Can two differently prepared mixed quantum-ensembles be discriminated via measurement variance ?

C S Sudheer Kumar[1, *]

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

Alice prepares two large qubit-ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ in the following states: She individually prepares each qubit of $\mathcal{E}_1$ in $|0\rangle$ or $|1\rangle$, the eigenstates of Pauli-z operator σ_z, depending on the outcome of an unbiased coin toss. Similarly, she individually prepares each qubit of $\mathcal{E}_2$ in $|+\rangle$ or $|-\rangle$ the eigenstates of σ_x. Bob, who is aware of the above states preparation procedures, but knows neither which of the two is $\mathcal{E}_1$ nor Alice's outcomes of coin tosses, needs to discriminate between the two maximally mixed ensembles. Here we argue that Bob can partially purify the mixed states ($\mathcal{E}_1$, $\mathcal{E}_2$), using the information supplied by central limit theorem. We will show that, subsequently Bob can discriminate between ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ by individually rotating each qubit state about the x-axis on Bloch sphere by a *random angle*, and then projectively measuring σ_z. By these operations, the variance of sample mean of σ_z measurement outcomes corresponding to the ensemble $\mathcal{E}_1$ gets reduced. On the other hand, qubit states in $\mathcal{E}_2$ are invariant under the x-rotations and therefore the variance remains unaltered. Thus Bob can discriminate between the two maximally mixed ensembles. We analyse the above problem both analytically as well as numerically, and show that the latter supports the former.

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

I. INTRODUCTION

A state vector is associated with a unique pure quantum system. However, a density matrix can be associated with two different mixed quantum ensembles where each is a mixture of different (complete) set of pure quantum (basis) states. It is generally believed that it is not possible to discriminate between such mixed ensembles, as density matrix is assumed to specify all the properties of a given quantum ensemble [1–3]. However there are discussions in the literature contrary to this belief [4]. Fano defines the state via the way it is prepared [5]. The purpose of the present paper is to re-examine this problem and show that discrimination is indeed possible.

The problem of discriminating between two ensembles of qubits where each is in a maximally mixed state but having different physical content i.e., having been prepared using two different procedures, is interesting as well as important. Alice prepares an ensemble $\mathcal{E}_1$ ($\mathcal{E}_2$) of N qubits in the following state: She tosses an unbiased coin and if the outcome is Head, then she prepares the j^{th} qubit in the state $|0\rangle$ ($|+\rangle$), else she prepares it in the state $|1\rangle$ ($|-\rangle$), $j = 1, 2, ..., N$, where $|0\rangle, |1\rangle$ are eigenkets of Pauli-z matrix σ_z with eigenvalues $+1, -1$ respectively, and $|\pm\rangle = [|0\rangle \pm |1\rangle]/\sqrt{2}$. N is sufficiently large enough to obtain at least approximately normally distributed sample mean. Alice gives Bob one of the two ensembles and Bob needs to find which of the two. Bob knows how Alice prepared the state of the qubits in the ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$, but he do not know Alice's outcomes of coin tosses. We are going to show that, even though Bob cannot know the exact state of each and every qubit in the given ensemble, still he can know whether it was the ensemble $\mathcal{E}_1$ or ensemble $\mathcal{E}_2$, provided the ensembles have certain extra features (i.e., Bob should be able to individually address and manipulate each qubit in the given ensemble, like Alice in the above case) than conventional ensembles.

A plausible implication of our discrimination protocol might be that Alice may signal Bob, provided they have pre agreed upon the time of communication. By this, Bob knows a priori that Alice is definitely going to measure on her entangled qubit states at a particular time instant. Hence it does not violate the no-signaling principle [6–8] per se, as the latter excludes any kind of a priori knowledge.

The rest of the introduction serves as a descriptive account of the notations and defining terms in developing our approach to the problem considered here, and gives a brief overview of the discrimination protocol.

We define following two kinds of ensembles: If Alice and Bob can (cannot) individually address and control each qubit in the ensemble, and if the qubits are non-interacting, we call it an individual control ensemble i.e., IC-ensemble (a collective control ensemble i.e., CC-ensemble). The above state preparation by coin tossing corresponds to IC-ensembles. A few more examples of IC-ensembles are the following: (a) Non-interacting qubits fixed at definite sites of a 2D crystal lattice (see Fig. (1)), where each qubit can be labeled with spatial coordinates and hence can be manipulated individually. (b) Consider two polarization directions of a single photon as a qubit. Alice prepares it either in $|0\rangle$ or $|1\rangle$ ($|+\rangle$ or $|-\rangle$) state, and sends it to Bob via optical fiber. She repeats this procedure N times. Here Alice and Bob can la-

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

bel each qubit with temporal coordinates, and obviously they can manipulate individual qubits at different points of time. An example of CC-ensemble is the following: Pseudo pure part of liquid state NMR qubit-ensemble [9] where individual labeling and hence individual manipulation of qubits is not possible. Here one can address and control all the qubits together i.e., only collective/mass control is possible.

Suppose in the above state preparation by coin tossing, Alice somehow looses information about label/address of individual qubits and also her ability to manipulate individual qubits, after she had prepared the ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$. Consequently Bob also looses individual addressability as well as his ability to manipulate individual qubits. Then $\mathcal{E}_1$ and $\mathcal{E}_2$ becomes CC-ensembles of qubits, and consequently both are in maximally mixed state (i.e., $\mathbb{1}_2/2$ where $\mathbb{1}_n$ is $n \times n$ identity matrix) even for Alice. This is because of the following two reasons: (1) Alice just knows that a given qubit in $\mathcal{E}_1$ ($\mathcal{E}_2$) is either in the state $|0\rangle$ or $|1\rangle$ ($|+\rangle$ or $|-\rangle$), but she is not sure of its exact state, as she has lost its label. (2) Alice has lost individual control. For Bob, qubits in the given CC-ensemble $\mathcal{E}_i$ are in the state $\mathbb{1}_2/2, i = 1$ or 2. In CC-ensembles, only collective/nonselective operations (unitary evolutions, measurements etc.) are possible. Hence every qubit state has to evolve under the same unitary operator i.e., $U\mathbb{1}_2U^\dagger/2 = \mathbb{1}_2/2$. Further if Bob measures some observable nonselectively on the state $\mathbb{1}_2/2$, then the post measurement state is also $\mathbb{1}_2/2$ (Appendix (C 5)). Hence Bob cannot purify his mixed state $\mathbb{1}_2/2$, and hence he obtains same mean and variance of sample mean of σ_z measurement outcomes from both the CC-ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$. Hence Bob cannot discriminate between CC-ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$.

However if $\mathcal{E}_1$ and $\mathcal{E}_2$ are IC-ensembles (see Fig. (1)), like in the above state preparation by coin tossing, then for Alice, qubits in them are in a pure state $|\phi_{1j}\rangle$ and $|\phi_{2j}\rangle$ respectively. This is because, she knows the exact state of each of the N qubits in the IC-ensemble $\mathcal{E}_i, i = 1, 2$. However for Bob, qubits in the given IC-ensemble $\mathcal{E}_i, i = 1$ or 2, are in a maximally mixed state $\mathbb{1}_{2^N}/2^N$. Mixedness represents his ignorance about whether the given ensemble is $\mathcal{E}_1$ or $\mathcal{E}_2$, and also his ignorance about the exact state of individual qubits in the given IC-ensemble. Bob partially purifies his mixed state by applying central limit theorem to his knowledge about Alice's states preparation procedures. Then Bob carries out following operations individually on each of the N qubit states in the given IC-ensemble $\mathcal{E}_i$ (i.e., $|\phi_{ij}\rangle$): (1) Rotates the qubit state about x-axis on Bloch sphere by an angle θ_q where θ_q is a *random number* (random x-rotation). (2) Measures σ_z projectively. By these operations, variance of sample mean of σ_z measurement outcomes corresponding to the IC-ensemble $\mathcal{E}_1$ gets reduced (see orange narrow Gaussian in the top row of Fig. (1)). This happens due to sort of convolution between probability distribution of θ_q and that of sample mean before random x-rotations. However if the given IC-ensemble is $\mathcal{E}_2$, then random x-rotations introduces an insignificant global phase to the qubit states $|+\rangle, |-\rangle$ in it. Hence the variance of sample mean remains unaltered (see green broad Gaussians in the bottom row of Fig. (1)). Hence Bob can discriminate between IC-ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ via variance of sample mean.

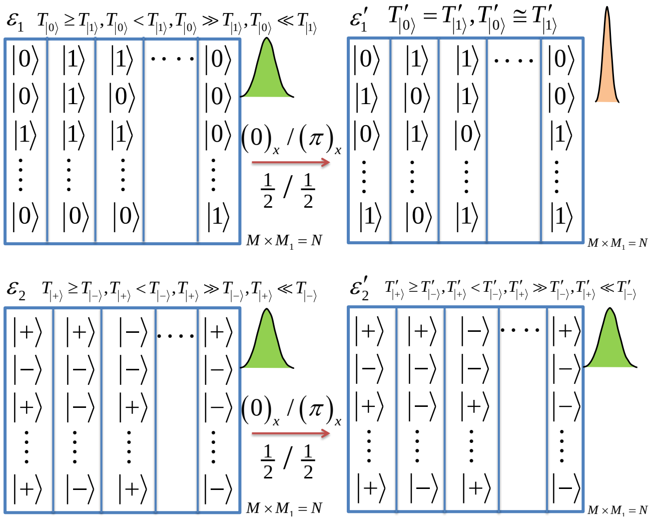
$\mathcal{E}_1$ $T_{|0\rangle} \ge T_{|1\rangle}, T_{|0\rangle} < T_{|1\rangle}, T_{|0\rangle} \gg T_{|1\rangle}, T_{|0\rangle} \ll T_{|1\rangle}$
$\mathcal{E}_1'$ $T'_{|0\rangle} = T'_{|1\rangle}, T'_{|0\rangle} \cong T'_{|1\rangle}$
$(0)_x / (\pi)_x$
$\frac{1}{2} / \frac{1}{2}$
$M \times M_1 = N$
$\mathcal{E}_2$ $T_{|+\rangle} \ge T_{|-\rangle}, T_{|+\rangle} < T_{|-\rangle}, T_{|+\rangle} \gg T_{|-\rangle}, T_{|+\rangle} \ll T_{|-\rangle}$
$\mathcal{E}_2'$ $T'_{|+\rangle} \ge T'_{|-\rangle}, T'_{|+\rangle} < T'_{|-\rangle}, T'_{|+\rangle} \gg T'_{|-\rangle}, T'_{|+\rangle} \ll T'_{|-\rangle}$
$(0)_x / (\pi)_x$
$\frac{1}{2} / \frac{1}{2}$
$M \times M_1 = N$

FIG. 1. (Color online) Non-interacting qubits fixed at definite sites of a 2D crystal lattice. Hence $\mathcal{E}_1$ and $\mathcal{E}_2$ are IC-ensembles. For Alice qubits in the IC-ensemble $\mathcal{E}_1$ ($\mathcal{E}_2$) are in a pure state $|\phi_{1j}\rangle = [|0\rangle|0\rangle|1\rangle...|0\rangle][|1\rangle|1\rangle|1\rangle...|0\rangle]...$ ($|\phi_{2j}\rangle = [|+\rangle|-\rangle|+\rangle...|+\rangle][|+\rangle|-\rangle|-\rangle...|-\rangle]...$). $(\theta_q)_x$ is random rotation about x-axis where $\theta_q = \{0, \pi\}$ with probability $\{1/2, 1/2\}$ respectively. $T_{|a_1\rangle}$ ($T_{|a_2\rangle}$) is the number of qubits in a given column of the IC-ensemble $\mathcal{E}_1$ ($\mathcal{E}_2$) which are in the state $|a_1\rangle$ ($|a_2\rangle$), $a_1 = 0, 1$ ($a_2 = +, -$), where $T_{|0\rangle} + T_{|1\rangle} = M$, and $T_{|+\rangle} + T_{|-\rangle} = M$. Gaussian is the probability density function of sample mean of M number of σ_z measurement outcomes. Bob uses M_1 sample mean points to construct the full Gaussian. Note that Bob measures only after applying $(\theta_q)_x$s.

We start section II by describing how Bob can partially purify his mixed state. In the beginning of section IIA, we give the motivation to introduce random x-rotations. Then by applying central limit theorem to independently distributed random variables, we obtain the resultant probability density of sample mean, in case of Bob getting IC-ensemble $\mathcal{E}_1$. In section IIB we consider the simplest case and show that resultant variance of sample mean gets reduced only in case of IC-ensemble $\mathcal{E}_1$, leading to discrimination. In section III we present results of a numerical simulation in support of the theoretical predictions, and we summarize and conclude in section IV. The details of the frameworks and related topics are elucidated in four appendices each of which are divided into subsections for purposes of clarity of presentation.

II. THEORY

Whenever we say measurement, we mean projective measurement unless stated otherwise.
Problem: Alice prepares an IC-ensemble $\mathcal{E}_1$ ($\mathcal{E}_2$) of N qubits in the following state: She tosses an unbiased coin and if the outcome is Head, then she prepares the j^{th} qubit in the state $|0\rangle$ ($|+\rangle$), else she prepares it in the state $|1\rangle$ ($|-\rangle$), $j = 1, 2, ..., N$. Alice gives Bob, IC-ensemble $\mathcal{E}_i$ with probability α_i, $i = 1, 2$, $\alpha_1 + \alpha_2 = 1$. Bob knows how Alice prepared the state of the qubits in the IC-ensembles $\mathcal{E}_1, \mathcal{E}_2$, but he do not know Alice's outcomes of coin tosses. Bob has to find out whether the given IC-ensemble is $\mathcal{E}_1$ or $\mathcal{E}_2$. Bob need not have to find out the exact state of each qubit in the given IC-ensemble $\mathcal{E}_i, i = 1$ or 2.

For notational convenience, almost every where we use same symbol for both random variable and its value. However they are distinguishable from the context.
Solution: Bob divides the given IC-ensemble $\mathcal{E}_i$ as follows to obtain M_1 sample mean points: $\mathcal{E}_i = \prod_{j=1}^{M_1} \otimes \mathcal{E}_{ij}$, $i = 1$ or 2. Each sample mean point is calculated with M number of σ_z measurement outcomes. Hence $N = M \times M_1$ where M, M_1 are sufficiently large enough to obtain at least approximately normally distributed sample mean. IC-ensemble $\mathcal{E}_{ij}$ is nothing but a column of the matrix in Fig. (1).

Let

$$\begin{aligned} \mathcal{F}_1 &= \{|0\rangle^{\otimes N}, |0\rangle^{\otimes N-1}|1\rangle, ..., |1\rangle^{\otimes N}\}, \text{ and} \\ \mathcal{F}_2 &= \{|+\rangle^{\otimes N}, |+\rangle^{\otimes N-1}|-\rangle, ..., |-\rangle^{\otimes N}\}. \end{aligned} \tag{1}$$

$\mathcal{F}_i$ is a complete set of orthonormal basis states in 2^N dimensional (2^N-D) Hilbert space, and also it corresponds to the sample space of N number of unbiased coin tosses where each of the 2^N possible outcomes are equally likely (Appendix C 1), $i = 1, 2$. E.g., for $N = 2$,

$$\begin{aligned} \mathcal{F}_1 &= \{|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle\}, \text{ and} \\ \mathcal{F}_2 &= \{|+\rangle|+\rangle, |+\rangle|-\rangle, |-\rangle|+\rangle, |-\rangle|-\rangle\}. \end{aligned}$$

Let $|\phi_{ij}\rangle \in \mathcal{F}_i$, $i = 1, 2$, where $j = 1, 2, ..., 2^N$. For Alice, qubits in the IC-ensemble $\mathcal{E}_i$ are in one of the pure states $|\phi_{ij}\rangle$s, as she knows the exact state of each of the N qubits in it, $i = 1, 2$. However for Bob, given qubits are in a mixed state

$$\begin{aligned} \rho^B &= \alpha_1\rho_1 + \alpha_2\rho_2, \\ \text{where } \rho_i &= \sum_{j=1}^{2^N} \frac{1}{2^N} |\phi_{ij}\rangle\langle\phi_{ij}| = \frac{\mathbb{1}_{2^N}}{2^N}, i = 1, 2 \end{aligned} \tag{2}$$

[10]. ρ^B represents the state of given qubits from Bob's mathematical perspective. Mixedness of ρ^B represents Bob's ignorance about whether the given IC-ensemble is $\mathcal{E}_1$ or $\mathcal{E}_2$, and also his ignorance about the exact state of each of the N qubits in the IC-ensemble $\mathcal{E}_i$ (which is represented by the mixedness of ρ_i), $i = 1, 2$. The fact that both ρ_1 and ρ_2 are maximally mixed implies that Bob is equally ignorant in both the cases. But it does not mean that Bob cannot discriminate between them. Although mathematically same, they are physically different (Appendix (C 3)). Bob assigns the value $+1$ (-1) to the outcome Head (Tail) in the above state preparation by Alice. As Bob knows Alice's states preparation procedures, he applies central limit theorem ([11], Appendix (A 1)) to them, which asserts that in the large M limit, probability of Alice preparing such a state $|\phi_{ij}\rangle$ which corresponds to the sample mean $S_i^A \rightarrow \text{ND} : 0, 1/M$ (i.e., S_i^A is a Normally Distributed random variable with mean 0 and variance $1/M$) tends to one, where $S_i^A = (T_i^{A+} - T_i^{A-})/M$, $T_i^{A\pm}$ is the number of ± 1 outcomes Alice obtains during the preparation of IC-ensemble $\mathcal{E}_i$ such that $T_i^{A+} + T_i^{A-} = M$, $i = 1, 2$. Using this information Bob projects ρ_i as follows:

$$\rho_i \rightarrow \hat{\rho}_i = \sum_j \hat{p}_j |\phi_{ij}\rangle\langle\phi_{ij}|$$

where $|\phi_{ij}\rangle$ is such a state which corresponds to the sample mean S_i^A, and $\sum_j \hat{p}_j \approx 1$, $i = 1, 2$. Renormalizing we get

$$\begin{aligned} \tilde{\rho}_i = \frac{\hat{\rho}_i}{\text{Tr}(\hat{\rho}_i)} &= \sum_j \tilde{p}_j |\phi_{ij}\rangle\langle\phi_{ij}| \neq \frac{\mathbb{1}_{2^N}}{2^N}, i = 1, 2. \\ \Rightarrow \rho^B \rightarrow \tilde{\rho}^B &= \sum_{i=1}^{2} \alpha_i \tilde{\rho}_i. \end{aligned} \tag{3}$$

Further $\text{Tr}(\tilde{\rho}_i^2) = \sum_j \tilde{p}_j^2 > \text{Tr}(\rho_i^2) = 1/2^N, i = 1, 2$. Hence the information supplied by central limit theorem has partially purified ρ_1, ρ_2 and hence ρ^B. In Appendix (C 6) we show how mixed state of qubits in an IC-ensemble can be purified solely via information.

Notations and definitions: (1) Bob rotates j^{th} qubit state in the IC-ensemble $\mathcal{E}_i$ about x-axis on Bloch sphere, through an angle θ_q (i.e., Bob evolves 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 which takes discrete values $\{\theta_1, \theta_2, ...\}$ with probability $\{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$ respectively ($\{\theta_1, \theta_2, ...\} \rightarrow \{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$), $\sum_q p^o_{\theta_q} = 1$, $i = 1$ or 2, $j = 1, 2, ..., N$.
(2) If Bob measures σ_z individually on each of the M qubits in the IC-ensemble $\mathcal{E}_{ij}$, then sample mean

$$S_i = (T_i^+ - T_i^-)/M$$

where $T_i^\pm$ is the number of ± 1 outcomes, and $T_i^+ + T_i^- = M$, $i = 1$ or 2. Value of S_i varies as j varies. However, Bob is going to measure only after applying $(\theta_q)_x$ individually to each of the M qubits in the IC-ensemble $\mathcal{E}_{ij}, i = 1$ or 2. S_i is defined just for the sake of calculations. Further $S_1 = S_1^A$ (Appendix (B 3)).
(3) Bob measures σ_z individually on each of the M qubits

in the IC-ensemble $\mathcal{E}'_{ij}$ to obtain sample mean

$$S'_i = (T'^{+}_i - T'^{-}_i)/M$$

where $T'^{\pm}_i$ is the number of ± 1 outcomes, $T'^{+}_i + T'^{-}_i = M$, and IC-ensemble $\mathcal{E}'_{ij}$ is got by applying $(\theta_q)_x$ individually to each of the M qubits in the IC-ensemble $\mathcal{E}_{ij}$, $i = 1$ or 2. Value of S'_i varies as j varies. But $\mathcal{E}'_{2j} = \mathcal{E}_{2j}$ ($\because$ $(\theta_q)_x$ introduces an insignificant global phase to the qubit states in the IC-ensemble $\mathcal{E}_{2j}$). Hence $S'_2 = S_2 + X = S_2$ where X corresponds to application of $(\theta_q)_x$s.

From Alice's state preparation procedure, it is easily evident that $S_1 \equiv S_2$ i.e., S_1 and S_2 are independent and identically distributed (i.e., they have same mean and variance) random variables (Appendix (B 3)). Further, sample mean S_i has mean $\langle S_i \rangle = \langle \sigma_z \rangle_{|+\rangle} = 0$, and variance

$$\Delta S_i^2 = (\Delta \sigma_z)^2_{|+\rangle}/M = 1/M,$$

(Appendix (B 3)), where $\langle X \rangle_{|\zeta\rangle} = \mathrm{Tr}(X|\zeta\rangle\langle\zeta|)$, and

$$(\Delta X)^2_{|\zeta\rangle} = \langle (\langle X \rangle_{|\zeta\rangle} - x)^2 \rangle = \langle X^2 \rangle_{|\zeta\rangle} - \langle X \rangle^2_{|\zeta\rangle}$$

[12], $i = 1, 2$. As $S_1^A = S_1 \equiv S_2$ (Appendix (B 3)) we have sample mean $S_i \to \mathrm{ND} : 0, 1/M$, $i = 1, 2$. Now 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 Bob can discriminate between IC-ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$.

A. General case

Motivation: Consider the following theorem: If $X_i \to \mathrm{ND} : \mu_i, \sigma_i^2$, then $Z = \sum_{i=1}^{\tilde{N}} X_i \to \mathrm{ND} : \sum_{i=1}^{\tilde{N}} \mu_i, \sum_{i=1}^{\tilde{N}} \sigma_i^2$ where X_is are normally distributed independent random variables [11]. 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 via $(\theta_q)_x$, into already present random variable S_1 in the IC-ensemble $\mathcal{E}_{1j}$, so that resultant probability distribution may turn out to be different from that of S_1. We are going to sort of convolute (Eq. (10) resembles convolution) two independent probability distributions: $S_1 \to \mathrm{ND} : 0, 1/M$ and $\{\theta_1, \theta_2, ...\} \to \{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$ to obtain $S'_1 \to \mathrm{ND} : 0, (1 - (\Delta \cos\theta_q)^2_{p^o_{\theta_q}})/M$. Note that in S'_1 there is a reduction in variance unlike in Z above. This is because in case of Z, as $\tilde{N}$ increases, number of independent random variables also increases. But it is not so in case of S'_1 (Appendix (B 13)).

Applying $(\theta_q)_x$ individually to each of the M qubit states in the IC-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 Bob obtains sample mean $S'_2 (= S_2) \to \mathrm{ND} : 0, 1/M$. Whereas in the IC-ensemble $\mathcal{E}_{1j}$, applying $(\theta_q)_x$ individually to each of the M qubit states, transforms $|0\rangle, |1\rangle$ to

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

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

respectively.

Application of $(\theta_q)_x$ individually to each of the N qubits in the unknown state $|\phi_{ij}\rangle$, transforms $\tilde{\rho}^B$ (Eq. (3)) as follows:

$$\begin{aligned}\rho'^B &= \sum_{l=1}^{d^N} P_l U_l \tilde{\rho}^B U_l^\dagger = \sum_{i=1}^{2} \alpha_i \sum_j \tilde{p}_j \sum_{l=1}^{d^N} P_l U_l |\phi_{ij}\rangle\langle\phi_{ij}| U_l^\dagger \\ &= \sum_{i=1}^{2} \alpha_i \sum_{j,l} \tilde{p}_j P_l |\phi'_{ijl}\rangle\langle\phi'_{ijl}| = \sum_{i=1}^{2} \alpha_i \rho'_i, \qquad (4)\end{aligned}$$

where $U_l = (\theta_{q_1})_x \otimes (\theta_{q_2})_x \otimes ... \otimes (\theta_{q_N})_x$, $q_1, q_2, ..., q_N = 1, 2, ..., d$, $\{\theta_1, \theta_2, ..., \theta_d\} \to \{p^o_{\theta_1}, p^o_{\theta_2}, ..., p^o_{\theta_d}\}$, P_l is the probability with which U_l is applied, and $\sum_{l=1}^{d^N} P_l = 1$ [13], Appendix (C 8). Upon Bob measuring σ_z individually on each of the N qubits in the state ρ'_i, the state ρ'_i gets projected onto a pure state ($\because$ Bob knows the post measurement state of each of the N qubits exactly), $i = 1$ or 2 (Appendix (C 4)). Using central limit theorem we are going to show that, in the large M limit, probability of ρ'_1 getting projected onto such a pure state which corresponds to the sample mean $S'_1 \to \mathrm{ND} : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$ tends to one. Using this information Bob can project as follows:

$$\rho'_1 \to \rho''_1 = \sum_{j,l} \tilde{p}_j P_l |\phi'_{1jl}\rangle\langle\phi'_{1jl}| = \sum_j p''_j |\phi'_{1j}\rangle\langle\phi'_{1j}|,$$

where $|\phi'_{1jl}\rangle, |\phi'_{1j}\rangle$ are such states which corresponds to the sample mean $S'_1 \to \mathrm{ND} : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$, and $\sum_j p''_j \approx 1$. Renormalizing

$$\hat{\rho}'_1 = \frac{\rho''_1}{\mathrm{Tr}(\rho''_1)} = \sum_j \hat{p}'_j |\phi'_{1j}\rangle\langle\phi'_{1j}|. \qquad (5)$$

In a special case one can easily show that $\mathrm{Tr}(\hat{\rho}'^2_1) > \mathrm{Tr}(\tilde{\rho}_1^2)$ (Appendix (C 7)), and hence there is an increase in purity. But $\rho'_2 = \tilde{\rho}_2$ ($\because$ $(\theta_q)_x$ introduces an insignificant global phase to $|+\rangle, |-\rangle$ in $|\phi_{2j}\rangle$s). $\Rightarrow \mathrm{Tr}({\rho'_2}^2) = \mathrm{Tr}(\tilde{\rho}_2^2)$. Hence

$$\rho'^B \to \hat{\rho}'^B = \alpha_1 \hat{\rho}'_1 + \alpha_2 \tilde{\rho}_2.$$

As $\hat{\rho}'_1$ corresponds to the sample mean S'_1, and $\tilde{\rho}_2$ to $S'_2 (= S_2 \equiv S_1)$, Bob can discriminate between ρ_1 and ρ_2.

Consider $|\phi'_{1jl}\rangle$. Measuring σ_z individually on each of the $|\theta_q\rangle$s and $|\theta_{q\perp}\rangle$s is equivalent to tossing differently biased coins. By these measurements, ρ'_1 is projected onto a pure state as explained above. In a special case Bob can even know the exact state of each of the N qubits in the given IC-ensemble $\mathcal{E}_1$ (Appendix (B 6)). After measurement, Bob calculates the sample mean (S_1') of the outcomes. We are now going to obtain the probability density function of S_1'. We have random variable means $\langle\sigma_z\rangle_{|\theta_q\rangle} = \cos\theta_q$, $\langle\sigma_z\rangle_{|\theta_{q\perp}\rangle} = -\cos\theta_q$, 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 (id) random variables [11], we obtain effective mean

$$\begin{aligned}\mu_{\text{eff}} &= \sum_q (p_q\langle\sigma_z\rangle_{|\theta_q\rangle} + p_{q\perp}\langle\sigma_z\rangle_{|\theta_{q\perp}\rangle})\\ &= \sum_q (p_q - p_{q\perp})\cos\theta_q, \qquad (6)\end{aligned}$$

where $p_q = M_q'(T_1^+, p_{\theta_q})/M$ ($p_{q\perp} = M_{q\perp}'(T_1^-, p_{\theta_q})/M$), and M_q' ($M_{q\perp}'$) is the total number of $|\theta_q\rangle$s ($|\theta_{q\perp}\rangle$s), $\sum_q[M_q' + M_{q\perp}'] = M$; $p_{\theta_q} = m_q/M$ where m_q is the total number of times $(\theta_q)_x$ is applied, and $\sum_q m_q = M$ (see Appendix (A 2) for derivation of μ_{eff}). Note that the basic probabilities $\cos^2(\theta_q/2), \sin^2(\theta_q/2)$ are fixed, whereas the derived probabilities $p_q, p_{q\perp}$ varies over M_1 number of IC-ensembles $\mathcal{E}_{1j}$s, because the numbers $M_q', M_{q\perp}'$ are not fixed. As $T_1^{\pm}$ and m_q are independent, we obtain using Bayes' rule

$$p_q = p_1^+ p_{\theta_q}, p_{q\perp} = p_1^- p_{\theta_q}, \qquad (7)$$

where $p_1^{\pm} = T_1^{\pm}/M$ [14]. Substituting $p_q, p_{q\perp}$ into μ_{eff} (Eq. (6)) we obtain

$$\mu_{\text{eff}} = S_1\langle\cos\theta_q\rangle_{p_{\theta_q}}, \qquad (8)$$

where $\langle\cos\theta_q\rangle_{p_{\theta_q}} = \sum_q p_{\theta_q}\cos\theta_q$. Note that probabilities $p_1^{\pm}, p_{\theta_q}$, and sample mean S_1 are normally distributed random variables with non-zero variance i.e.,

$$\begin{aligned}&p_1^{\pm} \rightarrow \text{ND} : 1/2, 1/(4M)\ (\because T_1^{\pm} \rightarrow \text{ND} : M/2, M/4),\\ &p_{\theta_q} \rightarrow \text{ND} : p^o_{\theta_q}, \sigma^2_{m_q}/M^2\ (\because m_q \rightarrow \text{ND} : p^o_{\theta_q}M, \sigma^2_{m_q}),\end{aligned}$$

and $S_1 \rightarrow \text{ND} : 0, 1/M$ where $\sigma^2_{m_q} \sim M$ (see Appendix (B 4) for derivation). Hence we need to take care of the variance (however small) present in them. Hence, first we shall 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.

Applying central limit theorem to id random variables, we obtain effective variance

$$\begin{aligned}(\Delta\sigma_z)^2_{\text{eff}} &= \sum_q (p_q(\Delta\sigma_z)^2_{|\theta_q\rangle} + p_{q\perp}(\Delta\sigma_z)^2_{|\theta_{q\perp}\rangle})\\ &= \sum_q (p_q + p_{q\perp})\sin^2\theta_q = 1 - \langle\cos^2\theta_q\rangle_{p_{\theta_q}}, \ (9)\end{aligned}$$

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 $(\Delta\sigma_z)^2_{\text{eff}}$). Note that even though μ_{eff} happens to coincide with $\langle\sigma_z\rangle_{\rho_{1j}'}$, $(\Delta\sigma_z)^2_{\text{eff}} \neq \langle\sigma_z^2\rangle_{\rho_{1j}'} - \langle\sigma_z\rangle^2_{\rho_{1j}'}$ where $\rho_{1j}' = \sum_q(p_q|\theta_q\rangle\langle\theta_q| + p_{q\perp}|\theta_{q\perp}\rangle\langle\theta_{q\perp}|)$, this is because in going from IC-ensemble $\mathcal{E}_{1j}'$ to CC-ensemble corresponding to ρ_{1j}', there is information loss (Appendix (A 2, A 3)). 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, $S_1' \rightarrow \text{ND} : \mu_{\text{eff}}, (\Delta\sigma_z)^2_{\text{eff}}/M$ [11], 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:

$$\begin{aligned}f(S_1') = \int\prod_{i,i\neq l}\{dp_{\theta_i}(\text{Nd}(p_{\theta_i}) : p^o_{\theta_i}, \sigma^2_{m_i}/M^2)\}\\ \times dS_1(\text{Nd}(S_1) : 0, 1/M)\left(\text{Nd}(S_1') : \mu_{\text{eff}}, (\Delta\sigma_z)^2_{\text{eff}}/M\right),(10)\end{aligned}$$

where $(\text{Nd}(x) : \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}}\exp(-(x-\mu)^2/(2\sigma^2))$ (i.e., Normal probability density function with mean μ and variance σ^2), $dx(\text{Nd}(x) : \mu, \sigma^2)$ is the probability of obtaining value x of normally distributed random variable x. In Eq. (10) 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. (10) we have product of probabilities because p_{θ_q}s and S_1 are independent random variables [15]. As there are no constraint equations in p_{θ_q}s and S_1, we have to integrate over the entire hyper volume spanned by p_{θ_q}s and S_1. Further, as we can integrate in any order (Appendix (B 16)), we can integrate out S_1 from $-\infty$ to ∞ [16] in Eq. (10). Integrating out S_1 we obtain

$$\begin{aligned}f(S_1') = \int\prod_{i,i\neq l}\{dp_{\theta_i}(\text{Nd}(p_{\theta_i}) : p^o_{\theta_i}, \sigma^2_{m_i}/M^2)\}\\ \times(\text{Nd}(S_1') : 0, (1 - (\Delta\cos\theta_q)^2_{p_{\theta_q}})/M) \quad (11)\end{aligned}$$

where $(\Delta\cos\theta_q)^2_{p_{\theta_q}} = \langle\cos^2\theta_q\rangle_{p_{\theta_q}} - \langle\cos\theta_q\rangle^2_{p_{\theta_q}}$ [17].

Now consider $\theta_q = \theta_0, \forall q$. Then Eq. (11) reduces to $f(S_1') = (\text{Nd}(S_1') : 0, (1-0)/M) = g(S_2')$, hence no discrimination. This is expected because, by rotating all M qubit states by same angle we are not introducing any new independent random variable. A random variable is characterized by having non zero variance. But here variance of random variable $\cos\theta_q$ is $(\Delta\cos\theta_q)^2_{p^o_{\theta_q}} = 0$. Hence no randomness. Hence we cannot change/distort the probability distribution of sample mean (S_1) corresponding to the IC-ensemble $\mathcal{E}_{1j}$. Hence for discrimination, we should take at least $\{\theta_1, \theta_2\} \rightarrow \{p^o_{\theta_1}, p^o_{\theta_2}\}$, $\theta_1 \neq \theta_2, p^o_{\theta_q} \neq 0, \forall q$. Further, when $\theta_0 = 0, f(S_1') = (\text{Nd}(S_1') : 0, 1/M) = g(S_1)$, probability density of S_1, as required.

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. (11). Eliminating p_{θ_2} from Eq. (11) we obtain

$$f(S'_1) = \int dp_{\theta_1} (\mathrm{Nd}(p_{\theta_1}) : p^o_{\theta_1}, \sigma^2_{m_1}/M^2)$$
$$\times (\mathrm{Nd}(S'_1)\ :\ 0, (1 - p_{\theta_1}(1 - p_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2)/M). \tag{12}$$

Direct evaluation of the integral in Eq. (12) is difficult (for an indirect evaluation see Appendix (B 5)). Note that $f(S'_1)$ in Eq. (12) is the weighted mean of many Gaussians each having mean zero. Hence there is no swaying of center of Gaussians (Appendix (B 1)) unlike in Eq. (10). Also as M is large, it is justifiable to replace the weighing Gaussian $(\mathrm{Nd}(p_{\theta_1}) : p^o_{\theta_1}, \sigma^2_{m_1}/M^2)$ in Eq. (12) with delta function $\delta(p_{\theta_1} - p^o_{\theta_1})$ (Appendix (A 1)) to obtain

$$f(S'_1) \approx (\mathrm{Nd}(S'_1) : 0, ((\Delta\sigma_z)^2_{|+\rangle} - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M), \tag{13}$$

where $(\Delta\sigma_z)^2_{|+\rangle} = 1$, and

$$(\Delta\cos\theta_q)^2_{p^o_{\theta_q}} = p^o_{\theta_1}(1 - p^o_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2.$$

Hence the resultant variance of sample mean S'_1 is $\Delta S'^2_1 \approx (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$. Note that this approximation does not work in Eq. (10), as there is swaying of center of Gaussians. In Eq. (10) if we replace the weighing Gaussian $(\mathrm{Nd}(S_1) : 0, 1/M)$ with $\delta(S_1 - 0)$, then we will be neglecting the swaying of center of Gaussians in $(\mathrm{Nd}(S'_1)\ :\ \mu_{\text{eff}}, (\Delta\sigma_z)^2_{\text{eff}}/M)$. This results in $f(S'_1) = (\mathrm{Nd}(S'_1) : 0, \sin^2 0/M)$, for $\theta_q = 0\ \forall q$, which is not correct. Hence swaying of center of Gaussians affects/contributes to the resultant/net variance. To conclude, as the variance of sample mean S'_2, $\Delta S'^2_2 = 1/M$ ($\because S'_2 = S_2$), is different from $\Delta S'^2_1$, Bob can discriminate between the two IC-ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$.

Nonlinearity in action: We will show how nonlinearity is reducing the variance. We have $\Delta S'^2_1 \approx (\langle\cos\theta_q\rangle^2_{p^o_{\theta_q}} + \langle\sin^2\theta_q\rangle_{p^o_{\theta_q}})/M$ (Eq. (13)). Let $\{\theta_1(=0), \theta_2(=\pi/2)\} \to \{p^o_0, p^o_{\pi/2}\}$. $\Rightarrow \Delta S'^2_1 \approx (p^{o2}_0 + p^o_{\pi/2})/M < 1/M$. Note that squaring of probabilities is a nonlinear operation (Appendix (B 10)).

In Appendix (C 1) we show that it is nothing but a 'deterministic but inexact nonorthogonal state discrimination' problem. For explanation using central limit theorem and Shannon entropy, see sections (B 8) and (B 9) respectively in Appendix.

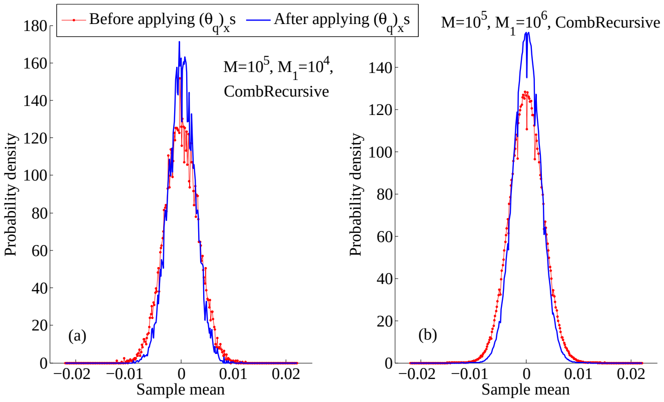

FIG. 2. (Color online) CombRecursive is the PRN generator. Red curve with dot marker is $g(S_1)(= g(S'_2) = g(S_2))$, and blue curve with no marker is $f(S'_1)$. (a) A_g (= Area under one standard deviation of $g(S_1)$ i.e., from $S_1 = -\Delta S_1 = -1/\sqrt{M}$ to $S_1 = \Delta S_1$) is 0.685 (theoretical prediction in the large M, M_1 limit is ≈ 0.683). A_f (= Area under $f(S'_1)$ corresponding to one standard deviation of $g(S_1)$) is 0.783 (as predicted by our protocol in the large M, M_1 limit is ≈ 1). Hence there is clear reduction in variance i.e., $\Delta S'^2_1 < \Delta S^2_1 (= \Delta S'^2_2)$. (b) $A_g = 0.684$, $A_f = 0.78$.

III. NUMERICAL SIMULATION

Reduction in variance: Standard uniformly distributed Pseudo Random Numbers (PRN), drawn from the open interval $(0, 1)$, were generated using MATLAB. These PRNs were used to 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 $\{\theta_1(= 0), \theta_2(= \pi)\} \to \{p^o_{\theta_1}(= 1/2), p^o_{\theta_2}(= 1/2)\}$ for various values of M, M_1. Application of $(\theta_q)_x$s was simulated as described in Appendix (B 4). Here we discriminate by comparing $f(S'_1)$ with $g(S_1)(= g(S_2) = g(S'_2))$, as both density functions can be obtained from the given IC-ensemble $\mathcal{E}_i, i = 1$ or 2. If the given IC-ensemble is $\mathcal{E}_1$, then $f(S'_1)$ corresponds to after applying $(\theta_q)_x$s, else $f(S'_1)$ (virtual) corresponds to before applying $(\theta_q)_x$s (Appendix (B 2)). Results are plotted in Fig. (2), and in Appendix Fig.s (3, 4, 5, 7(c)). There is a clear reduction in variance as predicted by theory. In Fig. (2), $g(S_1)$ is much closer to the corresponding theoretical prediction, but $f(S'_1)$ is not so close to the corresponding theoretical prediction (approximately a delta function). 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 $M = 10^2, 10^5, ...; M_1 = 2000, 10^4, 10^6, ...$ Reasons given in the section 'How hard it might be to reduce the variance?' in Appendix (B 14) may also apply here. (2) Theoretical calculations may not be exact/precise. E.g., we have not evaluated the integral in Eq. (12) exactly. (3) PRNs depends on the generator. (4) We might be missing something in theoretical calculations.

Instead of directly looking for reduction in variance, we can also look for reduction in population difference $(|T_1'^+ - T_1'^-| - |T_1^+ - T_1^-|)$ (Appendix (D 2)).

IV. SUMMARY AND CONCLUSION

We considered such kind of ensembles where Alice and Bob were able to individually address and control each qubit in the ensemble. Alice prepared two ensembles of a large number of qubits, $\mathcal{E}_1$ and $\mathcal{E}_2$, in the following states: Depending on the outcomes of unbiased coin tosses, she prepared the qubits in $\mathcal{E}_1$, in the eigenstates of Pauli-z matrix; while she prepared the qubits in $\mathcal{E}_2$, in the eigenstates of Pauli-x matrix. Alice gave Bob one of the two ensembles, $\mathcal{E}_1, \mathcal{E}_2$, and asked him to identify. Bob was aware of Alice's states preparation procedures, but was unaware of her outcomes of coin tosses. Qubits in each of the two ensembles were in maximally mixed state from Bob's mathematical perspective, although they had different physical content. We showed that Bob was able to partially purify the mixed states by applying central limit theorem to his knowledge of Alice's state preparation procedure, which gave him the hopes of discrimination. Then Bob individually rotated each qubit state in the given ensemble about x-axis by a random angle, and then projectively measured Pauli-z operator. We showed that, by these operations, variance of sample mean of measurement outcomes gets reduced if the given ensemble were $\mathcal{E}_1$. As random rotations about x-axis does nothing to the qubit states in the ensemble $\mathcal{E}_2$, variance of sample mean remains unaltered, leading to discrimination. We also exhibited numerical simulation results in support of theoretical predictions. A likely implication of our discrimination protocol might be that Alice may signal Bob provided Bob knows a priori that Alice is going to send a message at a pre agreed upon time instant. Hence it does not violate the no-signaling principle per se, as the latter excludes any kind of a priori knowledge. However the origin of nonlinear effect (reduction in variance) (II B) which leads to discrimination is not clear. Still it remains to be explored whether it is a genuine or pseudo nonlinear effect due to one or more or some combination of the following operations: selective projective measurements (Appendix (C 11)), information supplied by central limit theorem, selective random x-rotations, and statistical data analysis technique? Further analysis from an information theoretic perspective i.e., entropy, mutual information etc., may shed some light on this question, and also may give further insight. Our efforts are going on in this direction.

ACKNOWLEDGEMENTS

It is a pleasure to thank my supervisor Prof. T S Mahesh for initiating the project, ensemble description, and manuscript preparation. I express my deep gratitude to Prof. R Srikanth for very useful discussions, comments, and density matrix formulation. I also express my sincere gratitude to Prof. A K Rajagopal for insightful suggestions, critical comments, and scrutinizing part of the manuscript. Finally I thank Prof. L Vaidman, Aravinda, Govind Unnikrishnan, Prof. M Ozawa, Prof. D Home, Prof. A D Patel, Deepak Khurana, Anjusha V S, Soham Pal, Udaysinh T Bhosale, and Mohd A Siddiqui for useful discussions, valuable comments and suggestions.

[1] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic Publishers, 2002).

[2] J. Tolar and P. Hjek, Physics Letters A **353**, 19 (2006).

[3] S. K. Goyal, R. Singh, and S. Ghosh, Phys. Rev. A **93**, 012114 (2016).

[4] G. L. Long, Y.-F. Zhou, J.-Q. Jin, Y. Sun, and H.-W. Lee; quant ph/0408079v3, Foundations of Physics **36**, 1217 (2006).

[5] U. Fano, Rev. Mod. Phys. **29**, 74 (1957).

[6] A. Peres and D. R. Terno, Rev. Mod. Phys. **76**, 93 (2004).

[7] J. Kofler and i. c. v. Brukner, Phys. Rev. A **87**, 052115 (2013).

[8] Y.-C. Lee, M.-H. Hsieh, S. T. Flammia, and R.-K. Lee, Phys. Rev. Lett. **112**, 130404 (2014).

[9] D. G. Cory, A. F. Fahmy, and T. F. Havel, Proceedings of the National Academy of Sciences **94**, 1634 (1997).

[10] ρ_1 can also be rewritten as follows: $\rho_1 = \rho_{1\text{ind}}^{\otimes N}$ where $\rho_{1\text{ind}} = (\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|)$ is the state of an individual qubit. Similarly $\rho_2 = \rho_{2\text{ind}}^{\otimes N}$ where $\rho_{2\text{ind}} = (\frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -|)$.

[11] S. Ross, *A first course in probability* (Pearson, 2012).

[12] C. Cohen Tannoudji et al., *Quantum Mechanics vol. one* (John Wiley and Sons, 2005).

[13] Note that this kind of transformation is not possible in a CC-ensemble. Further $\{U_1, U_2, ...\} \rightarrow \{P_1, P_2, ...\}$ is equivalent to $\{\theta_1, \theta_2, ...\} \rightarrow \{p^o_{\theta_1}, p^o_{\theta_2}, ...\}$. Hence Bob can expect even U_ls to change the probability distribution of sample mean, similar to $(\theta_q)_x$s, leading to discrimination.

[14] $M'_q(T_1^+, p_{\theta_q}) = T_1^+ p_{\theta_q} = p_1^+ m_q$. 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. $\Rightarrow p_q = p_1^+ p_{\theta_q}, p_{q\perp} = p_1^- p_{\theta_q}$.

[15] Hence the joint probability $P(p_{\theta_1}, p_{\theta_2}, ..., p_{\theta_{l-1}}, p_{\theta_{l+1}}, ..., S_1) = P(p_{\theta_1})P(p_{\theta_2})...P(p_{\theta_{l-1}})P(p_{\theta_{l+1}})...P(S_1) = \prod_{i, i\neq l}\{dp_{\theta_i}(\text{Nd}(p_{\theta_i}) : p^o_{\theta_i}, \sigma^2_{m_i}/M^2)\} dS_1(\text{Nd}(S_1) : 0, 1/M)$.

[16] Actually $-1 \leq S_1 \leq 1$, but we are going to integrate from $-\infty$ to ∞. This is because, in the large M limit, $(\text{Nd}(S_1) : 0, 1/M)$ behaves like 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 ($\mathrm{erf}(x)$).

[17] S_1 was oscillating symmetrically about zero. Hence independent of the value of the coefficient of S_1 (in Eq. (10)), resultant mean has vanished. It is like $\langle CS_1\rangle_{\Omega_1} = C\langle S_1\rangle_{\Omega_1} = 0$ where $\Omega_1 = (\mathrm{Nd}(S_1) : 0, 1/M)$.

[18] J. Audretsch, *Entangled Systems: New Directions in Quantum Physics* (Wiley, 2007).

[19] G. M. D'Ariano and H. P. Yuen, Phys. Rev. Lett. **76**, 2832 (1996).

[20] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010) cambridge Books Online.

[21] D. Home, *Conceptual Foundations of Quantum Physics* (Springer Science and Business media, 1997).

[22] G. Lüders, Annalen der Physik **15**, 663 (2006).

[23] ComScire, True random number generator .

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: $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(x-\mu)^2/(2\sigma^2))$. $\Rightarrow$ $\langle X\rangle = \int\limits_{-\infty}^{\infty} xf(x)dx = \mu$, and variance $\langle X^2\rangle - \langle X\rangle^2 = \sigma^2$ [11]. 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 $Y = aX + b \to \mathrm{ND} : a\mu + b, a^2\sigma^2$ [11]. 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 [11]. 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 $\langle X_i\rangle = \mu$ 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 $\Delta S^2 = \sigma^2/n$ [11]. 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 outcome of each 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:

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

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 [11]. Then using Eq. (A1) we obtain

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

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_j have mean μ'_j and variance σ'^2_j, $j = 1, 2, ..., r$, 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 $\sigma_i^2, i = 1, 2, ..., n$.

$$\mu_i = \mu'_1, \sigma_i^2 = \sigma'^2_1 \text{ for } i = 1, 2, ..., n_1,$$
$$\mu_i = \mu'_2, \sigma_i^2 = \sigma'^2_2 \text{ for } i = n_1+1, n_1+2, ..., n_1+n_2,$$
$$\vdots$$
$$\mu_i = \mu'_r, \sigma_i^2 = \sigma'^2_r \text{ for } i = (n_1+n_2+...+n_{r-1}+1),$$
$$(n_1+n_2+...+n_{r-1}+2), ..., (n_1+n_2+...+n_{r-1}+n_r).$$

If X_is are uniformly bounded ([11] p. 399) and $\sum_{i=1}^\infty \sigma_i^2 = \infty$ then,

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

for given values of n_js (and hence p_js), $j = 1, 2, ..., r$ [11]. 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 $j^{\rm th}$ face is p_j^o. $p_j \to {\rm ND} : p_j^o, \sigma_{n_j}^2/n^2$, $\sigma_{n_j}^2 \sim n$), then it should be taken into account by integrating over all possible values of n_js (or p_js), after multiplying ${\rm Nd}(S) : \mu'_{\rm eff}, (\Delta X')^2_{\rm eff}/n$ by corresponding weighing factors/weights, to get the final effective probability density function of sample mean. When $\mu'_j = \mu, \sigma'^2_j = \sigma^2, \forall j$ 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 $\sigma'^2_j = 1 - \mu'^2_j (\because \sigma_i^2 = \langle X_i^2 \rangle - \langle X_i \rangle^2 = 1 - \mu_i^2)$. Using $\mu_i = p_{H_i} - p_{T_i}$, we can rewrite

$$\mu'_{\rm eff} = \sum_{j=1}^{r} p_j \mu'_j = p^+_{\rm eff} - p^-_{\rm eff} = \langle X_{\rm eff} \rangle$$

for given values of p_js. Now consider

$$(\Delta X')^2_{\rm eff} = \sum_{j=1}^{r} p_j \sigma'^2_j = \sum_{j=1}^{r} p_j (1 - \mu'^2_j) = 1 - \sum_{j=1}^{r} p_j \mu'^2_j$$
$$\neq 1 - (\sum_{j=1}^{r} p_j \mu'_j)^2 = 1 - \langle X_{\rm eff} \rangle^2 = \Delta X^2_{\rm eff},$$

where $\Delta X^2_{\rm eff} = \langle X^2_{\rm eff} \rangle - \langle X_{\rm eff} \rangle^2$. 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 $X_{\rm eff}$ (here previous r types of coins has been coalesced into a single effective coin $X_{\rm eff}$). Hence the two situations are different.

No effective state: Let the random variable X_i be measuring σ_z on the state $|P_i\rangle = \sqrt{P_i}|0\rangle + \sqrt{1 - P_i}|1\rangle, i = 1, 2, ..., n$. Let

$P_i = P'_1$ for $i = 1, 2, ..., n_1$,
$P_i = P'_2$ for $i = n_1 + 1, n_1 + 2, ..., n_1 + n_2$,
$\vdots$
$P_i = P'_r$ 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 $\sigma'^2_j = 1 - \mu'^2_j (\because \sigma_i^2 = \langle X_i^2 \rangle - \langle X_i \rangle^2 = 1 - \mu_i^2)$. We can rewrite

$$\mu'_{\rm eff} = \sum_{j=1}^{r} p_j (2P'_j - 1) = \sum_{j=1}^{r} p_j {\rm Tr}(\sigma_z \rho'_j) = {\rm Tr}(\sigma_z \rho'_{\rm eff})$$
$$= \langle \sigma_z \rangle_{\rho'_{\rm eff}}$$

for given values of p_js, where $\rho'_{\rm eff} = \sum_{j=1}^r p_j \rho'_j$, $\rho'_j = |P'_j\rangle\langle P'_j|$. Now consider

$$(\Delta X')^2_{\rm eff} = \sum_{j=1}^{r} p_j \sigma'^2_j = \sum_{j=1}^{r} p_j (1 - \mu'^2_j) = 1 - \sum_{j=1}^{r} p_j \mu'^2_j$$
$$\neq 1 - (\sum_{j=1}^{r} p_j \mu'_j)^2 = 1 - \langle \sigma_z \rangle^2_{\rho'_{\rm eff}} = \langle \sigma_z^2 \rangle_{\rho'_{\rm eff}} - \langle \sigma_z \rangle^2_{\rho'_{\rm eff}}.$$

Hence the concept of effective state $\rho'_{\rm eff}$ is not correct with respect to effective variance unless $r = 1$. This is because $\rho'_{\rm eff}$ corresponds to a CC-ensemble (Alice has only partial knowledge of the state of the qubits in the ensemble), whereas the formula $(\Delta X')^2_{\rm eff} = \sum_j p_j \sigma'^2_j$ corresponds to an IC-ensemble (Alice has full knowledge of the state of the qubits in the ensemble). For further justification see Appendix (A 3).

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

3. Justifying the formula $(\Delta X')^2_{\rm eff} = \sum_j p_j \sigma'^2_j$ (Appendix (A 2)) in the light of an IC-ensemble

Consider an IC-ensemble $\mathcal{E}_{11}$ having $T_{|0\rangle}$ number of qubits in the state $|0\rangle$ and $T_{|1\rangle}$ number of qubits in the state $|1\rangle$ such that $T_{|0\rangle} + T_{|1\rangle} = M$, and M is sufficiently large. Let the state of the qubits in the IC-ensemble $\mathcal{E}_{11}$ be $|0\rangle|1\rangle....$ Consider measuring σ_z. σ_z has no variance with respect to $|0\rangle$ and $|1\rangle$ as both are its eigenkets. Hence effective/resultant variance of σ_z measurement on the IC-ensemble $\mathcal{E}_{11}$ must also be zero. This is because however many times we repeat σ_z measurement on identically prepared IC-ensembles $\mathcal{E}_{11}$s (i.e., each having exactly $T_{|0\rangle}, T_{|1\rangle}$ number of $|0\rangle$s, $|1\rangle$s respectively), we always get exactly $T_{|0\rangle}$ number of $+1$ outcomes and $T_{|1\rangle}$ number of -1 outcomes. Hence no variance. Hence sample mean $S = (T_{|0\rangle} - T_{|1\rangle})/M = p_0 - p_1$ where $p_0 = T_{|0\rangle}/M, p_1 = T_{|1\rangle}/M$, and S has zero variance.

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

Instead let us define $(\Delta \sigma_z)^2_{\rm eff} = \langle \sigma_z^2 \rangle_{\rho_1} - \langle \sigma_z \rangle^2_{\rho_1}$ where $\rho_1 = p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|$ (as the state has become mixed, we have lost partial information. Where as the IC-ensemble $\mathcal{E}_{11}$ corresponds to a pure state, and hence we have full information). $\mu_{\rm eff} = \langle \sigma_z \rangle_{\rho_1} = p_0 - p_1$.

$\Rightarrow (\Delta\sigma_z)^2_{\text{eff}} = 1-(p_0-p_1)^2 = 1-\mu^2_{\text{eff}}$. Then sample mean S has variance $\Delta S^2 = (\Delta\sigma_z)^2_{\text{eff}}/M = (1-(p_0-p_1)^2)/M$ (it is the variance in initial state (ρ_1) preparation, see 'Coalesced' below) which is $\neq 0$ in general. Moreover it predicts that $S \rightarrow \text{ND} : p_0 - p_1, (1-(p_0-p_1)^2)/M$. 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 'Coalesced' below). Hence this definition of effective variance corresponds to a CC-ensemble whose qubits are in the state ρ_1, whereas the formula $(\Delta X')^2_{\text{eff}} = \sum_j p_j\sigma_j'^2$ corresponds to the IC-ensemble $\mathcal{E}_{11}$ whose qubits are in the pure state $|0\rangle|1\rangle\ldots$.

Consider another IC-ensemble $\mathcal{E}_{21}$ having $T_{|+\rangle}$ number of qubits in the state $|+\rangle$ and $T_{|-\rangle}$ number of qubits in the state $|-\rangle$ such that $T_{|+\rangle} + T_{|-\rangle} = M$, and M is sufficiently large. Let the state of the qubits in the IC-ensemble $\mathcal{E}_{21}$ be $|+\rangle|-\rangle\ldots$. Let $\rho_2 = p_+|+\rangle\langle+|+p_-|-\rangle\langle-|$ where $p_\pm = T_{|\pm\rangle}/M$. ρ_2 represents the state of the qubits in a CC-ensemble. State of the qubits in the IC-ensemble $\mathcal{E}_{21}$ is pure and hence we have full knowledge of the state of the qubits in IC-ensemble $\mathcal{E}_{21}$, where as ρ_2 is mixed, which represents our ignorance (incomplete knowledge) about the state of the qubits in the CC-ensemble. Hence there is loss of information in going from IC-ensemble $\mathcal{E}_{21}$ to ρ_2. This can be further justified as follows: Let $p_\pm = 1/2$. Then $\rho_2 = (|+\rangle\langle+| + |-\rangle\langle-|)/2 = \mathbb{1}_2/2$. But we also have $\rho_1 = (|0\rangle\langle 0| + |1\rangle\langle 1|)/2 = \mathbb{1}_2/2$ for $p_0 = p_1 = 1/2$. Hence ρ_1 and ρ_2 have become identical, even though ρ_1 contains $|0\rangle$s and $|1\rangle$s where as ρ_2 contains $|+\rangle$s and $|-\rangle$s. But IC-ensembles $\mathcal{E}_{11}$ and $\mathcal{E}_{21}$ are different (i.e., not identical). Hence in going from IC-ensemble $\mathcal{E}_{i1}$ to ρ_i there is loss of information, $i = 1, 2$. The extra information which was making IC-ensemble $\mathcal{E}_{11}$ different from IC-ensemble $\mathcal{E}_{21}$, has been lost in going from IC-ensemble $\mathcal{E}_{i1}$ to ρ_i, $i = 1, 2$, there by making ρ_1 and ρ_2 identical. Further $(\Delta\sigma_z)^2_{\text{eff},\mathcal{E}_{11}} = p_0(\Delta\sigma_z)^2_{|0\rangle} + p_1(\Delta\sigma_z)^2_{|1\rangle} = p_0 \times 0 + p_1 \times 0 = 0$, and $(\Delta\sigma_z)^2_{\text{eff},\mathcal{E}_{21}} = p_+(\Delta\sigma_z)^2_{|+\rangle} + p_-(\Delta\sigma_z)^2_{|-\rangle} = p_+ + p_- = 1$. This is because IC-ensembles $\mathcal{E}_{11}$ and $\mathcal{E}_{21}$ are different (for a similar example see [4]). Where as, $(\Delta\sigma_z)^2_{\text{eff},\rho_1} = \langle\sigma_z^2\rangle_{\rho_1} - \langle\sigma_z\rangle^2_{\rho_1} = 1 - 0 = 1$, and $(\Delta\sigma_z)^2_{\text{eff},\rho_2} = \langle\sigma_z^2\rangle_{\rho_2} - \langle\sigma_z\rangle^2_{\rho_2} = 1 - 0 = 1$. This is because ρ_1 and ρ_2 are identical.

Coalesced: Consider measuring σ_z nonselectively on a CC-ensemble of M identical copies of $|\psi\rangle = \sqrt{p_0}|0\rangle + \sqrt{p_1}|1\rangle$. Then the post measurement state is $\rho_1 = p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|$. Further $\langle\sigma_z\rangle_{|\psi\rangle} = p_0 - p_1 = \langle\sigma_z\rangle_{\rho_1}$ and $(\Delta\sigma_z)^2_{|\psi\rangle} = \langle\sigma_z^2\rangle_{|\psi\rangle} - \langle\sigma_z\rangle^2_{|\psi\rangle} = 1 - (p_0 - p_1)^2 = \langle\sigma_z^2\rangle_{\rho_1} - \langle\sigma_z\rangle^2_{\rho_1}$, hence this corresponds to variance in the initial state (ρ_1) preparation. Here sample mean, $S(= (T_+ - T_-)/M) \rightarrow \text{ND} : p_0 - p_1, (1 - (p_0 - p_1)^2)/M$ in the large M limit, where $T_\pm$ is the number of ± 1 outcomes, and $T_+ + T_- = M$.

4. Independent and normally distributed random variables

If $X_i \rightarrow \text{ND} : \mu_i, \sigma_i^2$ then $Z = \sum_{i=1}^{\tilde{N}} X_i \rightarrow \text{ND} : \sum_{i=1}^{\tilde{N}}\mu_i, \sum_{i=1}^{\tilde{N}}\sigma_i^2$ where X_is are normally distributed independent random variables [11]. 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 [11]. 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 $X_i \rightarrow \text{ND} : \mu_i, \sigma_i^2$ 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 $n \rightarrow \infty$, $i = 1, 2$. Then $Z = X_1 + X_2 = (1/n)(\sum_{j=1}^{n} X_{1j} + \sum_{j=1}^{n} X_{2j})$.

Appendix B: IC-Ensemble picture

1. Swaying of center of Gaussians

Another less rigorous way of arriving at the resultant variance in Eq. (11) i.e., $(1 - (\Delta\cos\theta_q)^2_{p_{\theta_q}})/M$ is the following: When we integrate over S_1 in Eq. (10), there is contribution to resultant variance from following two factors: (1) Swaying of center of Gaussians due to the varying mean i.e., $\mu_{\text{eff}} = \langle\cos\theta_q\rangle_{p_{\theta_q}} S_1$. (2) Variance arising from the measurement of σ_z on $|\theta_q\rangle$s and $|\theta_{q\perp}\rangle$s i.e., $(\Delta\sigma_z)^2_{\text{eff}}/M$ (below Eq. (9)). In Eq. (10) 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 $\mu_{\text{eff}} \rightarrow \text{ND} : 0, \langle\cos\theta_q\rangle^2_{p_{\theta_q}}/M$ (using the theorem in Appendix (B 4)). Hence the resultant variance is given by $\langle\cos\theta_q\rangle^2_{p_{\theta_q}}/M + (\Delta\sigma_z)^2_{\text{eff}}/M = (1 - (\Delta\cos\theta_q)^2_{p_{\theta_q}})/M$.

2. Discrimination via comparison

As $\Delta S_1'^2 < \Delta S_2'^2 (= 1/M \because S_2' = S_2)$, Bob can discriminate between the two IC-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., $\lim_{M\rightarrow\infty}\Delta S_1'^2/\Delta S_2'^2 \approx \lim_{M\rightarrow\infty}[(1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M]/[1/M] = 1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}}$, rather than via $\Delta S_i'^2$ alone, $i = 1$ or 2. This is possible in a special case: $\{\theta_1(=0), \theta_2(=\pi)\} \rightarrow \{p^o_0, p^o_\pi\}$. In this case Bob can obtain both $\Delta S_1'^2$ (and hence $f(S_1')$) and $\Delta S_2'^2$ (and hence $g(S_2') = g(S_1) = g(S_1)$) from the given IC-ensemble $\mathcal{E}_i, i = 1$ or 2 (Appendix (B 6)).

3. Equivalence of S_1 and S_2

Consider the following alternate preparation procedure which also gives exactly same state of the qubits in IC-ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ as that given in the main text (II): Alice prepares an IC-ensemble $\mathcal{E}_1$ ($\mathcal{E}_2$) of qubit states by measuring σ_z (σ_x, Pauli-x matrix) individually on each of the N qubits in an IC-ensemble where each qubit is in the state $|+\rangle$ ($|0\rangle$), and renormalizing the post measurement state. Then T_2^+, T_2^- are also the number of $|0\rangle$s, $|1\rangle$s respectively in the IC-ensemble $\mathcal{E}_{1j}$ ($\because$ $|+\rangle$ and $|-\rangle$ are equivalent with respect to σ_z measurement outcomes (Appendix (B 7)). Hence measuring σ_z individually on each of the M qubits in the IC-ensemble $\mathcal{E}_{2j}$ is equivalent to measuring σ_z individually on each of the M identical copies of $|+\rangle$ in an IC-ensemble. But IC-ensemble $\mathcal{E}_{1j}$ has also been obtained by measuring σ_z individually on each of the M identical copies of $|+\rangle$ in an IC-ensemble). $\Rightarrow$ $T_2^{\pm}$ is also the total number of ±1 outcomes obtained by measuring σ_z individually on each of the M qubits in the IC-ensemble $\mathcal{E}_{1j}$ ($\because$ variances $(\Delta\sigma_z)^2_{|0\rangle} = \langle\sigma_z^2\rangle_{|0\rangle} - \langle\sigma_z\rangle^2_{|0\rangle} = 0, (\Delta\sigma_z)^2_{|1\rangle} = 0$ where $\langle\sigma_z\rangle_{|0\rangle} = \mathrm{Tr}(\sigma_z|0\rangle\langle 0|)$) i.e., $T_2^{\pm} \equiv T_1^{\pm}$. Hence $S_1 \equiv S_2$. Further, sample mean S_i has mean $\langle S_i\rangle = \langle\sigma_z\rangle_{|+\rangle} = 0$, and variance $\Delta S_i^2 = (\Delta\sigma_z)^2_{|+\rangle}/M = 1/M$ ($\because$ $S_1 \equiv S_2$ as explained above, and S_1 corresponds to sample mean of outcomes of individual σ_z measurements on each of the M identical copies of $|+\rangle$ in an IC-ensemble (as explained above)). Now it is easily evident that $S_1 = S_1^A$ ($\because$ variances $(\Delta\sigma_z)^2_{|0\rangle} = 0, (\Delta\sigma_z)^2_{|1\rangle} = 0$).

4. Probability distribution of $T_1^{\pm}, p_1^{\pm}$, and p_{θ_q}

We have $S_1 \to \mathrm{ND} : 0, 1/M$ where $S_1 = (T_1^+ - T_1^-)/M, T_1^+ + T_1^- = M$. $\Rightarrow T_1^{\pm} = M(1 \pm S_1)/2$. Consider the following theorem: If $X \to \mathrm{ND} : \mu, \sigma^2$ then $Y(= aX + b) \to \mathrm{ND} : a\mu + b, a^2\sigma^2$ [11]. Using this we obtain $T_1^{\pm} \to \mathrm{ND} : M/2, M/4$. We have defined $p_1^{\pm} = T_1^{\pm}/M$. Again using the above theorem we obtain $p_1^{\pm} \to \mathrm{ND} : 1/2, 1/(4M)$.

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 $S(= (m_1 - m_2)/M) \to \mathrm{ND} : 2p^o_{\theta_1} - 1, (\Delta\sigma_z)^2_{|\zeta\rangle}/M$ where $m_1(m_2)$ is the number of $+1(-1)$ outcomes. $(\Delta\sigma_z)^2_{|\zeta\rangle} = 1 - \langle\sigma_z\rangle^2_{|\zeta\rangle}, m_1 + m_2 = M$. $\Rightarrow m_j(= M(1 + (-1)^{j+1}S)/2) \to \mathrm{ND} : p^o_{\theta_j}M, \sigma^2_{m_j}$ where $\sigma^2_{m_j} = (\Delta\sigma_z)^2_{|\zeta\rangle}M/4, j = 1, 2$ (using the theorem stated above). $\Rightarrow p_{\theta_j}(= m_j/M) \to \mathrm{ND} : p^o_{\theta_j}, \sigma^2_{m_j}/M^2$, $j = 1, 2$ (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 to work.

5. Indirect exact evaluation of the integral in Eq.(12)

We may indirectly evaluate the integral in Eq. (12) as follows: As both weighing function and the function being weighed are 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

$$\begin{aligned}\Delta S_1'^2 &= \int_{-\infty}^{\infty} dp_{\theta_1}(\mathrm{Nd}(p_{\theta_1}) : p^o_{\theta_1}, \sigma^2_{m_1}/M^2)(1 - p_{\theta_1}(1 - p_{\theta_1})(\cos\theta_1 - \cos\theta_2)^2)/M \\ &= (1 - (p^o_{\theta_1}(1 - p^o_{\theta_1}) - \sigma^2_{m_1}/M^2)(\cos\theta_1 - \cos\theta_2)^2)/M. \end{aligned} \tag{B1}$$

Hence $f(S_1') = (\mathrm{Nd}(S_1') : 0, \Delta S_1'^2)$. Here we are integrating from $-\infty$ to ∞ because of a reason similar to [16]. 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. (13) as required.

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

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

If Bob has got the IC-ensemble $\mathcal{E}_1$, then sample mean $S_1 \rightarrow \mathrm{ND} : 0, 1/M$ corresponds to before applying $(\theta_q)_x$s, where as the sample mean $S_1' \rightarrow \mathrm{ND} : 0, (1 - 4p_0^o(1 - p_0^o))/M$ (for $\{\theta_1(=0), \theta_2(=\pi)\} \rightarrow \{p_0^o, p_\pi^o\}$) corresponds to after applying $(\theta_q)_x$s. In this case Bob can construct the probability distribution of even S_1 via inverse mapping $\mathcal{E}_1 = F^{-1}(\mathcal{E}_1')$, as explained above.

However, if Bob has got the IC-ensemble $\mathcal{E}_2$, then sample mean $S_2' \rightarrow \mathrm{ND} : 0, 1/M$ 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, which leads to discrimination). From the IC-ensemble $\mathcal{E}_2''$ got by measuring σ_z individually on each of the N qubits in the IC-ensemble $\mathcal{E}_2'$, if Bob works backward via the sequence of $(\theta_q)_x$s that he had applied (i.e., the mapping $\mathcal{E}_2''' = F(\mathcal{E}_2'')$), then he obtains a virtual IC-ensemble $\mathcal{E}_2'''$ which gives sample mean $\rightarrow \mathrm{ND} : 0, (1 - 4p_0^o(1 - p_0^o))/M$. Hence Bob can construct the probability distribution of a virtual sample mean which is same as that of S_1'.

Hence Bob can obtain both $\Delta S_1'^2$ and $\Delta S_2'^2$ from the given IC-ensemble $\mathcal{E}_i, i = 1$ or 2. If $i = 1$ then $\Delta S_1'^2$ and $f(S_1')$ corresponds to after applying $(\theta_q)_x$s where as $\Delta S_2'^2 (= \Delta S_1^2 \because S_2' = S_2 \equiv S_1)$ and $g(S_2')(= g(S_2) = g(S_1))$ corresponds to before applying $(\theta_q)_x$s. If $i = 2$ then $\Delta S_1'^2$ and $f(S_1')$ (which are virtual) corresponds to before applying $(\theta_q)_x$s where as $\Delta S_2'^2$ and $g(S_2')$ corresponds to after applying $(\theta_q)_x$s. Hence Bob can also discriminate by comparing the entire probability density functions $f(S_1')$ and $g(S_1)(= g(S_2) = g(S_2'))$.

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

Consider an IC-ensemble having $T_\pm$ number of $|\pm\rangle$s. Let $p_\pm = T_\pm/M$ where $T_+ + T_- = M$, and M is sufficiently large. Then, $\mu_{\rm eff} = p_+\langle\sigma_z\rangle_{|+\rangle} + p_-\langle\sigma_z\rangle_{|-\rangle} = 0$, and hence independent of p_+, p_-. $(\Delta\sigma_z)^2_{\rm eff} = p_+(\Delta\sigma_z)^2_{|+\rangle} + p_-(\Delta\sigma_z)^2_{|-\rangle} = 1 (\because (\Delta\sigma_z)^2_{|+\rangle} = (\Delta\sigma_z)^2_{|-\rangle} = 1)$, again independent of p_+, p_-. $\Rightarrow$ variance of effective sample mean $\Delta S^2_{\rm eff} = (\Delta\sigma_z)^2_{\rm eff}/M = 1/M$. Hence all the results are same as measuring σ_z individually on each of the M copies of $|+\rangle$s or $|-\rangle$s in an IC-ensemble. 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.

8. Explanation using central limit theorem

We can explain the reduction in population difference (and hence variance) via central limit theorem as follows: Let $\{\theta_1(=0), \theta_2(=\pi)\} \rightarrow \{p_0^o(=1/2), p_\pi^o(=1/2)\}$. 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, whereas getting $T_1'^+ \approx T_1'^-$ 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. (9) (i.e., $S_1' \rightarrow \mathrm{ND} : \mu_{\rm eff}, (\Delta\sigma_z)^2_{\rm eff}/M$) is due to central limit theorem. Hence the results in Eq.s (12, 13) (with $\{\theta_1(=0), \theta_2(=\pi)\} \rightarrow \{p_0^o(=1/2), p_\pi^o(=1/2)\}$) 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 Appendix (D 2)).

9. 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 IC-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 $|1\rangle$ [18]. Then, $|T_1^+ - T_1^-| = M$ (extreme case) corresponds to $P_1 = 1$ or $P_2 = 1$ where $P_1 = T_1^+/M, P_2 = T_1^-/M$ ($\because T_1^+, T_1^-$ are also the number of $|0\rangle$s, $|1\rangle$s respectively). $\Rightarrow H = 0$ i.e., minimum entropy configuration. When we introduce a new independent random variable θ_q such that $\{\theta_1(=0), \theta_2(=\pi)\} \rightarrow \{p_0^o(=1/2), p_\pi^o(=1/2)\}$ 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 $|T_1'^+ - T_1'^-| \approx 0$ in the large M limit.

10. Nonlinearity in action

We saw in the main text (II B) that 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 $\langle\sin^2\theta_q\rangle_{p_{\theta_q}}$ they come out as such (linear operation). Hence we call this linear channel, and it corresponds to measuring σ_z selectively on $|\theta_q\rangle$s and $|\theta_{q\perp}\rangle$s (see the section 'Swaying of center of Gaussians' in Appendix (B 1)). Hence there is reduction in variance due to nonlinear effect.

Now we can justify the result obtained in Eq. (13) as follows: In Eq. (12) 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. (10) 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 $p^o_{\theta_q}$.

Asymmetry (nonlinear and linear) in the two channels mentioned above 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 IC-ensemble $\mathcal{E}_{1j}$ approximately $Mp^o_{\pi/2}$ number of qubit 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 qubit states on y-axis there is positive contribution to the resultant variance as $\sin(\pi/2) = 1$. Similarly the qubit 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 qubit states on z-axis do 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 $1/M$.

Further, the result $\Delta S'^2 = (p^{o2}_0 + p^o_{\pi/2})/M < 1/M$ (main text (II B)) is counter intuitive, because intuitively if Bob rotates $\tilde{N}$ number of qubit states on z-axis (on Bloch sphere) on to y-axis, it is as if he has measured σ_z on $M - \tilde{N}$ number of $|+\rangle$s ($\because$ IC-ensemble $\mathcal{E}_{1j}$ can also be obtained by measuring σ_z selectively on an IC-ensemble of M identical copies of $|+\rangle$ (Appendix (B 3))) and $\tilde{N}$ more are to be measured ($\because$ with respect to σ_z measurement outcomes, eigenkets of Pauli-y matrix σ_y, $|\pm\rangle_y$, are equivalent to $|\pm\rangle$). Hence after measuring $\tilde{N}$ more, Bob should get back variance $1/M$. A closer look shows that, when Bob is 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 Bob 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 that in previous paragraph.

11. Smoothing out non uniformities

Let $\{\theta_1(=0), \theta_2(=\pi)\} \to \{p^o_{\theta_1}(=1/2), p^o_{\theta_2}(=1/2)\}$. $\Rightarrow \Delta S_1'^2 \approx 0/M$. 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 ($\equiv (T_1'^+ \approx T_1'^-)$). Situation here is also 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 time averaged (dynamic) COM lies at $(\sum_i m_i\vec{r}(t_i))/\sum_i m_i = \langle\vec{r}(t)\rangle_{\delta t} = 0(\equiv (T_1'^+ - T_1'^- \approx 0))$ 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. (B1), 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 $\sigma^2_{m_1}/M^2$ of $p^o_{\theta_1}$), still nonuniformities will be smoothed out.
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.

12. Why the reduction in variance?

Variance of random variable σ_z in the state $|\psi\rangle$ 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 IC-ensemble $\mathcal{E}_1$ already corresponds to maximum possible variance ($\because S_1 \to$ ND : $0, 1/M$). Hence, introduction of a new independent random variable θ_q [which does not increase the number of qubit states ($= M \times M_1$)(see the section 'Motivation' in main text (II A))] can only decrease the variance. We can write $S_1' = h(S_1, \cos\theta_q)$ 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 $S_1' \to$ 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

$(+,-)$.

13. Why the reduction in variance unlike in Z ?

Let X_i be the 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 Z.

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

If Bob has IC-ensemble $\mathcal{E}_2$, then even if M, M_1 are not very large, he obtains 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 $\text{ND} : 0, 1/M$. But $f(S_1')$ in Eq. (10) is a complicated weighted mean of very large number of Gaussians. There seems to be no simple way of obtaining $f(S_1') \approx \text{Nd}(S_1') : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$. This may be shown as follows: Consider $|\delta\rangle = \cos(\delta/2)|0\rangle + \sin(\delta/2)|1\rangle$. Then $\langle\sigma_z\rangle_{|\delta\rangle} = \cos\delta$, $(\Delta\sigma_z)^2_{|\delta\rangle} = \sin^2\delta$. In the large M limit, sample mean $\rightarrow \text{ND} : \cos\delta, \sin^2\delta/M$. Let $\text{ND} : \cos\delta, \sin^2\delta/M = \text{ND} : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$. $\Rightarrow \delta = \pi/2$. 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/single coin with respect to effective variance (Appendix (A 2)). Hence it seems $\text{ND} : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$ can be got only as the resultant of complicated weighted mean of a large number of different Gaussians as given in Eq. (10). Even to realize one of the component Gaussians in Eq. (10) (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 $f(S_1') \approx \text{Nd}(S_1') : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$, 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, Bob may get $f(S_1') \approx g(S_2') (= \text{ND} : 0, 1/M)$, and hence no discrimination. Only when M, M_1 are really large, Bob may obtain $f(S_1') \approx \text{Nd}(S_1') : 0, (1 - (\Delta\cos\theta_q)^2_{p^o_{\theta_q}})/M$.
Bob is trying to change/deviate from $\text{ND} : 0, 1/M$. At low M, M_1 the deviation/change 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.

15. 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}s)

Reasons for this 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 (where measuring σ_z on qubit states is equivalent to X_{ij}s in Appendix (A 4)), which is unlike in Appendix (A 4) where Z corresponds to $2n$ number of X_{ij}s, whereas X_i corresponds to n number of X_{ij}s. (3) S_1 is continuous, where as θ_q is discrete.

16. Order of integration does not matter

In Eq. (10) 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. (11). Further multiple integration is nothing but sum over multiple indices. As long as indices are independent, order in which the indices are summed over should not matter.

Appendix C: Single copy picture

1. As a deterministic but inexact nonorthogonal state discrimination problem

Linear and unitary nature of quantum mechanics forbids cloning an unknown state chosen from a set of nonorthogonal states [19]. Hence, given a single copy of a pure state chosen from a set of nonorthogonal states, Bob cannot know the given state both *exactly* and *deterministically*. However, exact but probabilistic nonorthogonal state discrimination is possible. E.g., given a single copy of $|0\rangle$ or $|+\rangle$, Bob can know the unknown state exactly but only probabilistically using POVM measurement [20]. 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: Consider the two sets $\mathcal{F}_1, \mathcal{F}_2$ defined in the main text (Eq. (1)). The problem considered in the main text (II) is exactly equivalent to the following: Alice gives Bob, a *single copy* of $|\phi_{ij}\rangle$ chosen with probability $1/2^N$ from $\mathcal{F}_i$ (i.e., all the states in $\mathcal{F}_i$ are equally likely to be chosen), $i = 1$ or 2. Bob knows that the given state is either one of $|\phi_{1j}\rangle$s or one of $|\phi_{2j}\rangle$s. Bob has to find out whether the given state $|\phi_{ij}\rangle$ was chosen from $\mathcal{F}_1$ or $\mathcal{F}_2$. Qubits in the IC-ensemble $\mathcal{E}_i$ are in one of the states $|\phi_{ij}\rangle$s which Alice has chosen from $\mathcal{F}_i$, $i = 1$ or 2. What we have shown is that, in the large N limit, even though Bob cannot know the unknown state $|\phi_{ij}\rangle$ exactly ($\because$ he cannot know the state of each of the N qubits in it), still he can know deterministically (i.e., with probability tending to one, in the large N limit) whether it was chosen from $\mathcal{F}_1$ or $\mathcal{F}_2$ (and hence it is deterministic but inexact nonorthogonal state discrimination). Note that even though $|\langle\phi_{1j}|\phi_{2k}\rangle|$ decreases as N increases, $|\phi_{2k}\rangle$ can never become exactly orthogonal to $|\phi_{1j}\rangle$, because $\mathcal{F}_i$ is already a complete set of orthonormal basis states, $i = 1, 2$. Hence knowing whether $|\phi_{ij}\rangle$ was chosen from $\mathcal{F}_1$ or $\mathcal{F}_2$ is a nontrivial deterministic but inexact nonorthogonal state discrimination problem (Appendix (C 2)). Further, as each of $\mathcal{F}_1$, $\mathcal{F}_2$ is a complete set of basis states, this can also be viewed as a basis discrimination problem.

2. Nontrivial nonorthogonal states

Consider two sets $\mathcal{D}_1 = \{|0\rangle|0\rangle, |0\rangle|1\rangle\}$ and $\mathcal{D}_2 = \{|w\rangle|0\rangle, |w\rangle|w\rangle\}$ where $|w\rangle = \cos((\pi - \epsilon)/2)|0\rangle + \sin((\pi - \epsilon)/2)|1\rangle$, and $\epsilon \to 0$. $\langle 0|w\rangle = \cos((\pi - \epsilon)/2)$. $\mathcal{D}_i$ is not a complete set of orthonormal basis states (CSOBS), $i = 1, 2$. $|w\rangle \to |1\rangle$ in the limit $\epsilon \to 0$, and hence $\mathcal{D}_1$ and $\mathcal{D}_2$ together constitute a CSOBS. Hence given a state chosen from $\mathcal{D}_1$ or $\mathcal{D}_2$, knowing whether it was chosen from $\mathcal{D}_1$ or $\mathcal{D}_2$ is equivalent to (in the limit $\epsilon \to 0$) orthogonal state discrimination. Hence we call this, trivial nonorthogonal state discrimination. Moreover, by measuring a nondegenerate observable whose eigenkets are $|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle$, we can discriminate between the states in $\mathcal{D}_1, \mathcal{D}_2$ with probability tending to one (in the limit $\epsilon \to 0$).

However it is not so in case of $\mathcal{F}_1$ and $\mathcal{F}_2$. $\mathcal{F}_i$ is itself a CSOBS, $i = 1, 2$. Hence, even in the limit $N \to \infty$, $\mathcal{F}_1$ and $\mathcal{F}_2$ together cannot form a CSOBS unlike in the previous case. Hence knowing whether the given state $|\phi_{ij}\rangle$ was chosen from $\mathcal{F}_1$ or $\mathcal{F}_2$ cannot be reduced to orthogonal state discrimination unlike in previous case. Further, even though $|\langle\phi_{1j}|\phi_{2k}\rangle| \to 0$ in the limit $N \to \infty$, $|\phi_{1j}\rangle$ cannot be exactly orthogonal to $|\phi_{2k}\rangle$ because the set $\mathcal{F}_i$ already forms a CSOBS, $i = 1, 2$. Hence even in the limit $N \to \infty$, discriminating between $|\phi_{1j}\rangle$s and $|\phi_{2k}\rangle$s is a nontrivial nonorthogonal state discrimination problem. This is also justified by the fact that, 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 know even with probability just greater than zero, whether the given state was chosen from $\mathcal{F}_1$ or $\mathcal{F}_2$, unlike in the case of $\mathcal{D}_1, \mathcal{D}_2$. This is because $|\phi_{2k}\rangle$ is a state of equal superposition of all the states in $\mathcal{F}_1$. $|\phi_{2k}\rangle$ is the state which is equidistant from each of the states in $\mathcal{F}_1$, and hence in some sense most nonorthogonal simultaneously to each of the states in $\mathcal{F}_1$. This is easily evident for the case $N = 1$. The fact that $|\langle\phi_{1j}|\phi_{2k}\rangle|$ decreases as N increases, seems to be analogous to the following example: In d dimensional hyper space, the ratio of volume of a sphere of radius $1-\epsilon$ to the volume of unit sphere turns out to be $(1-\epsilon)^d$ ($\because$ volume is proportional to R^d where R is the radius). For a fixed ϵ such that $0 < \epsilon < 1$, $(1 - \epsilon)^d$ decreases as d increases i.e., most of the volume tends to accumulate near the surface of the unit sphere.

Suppose Alice tells Bob, her outcomes of coin tosses used to prepare the states $|\phi_{1j}\rangle, |\phi_{2j}\rangle$ (see the preparation procedure in the main text (II)). Then for Bob the problem reduces to discriminating between just the two states $|\phi_{1j}\rangle$ and $|\phi_{2j}\rangle$ where j is known to Bob. The problem reduces to sort of orthogonal state discrimination in the large N limit, as the inner product $\langle\phi_{1j}|\phi_{2j}\rangle$ tends to zero. Bob assigns value $+1$ to Alice's outcome Head and -1 to Tail. Then he measures σ_z individually on each of the qubits in the unknown state $|\phi_{ij}\rangle, i = 1$ or 2. If the outcomes match exactly with that of Alice's, then he comes to know that the given state is $|\phi_{1j}\rangle$, else he comes to know that it was $|\phi_{2j}\rangle$.

3. Physical difference between ρ_1 and ρ_2

$|\phi_{1j}\rangle$ ($|\phi_{2j}\rangle$) physically has $|0\rangle$s and $|1\rangle$s ($|+\rangle$s and $|-\rangle$s) arranged in some random sequence. Hence $|\phi_{1j}\rangle$ and $|\phi_{2j}\rangle$ are physically different even for Bob. This can be justified as follows: Suppose Alice prepares the qubits in the IC-ensemble $\mathcal{E}_1$ ($\mathcal{E}_2$) in the state $|\phi_{1j}\rangle$ ($|\phi_{2j}\rangle$) by measuring σ_z (σ_x) individually on each of the N identical copies of $|+\rangle$ ($|0\rangle$) in an IC-ensemble. This preparation procedure is exactly equivalent to that given in the main text (II). Let Alice give Bob $|\phi_{ij}\rangle$, and also let her tell Bob, her sequence of measurement outcomes, $i = 1$ or 2. Then Bob measures σ_z selectively on each of the N qubits in the unknown state $|\phi_{ij}\rangle$, $i = 1$ or 2. If his measurement outcomes exactly matches with that of Alice, then he comes to know that the given state is $|\phi_{1j}\rangle$, else he comes to know that it was $|\phi_{2j}\rangle$ [4] (this is sort of orthogonal state discrimination, see Appendix (C 2)). This clearly shows that ρ_1 (ρ_2) is just a mathematical description of the over all (i.e., taking into consideration all possibilities) state of the qubits in the IC-ensemble $\mathcal{E}_1$ ($\mathcal{E}_2$) from Bob's ignorant perspective, which physically has $|0\rangle$s and $|1\rangle$s ($|+\rangle$s and $|-\rangle$s). In our protocol, Alice do not tell Bob, her measurement outcomes. But this cannot change the actual physical content of ρ_i. Purification of ρ_i using the information supplied by central limit theorem, and subsequently by selectively applying $(\theta_q)_x$s, and measuring σ_z, implies gain of knowledge about the unknown state, there by leading to discrimination. Further it has been shown in [21] that, one can also discriminate between states similar to ρ_1 and ρ_2 using deterministic nonlinear evolution (but there ρ_1, ρ_2 corresponds to CC-ensembles).

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

Consider the following game: Alice has a single qubit. She tosses an unbiased coin and if the outcome is Head, she prepares the qubit in the state $|0\rangle$, else she prepares it in the state $|1\rangle$. Then she gives the qubit to Bob. Bob is aware of preparation procedure. He need to find out the state of the qubit. For Bob, state of the single qubit 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 the qubit is in the state $|0\rangle$. In density matrix language, $\rho_B \to \rho_B^0/Tr(\rho_B^0) = |0\rangle\langle 0|$, a renormalized pure state which indicates gain of full knowledge about the unknown state, where $\rho_B^0 = |0\rangle\langle 0|\rho_B|0\rangle\langle 0|$. $|0\rangle\langle 0|$ is a linear but nonunitary operator which does projection. Similarly, if he gets the outcome -1, then he comes to know that the qubit is in the state $|1\rangle$ i.e., $\rho_B \to \rho_B^1/Tr(\rho_B^1) = |1\rangle\langle 1|$ 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 CC-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 versus nonselective CC-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 a CC-ensemble of qubits initialized in the state $\rho_{in} = \mathbb{1}_2/2$, then post measurement state of the full CC-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 cannot purify the state unlike in single copy measurement.

5. Single copy versus nonselective CC-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 $\Pi_0 = |0\rangle\langle 0|$ [12]. In density matrix language 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 a CC-ensemble of n identical copies of $|m\rangle$. If we measure σ_z nonselectively, then the unnormalized post measurement 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 subensemble corresponding to -1 outcome is given by $\rho_1 = \Pi_1\rho_m\Pi_1 = \sin^2(\theta/2)|1\rangle\langle 1|$ where $\Pi_1 = |1\rangle\langle 1|$. Normalized state of the full CC-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 [22].

6. Purification solely via information

Bob's unknown state is given by $\rho^B = \alpha_1\rho_1 + \alpha_2\rho_2 = \mathbb{1}_{2^N}/2^N$ (Eq. (2)). Suppose Alice tells Bob, whether she has given him IC-ensemble $\mathcal{E}_1$ or $\mathcal{E}_2$, say, $\mathcal{E}_1$. Then Bob's state collapses as follows: $\rho^B \to \alpha_1\rho_1$. Renormalizing, Bob obtains ρ_1. Next suppose Alice also tells Bob, the exact state of each of the N qubits in the given IC-ensemble $\mathcal{E}_1$. Then even without Bob doing any operation (unitary/nonunitary, linear/nonlinear), his state further gets projected as follows: $\rho_1 \to |\phi_{1j}\rangle\langle\phi_{1j}|/2^N$ where value of j is known to Bob. Again renormalizing, Bob obtains $|\phi_{1j}\rangle\langle\phi_{1j}|$. This clearly shows that in case of an unknown pure state or IC-ensemble, mixedness represents mere ignorance, which can be purified via pure information (e.g., information supplied by central limit theorem in the main text (Eq. (3))), even without any kind of operations.

7. U_ls increases the purity of $\tilde{\rho}_1$ via central limit theorem

Consider $\{\theta_1(= 0), \theta_2(= \pi)\} \to \{p_0^o(= 1/2), p_\pi^o(= 1/2)\}$. Then, variance $\Delta S_1'^2 \approx 0/M$ (Eq. (13)). Using Eq. (5) we get $\hat{\rho}_1' = \sum_j \hat{p}_j'|\phi_{1j}\rangle\langle\phi_{1j}|$ (also see Appendix (B 8) for further justification). Obviously number of $|\phi_{1j}\rangle$s in $\hat{\rho}_1'$ will be much less than that in $\tilde{\rho}_1$ ($\because$ the former corresponds to variance $0/M$ whereas the latter corresponds to variance $1/M$), and hence purity (von Neumann entropy) of $\hat{\rho}_1'$ must be greater (less) than that of $\tilde{\rho}_1$. Hence $\mathrm{Tr}(\hat{\rho}_1'^2) > \mathrm{Tr}(\tilde{\rho}_1^2)$.

Note that the above result does not contradict the result in Appendix (B 9) where we showed that with $\{\theta_1(= 0), \theta_2(= \pi)\} \to \{p_0^o(= 1/2), p_\pi^o(= 1/2)\}$ Shannon entropy increases. There we were looking at the number of $|0\rangle$s and $|1\rangle$s in a particular $|\phi_{1j}\rangle$, but not the whole state $\tilde{\rho}_1$ (Eq. (3)). Note that even a state $|\phi_{1j}\rangle$ with equal number of $|0\rangle$s and $|1\rangle$s has maximum Shannon entropy (as calculated in Appendix (B 9)) where as minimum von Neumann entropy (because we are looking only at a particular $|\phi_{1j}\rangle$, and as we have assumed that we know how many $|0\rangle$s and $|1\rangle$s it has, and further if we assume that we also know their sequence, it is a pure state). There we introduced a new random variable $(\theta_q)_x$ into a particular pure state $|\phi_{1j}\rangle$ and hence obviously it should increase the randomness and hence Shannon entropy as calculated over there. But here we are pumping in information into the whole state $\tilde{\rho}_1$, consequently purity (von Neumann entropy) of the mixed state must increase (decrease). The thing we are concerned here is different from that in Appendix (B 9). Hence no contradiction.

8. Sort of nonlinear unitary evolution

We can also rewrite ρ_1' (Eq. (4)) as follows: $\rho_1' = \sum_{l=1}^{d^N} P_l U_l \tilde{\rho}_1 U_l^\dagger$. Then we obtain

$$\mathrm{Tr}(\rho_1'^2) = \mathrm{Tr}(\tilde{\rho}_1^2)\sum_l P_l^2 + \sum_{l,k\neq l} P_l P_k \mathrm{Tr}(U_l \tilde{\rho}_1 U_l^\dagger U_k \tilde{\rho}_1 U_k^\dagger)$$

$$\neq \mathrm{Tr}(\tilde{\rho}_1^2)(\because\ \tilde{\rho}_1 \neq \frac{\mathbb{1}_{2^N}}{2^N}, \text{ and } l > 1) \Rightarrow \mathrm{Tr}({\rho_1'}^2) \neq \mathrm{Tr}(\tilde{\rho}_1^2).$$

But $\rho_2' = \tilde{\rho}_2$ ($\because$ $(\theta_q)_x$ introduces an insignificant global phase to $|+\rangle, |-\rangle$ in $|\phi_{2j}\rangle$s). $\Rightarrow$ $\mathrm{Tr}({\rho_2'}^2) = \mathrm{Tr}(\tilde{\rho}_2^2)$. This clearly shows that $(\theta_q)_x$s does nothing to $\tilde{\rho}_2$ and does something to $\tilde{\rho}_1$ and hence may give rise to some observable effect which may lead to discrimination. Further

$$\mathrm{Tr}\left(\rho_i' \rho_j'\right)$$
$$= \mathrm{Tr}(\tilde{\rho}_i \tilde{\rho}_j)\sum_l P_l^2 + \sum_{l,m\neq l} P_l P_m \mathrm{Tr}(U_l \tilde{\rho}_i U_l^\dagger U_m \tilde{\rho}_j U_m^\dagger)$$
$$\neq \mathrm{Tr}(\tilde{\rho}_i \tilde{\rho}_j) \text{ in general for } i \neq j (\because\ \tilde{\rho}_i \neq \frac{\mathbb{1}_{2^N}}{2^N}, \text{ and } l > 1).$$

$\Rightarrow$ $\mathrm{Tr}((\rho'^B)^2) \neq \mathrm{Tr}((\tilde{\rho}^B)^2)$. Hence there is change in purity. Hence the net effect (of having individual control, working with a single copy of $|\phi_{ij}\rangle$, and introducing a new independent random variable θ_q via $(\theta_q)_x$) is as if Bob is evolving $\tilde{\rho}^B$ under a *nonlinear unitary operator*. Hence the purity of $\tilde{\rho}^B$ has changed. Note that no linear unitary operation can do this. Notice how purity further changed, as Bob pumped in the information that $\tilde{\rho}_1$ contains $|0\rangle$s and $|1\rangle$s and hence U_ls must affect them, and that $\tilde{\rho}_2$ contains $|+\rangle$s and $|-\rangle$s and hence U_ls cannot affect them.

9. 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.

10. 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 the linear operator 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 nonorthogonal state discrimination (because, after mapping if we measure σ_z and if the outcome is $+1$, then we come to know only that the given state was $|0\rangle$ or $|1\rangle$. Similar thing with -1 outcome). Hence deterministic but inexact nonorthogonal state 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 may do deterministic but inexact nonorthogonal state 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 (above) may require nonlinear evolution.

11. Nonlinear evolution seems to be necessary

A linear operator can clone at the most 2^N nonorthogonal states in 2^N-D Hilbert space (Appendix (C 9)). 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 10)). 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 in our protocol 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. Note that in Appendix (C 2) we showed that $|\phi_{1j}\rangle$s and $|\phi_{2k}\rangle$s are nontrivially nonorthogonal. Even if we discard those states among $|\phi_{1j}\rangle$s and $|\phi_{2j}\rangle$s, whose probability of Bob getting them tends to zero in the large N limit (e.g., $|0\rangle^{\otimes N}$), 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 10), even deterministic but inexact discrimination between $|\phi_{1j}\rangle$s and $|\phi_{2j}\rangle$s may also require nonlinear evolution.

We showed that it is the nonlinear effect (reduction in variance) (main text (II B)) which leads to discrimination. Further it makes sense to say that, nonlinear effect could be a consequence of some nonlinear evolution.

Following are the likely sources of nonunitary, nonlinear evolution in our protocol: (1) Individual/selective projective measurements. (2) Information supplied by central limit theorem which purifies the state at various stages of the protocol. (3) Application of random-x rotation i.e., $(\theta_q)_x$ individually to each of the qubit states in the given IC-ensemble (this is justified by the fact that $\mathrm{Tr}({\rho_1'}^2) \neq \mathrm{Tr}(\tilde\rho_1^2)$, see Appendix (C 8)).

12. Power of a single quantum system

To build a portable quantum computer we need to manipulate single quantum systems. NMR being a CC-ensemble, is considered only as a test bed for quantum information protocols, but not a candidate for ultimate portable quantum computer. Whereas NV center, SQUID, trapped ion, cold atoms etc., where we can manipulate single quantum systems, are considered as ultimate candidates to build a portable quantum computer. IC-ensembles can be realized in these architectures. 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, $j = 1$ or 2 or ... or 2^N.

Appendix D: Numerical simulation

1. Reduction in variance

(Continued from the main text (III)) A few more simulation results are plotted in Fig.s (3, 5, 7 (c)). Details about the seed corresponding to Fig. (2) is given in Fig. (4).

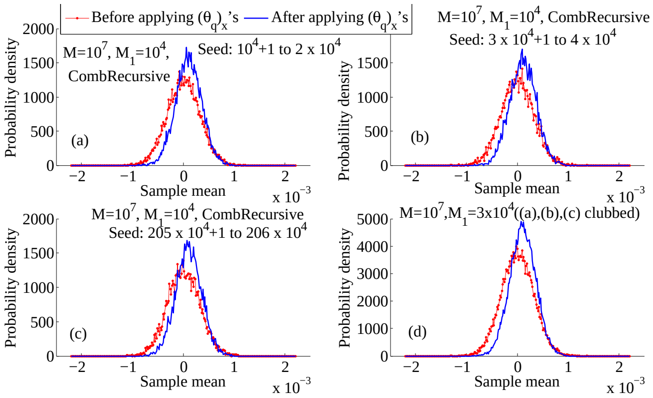

FIG. 3. (Color online) 'Seed' is the PRN generator's seed value. We used different seed for each of the M_1 number of sample mean points. (a) $A_g = 0.6795, A_f = 0.7445$. As it is evident from the figure (a), centers of two Gaussians are not perfectly coinciding (this may be due to small M_1). If we make them to coincide, 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) $A_g = 0.6811, A_f = 0.7442, A_g' = 0.6811, A_f' = 0.7824$. Seed corresponding to applying $(\theta_q)_x$s in (a) was from 1 to 10^4, in (b) was from $2 \times 10^4 + 1$ to 3×10^4, and in (c) was from $204 \times 10^4 + 1$ to 205×10^4.

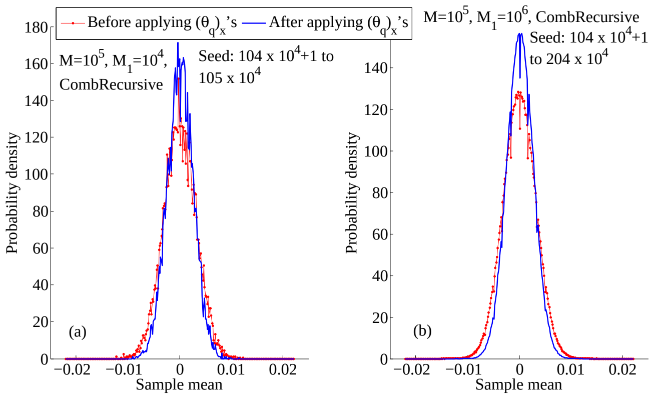

FIG. 4. (Color online) (b) $A_g' = 0.683286$, $A_f' = 0.780642$. Seed corresponding to applying $(\theta_q)_x$s in (a) was from $4 \times 10^4 + 1$ to 5×10^4, and in (b) was from $4 \times 10^4 + 1$ to 104×10^4.

2. 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 IC-ensemble $\mathcal{E}_i, i = 1$ or 2 (B 2). Here we considered the case $\{\theta_1(= 0), \theta_2(= \pi)\} \to \{p_0^o(= 1/2), p_\pi^o(= 1/2)\}$. Let $S_1' = \Delta S_1' \tilde S_1'$ where $\tilde S_1' \to \mathrm{ND} : 0, \Delta \tilde S_1'^2$. $\Rightarrow S_1' \to \mathrm{ND} : 0, \Delta S_1'^2 \Delta \tilde S_1'^2$ (using theorem in Appendix (B 4)). But we have $S_1' \to \mathrm{ND} : 0, \Delta S_1'^2$. $\Rightarrow \Delta \tilde S_1'^2 = 1$. Also $S_1' = (T_1'^+ - T_1'^-)/M$. Substituting $\theta_1 = 0, \theta_2 = \pi, p_0^o = p_\pi^o = 1/2, \sigma_{m_1}^2 = M/4$ in Eq. (B1) 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 $\tilde S_1 \to \mathrm{ND} : 0, 1$. $\Rightarrow (T_1^+ - T_1^-) = \sqrt{M}\tilde S_1$. Now consider

$$h(r) = \sum_{i=1}^{r} (|T_{1i}'^+ - T_{1i}'^-| - |T_{1i}^+ - T_{1i}^-|)$$
$$= (\langle|\tilde S_1'|\rangle_r - \sqrt{M}\langle|\tilde S_1|\rangle_r)r, \qquad \text{(D1)}$$

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

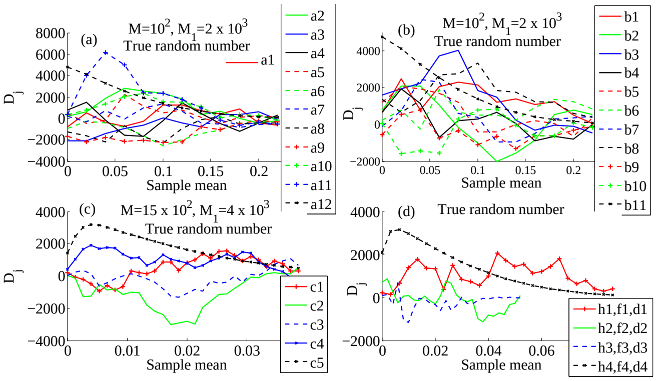

FIG. 5. (Color online) True random number generator [23] was interfaced with MATLAB. D_j is the difference in area under the Gaussians $\times 10^5$ i.e., $D_j = (\sum_{S_1'=-a_j}^{a_j} f(S_1') - \sum_{S_1=-a_j}^{a_j} g(S_1))\delta S \times 10^5$ where δS is the smallest element (step size) on x-axis (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 $\Delta S_1'^2 \approx (1 - (\Delta \cos\theta_q)^2_{p^o_{\theta_q}})/M = 0/M$. 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.

$r > r^o, r^o \sim 100$, we can neglect $\langle|\tilde{S}_1'|\rangle_r$ compared to $\sqrt{M}\langle|\tilde{S}_1|\rangle_r$ as $M \gg 1$. Hence $h(r) \approx -\sqrt{M}\langle|\tilde{S}_1|\rangle_r r$. $\langle|\tilde{S}_1|\rangle_r$ 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 $h(r) \approx -\sqrt{M}Cr$, 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 $h(r)$, $A(r) = \sum_{x=1}^{r} h(x)\delta x \approx -\sqrt{M}Cr(r+1)/2$, 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 (6, 7 (a,b), 8).

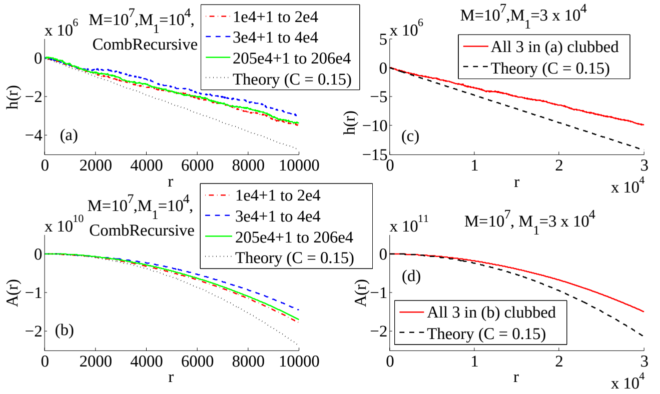

FIG. 6. (Color online) Number with the legend (where 'r'e'n' = r × 10$^{\mathrm{n}}$, 'r' is a real number and 'n' is an integer) is the PRN generator (CombRecursive)'s seed values. We used different seed for each of the M_1 number of sample mean points. Curves in (b) represents the area under the respective curves in (a). Similarly, the curves in (d) represents the area under the respective curves in (c). Consider (a): Seed corresponding to applying $(\theta_q)_x$s in the red curve (i.e., seed 10^4+1 to 2×10^4) was from 1 to 10^4, in the blue curve (i.e., seed $3\times10^4+1$ to 4×10^4) was from $2\times10^4+1$ to 3×10^4, and in the green curve (i.e., seed $205\times10^4+1$ to 206×10^4) was from $204\times10^4+1$ to 205×10^4.

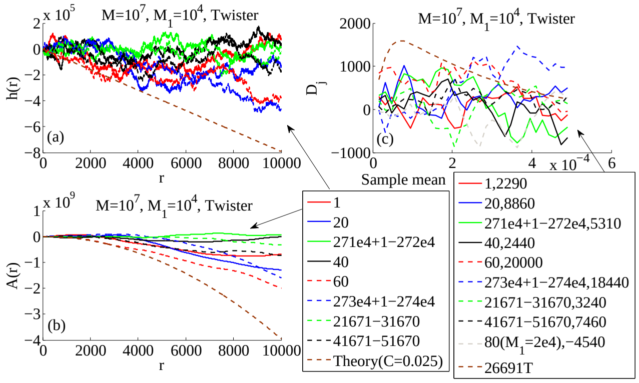

FIG. 7. (Color online) Twister is the PRN generator. (a) and (b) Values in the legend are the PRN generator's seed value(s). Only one seed value implies that we have used same seed for all M_1 number of sample mean points. A range (e.g., 271e4+1−272e4 i.e., 2710001 to 2720000) of seed values implies that 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. (5). In each legend (except the 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. Consider (a) and (c): Seed corresponding to applying $(\theta_q)_x$s in the curve represented by first legend (i.e., seed 1) was 1, second legend (i.e., seed 20) was 20, third legend (i.e., seed $271 \times 10^4 + 1$ to 272×10^4) was from $270 \times 10^4 + 1$ to 271×10^4, fourth legend (i.e., seed 40) was 40, fifth legend (i.e., seed 60) was 60, sixth legend (i.e., seed $273 \times 10^4 + 1$ to 274×10^4) was from $272 \times 10^4 + 1$ to 273×10^4, seventh legend (i.e., seed 21671 to 31670) was from 11671 to 21670, eighth legend (i.e., seed 41671 to 51670) was from 31671 to 41670, ninth legend of Fig. (c) (i.e., seed 80) was 80.

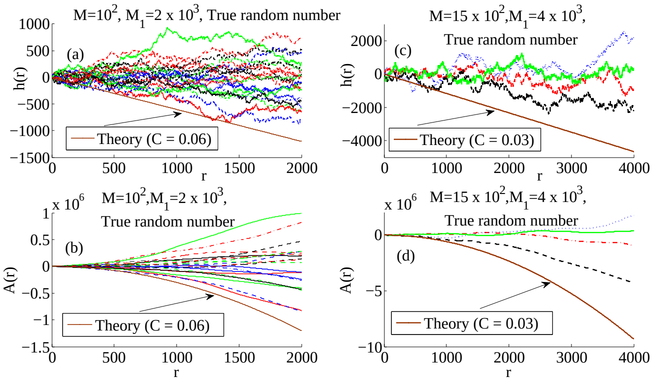

FIG. 8. (Color online) Simulation was done with true random numbers. Curves in (b) represents the area under the respective curves in (a). Similarly the curves in (d) represents the area under the respective curves in (c).