QUANTUM SLICED WASSERSTEIN EMBEDDING: GEOMETRY AND LEARNING FROM HOMODYNE DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Continuous-variable quantum optics underlies photonic computing, bosonic error correction and precision sensing. Information is obtained by homodyne detection of the quadratures of an optical mode, and each shot produces one sample from a quadrature distribution and destroys the state. Extracting what is needed from few shots is the central practical problem. We bring sliced optimal transport, developed in machine learning for exactly this kind of data, to the homodyne record, and compare states through their measured slices with no reconstruction. Our instrument is the quantum sliced Wasserstein embedding, a map from a homodyne record to a Euclidean vector whose distances are the sliced Wasserstein distance between quadrature marginals. That relationship gives the embedding a metric structure in which displacement and squeezing of the source are locally orthogonal. We confirm the geometry on a grid of displaced and squeezed states, whose multidimensional scaling image reproduces the displacement lattice at physical spacing and stretches along squeezing as the metric predicts. We then apply the embedding to cat states, a resource widely used in quantum computing, and regress their displacement and squeezing from shot records, comparing the error to the Cramér–Rao bound. Moment, energy-distance and kernel mean-embedding features of the same records serve as baselines. Under a linear map fitted on a few dozen labeled states the embedding comes within a factor two of the bound on both parameters and is the best feature at every shot budget. The kernel features match it only with a nonlinear regressor and ten times the training states. The embedding is therefore the representation to use when a Gaussian operation on a non-Gaussian state must be read from few shots and few labels.

1 INTRODUCTION

Light carries quantum information in the continuous amplitude and phase of a field mode. These continuous-variable states support photonic computing (Madsen et al., 2022), bosonic error correction (Gottesman et al., 2001; Mirrahimi et al., 2014) and precision sensing (The LIGO Scientific Collaboration, 2011; Weedbrook et al., 2012). Homodyne detection probes such a state by choosing a phase and measuring the corresponding field quadrature. Each shot consumes one copy of the state and returns one real number. Tomography reconstructs the full state from these records (Vogel & Risken, 1989; Lvovsky & Raymer, 2009), but many tasks need only a few parameters, such as how far a source has drifted or how strongly it is squeezed. This motivates a learning problem: how should a homodyne record be represented so that a simple predictor can recover physical parameters from few shots and few labeled states?

Interestingly, the measurement already supplies the structure used by sliced optimal transport commonly used in machine learning for a wide variety of tasks from generative modeling to representation learning. A homodyne record contains samples from a one-dimensional distribution at each measured phase (Figure 1c). Sliced transport compares distributions through their one-dimensional projections (Rabin et al., 2011; Bonneel et al., 2015; Kolouri et al., 2019); in homodyne detection, the experiment performs these projections. Although the Wigner function can be negative and is therefore unsuitable as a probability measure for ordinary optimal transport, its measured marginals are nonnegative. Comparing them directly gives a transport distance without reconstructing the state.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

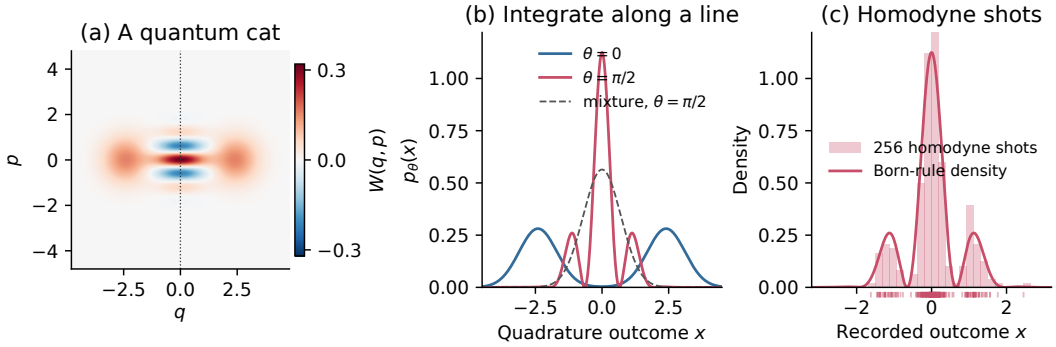


Figure 1: **From a cat state to a homodyne record.** (a) An even cat with real amplitude $\alpha = 1.7$ has two positive lobes and signed interference fringes. (b) Its projections are nonnegative probability distributions. The dashed curve is the momentum distribution of the incoherent mixture: projecting does not erase all evidence of coherence. (c) A simulated homodyne record consists of samples from one such distribution, here 256 independent shots at $\theta = \pi/2$. A "slice" means a line integral, not a pointwise cross-section of the image.

We construct the *quantum sliced Wasserstein embedding* (QSWE) by sorting the shots at each phase, averaging them into equal-mass bins, and stacking the bin means with a common normalization. At full resolution, Euclidean distances between these vectors equal the empirical sliced Wasserstein distance between the measured quadrature distributions. Fewer bins give a fixed-size representation by projecting the quantile functions onto a coarser grid. This applies the classical cumulative distribution transform and its Radon extension (Kolouri et al., 2016; Park et al., 2018; Shifat-E-Rabbi et al., 2021) to homodyne data. The resulting vector supports comparison, visualization and regression, without a kernel bandwidth.

Our analysis connects these coordinates to the physical controls of the source. A displacement shifts every quantile at a given phase by the same amount, so it acts affinely on the embedding for any state. The distance between displaced copies is exactly proportional to their physical separation, and quantile compression preserves this response. For displaced squeezed Gaussian states, we derive a local metric in which displacement and squeezing are orthogonal, while the length of a squeezing step depends on the existing squeezing. Finite-shot bounds quantify the error in estimating the distance from measured records. Experiments on displaced and squeezed state grids then test whether this geometry remains visible at finite measurement budgets.

The learning experiments test whether these coordinates also improve estimation precision. Moments already respond linearly to displacement, so the affine law alone does not establish an advantage. We compare QSWE with moment, energy-distance and kernel mean-embedding features of the same records, using both linear and nonlinear regressors. On simulated cat states with unknown shape and nuisance parameters, ridge regression with 48 training states and 24 validation states reaches approximately 1.6 and 1.9 times the square root of the nuisance-adjusted classical Cramér–Rao bound for displacement and squeezing, respectively, at 3,072 copies per state. QSWE gives the lowest errors among the tested feature families under this linear regressor; nonlinear kernel readouts narrow the gap as training data increase. These results support the embedding as a representation for learning physical controls when shots and labeled states are limited.

Position relative to prior work. Transport distances have been applied to optical tomograms and photon statistics (Paul et al., 2025a;b; 2026). Other quantum transport constructions include the Husimi-based Monge metric, which is Euclidean on coherent states (Życzkowski & Słomczyński, 1998), channel-based transport (De Palma & Trevisan, 2021) and quantum $\mathcal{W}_1$ on qudits (De Palma et al., 2021). Building on the established quantile transforms, we characterize the displacement and squeezing geometry of homodyne records and the effects of compression and finite sampling. The learning contribution is a controlled comparison of how readily these physical parameters can be recovered from different representations, with shared measurement budgets and a classical Cramér–Rao benchmark for the raw record.

2 QUANTUM STATES AND HOMODYNE MEASUREMENT

A single optical mode behaves as a quantum harmonic oscillator. Its position and momentum quadratures, $\hat{q}$ and $\hat{p}$, describe field components at phases separated by $\pi/2$. We use dimensionless units with $[\hat{q}, \hat{p}] = i$ and write (q, p) for the corresponding phase-space coordinates. A state ρ is represented on this plane by its Wigner function $W(q, p)$, a real function of unit mass whose marginals give the measurement distributions of $\hat{q}$ and $\hat{p}$, but which can be negative and is therefore not generally a joint probability density (Wigner, 1932). A coherent state is a Gaussian of covariance $I_2/2$ centred at a point d. Two operations move Gaussians without changing their shape: a displacement $D(d)$ shifts the centre by $d \in \mathbb{R}^2$, and a squeezing $S(\zeta)$ narrows q by $e^{-\zeta}$ and widens p by e^{ζ}. States built from the vacuum by these two are the Gaussian states. A cat state is the standard non-Gaussian example. Its Wigner function has two Gaussian lobes at $\pm d$ and, between them, a fringe that dips negative, as in Figure 1.

The two quadratures cannot be measured simultaneously with arbitrary precision. A homodyne detector at phase θ instead measures $\hat{X}_\theta = \hat{q}\cos\theta + \hat{p}\sin\theta$ and returns a real outcome x, whose distribution P_θ has density

$$p_\theta(x) = \int_{\mathbb{R}} W(x\cos\theta - y\sin\theta,\ x\sin\theta + y\cos\theta)\ \mathrm{d}y, \tag{1}$$

the integral of W along lines perpendicular to $u_\theta = (\cos\theta, \sin\theta)$. This Radon projection is nonnegative even where W is not, and the uncertainty principle is what forces the data to arrive as slices. A displacement shifts every projection by $u_\theta^\top d$ and a squeezing rescales each one. One shot consumes one copy of the state, so a record of n shots at phase θ is a sample of size n from P_θ.

The projections determine the state. Eq. (1) is the Radon transform of W, which is invertible, so knowing every projection fixes W (Vogel & Risken, 1989; Lvovsky & Raymer, 2009). Recovering the state from the projections is tomography. What follows compares states through the projections themselves.

3 QUANTUM SLICED WASSERSTEIN EMBEDDING

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, the space of Borel probability measures on $\mathbb{R}$ with finite second moments. For $\eta \in \mathcal{P}_2(\mathbb{R})$, write $F_\eta(x) = \eta((-\infty, x])$ for its cumulative distribution function and $F_\eta^{-1}(t) = \inf\{x \in \mathbb{R} : F_\eta(x) \geq t\}$, $t \in (0, 1)$, for its generalized quantile function. In one dimension, the squared 2-Wasserstein distance admits the explicit formula (Peyré & Cuturi, 2019)

$$\mathcal{W}_2^2(\mu, \nu) = \int_0^1 \left|F_\mu^{-1}(t) - F_\nu^{-1}(t)\right|^2 \mathrm{d}t. \tag{2}$$

For two empirical measures with n equally weighted shots each, this integral is $n^{-1}\sum_{i=1}^n |x_{(i)} - y_{(i)}|^2$, where $x_{(i)}$ and $y_{(i)}$ are the ordered shots. Thus it is the squared Euclidean distance between the sorted records divided by n. Sliced Wasserstein distances average this one-dimensional cost over projections of a distribution in $\mathbb{R}^d$ onto the directions of the unit sphere S^{d-1} (Bonneel et al., 2015; Kolouri et al., 2019). A homodyne detector performs the projection itself. Setting the phase θ selects the direction vector $u_\theta = (\cos\theta, \sin\theta)$, and the detector returns samples from the quadrature distribution P_ρ^θ of state ρ. We therefore define the distance between two states as the phase average of the one-dimensional cost between their quadrature distributions,

$$\mathrm{QSW}_2(\rho, \sigma) = \left[\frac{1}{\pi}\int_0^\pi \mathcal{W}_2^2(P_\rho^\theta, P_\sigma^\theta)\,\mathrm{d}\theta\right]^{1/2}. \tag{3}$$

We average over $[0, \pi)$ because a shift of π mirrors both distributions and leaves the cost unchanged. At each phase, the quantile function $F_{\rho,\theta}^{-1}$ is the optimal Monge map transporting the uniform reference measure $\lambda = \mathrm{Unif}(0, 1)$ to P_ρ^θ. Collecting these maps over phases gives the uncentred sliced Wasserstein embedding, or Radon-CDT in quantile coordinates (Kolouri et al., 2016; Park et al., 2018). Its L^2 geometry is exactly the distance in Eq. (3).

Theorem 1 (Metric and sliced Wasserstein embedding)**.** *On states with finite second quadrature moments,* QSW_2 *is a metric. Moreover, the map that sends a state to its quantile functions at every phase,*

$$\Phi(\rho)(\theta,t) = F^{-1}_{\rho,\theta}(t), \qquad (\theta,t) \in [0,\pi) \times (0,1), \tag{4}$$

is an isometric embedding into $L^2([0,\pi) \times (0,1), \pi^{-1}\,\mathrm{d}\theta\,\mathrm{d}t)$*, that is* $\mathrm{QSW}_2(\rho,\sigma) = \|\Phi(\rho) - \Phi(\sigma)\|_{L^2}$ *for all states* ρ, σ*.*

Thus Φ is injective. The assumption of finite second quadrature moments is the same as finite mean energy, since the phase average of $\mathbb{E}X_\theta^2$ equals the mean photon number plus one half. Every state an experiment can prepare has finite mean energy, so the restriction excludes only density operators that no finite apparatus produces. Appendix A gives the proof and the general order p and phase average.

For a finite record, use the same construction with a discrete reference. Given n shots $x_{\ell i}$ at each of L common phases θ_ℓ, let $\widehat{P}_\ell = n^{-1}\sum_{i=1}^n \delta_{x_{\ell i}}$ and choose $\lambda_r = r^{-1}\sum_{j=1}^r \delta_{t_j}$ with $t_j = (j-\frac{1}{2})/r$ and $1 \le r \le n$. Let Γ^ℓ be the one-dimensional optimal transport plan from λ_r to $\widehat{P}_\ell$, with row sums $1/r$ and column sums $1/n$. Its barycentric projection gives the quantum sliced Wasserstein embedding (QSWE), following the transport-based embedding construction of Liu et al. (2025, Section 3.6):

$$b_{\ell j} = r\sum_{i=1}^n \Gamma^\ell_{ji} x_{\ell i}, \qquad \Phi_{L,r}(\rho) = \frac{(b_{\ell j})_{\ell,j}}{\sqrt{Lr}} \in \mathbb{R}^{Lr}. \tag{5}$$

Each coordinate is the mean destination of one reference particle. When $r = n$, the monotone plan matches particles by rank, so the embedding is the normalized sorted record and its Euclidean distances equal empirical QSW_2. When $r < n$, reference particles can split their mass; the barycentric projection averages their destinations. In one dimension this gives $b_{\ell j} = r\int_{(j-1)/r}^{j/r} F^{-1}_{\widehat{P}_\ell}(t)\,\mathrm{d}t$, as derived in Appendix A. Thus equal-mass quantile bins follow from the uniform reference weights. If r divides n, each coordinate is the mean of n/r consecutive sorted shots. Proposition 3 describes what this finite reference preserves.

4 Single-mode geometry: displacement and squeezing

If two embeddings are one unit apart, what has changed in the state? Trace and Bures distances quantify distinguishability and are bounded. Transport instead measures how far the quadrature distributions have moved. For displacement and Gaussian squeezing, this geometry can be computed exactly. Section 6 tests how it survives finite data.

Proposition 1 (Exact displacement geometry)**.** *Let* $D(d)$ *displace the quadrature mean by* $d \in \mathbb{R}^2$*, and let* $D(d)^\dagger$ *be its adjoint. For any state* ρ*, the displaced state is* $\rho_d = D(d)\rho D(d)^\dagger$*. For any state with finite second quadrature moments,*

$$\mathrm{QSW}_2(\rho_d,\rho_e) = \tfrac{1}{\sqrt{2}}\|d-e\|_2. \tag{6}$$

This holds for Gaussian states, cats and mixed states alike. A displacement shifts every quantile at phase θ by $u_\theta^\top d$, so the embedding is affine in d and displacement differences admit an exact linear readout. Appendix A gives the general order p, with constant $c_p^p = \pi^{-1}\int_0^\pi |\cos\theta|^p\,\mathrm{d}\theta$.

For coherent states the squared fidelity is $F(\rho_d,\rho_e) = e^{-\|d-e\|^2/2}$. The trace distance $\sqrt{1-F}$ and Bures distance $\sqrt{2-2\sqrt{F}}$ approach 1 and $\sqrt{2}$, respectively, as displacement grows. Their sensitivity to further displacement therefore decays, while Eq. (6) keeps growing linearly.

Proposition 2 (Translation plus squeezing)**.** *For Gaussian states with means* m, m' *and covariances* V, V'*, the quadrature distribution at phase* θ *is* $\mathcal{N}(u_\theta^\top m, u_\theta^\top V u_\theta)$*, and*

$$\mathrm{QSW}_2^2((m,V),(m',V')) = \frac{1}{2}\|m-m'\|^2 + \frac{1}{\pi}\int_0^\pi \left(\sqrt{u_\theta^\top V u_\theta} - \sqrt{u_\theta^\top V' u_\theta}\right)^2 \mathrm{d}\theta. \tag{7}$$

For the displaced squeezed vacuum, $m = (a,0)$ *and* $V(\zeta) = \frac{1}{2}\,\mathrm{diag}(e^{-2\zeta}, e^{2\zeta})$*, the induced local line element is*

$$\mathrm{d}s^2 = \frac{1}{2}\,\mathrm{d}a^2 + \left[\frac{\cosh(2\zeta)}{2} - \frac{1}{4\cosh^2\zeta}\right]\mathrm{d}\zeta^2. \tag{8}$$

Eq. (7) separates changes of mean from changes of width. Eq. (8) makes their local geometry explicit: displacement and squeezing are orthogonal tangent directions. The displacement coefficient is constant, whereas the squeezing coefficient increases with $|\zeta|$, from $1/4$ at $\zeta = 0$. Thus equal small increments in squeezing have different lengths depending on the starting state.

The separation of mean and shape extends beyond Gaussian states. At each phase, subtracting the mean leaves a centred quantile function whose integral is zero. Consequently, every quantile difference splits orthogonally into its mean and centred parts (Appendix A). Squeezing about the mean changes only the centred part, so its tangent, when it exists in L^2, is orthogonal to displacement. For non-Gaussian states this shape term depends on the full quadrature distributions, not just their widths. Squeezing both rescales and relabels the slices; a local linearization requires L^2 differentiability, and the accuracy of a linear readout over a finite range must be assessed for the state family in question.

Proposition 3 (Reference size)**.** *Fix L phases and two records of n shots per phase. Under the identification with piecewise constant quantile functions, barycentric compression is an orthogonal projection, so for every $1 \le r \le n$, $\|\Phi_{L,r}(\rho) - \Phi_{L,r}(\sigma)\| \le \|\Phi_{L,n}(\rho) - \Phi_{L,n}(\sigma)\|$. Equality holds at $r = n$, and distances are nondecreasing along nested bin refinements. The population counterpart, obtained by using the exact quadrature laws, preserves displacement differences at every r. It satisfies Eq.* (6) *when $L^{-1}\sum_\ell u_{\theta_\ell} u_{\theta_\ell}^\top = I_2/2$, in particular for $L \ge 2$ equally spaced phases.*

Projection retains bin means and discards within-bin variation. It preserves the per-phase mean at every r, while it can attenuate width and shape differences. At $r = 1$ only the means remain, so centred shape differences disappear entirely. The contraction is exact for each pair of records; the displacement identity concerns population laws, since independently sampled records also contain shot noise.

5 Estimation from homodyne shots

We now quantify the error from estimating the population geometry with n shots at each of L common phases. At $r = n$, the empirical distance $\widehat{\mathrm{QSW}}_{2,L,n}$ is the Euclidean distance between the sorted records normalized by $1/\sqrt{Ln}$, computable in $O(Ln \log n)$ time. Unequal record sizes can be compared exactly through Eq. (2), or compressed using any common reference size $r \le \min(n, m)$; divisibility is unnecessary. Compression can reduce the distance, whereas truncating the longer sorted record does not implement the quantile integral. Acquisition costs $B = Ln$ copies per state. Each shot measures one phase on one copy; the resulting classical record can be reused across embeddings and learning tasks.

There are two sources of error. With exact laws but only L phases, the distance $\mathrm{QSW}_{2,L} = [L^{-1}\sum_\ell \mathcal{W}_2^2(P_\rho^{\theta_\ell}, P_\sigma^{\theta_\ell})]^{1/2}$ is generally only a pseudometric: agreement at the measured phases need not determine the state. Finite shots introduce a second error, so two independent records of the same state generally have a nonzero distance. This noise must be accounted for when interpreting small distances; it does not imply a universal upward bias of the unsquared distance or a fixed threshold below which states cannot be distinguished.

Proposition 4 (Conservative finite-shot rate)**.** *Let $\widehat{\mathrm{QSW}}_{2,L,n} = \|\Phi_{L,n}(\rho) - \Phi_{L,n}(\sigma)\|$. Suppose $q > 4$ and $\sup_\theta \int |x|^q P_\xi^\theta(\,\mathrm{d}x) \le K_q$ for each $\xi \in \{\rho, \sigma\}$. For phases drawn independently and uniformly from $[0, \pi)$, shared by both states, and independent shots conditional on these phases, there is a constant C_q, depending only on q, such that*

$$\mathbb{E}\big|\widehat{\mathrm{QSW}}_{2,L,n} - \mathrm{QSW}_2(\rho, \sigma)\big| \le C_q K_q^{1/q}\big(n^{-1/4} + L^{-1/2}\big). \tag{9}$$

The shot term transfers a one-dimensional empirical transport bound (Fournier & Guillin, 2015; Nadjahi et al., 2020); the phase term bounds Monte Carlo integration of the population cost. These are sufficient bounds from moments alone, with no root-n claim for the shot error. For fixed phases, as in our experiments, the same shot bound holds relative to $\mathrm{QSW}_{2,L}$. Its discrepancy from QSW_2 is a separate deterministic quadrature error. Appendix B proves the bound, extends the shot analysis to order p, and gives phase-grid guarantees under regularity assumptions and the resulting cost in copies.

6 EXPERIMENTS: GEOMETRY AND REGRESSION

We designed two experiments that answer different questions. The first asks whether the geometry of Section 4 survives finite data. The second tests the embedding in a downstream learning task. Both experiments compare QSWE with the same baselines, each a vector built from the same record phase by phase. Gaussian moments are the mean and standard deviation of each quadrature distribution. Hermite moments extend them to the first six moments. The energy-CDF features tabulate empirical distribution functions on a fixed grid. The RBF mean embedding uses weighted Fourier features at bandwidth h. Their Euclidean distances numerically approximate the energy and RBF-MMD distances of Figure 2; Appendix D.2 gives the discretizations. Bures distances are computed on the exact states and serve as references in the geometry experiment only, since they need the state.

6.1 THE GEOMETRY OF DISPLACEMENT AND SQUEEZING UNDER FINITE DATA

Propositions 1 and 2 prove how translation and squeezing behave in this embedding. However, as discussed in Section 5, real homodyne experiments do not have access to the raw distribution, and the embedding has to be computed from samples. We generate two state grids from combinations of translation and squeezing. The first is a 7×7 grid of coherent states with $(q, p) \in [-2, 2]^2$, where q and p are the two quadratures. The second is the displaced squeezed vacuum of Proposition 2, squeezed by $\zeta \in [-0.9, 0.9]$ and displaced along the squeezing axis by $q \in [-2, 2]$, with $p = 0$ throughout. For every state we collect 12 equally spaced phases with 64 shots per phase. We compute all pairwise distances and apply classical MDS to each distance matrix. Vertices are states and edges join neighbours on the parameter grid. Two numbers summarize how each MDS map compares to the true displacement and squeezing of the states. First, we slide, rotate, flip and uniformly rescale the MDS positions to line them up with the true ones as well as possible. Grid error is the mismatch left after that, as a fraction of the size of the grid. It grows when rows bend, spacing varies or the grid folds. Second, stress compares the pairwise distances between states with the pairwise distances between their points in the plane, again as a fraction of the total. Estimator conventions and the complete tables are in Appendices C and C.1.

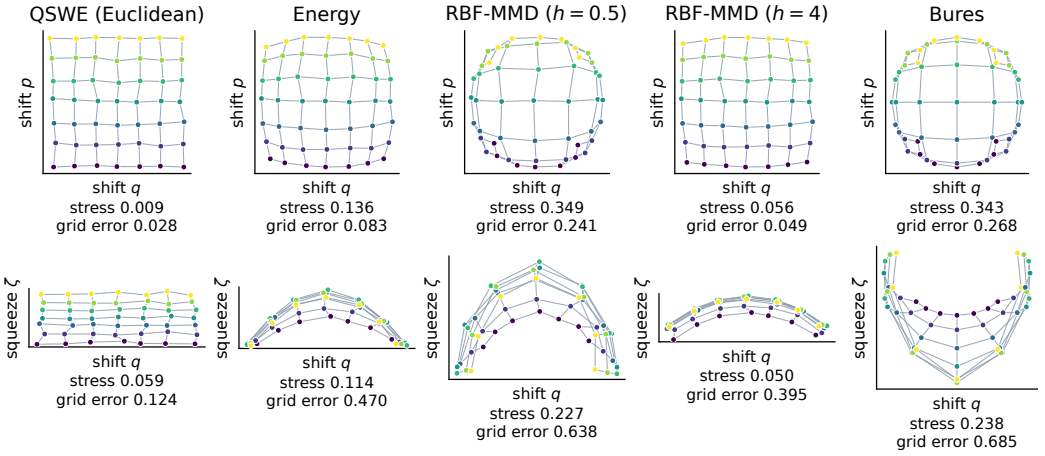


Figure 2: **Five distances on the same two grids of states.** Top row, a 7×7 grid of coherent states with $(q, p) \in [-2, 2]^2$. Bottom row, a 7×7 grid of squeezed states with squeezing $\zeta \in [-0.9, 0.9]$ and displacement $q \in [-2, 2]$ along the squeezing axis, with $p = 0$. Each panel is the classical MDS embedding of one distance matrix, aligned to the parameter grid so that the horizontal axis is the shift along q and the vertical axis is the shift along p (top) or the squeezing ζ (bottom). Lines join neighbours on the grid and point colour follows the vertical parameter. Displayed values are for the first seed; the text reports three repeats.

Displacement keeps its physical length and squeezing keeps the grid. Proposition 1 says that QSW_2 reports a move as a Euclidean distance with a fixed scale, so a square grid of displacements should come back as a square grid. With finite shots the measured grid error is 0.027 ± 0.002 over

three repeats. On the shift and squeeze grid, Eq. (8) says that a displacement and a squeezing move the embedding in orthogonal directions, so the rows and columns stay straight and perpendicular; embedding is not linear in squeezing, so the vertical spacing of the rows is not uniform. The grid error is 0.117 ± 0.007 against 0.101 for the exact states, so for squeezing the error is partly a property of the metric. Figure 2 shows both grids under the other distances. On the shift and squeeze grid every other distance folds the squeezing rows onto one another, so a step in squeezing and a step in displacement can no longer be told apart. Energy keeps the shift grid but shrinks its outer rows, since its response to a shift slows from linear to square root once the shift exceeds the width of the state. The kernel distance is the closest competitor, but only once its bandwidth is chosen. At $h = 4$ it recovers the displacement grid, with grid error 0.049, while at $h = 0.5$ it rounds the same grid into a disc because a kernel stops responding once two states are further apart than h. No bandwidth recovers the squeezing grid, where its best grid error is 0.393 against 0.117 ± 0.007 for QSW_2. The bandwidth therefore has to be matched to the scale of the differences being compared. Appendix C.2 gives the response laws behind this and the sweep over bandwidths.

Shots and reference size. A second set of shot records tests the effects of reference size. Seven even cats, squeezed by ζ from -0.9 to 0.9 and not displaced, give 21 pairwise distances, and Figure 3 compares each with its ideal, the twelve-phase QSW_2 of the exact quantile functions, as a fraction of the mean ideal distance. On the left the shots per phase vary at full reference size. The error falls from 28% at 16 shots to 0.5% at 4,096, faster than the $n^{-1/4}$ of Proposition 4. On the right the reference size varies at three budgets. Each curve ends at full size, and each dips below that value on the way, since a bin average smooths the empirical quantile function where the cat's density is small and the shot noise is largest. At $r = 1$ only the per-phase means remain, which carry nothing about the squeezing of a centred cat, and the error is the whole distance.

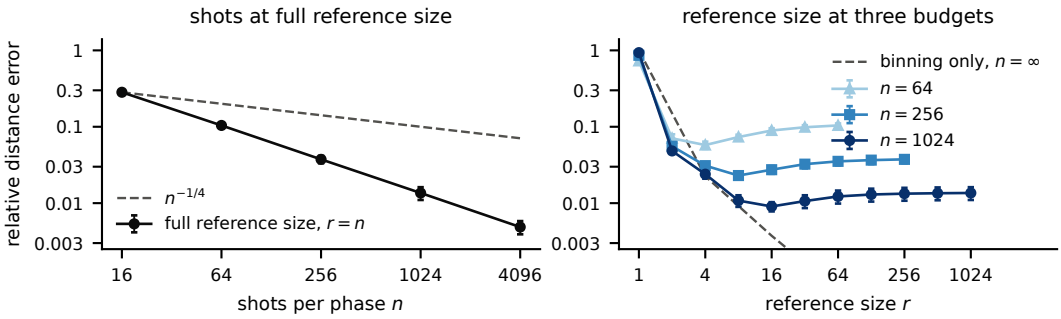


Figure 3: **Relative distance error against the ideal QSW_2 on squeezed cats**, mean over 21 pairs, mean $\pm$ 95% half-width over ten records, 12 phases. Left, shots per phase at full reference size $r = n$, with an $n^{-1/4}$ guide through the point at $n = 16$. Right, reference size at three budgets; each curve ends at $r = n$, and the dashed curve is the binning error with no shot noise.

6.2 Parameter regression on non-Gaussian states

We now ask what the geometry is worth for learning. The test is to regress a physical parameter of a state from its homodyne record and to compare the error with the Cramér–Rao bound of that record, the smallest error any unbiased estimator can reach from the same shots. We use the even cat of Section. 2, a state common in applications, squeezed by $S(\zeta)$ and displaced by $D(q, p)$, and regress q and ζ. The targets are drawn uniformly from a fixed box. The conjugate displacement p, the cat amplitude α and the shape of the cat are also drawn at random for every state and are never given to any method, so they act as nuisance parameters. A record is 12 phases and a fixed number of shots per phase. Each repeat draws a few dozen training states, a validation set and a test set, and every number is averaged over repeats. The exact ranges, budgets and counts are in the caption of Figure 4.

QSWE enters as the finite embedding of reference size $r = 16$ and at the full size $r = n$. Proposition 3 characterizes the quantile projection. Empirically, $r = 16$ matches the full reference on both targets; The comparison features are Gaussian moments, Hermite moments through order six, the energy-CDF features and the RBF mean embedding with its best performing bandwidth chosen on validation. Each feature set is fitted by two regressors. The first is ridge regression with the penalty chosen on

validation error for each target. The second reduces the features to a few principal components and fits a Gaussian process on them (Appendix D.2). Every error is compared with the classical Cramér–Rao bound of the raw record, computed with every parameter of the state free and averaged over the test states of a repeat. We call this the floor, and Appendices D.3 and D.4 give its computation and a likelihood check on these records. Chance is the error of a predictor that always returns the prior mean, for which RMSE equals the prior standard deviation σ. We report RMSE in percent of σ, the fraction of prior variance explained, and RMSE divided by the square root of the floor.

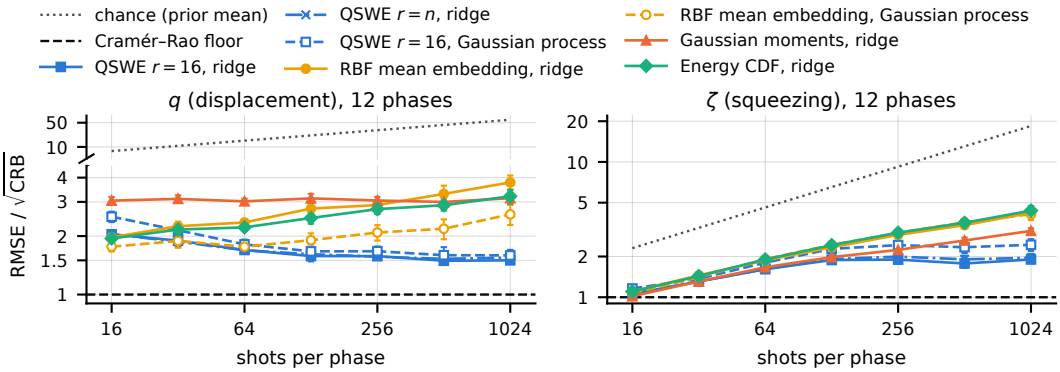


Figure 4: **Relative error against copies per state.** Protocol: $q \sim U[-1, 1]$, $\zeta \sim U[-0.3, 0.3]$, with nuisance $p \sim U[-0.3, 0.3]$, $\alpha \sim U[1.5, 1.9]$; 12 phases with 16 to 1,024 shots each, so 192 to 12,288 copies per state; 48 training, 24 validation and 96 test states per repeat; 20 repeats. RMSE as a fraction of the prior standard deviation on log axes. QSWE is shown at $r = 16$ and at the full size $r = n$; the two coincide within their intervals at every budget. The rows of Table 4 in the appendix are the 3,072-copy points with 30 repeats for ridge regression.

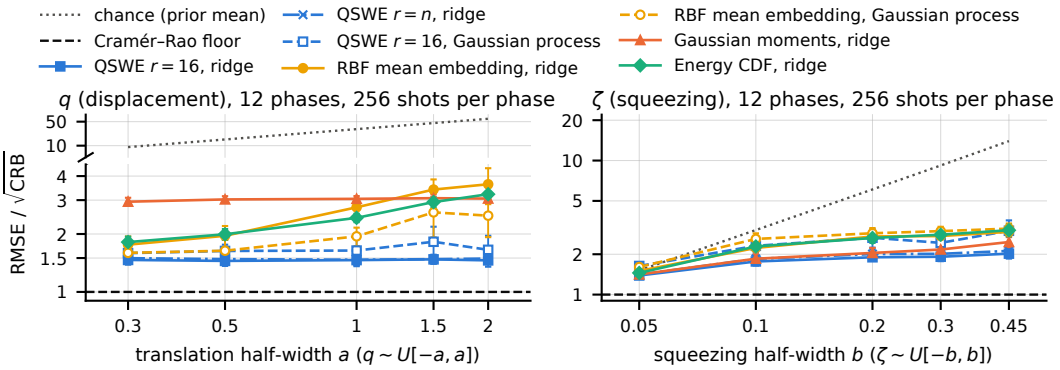


Figure 5: **Error against the size of the transformation**, 3,072 copies, 48 labeled states, 10 repeats per point. Left, $q \sim U[-a, a]$ for five translation half-widths a. Right, $\zeta \sim U[-b, b]$ for five squeezing half-widths b. Everything else as in Figure 4.

Three readings of the results. First, the advantage of QSWE grows with the number of shots. At the smallest budget in Figure 4 shot noise dominates and every feature set sits within a few percent of the others on both targets. As the budget grows, QSWE read by ridge regression runs parallel to the floor, at about 1.6 times it on q and 1.9 times on ζ, while the other feature sets under the same regressor flatten, so that at the largest budget their errors are two to three times QSWE's. Compressing the reference to $r = 16$ costs nothing at any budget. The $r = 16$ and $r = n$ curves coincide within their intervals; this preservation of regression accuracy is empirical, beyond the contraction in Proposition 3. Second, the advantage does not depend on the size of the transformation (Figure 5). Because a displacement shifts the embedding by an exact amount (Proposition 1), QSWE stays at about 1.5 times the floor on q as the translation half-width grows by a factor seven, while the kernel features read linearly fall from 1.8 to 3.6 times the floor. On ζ every feature set degrades as

	q, RMSE / prior s.d.			ζ, RMSE / prior s.d.		
Training states	QSWE (L)	RBF (L)	RBF (N)	QSWE (L)	RBF (L)	RBF (N)
floor	3.3	3.3	3.3	10.8	10.8	10.8
48	5.2 ± 0.2	9.5 ± 0.6	6.9 ± 0.6	20.6 ± 1.4	31.4 ± 1.8	32.4 ± 1.9
160	4.3 ± 0.2	7.0 ± 0.3	4.4 ± 0.2	14.8 ± 0.5	24.4 ± 0.7	19.1 ± 0.7
480	3.9 ± 0.1	5.5 ± 0.2	3.8 ± 0.1	13.4 ± 0.4	22.2 ± 0.6	14.7 ± 0.6

Table 1: **Error against the number of labeled training states**, 3,072 copies, 20 repeats, RMSE as a percentage of the prior standard deviation. QSWE $r = 16$ under ridge regression (L), the RBF mean embedding under ridge regression (L) and the Gaussian process (N). Floor ranges are given in Table 7.

the squeezing half-width grows, and QSWE degrades least. Third, QSWE needs the fewest labels (Table 1). The other feature sets carry the same information about displacement and squeezing, but it is not linear in their coordinates and is harder to extract. Because it is linear in QSWE's coordinates, ridge regression reads it from 48 states. A Gaussian process on the kernel features reaches parity on q once the training set is ten times larger, whereas ridge regression on the same features never catches up. The nonlinear regressor does not help QSWE, and on ζ it is clearly worse, which is the cost of a flexible fit on a feature that is already linear in the target.

7 Discussion and limitations

The experiments confirm the three predictions of Section 4 on finite records. The displacement grid comes back at physical spacing, the squeezing grid comes back straight and stretched, and a linear map reads both parameters of a cat at 1.6 and 1.9 times the Cramér–Rao floor from 48 labeled states. The advantage is one of linear readability from few labels, not of information content, since the kernel features reach parity once a nonlinear regressor and ten times the labels are supplied. Compressing the reference to $r = 16$ gives comparable errors on both parameters in these experiments.

The representation is suitable whenever a displacement or a squeezing is the quantity of interest, and that covers a large part of continuous-variable practice. A force or a field shifts the mean of a quadrature; A drift of the local oscillator phase rotates the state, which the record sees as a change of the measured displacement. A squeezed source is commonly used to push the noise of a sensor below the shot-noise limit (The LIGO Scientific Collaboration, 2011). Appendix E.1 gives a preliminary experiment in which a change of squeezing is detected from the running embedding by a cumulative-sum monitor. For a cat state specifically, explicit model-based reconstruction is the standard and reaches the floor, so the cat is a test bed here rather than the target application. The embedding earns its place when the family is not trusted or has no closed form. Appendix E.2 repeats the regression on a Kerr cat, whose tomograms have no closed form, and the advantage on displacement persists.

A transport distance is Lagrangian. It follows where probability mass goes and how far, which is why a shift and a stretch are read so directly. A change in the shape of a distribution at fixed location and width is an Eulerian difference, a change in the density at each point, and transport is not the natural measure of it. The visibility and fringe phase of the cat are parameters of this kind, and the tested QSWE predictors read them poorly (Table 9). We do not expect the embedding to beat energy or kernel features on such parameters. One way forward is to concatenate the embedding with a feature that is invariant to shift and scale but sensitive to shape. Appendix E.4 reports a first attempt, in which the two blocks show complementary responses.

Several things remain open. Every record here is simulated with an ideal detector. A preliminary study in Appendix E.3 suggests the ordering of the feature sets survives loss and detector noise, but the embedding's response to a detector model is not derived. The rate of Proposition 4 is conservative, and a bound at fixed r would say how the reference size trades shot noise against resolution. Finally, the phases were fixed in advance. Every shot is spent at one phase and cannot be reused, and choosing the phase direction to read a given parameter to reduce the number of shots is an open design problem.

REPRODUCIBILITY STATEMENT

Code accompanies all figures and simulations, with saved configurations, observations or reproducible acquisition seeds, distances, and predictions. Appendices provide proofs and protocols; tests check analytic Born laws, reference transport against an independent solver, and Euclidean distance identities.

AI USE STATEMENT

In this work, we used generative AI tools to assist with the writing of proofs, drafting, implementing methods code, and simulations. We have not used generative AI tools to formulate mathematical claims, provide critical ingredients for proving mathematical claims, design or provide feedback on research methodology or experiments, and interpret results. We have reviewed all AI-assisted work. We reviewed proofs for correctness, and LLM-generated code was verified and tested by the author. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Leonardo Banchi, Samuel L. Braunstein, and Stefano Pirandola. Quantum fidelity for arbitrary Gaussian states. *Physical Review Letters*, 115(26):260501, 2015. doi: 10.1103/PhysRevLett.115.260501. URL https://arxiv.org/abs/1507.01941.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015. doi: 10.1007/s10851-014-0506-3.

Giacomo De Palma and Dario Trevisan. Quantum optimal transport with quantum channels. *Annales Henri Poincaré*, 22:3199–3234, 2021. doi: 10.1007/s00023-021-01042-3. URL https://arxiv.org/abs/1911.00803.

Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd. The quantum Wasserstein distance of order 1. *IEEE Transactions on Information Theory*, 67(10):6627–6643, 2021. doi: 10.1109/TIT.2021.3076442.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3–4):707–738, 2015. doi: 10.1007/s00440-014-0583-7. URL https://arxiv.org/abs/1312.2128.

Daniel Gottesman, Alexei Kitaev, and John Preskill. Encoding a qubit in an oscillator. *Phys. Rev. A*, 64:012310, Jun 2001. doi: 10.1103/PhysRevA.64.012310. URL https://link.aps.org/doi/10.1103/PhysRevA.64.012310.

J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3–4):325–338, 1966. doi: 10.1093/biomet/53.3-4.325.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL https://jmlr.org/papers/v13/gretton12a.html.

Soheil Kolouri, Se Rim Park, and Gustavo K. Rohde. The Radon cumulative distribution transform and its application to image classification. *IEEE Transactions on Image Processing*, 25(3):920–934, 2016. doi: 10.1109/TIP.2015.2509419.

Soheil Kolouri, Kimia Nadjahi, Umut Şimşekli, Roland Badeau, and Gustavo K. Rohde. Generalized sliced Wasserstein distances. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/f0935e4cd5920aa6c7c996a5ee53a70f-Abstract.html.

Xinran Liu, Rocío Martín Díaz, Yikun Bai, Ashkan Shahbazi, Matthew Thorpe, Akram Aldroubi, and Soheil Kolouri. Expected sliced transport plans. In *International Conference on Learning Representations*, 2025.

A. I. Lvovsky and M. G. Raymer. Continuous-variable optical quantum-state tomography. *Reviews of Modern Physics*, 81(1):299–332, 2009. doi: 10.1103/RevModPhys.81.299.

L. S. Madsen, F. Laudenbach, M. Falamarzi Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, and J. Lavoie. Quantum computational advantage with a programmable photonic processor. *Nature*, 606(7912):75–81, 2022. doi: 10.1038/s41586-022-04725-x.

Mazyar Mirrahimi, Zaki Leghtas, Victor V Albert, Steven Touzard, Robert J Schoelkopf, Liang Jiang, and Michel H Devoret. Dynamically protected cat-qubits: A new paradigm for universal quantum computation. *New Journal of Physics*, 16(4):045014, 04 2014. doi: 10.1088/1367-2630/16/4/045014.

Kimia Nadjahi, Alain Durmus, Lénaïc Chizat, Soheil Kolouri, Shahin Shahrampour, and Umut Simsekli. Statistical and topological properties of sliced probability divergences. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 20802–20812. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/eefc9e10ebdc4a2333b42b2dbb8f27b6-Paper.pdf.

Se Rim Park, Soheil Kolouri, Shinjini Kundu, and Gustavo K. Rohde. The cumulative distribution transform and linear pattern classification. *Applied and Computational Harmonic Analysis*, 45(3):616–641, 2018. doi: 10.1016/j.acha.2017.02.002.

Soumyabrata Paul, Arman, S. Lakshmibala, Prasanta K. Panigrahi, S. Ramanan, and V. Balakrishnan. Optimal sensing of photon addition and subtraction on nonclassical light. *Physical Review A*, 111(4):L040601, 2025a. doi: 10.1103/PhysRevA.111.L040601.

Soumyabrata Paul, V. Balakrishnan, S. Ramanan, and S. Lakshmibala. Comparing probability distributions: Application to quantum states of light. *Journal of the Indian Institute of Science*, 105:123–130, 2025b. doi: 10.1007/s41745-025-00474-8. URL https://arxiv.org/abs/2506.10760.

Soumyabrata Paul, H. S. Subramania, S. Ramanan, V. Balakrishnan, and S. Lakshmibala. Machine learning approach to tomographic pattern generation and classification of quantum states of light. *APS Open Science*, 1:000044, 2026. doi: 10.1103/c7g1-yj2y. URL https://journals.aps.org/apsos/abstract/10.1103/c7g1-yj2y.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019. doi: 10.1561/2200000073.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pp. 435–446, 2011.

Mohammad Shifat-E-Rabbi, Xuwang Yin, Abu Hasnat Mohammad Rubaiyat, Shiying Li, Soheil Kolouri, Akram Aldroubi, Jonathan M. Nichols, and Gustavo K. Rohde. Radon cumulative distribution transform subspace modeling for image classification. *Journal of Mathematical Imaging and Vision*, 63:1185–1203, 2021. doi: 10.1007/s10851-021-01052-0.

Gábor J. Székely and Maria L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013. doi: 10.1016/j.jspi.2013.03.018.

The LIGO Scientific Collaboration. A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nature Physics*, 7(12):962–965, 2011. doi: 10.1038/nphys2083.

K. Vogel and H. Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Physical Review A*, 40(5):2847–2849, 1989. doi: 10.1103/PhysRevA.40.2847.

Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, and Seth Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84:621–669, 2012. doi: 10.1103/RevModPhys.84.621. URL https://arxiv.org/abs/1110.3234.

E. Wigner. On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40(5): 749–759, 1932. doi: 10.1103/PhysRev.40.749.

Karol Życzkowski and Wojciech Słomczyński. The Monge distance between quantum states. *Journal of Physics A: Mathematical and General*, 31(45):9095–9104, 1998. doi: 10.1088/0305-4470/31/45/009. URL https://arxiv.org/abs/quant-ph/9711011.

A PROOFS OF THE METRIC AND GEOMETRY STATEMENTS

Conventions. For an order $p \geq 1$ and a probability measure λ on phases, let $\mathcal{D}_{p,\lambda}$ be the states with $\int\int |x|^p P_\rho^\theta(\,\mathrm{d}x)\lambda(\,\mathrm{d}\theta) < \infty$ and define

$$\mathrm{QSW}_{p,\lambda}(\rho,\sigma) = \left[\int \mathcal{W}_p^p\big(P_\rho^\theta, P_\sigma^\theta\big)\lambda(\,\mathrm{d}\theta)\right]^{1/p}. \tag{10}$$

With $p = 2$ and λ uniform on $[0,\pi)$ this is Eq. (3). A phase shift of π reflects both distributions and leaves every transport cost unchanged, so the uniform average over the full circle equals the average over $[0,\pi)$. We use the phase circle $\mathbb{R}/(2\pi\mathbb{Z})$ whenever continuity in θ is needed.

Born laws and moments. If $\rho = \sum_j a_j|\psi_j\rangle\langle\psi_j|$ is a spectral decomposition, then $p_\theta(x) = \sum_j a_j|\psi_{j,\theta}(x)|^2$ with $\psi_{j,\theta}$ the wavefunction in the rotated quadrature representation. The sum is nonnegative and integrates to $\sum_j a_j = 1$, which is the nonnegativity of Eq. (1) without any assumption on the sign of W. For the number operator $\hat{N} = \hat{a}^\dagger\hat{a}$,

$$\frac{1}{\pi}\int_0^\pi \int x^2 P_\rho^\theta(\,\mathrm{d}x)\,\mathrm{d}\theta = \tfrac{1}{2}\,\mathrm{tr}[\rho(\hat{q}^2+\hat{p}^2)] = \mathrm{tr}(\rho\hat{N}) + \tfrac{1}{2}, \tag{11}$$

so finite mean energy is enough for the finite second moments assumed in Theorem 1.

The characteristic function. Write $\chi_\rho(v) = \mathrm{tr}[\rho\, e^{i(v_1\hat{q}+v_2\hat{p})}]$ for $v \in \mathbb{R}^2$. The Fourier transform of the quadrature distribution at phase θ is

$$\int e^{itx} P_\rho^\theta(\,\mathrm{d}x) = \chi_\rho(t\,u_\theta), \tag{12}$$

so the projections give χ_ρ on every ray through the origin. χ_ρ is continuous, by strong continuity of the Weyl operators and trace-class approximation of ρ, and it determines ρ, since the Weyl transform is injective (Vogel & Risken, 1989; Lvovsky & Raymer, 2009). This is the invertibility of the Radon transform quoted in Section 2.

Proof of Theorem 1. Write $M_p(\rho,\theta) = \int |x|^p P_\rho^\theta(\,\mathrm{d}x)$. The bound $\mathcal{W}_p^p(P_\rho^\theta, P_\sigma^\theta) \leq 2^{p-1}[M_p(\rho,\theta) + M_p(\sigma,\theta)]$ proves finiteness on $\mathcal{D}_{p,\lambda}$. Symmetry and nonnegativity hold at each phase. The one-dimensional triangle inequality applied pointwise in θ, followed by Minkowski's inequality in $L^p(\lambda)$, gives the triangle inequality for $\mathrm{QSW}_{p,\lambda}$.

If the distance is zero, the quadrature distributions of ρ and σ agree on a set S of full λ measure. For the uniform measure every nonempty open set of phases has positive measure and therefore meets S, so S is dense. By Eq. (12), equality of the distributions on S gives $\chi_\rho(tu_\theta) = \chi_\sigma(tu_\theta)$ for all $t \in \mathbb{R}$ and $\theta \in S$. Continuity extends the equality to every ray, hence to all of $\mathbb{R}^2$, and injectivity of the Weyl transform gives $\rho = \sigma$. The converse is immediate.

For $p = 2$, the quantile function satisfies $\int_0^1 |F_{\rho,\theta}^{-1}(t)|^2\,\mathrm{d}t = M_2(\rho,\theta)$, so $\Phi(\rho)$ of Eq. (4) lies in $L^2([0,\pi)\times(0,1), \pi^{-1}\,\mathrm{d}\theta\,\mathrm{d}t)$. Integrating the one-dimensional identity Eq. (2) over θ gives $\mathrm{QSW}_2(\rho,\sigma) = \|\Phi(\rho) - \Phi(\sigma)\|_{L^2}$, and the separation just proved is the injectivity of Φ. With λ supported on finitely many phases the same argument gives every metric axiom except separation, which then holds only on a class of states that the measured phases distinguish. This is the finite-phase pseudometric of Section 5. □

Proof of Proposition 1. The distribution of $\hat{X}_\theta$ in ρ_d is P_ρ^θ translated by $u_\theta^\top d$, so its quantile function is $F_{\rho,\theta}^{-1}(t) + u_\theta^\top d$. Eq. (2) gives $\mathcal{W}_p(P_{\rho_d}^\theta, P_{\rho_e}^\theta) = |u_\theta^\top(d-e)|$ for every p. Averaging $|u_\theta^\top(d-e)|^p$ over the uniform phase measure and using rotation invariance,

$$\mathrm{QSW}_p(\rho_d,\rho_e) = c_p\|d-e\|_2, \qquad c_p^p = \frac{1}{\pi}\int_0^\pi |\cos\theta|^p\,\mathrm{d}\theta, \tag{13}$$

and $c_2 = 1/\sqrt{2}$ is Eq. (6). Translating both states by a common displacement cancels from every quantile difference, and rotating both states re-indexes the uniform phase integral, so both are isometries of QSW_p. □

Proof of Proposition 2. The quantile function of $\mathcal{N}(b, s^2)$ is $b + s\Psi^{-1}(t)$ with Ψ the standard normal distribution function. Since Ψ^{-1} has mean zero and second moment one on $(0, 1)$, $\mathcal{W}_2^2(\mathcal{N}(b, s^2), \mathcal{N}(b', s'^2)) = (b - b')^2 + (s - s')^2$, and averaging over θ with $b = u_\theta^\top m$ and $s^2 = u_\theta^\top V u_\theta$ proves Eq. (7).

For the displaced squeezed vacuum put $s_\zeta(\theta) = \sqrt{(e^{-2\zeta}\cos^2\theta + e^{2\zeta}\sin^2\theta)/2}$, so the quantile function at phase θ is $a\cos\theta + s_\zeta(\theta)\Psi^{-1}(t)$. Its derivatives in a and in ζ are $\cos\theta$ and $\partial_\zeta s_\zeta(\theta)\Psi^{-1}(t)$. The squared L^2 norms and the inner product of these two functions are the coefficients of the line element: $g_{aa} = \pi^{-1}\int_0^\pi \cos^2\theta\,\mathrm{d}\theta = 1/2$, $g_{a\zeta} = 0$ because $\int_0^1 \Psi^{-1}(t)\,\mathrm{d}t = 0$, and $g_{\zeta\zeta} = \pi^{-1}\int_0^\pi (\partial_\zeta s_\zeta)^2\,\mathrm{d}\theta$. With $A = e^{-2\zeta}$ and $B = e^{2\zeta}$,

$$(\partial_\zeta s_\zeta)^2 = \frac{A\cos^2\theta + B\sin^2\theta}{2} - \frac{2AB\cos^2\theta\sin^2\theta}{A\cos^2\theta + B\sin^2\theta}, \tag{14}$$

$$\frac{1}{\pi}\int_0^\pi \frac{\cos^2\theta\sin^2\theta}{A\cos^2\theta + B\sin^2\theta}\,\mathrm{d}\theta = \frac{1}{2(\sqrt{A} + \sqrt{B})^2}, \tag{15}$$

the second by the substitution $t = \tan\theta$ on each half quadrant. Since $AB = 1$, $g_{\zeta\zeta} = \cosh(2\zeta)/2 - 1/(4\cosh^2\zeta)$, which is Eq. (8), with $g_{\zeta\zeta}(0) = 1/4$. □

Mean and shape beyond Gaussian states. For any finite-energy state, let m_ρ be its quadrature mean vector and write $\bar{Q}_{\rho,\theta}(t) = F^{-1}_{\rho,\theta}(t) - u_\theta^\top m_\rho$. Since $\int_0^1 \bar{Q}_{\rho,\theta}(t)\,\mathrm{d}t = 0$, expanding the squared quantile difference and integrating eliminates the cross term:

$$\mathrm{QSW}_2^2(\rho, \sigma) = \frac{1}{2}\|m_\rho - m_\sigma\|_2^2 + \frac{1}{\pi}\int_0^\pi\int_0^1 |\bar{Q}_{\rho,\theta}(t) - \bar{Q}_{\sigma,\theta}(t)|^2\,\mathrm{d}t\,\mathrm{d}\theta. \tag{16}$$

This identity requires no differentiability. Along any mean-preserving path whose quantile map is differentiable in L^2, its derivative remains in the closed subspace of functions with zero quantile mean at each phase. It is therefore orthogonal to every displacement tangent $u_\theta^\top\delta$. This applies to squeezing about the state's mean. Squeezing about another point can also change the mean, so orthogonality then need not hold.

For a centred state, Eq. (30) gives the squeezed quantile function as $c(\theta, \zeta)F^{-1}_{\rho,\theta'(\zeta)}(t)$, where $c(\theta, \zeta) = (e^{-2\zeta}\cos^2\theta + e^{2\zeta}\sin^2\theta)^{1/2}$ and $\theta'(\zeta)$ is the direction of $(e^{-\zeta}\cos\theta, e^{\zeta}\sin\theta)$. When differentiation and the chain rule are valid in L^2, the tangent at $\zeta = 0$ is $-\cos 2\theta\, F^{-1}_{\rho,\theta}(t) + \sin 2\theta\,\partial_\theta F^{-1}_{\rho,\theta}(t)$. Thus the response includes a change of phase as well as a scale change. Bounding linear-readout bias over a finite squeezing range additionally requires control of the higher-order response for the chosen state family.

Proof of Proposition 3. We first connect the discrete transport construction in Eq. (5) to quantile projection. Fix a phase and write its empirical measure as $\widehat{P} = n^{-1}\sum_i \delta_{x_{(i)}}$, where the shots are in increasing order. For the reference $\lambda_r = r^{-1}\sum_j \delta_{t_j}$, the monotone optimal coupling is $(F^{-1}_{\lambda_r}, F^{-1}_{\widehat{P}})_{\#}\,\mathrm{Unif}(0, 1)$. Writing $I_j = ((j-1)/r, j/r)$ and $J_i = ((i-1)/n, i/n)$, a matrix representation of this coupling has entries $\Gamma_{ji} = |I_j \cap J_i|$, where $|\cdot|$ denotes interval length. Its row and column sums are $1/r$ and $1/n$, respectively, and its barycentric projection is

$$b_j = r\sum_i \Gamma_{ji}x_{(i)} = r\int_{I_j} F^{-1}_{\widehat{P}}(t)\,\mathrm{d}t. \tag{17}$$

This identity holds for every r, n, including when r does not divide n; then boundary shots contribute according to the interval overlaps. At $r = n$ it reduces to $b_j = x_{(j)}$. For a population distribution $P \in \mathcal{P}_2(\mathbb{R})$, the same coupling with F_P^{-1} gives $b_j = r\int_{I_j} F_P^{-1}(t)\,\mathrm{d}t$. The displacement coordinates $b_j - t_j$ used in transport-based embeddings differ from our uncentred coordinates by a fixed vector and therefore give the same pairwise distances. Likewise, subtracting t from Eq. (4) does not change its L^2 distances.

Fix a phase and let Q denote an empirical or population quantile function. Let P_r be the operator on $L^2(0, 1)$ that replaces a function by its average on each of the r equal bins $I_j = ((j-1)/r, j/r)$.

P_r is the orthogonal projection onto functions constant on those bins, so $\|P_r f\| \le \|f\|$, and the jth coordinate of $P_r Q$ is the barycentric coordinate $r \int_{I_j} Q$ just derived. Applied at each of the L phases and scaled by $1/\sqrt{Lr}$, the projected functions have exactly the Euclidean geometry of $\Phi_{L,r}$ for empirical quantiles (and its population counterpart for exact laws). The phase measure here gives weight $1/L$ to each phase, so $\|\Phi_{L,r}(\rho) - \Phi_{L,r}(\sigma)\| = \|P_r(Q_\rho - Q_\sigma)\|_{L^2(\text{phases}\times(0,1))} \le \|Q_\rho - Q_\sigma\|$, which is the contraction. If r divides r' the bins of r' refine those of r, the range of P_r is contained in the range of $P_{r'}$, and $\|P_r f\| \le \|P_{r'} f\|$, which is the monotonicity along nested bins. An empirical quantile function built from n sorted shots is constant on the n intervals of mass $1/n$, so $P_n Q = Q$ and the distance at $r = n$ is the empirical finite-phase QSW_2 of Section 5. When r divides n, averaging on I_j is the arithmetic mean of the jth block of n/r order statistics. As $r \to \infty$ along nested bins, P_r converges strongly to the identity, for continuous functions by uniform continuity and for all of L^2 by density and contraction, so the compressed distance increases to the uncompressed one.

For the population displacement statement, let ρ_d and ρ_e be displacements of one base state. Their exact quantile difference at phase θ_ℓ is the constant $u_{\theta_\ell}^\top (d - e)$. Every P_r preserves constants, so for every r, including $r = 1$,

$$\|\Phi^{\mathrm{pop}}_{L,r}(\rho_d) - \Phi^{\mathrm{pop}}_{L,r}(\rho_e)\|^2 = (d-e)^\top \Big(\frac{1}{L} \sum_{\ell=1}^{L} u_{\theta_\ell} u_{\theta_\ell}^\top \Big) (d - e). \tag{18}$$

Here $\Phi^{\mathrm{pop}}_{L,r}$ uses the exact quadrature laws in place of the empirical measures in Eq. (5). For $L \ge 2$ equally spaced phases the bracket is $I_2/2$ and the right side is $\frac{1}{2}\|d - e\|^2$, the finite-phase form of Eq. (6). This is the identity behind the statement in Section 6 that the displacement grid error vanishes with unlimited shots. Independently sampled records need not differ by constant shifts, so this exact identity does not apply to them. □

B Finite-shot and finite-phase guarantees

Write $D = \mathrm{QSW}_p(\rho, \sigma)$, D_L for the same average over L independent uniform phases with the quadrature distributions known exactly, and $\widehat{D}_{L,n}$ for the distance between the sorted records of Proposition 4, extended to order $p \ge 1$ by replacing the squared costs and square root with pth powers and a pth root. Assume $\sup_\theta \int |x|^q P_\xi^\theta(\,\mathrm{d}x) \le K_q$ for both states $\xi \in \{\rho, \sigma\}$ and some $q > 2p$. Expectations are over the phases and the shots. Constants depend only on the displayed moment orders.

Proof of Proposition 4. For a one-dimensional distribution μ with qth moment at most K_q, $q > 2p$, and its empirical measure $\widehat{\mu}_n$ from n independent samples, Fournier & Guillin (2015, Theorem 1) gives

$$\mathbb{E}\, \mathcal{W}_p^p(\widehat{\mu}_n, \mu) \le C_{p,q} K_q^{p/q} n^{-1/2}, \tag{19}$$

since its second term $n^{-(q-p)/q}$ is at most $n^{-1/2}$ in this regime. Conditional on the phases, the reverse triangle inequality in the product of the L one-dimensional transport spaces gives

$$|\widehat{D}_{L,n} - D_L| \le \sum_{\xi \in \{\rho,\sigma\}} \left[\frac{1}{L} \sum_{\ell=1}^{L} \mathcal{W}_p^p\big(\widehat{P}^{\theta_\ell}_{\xi,n}, P^{\theta_\ell}_{\xi}\big) \right]^{1/p}, \tag{20}$$

and Jensen's inequality with Eq. (19) bounds its expectation by $2 C_{p,q}^{1/p} K_q^{1/q} n^{-1/(2p)}$, which is $n^{-1/4}$ at $p = 2$.

For the phase error put $h(\theta) = \mathcal{W}_p^p(P_\rho^\theta, P_\sigma^\theta)$. The moment assumption gives $0 \le h \le H := 2^p K_q^{p/q}$. For $p = 2$, set $a = \mathbb{E}h = D^2$. If $a > 0$, rationalizing the square-root difference and using $h^2 \le Hh$ gives

$$\mathbb{E}|D_L - D| \le \frac{\mathbb{E}|L^{-1}\sum_\ell h(\theta_\ell) - a|}{\sqrt{a}} \le \sqrt{\frac{\mathrm{Var}(h)}{La}} \le \sqrt{\frac{H}{L}}, \tag{21}$$

and if $a = 0$ then $h = 0$ almost surely and the error vanishes. Combining this with Eq. (20) at $p = 2$ proves Eq. (9). The shot bound is conditional on the phases and applies unchanged to fixed phases, relative to D_L. □

General order and copy budget. For any $p \geq 1$ with $q > 2p$, the bounded-variable variance inequality gives $\mathbb{E}|D_L^p - D^p| \leq H/(2\sqrt{L})$. Applying $|a^{1/p} - b^{1/p}| \leq |a-b|^{1/p}$ and Jensen's inequality yields $\mathbb{E}|D_L - D| \leq (H/(2\sqrt{L}))^{1/p}$. Together with the shot bound, this gives an expected error of order $K_q^{1/q}(n^{-1/(2p)} + L^{-1/(2p)})$. Choosing $L \asymp n \asymp \varepsilon^{-2p}$ therefore suffices for error $O(\varepsilon)$ with $B = O(\varepsilon^{-4p})$ copies at fixed moment scale. For $p = 2$, the stronger phase bound improves this allocation. For a budget $B = Ln$, the split $n \asymp B^{2/3}$, $L \asymp B^{1/3}$ gives an expected error $O(K_q^{1/q} B^{-1/6})$, or $B = O(\varepsilon^{-6})$ copies at fixed moment scale. These are sufficient bounds from moments alone, not optimal rates. They concern distance estimation, not the recovery of a density matrix or the error of a regressor.

Why sharper rates need regularity. For equally spaced phases $\theta_\ell = \theta_0 + (\ell-1)\pi/L$ modulo π, the natural alternative is deterministic quadrature. If the π-periodic integrand h is C^2 with $\|h''\|_\infty \leq C_h$, the composite trapezoid rule gives

$$|D_L^p - D^p| \leq \frac{\pi^2 C_h}{12L^2}, \qquad |D_L - D| \leq \left(\frac{\pi^2 C_h}{12L^2}\right)^{1/p}, \tag{22}$$

and for distances bounded away from zero the distance error is $O(L^{-2})$ as well. For each Gaussian pair with positive definite covariances, the $p = 2$ integrand is smooth. Non-Gaussian distributions with nodes need this checked separately. On the shot side, at a quantile where the density f is continuous and positive, the pointwise asymptotic variance of $\sqrt{n}(F_{\widehat{P}}^{-1}(t) - F_P^{-1}(t))$ is $t(1-t)/f(F_P^{-1}(t))^2$. Thin valleys and tails can therefore amplify quantile error, while nodes require separate analysis. A sharper $\mathcal{W}_2$ guarantee must control these effects across quantile levels, beyond the moment assumption used here.

C Geometry experiment: protocol and results

States and acquisition. Both grids have 49 Gaussian states: coherent states on $(q,p) \in [-2,2]^2$, and displaced squeezed vacua on $(a,\zeta) \in [-2,2] \times [-0.9, 0.9]$, each on a 7×7 uniform grid. The latter have mean $(a,0)$ and covariance $\frac{1}{2}\operatorname{diag}(e^{-2\zeta}, e^{2\zeta})$. Quadrature laws and sampling are analytic, with no Fock truncation. Each state supplies $n = 64$ shots at $L = 12$ equally spaced phases in $[0,\pi)$, or 768 copies, shared by all sample-based methods. There are three independent repeats. Population QSW_2 uses the Gaussian formula of Proposition 2 at 2,048 equally spaced phases. These deterministic grids do not satisfy the random-phase sampling assumption of Proposition 4.

Distances. QSWE at $r = n$ gives empirical QSW_2 exactly. For records $X_{\ell i}, Y_{\ell i}$ with empirical distribution functions $\widehat{F}_\ell, \widehat{G}_\ell$, the other record-based distances are

$$D^2_{\mathrm{energy}} = \frac{2}{L}\sum_\ell \int_{\mathbb{R}} |\widehat{F}_\ell(x) - \widehat{G}_\ell(x)|^2 \,\mathrm{d}x, \tag{23}$$

$$D^2_{\mathrm{MMD},h} = \frac{1}{L}\sum_\ell \left[\frac{1}{n^2}\sum_{i,j} k_h(X_{\ell i}, X_{\ell j}) + \frac{1}{n^2}\sum_{i,j} k_h(Y_{\ell i}, Y_{\ell j}) - \frac{2}{n^2}\sum_{i,j} k_h(X_{\ell i}, Y_{\ell j})\right], \tag{24}$$

where $k_h(x,y) = \exp[-(x-y)^2/(2h^2)]$ (Székely & Rizzo, 2013; Gretton et al., 2012). The MMD estimator includes the within-record diagonal terms and is nonnegative. The geometry experiment uses these empirical distances directly; the regression features approximate them numerically as specified in Appendix D.2. For pure Gaussian states, Bures distance is $\sqrt{2 - 2\sqrt{F}}$ with

$$F = \frac{\exp[-\frac{1}{2}(m-m')^\top (V+V')^{-1}(m-m')]}{\sqrt{\det(V+V')}}. \tag{25}$$

It uses the exact states, not the homodyne records (Banchi et al., 2015).

Diagnostics. Classical MDS keeps the two largest positive eigenvalues of $-\frac{1}{2}JD^{\circ 2}J$, where $J = I - \mathbf{1}\mathbf{1}^\top/49$ (Gower, 1966). For coordinates Y, the reported stress is $[\sum_{i<j}(D_{ij} - \|Y_i -$

$Y_j\|)^2/\sum_{i<j} D_{ij}^2]^{1/2}$, computed before alignment. Grid error is $\|\widetilde{Y} - Z\|_F/\|Z - \bar{Z}\|_F$, where Z contains the physical parameters and $\widetilde{Y}$ is aligned to Z by translation, an orthogonal transformation and one common scale. No anisotropic rescaling is allowed. Stress measures the loss in projecting a distance geometry into two dimensions; grid error measures agreement with the parameter grid. A local line element alone does not determine this global MDS projection. Figure 2 uses the first seed; tables report the mean and one sample standard deviation over repeats.

C.1 Geometry and reference-size tables

Table 2 collects the distances shown in Figure 2 and all tested kernel bandwidths. No bandwidth is selected using the grid coordinates. Population rows are deterministic.

	Displacement		Shift and squeeze	
Distance	Stress	Grid error	Stress	Grid error
QSW_2 (population)	0.0000	0.000	0.0374	0.101
QSW_2 (empirical)	0.0089 ± 0.0001	0.027 ± 0.002	0.0614 ± 0.0021	0.117 ± 0.007
Bures	0.3432	0.268	0.2382	0.685
Energy	0.1352 ± 0.0010	0.083 ± 0.002	0.1158 ± 0.0018	0.472 ± 0.004
RBF-MMD, $h = 0.5$	0.346 ± 0.003	0.241 ± 0.001	0.228 ± 0.003	0.640 ± 0.002
RBF-MMD, $h = 1$	0.271 ± 0.002	0.198 ± 0.002	0.156 ± 0.002	0.583 ± 0.004
RBF-MMD, $h = 2$	0.150 ± 0.002	0.115 ± 0.002	0.091 ± 0.002	0.472 ± 0.004
RBF-MMD, $h = 4$	0.055 ± 0.001	0.049 ± 0.002	0.053 ± 0.003	0.393 ± 0.004

Table 2: Geometry on the two 7×7 grids, 768 copies per state. The nonzero population squeezing-grid error is a property of the distance and its planar representation, not sampling error.

Reference compression uses exact empirical quantile integrals, checked against an independent linear-program transport solver. Table 3 reports the relative Frobenius difference from the full distance matrix and the resulting grid error. The nested references contract pairwise distances as in Proposition 3; this does not guarantee preservation of MDS coordinates or regression accuracy.

	Displacement		Shift and squeeze	
r	Distance error	Grid error	Distance error	Grid error
1	0.0059 ± 0.0004	0.027 ± 0.005	0.1578 ± 0.0080	0.403 ± 0.006
4	0.0033 ± 0.0003	0.027 ± 0.005	0.0233 ± 0.0016	0.135 ± 0.014
8	0.0023 ± 0.0001	0.027 ± 0.005	0.0115 ± 0.0010	0.123 ± 0.015
16	0.0014 ± 0.0001	0.027 ± 0.005	0.0058 ± 0.0003	0.119 ± 0.016
32	0.0007 ± 0.0000	0.027 ± 0.005	0.0026 ± 0.0003	0.117 ± 0.016
64	0.0000 ± 0.0000	0.027 ± 0.005	0.0000 ± 0.0000	0.117 ± 0.017

Table 3: Reference-size sweep, $L = 12$, $n = 64$, three repeats. Feature dimension is $12r$; distance error vanishes at $r = n$.

C.2 Kernel bandwidth and response laws

For one Gaussian slice $P = \mathcal{N}(0, s^2)$ and its translate $Q = \mathcal{N}(\delta, s^2)$, direct integration gives

$$E^2(P, Q) = 2|\delta| \operatorname{erf}\left(\frac{|\delta|}{2s}\right) + \frac{4s}{\sqrt{\pi}}\left(e^{-\delta^2/(4s^2)} - 1\right), \tag{26}$$

$$\mathrm{MMD}_h^2(P, Q) = \frac{2h}{\sqrt{h^2 + 2s^2}}\left(1 - e^{-\delta^2/[2(h^2+2s^2)]}\right), \tag{27}$$

whereas $\mathcal{W}_2(P, Q) = |\delta|$. For the first identity, use $E^2 = 2\mathbb{E}|X - Y| - \mathbb{E}|X - X'| - \mathbb{E}|Y - Y'|$ with independent draws; $X - Y \sim \mathcal{N}(-\delta, 2s^2)$ has a folded-normal absolute mean. For the second, integrate $e^{-(X-Y)^2/(2h^2)}$ under that same Gaussian. As $\delta \to 0$, $E^2 = \delta^2/(s\sqrt{\pi}) + O(\delta^4)$ and

$\mathrm{MMD}_h^2 = h\delta^2/(h^2+2s^2)^{3/2} + O(\delta^4)$. For large shifts, E grows as $\sqrt{2|\delta|}$, while MMD approaches a finite limit on scale $\sqrt{h^2+2s^2}$.

For Gaussian slices with means m_1, m_2 and variances v_1, v_2,

$$\mathrm{MMD}_h^2 = \frac{h}{\sqrt{h^2+2v_1}} + \frac{h}{\sqrt{h^2+2v_2}} - \frac{2h\exp[-(m_1-m_2)^2/(2(h^2+v_1+v_2))]}{\sqrt{h^2+v_1+v_2}}. \tag{28}$$

As $h \to \infty$, the leading term is $(m_1-m_2)^2/h^2$ when the means differ and $3(v_1-v_2)^2/(4h^4)$ at equal means. Thus a wide kernel is relatively less sensitive to width changes. On the tested grids, increasing h from 0.5 to 4 lowers displacement-grid error from 0.241 to 0.049, but squeezing-grid error remains 0.393, compared with 0.117 for empirical QSW_2 (Table 2). This is evidence about the tested bandwidths and two-dimensional representations, not failure of the Gaussian kernel to distinguish the underlying distributions.

D Parameter regression: protocol and supporting results

D.1 State family and acquisition

For real α, the partially coherent cat is

$$\rho(\alpha, c, \phi) = \frac{|\alpha\rangle\langle\alpha| + |-\alpha\rangle\langle-\alpha| + c(e^{i\phi}|\alpha\rangle\langle-\alpha| + e^{-i\phi}|-\alpha\rangle\langle\alpha|)}{2(1+ce^{-2\alpha^2}\cos\phi)}. \tag{29}$$

The even cat has $c=1$, $\phi=0$; $c=0$ gives the incoherent mixture. Let $g(x \mid \theta)$ denote its quadrature density. Under squeezing $S = \mathrm{diag}(e^{-\zeta}, e^{\zeta})$ followed by displacement $d=(q,p)$, the Wigner function becomes $W(S^{-1}(\xi-d))$. Changing variables $\xi = S\eta + d$ in its Radon marginal gives

$$p_\theta(x) = \frac{1}{s_\theta}\, g\left(\frac{x - q\cos\theta - p\sin\theta}{s_\theta}\,\middle|\,\theta'\right), \quad s_\theta = \|S^\top u_\theta\|, \quad u_{\theta'} = \frac{S^\top u_\theta}{s_\theta}. \tag{30}$$

Here $\det S = 1$, and the factor $1/s_\theta$ is the Jacobian of the one-dimensional rescaling. Explicitly, $s_\theta = (e^{-2\zeta}\cos^2\theta + e^{2\zeta}\sin^2\theta)^{1/2}$ and $\theta' = \mathrm{atan2}(e^{\zeta}\sin\theta, e^{-\zeta}\cos\theta)$. The resulting quantiles shift by $q\cos\theta + p\sin\theta$; squeezing changes both their scale and their phase index. An independent Wigner-marginal check of this rule agreed to 1.7×10^{-16} on 18 audit points.

All six parameters are drawn independently: $q \sim U[-1,1]$, $\zeta \sim U[-0.3, 0.3]$, $c \sim U[0,1]$, $\phi \sim U[-1,1]$ rad, $p \sim U[-0.3, 0.3]$, and $\alpha \sim U[1.5, 1.9]$. For the primary targets (q,ζ), the other four parameters are unknown nuisances. Each repeat draws 48 training, 24 validation and 96 test states and one common phase offset $\theta_0 \sim U[0,\pi)$. All states and methods use phases $\theta_\ell = \theta_0 + \ell\pi/12$, $\ell = 0,\ldots,11$. Sampling uses tabulated inverse CDFs on $8{,}193$ points in $[-9,9]$. The 768-copy records are prefixes of the $3{,}072$-copy records (64 versus 256 shots per phase). Ridge uses 30 repeats; Gaussian processes and training-size comparisons use 20. The budget and transformation-range sweeps have the counts stated in Figures 4 and 5. All reported intervals are Student-t 95% half-widths over independent repeats; comparisons use shared records within each repeat.

D.2 Features and regressors

QSWE uses the normalized barycentric coordinates of Eq. (5). Gaussian features concatenate per-phase sample means and standard deviations. The moment features concatenate empirical means of $H_k(X)/\sqrt{2^k k!}$, $k=1,\ldots,6$, with physicists' Hermite polynomials and fixed vacuum scaling. Both moment blocks carry the common phase normalization $1/\sqrt{L}$. Energy features evaluate empirical CDFs at 513 equally spaced points in $[-9,9]$, weighted for trapezoidal integration of $2L^{-1}\sum_\ell \int (F_\ell - G_\ell)^2$. RBF features use linear binning into $2{,}048$ bins over an interval of length 36 and a Fourier transform. At positive frequencies $\omega_k = 2\pi k/36$, real and imaginary characteristic-function components have weight $\sqrt{2a_k/L}$, where $a_k = \sqrt{2\pi}he^{-h^2\omega_k^2/2}/36$; frequencies with $a_k/a_0 \le 10^{-10}$ are omitted. Thus the energy and RBF feature distances approximate their empirical integral and kernel distances. RBF bandwidths $h \in \{0.25, 0.5, 1, 2\}$ are selected on validation data.

Both regressors centre features on the training mean, divide by one training RMS norm, and standardize targets using training states only. Predictions are transformed back without clipping to the prior box. Every model choice is made separately for each target on the 24 validation states. **L, ridge regression:** the intercept is unpenalized and the penalty is $n_{\mathrm{train}}\lambda$, with $\lambda \in \{10^{-8}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1, 10\}$; ties favour the larger penalty. **N, PCA and Gaussian-process regression:** training-set PCA retains $k \in \{5, 10, 20\}$ components, selected on validation error. Scores are scaled by their training standard deviations. The Gaussian process uses an RBF kernel with a separate length scale per component and a white-noise term, fitted by training marginal likelihood with two restarts.

D.3 Fisher benchmark with unknown nuisance parameters

For $\beta = (q, \zeta, c, \phi, p, \alpha)$ and n independent shots per phase,

$$I_{jk}(\beta) = n \sum_{\ell=1}^{12} \int \frac{\partial_{\beta_j} p_{\theta_\ell}(x)\, \partial_{\beta_k} p_{\theta_\ell}(x)}{p_{\theta_\ell}(x)}\, \mathrm{d}x, \qquad \mathrm{CRB}_j(\beta) = [I(\beta)^{-1}]_{jj}. \tag{31}$$

The full 6×6 matrix is inverted before extracting the target entries, so all nuisance parameters remain unknown. Translation derivatives are $-\cos\theta\, \partial_x p_\theta$ and $-\sin\theta\, \partial_x p_\theta$. The other derivatives use central finite differences with steps 2×10^{-3} for ζ, ϕ and α, and 10^{-3} for c (clipped to $[0, 1]$ with the actual step in the denominator). Halving steps and refining the integration grid changed the audited bounds by at most 1.4×10^{-5} and 8.7×10^{-6} relatively.

Within each repeat, the reported benchmark is $[96^{-1} \sum_{i=1}^{96} \mathrm{CRB}_j(\beta_i)]^{1/2}$. Tables average the per-repeat RMSE-to-benchmark ratios, rather than taking a ratio of aggregate means. This is a local classical bound for unbiased estimation from the specified record, not a universal lower bound on the risk of biased or prior-informed predictors. At 3,072 copies the mean benchmarks are 0.01896 for q and 0.01879 for ζ. Holding p and α known would reduce the latter to 0.007140: the large change is consistent with the mean Fisher correlation -0.9247 between squeezing and cat amplitude. Shape results below use absolute RMSE; a prior-averaged fringe-phase bound is unstable because ϕ becomes unidentifiable as $c \to 0$.

D.4 Likelihood check of the benchmark

A maximum-likelihood fit estimates all six parameters from the shots, without training labels or true nuisance values. It starts at $(c, \phi, p, \alpha) = (0.5, 0, 0, 1.7)$ with (q, ζ) chosen by a coarse search (steps 0.05 and 0.10), then uses L-BFGS-B with an analytic gradient. Bounds are $c \in [0, 1]$, $\alpha \in [0.5, 5]$ and $\phi \in [-\pi, \pi]$; q, ζ and p are unconstrained. Across 2,880 test-state fits, 99.86% converge and 4.06% reach a bound. At 3,072 copies, RMSE is 0.01911 ± 0.00051 on q and 0.01892 ± 0.00057 on ζ, each about 1.01 times the benchmark. This checks its scale on the simulated family; the estimator is given the correct parametric model, unlike the learned feature comparisons.

D.5 Results supporting the main comparisons

Tables 4 and 5 give the two main shot budgets. Table 6 retains intermediate reference sizes, Table 8 reports paired differences for the headline comparisons, and Table 7 gives the training-size comparison.

	RMSE / prior s.d. (%)		RMSE / $\sqrt{\mathrm{CRB}}$	
Feature set	q	ζ	q	ζ
chance	100.0	100.0	30.45	9.22
floor	3.3	10.8	1.00	1.00
linear ridge (L), 30 repeats				
QSWE $r = 16$	5.2 ± 0.2	20.2 ± 1.0	1.58	1.87
QSWE $r = n$	5.2 ± 0.2	21.1 ± 1.0	1.59	1.95
Gaussian moments	10.0 ± 0.3	24.4 ± 0.8	3.05	2.25
Moments 1–6	20.3 ± 1.8	53.5 ± 4.3	6.18	4.93
Energy CDF	9.2 ± 0.4	32.1 ± 1.5	2.79	2.96
RBF mean embedding	9.9 ± 0.6	30.5 ± 1.4	3.00	2.81
PCA + Gaussian-process regression (N), 20 repeats				
QSWE $r = 16$	5.5 ± 0.3	26.3 ± 2.0	1.67	2.43
QSWE $r = n$	5.7 ± 0.4	29.1 ± 1.9	1.72	2.69
Gaussian moments	10.5 ± 0.5	31.1 ± 3.1	3.18	2.86
Moments 1–6	32.2 ± 4.5	48.7 ± 3.5	9.80	4.49
Energy CDF	5.9 ± 0.4	39.5 ± 2.9	1.81	3.64
RBF mean embedding	6.9 ± 0.6	32.4 ± 1.9	2.09	2.99

Table 4: Displacement and squeezing at 3,072 copies per state. RMSE as a percentage of the prior standard deviation (mean $\pm$ 95% half-width) and the mean per-repeat ratio to the nuisance-adjusted local benchmark. L uses 30 repeats; N uses the first 20 repeats.

	RMSE / prior s.d. (%)		RMSE / $\sqrt{\mathrm{CRB}}$	
Feature set	q	ζ	q	ζ
chance	100.0	100.0	15.23	4.61
floor	6.6	21.7	1.00	1.00
linear ridge (L), 30 repeats				
QSWE $r = 16$	11.5 ± 0.4	35.2 ± 1.2	1.75	1.62
QSWE $r = n$	11.5 ± 0.4	35.2 ± 1.1	1.75	1.62
Gaussian moments	20.2 ± 0.5	36.0 ± 1.4	3.08	1.66
Moments 1–6	30.5 ± 2.7	58.1 ± 3.4	4.64	2.68
Energy CDF	14.9 ± 0.6	40.7 ± 1.2	2.28	1.88
RBF mean embedding	15.9 ± 0.7	40.0 ± 1.2	2.42	1.84
PCA + Gaussian-process regression (N), 20 repeats				
QSWE $r = 16$	11.9 ± 0.7	39.1 ± 1.4	1.81	1.80
QSWE $r = n$	12.2 ± 0.5	40.2 ± 1.9	1.86	1.85
Gaussian moments	19.8 ± 0.9	38.3 ± 1.5	3.01	1.77
Moments 1–6	37.1 ± 3.9	50.4 ± 3.5	5.63	2.32
Energy CDF	13.1 ± 1.0	43.6 ± 2.4	1.99	2.01
RBF mean embedding	11.6 ± 0.5	41.5 ± 1.8	1.76	1.92

Table 5: Displacement and squeezing at 768 copies per state, with 64 shots per phase from the same records. Conventions and repeat counts are those of Table 4.

	Ridge (L)		Gaussian process (N)	
r	q	ζ	q	ζ
4	0.03521 ± 0.0014	0.03734 ± 0.0013	0.03888 ± 0.0039	0.04321 ± 0.0043
16	0.02989 ± 0.0011	0.03505 ± 0.0017	0.03175 ± 0.0018	0.04563 ± 0.0035
64	0.03013 ± 0.00096	0.03627 ± 0.0018	0.03343 ± 0.0023	0.05074 ± 0.0043
n	0.03017 ± 0.00094	0.03660 ± 0.0018	0.03276 ± 0.0021	0.05046 ± 0.0032

Table 6: Reference-size comparison on cat regression, $n = 256$ shots per phase. Absolute RMSE, mean $\pm$ 95% half-width; 30 repeats for L and 20 for N. Accuracy at a smaller reference is an empirical result, not a consequence of distance contraction alone.

Feature set	Readout	48	160	480
QSWE $r = 16$	L	0.02997 ± 0.0012	0.02469 ± 0.0013	0.02246 ± 0.00085
	N	0.03175 ± 0.0018	0.02459 ± 0.0011	0.02211 ± 0.00073
Gaussian moments	L	0.05795 ± 0.0026	0.05643 ± 0.0024	0.05437 ± 0.0014
	N	0.06043 ± 0.0030	0.05679 ± 0.0025	0.05352 ± 0.0015
Energy CDF	L	0.05232 ± 0.0031	0.04032 ± 0.0014	0.03270 ± 0.0012
	N	0.03434 ± 0.0026	0.02576 ± 0.00092	0.02239 ± 0.00078
RBF mean embedding	L	0.05491 ± 0.0034	0.04036 ± 0.0020	0.03185 ± 0.0011
	N	0.03964 ± 0.0036	0.02564 ± 0.0011	0.02194 ± 0.00059

Feature set	Readout	48	160	480
QSWE $r = 16$	L	0.03564 ± 0.0025	0.02568 ± 0.00090	0.02322 ± 0.00073
	N	0.04563 ± 0.0035	0.02904 ± 0.0014	0.02607 ± 0.00082
Gaussian moments	L	0.04207 ± 0.0016	0.03454 ± 0.0013	0.03367 ± 0.0010
	N	0.05381 ± 0.0054	0.03335 ± 0.00099	0.03091 ± 0.00071
Energy CDF	L	0.05636 ± 0.0035	0.04305 ± 0.0012	0.03741 ± 0.0011
	N	0.06833 ± 0.0050	0.04002 ± 0.0025	0.02934 ± 0.0013
RBF mean embedding	L	0.05431 ± 0.0030	0.04224 ± 0.0012	0.03838 ± 0.00099
	N	0.05619 ± 0.0033	0.03314 ± 0.0012	0.02551 ± 0.0011

Table 7: Training-size study at 3,072 copies, 20 repeats. Absolute RMSE on q (upper) and ζ (lower), mean $\pm$ 95% half-width. The 48-state arm uses the main study's states; the 160- and 480-state arms have separate random streams. Chance RMSEs are 0.5774 and 0.1732; the per-arm benchmarks range from 0.0189 to 0.0190 on q and round to 0.0188 on ζ.

Comparator minus reference	q	ζ
QSWE $r = 16$ (L) $-$ RBF mean embedding (L)	-1.424 ± 0.206	-0.943 ± 0.121
QSWE $r = 16$ (L) $-$ Gaussian moments (L)	-1.477 ± 0.102	-0.382 ± 0.085
RBF mean embedding (N) $-$ QSWE $r = 16$ (L)	0.509 ± 0.203	1.095 ± 0.181

Table 8: Paired differences in RMSE-to-benchmark ratio at 3,072 copies. Mean $\pm$ 95% half-width over shared repeats; negative favours the comparator. Pairs involving N use its 20 repeats; L–L pairs use 30.

	Ridge (L)		Gaussian process (N)	
Feature set	c	ϕ	c	ϕ
QSWE $r = 16$	0.265 ± 0.011	0.534 ± 0.019	0.319 ± 0.012	0.629 ± 0.033
QSWE $r = n$	0.268 ± 0.0088	0.539 ± 0.019	0.314 ± 0.010	0.636 ± 0.030
Gaussian moments	0.299 ± 0.0061	0.589 ± 0.018	0.327 ± 0.017	0.635 ± 0.024
Moments 1–6	0.307 ± 0.016	0.592 ± 0.019	0.308 ± 0.011	0.613 ± 0.022
Energy CDF	0.292 ± 0.012	0.566 ± 0.021	0.314 ± 0.013	0.625 ± 0.029
RBF mean embedding	0.267 ± 0.0099	0.534 ± 0.020	0.206 ± 0.013	0.584 ± 0.038

Table 9: Shape-parameter RMSE at 3,072 copies: 30 repeats for L, 20 for N; mean $\pm$ 95% half-width. Chance RMSE is 0.2887 for c and 0.5774 for ϕ. QSWE shows limited improvement over chance; the nonlinear RBF predictor performs better on coherence. These results support the transport-versus-shape limitation in Section 7.

E Preliminary studies

These simulated studies support the applications and limitations discussed in Section 7. Intervals are 95% half-widths over repeats; they are descriptive and are not adjusted for multiple comparisons.

E.1 Drift monitoring with a running embedding

A squeezed vacuum with $\zeta_0 = 0.5$ supplies 16 shots at each of 12 phases per step. Steps 1–40 form a reference record; monitoring runs from step 41 to 160, with drift beginning at step 80. The statistic S_k compares each new record with the reference using QSW_2, Gaussian-moment distance, or a

Gaussian generalized likelihood ratio (GLR). Unequal record lengths use the full quantile integral. The GLR is twice the log likelihood ratio of the step-fitted Gaussian to the reference-fitted Gaussian, summed over phases. Each statistic enters $C_k = \max(0, C_{k-1} + S_k - \kappa)$, initialized at zero, with κ its null mean and a reset after an alarm. Thresholds target 1% false alarms per step on 50 calibration runs; 50 fresh null runs give 0.83–0.95% for the cumulative-sum monitors.

Table 10 covers abrupt squeezing changes, axis rotation, per-shot phase jitter, loss, a mixture of $\Delta\zeta = 0.05$ with 0.05 rad rotation and 0.1 rad jitter, and a linear squeezing ramp of 0.2 over 80 steps. QSWE is earlier on squeezing changes, whereas the Gaussian GLR is earlier on rotation, jitter and loss. This comparison uses one nominal state and one shot budget; thresholds require recalibration when the nominal state changes.

Drift	QSWE	Gaussian moments	Gaussian GLR
none	51–58	51–58	51–58
$\Delta\zeta = 0.05$	24.8 ± 6.9	28.8 ± 7.2	40.4 ± 9.1
$\Delta\zeta = 0.1$	10.3 ± 1.9	11.8 ± 2.5	15.7 ± 3.1
$\Delta\zeta = 0.2$	3.4 ± 0.6	3.9 ± 0.7	4.4 ± 0.8
axis rotation 0.2 rad	17.0 ± 3.7	15.1 ± 2.5	5.7 ± 0.8
phase jitter 0.2 rad	37.0 ± 7.7	38.0 ± 7.5	20.7 ± 5.1
loss $\eta = 0.9$	54.5 ± 8.9	46.7 ± 8.5	35.5 ± 7.8
mixed	19.4 ± 4.5	26.4 ± 5.3	28.5 ± 6.8
ramp	28.5 ± 4.3	30.6 ± 4.7	38.4 ± 5.3

Table 10: Cumulative-sum detection delays in steps, mean $\pm$ 95% half-width over 50 repeats, censored at 81 steps. The no-drift row reports the range of censored false-alarm delays.

E.2 Regression on a numerically simulated Kerr cat

An even cat evolves for unit time under Kerr dynamics and single-photon loss,

$$\dot{\rho} = -i\chi[\hat{a}^{\dagger 2}\hat{a}^2, \rho] + \kappa(\hat{a}\rho\hat{a}^\dagger - \tfrac{1}{2}\{\hat{a}^\dagger\hat{a}, \rho\}), \qquad \kappa = -\log\eta, \tag{32}$$

then is displaced by (q, p). The targets are $q \sim U[-1, 1]$, $\chi \sim U[0, 0.35]$ and $\eta \sim U[0.8, 1]$; nuisances are $p \sim U[-0.3, 0.3]$ and $\alpha \sim U[1.5, 1.9]$. The 12-phase acquisition, $48/24/96$ split and regressors follow Appendix D, with 10 repeats. Tomograms are evaluated numerically at Fock dimension 60; comparison with dimension 100 gives agreement to 10^{-16}. The full five-parameter Fisher matrix uses exact translation derivatives and finite differences for the remaining scores; halving the steps changes the audited bound by 0.12%.

Table 11 supports the displacement result cited in the discussion: QSWE retains its advantage under ridge regression. For Kerr phase, nonlinear regression improves both QSWE and RBF features; neither approaches the local benchmark. This is one numerically specified family, not a comparison with state reconstruction or numerical maximum-likelihood estimation.

		q		χ	
Regressor	Feature set	RMSE / s.d. (%)	ratio	RMSE / s.d. (%)	ratio
	floor	2.6	1.00	2.1	1.00
ridge	QSWE $r = 16$	6.3 ± 0.4	2.45	19.8 ± 0.9	9.52
ridge	QSWE $r = n$	6.4 ± 0.4	2.49	19.5 ± 0.8	9.35
ridge	Gaussian moments	8.7 ± 0.6	3.40	39.0 ± 2.7	18.77
ridge	Energy CDF	9.1 ± 0.9	3.54	21.6 ± 0.8	10.37
ridge	RBF mean embedding	9.5 ± 0.8	3.69	19.9 ± 1.1	9.54
Gaussian process	QSWE $r = 16$	7.4 ± 0.6	2.89	11.9 ± 2.5	5.76
Gaussian process	RBF mean embedding	9.6 ± 0.7	3.74	14.1 ± 1.5	6.76

Table 11: Kerr-cat regression, $3{,}072$ copies, 48 training states, 10 repeats. RMSE as a percentage of prior standard deviation (mean $\pm$ 95% half-width) and its ratio to the numerical local benchmark. For transmission η, all tested rows exceed 85% of the prior standard deviation, compared with a benchmark near 20%.

E.3 Detector loss and noise

The cat regression is repeated with independently drawn detector efficiency $\eta \sim U[0.85, 1]$ and additive-noise variance $\nu \sim U[0, 0.05]$ per state, both unknown to the predictors. The measured law is that of $\sqrt{\eta}X + Z$ for ideal X and independent $Z \sim \mathcal{N}(0, (1-\eta)/2 + \nu)$, in the convention that vacuum variance is $1/2$. Other priors, records, splits and regressors are unchanged, with 10 repeats. Table 12 retains the feature comparisons: QSWE has the lowest ridge error on both targets in this study. No phase jitter is included. We report prior-normalized RMSE because the nuisance-adjusted local bounds vary widely when loss and added noise are both unknown.

Regressor	Feature set	q	ζ
ridge	QSWE $r = 16$	5.8 ± 0.5	22.9 ± 2.2
ridge	Gaussian moments	10.4 ± 0.5	27.8 ± 1.7
ridge	Moments 1–6	23.5 ± 5.2	50.6 ± 4.0
ridge	Energy CDF	9.5 ± 0.6	32.9 ± 1.6
ridge	RBF mean embedding	10.8 ± 0.8	31.9 ± 2.2
Gaussian process	QSWE $r = 16$	6.6 ± 0.7	29.5 ± 4.6
Gaussian process	RBF mean embedding	9.3 ± 2.0	34.7 ± 2.1

Table 12: Regression with loss and detector noise, 3,072 copies, 48 training states, 10 repeats. RMSE as a percentage of prior standard deviation, mean $\pm$ 95% half-width. The ideal QSWE $r = 16$ ridge row is 5.2% on q and 20.2% on ζ (Table 4).

E.4 A shape feature and its concatenation with QSWE

Each phase's shots are standardized by their sample mean and standard deviation. Using the Fourier grid of Appendix D.2, we retain characteristic-function moduli at frequencies in $[2.5, 10]$, with equal weights $1/\sqrt{LK}$ for K retained frequencies. This block, denoted HP|CF|, removes per-slice location, scale and reflection. It is not invariant to general squeezing, which also relabels the phase. The hybrid concatenates this block and QSWE $r = 16$ with weights $1/\sqrt{2}$ after centring and RMS scaling each block on the training states.

Table 13 uses the same cat protocol as Appendix D. The shape block improves coherence estimation but loses displacement; the concatenation retains sensitivity to both, without matching the better individual block on either target. This is evidence of complementary responses, not disjoint information or an established optimal combined representation.

Feature set	q	ζ	c	ϕ
linear ridge (L), 30 repeats				
QSWE $r = 16$	0.02989 ± 0.0011	0.03505 ± 0.0017	0.265 ± 0.011	0.534 ± 0.019
HP|CF|	0.5971 ± 0.018	0.05223 ± 0.0017	0.0903 ± 0.0036	0.588 ± 0.014
QSWE $\oplus$ HP|CF|	0.04289 ± 0.0017	0.03862 ± 0.0011	0.0957 ± 0.0031	0.589 ± 0.015
PCA + Gaussian-process regression (N), 20 repeats				
QSWE $r = 16$	0.03175 ± 0.0018	0.04563 ± 0.0035	0.319 ± 0.012	0.629 ± 0.033
HP|CF|	0.6383 ± 0.025	0.04984 ± 0.0030	0.0822 ± 0.0036	0.635 ± 0.029
QSWE $\oplus$ HP|CF|	0.05268 ± 0.0029	0.04090 ± 0.0022	0.0890 ± 0.0047	0.630 ± 0.027

Table 13: Shape feature and concatenation, 3,072 copies. Absolute RMSE, mean $\pm$ 95% half-width. Chance levels are 0.5774, 0.1732, 0.2887 and 0.5774, respectively.