University of California, Irvine

Real-Time Full-Band AI Triggering for Radio-based HE ν Detectors

Albert Likai Zhang [1,*] Hank Tang [1] Adam Barletta [1] Steven W. Barwick [1]

[1]University of California, Irvine

Radio detection of ultra-high-energy neutrinos

- A neutrino interaction in polar ice produces a particle shower and a nanosecond Askaryan radio pulse detectable over kilometer-scale distances.
- Sparse radio arrays instrument large ice volumes, but their sensitivity near threshold depends mainly on the **trigger**.
- In remote, power-limited stations, waveform classification must run locally.

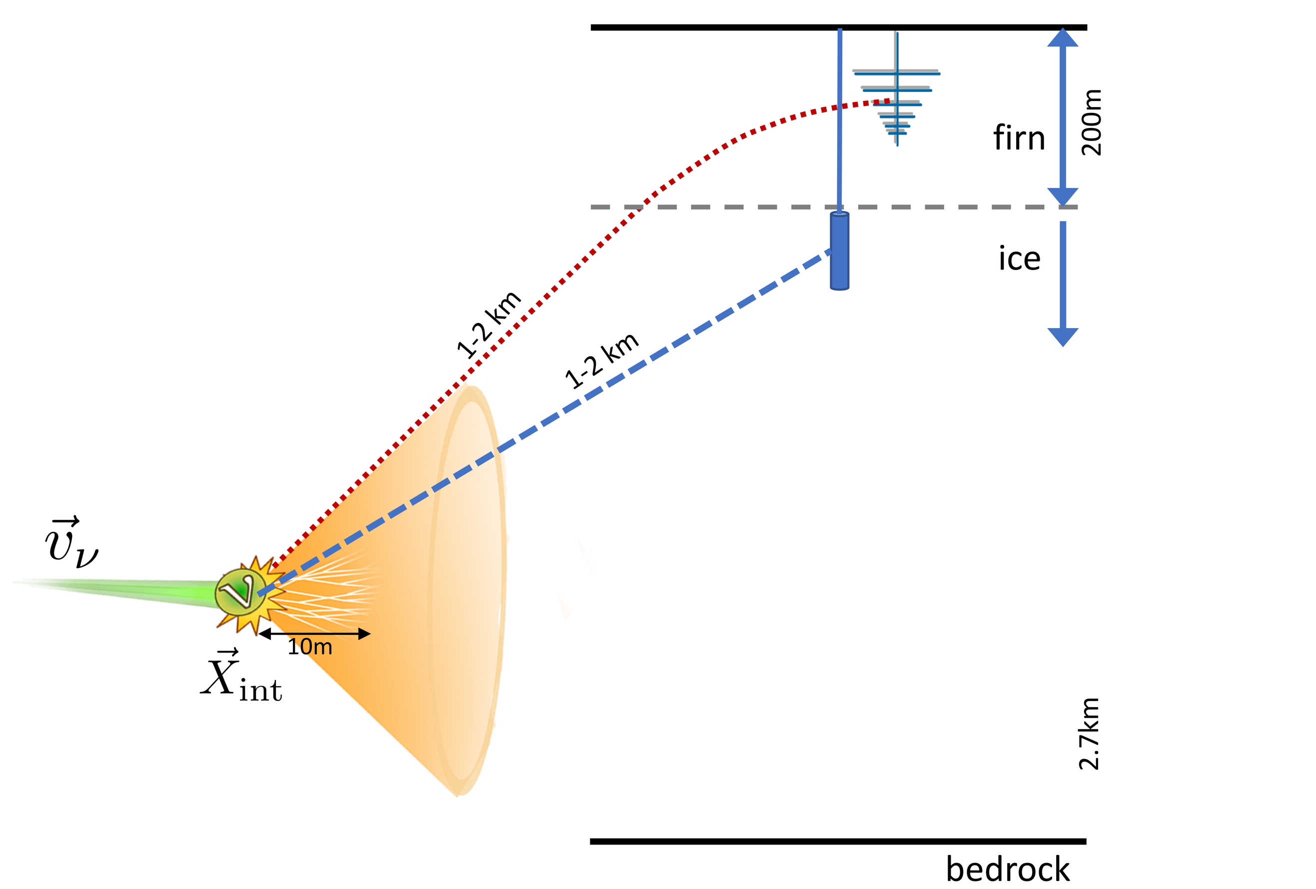

Figure 1. Askaryan radio emission in polar ice and representative antenna geometries. Adapted from Barwick and Glaser [1].

- **ARIANNA:** compact shallow stations with near-surface LPDAs.
- **RNO-G:** hybrid stations with surface LPDAs and $\sim$100 m-deep borehole antennas.
- **IceCube-Gen2:** larger South Pole array combining shallow and deep station components.

Why previous Hi-Lo threshold triggers hit a thermal-noise wall

- **Legacy trigger** use a bipolar Hi–Lo crossing within ~ 5 ns, followed by a 2-of-4 coincidence within ~ 30 ns. (however also implemented on FPGAs as a backup option)
- **Historical operating point:** $V_{\rm th} \approx 4V_{\rm RMS}$, with a global rate of 10^{-3}–10^{-2} Hz. Lowering every $\sim 0.13V_{\rm RMS}$ raises the thermal-noise rate by a factor of 10.

$$R_{\rm ch} \approx S\,Q\left(\frac{V_{\rm th}}{V_{\rm RMS}}\right), \qquad R_{N,m} \approx \binom{N}{m} R_{\rm ch}^m T^{m-1}. \tag{1}$$

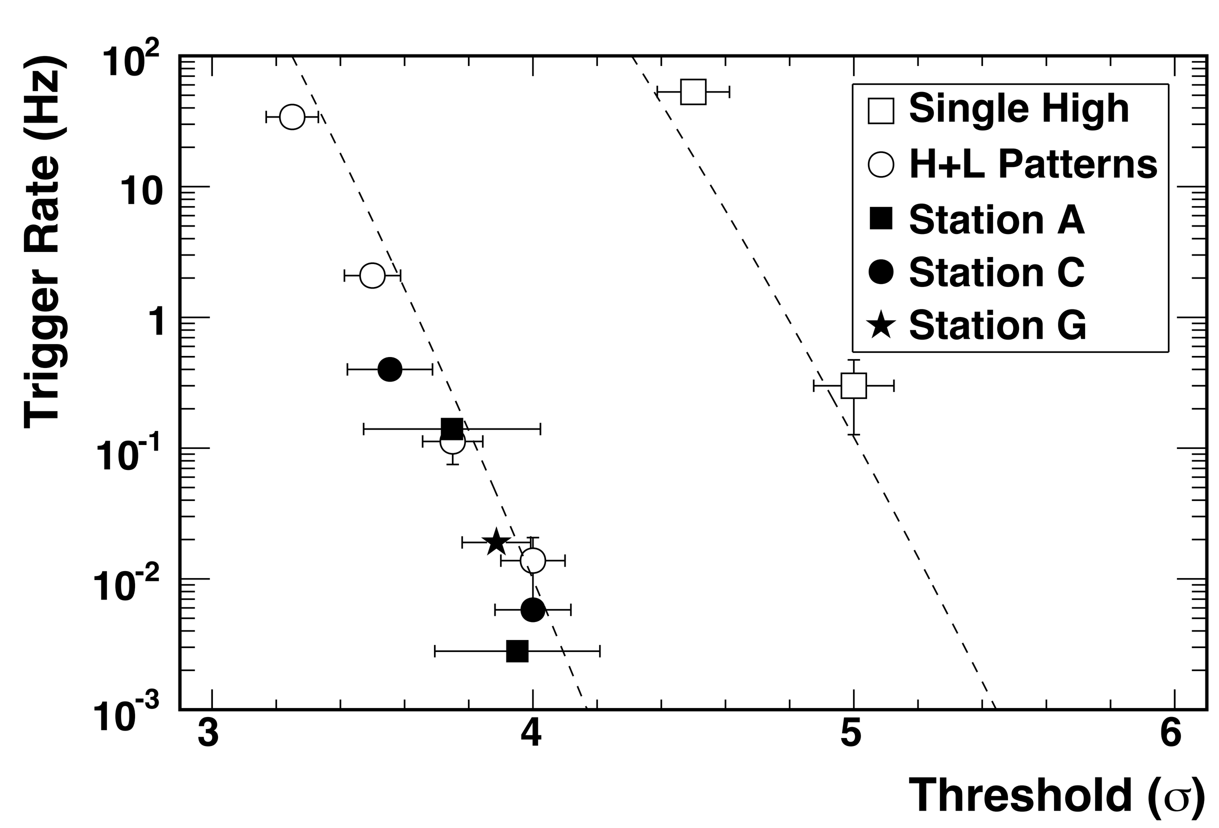

Figure 2. ARIANNA (Hi-Lo) trigger rate versus threshold. Hi–Lo patterns suppress thermal fluctuations, but the rate still rises steeply at lower thresholds. Around $1.8\times10^{8}\,\mathrm{s}^{-1}$ raw thermal crossings at $\sim 2V_{RMS}$. Adapted from Barwick et al. [2].

Our approach: continuous AI (CNN) inference, using information discarded by threshold logic, pushing to $\sim 2V_{\rm RMS}$ regime from $\sim 4V_{\rm RMS}$.

Continuous lane-parallel AI trigger system

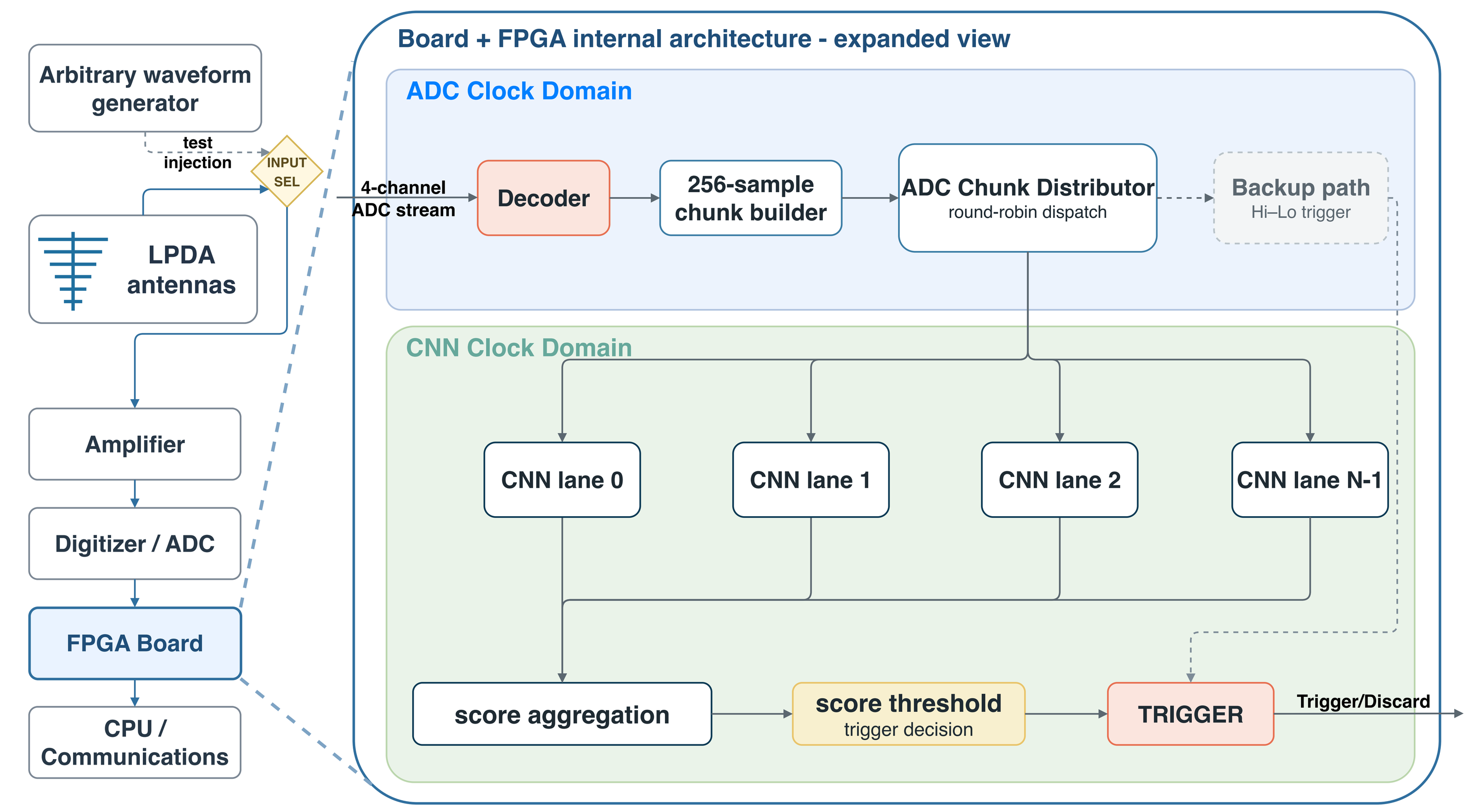

Figure 3. A typical readout chain. Continuous waveform chunks are dispatched to parallel FPGA inference lanes. Replication of a compact CNN core enables sustained multi-GSa/s processing.

- The CNN is the **only trigger**, every waveform chunk is classified continuously.
- Each input chunk contains four antenna channels and 256 time samples. Chunks are assigned round-robin to parallel CNN lanes.
- Each lane returns a scalar score. A final threshold comparison makes the decision.

From model to optimized CNN core on FPGA

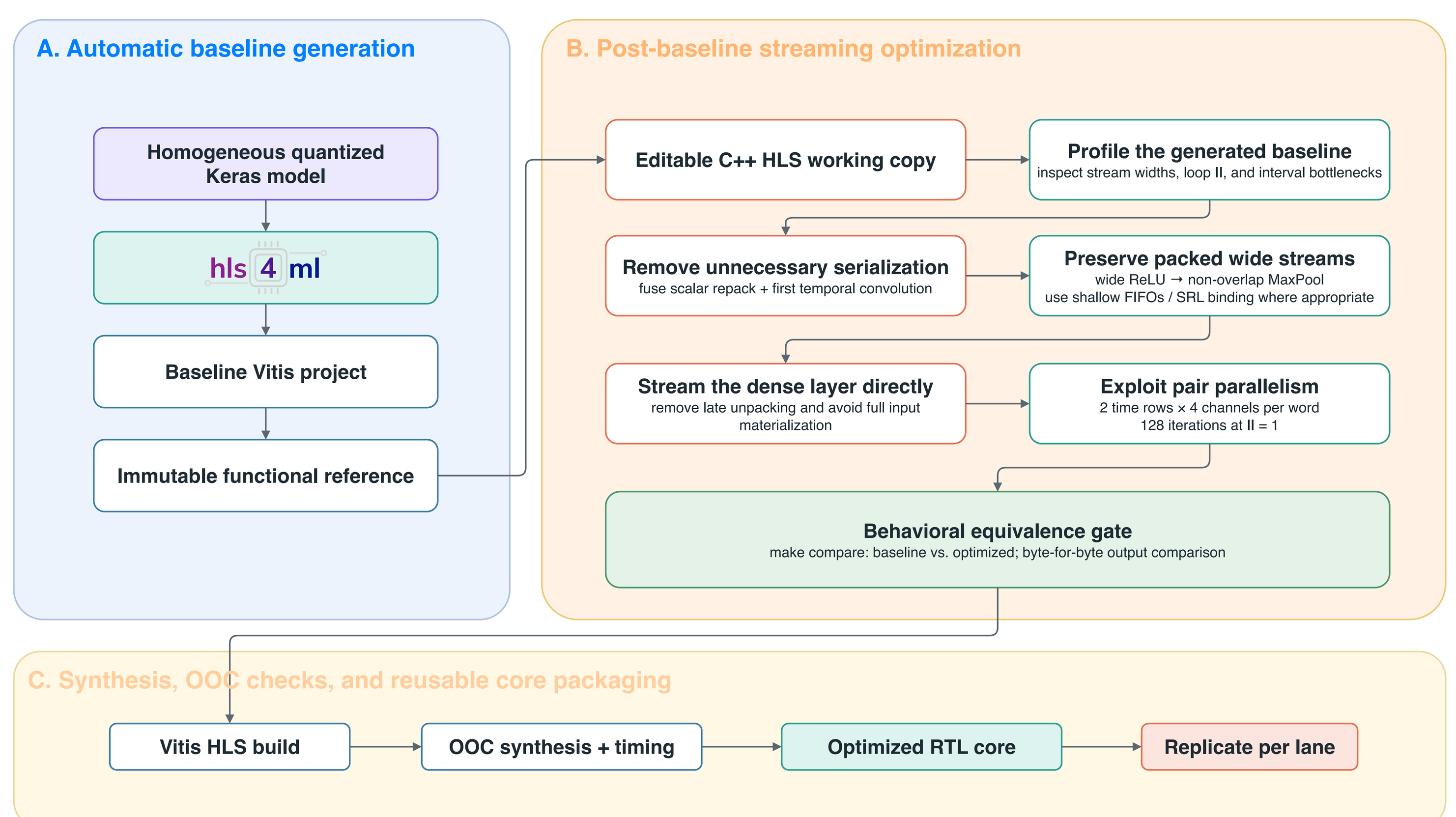

Figure 4. From a quantized Keras model to a reusable CNN core on FPGAs. The generated baseline is retained as a start point while the streaming dataflow is optimized for throughput.

- **Optimize the RTL dataflow, not the network semantics.** The baseline is refactored to remove unnecessary serialization and exploit pair-level parallelism.

Metric	Per-CNN-core result	% of FPGA
Input waveform chunk	4 channels $\times$ 256 samples $\times$ 12 bits	–
Inference interval per chunk	177 cycles $= 885$ ns at 200 MHz	–
Timing at 200 MHz	$+2.152$ ns WNS, timing met comfortably	–
LUT logic	$4{,}829$	2.2%
SRL shift registers	$1{,}093$	1.1%
Flip-flops	$2{,}953$	0.7%
DSP slices	7	0.4%
Block RAM	$2\times$ RAMB18	0.2%

Table 1. Out-of-context implementation result for one optimized CNN core on a Xilinx Kintex UltraScale+ XCKU5P FPGA. One lane classifies a 4-channel, 256-sample waveform chunk every 177 clock cycles. Further optimization is ongoing.

➤ This implementation demonstrates the feasibility of real-time full-band AI (CNN) processing on FPGAs, while leaving considerable room to scale both CNN capacity and lane parallelization.

Multi-GSa/s inference opens the low-SNR regime

2.0–2.4 **GSa/s/ch**	1.0–1.6 **GSa/s/ch**	$4V_{\rm RMS} \to 2V_{\rm RMS}$ **equivalent threshold**
projected inference capacity	typical ADC streaming range	vs. conventional Hi-Lo trigger

By exploiting waveform shape, this trigger can retain weak signal-like impulses that amplitude-only logic rejects. The implementation results demonstrate sufficient hardware headroom for an effective threshold near $2V_{\rm RMS}$.

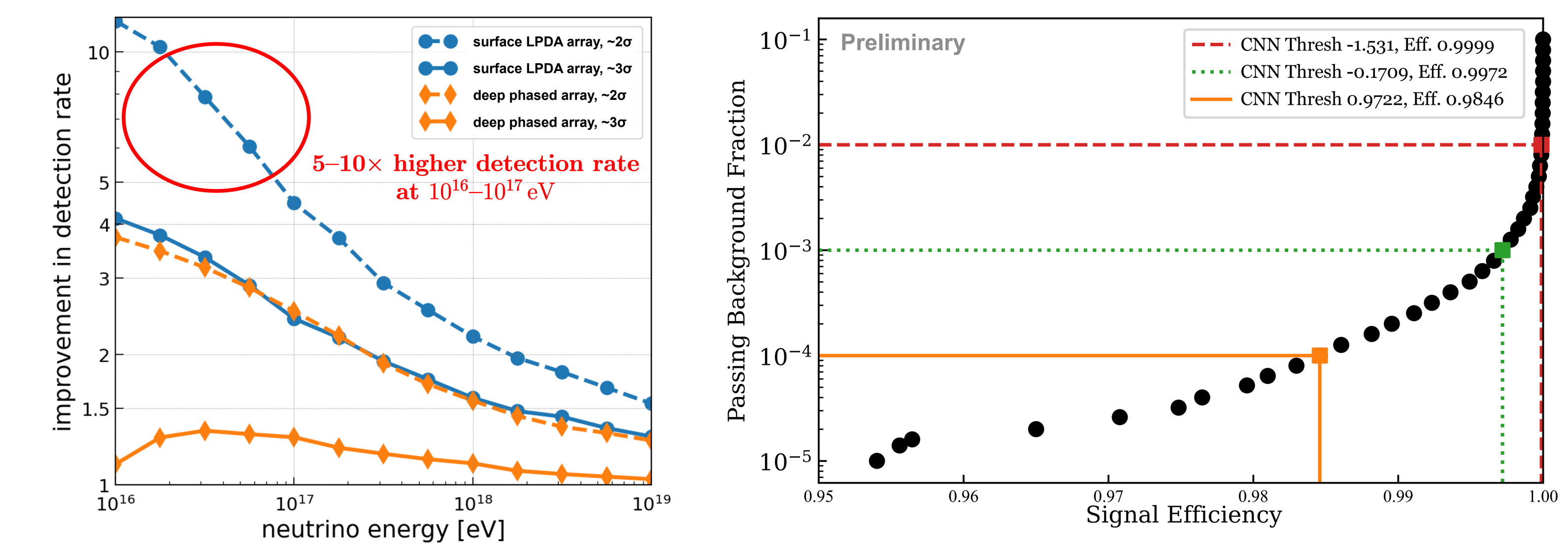

Figure 5. **Left:** Lowering the effective threshold toward $2V_{\rm RMS}$ yields a projected 5–$10\times$ detection-rate gain at $E_\nu \sim 10^{16}$–10^{17} eV. Adapted from Glaser et al. [3]. **Right:** CNN score cuts trade signal efficiency against thermal-noise rejection. At ~ 10 Hz trigger rate, the current compact CNN retains $\sim 95.4\%$ of signal-like events. Further optimization is expected to improve this operating point.

➤ First-ever low-power real-time FPGA implementation of full-band CNN triggering for high-energy physics:

- $\mathbf{5}$–$\mathbf{10\times}$ higher projected detection rate near threshold the largest gain occurs at $E_\nu \sim 10^{16}$–10^{17} eV, where weak Askaryan pulses are most easily lost.
- $\mathbf{0.9}$ **W** simulated dynamic FPGA power well within the 5 W power requirement.
- Performance successfully reaches optimistic goal

Note that this is far from the ultimate performance; with the demonstrated feasibility of deploying larger CNNs and further optimization, the performance is expected to improve further.

Outlook: toward a field-ready radio trigger

The present AI (CNN) trigger is a pathfinder, not the endpoint. Its sub-watt dynamic power consumption and high throughput leave enough headroom to move beyond this compact CNN: larger or different models, more parallel CNN cores, and detector-level optimization are all realistic next steps.

The immediate goal is optimization toward deployment. With the main improvement paths now identified, we will refine the trigger, install and test it with realistic input streams, followed by live operation in the lab and eventually in the field.

The remaining throughput and power headroom also leave room for more capable networks. A mature version of this technology is expected to be applied in ARIANNA, RNO-G, and future IceCube-Gen2, improving the sensitivity of high-energy neutrino searches near threshold.

Additional Notes: They are derived from generated neutrino-like waveforms and archival data collected with ARIANNA and earlier radio-detector systems; they are not yet an in-ice measurement of the new continuous FPGA trigger. Live testing and deployment will take place in the near future.

References

[1] Steven W Barwick and Christian Glaser. Radio detection of high energy neutrinos in ice. In *The Encyclopedia of Cosmology: Set 2: Frontiers in Cosmology Volume 2 Neutrino Physics and Astrophysics*, pages 237–302. World Scientific, 2024.

[2] SW Barwick, Eric C Berg, DZ Besson, T Duffin, JC Hanson, SR Klein, SA Kleinfelder, K Ratzlaff, C Reed, M Roumi, et al. Design and performance of the arianna hra-3 neutrino detector systems. *IEEE Transactions on Nuclear Science*, 62(5):2202–2215, 2015.

[3] Christian Glaser, Alan Coleman, and Thorsten Glüsenkamp. Nuradioopt: Optimization of radio detectors of ultra-high energy neutrinos through deep learning and differential programming. In *38th International Cosmic Ray Conference, ICRC 2023, Nagoya, Japan, July 26 2023-August 3 2023*. Sissa Medialab Srl, 2024.

albert.zhang@cern.ch