*Exceptional service in the national interest*

Sandia National Laboratories

# High-Accuracy Classification of Single-Qubit Noise via Machine Learning

Travis L Scholten, University of New Mexico & Sandia National Laboratories

## Machine learning (ML) could help us develop new tomographic protocols.

Standard tomographic protocols may not scale as the number of qubits in a device increases, nor is the path to developing new ones always clear.

```
My Data
Circuit(j)   P(j)
----------------
{}           1
Gx           .44
Gy           .45
GxGx         .9
GxGxGx       .68
GyGyGy       .70
```

→ Data processing (e.g., RB, GST) → Quantity of interest

***What are ML tools that can extract (new & useful) quantities of interest out of device data?***

## Noise affects circuits & manifests itself in their probabilities.

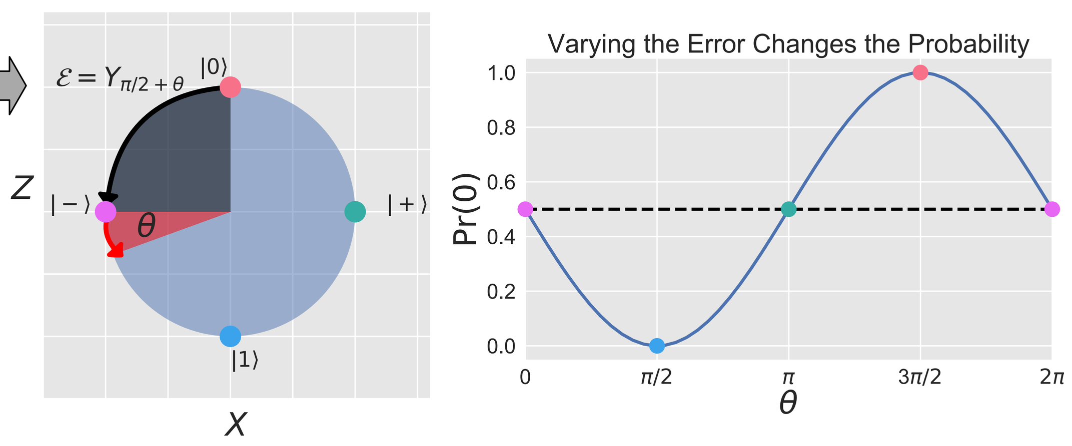
Noise affects circuit probabilities, and data sets (collections of circuits & their estimated probabilities) encode information about noise.

Gate set tomography (GST) data sets are useful in this regard, because the circuits in the data set *amplify all possible noise.*

Example: over-rotation error of a single-qubit gate

$|0\rangle$ — $Y_{\pi/2}$ — measurement   $\Pr(0) = \mathrm{Tr}(|0\rangle\langle 0|\mathcal{E}(|0\rangle\langle 0|)) = \frac{1}{2}(1-\sin\theta)$

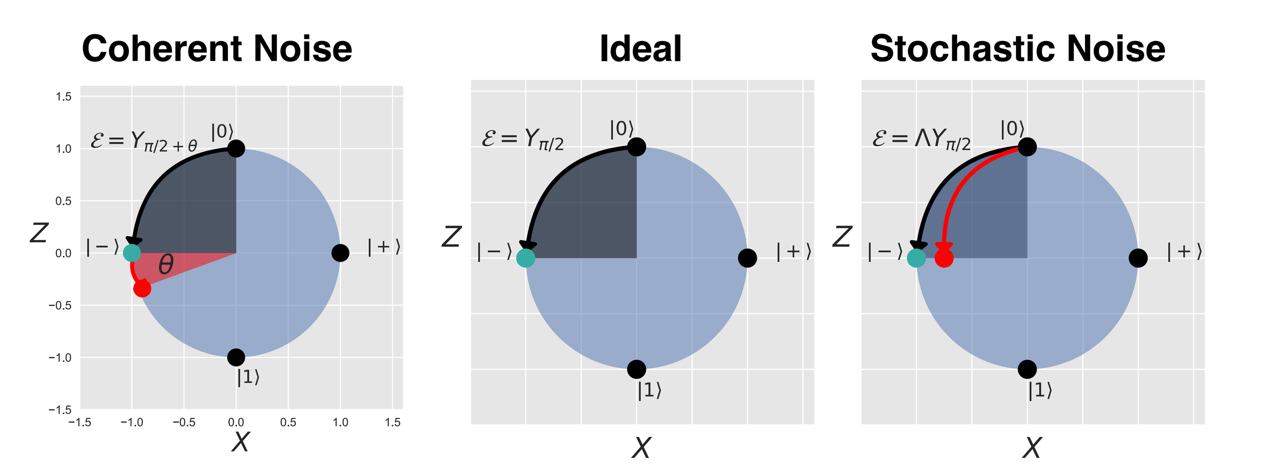
(The circuit we write down)   (Noise affects outcome probability)




## Overview: machine learning (ML) to classify coherent & stochastic single-qubit noise

Telling the difference between coherent noise ("systematic" errors) and stochastic noise ("random" errors) would be useful in debugging a qubit.

As the circuits used for GST amplify all possible noise, GST data sets are useful for classifying these kinds of noise.



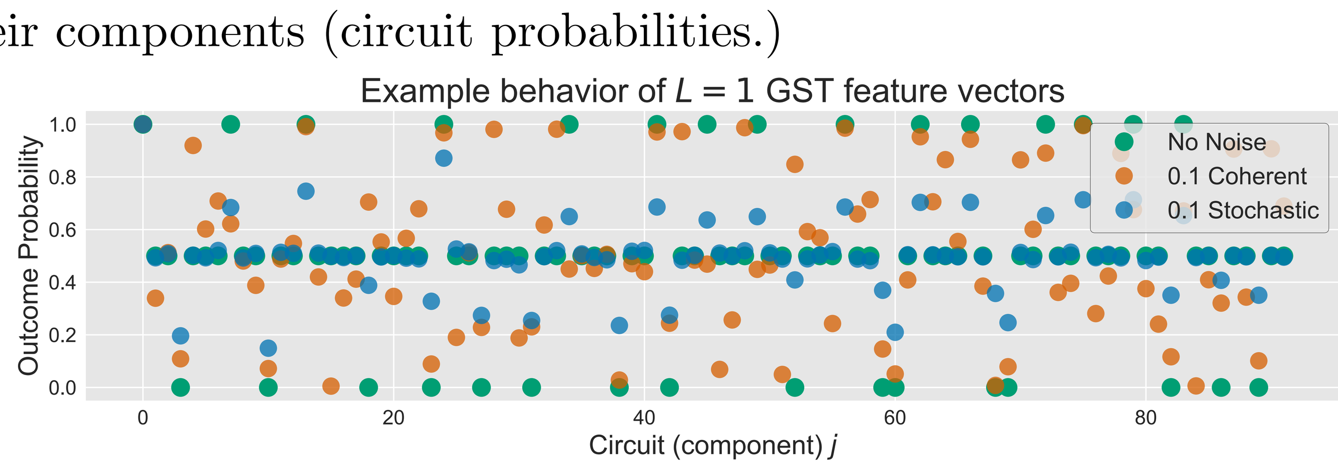

$\dot{\rho} = -i[H_0,\rho] - i[e,\rho]$   $\dot{\rho} = -i[H_0,\rho]$   $\dot{\rho} = -i[H_0,\rho] + A\rho A^\dagger - \frac{1}{2}\{A^\dagger A,\rho\}$

GST datasets (with counts normalized to [0, 1]) can be represented as *feature vectors* in a *feature space.*

```
## Columns = minus count, plus count
{}      100  0
Gx       44  56
Gy       45  55
GxGx      9  91
GxGxGx   68  32
GyGyGy   70  30
```

$\to \begin{pmatrix} 0 \\ .56 \\ .55 \\ .91 \\ .32 \\ .3 \end{pmatrix} \to \mathbf{f} = (f_1, f_2, \cdots, f_d)$

The feature vectors encode information about noise in their components (circuit probabilities.)




## In depth: Gate set tomography (GST) circuits

Each GST circuit consists of: a state prep $\rho$, the application of 1 of 6 fiducials $F_j$, $l$ repetitions of a gate or germ $g$ such $|g^l| \leq L$, the application of another fiducial $F_k$, and a measurement $E$.

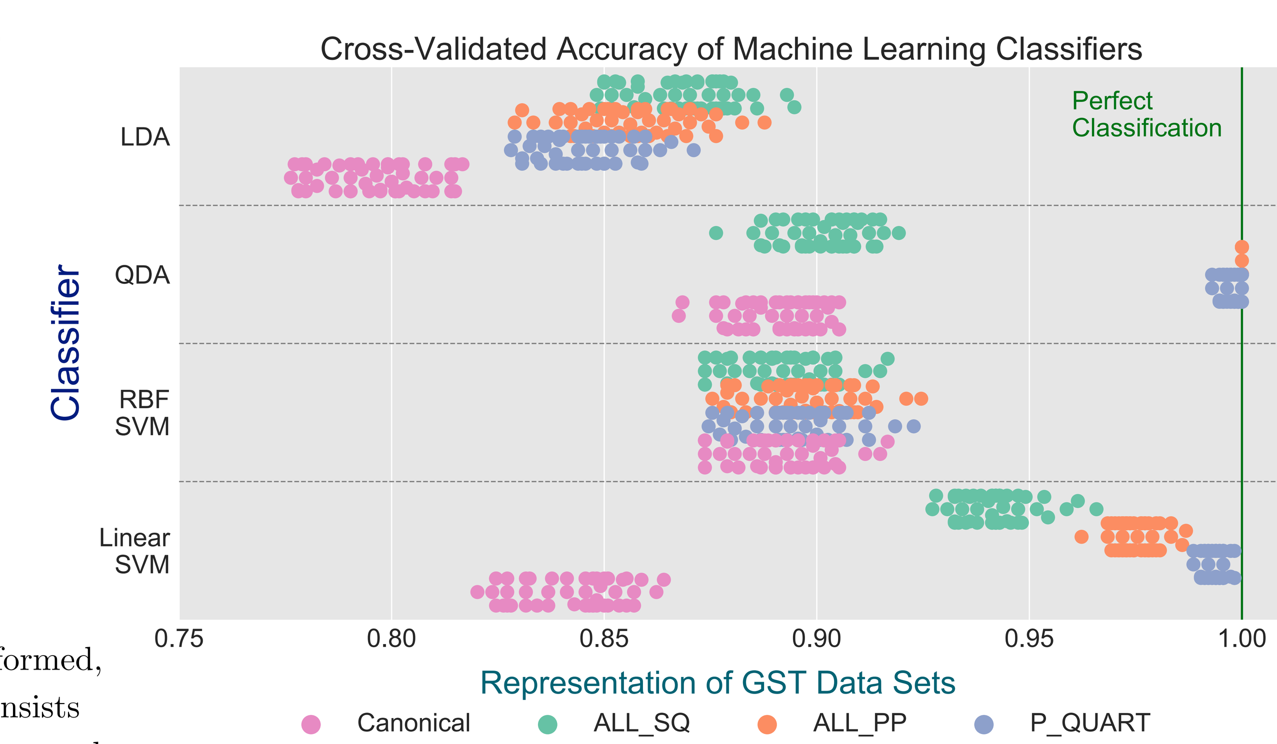
$\rho$ → $F_j$ → $g^l$ → $F_k$ → $E$

We'll focus on $L = 1$ ("linear GST"). For this value of $L$ there are 2 germs $(Gx, Gy)$ and the total number of circuits, $d$, is 92.

***GST feature vectors are "quartically separable". Why?***

We investigated 4 ML classifiers and 4 representations of GST data sets.

Classification accuracy depends on the **classifier** and the **representation** of the data set. (See "In Depth" below.)



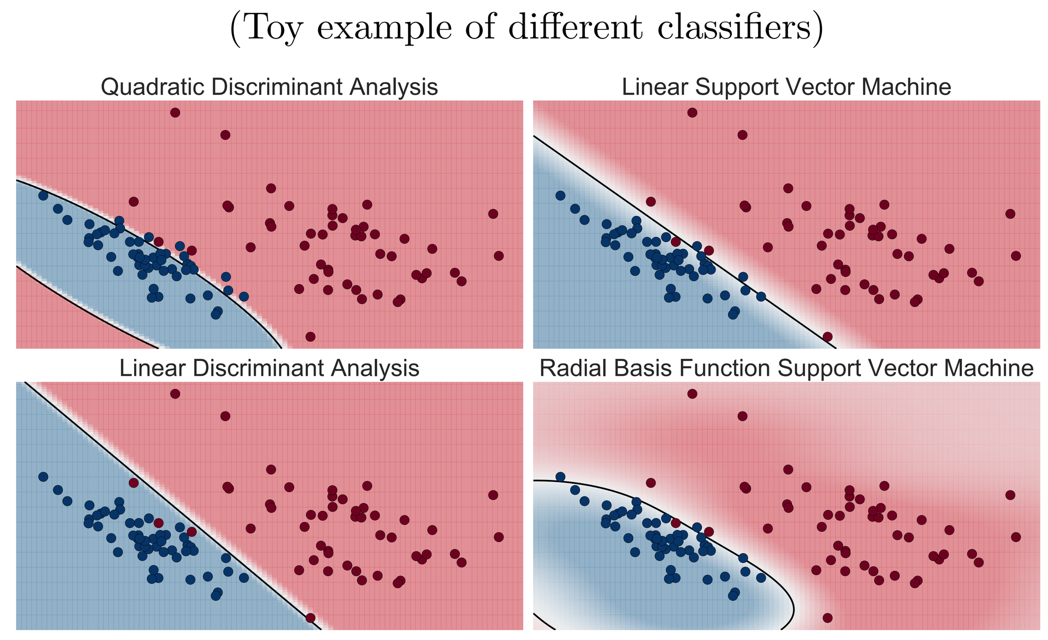

A 50-fold shuffle-split cross-validation was performed, with 10% of the data used for testing. Data consists of 300 random instances for each of 2 noise types and 19 unique noise strengths in $[10^{-4}, .5]$.

## In depth: ML classifiers and representations of GST data sets

*Quadratic Discriminant Analysis (QDA)*
Bayesian solution for two-class classification assuming data for each class is Gaussian, with unequal covariance matrices.

*Linear Discriminant Analysis (LDA)*
Similar idea as QDA, but assumes covariances for each class are the same.

(Toy example of different classifiers)




*Linear Support Vector Machine (Linear SVM)*
Finds a hyperplane by solving an optimization problem that trades off between accuracy on known data and accuracy on future data ("soft-margin").

*Radial Basis Function Support Vector Machine (RBF SVM)*
Similar idea as linear SVM, but classification depends on a Gaussian function of the data.

Different representations of the data give rise to new feature spaces with their own structure.

(Canonical)   $\mathbf{f} = (f_1, f_2, \cdots, f_d)$

(Squares of all components; ALL_SQ)   $\mathbf{f} \to \mathbf{f} \bigoplus_j f_j^2$   $\mathbb{R}^d \to \mathbb{R}^{2d}$   $(x,y) \to (x,y,x^2,y^2)$

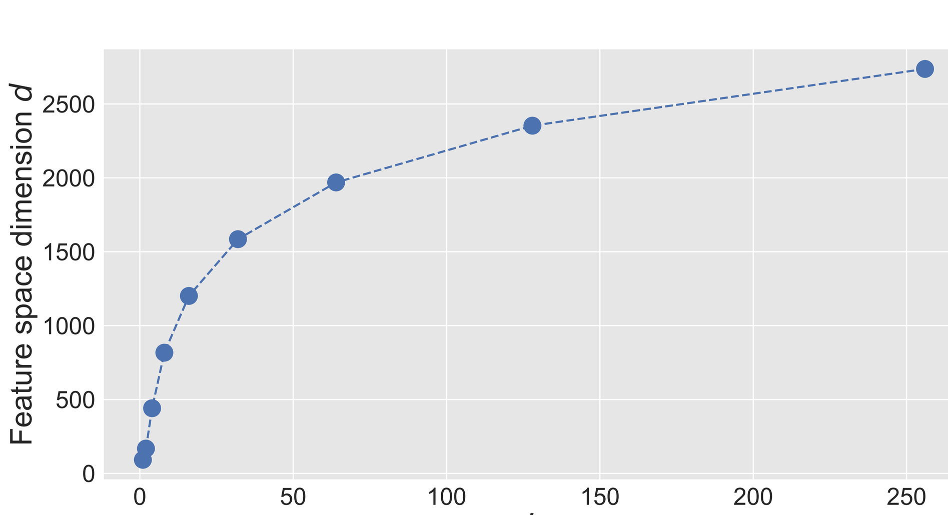
(All unique pairwise products; ALL_PP)   $\mathbf{f} \to \mathbf{f} \bigoplus_{j,k} f_j f_k$   $\mathbb{R}^d \to \mathbb{R}^{d(d+3)/2}$   $(x,y) \to (x,y,x^2,y^2,xy)$

(All unique pairwise products & squares; P_QUART)   $\mathbf{f} \to \mathbf{f} \bigoplus_{j,k} f_j f_k \bigoplus_{j,k} (f_j f_k)^2$   $\mathbb{R}^d \to \mathbb{R}^{d(d+2)}$   $(x,y) \to (x,y) \oplus (x^2,y^2,xy,x^4,y^4,(xy)^2)$

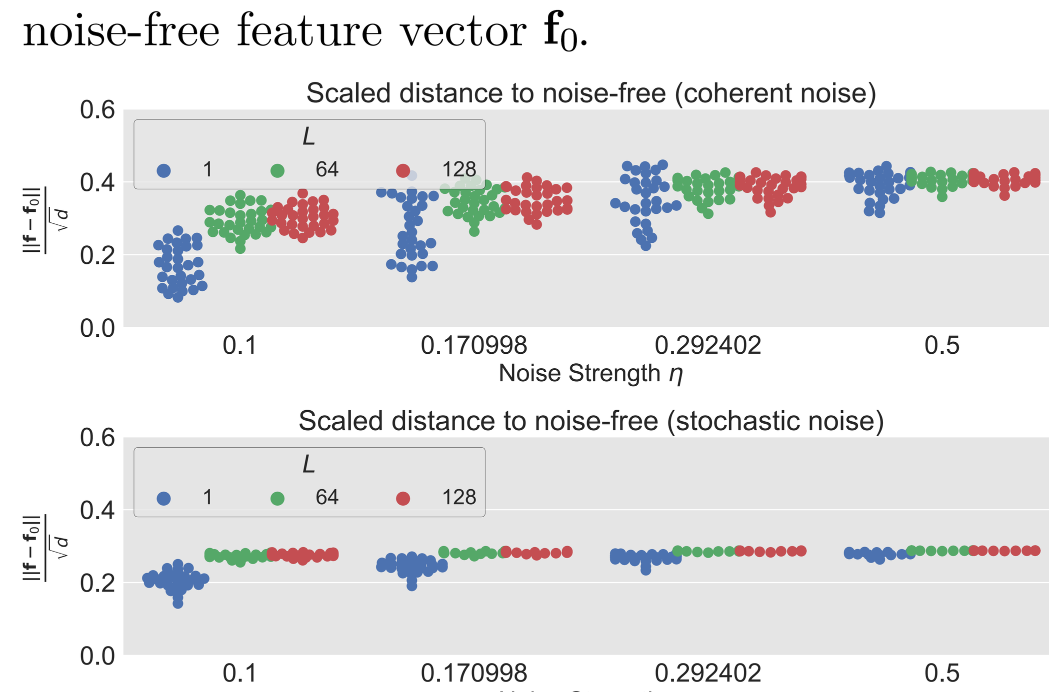
## In depth: structure of (canonical) GST feature space

***The feature space for GST data sets has structure which can be exploited in the development of new ML-based tomographic protocols.***

$L$ controls the number of circuits in a GST data set, which in turn sets the *dimension* of the feature space, $d$.



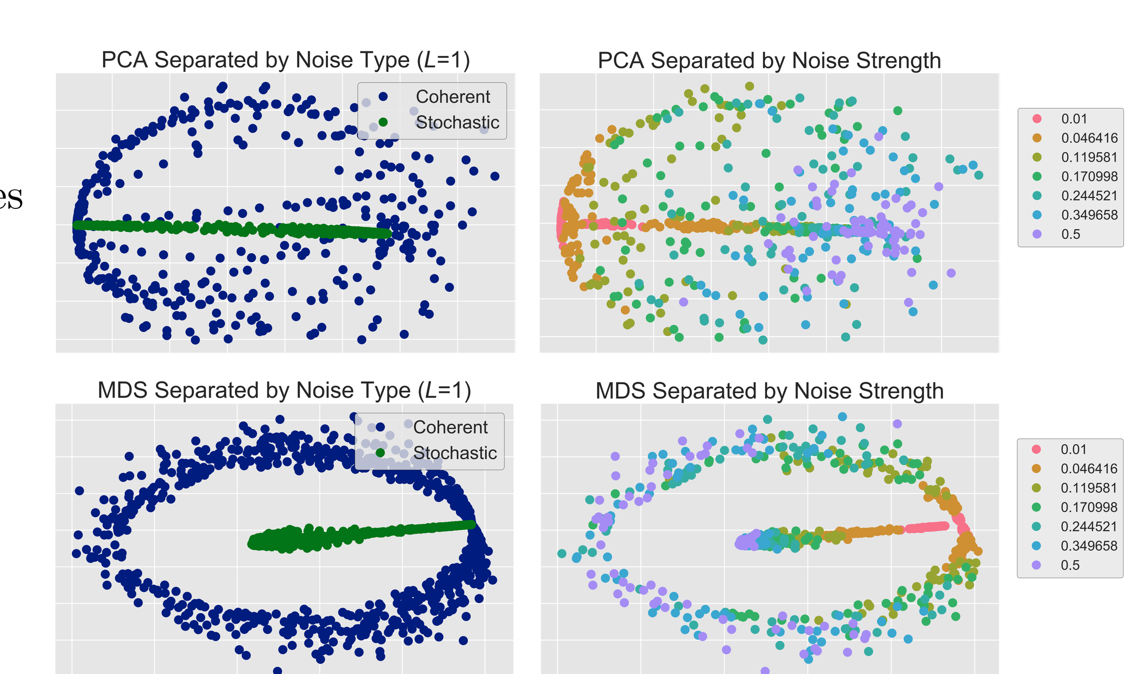

As $L$ increases, the noisy feature vectors "fill out" the feature space, moving away from the noise-free feature vector $\mathbf{f}_0$.




Low-dimensional representations of $L = 1$ feature vectors show why linear separability may not be possible in general.

*Principal Component Analysis (PCA)*
Representation captures maximal variance in the data.

*Multidimensional scaling (MDS)*
Representation maximally preserves pairwise distances.




U.S. DEPARTMENT OF ENERGY · NNSA National Nuclear Security Administration · Sandia National Laboratories · CCR Center for Computing Research

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.