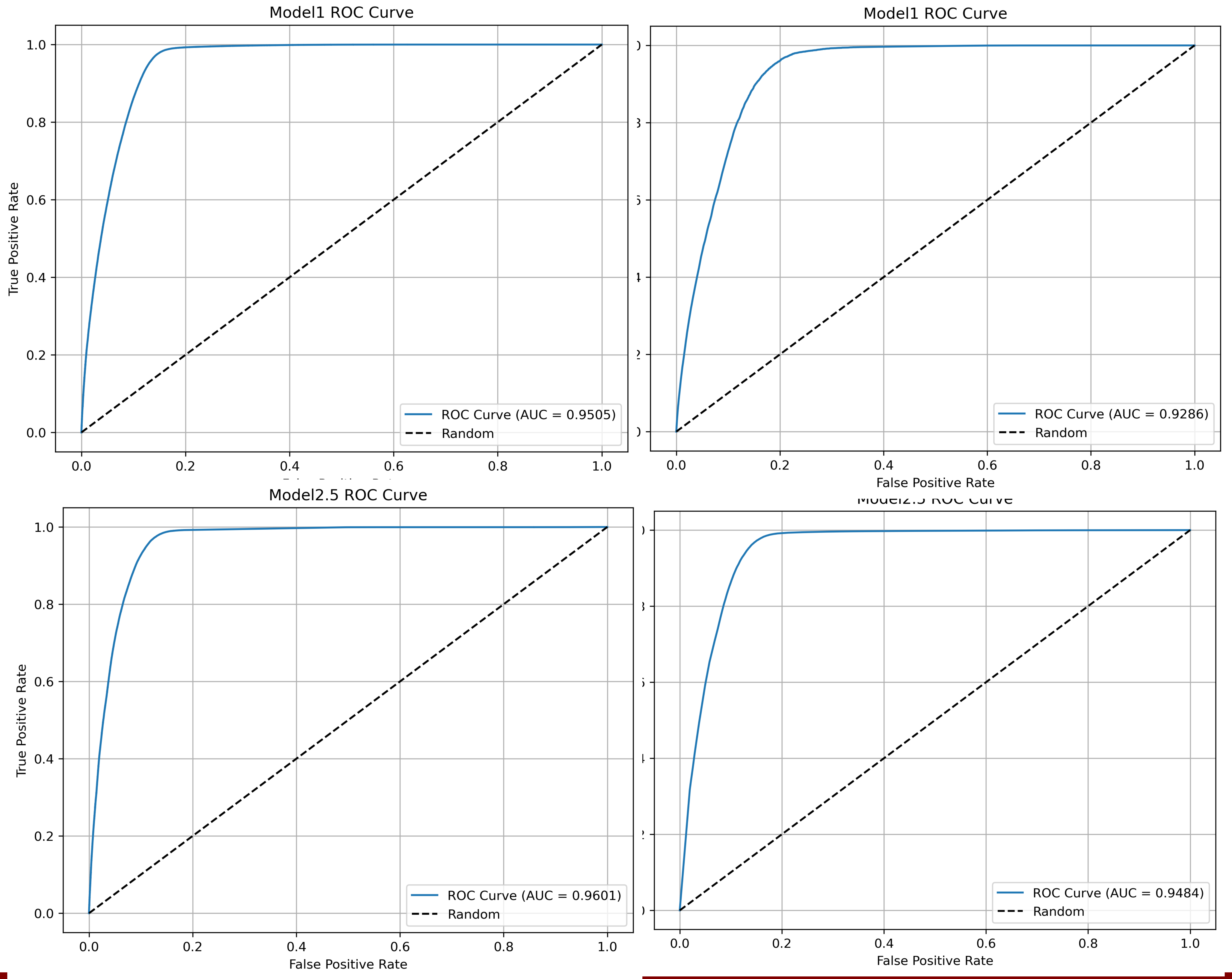
- N.B. Still missing input quantization for this study.

Unquantized

4 bit Quantized

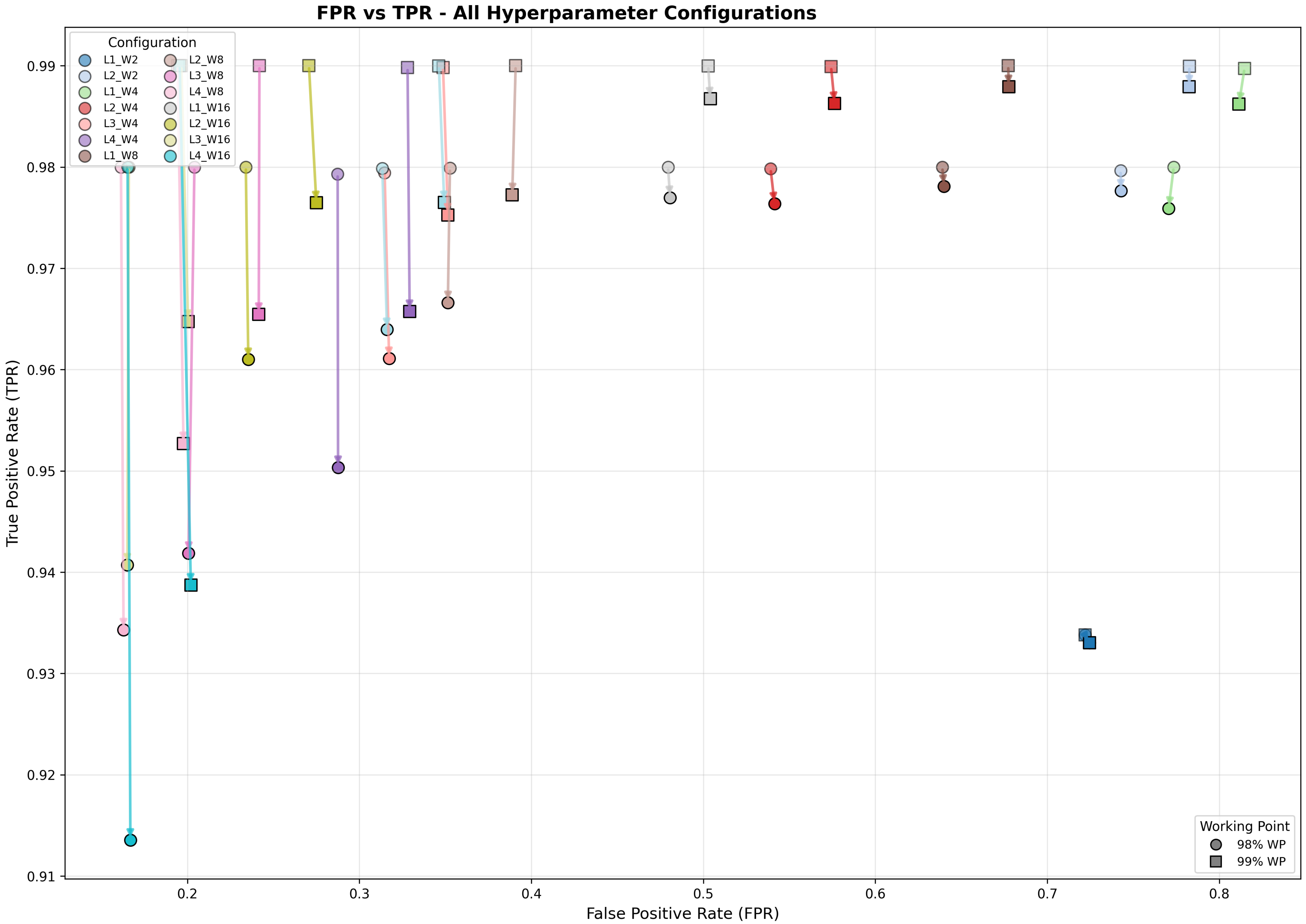
Simple Model

Intermediate Model



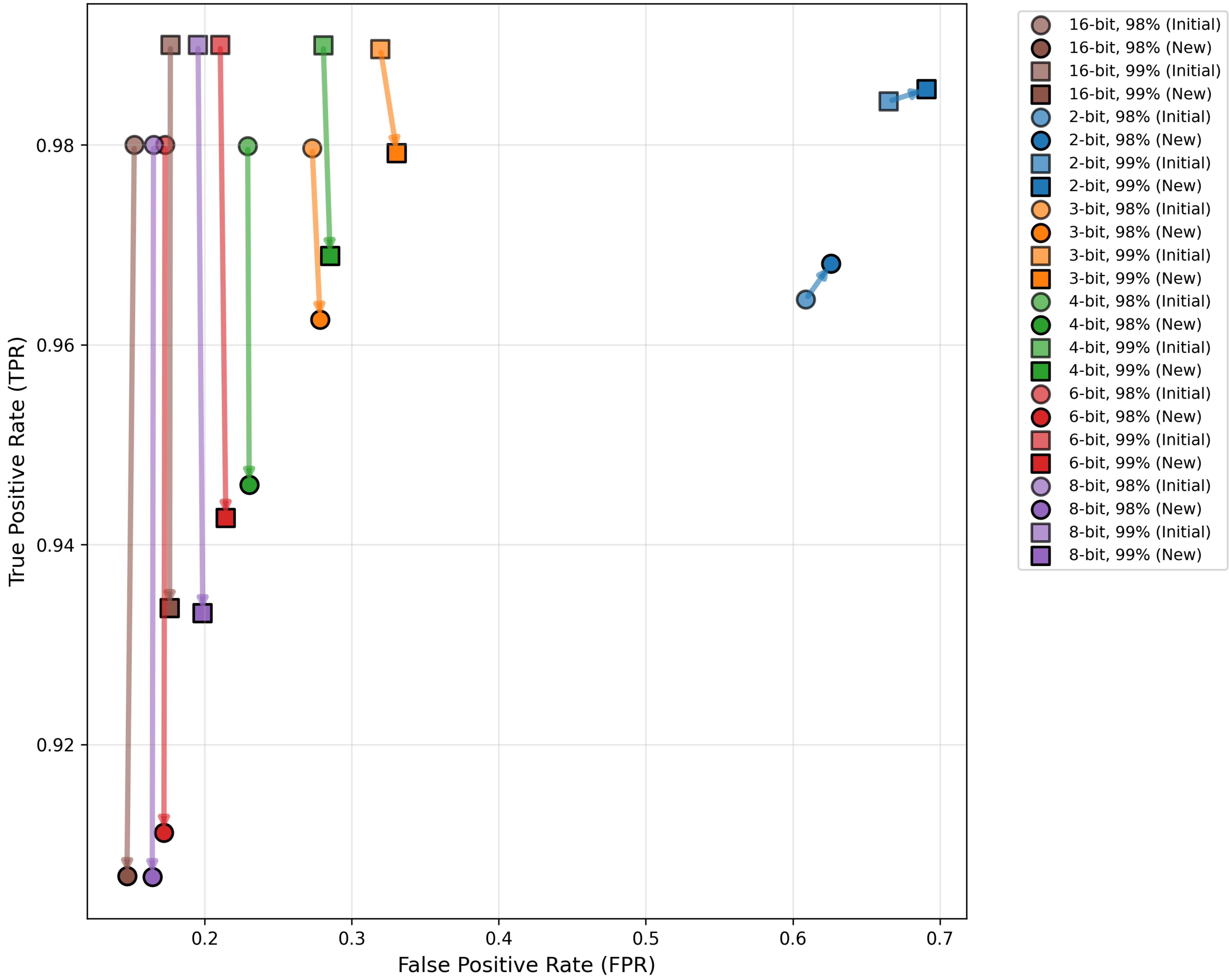
Random Accumulated Bit Flips

- Rough estimate: 1.2×10^{-5} peak SEU occupancy
  - RD53B SEU cross-section * BIB flux estimate / 5 Hz fill length
- Simulate at 10^{-3} — large safety factor and reduced statistical variation
- **Simple model shown here**
- Main effect is loss of signal efficiency.
- Especially pronounced for larger, better-performing models.



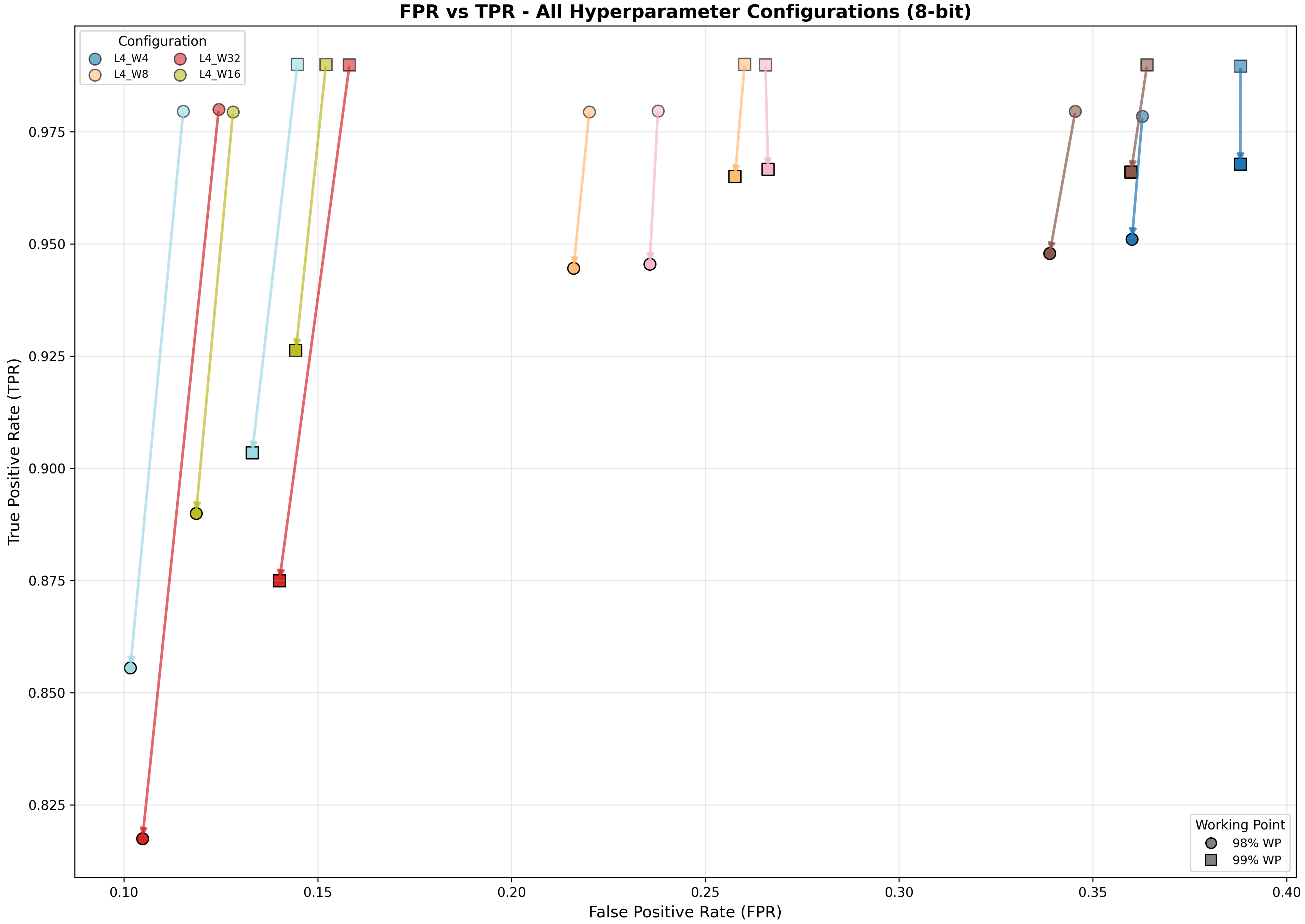
Quantization Dependence

- Less quantized models tend show lower susceptibility in signal efficiency
  - Speculate that this is because high-performing models cut very close to 1 on the score.



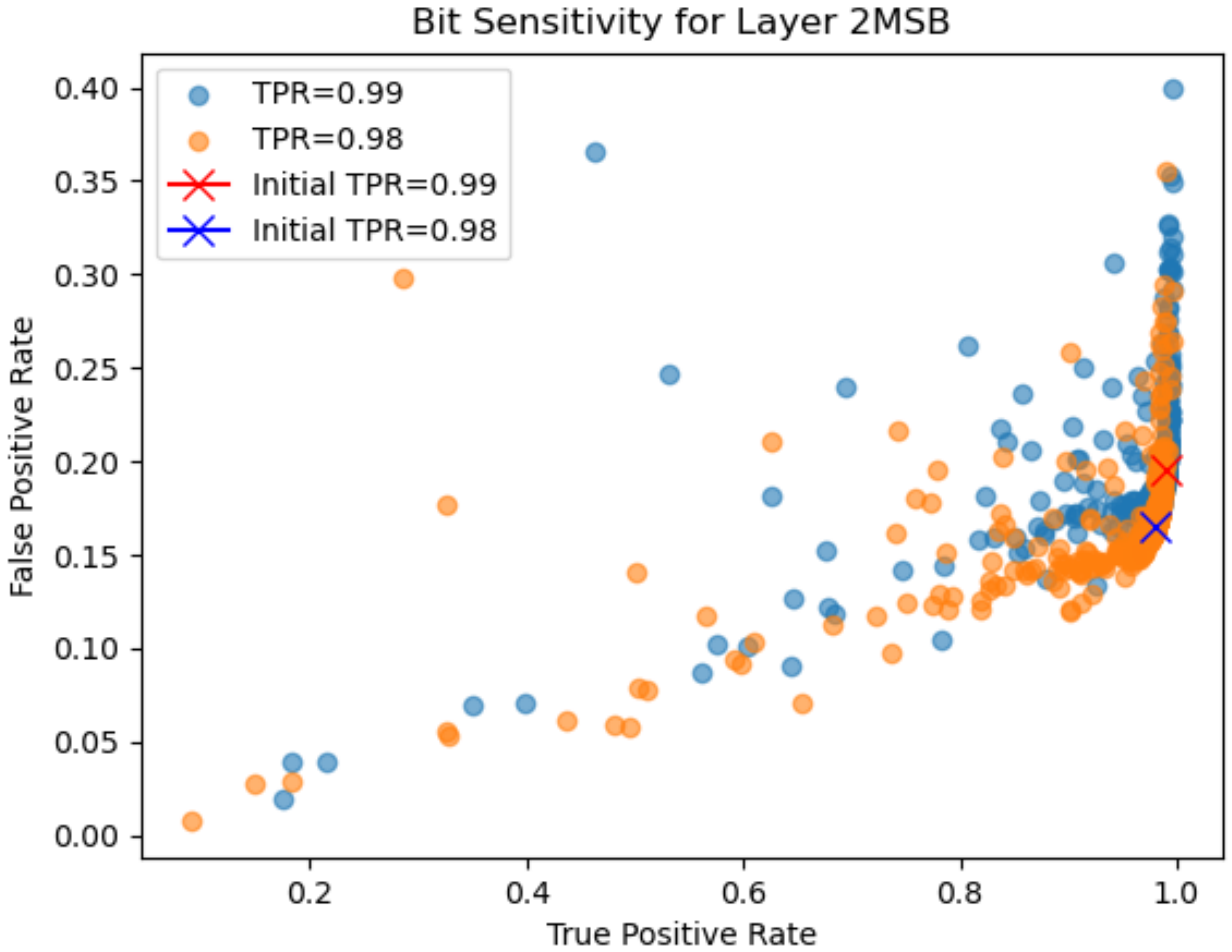
Random Accumulated Bit Flips

- **Intermediate model shown here**
- Background rates and signal efficiency both more sensitive
- Background rate decreases — harder to reach the hard cut being placed by the NN



Effect of Individual Bit Flips

- In addition to average performance, we want to study failure modes.
- Single bit flip can cause the data rate to spike or disable the chip.
- Most significant bits, bits in biases, and bits in the first layer have larger effects.
- Testing MSB of every parameter for large models may take ~1 day.
  - FKeras and other tools to guess most sensitive bits.



Result of flipping the 2 MSB of every weight in the simple model with 8-bit quantization

Takeaways and Future Directions

- Measurable effects on average performance are possible within a conservative safety factor.
- Individual bit flips can disable a chip for the entire fill.
- Possible mitigations to be studied:
  - Architecture dependence
  - Training procedure (including usage of dropout or injection of faults at training time)
  - Selective triplication (e.g. MSBs, most sensitive layers)
    - Must be robust to possible retraining.
- Expect these problems to be solvable at a minor cost to performance.

Backup

Random Accumulated Bit Flips

- Assume 5 Hz reconfiguration — collider fill length
- $\sigma_{\mathrm{SEU}} = 5 \times 10^{-14}\ \mathrm{cm}^2/\mathrm{bit}$ from RD53B
    - Strong process dependence
- Maximum flux ~$4 \times 10^4/\mathrm{BX}$
    - Flux used for SEU cross-section should not be the total particle flux
- Rough estimate: 1.2×10^{-5} peak SEU occupancy