Ensemble discrimination via selective random rotations and projective measurements

C S Sudheer Kumar[1,*]

[1] *NMR Research Center and Department of Physics,*
Indian Institute of Science Education and Research, Pune 411008, India

Consider two ensembles of N qubits each: $\mathcal{E}_1$ and $\mathcal{E}_2$. The ratio $N_{|a_1\rangle}/N$ ($N_{|a_2\rangle}/N$) is approximately 1/2 (1/2), where $N_{|a_1\rangle}$ ($N_{|a_2\rangle}$) is the total number of qubits in $\mathcal{E}_1$ ($\mathcal{E}_2$) which are in the state $|a_1\rangle$ ($|a_2\rangle$), $a_1 = 0,1$ ($a_2 = +,-$). $|0\rangle, |1\rangle$ are eigenkets of σ_z (Pauli-z matrix), and $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. If we cannot address and control each of the N qubits in the ensemble separately (i.e., no local control, but only global control), like in nuclear magnetic resonance (NMR) spin ensembles, then both $\mathcal{E}_1$ and $\mathcal{E}_2$ are said to be maximally mixed, and hence it is not possible to discriminate between them. This is because, both $\mathcal{E}_1$ and $\mathcal{E}_2$ give same expectation value of an arbitrary observable. As number of measurements increases, variance of sample mean (of measurement outcomes) decreases, and hence sample mean approaches expectation value of the observable being measured. We are going to show that, if we have local control (like in experiments with single photons), then selectively rotating about x-axis (on Bloch sphere) each of the N qubits in the ensemble by a *random* angle, reduces variance of sample mean in $\mathcal{E}_1$ (this is due to sort of convoluting two independent probability distributions). As random x-rotations does nothing (up to an insignificant global phase) to the states $|\pm\rangle$, variance of sample mean remains unaltered in $\mathcal{E}_2$. Without random x-rotations, both $\mathcal{E}_1$ and $\mathcal{E}_2$ give same variance of sample mean. Hence we can discriminate between $\mathcal{E}_1$ and $\mathcal{E}_2$, via variance of sample mean. We also show that, numerical simulation results support theoretical predictions.

PACS numbers: 03.67.-a, 03.65.Ta, 03.65.Ud, 03.67.Ac
Keywords: Random rotations, Projective measurements

I. INTRODUCTION

"There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy "-W Shakespeare [1]. If we have an ensemble of identical copies of an arbitrary unknown state of a qubit, then we can know the unknown state via tomography. Instead consider the two ensembles $\mathcal{E}_1, \mathcal{E}_2$ shown in Fig. (1). In $\mathcal{E}_1$ ($\mathcal{E}_2$) the ratio $T_{|a_1\rangle}/M$ ($T_{|a_2\rangle}/M$) is approximately 1/2 (1/2), where $T_{|a_1\rangle}$ ($T_{|a_2\rangle}$) is the total number of qubits in a given column of $\mathcal{E}_1$ ($\mathcal{E}_2$) which are in the state $|a_1\rangle$ ($|a_2\rangle$), $a_1 = 0,1$ ($a_2 = +,-$). $|0\rangle, |1\rangle$ are eigenkets of σ_z (Pauli-z matrix) with eigenvalues $+1, -1$ respectively, and $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. When we do not have local control (like in nuclear magnetic resonance (NMR) spin ensembles), then $\mathcal{E}_i$ is said to be maximally mixed, $i = 1,2$. Hence, expectation value of an arbitrary observable is same in both the ensembles $\mathcal{E}_1, \mathcal{E}_2$, and hence we cannot discriminate between $\mathcal{E}_1$ and $\mathcal{E}_2$.

However if we have local control (like in experiments with single photons), even though we obtain same expectation value of an arbitrary observable in both $\mathcal{E}_1$ and $\mathcal{E}_2$, we can still make the variance of sample mean different in $\mathcal{E}_1$ and $\mathcal{E}_2$ as follows: As far as we know, everywhere in literature when they talk of probabilities (and hence expectation value of respective observable), they implicitly neglect the variance of respective sample mean of measurement outcomes. This is because, as number of measurements increases, variance of sample mean decreases, and hence sample mean approaches expectation value of the observable being measured. In our discrimination protocol also, variance of sample mean in both the ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ decreases, as number of projective measurements increases, as required. However, because of applying *random* x-rotation ($(\theta_q)_x$) se-

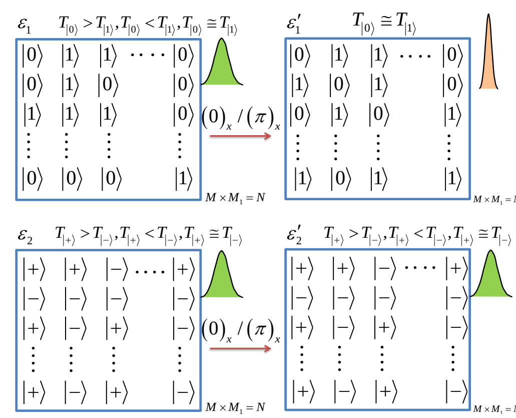

FIG. 1. (Color online) Matrix representation implies that we can address and control each of the $N(= M \times M_1)$ qubits in the ensemble $\mathcal{E}_i$ separately, i.e., we have local control, $i = 1,2$. Any qubit in $\mathcal{E}_i$ is always in a definite state, $i = 1,2$. $(\theta_q)_x$ is rotation about x-axis on Bloch sphere through a *random* angle $\theta_q = 0, \pi$ with probability $1/2, 1/2$ respectively. Gaussian is the probability density function of sample mean of M number of σ_z (Pauli-z matrix) measurement outcomes. M_1 sample mean points are used to construct the full Gaussian. Note that we are going to measure only after applying $(\theta_q)_x$s. Given one of $\mathcal{E}_1, \mathcal{E}_2$, we need to find out whether the given ensemble is $\mathcal{E}_1$ or $\mathcal{E}_2$.

* sudheer.kumar@students.iiserpune.ac.in

lectively to each of the qubit states in the ensemble using local control, variance of sample mean is reduced in the ensemble $\mathcal{E}_1$ (narrow Gaussian in Fig. (1)). This effect is due to sort of convoluting probability distribution of sample mean (before applying $(\theta_q)_x$s) and probability distribution of random x-rotation angle (θ_q). Applying $(\theta_q)_x$s is analogous to rotating a nonuniform (in mass distribution)($\equiv$ $(T_{|0\rangle} > T_{|1\rangle})$) disc at high speed. As random x-rotation does nothing (up to an insignificant global phase) to the states $|\pm\rangle$, variance of sample mean remains unaltered in the ensemble $\mathcal{E}_2$. Without random x-rotations, both $\mathcal{E}_1$ and $\mathcal{E}_2$ give same variance of sample mean. Hence we can discriminate between the two ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$, via variance of sample mean.

In section II we give theoretical details of the discrimination protocol. In section III we present MATLAB simulation results, and we conclude in section IV.

II. THEORY

Problem

Note: Whenever we say measurement, we mean projective measurement unless stated otherwise. There are two ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ of $N(= M \times M_1)$ qubits each, where M, M_1 are sufficiently large enough to obtain approximately normally distributed sample mean of M measurement outcomes. $\mathcal{E}_i$ is such an ensemble where we have local control, $i = 1, 2$. We divide $\mathcal{E}_i$ as follows to obtain M_1 sample mean points of M measurements each: $\mathcal{E}_i = \sum_{j=1}^{M_1} \mathcal{E}_{ij},\ i = 1, 2$. $\mathcal{E}_{1j}$ $(\mathcal{E}_{2j})$ has $T_{|0\rangle}$ $(T_{|+\rangle})$ number of $|0\rangle$s ($|+\rangle$s) and $T_{|1\rangle}$ $(T_{|-\rangle})$ number of $|1\rangle$s ($|-\rangle$s) where $T_{|0\rangle} + T_{|1\rangle} = M$ $(T_{|+\rangle} + T_{|-\rangle} = M)$. Value of $T_{|a_i\rangle}$ varies with j such that $T_{|a_i\rangle} \to \mathrm{ND} : M/2, M/4$ ($T_{|a_i\rangle}$ is a Normally Distributed random variable with mean $M/2$ and variance $M/4$), $p_{|a_i\rangle}(= T_{|a_i\rangle}/M) \to \mathrm{ND} : 1/2, 1/(4M)$ where $i = 1, 2$ and $a_1 = 0, 1,\ a_2 = +, -$. E.g., $\mathcal{E}_1$ $(\mathcal{E}_2)$ can be prepared by measuring σ_z (σ_x (Pauli-x matrix)) selectively on an ensemble of N identical copies of $|+\rangle$ ($|0\rangle$). Here 'Selectively' is used to stress the fact that we have local control. We are going to show how to discriminate between $\mathcal{E}_1$ and $\mathcal{E}_2$.

Solution

Notation and definition

For notational convenience, almost every where we use same symbol for both random variable (θ_q, S_1, etc.) and its value (θ_q, S_1, etc. respectively). Whether we are referring to random variable or its value is understandable from the context where it is used.
(1) We selectively rotate j^{th} qubit state $(j = 1, 2, ..., N)$ about x-axis on Bloch sphere, through an angle $\theta_q, q = 1, 2, ...$ (i.e., we evolve the j^{th} qubit state under the unitary operator $(\theta_q)_x = \exp(-i\theta_q\sigma_x/2)$), where θ_q is a random number drawn from the discrete set $\{\theta_1, \theta_2, ...\}$ with probability $\{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$ respectively ($\{\theta_1, \theta_2, ...\} \to \{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$), $\sum_q p^o_{\theta_q} = 1$.
(2) If we measure σ_z selectively on each of the M qubits in the ensemble $\mathcal{E}_{ij}$, then sample mean $S_i = (T^+_i - T^-_i)/M,\ T^+_i + T^-_i = M$ where $T^\pm_i$ is the total number of ± 1 outcomes, $i = 1, 2$. Value of S_i varies with j. However, we are going to measure only after selectively applying $(\theta_q)_x$ to each of the M qubits in $\mathcal{E}_{ij}, i = 1, 2$. Further $T^+_1 = T_{|0\rangle}, T^-_1 = T_{|1\rangle}$ ($\because |0\rangle, |1\rangle$ are eigenkets of σ_z, and hence no collapse upon measurement. In other words, variances $(\Delta\sigma_z)^2_{|0\rangle} = (\Delta\sigma_z)^2_{|1\rangle} = 0$ where $(\Delta X)^2_{|\zeta\rangle} = \langle(\langle X\rangle_{|\zeta\rangle} - x)^2\rangle = \langle X^2\rangle_{|\zeta\rangle} - \langle X\rangle^2_{|\zeta\rangle}$ [2]).
(3) Sample mean $S'_i = (T'^+_i - T'^-_i)/M,\ T'^+_i + T'^-_i = M$ where $T'^\pm_i$ is the total number of ± 1 outcomes obtained by measuring σ_z selectively on each of the M qubits in the ensemble $\mathcal{E}'_{ij}$. $\mathcal{E}'_{ij}$ is got by selectively applying $(\theta_q)_x$ to each of the M qubits in the ensemble $\mathcal{E}_{ij},\ i = 1, 2$. Value of S'_i varies with j. But $\mathcal{E}'_{2j} = \mathcal{E}_{2j}$ (neglecting global phase in qubit states). Hence $S'_2 = S_2 + X = S_2$ where X corresponds to application of $(\theta_q)_x$s.

T^+_2, T^-_2 are also the total number of $|0\rangle$s, $|1\rangle$s respectively in the ensemble $\mathcal{E}_{1j}$ ($\because |+\rangle$ and $|-\rangle$ are equivalent with respect to σ_z measurement outcomes (Appendix (B 3)). Hence measuring σ_z selectively on each of the M qubits in the ensemble $\mathcal{E}_{2j}$ is equivalent to measuring σ_z selectively on an ensemble of M identical copies of $|+\rangle$. But $\mathcal{E}_{1j}$ can also be obtained by measuring σ_z selectively on an ensemble of M identical copies of $|+\rangle$). $\Rightarrow T^\pm_2$ is also the total number of ± 1 outcomes obtained by measuring σ_z selectively on each of the M qubits in the ensemble $\mathcal{E}_{1j}$. Hence $S_1 \equiv S_2$ (S_1, S_2 are independent and identically distributed (i.e., they have same mean and variance) random variables).

According to central limit theorem, in the large M limit, sample mean $S_i \to \mathrm{ND} : 0, 1/M,\ i = 1, 2$ [3], (Appendix (A 1)). Here mean $\langle S_i\rangle = \langle\sigma_z\rangle_{|+\rangle} = 0$, and variance $\Delta S^2_i = (\Delta\sigma_z)^2_{|+\rangle}/M = 1/M$, because as explained above, S_i corresponds to measuring σ_z selectively on an ensemble of M identical copies of $|+\rangle,\ i = 1, 2$. We obtain one value of S_i from $\mathcal{E}_{ij}$. As $j = 1, 2, ..., M_1$ we get M_1 sample mean points and hence we can construct the entire normal (Gaussian) probability density function of $S_i,\ i = 1, 2$. We are going to show that for $\theta_1 \neq \theta_2, p^o_{\theta_1} \neq 0, p^o_{\theta_2} \neq 0$, variance of S'_1 will be less than that of S'_2. This is because, in this case $S'_1 \neq S_1 (\equiv S_2 = S'_2$ as shown above). Hence we can discriminate between $\mathcal{E}_1$ and $\mathcal{E}_2$.

A. General case

Motivation

Consider the following theorem: If $X_i \to \mathrm{ND} : \mu_i, \sigma^2_i$, then $Z = \sum_{i=1}^{\tilde{N}} X_i \to \mathrm{ND} : \sum_{i=1}^{\tilde{N}} \mu_i, \sum_{i=1}^{\tilde{N}} \sigma^2_i$ where X_is are normally distributed independent random variables [3]. Probability distribution of Z is the convolution of that of X_is. Note that Z has probability distribution different from that of X_is. This is the motivation behind introducing a new independent random variable θ_q into already present random variable S_1 in the ensemble $\mathcal{E}_{1j}$, so that resultant probability distribution might be different from that of S_1. Let X_i be a sample mean: $X_i = (1/n)\sum_{j=1}^{n} X_{ij}$ where X_{ij}s are independent random variables (Appendix (A 4)). Then, total number of X_{ij}s increases as $\tilde{N}$ increases. But in our protocol, number of qubit states on which σ_z is measured ($\equiv$ number of X_{ij}s) remains unaltered with the introduction of θ_q. Application of $(\theta_q)_x$s changes only the probability distribution of already present random variables (i.e., measuring σ_z on the qubit states). Also, θ_q is not normally distributed in general. Hence we are going to observe a reduction in variance unlike in the theorem stated above. We are going to sort of convolute (Eq. (6) resembles convolution) two independent probability distributions: $S_1 \to \text{ND} : 0, 1/M$ and $\{\theta_1, \theta_2, ...\} \to \{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$ to obtain $S_1' \to \text{ND} : 0, (1 - (\Delta \cos\theta_q)^2_{p^o_{\theta_q}})/M$.

Applying $(\theta_q)_x$ selectively to each of the M qubits in the ensemble $\mathcal{E}_{2j}$, introduces an insignificant global phase ($\because$ $(\theta_q)_x|\pm\rangle = \exp(\mp i\theta_q/2)|\pm\rangle$), and hence we obtain sample mean $S_2' = S_2 \to \text{ND} : 0, 1/M$. Whereas in the ensemble $\mathcal{E}_{1j}$, applying $(\theta_q)_x$ selectively to each of the M qubits, transforms $|0\rangle, |1\rangle$ to

$$|\theta_q\rangle = (\theta_q)_x|0\rangle = \cos(\theta_q/2)|0\rangle + e^{-i\pi/2}\sin(\theta_q/2)|1\rangle,$$
$$|\theta_{q\perp}\rangle = (\theta_q)_x|1\rangle = -i(\sin(\theta_q/2)|0\rangle + e^{i\pi/2}\cos(\theta_q/2)|1\rangle)$$

respectively. We will neglect global phases. Loosely speaking ($\because \rho_{1j}, \rho'_{1j}$ (defined below) represents the states of ensembles where there is no local control. Hence, rigorous method (which is not essential here) is to formulate the problem in 2^N dimensional Hilbert space (Appendix (C 1))), using local control, applying $(\theta_q)_x$ selectively to each of the M qubits in the ensemble $\mathcal{E}_{1j}$, transforms the state $\rho_{1j} = p_1^+|0\rangle\langle 0| + p_1^-|1\rangle\langle 1|$ to

$$\rho'_{1j} = \sum_q (p_q|\theta_q\rangle\langle\theta_q| + p_{q\perp}|\theta_{q\perp}\rangle\langle\theta_{q\perp}|), \tag{1}$$

where $p_1^\pm = T_1^\pm/M$, $p_q = M_q'(T_1^+, p_{\theta_q})/M$ ($p_{q\perp} = M'_{q\perp}(T_1^-, p_{\theta_q})/M$), M_q' ($M'_{q\perp}$) is the total number of $|\theta_q\rangle$s ($|\theta_{q\perp}\rangle$s). $\sum_q(M_q' + M'_{q\perp}) = M$. Note that probabilities $p_1^\pm, p_{\theta_q}$, and sample mean S_1 are normally distributed random variables which converge to definite values only in the limit $M \to \infty$ i.e., $S_1 \to \text{ND} : 0, 1/M$, $p_1^\pm \to \text{ND} : 1/2, 1/(4M)$ ($\because T_1^\pm \to \text{ND} : M/2, M/4$), and $p_{\theta_q} (= m_q/M) \to \text{ND} : p^o_{\theta_q}, \sigma^2_{m_q}/M^2$ ($\because m_q \to \text{ND} : p^o_{\theta_q}M, \sigma^2_{m_q}$) where m_q is the total number of times $(\theta_q)_x$ is applied, $\sum_q m_q = M$, and $\sigma^2_{m_q} \sim M$ (see Appendix (B 1) for derivation). Hence we need to take care of the variance (however small) present in them. First we will do calculations for given values of p_{θ_q}s and S_1, and later we will integrate the results obtained over all possible values of p_{θ_q}s and S_1 after multiplying by the corresponding weighing factor.

Measuring σ_z selectively on $|\theta_q\rangle$s and $|\theta_{q\perp}\rangle$s is equivalent to tossing differently biased coins. We have random variable means $\langle\sigma_z\rangle_{|\theta_q\rangle} = \cos^2(\theta_q/2) \times +1 + \sin^2(\theta_q/2) \times -1 = \cos\theta_q$, $\langle\sigma_z\rangle_{|\theta_{q\perp}\rangle} = -\cos\theta_q$ and variances $(\Delta\sigma_z)^2_{|\theta_q\rangle} = (\Delta\sigma_z)^2_{|\theta_{q\perp}\rangle} = \sin^2\theta_q$. By applying central limit theorem to independently distributed random variables [3], we obtain effective mean of random variables

$$\mu_{eff} = \sum_q (p_q\langle\sigma_z\rangle_{|\theta_q\rangle} + p_{q\perp}\langle\sigma_z\rangle_{|\theta_{q\perp}\rangle}) = \langle\sigma_z\rangle_{\rho'_{1j}}$$
$$= \sum_q (p_q - p_{q\perp})\cos\theta_q, \tag{2}$$

(see Appendix (A 2) for derivation of μ_{eff}). Note that probabilities $\cos^2(\theta_q/2), \sin^2(\theta_q/2)$ are fixed, whereas probabilities $p_q, p_{q\perp}$ varies over M_1 ensembles, because the numbers $M_q', M'_{q\perp}$ are not fixed. We have $M_q'(T_1^+, p_{\theta_q}) = T_1^+ p_{\theta_q} = p_1^+ m_q (\because p_{\theta_q} = m_q/M)$. Similarly $M'_{q\perp} = T_1^- p_{\theta_q} = p_1^- m_q$. $\Rightarrow \sum_q(M_q' + M'_{q\perp}) = (T_1^+ + T_1^-)\sum_q p_{\theta_q} = (p_1^+ + p_1^-)\sum_q m_q = M$ as required. Hence we obtain

$$p_q = p_1^+ p_{\theta_q}, p_{q\perp} = p_1^- p_{\theta_q} \tag{3}$$

which are nothing but joint probabilities. This makes sense in the light of the fact that $M_q'(T_1^+, p_{\theta_q})$ (and hence p_q) depends on two independent random variables T_1^+ and θ_q. Similarly $p_{q\perp}$ depends on T_1^- and θ_q. Then Eq. (3) follows from Bayes rule. Substituting $p_q, p_{q\perp}$ into μ_{eff} (Eq. (2)) and simplifying we obtain

$$\mu_{eff} = S_1\langle\cos\theta_q\rangle_{p_{\theta_q}} \tag{4}$$

where $\langle\cos\theta_q\rangle_{p_{\theta_q}} = \sum_q p_{\theta_q}\cos\theta_q$.

Effective variance of random variables is given by

$$(\Delta\sigma_z)^2_{eff} = \sum_q (p_q(\Delta\sigma_z)^2_{|\theta_q\rangle} + p_{q\perp}(\Delta\sigma_z)^2_{|\theta_{q\perp}\rangle})$$
$$= \sum_q (p_q + p_{q\perp})\sin^2\theta_q = 1 - \langle\cos^2\theta_q\rangle_{p_{\theta_q}}, \tag{5}$$

where $\langle\cos^2\theta_q\rangle_{p_{\theta_q}} = \sum_q p_{\theta_q}\cos^2\theta_q$ (see Appendix (A 2) for derivation of $(\Delta\sigma_z)^2_{eff}$). Note that $(\Delta\sigma_z)^2_{eff} \neq \langle\sigma_z^2\rangle_{\rho'_{1j}} - \langle\sigma_z\rangle^2_{\rho'_{1j}}$ (see Appendix (A 2) for proof). Then according to central limit theorem, in the large M limit, probability distribution of effective sample mean S_1', for given values of p_{θ_q}s and S_1 (i.e., for given values of m_qs and T_1^+), tends to normal distribution i.e, $S_1' \to \text{ND} : \mu_{eff}, (\Delta\sigma_z)^2_{eff}/M$ [3], Appendix (A 2). Now we shall integrate over all possible values of p_{θ_q}s and S_1 after multiplying the component probability density function by corresponding weighing factor (joint probability), to get the resultant probability density of S_1', as follows:

$$f(S'_1) = \int \prod_{i, i\neq l} \{dp_{\theta_i}(\mathrm{Nd}(p_{\theta_i}) : p^o_{\theta_i}, \sigma^2_{m_i}/M^2)\} dS_1(\mathrm{Nd}(S_1) : 0, 1/M)\left(\mathrm{Nd}(S'_1) : S_1\langle\cos\theta_q\rangle_{p_{\theta_q}}, (1 - \langle\cos^2\theta_q\rangle_{p_{\theta_q}})/M\right), \quad (6)$$

where $(\mathrm{Nd}(x) : \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(x-\mu)^2/(2\sigma^2))$ (i.e., Normal probability density function with mean μ and variance σ^2), $dx(\mathrm{Nd}(x) : \mu, \sigma^2)$ is the probability of obtaining value x of normally distributed random variable x. In Eq. (6) index $i \neq l$ is because of the constraint equation $p_{\theta_l} = 1 - \sum_{j, j\neq l} p_{\theta_j}$. Using this constraint equation we should eliminate p_{θ_l} from $\langle\cos\theta_q\rangle_{p_{\theta_q}}$ and $\langle\cos^2\theta_q\rangle_{p_{\theta_q}}$, before integrating. In Eq. (6) we have product of probabilities because p_{θ_q}s and S_1 are independent random variables, and hence joint probability $P(p_{\theta_1}, p_{\theta_2}, ..., p_{\theta_{l-1}}, p_{\theta_{l+1}}, ..., S_1) = P(p_{\theta_1})P(p_{\theta_2})...P(p_{\theta_{l-1}})P(p_{\theta_{l+1}})...P(S_1) = \prod_{i, i\neq l}\{dp_{\theta_i}(\mathrm{Nd}(p_{\theta_i}) : p^o_{\theta_i}, \sigma^2_{m_i}/M^2)\} dS_1(\mathrm{Nd}(S_1) : 0, 1/M)$. As p_{θ_q}s and S_1 are independent of each other (i.e., no constraint equations in p_{θ_q}s and S_1 i.e., we have to integrate over the entire hyper volume spanned by p_{θ_q}s and S_1), and as we can integrate in any order, we can simply integrate out S_1 from $-\infty$ to ∞ in Eq. (6). Actually $-1 \leq S_1 \leq 1$ but we are integrating from $-\infty$ to ∞. This is because, in the limit $M \to \infty$, $(\mathrm{Nd}(S_1) : 0, 1/M)$ tends to delta function $\delta(S_1 - 0)$ (Appendix (A 1)). Hence both integration intervals will give same result. Advantage of the interval $(-\infty, \infty)$ is, we can get rid of error functions ($\mathrm{erf}(x)$). Integrating we get

$$f(S'_1) = \int \prod_{i, i\neq l} \{dp_{\theta_i}(\mathrm{Nd}(p_{\theta_i}) : p^o_{\theta_i}, \sigma^2_{m_i}/M^2)\}(\mathrm{Nd}(S'_1) : 0, (1 - (\Delta\cos\theta_q)^2_{p_{\theta_q}})/M) \quad (7)$$

where $(\Delta\cos\theta_q)^2_{p_{\theta_q}} = \langle\cos^2\theta_q\rangle_{p_{\theta_q}} - \langle\cos\theta_q\rangle^2_{p_{\theta_q}}$. S_1 was oscillating symmetrically about zero. Hence independent of the value of coefficient of S_1 (in Eq. (6)), resultant mean has vanished. It is like $\langle CS_1\rangle_{\Omega_1} = C\langle S_1\rangle_{\Omega_1} = 0$ where $\Omega_1 = (\mathrm{Nd}(S_1) : 0, 1/M)$.

$f(S'_1)$ in Eq. (7) is the weighted mean of many Gaussians all having mean zero. Hence center/mean of the probability density function $f(S'_1)$ must also be zero. We are going to show that $f(S'_1)$ is also normally distributed. For discrimination we should get $f(S'_1)$ different from $g(S'_2) = (\mathrm{Nd}(S'_2) : 0, 1/M)$ ($\because S'_2 = S_2$). Hence discrimination, if possible, can only be through a change in variance. Mean (effective) is linear in probabilities: $\langle X\rangle = \sum_i P_i x_i$, whereas variance (effective) is a nonlinear (quadratic) function of probabilities: $(\Delta X)^2 = \langle X^2\rangle - \langle X\rangle^2 = \sum_i P_i x_i^2 - (\sum_i P_i x_i)^2$. Here we have treated outcomes x_is as constant, and probabilities P_is as variable (e.g., in measuring σ_z on $|\theta_q\rangle$s, $|\theta_{q\perp}\rangle$s). Hence a change in mean corresponds to a linear change/effect, where as a change in variance corresponds to a nonlinear change/effect. This seems to be justifying the section 'Nonlinear evolution seems to be necessary' in Appendix (C 8).

Swaying of center of Gaussians

Another less rigorous way of arriving at the resultant variance in Eq. (7) i.e., $(1 - (\Delta\cos\theta_q)^2_{p_{\theta_q}})/M$ is the following: When we integrate over S_1 in Eq. (6), there is contribution to resultant variance from following two factors: (1) Swaying of center of Gaussians due to the varying mean i.e., $\mu_{eff} = \langle\cos\theta_q\rangle_{p_{\theta_q}} S_1$. (2) Variance arising from the measurement of σ_z on $|\theta_q\rangle$s, $|\theta_{q\perp}\rangle$s i.e., $(\Delta\sigma_z)^2_{eff}/M$ (Eq. (5)). In Eq. (6) we are integrating with respect to S_1 for given values of p_{θ_q}s. Hence $\langle\cos\theta_q\rangle_{p_{\theta_q}}$ in μ_{eff} can be treated as a constant. Then $\mu_{eff} \to \mathrm{ND} : 0, \langle\cos\theta_q\rangle^2_{p_{\theta_q}}/M$ (using the theorem in Appendix (B 1)). Hence resultant variance is given by $\langle\cos\theta_q\rangle^2_{p_{\theta_q}}/M + (\Delta\sigma_z)^2_{eff}/M = (1 - (\Delta\cos\theta)^2_{p_{\theta_q}})/M$.

Now consider $\theta_q = \theta_0, \forall q$. Then Eq. (7) reduces to $f(S'_1) = (\mathrm{Nd}(S'_1) : 0, (1-0)/M) = g(S'_2)$, hence no discrimination. When $\theta_0 = 0, f(S'_1) = (\mathrm{Nd}(S'_1) : 0, 1/M) = g(S_1)$, as required. This is expected because, by rotating all M qubit states by same angle we are not introducing any new random variable. A random variable is characterized by having non zero variance. But here variance of random variable $\cos\theta_q$: $(\Delta\cos\theta_q)^2_{p_{\theta_q}} = 0$. Hence no randomness. Hence we can not change/disturb/distort/deviate the probability distribution of sample mean (S_1) corresponding to the ensemble $\mathcal{E}_{1j}$. Hence for discrimination, we should take at least $\{\theta_1, \theta_2\} \to \{p^o_{\theta_1}, p^o_{\theta_2}\}$, $\theta_1 \neq \theta_2, p^o_{\theta_q} \neq 0, \forall q$.

B. Specific case

Let us consider the simplest possible case: $\{\theta_1, \theta_2\} \to \{p^o_{\theta_1}, p^o_{\theta_2}\}$. p_{θ_q}s are constrained by $p_{\theta_1} + p_{\theta_2} = 1$. Let $l = 2$ in Eq. (7). Eliminating p_{θ_2} from Eq. (7) we obtain

$$f(S_1') = \int dp_{\theta_1} (\text{Nd}(p_{\theta_1}) : p^o_{\theta_1}, \sigma^2_{m_1}/M^2)(\text{Nd}(S_1') : 0, (1 - p_{\theta_1}(1 - p_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2)/M). \tag{8}$$

Direct evaluation of integral in Eq. (8) is difficult. Note that $f(S_1')$ in Eq. (8) is the weighted mean of many Gaussians all having mean zero. Hence there is no swaying of center of Gaussians unlike in Eq. (6). Also as M is large, it is justifiable to replace the weighing Gaussian with delta function (Appendix (A 1)) as follows

$$f(S_1') \approxeq \int_{p^o_{\theta_1}-\epsilon}^{p^o_{\theta_1}+\epsilon} dp_{\theta_1} \delta(p_{\theta_1} - p^o_{\theta_1})(\text{Nd}(S_1') : 0, (1 - p_{\theta_1}(1 - p_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2)/M)$$
$$= (\text{Nd}(S_1') : 0, (1 - p^o_{\theta_1}(1 - p^o_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2)/M) = (\text{Nd}(S_1') : 0, ((\Delta\sigma_z)^2_{|+\rangle} - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M), \tag{9}$$

where $\epsilon > 0$. Note that this method does not work in Eq. (6), as there is swaying of center of Gaussians. In Eq. (6) if we replace the weighing Gaussian $(\text{Nd}(S_1) : 0, 1/M)$ with $\delta(S_1 - 0)$, then we will be neglecting swaying of center of Gaussians in $(\text{Nd}(S_1') : S_1 ...)$. This results in $f(S_1') = (\text{Nd}(S_1') : 0, \sin^2 0/M)$ for $\theta_q = 0, \forall q$, which is not correct. Hence swaying of center of Gaussians affects/contributes to the resultant/net variance.

We may indirectly evaluate the integral in Eq. (8) as follows: As the weighing function is normally distributed, it is justifiable to assume that $f(S_1')$ will also be normally distributed. As there is no swaying of center of Gaussians, contribution to the net/resultant variance comes only from $(1 - p_{\theta_1}(1 - p_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2)/M$. Hence resultant variance might be the following

$$\Delta S_1'^2 = \int_{-\infty}^{\infty} dp_{\theta_1} (\text{Nd}(p_{\theta_1}) : p^o_{\theta_1}, \sigma^2_{m_1}/M^2)(1 - p_{\theta_1}(1 - p_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2)/M$$
$$= (1 - (p^o_{\theta_1}(1 - p^o_{\theta_1}) - \sigma^2_{m_1}/M^2)(\cos\theta_1 - \cos\theta_2)^2)/M. \tag{10}$$

Hence $f(S_1') = (\text{Nd}(S_1') : 0, \Delta S_1'^2)$. In the large M limit, we can neglect $\sigma^2_{m_1}/M^2$ compared to $p^o_{\theta_1}(1-p^o_{\theta_1})$, and we recover variance in Eq. (9) as required. In Eq. (10) $p^o_{\theta_1} - \epsilon < p_{\theta_1} < p^o_{\theta_1} + \epsilon, \epsilon > 0$. But we are integrating from $-\infty$ to ∞. This is because, in the limit $M \to \infty$, $(\text{Nd}(p_{\theta_1}) : p^o_{\theta_1}, \sigma^2_{m_1}/M^2)$ tends to delta function $\delta(p_{\theta_1} - p^o_{\theta_1})$. Hence both integration intervals will give same result. We prefer $(-\infty, \infty)$ to get rid of $\text{erf}(x)$.

Summary: $\mathcal{E}_i \to$ apply $(\theta_q)_x$ selectively to each of the $N(= M \times M_1)$ qubits in the ensemble $\mathcal{E}_i \to$ do selective σ_z measurement $\to$ we get variance of sample mean of measurement outcomes: $(1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$ if $i = 1$, $1/M$ if $i = 2$.

Discrimination via comparison

As $\Delta S_1'^2 < \Delta S_2'^2 (= 1/M \because S_2' = S_2)$, we can discriminate between the two ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$. However in the large M limit, both $\Delta S_1'^2$ and $\Delta S_2'^2$ becomes smaller. Hence it is easier to discriminate via their ratio i.e., $\lim_{M\to\infty} \Delta S_1'^2/\Delta S_2'^2 \approxeq \lim_{M\to\infty}[(1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M]/[1/M] = 1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}}$, rather than via $\Delta S_i'^2$ alone, $i = 1$ or 2. This is possible in a special case: $\{\theta_1(= 0), \theta_2(= \pi)\} \to \{p^o_0, p^o_\pi\}$. In this case we can obtain both $\Delta S_1'^2$ and $\Delta S_2'^2$ from the given ensemble $\mathcal{E}_i, i = 1$ or 2. If the given ensemble is $\mathcal{E}_1$, then $\Delta S_2'^2 (= \Delta S_1^2 \because S_2' = S_2 \equiv S_1)$ and $g(S_2')(= g(S_1))$ corresponds to $\mathcal{E}_1$ which is the ensemble before applying $(\theta_q)_x$s. If the given ensemble is $\mathcal{E}_2$, then $\Delta S_1'^2$ and $f(S_1')$ corresponds to a virtual ensemble before applying $(\theta_q)_x$s (Appendix (B 2)). Hence we can also discriminate by comparing $f(S_1')$ and $g(S_1)(= g(S_2) = g(S_2'))$.

Also, as $\Delta S_1'^2 < \Delta S_2'^2$ we can say that $f(S_1')$ has more affinity/tendency/inclination towards mean zero of Gaussian, than $g(S_2')$ has.

Intuitively, introduction of a new random variable θ_q should increase the randomness and hence variance. We will try to explain this counter intuitive phenomenon in the following two sections and in Appendix (B 6).

Nonlinearity in action

We will show how nonlinearity is reducing the variance. Consider $\Delta S_1'^2 \approxeq (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M = (\langle\cos\theta_q\rangle^2_{p^o_{\theta_q}} +$

$\langle\sin^2\theta_q\rangle_{p^o_{\theta_q}})/M$ (Eq. (9)). Let $\{\theta_1(=0),\theta_2(=\pi/2)\}\rightarrow\{p^o_0,p^o_{\pi/2}\}$. $\Rightarrow$ $\Delta S_1'^2\approxeq[(p^o_0\cos 0+p^o_{\pi/2}\cos(\pi/2))^2+p^o_0\sin^2 0+p^o_{\pi/2}\sin^2(\pi/2)]/M=(p^{o\,2}_0+p^o_{\pi/2})/M<1/M$. It is counter intuitive, because intuitively if we rotate $\tilde{N}$ states on z-axis (on Bloch sphere) on to y-axis, it is as if we have measured σ_z on $M-\tilde{N}$ number of $|+\rangle$s ($\because\mathcal{E}_{1j}$ can be obtained by measuring σ_z selectively on an ensemble of M identical copies of $|+\rangle$) and $\tilde{N}$ more are to be measured. Hence after measuring $\tilde{N}$ more, we should get back variance $1/M$. A closer look shows that, when we are measuring σ_z first on $M-\tilde{N}$ number of $|+\rangle$s and then on $\tilde{N}$ more, there is only one random variable i.e., σ_z. Hence we are neglecting the way we brought $\tilde{N}$ states on z-axis onto y-axis i.e., via *random* rotations about x-axis by angle θ_qs. This new random variable is reducing the variance. More rigorous explanation is the following: Probabilities (corresponding to the new random variable θ_q) get squared (nonlinear operation) when they enter through $\langle\cos\theta_q\rangle^2_{p^o_{\theta_q}}$, and hence we call this nonlinear channel. This channel corresponds to swaying of center of Gaussians. Where as when probabilities enter through $\langle\sin^2\theta_q\rangle_{p^o_{\theta_q}}$ they come out as such (linear operation). Hence we call this linear channel. This corresponds to measurement of σ_z on $|\theta_q\rangle$s, $|\theta_{q\perp}\rangle$s (see the section 'Swaying of center of Gaussians' (II A) above). Hence there is reduction in variance. For more details see Appendix (B 9).

Now we can justify the result obtained in Eq. (9) as follows: In Eq. (8) there is no swaying of center of Gaussians. Hence p_{θ_q}s are contributing to resultant variance only via linear channel unlike in Eq. (6) where they were contributing via both linear and nonlinear channels. As the channel is linear, in the large M limit, we can simply replace p_{θ_q} with $p^o_{\theta_q}$.

Smoothing out non uniformities

Let $\{\theta_1(=0),\theta_2(=\pi)\}\rightarrow\{p^o_{\theta_1}(=1/2),p^o_{\theta_2}(=1/2)\}$. $\Rightarrow\Delta S_1'^2\approxeq[(1/2\times1+1/2\times-1)^2+1/2\times0+1/2\times0]/M=0$. This is saying that in the large M limit, random flippings removes/smooths out the population difference $T_1^+-T_1^-$ (nonuniformity). Situation here is analogous to the following example: If we rotate a nonuniform (in mass distribution) disc ($\equiv(T_1'^+>T_1'^-)$) at high speed ($\equiv$ random flippings), it starts behaving as if it were uniform ($\equiv(T_1'^+\approxeq T_1'^-)$). For more details see Appendix (B 10).

For explanation using central limit theorem and Shannon entropy, see sections (B 4) and (B 5) respectively in Appendix.

III. MATLAB SIMULATION RESULTS

A. Reduction in variance

MATLAB generates standard uniformly distributed Pseudo Random Numbers (PRN) drawn from the open interval $(0,1)$. Hence we can simulate measuring σ_z on the state $|\chi\rangle=\cos(\theta/2)|0\rangle+e^{i\phi}\sin(\theta/2)|1\rangle$ as follows: if we get a PRN in the interval $(0,\cos^2(\theta/2))$, then it is equivalent to getting outcome $+1$. Else it is equivalent to getting outcome -1. We simulated the case considered in the section 'Smoothing out non uniformities' (II B) for various values of M,M_1. Application of $(\theta_q)_x$s was simulated as described in Appendix (B 1). Here we discriminate by comparing $f(S_1')$ with $g(S_1)(=g(S_2)=g(S_2'))$ (see the section 'Discrimination via comparison' (II B) for more details). Results are plotted in Fig.s (2, 3, 4, 6(c)). There is clear reduction in variance as predicted by theory. $g(S_1)$ is much closer to corresponding theoretical curve, but $f(S_1')$ is not so close to corresponding theoretically predicted curve. Reasons for this lower reduction in variance than theoretically predicted might be the following: (1) Theoretical predictions are in the large M,M_1 limit, where as simulation results are for $M=10^2,10^7,...,M_1=2000,10^4,....$ Reasons given in the section 'How hard it might be to reduce the variance?' in Appendix (B 7) also applies here. (2) Theoretical calculations may not be exact/precise. E.g., we have not evaluated the integral in Eq. (8) exactly. (3) PRNs depends on generator.

B. To look for reduction in population difference

Instead of directly looking for reduction in variance, we can also look for total amount of reduction in population difference, as it is possible to obtain both $f(S_1')$ and $g(S_1)(=g(S_2)=g(S_2'))$ from the given ensemble $\mathcal{E}_i, i=1$ or 2 (II B). Let $S_1'=\Delta S_1'\tilde{S}_1'$ where $\tilde{S}_1'\rightarrow\text{ND}:0,\Delta\tilde{S}_1'^2$. $\Rightarrow$ $S_1'\rightarrow\text{ND}:0,\Delta S_1'^2\Delta\tilde{S}_1'^2$ (using theorem in Appendix (B 1)). But we have $S_1'\rightarrow\text{ND}:0,\Delta S_1'^2$. $\Rightarrow\Delta\tilde{S}_1'^2=1$. Also $S_1'=(T_1'^+-T_1'^-)/M$. Substituting $\theta_1=0,\theta_2=\pi,p^o_0=p^o_\pi=1/2,\sigma^2_{m_1}=M/4$ in Eq. (10) we obtain $\Delta S_1'=1/M$. $\Rightarrow(T_1'^+-T_1'^-)=\tilde{S}_1'$. Similarly we obtain $S_1=(T_1^+-T_1^-)/M=\tilde{S}_1/\sqrt{M}$ where $\tilde{S}_1\rightarrow\text{ND}:0,1$. $\Rightarrow(T_1^+-T_1^-)=\sqrt{M}\tilde{S}_1$. Now consider

$$\begin{aligned}h(r)&=\sum_{i=1}^{r}(|T_{1i}'^+-T_{1i}'^-|-|T_{1i}^+-T_{1i}^-|)\\&=(\langle|\tilde{S}_1'|\rangle_r-\sqrt{M}\langle|\tilde{S}_1|\rangle_r)r,\end{aligned}\tag{11}$$

where $r=1,2,...,M_1$, $\langle|\tilde{S}_1'|\rangle_r=(1/r)\sum_{i=1}^{r}|\tilde{S}_{1i}'|$, $\langle|\tilde{S}_1|\rangle_r=(1/r)\sum_{i=1}^{r}|\tilde{S}_{1i}|$. Note that even though both $\tilde{S}_1$ and $\tilde{S}_1'$ are identically distributed, they are independent. Therefore $\langle|\tilde{S}_1'|\rangle_r\neq\langle|\tilde{S}_1|\rangle_r$ in general. For $r>r^o,r^o\sim100$, we can neglect $\langle|\tilde{S}_1'|\rangle_r$ compared to

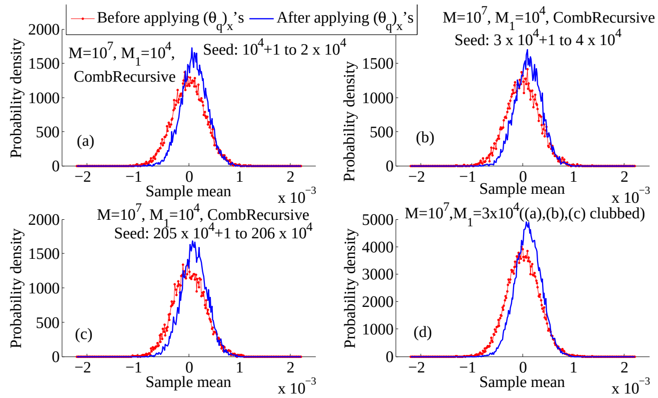

FIG. 2. (Color online) CombRecursive is the PRN generator. 'Seed' is the PRN generator's seed value. We used different seed for each of the M_1 number of sample mean points. Red curve with dot marker is $g(S_1)(= g(S_2') = g(S_2))$, and blue curve with no marker is $f(S_1')$. (a) A_g (= Area under one standard deviation of $g(S_1)$ (i.e., from $S_1 = -\Delta S_1 = -1/\sqrt{M}$ to $S_1 = \Delta S_1$)) is 0.6795 (theoretical prediction in the large M, M_1 limit is 0.6826895). A_f (= Area under $f(S_1')$ corresponding to one standard deviation of $g(S_1)$) is 0.7445 (as predicted by our protocol in the large M, M_1 limit is $\approxeq 1$). Hence there is clear reduction in variance i.e., $\Delta S_1'^2 < \Delta S_1^2 (= \Delta S_2'^2)$. As it is evident from the figure, there is slight offset in the centres of two curves (this may be due to small M_1). If we align them, we get A_g' (= aligned area under one standard deviation of $g(S_1)$) to be 0.6795, and A_f' (= aligned area under $f(S_1')$ corresponding to one standard deviation of $g(S_1)$) to be 0.785. (b) $A_g = 0.6787, A_f = 0.739, A_g' = 0.6793, A_f' = 0.777$. (c) $A_g = 0.685, A_f = 0.7492, A_g' = 0.6828, A_f' = 0.7855$. (d) $A_g = 0.6811, A_f = 0.7442, A_g' = 0.6811, A_f' = 0.7824$.

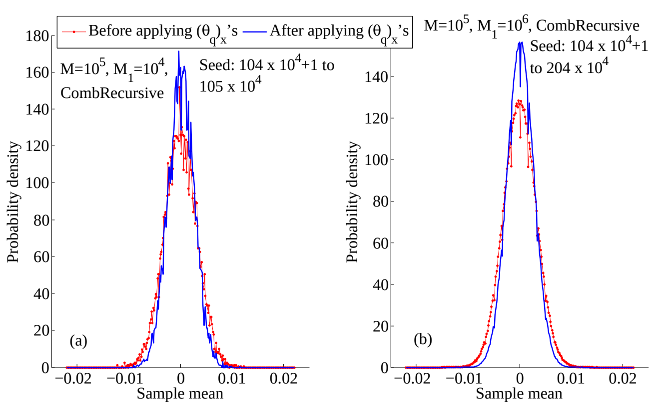

FIG. 3. (Color online) (a) $A_g = 0.6854, A_f = 0.7833$. (b) $A_g = 0.683967, A_f = 0.77997, A_g' = 0.683286, A_f' = 0.780642$.

$\sqrt{M}\langle|\tilde{S}_1|\rangle_r$ as $M \gg 1$. Hence $h(r) \approxeq -\sqrt{M}\langle|\tilde{S}_1|\rangle_r r$. $\langle|\tilde{S}_1|\rangle_r$ is also a random variable with certain mean and a small variance. We can replace $\langle|\tilde{S}_1|\rangle_r$ with a further averaged value $C = \langle\langle|\tilde{S}_1|\rangle_r\rangle$. Then $h(r) \approxeq -\sqrt{M}Cr$, which is a straight line with negative slope. Hence $h(r)$ diverges to $-\infty$ as $r \to \infty$. Now consider the area under the curve $h(r)$, $A(r) = \sum_{x=1}^{r} h(x)\delta x \approxeq -\sqrt{M}Cr(r+1)/2$, where step size $\delta x = 1, r = 1, 2, ..., M_1$. $A(r)$ is a downward opening parabola. Hence area under $h(r)$ also diverges. Corresponding MATLAB simulation results are plotted in Fig.s (5, 6, 7).

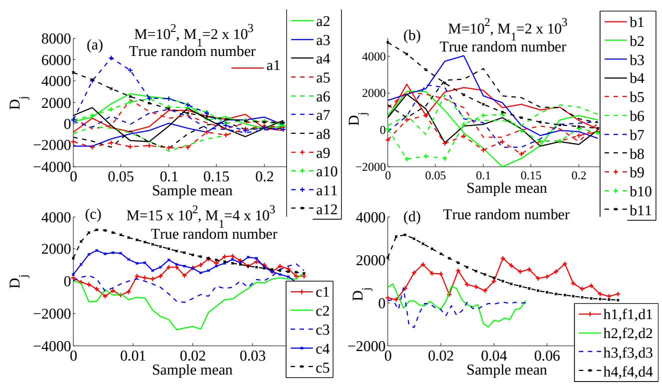

FIG. 4. (Color online) True random number generator [4] was interfaced with MATLAB. D_j is the difference in area under the Gaussians $\times 10^5$ i.e., $D_j = (\sum_{S_1'=-a_j}^{a_j} f(S_1') - \sum_{S_1=-a_j}^{a_j} g(S_1))\delta S \times 10^5$ where δS is the smallest element (step size) on x-axis (sample mean) considered for plotting, and $a_j = j \times \delta S, j = 1, 2, ...$ In the summation, S_1', S_1 increases in steps of δS. Following values indicate respective $\sum_j D_j$ (=area under respective curve divided by δS): a1=1800, a2=12550, a3=-7150, a4=-1200, a5=4550, a6=-11900, a7=650, a8=-12150, a9=-17200, a10=11550, a11=21300, a12=20624T ('T' stands for approximate theoretical prediction, and it has been scaled down by a factor of 10 (approximately) i.e., theoretical curve corresponds to $(D_j)_{theory}/10$. Also the theoretical curve corresponds to $f(S_1') = \mathrm{Nd}(S_1') : 0, \Delta S_1'^2$ where $\Delta S_1'^2$ was taken to be $(0.1^2/M)$ instead of $\Delta S_1'^2 \approxeq (1 - (\Delta \cos\theta_q)^2_{p^o_{\theta_q}})/M = 0/M$. This is for the sake of better comparison of simulation results with theoretical prediction, as simulation was done with small values of M, M_1). b1=15950, b2=2400, b3=16300, b4=1900, b5=7250, b6=9000, b7=5100, b8=19150, b9=-4950, b10=-3900, b11=20624T. c1=14225, c2=-32050, c3=-4725, c4=29450, c5=51683T. In the following (hj,fj,dj) represents the values of $(M, M_1, \sum_k D_k)$ respectively: h1=6e2, f1=3e3, d1=27667. h2=1e3, f2=4e3, d2=-3425. h3=2e2, f3=45e2, d3=-5378. h4=6e2, f4=3e3, d4=34562T.

IV. CONCLUSION

We showed that, if we have local control, we can discriminate between two ensembles, which otherwise (i.e., without local control) cannot be discriminated, as both are maximally mixed. However the origin of nonlinear effect (reduction in variance) which leads to discrimination is not clear. It is interesting to explore whether it is genuine nonlinear effect perhaps due to projective measurement (Appendix (C 8)) or it is just a consequence of statistical data analysis technique.

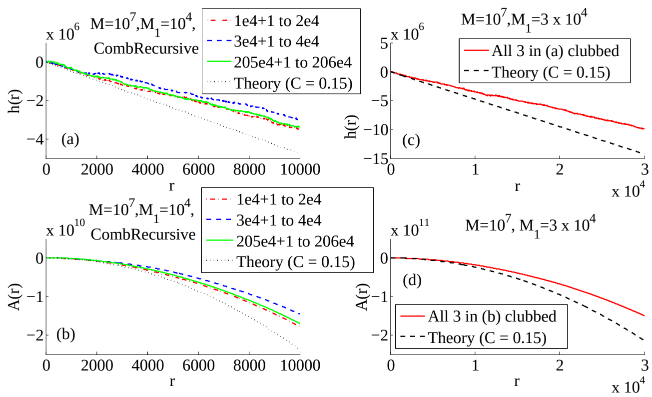

FIG. 5. (Color online) Number with the legend (where 'r'e'n' = r × 10^n, 'r' is a real number and 'n' is an integer) is the PRN generator (CombRecursive)'s seed values. We used different seed for each of the M_1 number of sample mean points. Curves in (b) represents the area under the respective curves in (a). Similarly, the curves in (d) represents the area under the respective curves in (c).

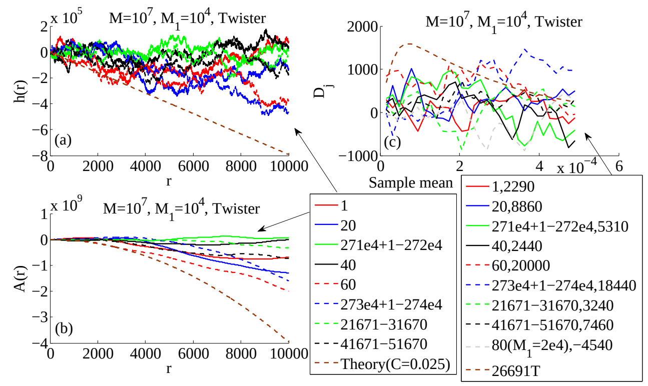

FIG. 6. (Color online) Twister is the PRN generator. (a) and (b) Values in the legend are the PRN generator's seed value(s). Only one seed value implies we have used same seed for all M_1 number of sample mean points. A range (e.g., 271e4+1−272e4 i.e., 2710001 to 2720000) of seed values implies we have used different seed value for each of the M_1 number of sample mean points. Curves in (b) represents the area under the respective curves in (a). (c) D_j and 'T' are defined in Fig. (4). In each legend (except last, which has no seed), first value represents PRN generator's seed value(s) (see above for description), and the second value represents $\sum_j D_j$. Theoretical curve has been scaled down by a factor of 100 (approximately), and $\Delta S_1'^2$ was taken to be $(0.1^2/M)$ to plot the theoretical curve.

ACKNOWLEDGEMENTS

It is a pleasure to thank my supervisor Prof. T S Mahesh for useful discussions and valuable suggestions. I express my deep gratitude to Prof. R Srikanth for very useful discussions and comments which gave me deeper insight. I also thank Prof. L Vaidman, Aravinda, Govind

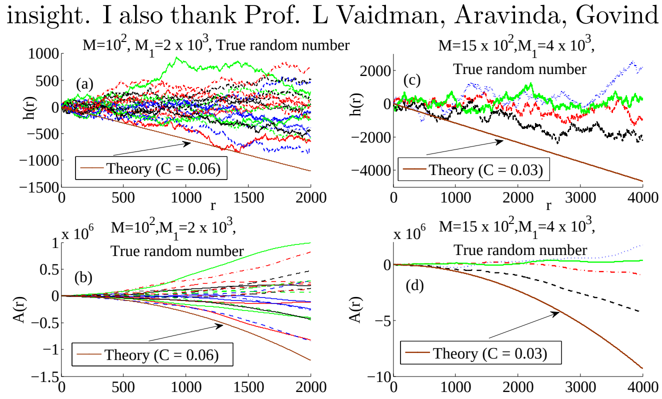

FIG. 7. (Color online) Simulation was done with true random numbers. Curves in (b) represents the area under the respective curves in (a). Similarly the curves in (d) represents the area under the respective curves in (c).

Unnikrishnan, Prof. M Ozawa, Deepak Khurana, Anjusha V S for useful discussions, and valuable comments and suggestions.

[1] *Hamlet (1,5,159-167)*.
[2] C. Cohen Tannoudji et al., *Quantum Mechanics vol. one* (John Wiley and Sons, 2005).
[3] S. Ross, *A first course in probability* (Pearson, 2012).
[4] ComScire, True random number generator .
[5] J. Audretsch, *Entangled Systems: New Directions in Quantum Physics* (Wiley, 2007).
[6] G. M. D'Ariano and H. P. Yuen, Phys. Rev. Lett. **76**, 2832 (1996).
[7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010) cambridge Books Online.
[8] D. Home, *Conceptual Foundations of Quantum Physics* (Springer Science and Business media, 1997).
[9] G. Lüders, Annalen der Physik **15**, 663 (2006).
[10] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge university press, 1989).
[11] J. D. Barrow, P. C. W. Davies, and J. Charles L. Harper, eds., *Science and Ultimate Reality* (Cambridge University Press, 2004) cambridge Books Online.
[12] Y. Aharonov and L. Vaidman, Physics Letters A **178**, 38 (1993).
[13] Y. Aharonov, J. Anandan, and L. Vaidman, Phys. Rev. A **47**, 4616 (1993).
[14] N. D. Hari Dass and T. Qureshi, Phys. Rev. A **59**, 2590 (1999).
[15] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).

Appendix A: Central limit theorem

1. Independent and identically distributed (iid) random variables

Let X be a normal random variable. Then, probability density of X is given by: $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(x-\mu)^2/(2\sigma^2))$. Probability of getting a value between x and $x+dx$ is given by $f(x)dx$. $\Rightarrow \langle X\rangle = \int\limits_{-\infty}^{\infty} xf(x)dx = \mu$, and variance $\langle X^2\rangle - \langle X\rangle^2 = \sigma^2$ [3]. One can verify that $\int\limits_{-\infty}^{\infty} f(x)dx = 1$. $\lim\limits_{\sigma\to 0} f(x) = 0$ for $x \neq \mu$, $\lim\limits_{\sigma\to 0} f(x) = \infty$ for $x = \mu$. Therefore $f(x)$ behaves like a delta function in the limit $\sigma \to 0$. One can show that, if $X \to \mathrm{ND}: \mu, \sigma^2$, then $Y = aX + b \to \mathrm{ND}: a\mu + b, a^2\sigma^2$. Let $a\mu + b = 0, a^2\sigma^2 = 1$. $\Rightarrow Y = (X-\mu)/\sigma$ and it is known as standard or unit normal random variable [3]. Consider $I = \int\limits_{-\infty}^{\infty} f(x)dx$. Put $(x-\mu)/\sigma = y$. $\Rightarrow I = \int\limits_{-\infty}^{\infty} g(y)dy$ where $g(y) = (2\pi)^{-1/2}e^{-y^2/2}$ is the probability density of Y. Note that, even in the limit $\sigma \to 0$, $g(y)$ does not behave like delta function. This is because, in the limit $\sigma \to 0$, it is like mapping an infinite plane ($f(x)$) on to Riemann sphere ($g(y)$).

Consider independent and identically distributed random variables $X_1, X_2, ..., X_n$ having mean $\langle X_i\rangle = \mu, \forall i$ and variance $\Delta X_i^2 = \langle X_i^2\rangle - \langle X_i\rangle^2 = \sigma^2, \forall i$. Sample mean is defined as $S = (1/n)\sum_{i=1}^n X_i$. S has mean $\langle S\rangle = \mu$, and variance $\Delta S^2 = \sigma^2/n$ [3]. Note that even though all X_is have same mean and variance, we cannot take $X_i = X', \forall i$, because they are independent events, and their outcome is random.

Let

$$J = \frac{1}{\sqrt{2\pi}\Delta S}\int\limits_{-\infty}^{c} dS\ \exp\left(\frac{-1}{2\Delta S^2}(S-\mu)^2\right)$$
$$= \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{b} dy\ e^{-y^2/2} = \Omega(b), \text{where } -\infty < c, b < \infty.$$

According to central limit theorem, probability distribution:

$$P\{\frac{1}{\Delta S}(S-\mu) \leq b\} \to \Omega(b),\ \text{as } n \to \infty, \tag{A1}$$

i.e., in the limit $n \to \infty$, probability distribution of random variable S tends to normal (Gaussian) distribution with mean μ and variance σ^2/n [3]. Then using Eq. (A1) we obtain

$$P\{-\epsilon \leq (S-\mu) \leq \epsilon\} = P\{\frac{-\epsilon}{\Delta S} \leq \frac{1}{\Delta S}(S-\mu) \leq \frac{\epsilon}{\Delta S}\}$$
$$\approx \Omega(\frac{\epsilon\sqrt{n}}{\sigma}) - \Omega(\frac{-\epsilon\sqrt{n}}{\sigma}) = \frac{1}{2}(\mathrm{erf}(\frac{\epsilon\sqrt{n}}{\sqrt{2}\sigma}) - \mathrm{erf}(\frac{-\epsilon\sqrt{n}}{\sqrt{2}\sigma}))$$
$$= 2\Omega(\frac{\epsilon\sqrt{n}}{\sigma}) - 1,$$

where $\mathrm{erf}(x) = (2/\sqrt{\pi})\int_0^x dt\ e^{-t^2}$. Approximation in the second line is based on the assumption that n is large. How large n should be for this to be a good approximation depends on probability distribution of X_i.

2. Independently distributed (id) random variables

Consider n biased coins out of which n_1 have mean μ_1' and variance $\sigma_1'^2$, n_2 have mean μ_2' and variance $\sigma_2'^2$,..., n_r have mean μ_r' and variance $\sigma_r'^2$ where $\sum_{j=1}^r n_j = n$. In other words, we have n independent random variables $X_i, i = 1, 2, ..., n$. X_i has mean μ_i and variance $\sigma_i^2, i = 1, 2, ..., n$. $\mu_i = \mu_1', \sigma_i^2 = \sigma_1'^2$ for $i = 1, 2, ..., n_1$, $\mu_i = \mu_2', \sigma_i^2 = \sigma_2'^2$ for $i = n_1+1, n_1+2, ..., n_1+n_2$,..., $\mu_i = \mu_r', \sigma_i^2 = \sigma_r'^2$ for $i = (n_1+n_2+...+n_{r-1}+1), (n_1+n_2+...+n_{r-1}+2), ..., (n_1+n_2+...+n_{r-1}+n_r)$.
If X_is are uniformly bounded (see [3] p399) and

$\sum_{i=1}^{\infty}\sigma_i^2=\infty$ then,

$$P\{n\times\frac{(1/n)\sum_{i=1}^{n}X_i-\sum_{j=1}^{r}(n_j/n)\mu'_j}{\sqrt{\sum_{j=1}^{r}n_j\sigma_j'^2}}\le b\}$$
$$=P\{\frac{S-\sum_{j=1}^{r}p_j\mu'_j}{\sqrt{(1/n)\sum_{j=1}^{r}p_j\sigma_j'^2}}\le b\}$$
$$=P\{\frac{S-\mu'_{eff}}{\sqrt{(\Delta X')^2_{eff}/n}}\le b\}\to\Omega(b),as\ n\to\infty \quad (A2)$$

for given values of n_js (and hence p_js), $j=1,2,...,r$ [3]. Further if n_js are varying (e.g., if n_j is got by throwing n times, a r faced biased dice such that probability of getting its j^{th} face is p_j^o. $p_j\to$ ND : $p_j^o,\sigma_{n_j}^2/n^2$, $\sigma_{n_j}^2\sim n$), then it should be taken into account by integrating over all possible values of n_js (or p_js), after multiplying Nd(S) : $\mu'_{eff},(\Delta X')^2_{eff}/n$ by corresponding weighing factors/weights, to get the final effective probability density function of sample mean. When $\mu'_j=\mu,\sigma_j'^2=\sigma^2,\forall j$ i.e., all n coins have same probability distribution, then Eq. (A2) reduces to Eq. (A1) as required.

No effective single coin: Let X_i be a coin with probability p_{H_i} of getting head and probability $p_{T_i}(=1-p_{H_i})$ of getting tail, $i=1,2,...,n$. Let us assign value $+1$ to head and value -1 to tail. Then we get $\sigma_j'^2=1-\mu_j'^2(\because\ \sigma_i^2=\langle X_i^2\rangle-\langle X_i\rangle^2=1-\mu_i^2)$. Using $\mu_i=p_{H_i}-p_{T_i}$, we can rewrite $\mu'_{eff}=\sum_{j=1}^{r}p_j\mu'_j=p^+_{eff}-p^-_{eff}=\langle X_{eff}\rangle$ for given values of p_js. Now consider $(\Delta X')^2_{eff}=\sum_{j=1}^{r}p_j\sigma_j'^2=\sum_{j=1}^{r}p_j(1-\mu_j'^2)=1-\sum_{j=1}^{r}p_j\mu_j'^2\neq1-(\sum_{j=1}^{r}p_j\mu'_j)^2=1-\langle X_{eff}\rangle^2=\langle X_{eff}^2\rangle-\langle X_{eff}\rangle^2=\Delta X^2_{eff}$. It is true for any given set of values of p_js including $p_j=n_j/n=1/n,\forall j$ i.e., $r=n$, and $p_j=n_j/n=(n/c)/n=1/c,\forall j$ where c is an integer. $\Rightarrow r=c$. Hence concept of single effective coin is not correct with respect to effective variance unless $r=1$. This is because when we are tossing r different types of independent coins, we are sort of convoluting r different probability distributions, which is absent in the case of tossing only one type of effective coin (X_{eff}). Hence the two situations are different.

No effective state: Let $X_i=\sigma_z$ measured on the state $|P_i\rangle=\sqrt{P_i}|0\rangle+\sqrt{1-P_i}|1\rangle,i=1,2,...,n$. $P_i=P'_1$ for $i=1,2,...,n_1$, $P_i=P'_2$ for $i=n_1+1,n_1+2,...,n_1+n_2$,..., $P_i=P'_r$ for $i=(n_1+n_2+...+n_{r-1}+1),(n_1+n_2+...+n_{r-1}+2),...,(n_1+n_2+...+n_{r-1}+n_r)$. Then we get $\sigma_j'^2=1-\mu_j'^2(\because\ \sigma_i^2=\langle X_i^2\rangle-\langle X_i\rangle^2=1-\mu_i^2)$. We can rewrite $\mu'_{eff}=\sum_{j=1}^{r}p_j(2P'_j-1)=\sum_{j=1}^{r}p_j\mathrm{Tr}(\sigma_z\rho'_j)=\mathrm{Tr}(\sigma_z\rho'_{eff})=\langle\sigma_z\rangle_{\rho'_{eff}}$ for given values of p_js, where $\rho'_{eff}=\sum_{j=1}^{r}p_j\rho'_j,\rho'_j=|P'_j\rangle\langle P'_j|$. Now consider $(\Delta X')^2_{eff}=\sum_{j=1}^{r}p_j\sigma_j'^2=\sum_{j=1}^{r}p_j(1-\mu_j'^2)=1-\sum_{j=1}^{r}p_j\mu_j'^2\neq1-(\sum_{j=1}^{r}p_j\mu'_j)^2=1-\langle\sigma_z\rangle^2_{\rho'_{eff}}=\langle\sigma_z^2\rangle_{\rho'_{eff}}-\langle\sigma_z\rangle^2_{\rho'_{eff}}$. Hence concept of effective state ρ'_{eff} is not correct with respect to effective variance unless $r=1$. For further justification see Appendix (A 3).

Also note that, even though j^{th} type of coin is thrown only $n_j(\le n)$ times, $j=1,2,...,r$, variance of effective/resultant sample mean is calculated considering all n measurements together i.e., $\Delta S^2_{eff}=(\Delta X')^2_{eff}/n=\sum_{j=1}^{r}p_j(\sigma_j'^2/n)$. But $\Delta S^2_{eff}\neq\sum_{j=1}^{r}p_j(\sigma_j'^2/n_j)$ $\because$ it gives inconsistent result as follows: Assume $\Delta S^2_{eff}=\sum_{j=1}^{r}p_j(\sigma_j'^2/n_j)$. Then for $\mu'_j=\mu,\sigma_j'^2=\sigma^2,\forall j$ we get $\Delta S^2_{eff}=\sigma^2\sum_{j=1}^{r}p_j/n_j=r\sigma^2/n\neq\sigma^2/n$ unless $r=1$.

3. Justifying the formula $(\Delta X')^2_{eff}=\sum_j p_j\sigma_j'^2$ derived using central limit theorem, in section (A 2)

Consider $\rho=p_0|0\rangle\langle0|+p_1|1\rangle\langle1|$ where $p_0=T_+/M,p_1=T_-/M,T_++T_-=M,M\to\infty$ or large. Let $T_\pm$ be fixed i.e., value of $T_\pm$ will not vary over, say, M_1 repetitions of the experiment. Consider measuring σ_z. σ_z has no variance with respect to $|0\rangle$ and $|1\rangle$ as both are its eigenkets. Hence σ_z must have zero variance even with respect to ρ, a statistical/classical mixture of $|0\rangle$ and $|1\rangle$ (i.e., has no superposition/coherence/off diagonal terms). In other words, effective/resultant variance of σ_z measurement on ρ must also be zero. This is because in ρ we have exactly T_+ number of $|0\rangle$s and exactly T_- number of $|1\rangle$s. Hence however many times (say, M_1 times) we repeat σ_z measurement on identically prepared states ρ, we always get exactly T_+ number of $+1$ outcomes and T_- number of -1 outcomes. Hence no variance. Hence sample mean $S=(T_+-T_-)/M=p_0-p_1$ and S has zero variance.

Now consider $\langle\sigma_z\rangle_{|0\rangle}=1,\langle\sigma_z\rangle_{|1\rangle}=-1$. $\Rightarrow\mu_{eff}=p_0\langle\sigma_z\rangle_{|0\rangle}+p_1\langle\sigma_z\rangle_{|1\rangle}=p_0-p_1=\langle\sigma_z\rangle_\rho$. $(\Delta\sigma_z)^2_{|0\rangle}=\langle\sigma_z^2\rangle_{|0\rangle}-\langle\sigma_z\rangle^2_{|0\rangle}=0$, $(\Delta\sigma_z)^2_{|1\rangle}=\langle\sigma_z^2\rangle_{|1\rangle}-\langle\sigma_z\rangle^2_{|1\rangle}=0$. Let us define $(\Delta\sigma_z)^2_{eff}=p_0(\Delta\sigma_z)^2_{|0\rangle}+p_1(\Delta\sigma_z)^2_{|1\rangle}$. Substituting previous results we obtain $(\Delta\sigma_z)^2_{eff}=0$. Hence sample mean $S=(T_+-T_-)/M=p_0-p_1=\mu_{eff}$, and S has variance $\Delta S^2=(\Delta\sigma_z)^2_{eff}/M=0$ as required.

Instead let us define $(\Delta\sigma_z)^2_{eff}=\langle\sigma_z^2\rangle_\rho-\langle\sigma_z\rangle^2_\rho$. Then substituting from above we obtain $(\Delta\sigma_z)^2_{eff}=1-(p_0-p_1)^2=1-\mu^2_{eff}$. Then sample mean S has variance $\Delta S^2=(\Delta\sigma_z)^2_{eff}/M=(1-(p_0-p_1)^2)/M\neq0$ in general. Hence this definition of effective variance is not correct. Moreover it predicts that $S\to$ ND : $p_0-p_1,(1-(p_0-p_1)^2)/M$. But from Eq. (A2) it is evident that, as the condition $\sum_{i=1}^{\infty}\sigma_i^2=\infty$ is not satisfied (actually $\sum_{i=1}^{\infty}\sigma_i^2=0$ as $(\Delta\sigma_z)^2_{|0\rangle}=(\Delta\sigma_z)^2_{|0\rangle}=0$), S cannot be normally distributed. This definition contains correlation term $p_0p_1\langle\sigma_z\rangle_{|0\rangle}\langle\sigma_z\rangle_{|1\rangle}$. But measurement of σ_z on $|0\rangle$ and $|1\rangle$ are uncorrelated. Here no convolution (sort of) of two independent probability distributions unlike in previous case. Here two independent events i.e., measuring σ_z on $|0\rangle$ and measuring σ_z on $|1\rangle$, has been coalesced into one single event i.e., measuring

σ_z on $\sqrt{p_0}|0\rangle + \sqrt{p_1}|1\rangle$ (see below).

This also seems to show the linear nature of mean (hence $\mu_{eff} = \langle\sigma_z\rangle_\rho$) and nonlinear nature of variance (hence $(\Delta\sigma_z)^2_{eff} \neq \langle\sigma_z^2\rangle_\rho - \langle\sigma_z\rangle^2_\rho$).

However if we measure σ_z on $|\psi\rangle = \sqrt{p_0}|0\rangle + \sqrt{p_1}|1\rangle$, then $\langle\sigma_z\rangle_{|\psi\rangle} = p_0 - p_1$ and $(\Delta\sigma_z)^2_{|\psi\rangle} = \langle\sigma_z^2\rangle_{|\psi\rangle} - \langle\sigma_z\rangle^2_{|\psi\rangle} = 1-(p_0-p_1)^2$. Here sample mean, $S(=(T'_+ - T'_-)/M) \to$ ND : $p_0 - p_1, (1-(p_0-p_1)^2)/M$ in the limit $M \to \infty$, where $T'_\pm$ is the total number of ± 1 outcomes.

4. Independent normally distributed random variables

If $X_i \to$ ND : μ_i, σ_i^2 then $Z = \sum_{i=1}^{\tilde{N}} X_i \to$ ND : $\sum_{i=1}^{\tilde{N}} \mu_i, \sum_{i=1}^{\tilde{N}} \sigma_i^2$ where X_is are normally distributed independent random variables [3]. This result is based on the following convolution relation: $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$ where f_X, f_Y, f_{X+Y} are probability density functions of $X, Y, X+Y$ respectively [3]. Let $\tilde{N} = 2$, $X_i = (1/n)\sum_{j=1}^n X_{ij}$ where X_{ij} is an independent random variable with mean μ_{ij} and variance σ_{ij}^2, $i = 1, 2$. Then using central limit theorem (Eq. (A2)) we get $X_i \to$ ND : μ_i, σ_i^2 where $\mu_i = \sum_{j=1}^n (1/n)\mu_{ij}, \sigma_i^2 = (1/n)\sum_{j=1}^n (1/n)\sigma_{ij}^2$ in the limit $n \to \infty$, $i = 1, 2$. Then $Z = X_1 + X_2 = (1/n)(\sum_{j=1}^n X_{1j} + \sum_{j=1}^n X_{2j})$.

Appendix B: Ensemble (with local control) picture

1. Distribution of $T_1^\pm, p_1^\pm$, and p_{θ_q}

We have $S_1 \to$ ND : $0, 1/M$ where $S_1 = (T_1^+ - T_1^-)/M, T_1^+ + T_1^- = M$. $\Rightarrow T_1^\pm = M(1 \pm S_1)/2$. Theorem: If $X \to$ ND : μ, σ^2 then $Y(= aX + b) \to$ ND : $a\mu + b, a^2\sigma^2$ [3]. Using this we obtain $T_1^\pm \to$ ND : $M/2, M/4$. We have defined $p_1^\pm = T_1^\pm/M$. Again using the above theorem we obtain $p_1^\pm \to$ ND : $1/2, 1/(4M)$.

Let $\{\theta_1, \theta_2\} \to \{p^o_{\theta_1}, p^o_{\theta_2}\}$. It is equivalent to measuring σ_z on $|\zeta\rangle = \sqrt{p^o_{\theta_1}}|0\rangle + \sqrt{p^o_{\theta_2}}|1\rangle$, and if the outcome is +1 apply θ_1, else apply θ_2. Then sample mean $S(= (m_1 - m_2)/M) \to$ ND : $2p^o_{\theta_1} - 1, (\Delta\sigma_z)^2_{|\zeta\rangle}/M$ where $m_1(m_2)$ is the number of $+1(-1)$ outcomes. $(\Delta\sigma_z)^2_{|\zeta\rangle} = 1 - \langle\sigma_z\rangle^2_{|\zeta\rangle}, m_1 + m_2 = M$. $\Rightarrow m_j(= M(1 + (-1)^{j+1}S)/2) \to$ ND : $p^o_{\theta_j}M, \sigma^2_{m_j}$ where $\sigma^2_{m_j} = (\Delta\sigma_z)^2_{|\zeta\rangle}M/4, j = 1, 2$. $\Rightarrow p_{\theta_j}(= m_j/M) \to$ ND : $p^o_{\theta_j}, \sigma^2_{m_j}/M^2$ (using the theorem stated above). We assume that this is true even when $j > 2$, where $\sigma^2_{m_j} \sim M$. But this assumption is not important, because $j = 2$ is necessary and sufficient for our protocol.

2. Knowing the state of each of the N qubits in the ensemble $\mathcal{E}_1$ exactly

Let $\{\theta_1(=0), \theta_2(=\pi)\} \to \{p^o_0, p^o_\pi\}$. Then if the given ensemble is $\mathcal{E}_1$, then no collapse of the qubit state upon measuring σ_z after applying $(\theta_q)_x$s. Sequence of $(\theta_q)_x$s (say, function F) maps $\mathcal{E}_1$ to $\mathcal{E}'_1$ (i.e., $\mathcal{E}'_1 = F(\mathcal{E}_1)$) where $\mathcal{E}'_i$ is the ensemble got by applying $(\theta_q)_x$ selectively to each of the N qubits in the ensemble $\mathcal{E}_i, i = 1, 2$. As we have local control, we can know the state of each of the N qubits in the ensemble $\mathcal{E}'_1$ exactly, by σ_z measurement. Then working backward using the sequence of $(\theta_q)_x$s (i.e., inverse mapping, $\mathcal{E}_1 = F^{-1}(\mathcal{E}'_1)$), we can know exactly what was the state of each of the N qubits in the given ensemble $\mathcal{E}_1$. Note that if the given ensemble is $\mathcal{E}_2$, then we cannot know the state of each of the N qubits, as there will be collapse upon σ_z measurement. We can know only that the given ensemble was $\mathcal{E}_2$.

If the given ensemble is $\mathcal{E}_1$, then sample mean $S_1 \to$ ND : $0, 1/M$ corresponds to before applying $(\theta_q)_x$s, where as the sample mean $S'_1 \to$ ND : $0, (1 - 4p^o_0(1 - p^o_0))/M$ (for $\{\theta_1(=0), \theta_2(=\pi)\} \to \{p^o_0, p^o_\pi\}$) corresponds to after applying $(\theta_q)_x$s. However, if the given ensemble is $\mathcal{E}_2$, then sample mean $S'_2 \to$ ND : $0, 1/M$ corresponds to after applying $(\theta_q)_x$s (note that in the previous case this probability distribution was present before applying $(\theta_q)_x$s). From the ensemble $(\mathcal{E}''_2)$ got by measuring σ_z selectively on each of the N qubits in the ensemble $\mathcal{E}'_2$, if we work backward via the sequence of $(\theta_q)_x$s that we had applied (i.e., the mapping $\mathcal{E}'''_2 = F(\mathcal{E}''_2)$), we obtain a virtual ensemble $(\mathcal{E}'''_2)$ which corresponds to sample mean $\to$ ND : $0, (1 - 4p^o_0(1 - p^o_0))/M$.

3. Equivalence of states $|+\rangle$ and $|-\rangle$ with respect to σ_z measurement

Consider an ensemble in the state

$$\rho = p_+|+\rangle\langle+| + p_-|-\rangle\langle-|. \tag{B1}$$

Let $p_\pm = \lim_{M\to\infty} T_\pm/M$ where $T_+ + T_- = M$. Then, $\mu_{eff} = p_+\langle\sigma_z\rangle_{|+\rangle} + p_-\langle\sigma_z\rangle_{|-\rangle} = 0 = \langle\sigma_z\rangle_\rho$, and hence independent of p_+, p_-. $(\Delta\sigma_z)^2_{eff} = p_+(\Delta\sigma_z)^2_{|+\rangle} + p_-(\Delta\sigma_z)^2_{|-\rangle} = 1(\because (\Delta\sigma_z)^2_{|+\rangle} = (\Delta\sigma_z)^2_{|-\rangle} = 1)$, again independent of p_+, p_-. $\Rightarrow$ variance of effective sample mean $\Delta S^2_{eff} = (\Delta\sigma_z)^2_{eff}/M = 1/M$. Hence all the results are same as measuring σ_z on M number of $|+\rangle$s or $|-\rangle$s. Hence the states $|+\rangle$ and $|-\rangle$ are equivalent as far as σ_z measurement outcomes are concerned. In other words, probabilities of getting outcomes $+1, -1$ upon measuring σ_z, is same in both the states: $|+\rangle, |-\rangle$. Hence the two states are equivalent with respect to σ_z measurement outcomes.

4. Explanation using central limit theorem

We can explain reduction in population difference (and hence variance) via central limit theorem as follows: Let $\{\theta_1(=0), \theta_2(=\pi)\} \to \{p^o_0(=1/2), p^o_\pi(=1/2)\}$. Consider the case where $T^+_1 = M$. Then it is obvious that getting $T'^+_1 \gg T'^-_1$ or $T'^+_1 \ll T'^-_1$ is very unlikely, where as getting $T'^+_1 \approxeq T'^-_1$ is very likely. Consider the case where $T'^+_1 = M$. There is only one sequence of $(\theta_q)_x$s which can give this i.e., all θ_qs being 0 radians. But there are very large number of sequences of $(\theta_q)_x$s which do not give $T'^+_1 = M$. Hence according to central limit theorem, probability of getting $T'^+_1 = M$ tends to zero in the large M limit. This extreme case clearly explains how and why there is reduction in population difference (and hence variance of sample mean) (i.e., $|T'^+_1 - T'^-_1| \ll |T^+_1 - T^-_1|$) upon applying $(\theta_q)_x$s. Similar thing happens even when $T^+_1 \gg T^-_1$ or $T^+_1 \ll T^-_1$, and it is also obvious. What is not obvious is the prediction that similar thing happens even when $T^+_1 > T^-_1$ or $T^+_1 < T^-_1$. This may be explained as follows: The result below Eq. (5) is due to central limit theorem. Hence the results in Eq.s (8-10) (with $\{\theta_1(=0), \theta_2(=\pi)\} \to \{p^o_0(=1/2), p^o_\pi(=1/2)\}$) are also a consequence of central limit theorem. Hence in the spirit of central limit theorem we can say that, total number of possible sequences of $(\theta_q)_x$s which transforms $|T^+_1 - T^-_1|(=\sqrt{M}|\tilde{S}_1|)$ to $|T'^+_1 - T'^-_1|(=|\tilde{S}'_1|)$, is much greater than sum of other possible sequences which do not do this transformation i.e., probability of this transformation tends to one in the large M limit, where $|\tilde{S}_1|, |\tilde{S}'_1|$ varies between 0 and 10 (approximately) (see section (III B)).

5. Explanation using Shannon entropy

We can also explain the phenomenon of reduction in population difference (and hence variance) in terms of entropy as follows: As the population difference $|T^+_1 - T^-_1|$ increases towards M, the sequence of $|0\rangle$s, $|1\rangle$s (in the ensemble $\mathcal{E}_{1j}$) becomes more and more ordered and hence entropy decreases. More rigorously consider Shannon entropy $H = -\sum_{i=1}^{2} P_i \log_2 P_i$, $\sum_i P_i = 1$, where P_1 is the probability of occurrence of $|0\rangle$, and P_2 that of $|1\rangle$ [5]. Then, $|T^+_1 - T^-_1| = M$ corresponds to $P_1 = 1$ or $P_2 = 1$ where $P_1 = T^+_1/M, P_2 = T^-_1/M$. $\Rightarrow H = 0$ i.e., minimum entropy configuration. When we introduce a new random variable θ_q such that $\{\theta_1(=0), \theta_2(=\pi)\} \to \{p^o_0(= 1/2), p^o_\pi(= 1/2)\}$ via the application of $(\theta_q)_x$s, naturally it will try to make the sequence of $|0\rangle$s, $|1\rangle$s disordered, which is typical of any random operation. This corresponds to increasing entropy. In other words, $|T'^+_1 - T'^-_1| = 0$ corresponds to $P_1 = P_2 = 1/2$ where $P_1 = T'^+_1/M, P_2 = T'^-_1/M$. $\Rightarrow H = 1$ i.e., maximum entropy configuration. Hence application of $(\theta_q)_x$s increases disorder (entropy) and hence reduces the population difference $|T^+_1 - T^-_1|$ towards zero. Hence $|T'^+_1 - T'^-_1| \approxeq 0$ in the large M limit.

6. Why the reduction in variance?

Variance of random variable σ_z is $(\Delta\sigma_z)^2_{|\psi\rangle} \leq 1$. For given M, sample mean has variance $(\Delta\sigma_z)^2_{|\psi\rangle}/M \leq 1/M$. Hence the ensemble $\mathcal{E}_1$, already corresponds to maximum possible variance ($\because S_1 \to$ ND : $0, 1/M$). Hence, introduction of a new random variable θ_q [which does not increase the number of qubit states $(= M \times M_1)$(see the section 'Motivation' (II A))] can only decrease the variance. Not rigorously we can write $Y = S_1 + \cos\theta_q$ where $S_1 \to$ ND : $0, 1/M = ND : 0, (\Delta\sigma_z)^2_{|+\rangle}/M$, and $\{\theta_1, \theta_2, ...\} \to \{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$. New random variable $\cos\theta_q$ has variance $(\Delta\cos\theta_q)^2_{p^o_{\theta_q}}$. Then $Y \to$ ND : $0, ((\Delta\sigma_z)^2_{|+\rangle} - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$. Variance here seems to have pseudo-Riemannian metric signature $(+, -)$.

7. How hard it might be to reduce the variance?

If the given ensemble is $\mathcal{E}_2$, then even if M, M_1 are not very large, we obtain sample mean S'_2 which is at least approximately normally distributed with mean 0 and variance $1/M$. This is because it is simple i.e., it is not a complicated weighted mean of very large number of Gaussians, whose resultant is ND : $0, 1/M$. But $f(S'_1)$ in Eq. (6) is a complicated weighted mean of very large number of Gaussians. There seems to be no simple way of obtaining $f(S'_1) \approxeq \text{Nd}(S'_1) : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$. This may be shown as follows: We have $\langle\sigma_z\rangle_{|\delta\rangle} = \cos\delta$, $(\Delta\sigma_z)^2_{|\delta\rangle} = \sin^2\delta$ where $|\delta\rangle = \cos(\delta/2)|0\rangle + \sin(\delta/2)|1\rangle$. In the large M limit, sample mean $\to$ ND : $\cos\delta, \sin^2\delta/M$. Let ND : $\cos\delta, \sin^2\delta/M =$ ND : $0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$. $\Rightarrow \delta = \pi/2$. But $\sin^2(\pi/2) \neq 1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}}$ in general. This is also justified by the fact that there is no effective state with respect to effective variance (Appendix (A 2)). Hence it seems ND : $0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$ can be got only as the resultant of complicated weighted mean of a large number of different Gaussians as given in Eq. (6). Even to realize one of the component Gaussians in Eq. (6) (e.g., $\text{Nd}(S'_1) : S_1\langle\cos\theta_q\rangle_{p_{\theta_q}}, (1 - \langle\cos^2\theta_q\rangle_{p_{\theta_q}})/M$ for given values of S_1, p_{θ_q}s and θ_qs) we require large set of M measurements each. Hence to obtain the resultant probability density $f(S'_1) \approxeq \text{Nd}(S'_1) : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$, it seems, M_1 should be really large. As we are working with perfect Gaussians, we have implicitly assumed that M is very large. When M, M_1 are small, we may get $f(S'_1) \approxeq g(S'_2)(=$ ND : $0, 1/M)$, and hence no discrimination. Only when M, M_1 are really large, we obtain $f(S'_1) \approxeq \text{Nd}(S'_1) : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$.

We are trying to distort/change/deviate from ND : $0, 1/M$. At low M, M_1 the deviation/distortion is not large enough to give rise to observable effect in the form of reduction in variance below $1/M$. Only as M, M_1 increases, deviations accumulate drop by drop and results in appreciable reduction in variance. Reduction in variance is the resultant of many operations viz., application of $(\theta_q)_x$s, measurement of σ_z on different states: $|\theta_q\rangle$s, $|\theta_{q\perp}\rangle$s.

8. We cannot directly convolute probability distribution of S_1 (i.e., ND $: 0, 1/M$) with that of θ_q (i.e., p_{θ_q})

Reasons are the following: (1) Effective/resultant random variable, S_1', is not a simple straight forward function of S_1 and θ_q i.e., no simple straight forward relation connecting them, even though S_1 and θ_q are independent. (2) Both S_1' and S_1 correspond to M number of qubit states ($\equiv X_{ij}$s in Appendix (A 4)), which is unlike in Appendix (A 4). (3) S_1 is continuous where as θ_q is discrete.

9. Nonlinearity in action

(Continued from the main text (II B)) Asymmetry (nonlinear and linear) in the two channels might be due to the following reason: It seems it is more difficult to change the variance via swaying of center (it requires undulating the entire Gaussian), than via throwing out/in a few sample mean points symmetrically about the center, with center fixed. In the ensemble $\mathcal{E}_{1j}$ approximately $Mp^o_{\pi/2}$ of the states were rotated on to y-axis. Hence there is reduction in swaying of center of Gaussians which in turn reduces resultant/net variance as $\cos(\pi/2) = 0$. When we measure σ_z on the states on y-axis there is positive contribution to the resultant variance as $\sin(\pi/2) = 1$. Similarly the states on z-axis will contribute positively to net/resultant variance via swaying of center of Gaussian, as $\cos 0 = 1$. But measurement of σ_z on the states on z-axis does not contribute to resultant variance, as $\sin 0 = 0$. Because of nonlinear nature (with respect to variance) of swaying of center of Gaussian, sum of contributions to variance from nonlinear and linear channels fails to reach back to $1/M$.

10. Smoothing out nonuniformities

(Continued from main text (II B)) We saw that in the large M limit, random flippings removes/smooths out the population difference $T_1^+ - T_1^-$ (nonuniformity). Situation here is analogous to the following example: Consider a small metallic sphere of mass m tied to a string of length L. At time $t = 0$ it is on z-axis pivoted at the origin. Its center of mass (COM) lies at $z = L(\equiv (T_1^+ - T_1^-))$. Now rotate the sphere about x-axis at high speed ($\equiv$ random flipping i.e., applying $(\theta_q)_x$s). Its dynamic COM lies at $(\sum_i m_i\vec{r}(t_i))/\sum_i m_i = \langle\vec{r}(t)\rangle_{\delta t} = 0(\equiv (T_1'^+ - T_1'^- \approxeq 0))$ where $\vec{r}(t_i)$ is the position vector at time t_i, and δt is a small time interval. Note that we cannot make $T_1'^+ - T_1'^- = 0$ always, because as evident from Eq. (10), even when $\{\theta_1(= 0), \theta_2(= \pi)\} \to \{p^o_0(= 1/2), p^o_\pi(= 1/2)\}$, and even in the large M limit, variance is non zero (however small). This will cause $T_1'^+ \neq T_1'^-$. The analogy used here is just for illustration.
Similarly if we spin a nonuniform (in mass distribution) disc at high speed, it starts behaving as if it were uniform. Fast spinning smooths out nonuniformities in mass distribution. Even if the angular speed varies slightly over time ($\equiv$ variance σ^2_{m1}/M^2 of $p^o_{\theta_1}$), still nonuniformities will be smoothed out.
Consider a diesel engine generator's fly wheel whose COM is slightly offset from its axis of rotation (due to some manufacturing defect or something else). Then one can observe decrease in vibrations with the increase in rotational speed of the fly wheel ($\because$ dynamic COM shifts towards the axis of rotation as rotational speed increases). This effect is also observable in vehicles. Situation here is analogous to this.
When $T_1^+ > T_1^-$ more number of $|0\rangle$s will be rotated by $\theta_2(= \pi)$ than $|1\rangle$s, there by equalizing the population difference. When $T_1^- > T_1^+$ its the other way round, there by equalizing the population difference again.

11. Rate of change of variance

We can also study the rate at which variance changes with increasing M. Let $R_{var} = |d(variance)/dM|$. Hence $R_{var}(S_1') \approxeq |d((1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M)/dM| = |-(1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M^2|$ (using Eq.(9)), where as $R_{var}(S_1) = |d(1/M)/dM| = |-1/M^2|$. Negative sign indicates that variance decreases with increasing M. Clearly $R_{var}(S_1) > R_{var}(S_1')$. Hence variance of S_1 decreases faster than that of S_1' with increasing M in the large M limit. Using this also we can discriminate.

12. Rate of change of sample mean

Following may not be observable at small M for the reason mentioned in Appendix (B 7). Consider $|S_1'| = |\tilde{S}_1'/M| \sim 1/M$ (using section (III B)). $\Rightarrow d(|S_1'|)/dM \sim -1/M^2$. Similarly $|S_1| = |\tilde{S}_1/\sqrt{M}| \sim 1/\sqrt{M}$ (using section (III B)). $\Rightarrow d(|S_1|)/dM \sim -1/(2M^{3/2})$. Taking the absolute values, we see that, $|S_1|$ decreases at a faster rate than $|S_1'|$ for a small increase in M. Using this also we can discriminate.

13. Order of integration does not matter

In Eq. (6) even if we had integrated first with respect to p_{θ_q}s, there would have been oscillations symmetrically about a given S_1. Next when we integrate with respect to S_1, all the previous oscillations will oscillate symmetrically about zero. Hence resultant mean vanishes as in Eq. (7).

Appendix C: Single copy picture

1. As a nonorthogonal state discrimination problem

Linear and unitary nature of quantum mechanics forbids cloning an unknown state chosen from a set of nonorthogonal states [6]. Hence, given a single copy of a pure state chosen from a set of nonorthogonal states, we cannot both *exactly* and with probability tending to one (*deterministically*) know the given state (unknown state). However, exact but probabilistic nonorthogonal state discrimination is possible. E.g., given a single copy of $|0\rangle$ or $(|0\rangle + |1\rangle)/\sqrt{2}$, we can know the unknown state exactly but only probabilistically using POVM measurement [7]. Natural next question to ask is the following: Is deterministic but inexact nonorthogonal state discrimination possible? To answer this question, we consider the following problem: There are two sets: $\mathcal{F}_1 = \{|0\rangle^{\otimes N}, |0\rangle^{\otimes N-1}|1\rangle, ..., |1\rangle^{\otimes N}\}$, and $\mathcal{F}_2 = \{|+\rangle^{\otimes N}, |+\rangle^{\otimes N-1}|-\rangle, ..., |-\rangle^{\otimes N}\}$. $\mathcal{F}_i$ is a complete set of orthonormal basis states in 2^N dimensional (2^N-D) Hilbert space, $i = 1, 2$. E.g., for $N = 2$, $\mathcal{F}_1 = \{|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle\}$, and $\mathcal{F}_2 = \{|+\rangle|+\rangle, |+\rangle|-\rangle, |-\rangle|+\rangle, |-\rangle|-\rangle\}$. Let $|\phi_{ij}\rangle \in \mathcal{F}_i$, $i = 1, 2$, $j = 1, 2, ..., 2^N$. Even though $|\langle\phi_{1j}|\phi_{2k}\rangle|$ tends to zero in the limit $N \to \infty$, $|\phi_{1j}\rangle$ can never become perfectly orthogonal to $|\phi_{2k}\rangle$, because the set $\mathcal{F}_i$ is already complete, $i = 1, 2$. Hence $\mathcal{F}_1$ and $\mathcal{F}_2$ together constitute a set of nontrivial nonorthogonal states (Appendix (C 2)). Alice gives Bob, a *single copy* of $|\phi_{ij}\rangle$ chosen with probability $1/2^N$ (i.e., all the states are equally likely to be chosen) from $\mathcal{F}_i$, $i = 1$ or 2. Hence $|\phi_{1j}\rangle$ ($|\phi_{2j}\rangle$) is nothing but the renormalized post measurement state of measuring σ_z (σ_x) selectively (i.e., locally) on each of the N qubits in the state $|+\rangle^{\otimes N}$ ($|0\rangle^{\otimes N}$), $|0\rangle = (|+\rangle + |-\rangle)/\sqrt{2}$. Alice tells Bob the way she chose the state from one of $\mathcal{F}_1, \mathcal{F}_2$, but she do not tell him exactly from which set she chose the state. Hence Bob is aware of $\mathcal{F}_1, \mathcal{F}_2$, and Alice's state choosing procedure. Bob has a single copy of the unknown state $|\phi_{ij}\rangle, i = 1$ or 2. We are going to show that, in the limit $N \to \infty$, even though Bob cannot know the unknown state exactly, still he can know deterministically whether it was chosen from $\mathcal{F}_1$ or $\mathcal{F}_2$ (and hence it is deterministic but inexact nonorthogonal state discrimination). This requires selective random x-rotations (unitary evolutions) and projective measurements. Variance of sample mean of σ_z measurement outcomes is reduced if the unknown state belongs to $\mathcal{F}_1$, else variance remains unaltered.

The protocol that we are going to describe is feasible, because there is no necessity of interacting one qubit with the other. Purpose of working in 2^N-D Hilbert space is to gain addressability and control over each of the N qubits (which is not possible with nonselective ensemble measurement (Appendix (C 5, C 10))), so that we can selectively apply random x-rotation to each of the N qubits in the unknown state. Another purpose is to get exact expression for variance of sample mean. Let $N = M \times M_1$, $M, M_1 \to \infty$, so that Bob can obtain M_1 sample mean points where each point corresponds to M measurements. Note that it is sufficient if M, M_1 are large enough to obtain approximately normally distributed sample mean. Hence we rewrite

$$|\phi_{ij}\rangle = |\psi_{ij1}\rangle|\psi_{ij2}\rangle...|\psi_{ijM_1}\rangle, \ i = 1, 2, \ j = 1, 2, ..., 2^N, \tag{C1}$$

where $|\psi_{ijk}\rangle$ is a M qubit state, $k = 1, 2, ..., M_1$. Hence $|\psi_{1jk}\rangle$ ($|\psi_{2jk}\rangle$) is nothing but the renormalized post measurement state of measuring σ_z (σ_x) selectively on each of the M qubits in the state $|+\rangle^{\otimes M}$ ($|0\rangle^{\otimes M}$). In the ensemble (with local control) picture just replace $\mathcal{E}_i$ with $|\phi_{ij}\rangle$, and $\mathcal{E}_{ik}$ with $|\psi_{ijk}\rangle$, $i = 1, 2$, $j = 1, 2, ..., 2^N$, $k = 1, 2, ..., M_1$. Then we get all the calculations in single copy picture i.e., as a nonorthogonal state discrimination problem.

2. Nontrivial nonorthogonal states

Consider the set $\{|0\rangle, |w\rangle\}$ where $|w\rangle = \cos((\pi - \epsilon)/2)|0\rangle + \sin((\pi - \epsilon)/2)|1\rangle, \epsilon \to 0$. $\langle 0|w\rangle = \cos((\pi - \epsilon)/2)$. We call this a set of trivial nonorthogonal states, because by measuring σ_z, we can discriminate between them deterministically. Instead consider the set $\mathcal{F} = \{|\phi_{11}\rangle, |\phi_{12}\rangle, ..., |\phi_{12^N}\rangle, |\phi_{21}\rangle, |\phi_{22}\rangle, ..., |\phi_{22^N}\rangle\}$. Alice gives Bob, a single copy of $|\phi_{ij}\rangle$ chosen with probability $1/2^N$ from $\mathcal{F}_i$, $i = 1$ or 2. $|\langle\phi_{1j}|\phi_{2k}\rangle| = 1/2^{N/2}$. Then, even in the limit $N \to \infty$, by direct measurement of whatever observable (e.g., an observable whose nondegenerate eigenkets are $|\phi_{1j}\rangle$s), Bob cannot say deterministically, whether the given state was chosen from $\mathcal{F}_1$ or $\mathcal{F}_2$. This is because $|\phi_{2k}\rangle$ is a state of equal superposition of all the states in $\mathcal{F}_1$. Hence we call $\mathcal{F}$ a set of nontrivial nonorthogonal states.

3. Density matrix formulation

In density matrix formulation, Bob's unknown state is given by:

$$\rho_i = \sum_{j=1}^{2^N} \frac{1}{2^N} |\phi_{ij}\rangle\langle\phi_{ij}| = \frac{\mathbb{1}_{2^N}}{2^N}, i = 1 \text{ or } 2 \tag{C2}$$

where $\mathbb{1}_n$ is $n \times n$ identity matrix. Note that ρ_i represents the state of a single copy of one of $|\phi_{ij}\rangle$s, $j = 1, 2, ..., 2^N$, which Bob has got, taking into consideration the probability ($1/2^N$) with which he obtains it. But ρ_i does not represent the state of an ensemble with no local control. ρ_i represents the state of an ensemble with local control.

Mixedness of ρ_i represents Bob's ignorance about the single copy of the state he has got. Hence it can be purified by selective projective measurement unlike in nonselective ensemble measurement (see Appendix (C 4) for more details).

Hence the fact that both ρ_1 and ρ_2 are maximally mixed does not imply that it is not possible to discriminate between them. It just says that, the amount of ignorance is same in both cases. $|\phi_{1j}\rangle$ is a random sequence of $|0\rangle$s and $|1\rangle$s, where as $|\phi_{2j}\rangle$ is a random sequence of $|+\rangle$s and $|-\rangle$s. Purification of ρ_i by selective projective measurement implies gain of knowledge about the unknown state. Consider the single copy of one of $|\phi_{1j}\rangle$s which Bob has got i.e., ρ_1. Bob applies $(\theta_q)_x$ and then measures σ_z selectively on each of the N qubits in the unknown state $|\phi_{1j}\rangle$. By this ρ_1 is projected onto a pure state, because post measurement state is completely known to Bob. In a special case $(\{\theta_1(=0), \theta_2(=\pi)\} \rightarrow \{p_0, p_\pi\})$ he can even know the exact state given to him by Alice (see Appendix (B 2)). Hence *maximally mixed state, ρ_1, has been transformed into a pure state via selective projective measurement* (same thing happens with ρ_2 as well, but here Bob cannot know the exact state given to him by Alice). Hence projection is nonunitary as no unitary operation can purify. Note that, as $\rho_i, i = 1$ or 2, represents the state of a single copy, we should not sum over all possible outcomes in finding the post σ_z measurements state (see Appendix (C 4, C 5) for justification). We showed that application of $(\theta_q)_x$s reduces variance of sample mean of σ_z measurement outcomes, only if the unknown state is $|\phi_{1j}\rangle$ i.e., we got $\Delta S_1'^2 < \Delta S_2'^2$. Hence Bob was able to discriminate between ρ_1 and ρ_2 after purification by σ_z measurement. It is shown in [8] that, one can also discriminate between states similar to ρ_1 and ρ_2 using deterministic nonlinear evolution (but there density matrices represent the states of ensembles with no local control).

4. Mixed state of a closed single quantum system can be purified by projective measurement

Consider the following game: Alice projectively measures σ_z on a single copy of $|+\rangle$, and gives the post measurement state to Bob. She tells Bob that she has measured σ_z on a single copy of $|+\rangle$, but she do not tell him, her measurement outcome. Now Bob should find out what was her outcome and hence the state given to him. For Bob, state of the single qubit given to him is the following: $\rho_B = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \mathbb{1}/2$. Mixedness is a measure of his ignorance about the state. No unitary operation can purify it because $U\rho_B U^\dagger = \mathbb{1}/2$. Now Bob projectively measures σ_z. Of course there is no collapse upon measurement, as his pre-measurement state was an eigenstate of σ_z. If he gets $+1$ outcome then he comes to know that Alice had got outcome $+1$ and the state given to him was $|0\rangle$. In density matrix language, $\rho_B \rightarrow \rho_B^0/Tr(\rho_B^0) = |0\rangle\langle 0|$ where $\rho_B^0 = |0\rangle\langle 0|\rho_B|0\rangle\langle 0|$, which is renormalised pure state. $|0\rangle\langle 0|$ is linear but nonunitary operator which does projection. Note that before Bob measuring σ_z, the state of the single qubit to him was maximally mixed, which represents his complete ignorance about the state. But after measurement, the state became pure, which indicates his gain of knowledge about the state. Similarly, if he gets the outcome -1, he comes to know that the state given to him was $|1\rangle$ i.e., $\rho_B \rightarrow \rho_B^1/Tr(\rho_B^1) = |1\rangle\langle 1|$ where $\rho_B^1 = |1\rangle\langle 1|\rho_B|1\rangle\langle 1|$. For Bob post measurement state is not the following: $\rho_B^f = \rho_B^0 + \rho_B^1 = \mathbb{1}/2$, because *he is no more ignorant of the state*. Hence ρ_B^f corresponds to nonselective ensemble measurement but not to a single copy measurement. Hence in case of single copy measurement we should not sum over all possibilities (also see the section 'Single copy verses nonselective ensemble measurement' in Appendix (C 5)). This is nothing but orthogonal state discrimination.

However if Bob measures $A = a_0\Pi_0 + a_1\Pi_1, \Pi_0 = |0\rangle\langle 0|, \Pi_1 = |1\rangle\langle 1|$, nonselectively on an ensemble of qubits initialised in the state $\rho_{in} = \mathbb{1}_2/2$, then post measurement state of the full ensemble is given by $\rho_f = \sum_i \Pi_i\rho_{in}\Pi_i = \mathbb{1}_2/2$. Note that it is true even if one of a_is is zero. Hence even if we measure an arbitrary observable, post measurement state remains maximally mixed. Hence measurement can not purify the state unlike in single copy measurement.

5. Single copy versus nonselective ensemble measurement

Consider a single copy of the state $|m\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$. If we measure σ_z and obtain the outcome $+1$, then the normalized state immediately after measurement is given by $\Pi_0|m\rangle/\sqrt{\langle m|\Pi_0|m\rangle} = |0\rangle$ where $\Pi_0 = |0\rangle\langle 0|$ [2]. In density matrix formulation it is given by $\Pi_0\rho_m\Pi_0/Tr(\Pi_0\rho_m\Pi_0) = |0\rangle\langle 0|$, a pure state, where $\rho_m = |m\rangle\langle m|$. Similarly if the outcome is -1, post measurement state turns out to be $|1\rangle$.

Now consider an ensemble of identical copies of $|m\rangle$. If we measure σ_z nonselectively (i.e., nonselective ensemble measurement like in NMR spin ensembles, where we cannot address and control each and every qubit in the ensemble separately (i.e., no local control), but we can only address and control them as a whole (i.e., only global control) i.e., same radio frequency (rf) pulse is applied to all of them. Also qubits are continuously interacting with environment), then the post measurement unnormalized state of the subensemble corresponding to $+1$ outcome is given by $\rho_0 = \Pi_0\rho_m\Pi_0 = \cos^2(\theta/2)|0\rangle\langle 0|$. State of the sub ensemble corresponding to -1 outcome is given by $\rho_1 = \Pi_1\rho_m\Pi_1 = \sin^2(\theta/2)|1\rangle\langle 1|$ where $\Pi_1 = |1\rangle\langle 1|$. Normalised state of the full ensemble is given by $\rho_f = \rho_0 + \rho_1 = \cos^2(\theta/2)|0\rangle\langle 0| + \sin^2(\theta/2)|1\rangle\langle 1|$ which is mixed [9].

6. A linear operator can clone at the most two nonorthogonal states in 2-D Hilbert space

Consider a linear operator L such that $L|0\rangle|0\rangle = |0\rangle|0\rangle$ and $L|1\rangle|0\rangle = |1\rangle|1\rangle$. Assume $L(\alpha|0\rangle + \beta|1\rangle)|0\rangle = (\alpha|0\rangle + \beta|1\rangle)(\alpha|0\rangle + \beta|1\rangle)$. Substituting the above transformations we obtain the following solutions: $\alpha = 1, \beta = 0$ or $\alpha = 0, \beta = 1$ or $\alpha = 0, \beta = 0$.

Now instead consider the following transformations: $L|0\rangle|0\rangle = |0\rangle|0\rangle$, $L|+\rangle|0\rangle = |+\rangle|+\rangle$. Then we obtain the following constraint equations: $L_{00,00} = 1, L_{00,10} = 1/\sqrt{2} - 1, L_{10,00} + L_{10,10} = 1/\sqrt{2}, L_{01,00} + L_{01,10} = 1/\sqrt{2}, L_{11,00} + L_{11,10} = 1/\sqrt{2}$ where $L_{00,10} = \langle 00|L|10\rangle$ etc. It has infinitely many solutions. This and previous results together imply that L can at the most clone two nonorthogonal states. We assume that a similar result holds even for N qubit state i.e., a linear operator can at the most clone 2^N nonorthogonal states.

7. A linear operator in 2-D Hilbert space can map at the most two nonorthogonal states into orthogonal states

Consider the following transformation: $L|0\rangle = |0\rangle, L|+\rangle = |1\rangle$. $\Rightarrow L_{11} = 1, L_{12} = -1, L_{21} = 0, L_{22} = \sqrt{2}$ where L_{ij}s are matrix elements of L. $\Rightarrow L|1\rangle = -|0\rangle + \sqrt{2}|1\rangle \neq |0\rangle$ and $L|-\rangle = \sqrt{2}|0\rangle - |1\rangle \neq |1\rangle$. Hence L can map at the most two nonorthogonal states into orthogonal states. Hence we require nonlinear evolution to map $|0\rangle, |1\rangle$ to $|0\rangle$, and $|+\rangle, |-\rangle$ to $|1\rangle$. This corresponds to deterministic but inexact discrimination (because, after mapping if we measure σ_z, and if we get $+1$ outcome, then we come to know only that the given state was $|0\rangle$ or $|1\rangle$. Similar thing with -1 outcome). Hence deterministic inexact discrimination also seems to be demanding nonlinear evolution. Of course there may be ways other than the mapping technique that we are using here, which can do deterministic inexact discrimination with linear evolution and measurement. We assume that a similar result holds even in 2^N-D Hilbert space i.e., a linear operator in 2^N-D Hilbert space can at the most map 2^N nonorthogonal states into orthogonal states. This is also justified by the fact that, maximum possible number of mutually orthogonal states in 2^N-D Hilbert space is 2^N. Hence a map similar to that described in 2-D Hilbert space may require nonlinear evolution.

8. Nonlinear evolution seems to be necessary

$|\langle\phi_{1j}|\phi_{2k}\rangle| = 1/2^{N/2} \to 0$ (but never becomes exactly equal to zero) in the limit $N \to \infty$. However, $|\phi_{1j}\rangle$ cannot be orthogonal to $|\phi_{2k}\rangle$, because $\mathcal{F}_i$ is already complete, $i = 1, 2$. A linear operator can clone at the most 2^N nonorthogonal states in 2^N-D Hilbert space (Appendix (C 6)). Hence we can discriminate between them exactly and deterministically, via tomography. Another method is the following: In 2^N-D Hilbert space, a linear operator can at the most map 2^N nonorthogonal states into orthogonal states (Appendix (C 7)). Then, by projectively measuring an observable whose eigenkets (with nondegenerate eigenvalue) are these orthogonal states, we can discriminate between 2^N number of nonorthogonal states, both exactly and deterministically. However here we have $2^N(|\phi_{1j}\rangle\text{s}) + 2^N(|\phi_{2j}\rangle\text{s})$ number of nonorthogonal states. Hence $|\phi_{1j}\rangle$s and $|\phi_{2j}\rangle$s together constitute a set of 2^{N+1} number of nonorthogonal states. Even if we discard those states among $|\phi_{1j}\rangle$s and $|\phi_{2j}\rangle$s, probability of getting which tends to zero in the limit of $N \to \infty$, still one can easily show that the set of nonorthogonal states will have much more than 2^N number of states. In this case, as shown in Appendix (C 7), even deterministic but inexact discrimination between $|\phi_{1j}\rangle$s and $|\phi_{2j}\rangle$s may require nonlinear evolution. If it is true, then our protocol establishes that projective measurement is a genuine probabilistic nonlinear evolution, as we are able to say whether the unknown state $|\phi_{ij}\rangle, i = 1$ or 2, belongs to $\mathcal{F}_1$ or $\mathcal{F}_2$, using projective measurement. Also it makes sense to say that, nonlinear effect (reduction in variance) could be a consequence of some nonlinear evolution. Hence projective measurement might be nonlinear. Discrimination is inexact because, if the unknown state is $|\phi_{1j}\rangle$, and $\{\theta_1(=0), \theta_2(=\pi)\} \to \{p_0^o, p_\pi^o\}$, then we can exactly know the unknown state (Appendix (B 2)). However if the unknown state is $|\phi_{2j}\rangle$, then for arbitrary $(\theta_q)_x$s, we can know only that the unknown state belongs to $\mathcal{F}_2$. However if projective measurement is also linear and unitary evolution (many worlds interpretation [10, 11]), then the above argument, which shows that nonlinear evolution might be necessary for discrimination, is not the most general, because our protocol shows that discrimination is possible via random unitary evolutions and projective measurements.

Following are the likely sources of nonunitary, nonlinear evolution in our protocol: (1) Nonunitary, nonlinear effects are already present in the unknown state $|\phi_{1j}\rangle$, in the form of variance of S_1. This came via the collapse (a nonunitary, nonlinear evolution [8]) that occurred during Alice measuring σ_z. Through selective random rotations about x-axis on Bloch sphere, we are reducing this variance (a nonlinear change, as variance is a quadratic function) in case of unknown state $|\phi_{1j}\rangle$. (2) Selectively applying $(\theta_q)_x$ and then measuring σ_z on each of the N qubits, projects ρ_i (Eq. (C2)) onto a pure state, which is a nonunitary transformation.

9. Is not 2^M-D Hilbert space sufficient?

Random rotations about x-axis on Bloch sphere applied selectively to each of the M qubits in the unknown state $|\psi_{ijk}\rangle$ varies over M_1 sets (i.e., $k = 1, 2, ..., M_1$). Hence, random rotation operator $U_l = U_{\theta_{c_{1l}}} \otimes U_{\theta_{c_{2l}}} \otimes$

$... \otimes U_{\theta_{c_{Ml}}}$ where $U_{\theta_{c_{nl}}} = \exp(-i\theta_{c_{nl}}\sigma_x/2) = (\theta_q)_x$, $c_{nl} = 1, 2, ...,$ and $n = 1, 2, ..., M$, applied to r^{th} set is in general different from that applied to t^{th} set, $r, t = 1, 2, ..., M_1$. Even in the simplest case i.e., $c_{nl} = 1, 2$, we have 2^M different U_ls possible. Hence it is not like applying same rf field to all the qubits in an NMR spin ensemble. Here we require addressability and control over each of the MM_1 number of qubits. Hence Bob must work in 2^{MM_1}-D Hilbert space.

10. What it is not

Alice is not measuring σ_z nonselectively on an ensemble of identical copies of $|+\rangle$s. Similarly she is also not measuring σ_z nonselectively on an ensemble of identical copies of $|0\rangle$s. If she does nonselective ensemble measurement, then Bob obtains $\rho_1 = |0\rangle\langle 0|/2 + |1\rangle\langle 1|/2 = \mathbb{1}_2/2$ in the former case, and $\rho_2 = |+\rangle\langle +|/2 + |-\rangle\langle -|/2 = \mathbb{1}_2/2$ in the latter case, instead of that given in Eq. (C2). Then, ρ_i cannot be purified by subsequent σ_z measurement, $i = 1, 2$. Also Bob cannot apply random rotations about x-axis on Bloch sphere ($(\theta_q)_x$) selectively to each of the qubits in the ensemble. Hence variance cannot be reduced. Hence discrimination not possible.

11. Power of single quantum system

To build a portable quantum computer we need to manipulate single quantum system. NMR being an ensemble with no local control, is considered only as a test bed for quantum information protocols, but not a candidate for ultimate quantum computer. Hence, NV center, SQUID, trapped ion, cold atoms, where we can manipulate single quantum systems, are considered as ultimate candidates to build a quantum computer. Similarly, in our protocol, Bob is able to discriminate because he works with a single quantum system in the state $|\phi_{ij}\rangle, i = 1$ or 2.

12. Making sense of probability amplitudes of a single quantum system

Given a single copy of one of the following two states: $|m\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle$, $|m_\perp\rangle = \sin(\theta/2)|0\rangle + e^{i(\phi+\pi)}\cos(\theta/2)|1\rangle$, where θ, ϕ (but not the state) is known a priori, it is possible to measure probability amplitudes via protective measurement [12–14]. This shows that probability amplitude (and hence expectation value of an arbitrary observable) is also a property of a single quantum system, but not just of an ensemble. We cannot do it for arbitrary unknown θ, ϕ, as we do not have control/access to nonlinear evolution i.e., because of no-cloning theorem [15], [6].