

Pauli Observables for Collider Event Topology

What spatial information survives beyond global event properties?

Irais Bautista Guzmán

QEX
IGFAE
INSTITUTO GALEGO DE FÍSICA DE ALTAS ENERXÍAS
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Motivation

High-multiplicity collisions exhibit long range correlations and collective-like anisotropies.

We test whether a compact circuit can systematically generate topology sensitive correlation candidates for future pp, pPb and PbPb studies.

CMS Run2016G. The target is the source dataset; it is not a label for collectivity.

Equal activity can hide different spatial structure

Diffuse map

10	10	10
10	20	10
10	10	10

Total cell activity: 100

Concentrated map

0	0	10
0	60	10
0	10	10

Total cell activity: 100

A single total cannot distinguish these two maps. Regional observables can.

Illustrative maps in arbitrary units. These are teaching examples, not CMS events.

Motivation

CMS Open Data samples

- **ZeroBias Run2016G UL2016 v1**: enhanced NanoAOD with particle-flow candidates.
- **JetHT Run2016G UL2016 v2**: enriched in energetic jets and high-activity topologies.
- Candidate inputs: p_T, η and ϕ.

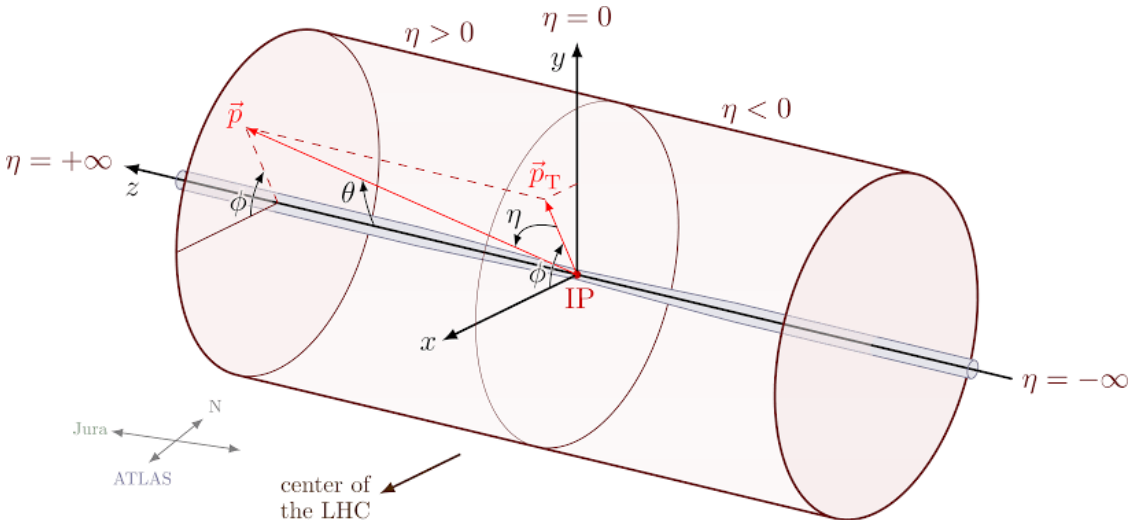
The class label identifies the **source dataset**. To reduce trivial discrimination from the global scale, the samples are matched in

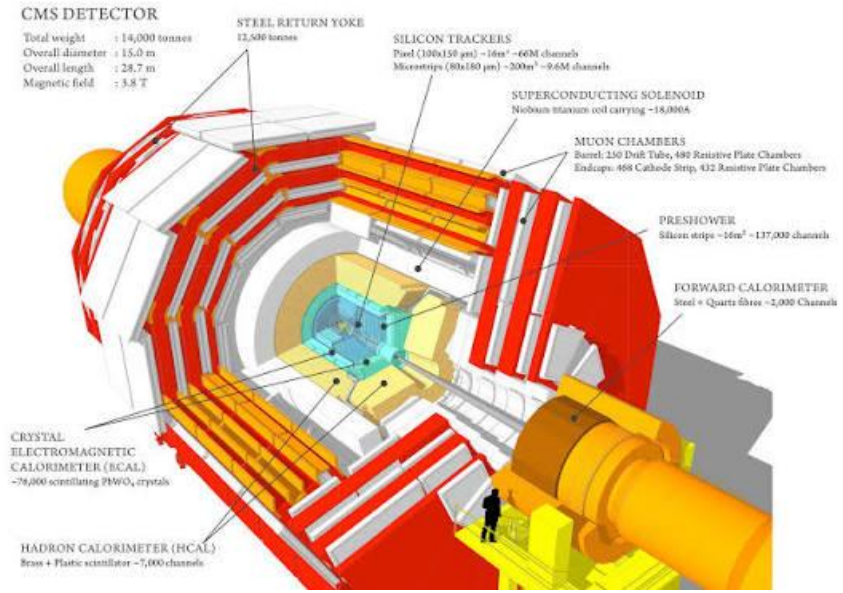
$$6.8521 \lesssim \log H_T \lesssim 6.9668,$$

giving 440 ZeroBias and 440 JetHT events. The common feature comparison uses the 700 events with all circuit observables available.

CONTROL PRINCIPLE

Identical events, train/test partitions and downstream classifiers are used for every feature family.





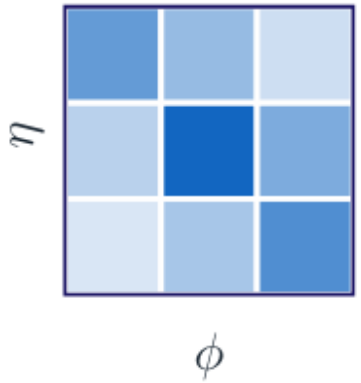
CMS Run2016G. The target is the source dataset; it is not a label for collectivity.

The transverse-momentum density in spatial bin B_{ab} is

$$\rho_{ab} = \sum_{k \in \text{event}} p_{T,k}\, \mathbf{1}[(\eta_k, \phi_k) \in B_{ab}] \,.$$

Azimuth is treated periodically. The event is normalised before angle encoding:

$$x_{ab} = \frac{\rho_{ab}}{\Sigma_{c,d}\, \rho_{cd} + \varepsilon}, \qquad \theta_i = \pi\, \mathrm{clip}(x_i, 0, 1).$$



3×3 map
nine spatial regions
one region per qubit

CMS Run2016G. The target is the source dataset; it is not a label for collectivity.

A controlled topology benchmark with CMS

ZeroBias

Collision events recorded with
a minimally selective trigger.

JetHT

Events selected by energetic jets
or large transverse activity.

$$6.852 \lesssim \log H_T \lesssim 6.967$$

Same activity window

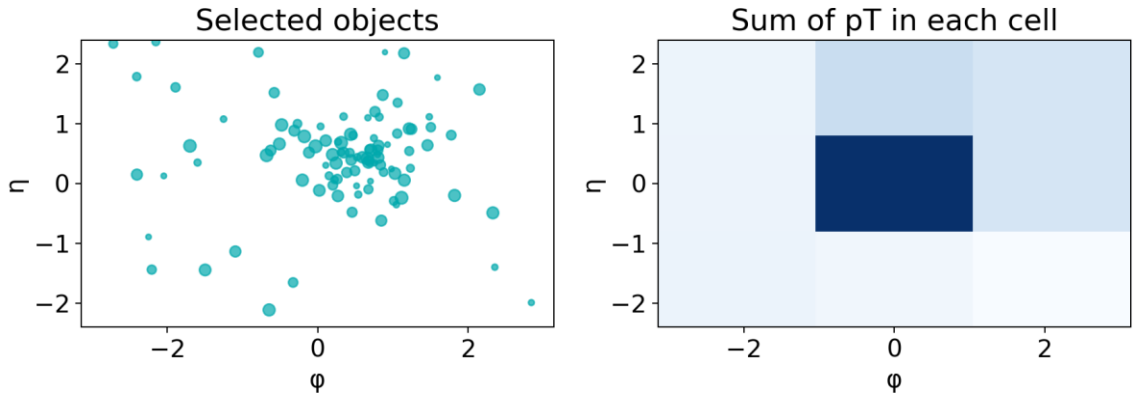
440 + 440 events

Balanced poster sample

> **Can spatial observables distinguish these samples beyond global activity?**

CMS Run2016G. The target is the source dataset; it is not a label for collectivity.

One angular map for every representation



$$\rho_i = \sum_{a \in B_i} p_{T,a}$$

Sum transverse momentum inside angular region B_i.

$$x_i = \frac{\rho_i}{\sum_{j=1}^{9} \rho_j + \varepsilon}, \qquad \theta_i = \pi x_i$$

Normalize the nine cells.
Convert each value to an angle.

The 3 × 3 map describes shape; global activity remains a separate control.

Schematic event. η is pseudorapidity; φ is periodic azimuth; ε regularizes the denominator.

Regional activity determines a fixed quantum state

One cell sets one rotation

$$R_y(\theta_i)|0\rangle = \cos\frac{\theta_i}{2}|0\rangle + \sin\frac{\theta_i}{2}|1\rangle$$

Nine rotations, followed by fixed entangling gates

$$|\psi(\mathbf{x})\rangle = U_{\text{ent}}\left[\bigotimes_{i=1}^{9} R_y(\theta_i)\right]|0\rangle^{\otimes 9}$$

> **The data change the angles. The circuit layout stays fixed.**

Uent denotes the entangling circuit. The poster reports nine qubits and layouts A/B.

Measurements turn the state into numerical features

$$z_i = \langle \psi | Z_i | \psi \rangle$$

Local expectation
Response of one measured qubit

$$z_{ij} = \langle \psi | Z_i Z_j | \psi \rangle$$

Pair expectation
Agreement of two Z measurements

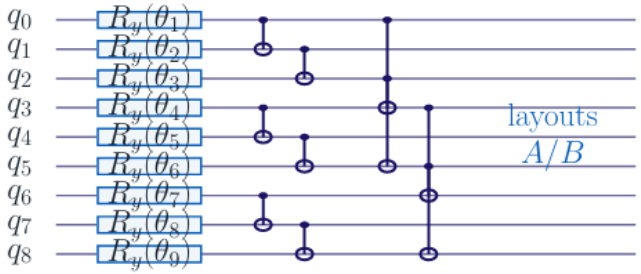
$$z_{ij} = p_{00} + p_{11} - p_{01} - p_{10}$$

Repeat this readout for an indexed set of operators to build the feature vector.

pab is the joint probability of outcomes a,b ∈ {0,1}. An expectation lies between −1 and +1.

Each normalised bin is encoded through a single-qubit rotation. Two fixed entangling layouts probe neighbouring and selected longer-range relations:

$$|\psi(\boldsymbol{x})\rangle = U_{\text{ent}}^{(A/B)} \prod_{i=1}^{9} R_y^{(i)}(\theta_i)|0\rangle^{\otimes 9}.$$



Measured observables

$$z_i = \langle Z_i \rangle, \qquad c_{ij} = \langle Z_i Z_j \rangle,$$
$$C_{ij} = c_{ij} - z_i z_j.$$

Local expectations, pairwise ZZ correlations and topology-dependent combinations form a 58-dimensional feature vector. **Ideal state-vector simulation** is used deliberately: at nine qubits it provides exact observables without shot noise, queueing or hardware latency.

CMS Run2016G. The target is the source dataset; i

The classical comparison must include spatial structure

Family	Information supplied	count
Standard S	HT and Global moments scale, width and activity	9
Spatial contrasts C	Regional differences and direct two-point correlations	11
Circuit Q	Pauli expectations and regional combinations	58

The relevant test is whether Q adds information after S and C.

Counts follow the poster. The nine-qubit dictionary requires the original CMS implementation.

The same classifier is evaluated on held-out events. Preprocessing is fitted on the training set only, and uncertainty is obtained with paired bootstrap resampling.
For an added family F and baseline B,

$$\Delta\mathrm{AUC}(F \mid B) = \mathrm{AUC}(B \cup F) - \mathrm{AUC}(B).$$

This conditional test distinguishes strong standalone performance from genuinely new information.

DECISION CRITERION

A positive 95% interval excluding zero indicates resolved incremental information within this benchmark.

CMS Run2016G. The target is the source dataset; it is not a label for collectivity.

Added information requires a paired comparison

$$\Delta\mathrm{AUC}(F \mid B) = \mathrm{AUC}(B \cup F) - \mathrm{AUC}(B)$$

B: existing baseline

Same held-out events
Same classifier and tuning budget

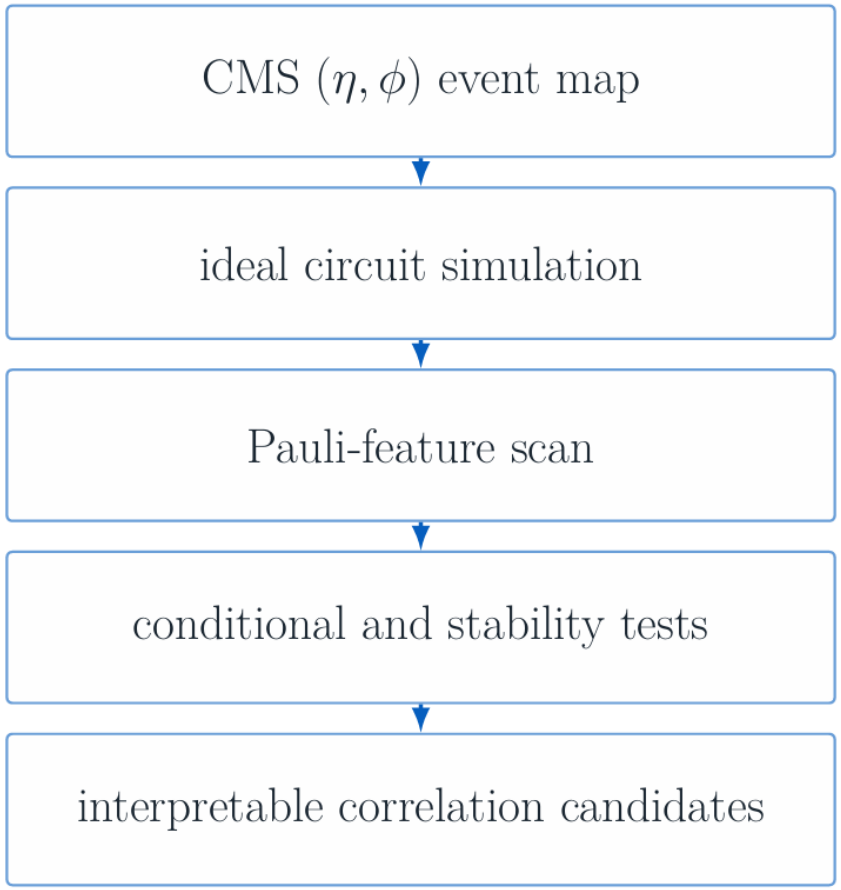
F: added feature family

Preprocessing and selection
fit within the training folds

> **High standalone AUC can coexist with zero incremental information.**

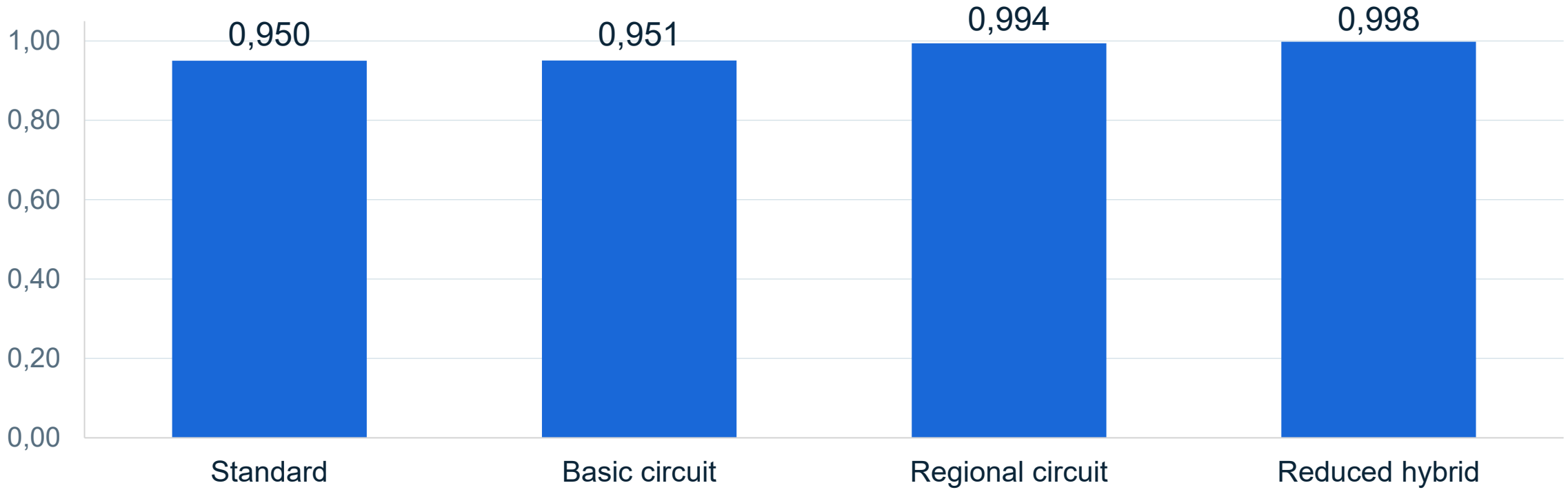
AUC evaluates ranking. Paired uncertainty must respect shared-event or matched-pair groups.

Analysis flow



CMS Run2016G. The target is the source dataset; it is not a label for collectivity.

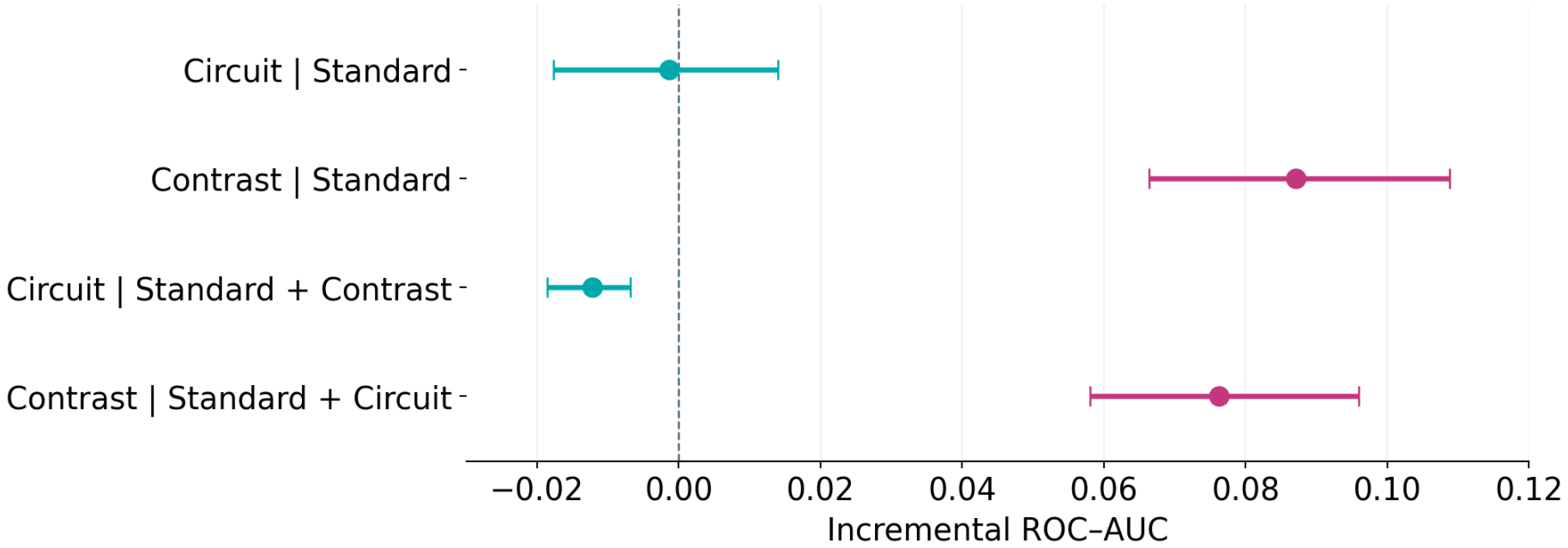
CMS establishes discrimination, not yet independent gain



The basic circuit matches the standard baseline; regional terms motivate further tests.

Preliminary poster point estimates. “Reduced hybrid” includes regional features. CMS uncertainties are unavailable here.

The executed toy test locates the gain in spatial contrasts



Classical contrasts add information. This circuit supplies no resolved extra gain.

700 synthetic events, 350 related pairs, 16 qubits. Original intervals bootstrap fixed OOF scores by event within class.

A two-qubit example reveals what the readout measures

$$|\psi(a,b)\rangle = \mathrm{CNOT}_{1\to 2}\left[R_y(a)\otimes R_y(b)\right]|00\rangle$$

$$\langle Z_1\rangle = \cos a \qquad \langle Z_2\rangle = \cos a \cos b$$

$$\langle Z_1 Z_2\rangle = \cos b$$

A two-qubit operator can depend on only one input angle.

> **Operator weight alone does not determine the spatial information in a feature.**

Exact result for Ry(a), Ry(b), then CNOT with qubit 1 as control. This is an analytic illustration.

Connected correlations carry a different dependence

$$C_{ij} = z_{ij} - z_i z_j$$

Subtract the product of local expectations

$$C_{12} = \sin^2 a \, \cos b$$

> **The definition of the observable changes which combinations of regional inputs survive.**

Same two-qubit example. Quantum-state covariance is not automatically a final-particle cumulant.

For this architecture, every feature has a classical formula

$$U_{\text{ent}}^{\dagger} P U_{\text{ent}} = s_P \bigotimes_i P_i'$$

Propagate the measured Pauli operator through the Clifford entangler.

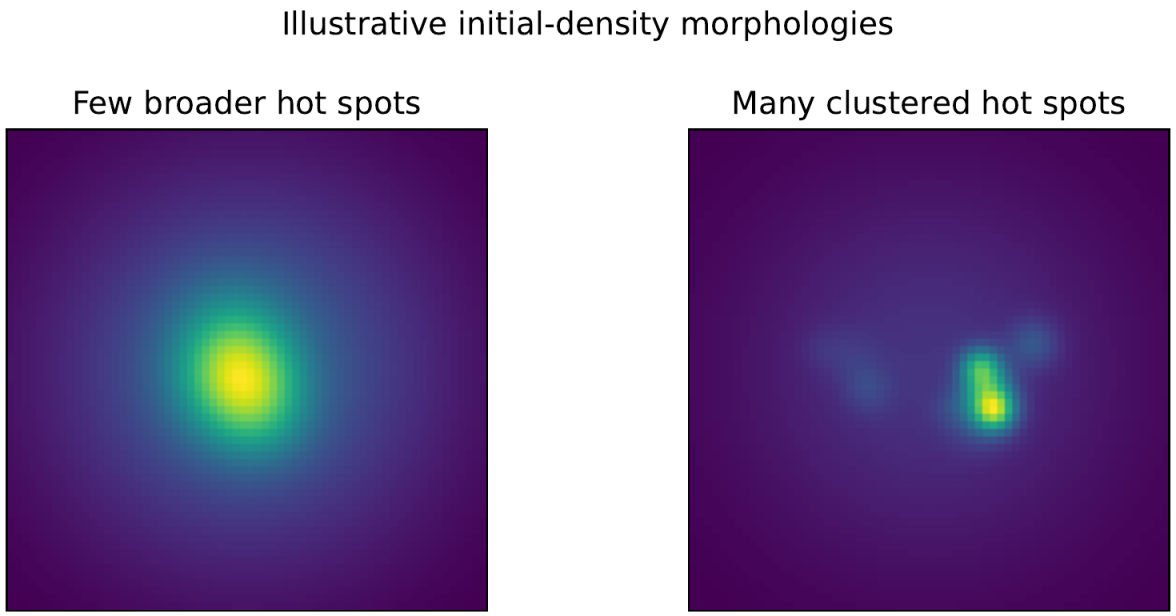
$$f_P(\mathbf{x}) = s_P \prod_i \langle 0 | R_y^{\dagger}(\theta_i) P_i' R_y(\theta_i) | 0 \rangle$$

$$I \mapsto 1, \quad X \mapsto \sin\theta_i, \quad Y \mapsto 0, \quad Z \mapsto \cos\theta_i$$

> **The circuit becomes a structured dictionary of explicit trigonometric functions.**

Applies to one layer of product Ry encoding followed by a Clifford entangler; sP = ±1.

Known hot spots test whether the response is genuinely spatial



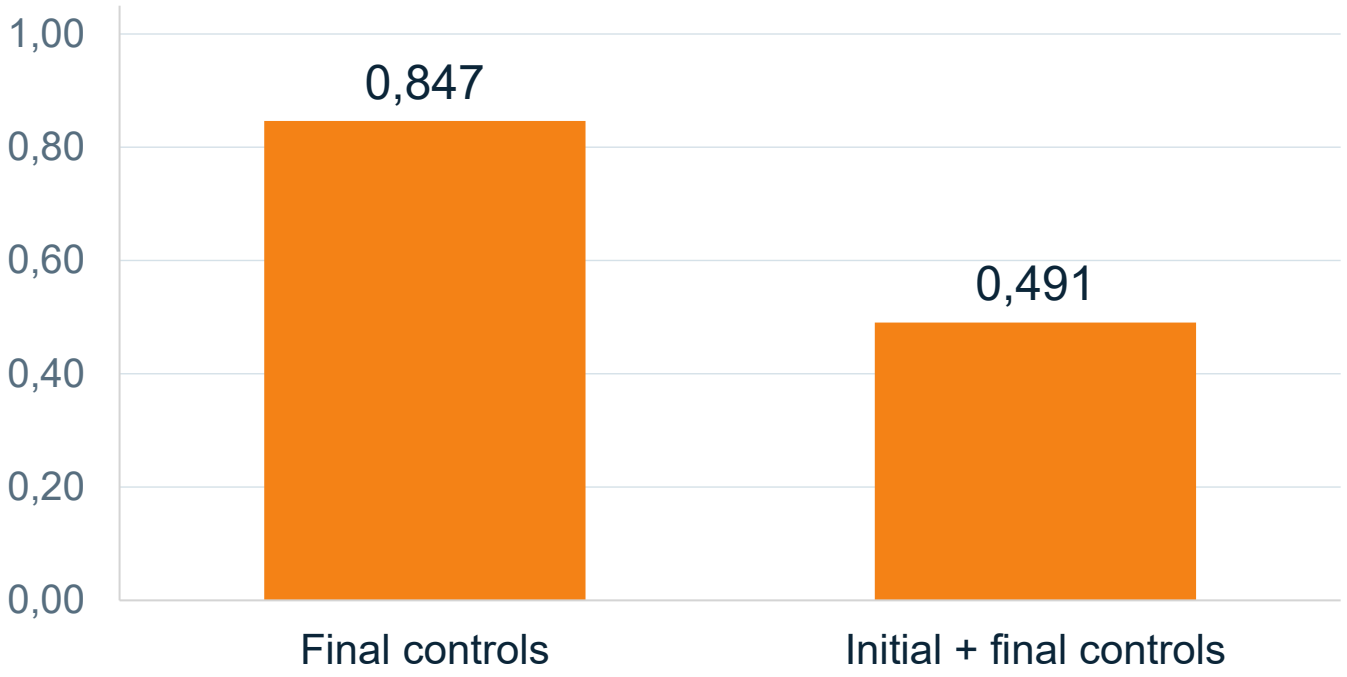
Few broad hot spots versus clustered small hot spots

A transparent response smears and samples the initial density.

Control global morphology before attributing discrimination to local granularity.

Project toy model. Gaussian response and sampling approximate information loss; this is not hydrodynamic evolution.

Stronger controls remove the apparent hot-spot signal



Final controls

Entropy and radial width

Additional initial controls

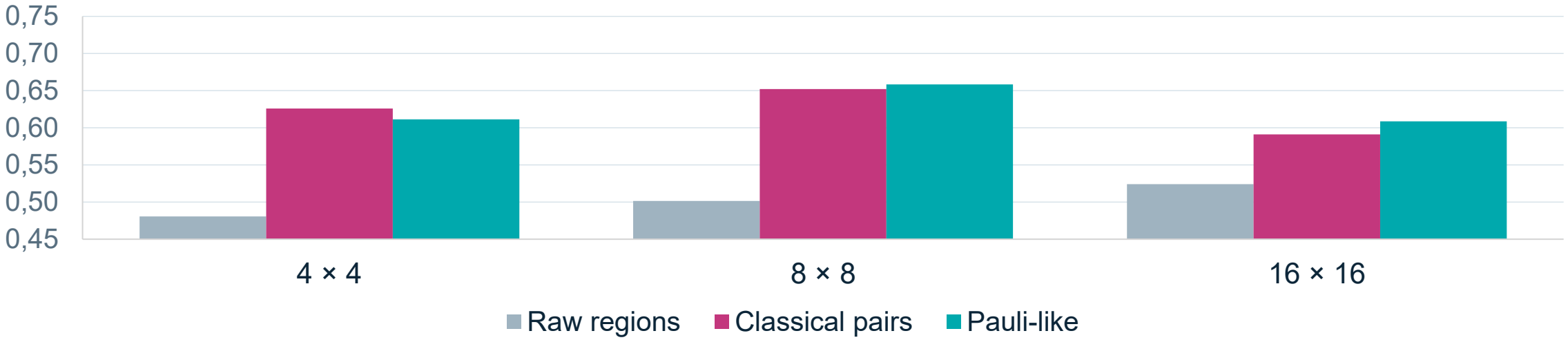
Eccentricities and density variance

Sensitivity is conditional on what the baseline has already fixed.

Draft point estimates, not rerun here. Matched samples differ: N = 3,452 and 1,222.

Changing only correlation length gives a cleaner physics test

$$\ell_{\mathrm{corr}} = 0.25 \text{ fm} \quad \text{vs.} \quad 0.80 \text{ fm}$$



Both pair representations retain sensitivity beyond raw regional intensities.

Draft, zero-smearing subset after global matching. Cropped AUC axis. Grid sizes count regions, not simulated qubits.

Regional relations explain the shared pattern

CMS benchmark	Separation motivates a test of added information.
Executed circuit toy	Direct classical contrasts account for the resolved gain.
Controlled morphology	Spatial sensitivity depends on the controls and the injected structure.

> **The next output should be an interpretable observable, with a validated physical response.**

Different tasks and feature maps. These results support a research direction, not a pooled performance claim.

Proposed contribution: interpretable spatial observables

Candidate

$$\langle Z_1 Z_2 \rangle = \cos b$$

Exact dependence

$$\frac{\partial f}{\partial a} = 0, \qquad \frac{\partial f}{\partial b} = -\sin b$$

Evidence still required

Stable selection and independent gain
Response to a controlled spatial change

> **A useful deliverable is a formula with a defined support, uncertainty and physical test.**

Analytic illustration of the proposed output. This term is not claimed as a discovered CMS observable.

A known spatial change and an independent test

Scientific test	Concrete requirement
Inject spatial structure	Vary correlation length while balancing global controls
Discover candidates	Derive and select formulas using training events only
Freeze and test	Compare with classical pair features on independent events
Test physical survival	Vary smearing, multiplicity and acceptance; then full evolution

Success means stable added information with a reproducible physical interpretation.

Proposed programme. Full initial-condition–MUSIC–UrQMD validation remains to be demonstrated.

CONCLUSIONS

Regional structure matters

Global activity alone does not exhaust event topology.

Pauli observables provide a structured set of functions

Their value must be tested against equivalent classical correlations.

> **The contribution to establish: a small set of interpretable observables with independent physical validation.**

Current evidence does not establish quantum advantage or a collectivity measurement.

The project contains distinct feature maps

Study	Representation	Interpretation
CMS poster	3 × 3 map, 9-qubit fixed circuit	32 features reported
Executed synthetic notebook	4 × 4 map, 16-qubit fixed circuit	58 features with saved outputs
Morphology and length drafts	Direct trigonometric regional functions	No entangled state simulation

$$\underbrace{16}_{Z_i} + \underbrace{16}_{X_i} + \underbrace{24}_{Z_i Z_j} + \underbrace{2}_{\text{parity}} = 58 \text{ features}$$

The bottom equation is the 16-qubit notebook dictionary. Equal feature counts do not imply equal methods.

Evidence and reproducibility record

Source	Status used in this talk
CMS poster	Reported preliminary data benchmark; original predictions needed
Final-state synthetic notebook	Executed saved outputs; limited original bootstrap
Hot-spot and correlation-length manuscripts	Draft results; not re-executed for this talk
Article and production proposals	Research programme and validation route

Equations: compiled LaTeX, with source expressions in image descriptions.