Quantum Track Reconstruction in the LHCb VELO

From the Hamiltonian to HHL to the 1-Bit Quantum Filter

George Scriven

Nikhef · Maastricht University · Hasselt University

Quantum Tracking Mini Workshop, Maastricht — 11 June 2026

Based on TrackHHL: arXiv:2511.11458 (HHL) · arXiv:2601.07766 (1-Bit Quantum Filter)

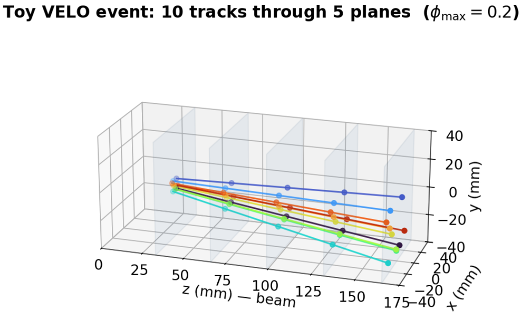
Outline

The problem: tracking is a combinatorial needle-in-a-haystack

- The LHCb **VELO** reconstructs charged particles (**tracks**) from **hits** on silicon planes.
- HL-LHC pile-up (many more simultaneous collisions) ⇒ reconstruction cost threatens to outpace classical resources.
- With T hits per plane each plane-gap offers T^2 candidate links, only T real:

$$n_{\text{seg}} = 4T^2 \quad \text{vs} \quad n_{\text{true}} = 4T \quad \text{(signal fraction } 1/T\text{)}.$$

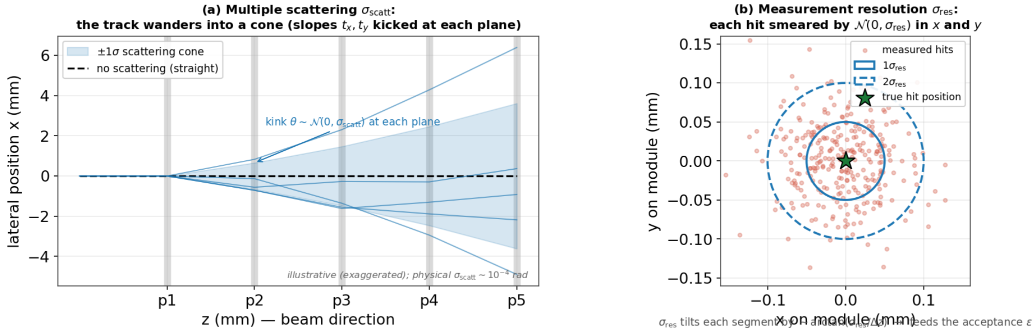
- At $T=400$: **640,000** candidates, **1,600** true (0.25%).
- **Idea:** cast track finding as a *linear system* $A\mathbf{x} = \mathbf{b}$, solvable classically *or* on a quantum computer.



A generated toy VELO event: tracks fan from the collision point (primary vertex) through 5 planes.

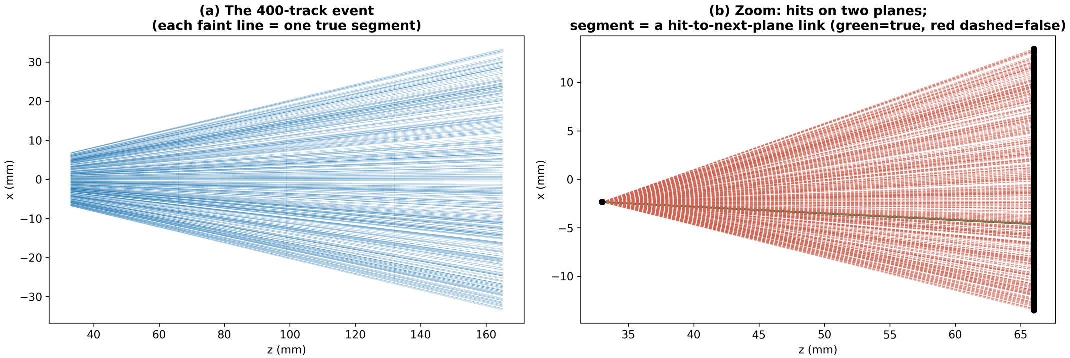
The toy injects the two physical detector noises

The two detector noises the toy injects — both widen the triplet kink angle, set by $\varepsilon = \sqrt{2(s\sigma_{\mathrm{scatt}})^2 + 12\arctan^2(s\sigma_{\mathrm{res}}/\Delta z) + \ldots}$



- **Multiple scattering** σ_{scatt}: kicks the slopes (t_x, t_y) at each plane ⇒ a cone around the ideal track.
- **Resolution** σ_{res}: smears each measured hit in x, y. Both *widen the triplet kink angle* ⇒ they set ε.

Step 1: from hits to segments



- A **segment** is a candidate hit→hit link between adjacent planes; assign each a variable x_i (its *activation*).
- **Green** = true (both hits from the same particle); **red dashed** = false (cross-track accident).
- The reconstruction task: switch ON the green, OFF the red — $4T^2$ binary decisions.

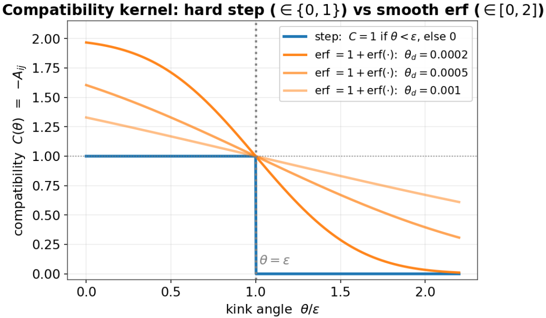
Step 2: the kink-angle cost — which segments support each other

- Two segments sharing a middle hit form a **triplet** with kink angle $\theta_{ij} = \arccos(\hat{v}_i \cdot \hat{v}_j)$.
- A real particle goes nearly straight $\Rightarrow$ small θ. The **cost/compatibility kernel** is explicit:

$$C(\theta_{ij}) = \begin{cases} 1 & \text{share a hit and } \theta_{ij} < \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

- smooth variant: $C = 1 + \mathrm{erf}\left(\frac{\varepsilon - \theta}{\theta_d \sqrt{2}}\right) \in [0, 2]$.
- ε is *derived from the detector noise*:

$$\varepsilon = \sqrt{2(s\sigma_{\mathrm{scatt}})^2 + 12 \arctan^2 \frac{s\sigma_{\mathrm{res}}}{\Delta z} + 2\theta_{\mathrm{min}}^2}.$$



Step 3: the explicit Hamiltonian

Reward mutually-compatible segments, penalise activation, bias every segment on:

$$H(\mathbf{x}) = \underbrace{\frac{\gamma+\delta}{2}\sum_i x_i^2}_{\text{activation penalty}} - \underbrace{\frac{1}{2}\sum_{i\neq j} C(\theta_{ij})\, x_i x_j}_{\text{triplet reward (attractive)}} - \underbrace{\delta \sum_i x_i}_{\text{bias}}$$

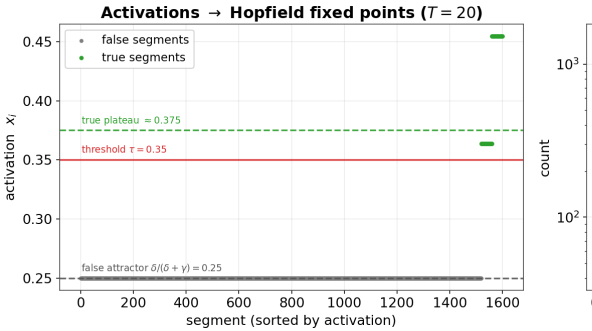
- An Ising/Hopfield-type energy over segment activations (Denby–Peterson lineage); defaults $\gamma=3$, $\delta=1$.
- Physically: each active segment "pulls up" the segments that continue it through a shared hit at a small kink angle; lone segments are pulled down.

Step 4: minimising H is a matrix inversion

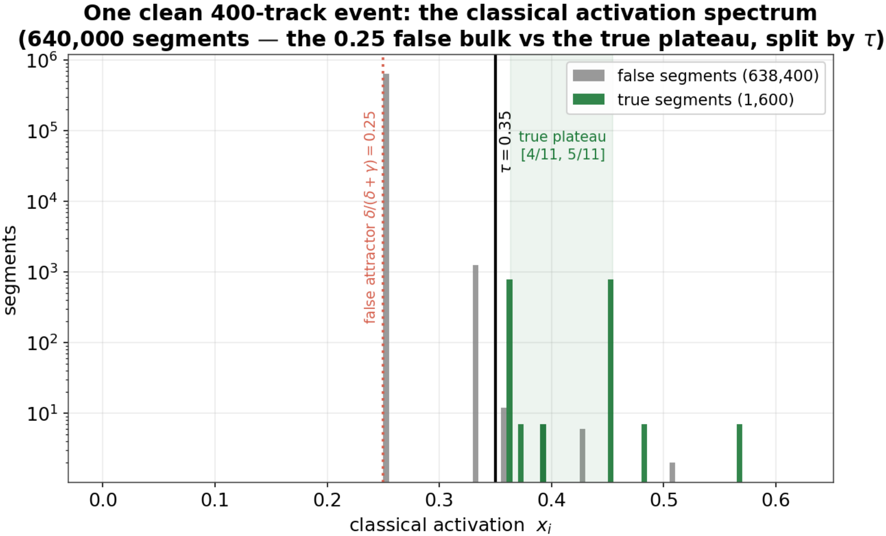
Stationarity, $\partial H / \partial x_i = 0$ for every segment i:

$$(\gamma+\delta)\,x_i \;-\; \sum_j C(\theta_{ij})\,x_j \;-\; \delta = 0 \qquad \Longleftrightarrow \qquad \boxed{A\,\mathbf{x} = \delta\mathbf{1}, \quad A = (\gamma+\delta)I - C}$$

- Reconstruction = **solve a sparse linear system**; classically: $\mathbf{x} = A^{-1}\delta\mathbf{1}$ (LU/CG).
- **Hopfield fixed points** give an absolute scale: false/isolated $\to \delta/(\delta+\gamma) = 0.25$; true chain $\to \left[\frac{4}{11}, \frac{5}{11}\right]$; threshold $\tau{=}0.35$.
- Classical solve = state-of-the-art efficiency [arXiv:2511.11458].



What the inversion produces on one clean 400-track event



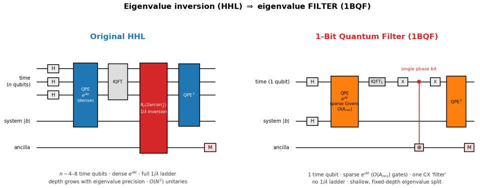
HHL: the quantum linear-system algorithm

Harrow–Hassidim–Lloyd solves $A\mathbf{x} = \mathbf{b}$:

1. prepare $|\mathbf{b}\rangle$ over the segment basis;
2. **QPE** (quantum phase estimation): e^{iAt} writes each eigenvalue λ_k onto an n_{time}-qubit clock register;
3. controlled R_y: amplitude $\propto 1/\lambda_k$ onto an ancilla;
4. uncompute, post-select ancilla $= 1 \Rightarrow$ state $\propto A^{-1}\mathbf{b}$.

Exponential speedup in the number of hits — *contingent on* efficient QPE & readout.

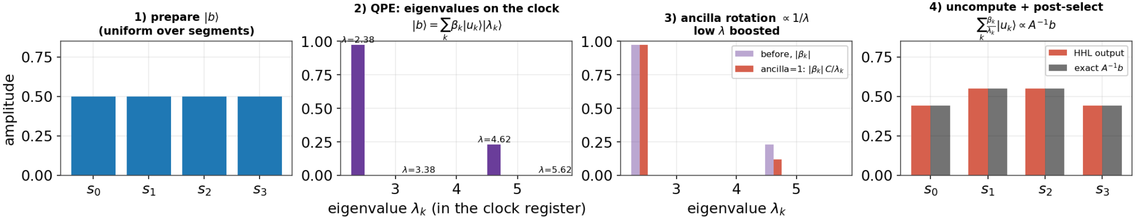
Catch: dense e^{iAt} + high-precision QPE $\Rightarrow$ prohibitive depth on near-term hardware.



Left: the HHL circuit. Right: where we are heading.

HHL stage by stage: the statevector at every step

HHL stage by stage on a true-track block (P_4): the statevector at every step

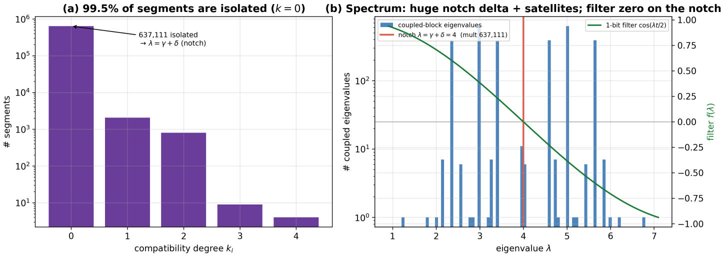


On a true-track block (P_4): $|b\rangle \xrightarrow{\text{QPE}} \sum_k \beta_k |u_k\rangle|\lambda_k\rangle \xrightarrow{R_y(1/\lambda)}$ ancilla $= 1$ branch $\xrightarrow{\text{post-select}} \propto A^{-1}\mathbf{b}$.

TrackHHL: restrict the phase estimation to one bit

[arXiv:2511.11458 — Chiotopoulos, Lucio Martinez, Nicotra *et al.*]

- For this Hamiltonian the spectrum is **bimodal**: a huge bulk at $\lambda = \gamma + \delta$ (lone/false segments) and a narrow band of track modes away from it.
- One bit of eigenvalue information — *on the bulk or off it* — is already the answer.
- Restricting QPE to **one clock qubit**: cuts circuit depth by up to $\mathbf{10^4}$; resolves the read-out problem (the output is binary); enables primary-vertex post-processing.



The event spectrum: 637k eigenvalues exactly at $\gamma + \delta$ (the false bulk) + coupled satellites.

The 1-Bit Quantum Filter: the maths

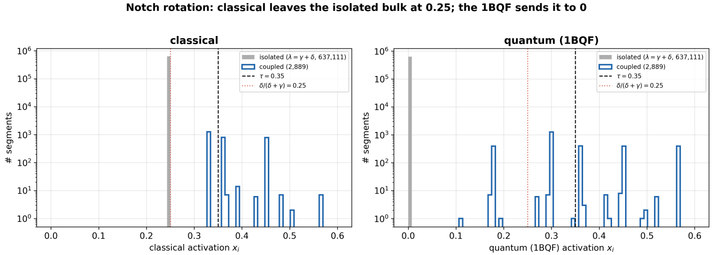
[arXiv:2601.07766 — Chiotopoulos, Nicotra, Scriven *et al.*]

Choose the evolution time $t = \pi/(\gamma + \delta)$. One-qubit QPE on e^{iAt} leaves, per eigencomponent,

$$|u_k\rangle \left(\cos\frac{\lambda_k t}{2}\,|0\rangle_{\text{clock}} - i\,\sin\frac{\lambda_k t}{2}\,|1\rangle_{\text{clock}} \right),$$

and the `X--CX--X` ancilla flip + post-selection keeps each mode with weight

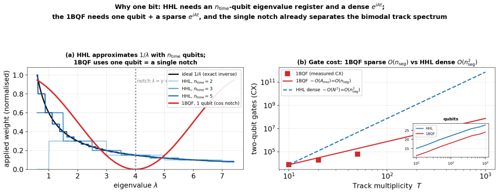
$$\boxed{w_Q(\lambda) = \cos\left(\frac{\pi}{2}\,\frac{\lambda}{\gamma+\delta}\right)} \qquad w_Q(\gamma+\delta) = 0 \ \ \text{(the notch, placed } \textit{exactly} \text{ on the false bulk).}$$



The 0.25 isolated-false bulk (637k segments) is rotated to 0; the true band survives.

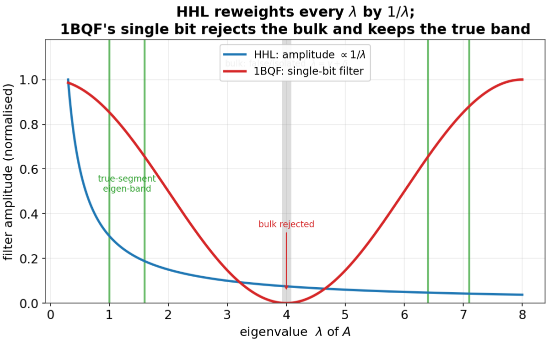
Why the 1BQF wins — in brief

- **One clock qubit** instead of an n_{time}-qubit eigenvalue register.
- **Depth** cut by up to 10^4 vs full HHL.
- **Sparse** e^{iAt} (two-level/Givens decomposition) $\Rightarrow$ gate complexity $O(\sqrt{N} \log N)$.
- **Read-out solved**: the answer is binary (on/off the notch), no amplitude tomography.
- **NISQ-validated** (today's noisy devices): Quantinuum H2 and IBM Heron noise models.
- **Enables PV reconstruction** as classical post-processing.



The conceptual reformulation: inversion $\rightarrow$ filtering

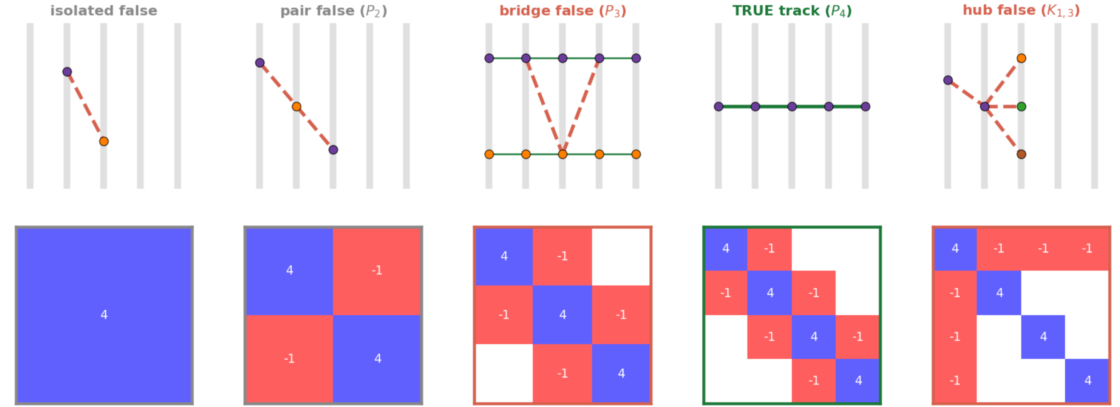
- We never needed the *numbers* in $A^{-1}\mathbf{b}$ — only a **binary decision** per segment: on or off.
- The spectrum is **bimodal**: false bulk *at* $\gamma + \delta$, true band *away* from it.
- So **one bit of spectral information classifies**: "matrix inversion" becomes "**project out the known-bad eigenspace**".
- HHL's eigenvalue *precision* was wasted effort for this problem — the 1BQF spends one bit exactly where the physics puts the information.



$1/\lambda$ reweighting (HHL) vs the single-notch filter (1BQF).

Every cluster type in segment space — and its block of A

Each cluster in segment space (top) and its block of A (bottom): $A = (\gamma + \delta)I - C$ is block-diagonal over clusters



$\lambda = \{4\}$
1 on the notch — quantum: erased ✓

$\lambda = \{3, 5\}$
0 on the notch — quantum: rejected by τ

$\lambda = \{2.59, 4, 5.41\}$
1 on the notch — quantum: SURVIVES

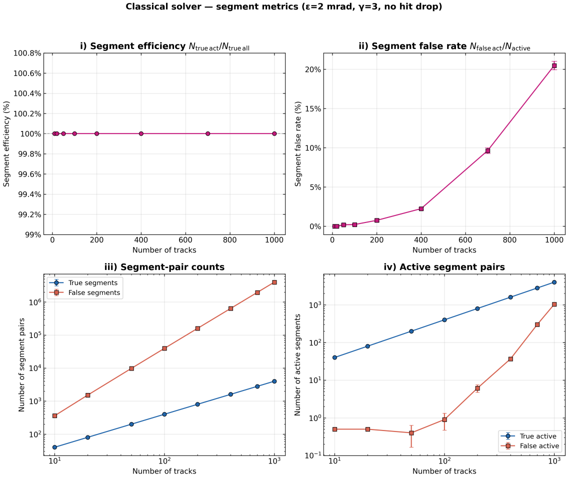
$\lambda = \{2.38, 3.38, 4.62, 5.62\}$
0 on the notch — quantum: kept ✓

$\lambda = \{2.27, 4, 4, 5.73\}$
2 on the notch — quantum: SURVIVE

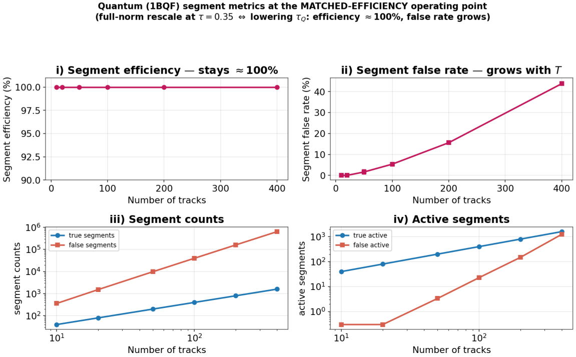
green solid = true segment · red dashed = false segment · dot colour = which real track the hit belongs to

Results: segment efficiency stays high; the false rate grows

Classical

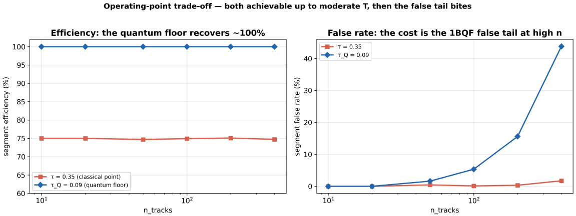


1BQF at matched efficiency



Both solvers: efficiency ≈ 100% **across multiplicity; the false rate grows with** T (classical $0 \to 21\%$ by T=1000; quantum $0 \to 44\%$ by T=400) — the coupled cross-track clusters.

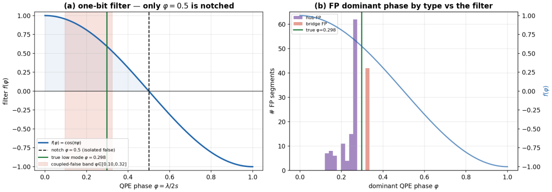
The efficiency–false-rate trade-off



- Recovering *every* true segment forces the threshold into the false band: **high efficiency and high false rate come together**.
- Quantum at matched efficiency ($\approx 100\%$): false rate $\rightarrow$ 44% at T=400.
- The alternative fixed-τ point (eff 75%, far 1.7%) only *hides* the trade-off by discarding a quarter of the true segments.
- The **operating curve**, not a single point, is the result.

Why the trade-off is fundamental

- The notch removes only the **isolated** false class (it *is* the bulk eigenvalue).
- Surviving false positives = **coupled cross-track bridges/hubs**: *good*, off-notch eigenvalues.
- In QPE phase they fall in the **same band as true segments —** one zero cannot separate them.
- **Topological degeneracy**: a false chain shaped like a true track has an *identical* A-block $\Rightarrow$ no threshold, classical or quantum, can split them.

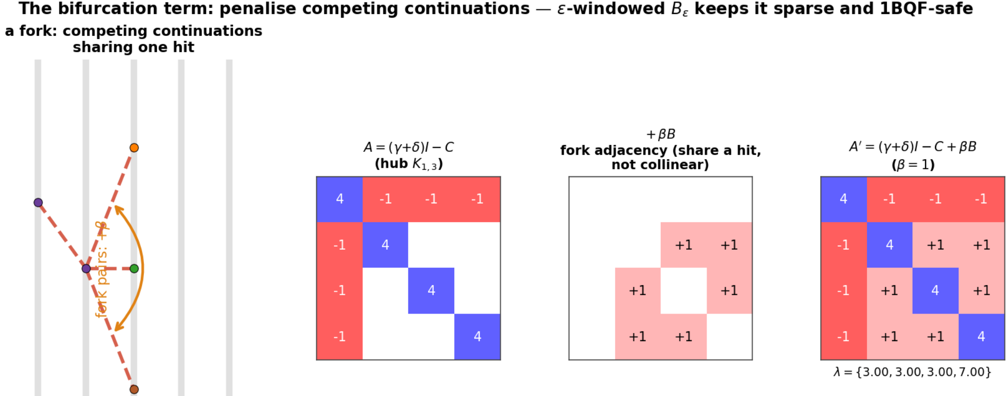


Add a bifurcation (fork) penalty to the Hamiltonian

The survivors are **competing continuations** — segments fanning from a shared hit. Penalise them:

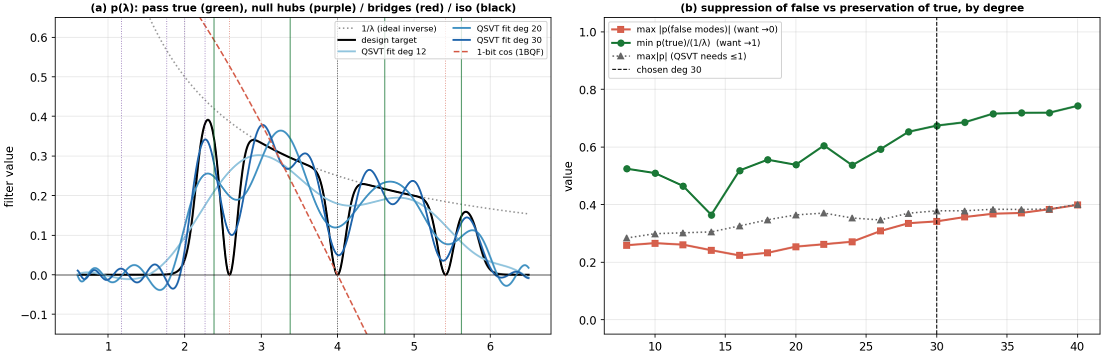
$$\boxed{A' = (\gamma + \delta)\, I - C + \beta B}, \qquad \mathbf{b} = \delta \mathbf{1},$$

with B the **fork adjacency**: segments sharing a start/end hit that are *not* collinear continuations. Every segment touches exactly two hits $\Rightarrow$ the diagonal stays constant $\Rightarrow$ **still 1BQF-compatible**.



Beyond one notch: Quantum Singular Value Transformation

- The 1BQF is QSVT restricted to $p = \cos$ — **one** notch.
- **QSVT** (Gilyén–Su–Low–Wiebe) realises *any* bounded polynomial $p(A)$ from a block-encoding of A + d phase rotations.
- Target: a **band-limited inverse** — $p(\lambda) \approx 1/\lambda$ on true-track eigenvalues, $p(\lambda) \approx 0$ on every false mode (*many* notches).
- Works *iff* false/true eigenvalues separate; first results: degree-$\sim$30 lifts AUC 0.49/0.64 $\rightarrow$ **0.88** [backup].



Summary

- Hits $\to$ segments $\to$ explicit kink-angle cost $C(\theta)$ $\to$ Ising-like $H(\mathbf{x})$ $\to$ stationarity $\Rightarrow$ **matrix inversion** $A\mathbf{x} = \delta\mathbf{1}$, $A = (\gamma + \delta)I - C$.
- **HHL** inverts it but needs deep circuits; **TrackHHL** restricts QPE to one bit (10^4 depth cut); the **1BQF** reframes inversion as a **single-notch spectral filter** — $O(\sqrt{N} \log N)$, NISQ-ready.
- On the toy: **efficiency stays ≈100%, the false rate grows with density** — a fundamental trade-off set by topologically-degenerate cross-track clusters.
- Hamiltonian-level fix: the ε-**windowed fork term** (targeted, sparse, 1BQF-safe). Algorithm-level next step: **QSVT** polynomial filters (AUC $\to$ 0.88).

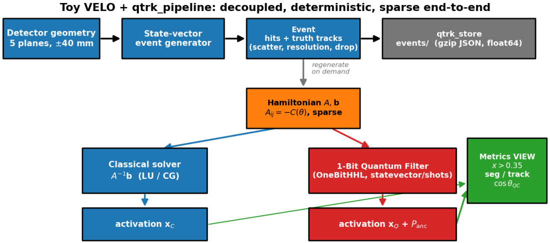
arXiv:2511.11458 (HHL) · arXiv:2601.07766 (1BQF) · toy: `LHCb_VeLo_Toy_Model`

Thank you

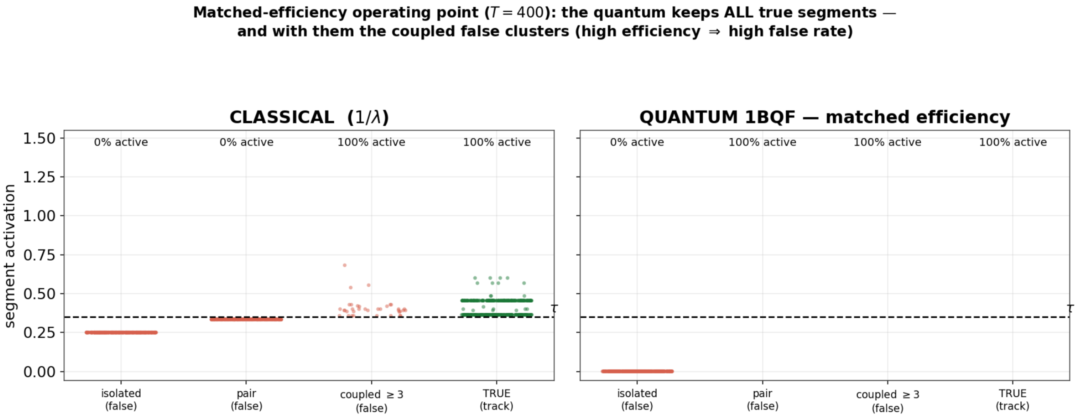
Questions?

Backup: a controlled toy VELO — our testbed

- `LHCb_VeLo_Toy_Model`: **5 planes** at $z = 33, 66, 99, 132, 165$ mm, ± 40 mm.
- Deterministic and fully controlled: sweep multiplicity T, noise, density, kernel.
- Decoupled pipeline: events stored once; the matrix A regenerated on demand (sparse); metrics recomputed as a *view*.
- Purpose: isolate *how each physical factor* moves the algorithm.

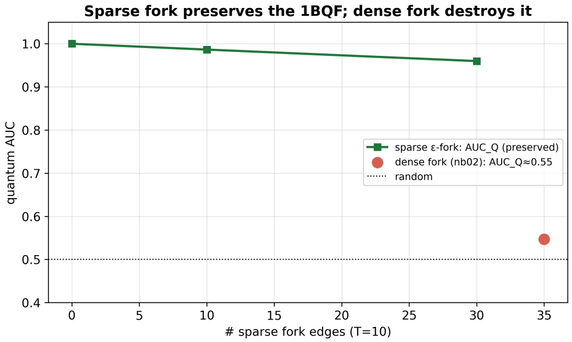


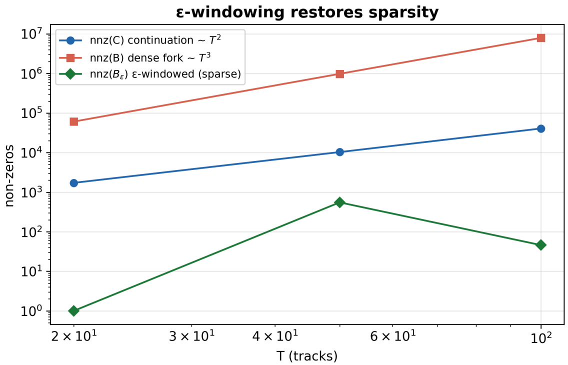
Backup: class by class at matched efficiency (T=400)



The quantum keeps **all** true segments — and with them the coupled false clusters: the price of efficiency.

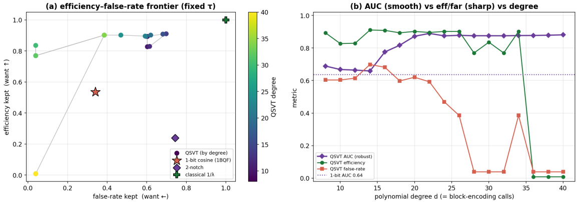
Backup: the fork term — dense breaks the quantum, ε-windowed fixes it





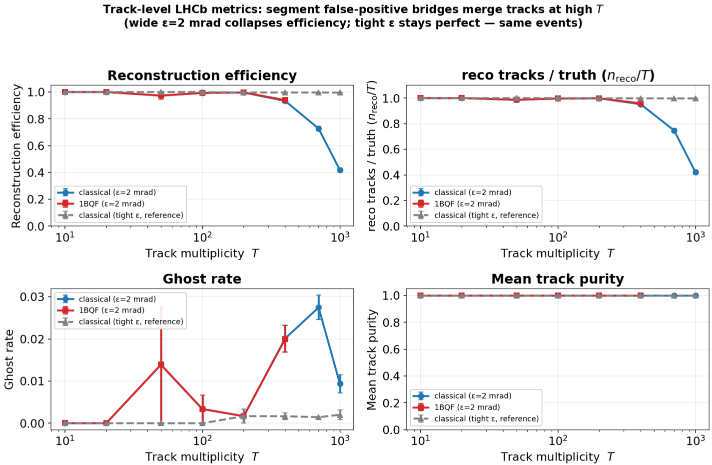
- **Dense** fork ($O(T^3)$ pairs): classically a benign uniform down-scaling (AUC ≈ 1) but **destroys the 1BQF** (AUC $\to$ 0.55) and the sparsity ($\sim$20$\times$ slower).
- ε**-windowed** B_ε (only near-collinear competitors): **sparse**, **targeted** (false rate $\to$ 0 at $\sim$2% efficiency cost), and **1BQF-safe** (AUC 0.96–0.99).

Backup: QSVT — a polynomial filter lifts true/false separation



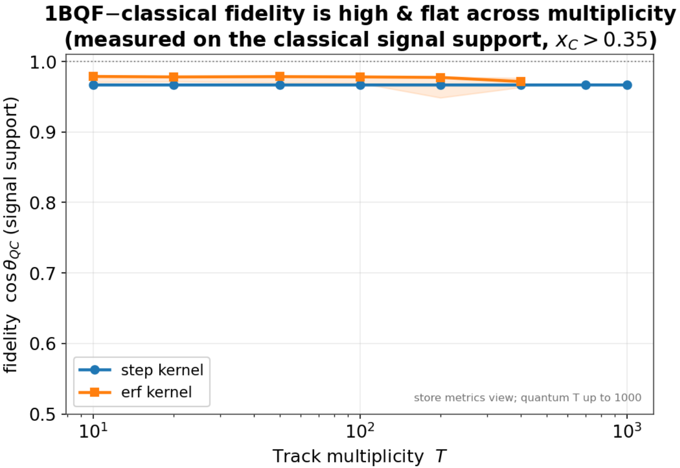
- A degree-$\sim$30 polynomial lifts true/false **AUC** from 0.49 (classical) / 0.64 (1-bit) to **0.88**.
- Nulls *all* off-length bridges and $\sim$95% of hubs.
- **Cost:** $\sim$30 block-encoding calls $\times$ $\sim$2$\times$ amplitude-amplification $\approx$ tens$\times$ the 1-bit depth; sparsity preserved.
- **Irreducible floor:** a false bridge of the *same length* as a true track (0% at $T \leq 200$; grows with density).

Backup: track-level metrics — segments propagate to tracks



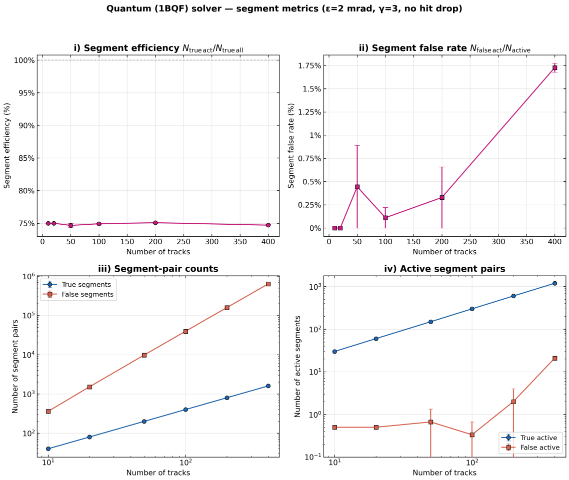
Wide $\varepsilon \Rightarrow$ bridges merge tracks $\Rightarrow$ reconstruction efficiency collapses ($1.0 \to 0.42$); tight ε stays perfect.

Backup: 1BQF–classical fidelity vs multiplicity



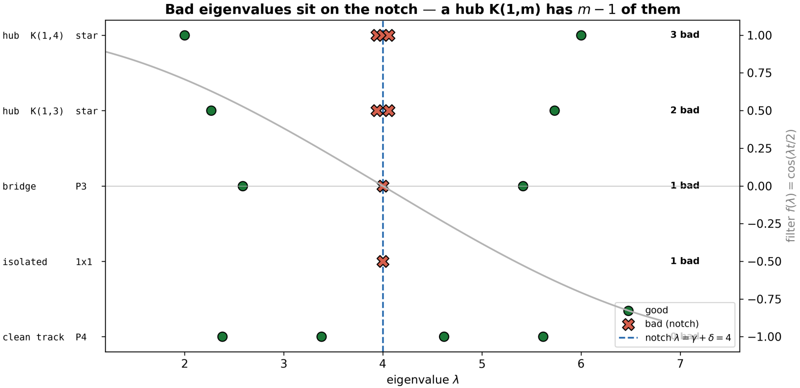
On the classical signal support the 1BQF is faithful: $\cos\theta_{QC} \approx 0.97$, flat to $T{=}1000$.

Backup: fixed-threshold operating point ($\tau = 0.35$)

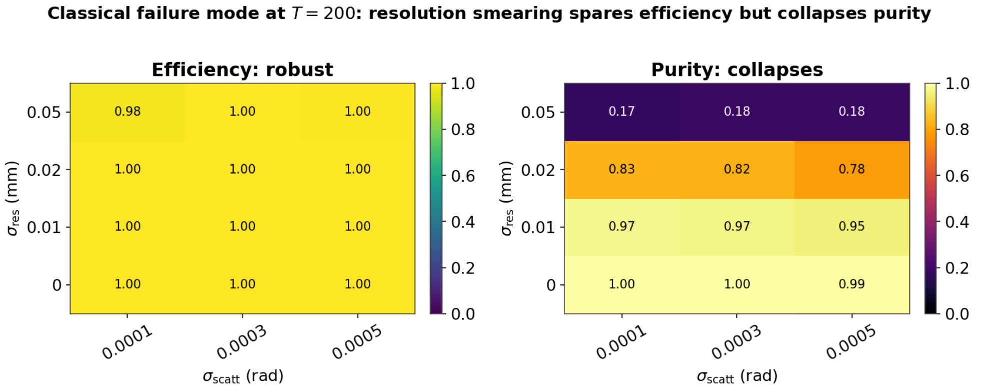


The same data at fixed τ: eff 75%, far 1.7% — the conservative end of the same operating curve.

Backup: the bad-eigenvalue ladder

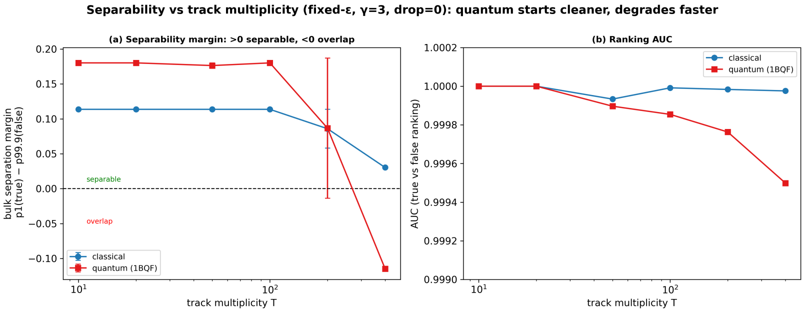


Backup: classical failure mode under noise



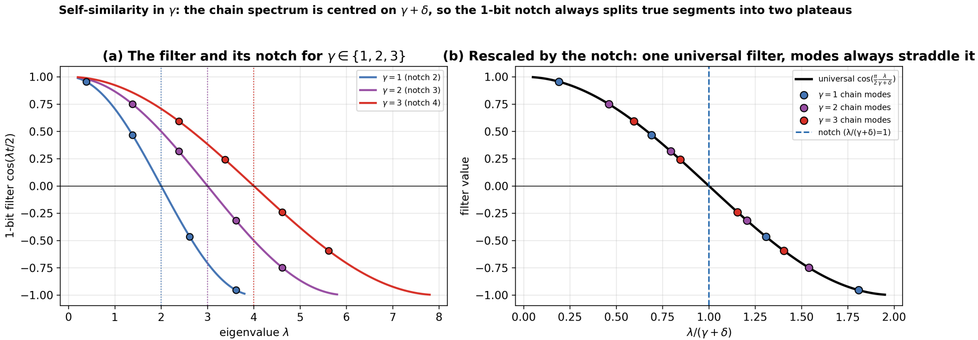
ε scales with noise $\Rightarrow$ efficiency robust, but **purity** collapses under resolution smearing.

Backup: separability vs multiplicity



Quantum starts cleaner, degrades faster; margin negative $\sim T{=}400$ — the coupled-false ceiling.

Backup: γ self-similarity of the notch



Backup: classical vs quantum per cluster (analytic)