Opening up three quantum boxes causes classically undetectable wavefunction collapse

Richard E. George[a,1,2], Lucio M. Robledo[b], Owen J. E. Maroney[c], Machiel S. Blok[b], Hannes Bernien[b], Matthew L. Markham[d], Daniel J. Twitchen[d], John J. L. Morton[e,f], G. Andrew D. Briggs[a], and Ronald Hanson[b]

[a]Department of Materials, University of Oxford, Oxford OX1 3PH, United Kingdom; [b]Kavli Institute of Nanoscience Delft, Delft University of Technology, 2600 GA, Delft, The Netherlands; [c]Faculty of Philosophy, University of Oxford, Oxford OX1 4JJ, United Kingdom; [d]Element Six, Ltd., Ascot, Berkshire SL5 8BP, United Kingdom; [e]London Centre for Nanotechnology, University College London, London WC1H 0AH, United Kingdom; and [f]Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, United Kingdom

Edited by Yakir Aharonov, Tel Aviv University, Tel Aviv, Israel, and approved January 15, 2013 (received for review May 24, 2012)

One of the most striking features of quantum mechanics is the profound effect exerted by measurements alone. Sophisticated quantum control is now available in several experimental systems, exposing discrepancies between quantum and classical mechanics whenever measurement induces disturbance of the interrogated system. In practice, such discrepancies may frequently be explained as the back-action required by quantum mechanics adding quantum noise to a classical signal. Here, we implement the "three-box" quantum game [Aharonov Y, et al. (1991) *J Phys A Math Gen* 24(10):2315–2328] by using state-of-the-art control and measurement of the nitrogen vacancy center in diamond. In this protocol, the back-action of quantum measurements adds no detectable disturbance to the classical description of the game. Quantum and classical mechanics then make contradictory predictions for the same experimental procedure; however, classical observers are unable to invoke measurement-induced disturbance to explain the discrepancy. We quantify the residual disturbance of our measurements and obtain data that rule out any classical model by $\gtrsim$7.8 standard deviations, allowing us to exclude the property of macroscopic state definiteness from our system. Our experiment is then equivalent to the test of quantum noncontextuality [Kochen S, Specker E (1967) *J Math Mech* 17(1):59–87] that successfully addresses the measurement detectability loophole.

Leggett–Garg | quantum contextuality | quantum non-demolition measurement

Classical physics describes the nature of systems that are "large" enough to be considered as occupying one definite state in an available state space at any given time. Macrorealism (MR) applies whenever it is possible to perform nondisturbing measurements that identify this state without significantly modifying the system's subsequent behavior (1). MR allows the assignment of a definite history (or probabilities over histories) to classical systems of interest, but the MR condition can break down for systems "small" enough to be quantum mechanical during times "short" enough to be quantum coherent: times and distances that now exceed seconds (2) and millimeters (3) in the solid state. How can we tell whether a particular case is better described by quantum mechanics (QM) or MR? If there is a crossover between these, what does it represent?

One explanation for the breakdown of MR is that measurement back-action (either deliberate measurements by an experimenter or effective measurements from the environment) unavoidably change the state in the quantum limit, excluding MR due to a breakdown of nondisturbing measurability. This position is supported by "weak value" experiments (4, 5) that explore the transition from quantum to classical behavior as a measurement coupling is varied. Quantum behavior is found under weak coupling, whereas MR-compatible behavior is recovered when strong projective measurements effectively "impose" a classical value onto the measured quantum system (4).

We examine a case in which the back-actions of sequential "strong" projective measurements impose new quantum states that provide no detectable indication of disturbance to a "macrorealist" observer. We show that these states are still incompatible with MR, however, because no possible MR-compatible history can be assigned to the process as a whole. Our experiment can be described as a game played by two opponents (Alice and Bob) who take alternate turns to measure a shared system. The system they share may or may not obey the axioms of MR. For the purposes of the game, Bob assumes he may rely on the MR assumptions being true and only Alice is permitted to manipulate the system between measurements. If Bob is correct to assume MR holds, the game they play is constructed in his favor; however, "paradoxically," the exact same sequence of operations will define a game that favors Alice when a quantum-coherent description of the system is valid (6).

Experimentally, we use the ^{14}N nuclear spin of the nitrogen vacancy (^{14}NV$^-$) center ($S = 1$, $I = 1$) in diamond as Alice and Bob's shared system, enabling us to maintain near-perfect undetectability by Alice of Bob's observations. The experiment involves pre- and postselection (5, 7) on a three-level quantum system that is known to be equivalent to a Kochen–Specker test of quantum noncontextuality (8). Such tests are only possible in $d \geq 3$ Hilbert spaces (9); here, we use recent advances in the engineering (10) and control (11) of the NV$^-$ system that enable the multiple projective nondemolition measurements that are crucial to observing Alice's quantum advantage in the laboratory. We describe the game (12) and Bob's verification of it from the MR perspective, and we then discuss the experiment and results from the QM position. We quantify the incompatibility of our results with MR through use of a Leggett–Garg inequality (1) and discuss the implications of our result.

In the "three-box" quantum game (12), Alice and Bob each inspect a freshly prepared three-state system (classically, three separate boxes hiding one ball) using an apparatus that answers the question "Is the system now in state j?" ("Is the ball in box j?") for $j = 1, 2, 3$ by responding either "true" (1) or "false" (0). The question is answered by performing one of three mutually orthogonal measurements M_j. The game allows Bob a single use of either M_1 or M_2. Alice is allowed to use only M_3, and, additionally,

PHYSICS

Author contributions: R.E.G., L.M.R., O.J.E.M., J.J.L.M., G.A.D.B., and R.H. designed research; R.E.G., L.M.R., and M.S.B. performed research; R.E.G., L.M.R., O.J.E.M., M.S.B., H.B., M.L.M., D.J.T., and R.H. contributed new reagents/analytic tools; R.E.G. and L.M.R. analyzed data; and R.E.G., L.M.R., O.J.E.M., J.J.L.M., G.A.D.B., and R.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: ucanrge@live.ucl.ac.uk.

[2]Present address: London Centre for Nanotechnology, University College London, London WC1H 0AH, United Kingdom.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1208374110/-/DCSupplemental.

she is allowed to manipulate the system. Alice is allowed one turn (a manipulation either before or after an M_3 measurement) before Bob to prepare the system and one turn following him. Alice attempts to guess Bob's measurement result, and the pair bet on Alice correctly answering the question, "Did Bob find his M_j to be true?" Alice offers Bob $\geq$50% odds to predict when his M_j is true, although she may "pass" on any given round at no cost when she is undecided.

Bob realizes that if the M_j measurements are performed on a system following MR axioms, Alice must bet incorrectly $\geq$50% of the time, even if Alice could "cheat" by knowing which j-value will be presented (classically, knowing which box contains the ball); with three boxes and his free choice between M_1 and M_2, Alice is prevented from using her prior knowledge to win with a >50% success rate. Bob expects to win if the M_j measurements reproduce the behavior of a ball hidden in one of the three boxes. The conditions for this are (a) the M_j measurements are repeatable and mutually exclusive, such that $M_j \wedge M_k = \delta_{jk}$ (classically, the ball does not move when measured); (b) for any trial, $M_1 \vee M_2 \vee M_3 = 1$ (there is only one ball, and it is definitely in one of the boxes); (c) Bob has an equal probability of finding each j-value when measuring a fresh state, with $P_{M_j}(B) = 1/3 \; j \in 1, 2, 3$ (the ball is placed at random); and (d) Alice has no additional means to determine which, if any, M_j measurement Bob has chosen to perform. The conditions a–d serve to prevent Alice from learning Bob's M_j result in any macroreal system. Before accepting Alice's invitation to play, Bob verifies that properties a–d hold experimentally by carrying out M_j measurements. During verification, the game rules are relaxed and Bob is permitted to make pairs of sequential measurements, checking $M_j \wedge M_k = \delta_{jk}$. He is also allowed to measure every M_j, including M_3, which will be reserved for Alice once betting commences, or he may opt to perform no measurement at all and monitor Alice's response to determine if she can detect a disturbance caused by his measurement (*SI Text*).

When Bob is satisfied that a–d hold, the game appears fair from his macrorealist standpoint. Bob accepts Alice's wager, and play commences with Alice preparing a state, which Bob measures using either M_1 or M_2, while keeping his j-choice and M_j result secret. Alice manipulates the system, uses her M_3 measurement, and bets whenever her M_3 result is true. Believing that Alice could only guess his secret result, Bob accepts Alice's wager. Doing so, he finds that Alice's probability of obtaining a true M_3 result is $P_{M_3}(A) \simeq 1/9$, independent of his j-choice between M_1, M_2, or no measurement. Under MR, Bob could account for this only through Alice using a nondeterministic manipulation that would reduce the information available to her from the M_3 result. To Bob's surprise, when Alice plays, her true M_3 results coincide with the rounds on which Bob's M_j-result was also true. She passes whenever Bob's M_j result was false. In a perfect experiment, she would win every round she chose to play; in our practical realization, she achieves significantly more than the 50% success rate that would be predicted by MR. To understand Alice's advantage, we must examine the game from a QM perspective.

Alice uses the initial M_3 measurement to obtain the pure quantum state $|3\rangle$, passing on all rounds in which her initial M_3 measurement is false. She applies the unitary $\hat{U}_{\mathrm{I}}$, which operates as $\hat{U}_{\mathrm{I}} = |I\rangle\langle 3| + (\textit{orthogonal terms})$, to produce the initial state:

$$|I\rangle = \frac{|1\rangle + |2\rangle + |3\rangle}{\sqrt{3}} \qquad [1]$$

Her first turn presents the state $|I\rangle$ to Bob, who next measures M_j on $|I\rangle$, performing a projection. If Bob's M_j result is true, he has applied the quantum projector $\hat{P}_j = |j\rangle\langle j|$, and by finding an M_j result that is false, he has applied $\hat{P}_j^{\perp} = 1 - |j\rangle\langle j|$. Alice uses her final turn to measure the component of the state left by Bob's measurement along the state $|F\rangle = (|1\rangle + |2\rangle - |3\rangle)/\sqrt{3}$. Bob's projectors on Alice's initial and final states $|I\rangle$ and $|F\rangle$ obey:

$$\left|\langle F|\hat{P}_1|I\rangle\right|^2 = \left|\langle F|\hat{P}_2|I\rangle\right|^2 = 1/9 \qquad [2]$$

$$\left|\langle F|\hat{P}_1^{\perp}|I\rangle\right|^2 = \left|\langle F|\hat{P}_2^{\perp}|I\rangle\right|^2 = 0 \qquad [3]$$

for both $j = 1$ and $j = 2$. Alice cannot directly measure $|F\rangle$ but is able to transform state $|F\rangle$ into state $|3\rangle$ with a unitary $\hat{U}_{\mathrm{F}} = |3\rangle\langle F| + (\textit{orthogonal terms})$, and she uses her measurement of M_3 as an effective M_F measurement. Alice therefore obtains M_3-true when Bob's M_j result is true with probability $P_{M_3}(A \cap B) = |\langle 3|\hat{U}_{\mathrm{F}}\hat{P}_j\hat{U}_{\mathrm{I}}|3\rangle|^2 = |\langle F|\hat{P}_j|I\rangle|^2 = 1/9$ and when Bob's M_j result is false with probability $P_{M_3}(A \cap \neg B) = |\langle 3|\hat{U}_{\mathrm{F}}\hat{P}_j^{\perp}\hat{U}_{\mathrm{I}}|3\rangle|^2 = |\langle F|\hat{P}_j^{\perp}|I\rangle|^2 = 0$. Alice finds that her M_3 result being true is conditional on Bob leaving a component of $|\psi_j\rangle$ along $|F\rangle$; to do so, his M_j result cannot have been false. Alice's probability conditioned on Bob is then $P_{M_j}(B|A) = 1$. Alice bets whenever her M_3 result is true, playing one-ninth of the rounds and winning each round she plays.

Materials and Methods

Our implementation of this game uses the NV$^-$ center, which hosts an excellent three-level quantum system for the three-box game: the ^{14}N nucleus, which has $(2I + 1) = 3$ quantum states (Fig. 1*A*). Although we cannot (yet)

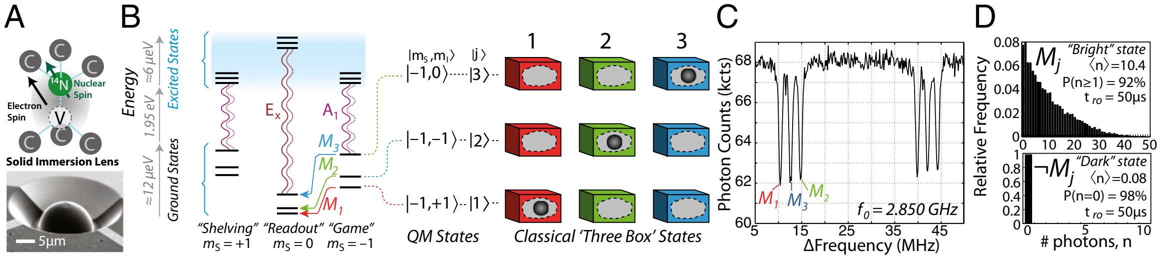


Fig. 1. Three-box game is implemented using the ^{14}N nuclear spin of the NV$^-$ center in diamond, measured using the electron spin. (*A*) Schematic of the NV$^-$ defect in diamond and representative diamond lens used in the measurements. (*B*) Magnetic moment of the electron spin is quantized into one of three values: $m_S = -1$, 0, or +1. These states split into a further three ($m_I = -1$, 0, or +1) according to the magnetic moment of the ^{14}N nuclear spin. The $m_S = \pm 1$ states fluoresce via the A_1 transition, whereas $m_S = 0$ fluoresces via the E_x transition. We use the $m_S = -1$ manifold to hold the three states in the game, conditionally moving the state between $m_S = -1$ and $m_S = 0$ dependent on the nuclear spin sublevel m_I. These three m_I states are taken to correspond to the configurations of a hidden ball. (*C*) We identify the allowed microwave transitions ($\Delta m_S = 1$, $\Delta m_I = 0$) that provide the M_j readouts. (*D*) Photon counting statistics, in each case from 10,000 trials, observed during a typical projective readout indicate the presence (*Upper*) or absence (*Lower*) of optical fluorescence, corresponding to outcomes M_j and $\neg M_j$, respectively.

superpose a physical ball under three separate boxes, real-space separation is not essential to the three-box argument. Alice and Bob can bet on any physical property of a system for which MR assigns mutually exclusive outcomes; for instance, a classical gyroscope revolving about one of three possible axes is not simultaneously revolving about the second and third axes. By using rf pulses (13), we can readily prepare the ^{14}N angular momentum into a superposition of alignment along three distinct spatial axes, providing three "box states" that are presumed to be mutually exclusive in the macrorealist picture. We work in the electron spin $m_S = -1$ manifold and assign eigenvalues of nitrogen nuclear spin m_I to the box-states j according to (*a*) $|m_I = -1\rangle \sim |j = 1\rangle$, (*b*) $|m_I = +1\rangle \sim |j = 2\rangle$, and (*c*) $|m_I = 0\rangle \sim |j = 3\rangle$ (Fig. 1*B*).

Preparation and readout of the ^{14}N nuclear spin is provided via the NV$^-$ electronic spin ($S = 1$). We use selective microwave pulses to change m_S conditioned on m_I, reading out the electron spin in a single shot and with high fidelity (11), by exploiting the electron spin-selective optical transitions of the NV$^-$ center. The spin readout achieves 96% fidelity and takes ≈20 μs, which is much shorter than the nuclear spin inhomogeneous coherence lifetime of $T_2^* \gg 1$ ms at $T = 8.7$ K, enabling three sequential readout operations during a single coherent evolution of the system, as required for our three-box implementation. We achieve all steps of the quantum experiment well within the coherence time of our system, and therefore make no use of refocusing rf pulses.

The full experimental sequence is shown in Fig. 2, with further details provided in *SI Text*. The initial state $|3\rangle$ is prepared by projective nuclear spin readout using a short-duration (≃200 ns) optical excitation. The subsequent experiment is then conditioned on detection of at least one photon during the preparation phase, which heralds $|3\rangle$ with ≳95% fidelity (Fig. 1*D*) at the expense of ≲1% preparation success rate. Once $|3\rangle$ is heralded, all subsequent data are accepted unconditionally. After initialization, Alice transforms the state $|3\rangle$ into $|I\rangle$ via two rf pulses (*SI Text*) and hands the system to Bob, who measures M_1 or M_2. A further four rf pulses transform $|F\rangle$ to $|3\rangle$, and Alice performs her final M_3 measurement while statistics about Alice and Bob's relative successes are recorded.

We quantify the discrepancy between MR and QM by constructing a Leggett-Garg function for our system, defined as

$$\langle K\rangle = \langle Q_1 Q_2\rangle + \langle Q_2 Q_3\rangle + \langle Q_1 Q_3\rangle \qquad [4]$$

where Q_j are observables of our system with values ±1, recorded at three different times, derived from Alice and Bob's measurements (1). We assign $Q_j = +1$ whenever an M_3 result is true (or could be inferred true in the MR picture) and assign $Q_j = -1$ otherwise. The initially heralded state $|3\rangle$ fixes the value of $Q_1 = +1$ always, and values for Q_2 and Q_3 are taken directly from Bob and Alice's measurement results. The Leggett–Garg function is known to satisfy $-1 \le \langle K\rangle \le +3$ for all MR systems (1), and for the present system, we can show that $\langle K\rangle$ is related to Bob and Alice's statistics (*SI Text*) as follows:

$$\langle K\rangle = \frac{4}{9}\left(1 - P_{M_1}(B|A) - P_{M_2}(B|A)\right) - 1 \qquad [5]$$

where $P_{M_j}(B|A)$ is the probability that Bob finds the M_j result true, given that Alice has also found her final M_3 result true. MR asserts that M_1 and M_2 are mutually exclusive events, whereas QM does not, such that:

$$\text{MR}: \; P_{M_1}(B|A) + P_{M_2}(B|A) \le 1 \qquad [6]$$

$$\text{QM}: \; P_{M_1}(B|A) + P_{M_2}(B|A) \le 2 \qquad [7]$$

Under QM assumptions, Eq. **5** satisfies $\langle K\rangle \ge -13/9 = -1.4\dot{4}$, possibly lying outside the range compatible with MR.

Results

Bob picks a secret j-value and maps the corresponding nuclear spin projection to the electron spin by applying a microwave π-pulse to drive a transition from one of the $m_S = -1$ states ($|j\rangle$ is $|1\rangle$ or $|2\rangle$) into the $m_S = 0$ manifold. He then uses optical measurement of the E_x fluorescence to determine m_S. Absence of fluorescence ("E_x-dark" NV$^-$) implies $\neg M_j$ and collapses the electron state into $m_S = -1$ while performing $\hat{P}_j^{\perp}$ on the nuclear spin (Fig. 3*A*, *ii*). We find that nuclear spin coherences within $m_S = -1$ are unaffected by the $\neg M_j$ readout process.

Detection of $n \ge 1$ photons during Bob's 20-μs readout projects the electron into $m_S = 0$ and corresponds to an M_j result that is true. In such events, there is an ≃70% chance the electronic spin will be left in an incoherent mixture of $m_S = \pm 1$ following readout, due to optical pumping (11). Conditional on Bob's M_j result being true, we take care to undo the mixing effect as follows. We first pump the electron spin to $m_S = 0$ by selective optical excitation of $m_S = \pm 1$ (via a laser resonant with the A_1 transition), followed by driving a selective a microwave pulse from $m_S = 0$ to $m_S = -1$ (Fig. 1*C*). This procedure is effective because the optical fluorescence preserves the nuclear spin populations m_I that encode the game eigenstates in ≳70% of cases (Fig. 3*B*). Bob performs repeated pairs of measurements, verifying from a macrorealist's perspective that performing M_j is equivalent to opening one of the three boxes containing a hidden ball. Bob finds the probability for each M_j is ≃ ⅓ (Fig. 3*A*, *i*). Bob performs consecutive M_j observations and verifies that finding M_j ($\neg M_j$) true on one run implies that the subsequent measurement of M_j ($\neg M_j$) will also be true (Fig. 3 *B* and *C*), gathering statistics over n = 1,200 trials for each combination.

Once Bob has measured in secret, Alice predicts his result by mapping $|F\rangle$ to $|3\rangle$ and performing M_3. Alice accomplishes this via: $|F\rangle \to |I\rangle \to |3\rangle$. The Berry's phase associated with 2π rotations (14) provides the map $|F\rangle \to |I\rangle$ via two rf pulses that change the signs of the $\{|1\rangle, |3\rangle\}$ and then $\{|2\rangle, |3\rangle\}$ states. State $|3\rangle$ then acquires two sign changes yielding $|F\rangle$ up to a global phase. The map $\hat{U}_I^{-1}$ from $|I\rangle$ to $|3\rangle$ is then achieved by inverting the order and phase of Alice's initial $\hat{U}_I$ pulses (*SI Text*).

Alice and Bob compare their measurement results during n = 2 * 1,200 rounds of play, distributed evenly across Bob's two choices of M_j measurement, as well as during a further 1,200 rounds in which Bob performs no measurement whatsoever.

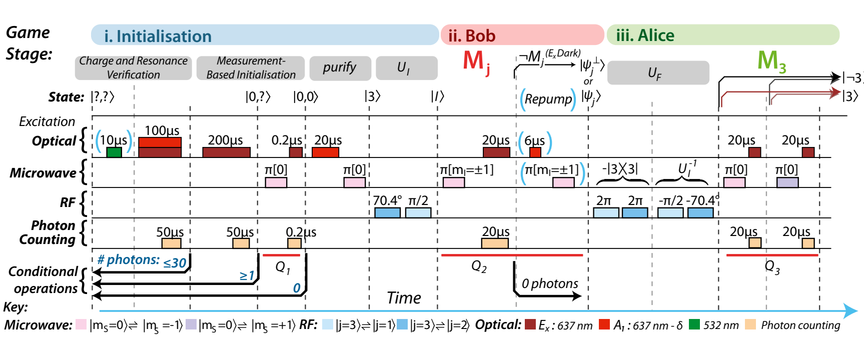


Fig. 2. Microwave, rf (RF), and optical pulse sequence implementation in the three-box experiment. (*i*) Initialization consists of preparing the NV$^-$ state via charge-state verification and measurement-based initialization into state $|3\rangle$, followed by purification of $m_S = 0$ and application of $\hat{U}_I$. (*ii*) Bob's measurement M_j consists of moving the population from $m_S = -1$ to $m_S = 0$ conditioned on m_I, indicated as $\pi[m_I]$ in the figure, followed by monitoring of E_x fluorescence. If fluorescence is observed, a "repopulating" sequence via the spectrally resolved A_1 fluorescence ($\lambda_{A_1} - \lambda_{E_x} = \delta = 1.89$ pm) resets $m_S = -1$ while leaving m_I unchanged and ready for Alice's measurement. (*iii*) Alice's measurement consists of the unitary $\hat{U}_F$, followed by readout of M_3 in the $m_S = -1$ and $m_S = +1$ sublevels. Further details on the experimental sequence are provided in *SI Text*.

Alice finds her final M_3 result is true in $\simeq 15\%$ of cases, independent of Bob's choice of measurement context between M_1 and M_2 or neither measurement (Fig. 4*A*). Among those $\simeq 15\%$ of cases in which Alice's M_3 result is true and she chooses to bet, Bob finds she wins $\gtrsim 67\%$ of such rounds for either of Bob's choices between measuring M_1 and M_2 (Fig. 4*B*), confounding the macrorealist expectation. The principle source of error in our experiment arises from imperfect control of the nuclear spin (*SI Text*).

We quantify the Leggett–Garg inequality violation in our experiment by determining $\langle K \rangle$ from estimates of $P_{M_j}(B|A)$, finding $\langle K \rangle = -1.265 \pm 023$, corresponding to a $\simeq 11.3$ σ-violation of the Leggett–Garg inequality under fair sampling assumptions and to a $\simeq 7.8$ σ-violation in a "maximally adverse" macrorealist position in which all undetermined measurements are assumed to represent Alice "cheating" and are reassigned to minimize the discrepancy between QM and MR predictions (*SI Text*).

Discussion

Our results unite two concepts in foundational physics: Leggett–Garg inequalities (1) and pre- and postselected effects (7) in a quantum system to which the Kochen–Specker no-go theorem applies (9). Previous experimental studies of the Leggett–Garg inequality have used ensembles (15, 16), have made assumptions regarding process stationarity (17, 18), or have required weak measurements (4) to draw conclusions, whereas the existing studies of the three-box problem cannot incorporate measurement nondetectability (19, 20), presenting a loophole that allows classical noncontextual models to reproduce the quantum statistics (8). We have studied the three-box experiment on a matter system, as originally conceived (12) and developed (6) in terms of sequential, projective nondemolition measurements, and we therefore reexamine the conclusions that can be drawn when using this improved measurement capability.

Two assumptions underpin MR: (*i*) macroscopic state definiteness and (*ii*) nondisturbing measurability. In previous studies, it has been possible to assign violations of the Leggett–Garg inequality to a loss of nondisturbing measurability in both optical (4) and spin-based (16) experiments. The disturbance due to measurement can sometimes be surprisingly nonlocal (21), and it has been suggested that detectable disturbance is a necessary condition for violating a Leggett–Garg inequality in all cases (22, 23). We improve this result, clarifying that detectable disturbance is a necessary condition for violating the Leggett–Garg inequality in two-level quantum systems but is not required in the three-level system studied here (*SI Text*).

We show from the statistics of the measurement outcomes that Alice cannot detect Bob's choice to measure or not (Fig. 4*A*); thus, our measurements involve no detectable disturbance, whereas the statistics from the three-box game violate a Leggett–Garg inequality. We are therefore able to rule out the macrorealist's assumption *i* of state definiteness, a result unobtainable from previous studies of two-level quantum systems.

Our experiment makes use of a three-level quantum system in which Bob's choice between M_1 and M_2 represents a choice of measurement "context" in the language of Kochen and Specker (9). If Bob is able to keep his measurement context secret, a macrorealist Alice could only use a "noncontextual" classical theory to describe the experiment. It is known that every pre- and postselection paradox implies a Kochen–Specker proof of quantum contextuality (8). It has been argued that measurement disturbance provides a loophole to admit noncontextuality into classical models [in addition to finite measurement precision (24, 25)]; all classical models presented to date that exploit this loophole give rise to detectable measurement disturbances. In

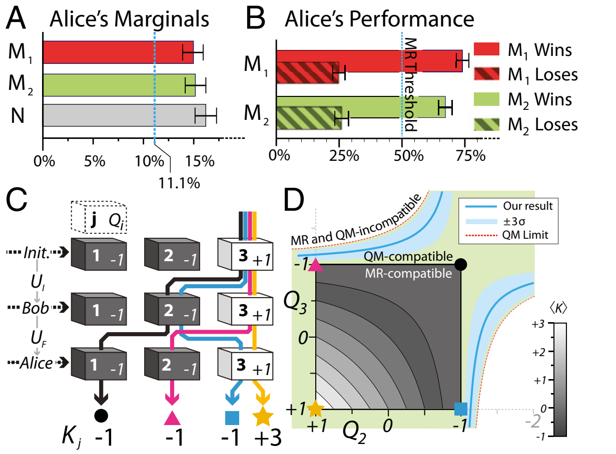


Fig. 4. Violation of a Leggett–Garg inequality in the three-box game. (*A*) Alice's measurement M_3 is independent of Bob's choice to perform measurement M_1, M_2, or neither (*N*). (*B*) Observations of Bob and Alice are correlated to indicate the probability that Bob has (or has not) seen state M_j, given that Alice has seen M_3, determining who "wins" the game. Alice's probability of winning exceeds 50% for both of Bob's choices M_1 and M_2. (*C*) Four MR-compatible histories that extremize $\langle K \rangle$ are illustrated by four trajectories passing through different boxes during the game. A trajectory entirely within the white $j = 3$ boxes has $Q_{\{1,\,2,\,3\}} = +1$ and yields $\langle K \rangle = +3$. Histories that visit other boxes yield $\langle K \rangle = -1$. (*D*) $\langle K \rangle$ values of the four paths are shown in the corners of the (Q_2, Q_3) graph. Values for $\langle Q_2 \rangle$ and $\langle Q_3 \rangle$ from MR-compatible experiments must lie inside the shaded square satisfying $-1 \le \langle K \rangle \le 3$. Our measurements lie on the cyan curve outside this region but within the region allowed by QM.

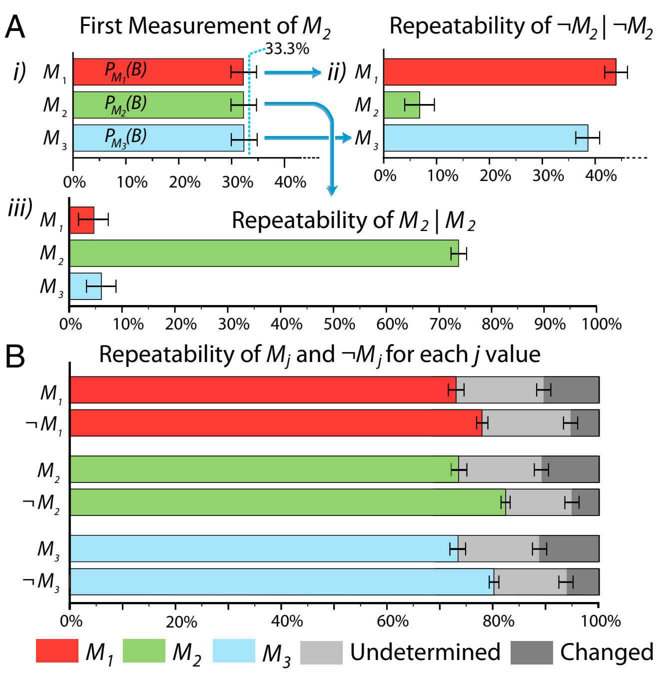


Fig. 3. Bob verifies the nondetectable nature of the M_j measurements. (*A*) Measurement within the $m_S = -1$ manifold only. (*i*) Bob's measurement results when observing the state $|I\rangle$ in the $|j\rangle$ basis are independent of the j-value selected to within experimental error. Repeatability is illustrated by plotting the result of a second M_j measurement within $m_S = -1$, conditioned on (*ii*) the result $\neg M_2$, or (*iii*) the result M_2. (*B*) Repeatability of each M_j measurement is studied within the $m_S = -1$ manifold; a finite probability exists for the electron spin to branch into the $m_S = +1$ manifold, yielding an undetermined reading, and for the nuclear spin to flip producing a "definitely changes" outcome. Error bars show ± 2 σ/95% confidence intervals.

our experiment, Bob's intervening measurement introduces no disturbances detectable by Alice and cannot be accounted for by existing classical models.

ACKNOWLEDGMENTS. R.E.G., O.J.E.M., and G.A.D.B. thank the John Templeton Foundation for supporting this work. This work is also supported by the Defense Advanced Research Planning Agency Quantum Entanglement Science and Technology (QuEST) and Quantum Assisted Sensing and Readout (QuASAR) programs, the Dutch Organization for Fundamental Research on Matter, the Netherlands Organization for Scientific Research, and the European Commission Seventh Framework Programme on Diamond based atomic nanotechnologies (DIAMANT). J.J.L.M. is supported by the Royal Society.

1. Leggett AJ, Garg A (1985) Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Phys Rev Lett* 54(9):857–860.
2. Tyryshkin AM, et al. (2012) Electron spin coherence exceeding seconds in high-purity silicon. *Nat Mater* 11(2):143–147.
3. Dicarlo L, et al. (2010) Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature* 467(7315):574–578.
4. Goggin ME, et al. (2011) Violation of the Leggett-Garg inequality with weak measurements of photons. *Proc Natl Acad Sci USA* 108(4):1256–1261.
5. Massar S, Popescu S (2011) Estimating preselected and postselected ensembles. *Phys Rev A* 84(5):052106.
6. Aharon N, Vaidman L (2008) Quantum advantages in classically defined tasks. *Phys Rev A* 77(5):052310.
7. Aharonov Y, Bergmann PG, Lebowitz JL (1964) Time symmetry in quantum process of measurement. *Phys Rev B* 134(6B):1410–1416.
8. Leifer M, Spekkens R (2005) Logical pre- and post-selection paradoxes, measurement-disturbance and contextuality. *Int J Theor Phys* 44(11):1977–1987.
9. Kochen S, Specker E (1967) The problem of hidden variables in quantum mechanics. *J Math Mech* 17(1):59–87.
10. Bernien H, et al. (2012) Two-photon quantum interference from separate nitrogen vacancy centers in diamond. *Phys Rev Lett* 108(4):043604.
11. Robledo L, et al. (2011) High-fidelity projective read-out of a solid-state spin quantum register. *Nature* 477(7366):574–578.
12. Aharonov Y, Vaidman L (1991) Complete description of a quantum system at a given time. *J Phys A Math Gen* 24(10):2315–2328.
13. Schweiger A, Jeschke G (2001) *Principles of Pulse Electron Paramagnetic Resonance* (Oxford Univ Press, Oxford).
14. Morton J, et al. (2006) Bang-bang control of fullerene qubits using ultrafast phase gates. *Nat Phys* 2(1):40–43.
15. Moussa O, Ryan CA, Cory DG, Laflamme R (2010) Testing contextuality on quantum ensembles with one clean qubit. *Phys Rev Lett* 104(16):160501.
16. Knee GC, et al. (2012) Violation of a Leggett-Garg inequality with ideal non-invasive measurements. *Nat Commun* 3:606.
17. Palacios-Laloy A, et al. (2010) Experimental violation of a Bell's inequality in time with weak measurement. *Nat Phys* 6(6):442–447.
18. Waldherr G, Neumann P, Huelga SF, Jelezko F, Wrachtrup J (2011) Violation of a temporal bell inequality for single spins in a diamond defect center. *Phys Rev Lett* 107(9):090401.
19. Resch KJ, Lundeen JS, Steinberg AM (2004) Experimental realization of the quantum box problem. *Phys Lett A* 324(2–3):125–131.
20. Kolenderski P, et al. (2012) Aharon-Vaidman quantum game with a Young type photonic qutrit. *Phys Rev A* 86(1):012321.
21. Kwait P, Weinfurter H, Zeilinger A (1996) Quantum seeing in the dark. *Sci Am* 375(11):72–78.
22. Benatti F, Ghirardi G, Grassi R (1994) On some recent proposals for testing macro-realism versus quantum-mechanics. *Foundations of Physics Letters* 7(1):105–126.
23. Lahti P, Mittelstaedt P, eds (1990) *Symposium on the Foundations of Modern Physics 1990* (World Scientific, Singapore).
24. Meyer D (1999) Finite precision measurement nullifies the Kochen-Specker theorem. *Phys Rev Lett* 83(19):3751–3754.
25. Clifton R, Kent A (2001) Simulating quantum mechanics by non-contextual hidden variables. *Proc R Soc Lond A Math Phys Sci* 456(1):2101–2114.

Supporting Information

George et al. 10.1073/pnas.1208374110

In this document, we detail the projections performed by Bob's measurements and provide a detailed account of the states prepared during the experiment. We detail how Alice is able to transform her initial state $|3\rangle$ into the state $|I\rangle$, and subsequently transform $|F\rangle$ back into state $|3\rangle$. We describe our notation for probabilities that allow us to describe the "three-box" game from both a classical and quantum perspective. We derive the Leggett–Garg function (1) for this system as calculated by an observer who assumes the system obeys the axioms of macrorealism (MR), namely, state definiteness and noninvasive measurability. We describe the sample fabrication and measurement setup and discuss the practicalities of the experimental measurements involving reading out the nuclear spin, and we discuss the significance of finite measurement precision and measurement errors.

I. Experimental Details

A. Projections Performed by Bob's Measurements. In the main text, we state that Bob finding a measurement result M_j-true prepares the state $|j\rangle$ by performing projector $\hat{P}_j$ on state $|I\rangle$, whereas Bob finding the result M_j-false prepares an orthogonal state $|\psi'_j\rangle$ by performing projector $\hat{P}_j^{\perp}$. Here, we give explicit vector representations of these states, and matrix representations of the projectors $\hat{P}_j$ and $\hat{P}_j^{\perp}$, to aid understanding. We can write a column vector to represent the general state $|\psi\rangle$ of the three-box problem as:

$$|\psi\rangle = a|1\rangle + b|2\rangle + c|3\rangle = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \textbf{[S1]}$$

The initial and final states used by Alice are then written as:

$$|I\rangle = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \langle F| = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 1 & -1 \end{pmatrix} \quad \textbf{[S2]}$$

We write the identity matrix as:

$$1 = \sum_j |j\rangle\langle j| = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \textbf{[S3]}$$

and we write the projectors $\hat{P}_j$ and $\hat{P}_j^{\perp}$ explicitly as:

$$\hat{P}_1 = |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\hat{P}_1^{\perp} = 1 - \hat{P}_1 = |2\rangle\langle 2| + |3\rangle\langle 3| = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \textbf{[S4]}$$

$$\hat{P}_2 = |2\rangle\langle 2| = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\hat{P}_2^{\perp} = 1 - \hat{P}_2 = |1\rangle\langle 1| + |3\rangle\langle 3| = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \textbf{[S5]}$$

Using this representation, it is straightforward to verify the claims in the main text, that:

$$P_{M_1}(A \cap B) = |\langle F|\hat{P}_1|I\rangle|^2 = P_{M_2}(A \cap B) = |\langle F|\hat{P}_2|I\rangle|^2 = 1/9 \quad \textbf{[S6]}$$

$$P_{M_1}(A \cap \neg B) = |\langle F|\hat{P}_1^{\perp}|I\rangle|^2 = P_{M_2}(A \cap \neg B) = |\langle F|\hat{P}_2^{\perp}|I\rangle|^2 = 0 \quad \textbf{[S7]}$$

These expressions describe Alice's ability to win $\gg$50% of rounds in the quantum version of the game.

B. Alice's Unitary Operations. ***1. Preparing the initial state.*** Alice would like to measure $|I\rangle$ and $|F\rangle$ but only has access to M_3. She performs effective M_I and M_F measurements by performing unitaries that map $|I\rangle \rightarrow |3\rangle$ and $|F\rangle \rightarrow |3\rangle$, followed by M_3 measurement. We define the unitary operation applied by Alice to transform between the states $|3\rangle$ and $|I\rangle$ in terms of its ability to split a population initially prepared in level $|3\rangle$ into an equal superposition of the states $|1\rangle$, $|2\rangle$, and $|3\rangle$. We construct $\hat{U}_I$ by concatenating two unitaries that can be implemented as rf pulses. The first step in performing $\hat{U}_I$ represents a rotation through angle θ in the $\{|3\rangle, |2\rangle\}$ plane, and the second step represents a rotation through angle $90° = \pi/2$ in the $\{|3\rangle, |1\rangle\}$ plane.

The first rotation (θ in the $\{|3\rangle, |1\rangle\}$ plane) must transfer one-third of the population from state $|3\rangle$ to state $|1\rangle$, leaving two-thirds of the population in state $|3\rangle$. The subsequent rotation must split the population in level $|3\rangle$ equally between $|3\rangle$ and $|2\rangle$, producing an equal population of one-third in each of the three $|j\rangle$ states.

Considering the coherent rotation in the $\{|3\rangle, |1\rangle\}$ plane, a rotation through θ transfers a fraction $\sin^2(\theta/2)$ into state $|1\rangle$, while leaving a fraction $\cos^2(\theta/2)$ in state $|3\rangle$; thus, to place one-third of the population in state $|1\rangle$, we have:

$$\sin^2(\theta/2) = 1/3 \qquad \cos^2(\theta/2) = 2/3 \quad \textbf{[S8]}$$

implying that:

$$\sin(\theta/2) = \sqrt{1/3} \qquad \cos(\theta/2) = \sqrt{2/3} \quad \textbf{[S9]}$$

$$\tan(\theta/2) = \sqrt{1/2} \qquad \theta = 2\tan^{-1}\left(\sqrt{1/2}\right) \quad \textbf{[S10]}$$

with the result that:

$$\theta = 70.6°(= 1.23 \text{ radians}) \quad \textbf{[S11]}$$

Alice prepares state $|3\rangle$ and performs $\hat{U}_I$ as two rotations: $\theta = 70.6°$ in the $\{|3\rangle, |1\rangle\}$ plane and $\pi/2 = 90°$ in the $\{|3\rangle, |2\rangle\}$ plane.

2. Alice's measurement of $|F\rangle$. The states $|I\rangle$ and $|F\rangle$ are defined in Vaidman's paper (2) as:

$$|I\rangle = \frac{|1\rangle + |2\rangle + |3\rangle}{\sqrt{3}} \qquad |F\rangle = \frac{|1\rangle + |2\rangle - |3\rangle}{\sqrt{3}} \quad \textbf{[S12]}$$

Because quantum states are defined only up to an overall multiplicative scalar (states such as $|F\rangle$ are rays in the Hilbert space), we can choose to write:

$$|F\rangle = \frac{-|1\rangle - |2\rangle + |3\rangle}{\sqrt{3}} \quad \textbf{[S13]}$$

A rotation through 2π radians introduces a sign change, such that there are two combined rotations through 2π, first on the $\{|3\rangle, |1\rangle\}$ level and then on the $\{|3\rangle, |2\rangle\}$ level. We have:

$$\begin{pmatrix} |1'\rangle \\ |3'\rangle \\ |2'\rangle \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\hat{U}_{IF}} \begin{pmatrix} |1\rangle \\ |3\rangle \\ |2\rangle \end{pmatrix} \quad \textbf{[S14]}$$

The two rotations, each through 2π, have the combined effect of flipping the signs of the states $|1\rangle$ and $|2\rangle$ relative to state $|3\rangle$; specifically, we have:

$$|F\rangle = \hat{U}_{IF}|I\rangle \quad \textbf{[S15]}$$

Therefore, by applying these two rotations, Alice can map $|F\rangle \to |I\rangle \to |3\rangle$ and measure M_3 as per the main text.

C. Experimental Implementation of Nuclear Spin Readout. ***1. Sample.*** We use a naturally occurring nitrogen vacancy (NV^-) center in high-purity (spin-bearing impurities controlled below 1 part per billion) type IIa diamond grown by chemical vapor deposition, with a $\langle 111\rangle$ crystal orientation obtained by cleaving a $\langle 100\rangle$ substrate. We optimize the photon collection efficiency through use of a solid immersion lens deterministically fabricated by focused ion beam milling (3) to focus light onto the selected NV^- center. Microwave and rf pulses for the spin manipulation are applied through a lithographically defined strip-line adjacent to the solid immersion lens (3).

2. Measurement setup. We use a home-built, low-temperature confocal microscope that has been described in detail by Robledo et al. (4). All experiments are performed at a sample temperature of T = 8.7 K. A small magnetic field ($B \approx 5$ G, oriented along the NV^- symmetry axis) is applied by means of four permanent magnets arranged around the cryostat.

3. General. In the course of this experiment, we use different variations of single-shot nuclear spin readout, adapted to our specific purpose. In general, nuclear spin readout is implemented according to the following protocol (4):

i) Optional: Electron spin initialization by optical pumping into $m_S = 0$ (excitation of A_1 transition) or $m_S = \pm 1$ (excitation of E_x transition)
ii) Map nuclear spin onto electron spin: Selective microwave (MW) excitation of the hyperfine transition representing the state to be probed (in general, effecting a π rotation)
iii) Readout of the electron spin: Resonant optical excitation on E_x transition (for maximum contrast, $t_{ro} \approx$ 15– 25 μs)
iv) Optional: Restore the electron spin state by optical pumping (for deterministic preparation of $m_S = +1$ or $m_S = -1$; optical pumping into $m_S = 0$, followed by a MW π-pulse)

If readout of the electron spin yields a result different from its initial state, we conclude that the nuclear spin occupies the probed state. The readout can be repeated using different MW frequencies, allowing us to perform population tomography on the full electron-nuclear spin state. We now outline the readout variations used.

4. Nuclear spin initialization. Initialization of the ^{14}N nuclear spin into $m_I = 0$ represents the first measurement of the Leggett–Garg test Q_1. This first measurement is probabilistic; we choose parameters that maximize the preparation fidelity with respect to the postmeasurement state, accepting a reduced preparation success probability:

The electron spin is initialized in the $m_S = \pm 1$ manifold by optical pumping, implemented by a pulse of 200 μs in duration, resonant with the E_x transition (fidelity F = 99.4%). The initialization fidelity is further increased to F > 99.9% by postselecting only experimental runs in which no photon is detected during the last 50 μs of the optical pumping pulse (avoiding accidental repopulation of $m_S = 0$).

We then apply a MW π-pulse resonant with the transition $|m_S = -1, m_I = 0\rangle \to |m_S = 0, m_I = 0\rangle$ with a state selectivity of ≈98%, limited by the proximity of other hyperfine transitions.

We probe successful initialization into $m_I = 0$ by requiring >0 detected photons during E_x excitation. To maximize fidelity, we keep the readout duration short (200 ns).

During the electron spin readout, there is a finite chance of optically induced electron spin flips. To ensure that the electron occupies the $m_S = -1$ state, we first optically pump it into $m_S = 0$ and then apply a selective MW π-pulse resonant with $|m_S = 0, m_I = 0\rangle \to |m_S = -1, m_I = 0\rangle$.

After successful initialization, we estimate an overlap with $|m_S = -1, m_I = 0\rangle$ of >95%.

All runs of the three-box experiment use this initialization step.

5. Three-box game: Bob's readout. The second readout (Bob's readout) consists of a selective MW π-pulse, resonant with:

$$|m_S = -1, m_I = -1\rangle \to |m_S = 0, m_I = -1\rangle \quad (M_1) \quad \textbf{[S16]}$$

$$|m_S = -1, m_I = +1\rangle \to |m_S = 0, m_I = +1\rangle \quad (M_2) \quad \textbf{[S17]}$$

$$|m_S = -1, m_I = 0\rangle \to |m_S = 0, m_I = 0\rangle \quad (M_3) \quad \textbf{[S18]}$$

depending on Bob's choice of measurement. Subsequently, the electron spin state is probed by a t_{ro} = 20-μs pulse resonant with E_x. This readout gives a large contrast (F = 96%), but if $m_S = 0$ is detected, many excitation cycles may have occurred, and due to optically induced spin flips, the electron spin may be left in an undefined state. As a remedy, conditional on obtaining an $m_S = 0$ readout result, we restore the spin into $m_S = -1$ by optical pumping into $m_S = 0$, followed by a selective MW π-pulse, $|m_S = 0, m_I = +1(-1)\rangle \to |m_S = -1, m_I = +1(-1)\rangle$ ($M_{1(2)}$). This procedure ensures the electron is found deterministically in $m_S = -1$ after the readout, leaving nuclear spin coherence unaffected.

6. Three-box game: Alice's readout. Although for the last readout (Alice's readout), we could, in principle, apply the same protocol as in Bob's readout, we decided to read out all three nuclear spin states (box states) for each measurement, also allowing us to identify the few cases in which we do not find the ball in any of the boxes (e.g., to determine the finite detection efficiency of the nuclear spin readout).

For each probed nuclear spin state, we repeat two readout iterations consisting of a selective MW π-pulse and a 20-μs E_x readout pulse. This is repeated for the three hyperfine lines corresponding to the $m_S = -1$ manifold (implementing M_1, M_2, and M_3). The first probed state found to emit a photon is identified as the readout result; if no photon is detected, we consequently assign no result. To avoid a readout bias due to the order of the probed states, we permute the order between measurements.

7. Probing the initial state $|I\rangle$. To test successful generation of state $|I\rangle = \frac{1}{\sqrt{3}}(|1\rangle + |2\rangle + |3\rangle)$ (data in Fig. 3*A*, *i*), we show data from Bob's readout of the three-box game, with 1,200 repetitions of measuring each M_1, M_2, and M_3.

8. Probing repeatability. Data shown in Fig. 3*A*, *ii* and *iii* and in Fig. 3 *B* and *C* are obtained by correlating two successive readouts, implemented as Bob and Alice's readouts in the three-box game. However, here, we omit the NMR manipulation between Bob and Alice's readouts, such that both readout instances probe in the same basis. For each choice of Bob's measurement (M_1, M_2, or M_3) in the following readout, we probe all nuclear spin states within the $m_S = -1$ manifold in the same measurement run.

II. Analysis of the Leggett–Garg Inequality

A. Leggett–Garg Function Is Satisfied for MR Systems. The Leggett–Garg function $\langle K\rangle$ is defined in terms of three sequential measurements Q_1, Q_2, and Q_3, having eigenvalues ± 1, as:

$$\langle K\rangle = \langle Q_1Q_2\rangle + \langle Q_2Q_3\rangle + \langle Q_1Q_3\rangle$$

If the observables Q_j are classical values that satisfy the assumptions of MR [e.g., they have a definite value (state definiteness (SD)] and measurement of one value does not change the subsequent values [nondisturbing measurability (NDM)], there are then eight possible combinations of $Q_1 \ldots Q_3$. The Leggett–Garg function can be thought of as the sum of three parity checks on three classical bits, of which two parity checks, at most, can be odd. Enumerating the combinations of Q_j shows that $\langle K\rangle$ will lie in the range $-1 \le \langle K\rangle \le 3$ in every case, such that the inequality is satisfied (Table S2).

B. Quantum Systems Can Violate the Leggett–Garg Inequality. The Leggett–Garg inequality is violated for quantum systems, however, if coherence persists between the times that different Q_j values are measured. Typically, a violation is observed by evaluating each term in the Leggett–Garg sum during separate runs of an experiment, as follows:

$$\langle K\rangle = \langle Q_1Q_2\rangle|_{\text{runs excluding } Q_3} + \langle Q_2Q_3\rangle|_{\text{runs excluding } Q_1} + \langle Q_1Q_3\rangle|_{\text{runs excluding } Q_2}$$

A macrorealist does not object to neglecting to measure Q_2 on a run that evaluates $\langle Q_1Q_3\rangle$ because he assumes that measurements do not disturb the state of the system (i.e., NDM holds); however, in the quantum case, measuring Q_2 can change the expectation of $\langle Q_1Q_3\rangle$.

For a system with two states $|1\rangle$ and $|2\rangle$, we can define $Q_j = +1$ if the system is found in the state $|1\rangle$ and $Q_j = -1$ if the system is found in the state $|2\rangle$. A spin-½ electron is an example of a physical system possessing two states: We can take spin up as $|1\rangle$, spin down as $|2\rangle$, and $Q_j = \sigma_z$ where σ_z is a Pauli matrix. Suppose that we prepare the state $|1\rangle\rangle$at time t_1 and that, during the interval $\Delta t = t_1 \ldots t_2$, we allow the system to evolve according to a unitary $U(\Delta t)$ acting as:

$$U(\Delta t) = U = \exp\left(\frac{2\pi i}{3}\frac{\sigma_y}{2}\right)$$

Writing in matrix notation, we have:

$$|1\rangle = \begin{pmatrix}1\\0\end{pmatrix} \quad |2\rangle = \begin{pmatrix}0\\1\end{pmatrix} \quad Q = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$$

and for the time evolutions:

$$U(\Delta t) = U = \exp\left(\frac{2\pi i}{3}\begin{pmatrix}0 & -i/2\\ i/2 & 0\end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix}1 & \sqrt{3}\\ -\sqrt{3} & 1\end{pmatrix}$$

$$U(2\Delta t) = UU = \exp\left(\frac{4\pi i}{3}\begin{pmatrix}0 & -i/2\\ i/2 & 0\end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix}-1 & -\sqrt{3}\\ \sqrt{3} & -1\end{pmatrix}$$

The initial state at time t_1 is represented by the density matrix ρ_1:

$$\rho_1 = |1\rangle\langle 1| = \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}$$

The Leggett–Garg correlators $\langle Q_1Q_2\rangle$ and $\langle Q_2Q_3\rangle$ are evaluated as:

$$\begin{aligned}\langle Q_1Q_2\rangle &= \langle Q(t_1)Q(t_2)\rangle = \mathrm{Tr}\ \left((\rho_1)(Q)\left(UQU^\dagger\right)\right)\\ \langle Q_2Q_3\rangle &= \langle Q(t_2)Q(t_3)\rangle = \mathrm{Tr}\ \left(\left(U\rho_1U^\dagger\right)\left(UQU^\dagger\right)\left(UUQU^\dagger U^\dagger\right)\right)\\ &= \mathrm{Tr}\ \left((\rho_1)(Q)\left(UQU^\dagger\right)\right) = \langle Q_1Q_2\rangle\end{aligned}$$

The correlator $\langle Q_1Q_3\rangle$ is evaluated as:

$$\langle Q_1Q_3\rangle = \langle Q(t_1)Q(t_3)\rangle = \mathrm{Tr}\ \left((\rho_1)(Q)\left(UUQU^\dagger U^\dagger\right)\right)$$

Explicit calculation then yields:

$$\langle Q_1Q_2\rangle = \mathrm{Tr}\left(\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}\frac{1}{2}\begin{pmatrix}1 & -\sqrt{3}\\ \sqrt{3} & 1\end{pmatrix}\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix} \times \frac{1}{2}\begin{pmatrix}1 & \sqrt{3}\\ -\sqrt{3} & 1\end{pmatrix}\right) = -\frac{1}{2}$$

$$\langle Q_1Q_3\rangle = \mathrm{Tr}\left(\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}\frac{1}{2}\begin{pmatrix}-1 & -\sqrt{3}\\ \sqrt{3} & -1\end{pmatrix}\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix} \times \frac{1}{2}\begin{pmatrix}-1 & \sqrt{3}\\ -\sqrt{3} & -1\end{pmatrix}\right) = -\frac{1}{2}$$

and the Leggett–Garg function evaluates as:

$$\langle K\rangle = \langle Q_1Q_2\rangle + \langle Q_2Q_3\rangle + \langle Q_1Q_3\rangle = -\frac{3}{2},$$

which lies outside the range -1 to $+3$, violating the inequality.

III. Detectable Disturbance Is a Prerequisite to Violating the Leggett–Garg Inequality in 2D Hilbert Spaces

We are interested to know which of the hypotheses underpinning MR (SD or NDM) fails to hold when the Leggett–Garg inequality is violated. In a pre- and postselected system, we can define "detectable disturbance" as the change in the marginal statistics of the pre- and postselection, compared between the cases when an intervening measurement is performed or not. We demonstrate that for a two-level quantum system, violating the Leggett–Garg inequality requires detectable disturbance. We therefore show that the failure of NDM can explain all Leggett–Garg inequality violations in a two-level quantum system.

A. Leggett–Garg Inequality Specific to a Two-Level Quantum System. In a two-level quantum system, the Leggett–Garg function $\langle K\rangle$ is defined in terms of three measurements, Q_1, Q_2, and Q_3, with eigenvalues ± 1, taken sequentially at times t_1, t_2, and t_3. To observe a violation of the Leggett–Garg inequality, the averages $\langle Q_iQ_j\rangle$ must be performed using measurements at times t_i and t_j only. We write a subscript on the angle brackets to indicate the measurement times, such that:

$$\langle K\rangle = \langle Q_1Q_2\rangle_{(t_1,t_2)} + \langle Q_2Q_3\rangle_{(t_2,t_3)} + \langle Q_1Q_3\rangle_{(t_1,t_3)}$$

The Leggett–Garg function is then understood as the sum of results taken from three different ensemble averages in the quantum case. The hypotheses of MR (SD and NDM) lead a macrorealist to assume that each of the averages will be drawn from the same ensemble whenever a system obeying MR is measured.

1. Measurements of Q_j in the two-level case. A general measurement on a 2D quantum system is represented by a Pauli operator. We can assume that the two-level quantum system is degenerate and has no internal dynamics between the measurement times. If this is not the case, we simply absorb the dynamics during intervals $t_1 \ldots t_2$ and $t_2 \ldots t_3$ into the definitions of the Q_j measurements.

We then write Q_1, Q_2, and Q_3 as measurements along three directions $\hat{n}_1$, $\hat{n}_2$, and $\hat{n}_3$, such that $\hat{Q}_j = \hat{\sigma}\cdot\hat{n}_j$. For the purposes of evaluating the Leggett–Garg function, the important quantity is the inner product between the measurement directions. We define $\cos\theta_{ij} = \hat{n}_i\cdot\hat{n}_j$, and analysis shows that $\langle Q_iQ_j\rangle_{(t_i,t_j)} = \cos\theta_{ij}$. In terms of this, the Leggett–Garg function becomes:

$$\langle K\rangle = \cos\theta_{12} + \cos\theta_{23} + \cos\theta_{13} \quad \textbf{[S19]}$$

In the quantum case, we can pick three directions for $\hat{n}_j$, such that $\theta_{12} = \theta_{23} = \theta_{13} = 120° = 2\pi/3$ radians. Because $\cos(2\pi/3) = -1/2$, this choice obtains $\langle K\rangle = -3/2$ when the quantum system is measured, violating the inequality.

2. Detectable disturbance during measurement. We define the detectable disturbance D as the difference in $\langle K\rangle$ induced by performing pairs of measurements, compared with performing all three measurements:

$$D = \langle Q_1Q_2\rangle_{(t_1,t_2)} - \langle Q_1Q_2\rangle_{(t_1,t_2,t_3)} + \langle Q_2Q_3\rangle_{(t_2,t_3)} - \langle Q_2Q_3\rangle_{(t_1,t_2,t_3)} + \langle Q_1Q_3\rangle_{(t_1,t_3)} - \langle Q_1Q_3\rangle_{(t_1,t_2,t_3)}$$

Analysis shows that the nonzero contributions to D arise from the $\langle Q_1Q_3\rangle$ terms; thus, D is a comparison between evaluating $\langle Q_1Q_3\rangle$ in an experiment in which the system is measured at times t_1 and t_3 only and an experiment in which the system is measured at each of the times t_1, t_2, and t_3, but with the Q_2 measurement result discarded. We have:

$$\begin{aligned} D = \langle Q_1Q_3\rangle_{(t_1,t_3)} - \langle Q_1Q_3\rangle_{(t_1,t_2,t_3)} = \langle Q_1Q_3\rangle_{(t_1,t_3)} \\ -P(Q_2=+1)\langle Q_1Q_3|Q_2=+1\rangle_{(t_1,t_2,t_3)} \\ -P(Q_2=-1)\langle Q_1Q_3|Q_2=-1\rangle_{(t_1,t_2,t_3)} \end{aligned}$$

The difference between $\langle Q_1Q_3\rangle_{(t_1,t_3)}$ and $\langle Q_1Q_3\rangle_{(t_1,t_2,t_3)}$ arises because including a Q_2 measurement at time t_2 and discarding the result will nevertheless project the quantum system onto the eigenstates of $\hat{\sigma}\cdot\hat{n}_2$ before the Q_3 measurement. We find that:

$$\begin{aligned} &P(Q_2=+1)\langle Q_1Q_3|Q_2=+1\rangle_{(t_1,t_2,t_3)} \\ &+P(Q_2=-1)\langle Q_1Q_3|Q_2=-1\rangle_{(t_1,t_2,t_3)} \\ &=\cos^2(\theta_{12}/2)\cos\theta_{23} - \sin^2(\theta_{12}/2)\cos\theta_{23} \end{aligned}$$

The expression for the disturbance then becomes:

$$\begin{aligned} D &= \langle Q_1Q_3\rangle_{(t_1,t_3)} - P(Q_2=+1)\langle Q_1Q_3|Q_2=+1\rangle_{(t_1,t_2,t_3)} \\ &\quad -P(Q_2=-1)\langle Q_1Q_3|Q_2=-1\rangle_{(t_1,t_2,t_3)} \\ &= \cos\theta_{13} - \cos^2(\theta_{12}/2)\cos\theta_{23} + \sin^2(\theta_{12}/2)\cos\theta_{23} \\ &= \cos\theta_{13} - \left(\cos^2(\theta_{12}/2) - \sin^2(\theta_{12}/2)\right)\cos\theta_{23} \\ &= \cos\theta_{13} - \cos\theta_{12}\cos\theta_{23} \end{aligned} \quad \textbf{[S20]}$$

The condition for obtaining no detectable disturbance ($D = 0$) is therefore:

$$\cos\theta_{13} = \cos\theta_{12}\cos\theta_{23} \quad \textbf{[S21]}$$

If the condition in Eq. **S21** can be satisfied, all observers will agree that a nondetectable measurement of Q_2 has taken place, whereas if the condition in Eq. **S21** is violated, no one could believe that NDM has taken place.

B. Zero Detectable Disturbance Implies the Leggett–Garg Inequality Is Satisfied. We now show that a measurement sequence that has no detectable disturbance will necessarily satisfy the Leggett–Garg inequality in a two-level quantum system. Starting with Eq. **S19**, we substitute in the condition $\cos\theta_{13} = D + \cos\theta_{12}\cos\theta_{23}$ from Eq. **S21**, obtaining:

$$\langle K\rangle = \cos\theta_{12} + \cos\theta_{23} + \cos\theta_{13}$$

$$\langle K\rangle = \cos\theta_{12} + \cos\theta_{23} + \cos\theta_{12}\cos\theta_{23} + D$$

Rearranging and collecting terms, we have:

$$\langle K - D + 1\rangle = \cos\theta_{12}\cos\theta_{23} + \cos\theta_{12} + \cos\theta_{23} + 1$$

$$\langle K - D + 1\rangle = (\cos\theta_{12} + 1)(\cos\theta_{23} + 1)$$

We find $\langle K - D + 1\rangle$ consists of the product of two terms, each of which is in the range $0 \le \langle\cos\theta_{ij} + 1\rangle \le 2$, such that the whole Leggett–Garg function is in the range $D - 1 \le \langle K\rangle \le D + 3$, or:

$$D = 0 \Rightarrow -1 \le \langle K\rangle \le +3$$

The condition for zero-disturbance is therefore identical to the condition that the Leggett–Garg inequality is satisfied, and violation of the Leggett–Garg inequality must be accompanied by detectable disturbance in the two-level case.

No measurement of a two-level quantum system could convince a stubborn macrorealist that a failure of SD alone has taken place, because NDM will necessarily have failed in any successful Leggett–Garg inequality violation demonstrated in a two-level quantum system. This is in comparison to our work on a three-level quantum system, in which we show that NDM remains valid, whereas a Leggett–Garg inequality is violated.

1. Analysis of the detectable disturbance. Here, we derive some results asserted in the proof above. The detectable disturbance D is the change in $\langle K\rangle$ induced by measuring at pairs of times vs. measuring at all three times. D contains three terms. We have:

$$D = \underbrace{\langle Q_1Q_2\rangle_{(t_1,t_2)} - \langle Q_1Q_2\rangle_{(t_1,t_2,t_3)}}_{D_{12}} + \underbrace{\langle Q_2Q_3\rangle_{(t_2,t_3)} - \langle Q_2Q_3\rangle_{(t_1,t_2,t_3)}}_{D_{23}} + \underbrace{\langle Q_1Q_3\rangle_{(t_1,t_3)} - \langle Q_1Q_3\rangle_{(t_1,t_2,t_3)}}_{D_{13}}$$

where D_{ij} represents the change to the expected value of the two-measurement correlation $\langle Q_iQ_j\rangle$ induced by performing the third measurement $Q_{k\neq i,j}$ while ignoring the Q_k result.

2. Evaluating D_{12} and D_{23}. Clearly, $D_{12} = 0$, because the measurement at time t_3 would otherwise influence the result of past events at time t_1 or t_2. We might suspect by symmetry that $D_{23} = \langle Q_2Q_3\rangle_{(t_2,t_3)} - \langle Q_2Q_3\rangle_{(t_1,t_2,t_3)}$ will also be zero, and we can show this explicitly. The overlaps of the Pauli operators $\hat{Q}_2 = \hat{\sigma}\cdot\hat{n}_2$ and $\hat{Q}_3 = \hat{\sigma}\cdot\hat{n}_3$ yield:

$$P(Q_3 = \pm1|Q_2 = \pm1) = \cos^2(\theta_{23}/2)$$

$$P(Q_3 = \pm1|Q_2 = \mp1) = \sin^2(\theta_{23}/2)$$

and:

$$\begin{aligned} \langle Q_2Q_3\rangle_{(t_2,t_3)} &= P(Q_3=\pm1|Q_2=\pm1) - P(Q_3=\pm1|Q_2=\mp1) \\ &= \cos^2(\theta_{23}/2) - \sin^2(\theta_{23}/2) = \cos\theta_{23} \end{aligned}$$

We now need to evaluate $\langle Q_2Q_3\rangle_{(t_1,t_2,t_3)}$. Suppose that the Q_1 preparation followed by Q_2 measurements yields $Q_2 =$

+1 with probability p and $Q_2 = -1$ with probability $1 - p$. We have:

$$\begin{aligned}\langle Q_2Q_3\rangle_{(t_1,t_2,t_3)} &= p[P(Q_3=+1|Q_2=+1)-P(Q_3=-1|Q_2=+1)] \\ &\quad +(1-p)[P(Q_3=-1|Q_2=-1)-P(Q_3=+1|Q_2=-1)] \\ &= p\cos^2(\theta_{23}/2)-p\sin^2(\theta_{23}/2)+(1-p)\cos^2(\theta_{23}/2) \\ &\quad -(1-p)\sin^2(\theta_{23}/2)=\cos^2(\theta_{23}/2)-\sin^2(\theta_{23}/2)=\cos\theta_{23}\end{aligned}$$

from which the influence of the Q_1 measurement represented by p cancels, implying $D_{23} = 0$.

3. Evaluating D_{13}. We can see that $\langle Q_1Q_3\rangle_{(t_1,t_3)}$ is insensitive to the state before Q_1, by substituting $t_1 \to t_2$ and following a similar argument as for $\langle Q_2Q_3\rangle_{(t_2,t_3)}$ above. We show that $\langle Q_1Q_3\rangle_{(t_1,t_2,t_3)}$ is also insensitive to the initial state by assuming that the state before Q_1 measurement yields $Q_1 = +1$ with probability q and $Q_1 = -1$ with probability $1 - q$. We have:

$$\begin{aligned}\langle Q_1Q_3\rangle_{(t_1,t_2,t_3)} &= P(Q_2=+1)\langle Q_1Q_3|Q_2=+1\rangle \\ &\quad +P(Q_2=-1)\langle Q_1Q_3|Q_2=-1\rangle \\ &= qP(Q_2=+1|Q_1=+1)[P(Q_3=+1|Q_2=+1) \\ &\quad -P(Q_3=-1|Q_2=+1)]+qP(Q_2=-1|Q_1=+1) \\ &\quad \times[P(Q_3=+1|Q_2=-1)-P(Q_3=-1|Q_2=-1)] \\ &\quad +(1-q)P(Q_2=+1|Q_1=-1)[P(Q_3=+1|Q_2=+1) \\ &\quad -P(Q_3=-1|Q_2=+1)]+(1-q)P(Q_2=-1|Q_1=-1) \\ &\quad \times[P(Q_3=+1|Q_2=-1)-P(Q_3=-1|Q_2=-1)]\end{aligned}$$

This expression contains 16 terms, yielding:

$$\begin{aligned}\langle Q_1Q_3\rangle_{(t_1,t_2,t_3)} &= q\cos^2(\theta_{12}/2)\left[\cos^2(\theta_{23}/2)-\sin^2(\theta_{23}/2)\right] \\ &\quad +q\sin^2(\theta_{12}/2)\left[\sin^2(\theta_{23}/2)-\cos^2(\theta_{23}/2)\right] \\ &\quad +(1-q)\sin^2(\theta_{12}/2)\left[-\cos^2(\theta_{23}/2)+\sin^2(\theta_{23}/2)\right] \\ &\quad +(1-q)\cos^2(\theta_{12}/2)\left[-\sin^2(\theta_{23}/2)+\cos^2(\theta_{23}/2)\right]\end{aligned}$$

The initial preparation represented by q again cancels:

$$\begin{aligned}\langle Q_1Q_3\rangle_{(t_1,t_2,t_3)} &= \cos^2(\theta_{12}/2)\left[\cos^2(\theta_{23}/2)-\sin^2(\theta_{23}/2)\right] \\ &\quad -\sin^2(\theta_{12}/2)\left[\cos^2(\theta_{23}/2)-\sin^2(\theta_{23}/2)\right] \\ &= \left[\cos^2(\theta_{12}/2)-\sin^2(\theta_{12}/2)\right]\left[\cos^2(\theta_{23}/2)\right. \\ &\quad \left.-\sin^2(\theta_{23}/2)\right]=\cos\theta_{12}\cos\theta_{23}\end{aligned}$$

The total expression for the detectable disturbance is then:

$$\begin{aligned}D &= \underbrace{D_{12}}_{0}+\underbrace{D_{23}}_{0}+D_{13}=\langle Q_1Q_3\rangle_{(t_1,t_3)}-\langle Q_1Q_3\rangle_{(t_1,t_2,t_3)} \\ &= \cos\theta_{13}-\cos\theta_{12}\cos\theta_{23}\end{aligned}$$

This is Eq. **S21** in the text above. Detectable disturbance is therefore a necessary consequence of violating the Leggett–Garg inequality in a two-level quantum system.

IV. Classical Models of the Three-Box Problem

A classical probability model of the three-box problem can be constructed using Kolmogorov's axioms by assuming that the system of three boxes exists at all times in a definite state of having one box occupied and the other two empty. States of the system are conventionally labeled "λ" in this context. The simplest model of three boxes sharing one ball assumes a one-to-one correspondence between the system states λ_j and the available boxes, such that if λ is known, all measurement results can be inferred with certainty; the system with three states λ_1, λ_2, and λ_3 behaves such that being in state λ_j corresponds to finding M_j-true and $M_{k\neq j}$-false. On any particular run of the experiment, the system is in a definite state λ, but we may not know what this state is. We can describe our knowledge of the state with the probability distribution over the λ, writing $P(\lambda_i)$ with $\sum_i P(\lambda_i)=1$. We can also assume many equivalent microstates $\{\lambda_1, \lambda_1', \lambda''_2, \cdots\}$ exist, which produce identical results when studied with the M_j measurements. This situation is illustrated in Fig. S1.

We can consider a continuous set of states $\{\lambda\}$ and a probability distribution over these as $\mu(\lambda)$, such that:

$$\mu(\lambda)\in\mathcal{R} \qquad \mu(\lambda)\geq 0\ \forall\ \lambda \qquad \int d\lambda\ \mu(\lambda)=1$$

Given a probability distribution $\mu(\lambda)$, we can define a measurement function $\xi_{\text{context}}(\text{result}|\lambda)$ that describes how each state λ will respond when measured. The probability of observing a particular result is $P_{\text{context}}(\text{result})$; thus, for instance, if Alice is measuring M_1, we write $P_{M_1}(A)$ as the probability that Alice finds M_1-true, and $P_{M_1}(\neg A)$ as the probability that Alice finds M_1-false:

$$P_{M_1}(A)=\int d\lambda\ \xi_{M_1}(A|\lambda)\mu(\lambda) \qquad P_{M_1}(\neg A)=\int d\lambda\ \xi_{M_1}(\neg A|\lambda)\mu(\lambda)$$

The key difference between such a "classical" model and the quantum picture of the experiment is that quantum interference is not possible because $\mu(\lambda)\geq 0$. These concepts are explored in related work (5).

V. Deriving a Leggett–Garg Function for the Three-Box Game

For a two-level quantum system, we have shown that it is necessary to perform the measurements of $\langle Q_1Q_2\rangle$, $\langle Q_2Q_3\rangle$, and $\langle Q_1Q_3\rangle$ independently to obtain a violation of the Leggett–Garg inequality. In the three-level case, a measurement result such as $\neg M_1$ can preserve the coherence between states $|2\rangle$ and $|3\rangle$, and it is therefore possible to violate a Leggett–Garg inequality while making three sequential measurements during the same run of an experiment. To derive the Leggett–Garg function specific to the three-level case, we must understand that a macrorealist can make counterfactual inferences that differ from the inferences that are valid for a quantum system, for example:

$$\neg M_1 \Rightarrow (M_2\wedge\neg M_3)\vee(\neg M_2\wedge M_3) \qquad \textbf{[S22]}$$

Seeing box 1 empty means either that the ball is in box 2 and not in box 3 or that the ball is in box 3 and not in box 2. Superpositions between boxes 2 and 3 are not allowed in the MR picture, due to SD, but superpositions are allowed before measurement in the quantum case and can survive following a partial measurement in the unobserved states. We can exploit the difference between the MR and quantum mechanics (QM) positions to derive an expression for $\langle K\rangle$ using counterfactual inferences, which would be valid if MR (and specifically SD) holds; we can then look for a violation of the Leggett–Garg inequality using an expression derived using counterfactual inferences.

A basic property of the three-box problem is that Alice is unable to detect any effect of measurements made by Bob; by definition, a successful demonstration of the three-box problem uses nondisturbing measurements. If we can derive a Leggett–Garg function specific to the three-box case and show that the inequality is violated by an experiment that successfully implements the three-box problem, the assumption of SD is shown to be invalid.

We have shown that all successful violations of the Leggett–Garg inequality in a two-level quantum system must involve measurements that cause disturbance; therefore, by extending the Leggett–Garg inequality to a three-level system, we can show that the absence of state definiteness alone is responsible for the Leggett–Garg inequality violation. In this way, we go beyond existing studies in the literature. We now derive the Leggett–Garg inequality for the three-level system.

A. Probability Notation. In the macrorealist picture, finding the system in state j corresponds to finding a macroscopic object, such as a hidden ball, in a location, such as box j. We write probabilities $P_{M_j}(B)$ to indicate the chance that when Bob performs measurement M_j, he sees a full box (finds state j) and $P_{M_j}(\neg B)$ as the probability that he finds box j is empty (measures "not state j"). The probability of the combined event, where both Bob and Alice see full boxes (Bob finds state j, followed by Alice finding M_3-true on her final measurement), is $P_{M_j}(B\cap A)=P_{M_j}(B|A)P_{M_j}(A)$, whereas Alice's probability of finding M_3-true when Bob has made no intervening measurement is written as $P_N(A)$.

The probabilities $P_{M_j}(\cdots)$ and $P_N(\ldots)$ are well-defined in both quantum and macrorealist theories, but our objective is to highlight the differences between these two theoretical descriptions. A macrorealist further believes that "counterfactual" expressions take a definite value. He defines quantities, such as $\tilde{P}_{\tilde{M}_j}(B)$, that give the probability for Bob to have found the object had he performed M_j. This allows a macrorealist to insert a resolution of the identity into his expressions for probabilities as:

$$\tilde{P}_{\tilde{M}_1}(B)+\tilde{P}_{\tilde{M}_2}(B)+\tilde{P}_{\tilde{M}_3}(B)=1 \quad \textbf{[S23]}$$

wherever he chooses. (We track quantities that are undefined in quantum mechanics with tilde symbols.)

B. Leggett–Garg Function for the Three-Level System. Given the definition of:

$$\langle K\rangle=\langle Q_1Q_2\rangle+\langle Q_2Q_3\rangle+\langle Q_1Q_3\rangle \quad \textbf{[S24]}$$

we apply the Leggett–Garg analysis to our system as follows. Our experiment uses measurement-based initialization (4) to prepare the initial state for Alice, and we therefore take $Q_1 = +1$ in all cases. We assign $Q_2 = -1$ whenever Bob observes the object in box 1 or box 2 and $Q_2 = +1$ whenever Bob should infer the object is in box 3. We assign $Q_3 = +1$ whenever Alice's final M_3 result is true and assign $Q_3 = -1$ whenever the M_3 result is false.

If the macrorealist framework is applicable, one of six possible measurement histories (*a*–*f*) must account for each particular run of the experiment (Table S1). To assign the probabilities that a given history occurred, the macrorealist must calculate the unobserved quantities $\tilde{P}_{\tilde{M}_3}(B\cap A)$ and $\tilde{P}_{\tilde{M}_3}(B\cap\neg A)$. If the measurements are operationally nondisturbing (a property that we check experimentally), it is possible to substitute $\tilde{P}_{\tilde{M}_1}(B|A)\Rightarrow P_{M_1}(B|A)$ and $\tilde{P}_{\tilde{M}_2}(B|A)\Rightarrow P_{M_2}(B|A)$ (etc.) for the measurements that are made, such that:

$$\tilde{P}_{\tilde{M}_3}(B)=1-P_{M_1}(B)-P_{M_2}(B) \quad \textbf{[S25]}$$

$$\tilde{P}_{\tilde{M}_3}(B\cap A)=P_N(A)-P_{M_1}(B\cap A)-P_{M_2}(B\cap A) \quad \textbf{[S26]}$$

$$\tilde{P}_{\tilde{M}_3}(B\cap\neg A)=P_N(\neg A)-P_{M_1}(B\cap\neg A)-P_{M_2}(B\cap\neg A) \quad \textbf{[S27]}$$

Using these definitions, the macrorealist framework deduces the expression for $\langle K\rangle$ (Table S1) as:

$$\begin{aligned}\langle K\rangle=&-P_{M_1}(B\cap A)-P_{M_2}(B\cap A)+3\tilde{P}_{\tilde{M}_3}(B\cap A)\\&-P_{M_1}(B\cap\neg A)-P_{M_2}(B\cap\neg A)-\tilde{P}_{\tilde{M}_3}(B\cap\neg A)\end{aligned} \quad \textbf{[S28]}$$

which, in terms of observable quantities, is:

$$\begin{aligned}\langle K\rangle=&-P_{M_1}(B\cap A)-P_{M_1}(B\cap\neg A)-P_{M_2}(B\cap A)-P_{M_2}(B\cap\neg A)\\&+3P_N(A)-3P_{M_1}(B\cap A)-3P_{M_2}(B\cap A)\\&-P_N(\neg A)+P_{M_1}(B\cap\neg A)+P_{M_2}(B\cap\neg A)\end{aligned} \quad \textbf{[S29]}$$

This expression simplifies to:

$$\langle K\rangle=4P_N(A)-4P_{M_1}(B\cap A)-4P_{M_2}(B\cap A)-1 \quad \textbf{[S30]}$$

We know that the quantum expressions for these occurrences are:

$$P_N(A)=|\langle F|I\rangle|^2=1/9 \quad \textbf{[S31]}$$

$$P_{M_1}(B\cap A)=|\langle F|\hat{P}_1|I\rangle|^2=1/9 \quad \textbf{[S32]}$$

$$P_{M_2}(B\cap A)=|\langle F|\hat{P}_2|I\rangle|^2=1/9 \quad \textbf{[S33]}$$

We observe two points here, given the probabilities above. The first is that Alice is unable to determine whether Bob has chosen to perform measurement M_1, measurement M_2, or neither measurement (N), and Alice's result is independent of measurement context. We have:

$$P_{M_1}(A)=P_{M_2}(A)=P_N(A) \quad \textbf{[S34]}$$

The second probability is that:

$$P_{M_1}(B\neg\cap A)=P_{M_2}(B\neg\cap A)=0 \quad \textbf{[S35]}$$

Alice will never find her M_3 result true when Bob has found his M_j result false, which is the key feature that enables Alice to win the three-box game. This implies:

$$P_{M_1}(B\cap A)=P_{M_1}(A) \quad \textbf{[S36]}$$

$$P_{M_2}(B\cap A)=P_{M_2}(A) \quad \textbf{[S37]}$$

We can extract the probability of Alice's measurement from each term via Bayes theorem:

$$P((B\cap A))=P(B|A)P(A) \quad \textbf{[S38]}$$

yielding:

$$\begin{aligned}\langle K\rangle&=4P(A)(1-P_{M_1}(B|A)-P_{M_2}(B|A))-1\\&=\frac{4}{9}(1-P_{M_1}(B|A)-P_{M_2}(B|A))-1\end{aligned} \quad \textbf{[S39]}$$

Given the macrorealist's hypothesis, the events under M_1 and M_2 should be mutually exclusive, and sums of events under these cases will obey an inequality:

$$P_{M_1}(\cdots)+P_{M_2}(\cdots)\le 1 \quad \textbf{[S40]}$$

The equality holds when the Leggett–Garg function in Eq. **S39** takes as its limiting value $\langle K\rangle = -1$. In the quantum case, meanwhile, $P_{M_1}(\cdots)$ and $P_{M_2}(\cdots)$ are independent, and we have:

$$P_{M_1}(\cdots)+P_{M_2}(\cdots)\le 2 \quad \textbf{[S41]}$$

allowing the Leggett–Garg function to reach a value of:

$$\langle K\rangle=-\frac{13}{9}=-1.4\dot{4} \quad \textbf{[S42]}$$

This is outside the range $-1 \leq \langle K \rangle \leq 3$, providing an opportunity to detect an inconsistency with MR.

VI. Error Analysis of the Experimental Results

We find small deviations from the values expected from an ideal implementation. In the following, we give a brief description of the origin of these discrepancies and discuss their consequences on the macrorealist's possible conclusions.

i) We find $\sum_j P_{M_j}(B) < 1$ (i.e., there is not always a ball found in all the boxes). This is a consequence of a smaller than unity probability of correctly identifying the electronic $m_S = 0$ state, resulting in an effective detection efficiency of $P_{det} \approx 90\%$. Although the macrorealist might conclude that there is not always an object hidden in the boxes, he still finds an unbiased initial state (within statistical uncertainty). Therefore, he cannot expect Alice to take advantage of this discrepancy. Based on his secret choice of M_1 or M_2 and the reduced probability of finding an object, he expects a maximum probability of $\frac{P_{det}}{2} \leq \frac{1}{2}$ of Alice predicting his positive measurement outcome correctly; thus, the macrorealist finds an even stronger violation of his expectations.

ii) For sequential measurements $i, i+1$, we find both $P(M_{j,i+1}|M_{j,i})$ and $P(\neg M_{j,i+1}|\neg M_{j,i}) < 100\%$ (Fig. 3*B*) (i.e., after measuring its position, with a small probability, the object is moved to a different box). This finding could indeed explain correlations between Bob and Alice's measurements: As a worst-case scenario, Bob could assume a hidden mechanism in the game whereby his successful measurement "moves" the object, deterministically storing it in the box Alice is probing and maximizing her conditional probability $P_{M_j}(B|A)$. He would deduce an upper limit for her success probability of $P_{M_j}(B|A) \leq \frac{1}{3} + P(\text{object moves})$. Taking into account all "Changed" and "Undetermined" events (Fig. 3*B*), he finds $P(\text{object moves}) \leq 28\%$ and $P_{M_j}(B|A) \leq 61\%$, clearly violated by the experimental findings.

iii) We find $P(A) \approx 14\% > 1/9$ (Fig. 4*A*). However, from the QM description, we expect:

$$P_N(A) = |\langle F|I\rangle|^2 = 1/9 \qquad \textbf{[S43]}$$

$$P_{M_1}(B \cap A) = |\langle F|\hat{P}_1|I\rangle|^2 = 1/9 \qquad \textbf{[S44]}$$

$$P_{M_2}(B \cap A) = |\langle F|\hat{P}_2|I\rangle|^2 = 1/9 \qquad \textbf{[S45]}$$

In our implementation, between measurement A and B, we apply the transformation $|F\rangle \to |I\rangle \to |3\rangle$, consisting of NMR pulses of a total duration of ≈750 μs. The rf-induced heating of the sample and nuclear spin dephasing limit the fidelity of this operation, leading to an increased probability $P(A)$. In the QM picture, Alice detects more positive results than she should (unconditional on measurement B); thus, her conditional probability $P_{M_j}(B|A)$ to predict Bob's measurement outcome correctly must drop below the theoretical maximum of 100% (Fig. 4*B*).

A. Statistical Error Analysis. For each particular run of our experiment, we can either count one or more photons ($n \geq 1$) or no photons ($n = 0$), inferring that the electron is in the $m_S = 0$ or $m_S = \pm 1$ state. A detailed analysis of the inferences between photon number and spin state was presented by Robledo et al. (4) as a combination of geometric distribution (accounting for the spin flip rate), binomial distribution (accounting for photon detector efficiency), and the poissonian background rate. For the purposes of our analysis, we define a variable $P = |m_S|$, which is the value of Bob's or Alice's M_j result on any particular round of the experiment. We assign $P = 0$ when we count $n = 0$ photons, and we assign $P = 1$ when we count $n \geq 1$ photons. We then define the probability p of finding $m_S = 0$ during a particular shot of the experiment as f, so that during N trials of the experiment, we expect to observe statistics:

$$\text{Mean}[p] = Nf \qquad \textbf{[S46]}$$

$$\text{Var}[p] = \sigma^2(p) = Nf(1-f) \qquad \textbf{[S47]}$$

$$\text{Standard Deviation}[p] = \sigma(p) = \sqrt{Nf(1-f)} \qquad \textbf{[S48]}$$

We use this to calculate the statistical significance of our results (e.g., the chance that the Leggett–Garg function we measured is compatible with MR and that counting statistics have produced a violation by chance).

B. Fair Sampling vs. Adversarial Macrorealist Positions. In our experiment, we have the option to measure either the population in electron spin sublevel $m_S = -1$ or in the electron spin sublevels $m_S = -1$ and $m_S = +1$ when performing Bob and Alice's measurements M_j. Measuring the $m_S = -1$ populations only, we have the possibility of obtaining "undetermined" outcomes in which the population branches from $m_S = -1$ to the uninspected $m_S = +1$ levels during measurement, whereas by measuring the $m_S = -1$ and $m_S = +1$ levels, we minimize these undetermined events while increasing the number of Δm_I nuclear spin flips, which corresponds to Bob measuring that the state has definitely changed between subsequent measurements.

There are two approaches that we could use to interpret the undetermined outcomes. The default assumption is that the unmeasured values are distributed fairly and will follow the same distribution as the measured values, whereas the more extreme assumption is that each unmeasured value somehow represents Alice "cheating" by hiding values that favor the macrorealist hypothesis. If we take this extreme position, it is interesting to know whether a result compatible with MR could be recovered by allowing Bob to assign a value to each undetermined result as he pleases (6). We then define quantities such as:

$$P(A \cap B)^{\min}_{M1} = \frac{N_{M1}(B \cap A)}{N_{M1}(B \cap A) + N_{M1}(\neg B \cap A) + N_{M_1}(U)} \qquad \textbf{[S49]}$$

$$P(A \cap B)^{\text{fair}}_{M1} = \frac{N_{M1}(B \cap A)}{N_{M1}(B \cap A) + N_{M1}(\neg B \cap A)} \qquad \textbf{[S50]}$$

$$P(A \cap B)^{\max}_{M1} = \frac{N_{M1}(B \cap A) + N_{M_1}(U)}{N_{M1}(B \cap A) + N_{M1}(\neg B \cap A) + N_{M_1}(U)} \qquad \textbf{[S51]}$$

where $N_{M_1}(U)$ is the number of undetermined measurement readings, given that Bob has performed M_1. This bounds the possibilities for Bob to reassign undetermined readings. In fact, both in the case in which we assume fair sampling and without, we find that $\langle K \rangle \leq -1$ and that our results are therefore incompatible with MR. We calculate each case and include errors as per our statistical analysis above. In the case that we include only the $m_S = -1$ readout, we find:

$$K^{\min}_{|m_S=-1} = -1.2026 \quad \sigma^{\min}_{|m_S=-1} = 0.0259 \ \ (7.81\sigma \text{ violation}) \qquad \textbf{[S52]}$$

$$K^{\text{fair}}_{|m_S=-1} = -1.2647 \quad \sigma^{\text{fair}}_{|m_S=-1} = 0.0234 \ \ (11.29\sigma \text{ violation}) \qquad \textbf{[S53]}$$

$$K^{\max}_{|m_S=-1} = -1.3494 \quad \sigma^{\max}_{|m_S=-1} = 0.0173 \ \ (20.19\sigma \text{ violation}) \qquad \textbf{[S54]}$$

When using the complete register readout on $m_S = -1$ and $m_S = +1$, we have:

$$K^{\min}_{|m_S=\pm1} = -1.1373 \quad \sigma^{\min}_{|m_S=\pm1} = 0.0252 \quad (5.46\sigma \text{ violation}) \quad \textbf{[S55]}$$

$$K^{\text{fair}}_{|m_S=\pm1} = -1.1833 \quad \sigma^{\text{fair}}_{|m_S=\pm1} = 0.0241 \quad (7.60\sigma \text{ violation}) \quad \textbf{[S56]}$$

$$K^{\max}_{|m_S=\pm1} = -1.2531 \quad \sigma^{\max}_{|m_S=\pm1} = 0.0210 \quad (12.07\sigma \text{ violation}) \quad \textbf{[S57]}$$

In the event, we found that the undetermined measurement outcomes do not give Bob sufficient leeway to explain the discrepancy of our result from the range predicted by MR, even when taking the most adversarial position permissible with respect to our data.

1. Leggett AJ, Garg A (1985) Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Phys Rev Lett* 54(9):857–860.
2. Aharon N, Vaidman L (2008) Quantum advantages in classically defined tasks. *Phys Rev A* 77(5):052310.
3. Bernien H, et al. (2012) Two-photon quantum interference from separate nitrogen vacancy centers in diamond. *Phys Rev Lett* 108(4):043604.
4. Robledo L, et al. (2011) High-fidelity projective read-out of a solid-state spin quantum register. *Nature* 477(7366):574–578.
5. Maroney OJE (2012) Detectability, invasiveness and the quantum three box paradox. *arXiv quant-ph*, 1207.3114v1.
6. Knee GC, et al. (2012) Violation of a Leggett-Garg inequality with ideal non-invasive measurements. *Nat Commun* 3:606.

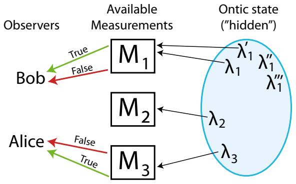


Fig. S1. Classical model of the three-box problem. In a simple classical model, the system is assumed to exist in a definite state λ_j. The specific state λ_j then determines how the system will respond to each measurement M_j.

Table S1. Assignment Q_j values for each run of the experiment

Case	Q_1	Bob measures	Q_2	Alice measures	Q_3	K	Probability
a	+1	M_1	−1	M_3	+1	−1	$P_{M_1}(B\cap A)$
b	+1	M_2	−1	M_3	+1	−1	$P_{M_2}(B\cap A)$
c	+1	Infers M_3	+1	M_3	+1	+3	$\tilde{P}_{\tilde{M}_3}(B\cap A)$
d	+1	M_1	−1	$\neg M_3$	−1	−1	$P_{M_1}(B\cap\neg A)$
e	+1	M_2	−1	$\neg M_3$	−1	−1	$P_{M_2}(B\cap\neg A)$
f	+1	Infers M_3	+1	$\neg M_3$	−1	−1	$\tilde{P}_{\tilde{M}_3}(B\cap\neg A)$

According to the MR picture, one of the six cases above must account for each run of the experiment. The measured probabilities P_{M_1} and P_{M_2} and inferred (counterfactual) probabilities $\tilde{P}_{\tilde{M}_3}$ that weight the value of K corresponding to each history are listed in the table.

Table S2. Enumerating values of Q_1, Q_2, and Q_3 for the Leggett–Garg function in a classical system

Measurements			Parity checks/correlators			Leggett–Garg function
Q_1	Q_2	Q_3	Q_1Q_2	Q_2Q_3	Q_1Q_3	K
+1	+1	+1	+1	+1	+1	3
−1	−1	−1	+1	+1	+1	3
+1	+1	−1	+1	−1	−1	−1
+1	−1	+1	−1	−1	+1	−1
−1	+1	+1	−1	+1	−1	−1
−1	−1	+1	+1	−1	−1	−1
−1	+1	−1	−1	−1	+1	−1
+1	−1	−1	−1	+1	−1	−1

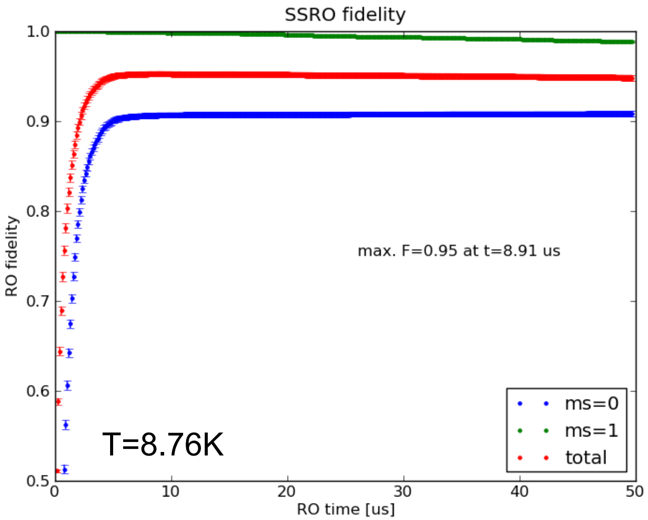
Each combination of Q_j yields a K value between −1 and +3.

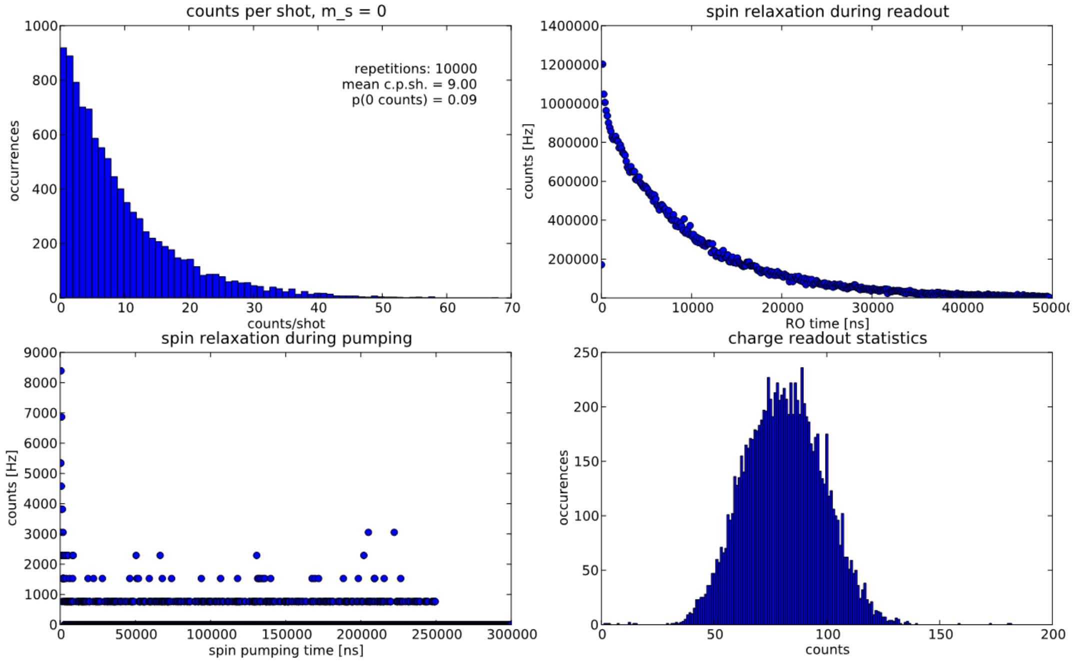
Other Supporting Information Files

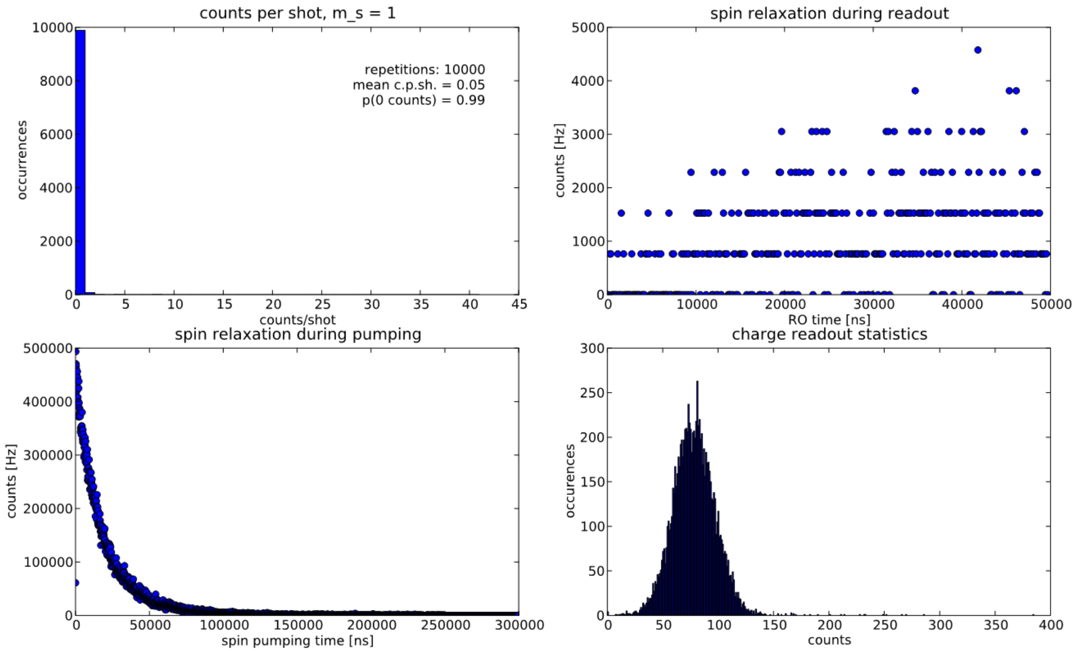
Dataset S1 (PDF)
Dataset S2 (PDF)

SIL2 – 3box paradox: electron spin readout

17/12/11

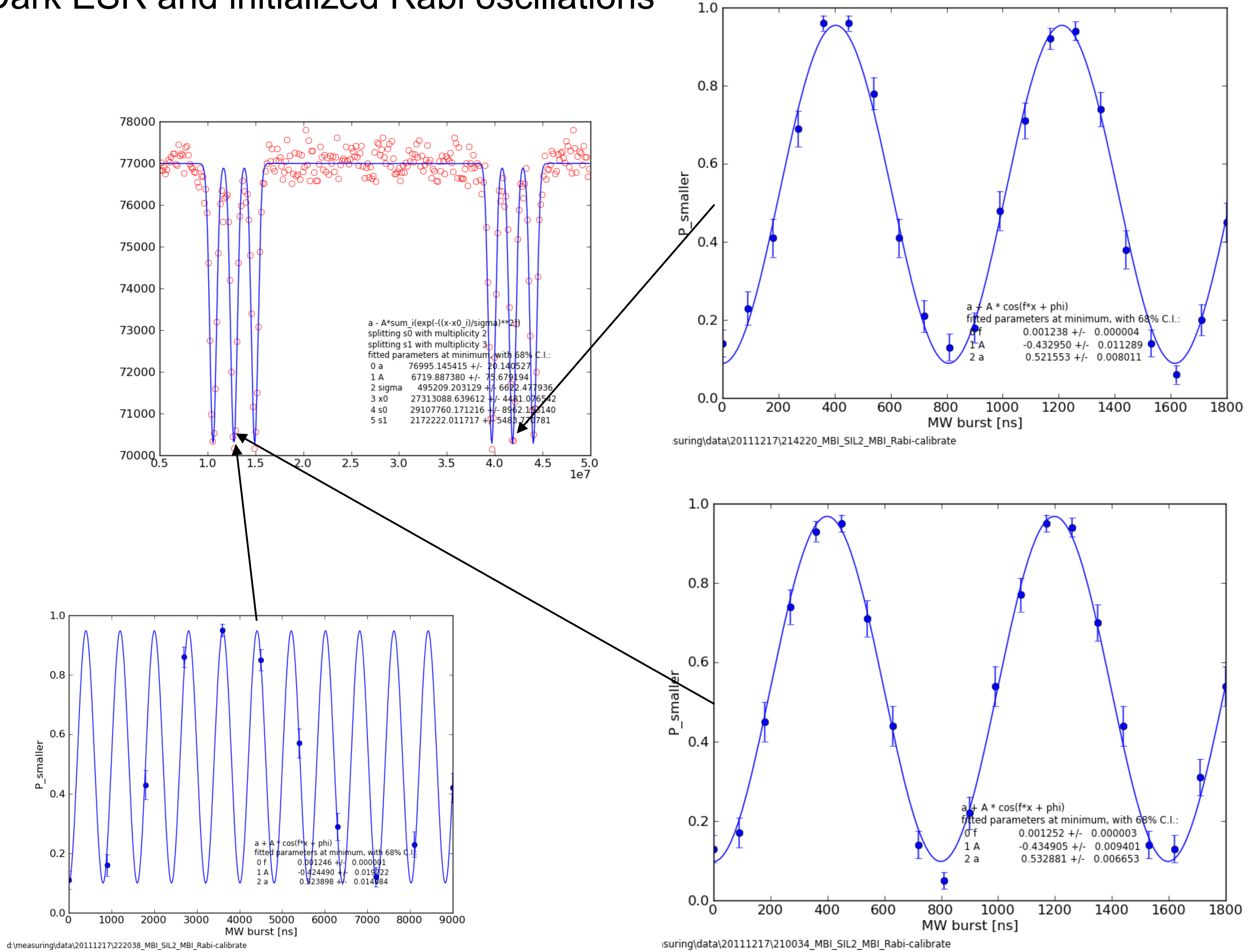






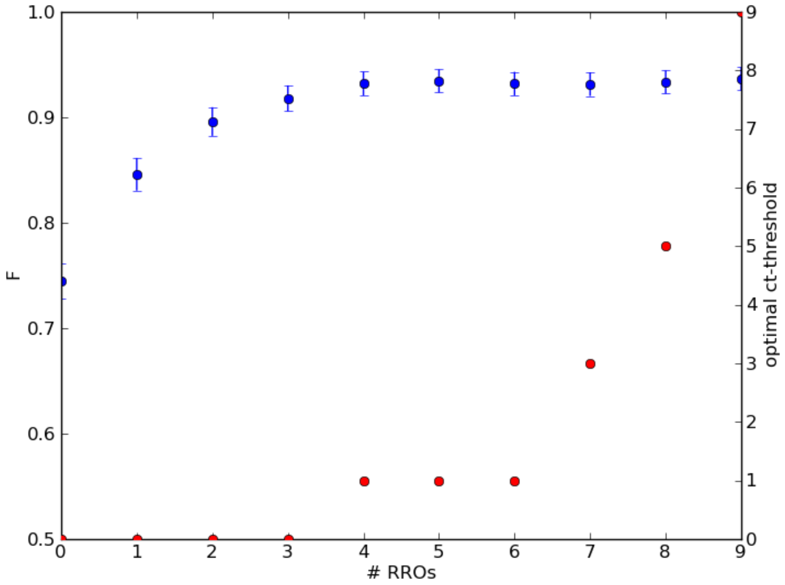
Here, we were reading out on Ey!

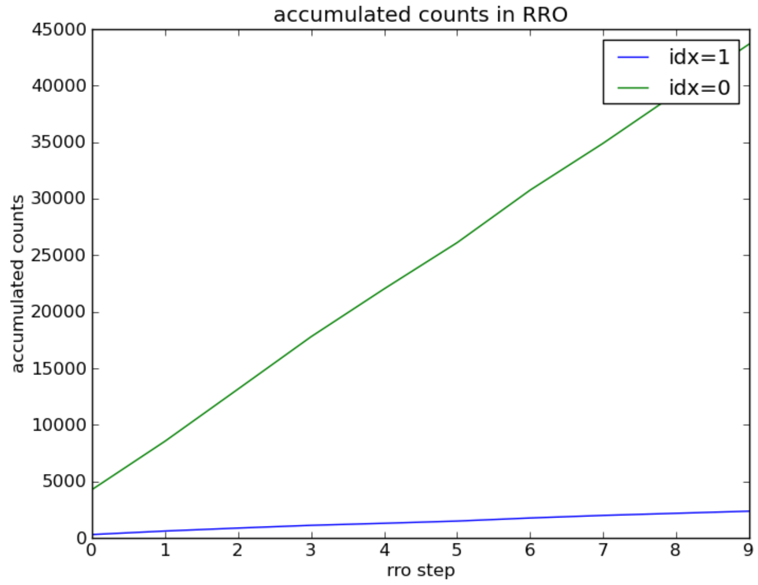
Dark ESR and initialized Rabi oscillations



Repetitive readout

Init in mI=0, 10x RO (sel. Pi, 20μs Ex, 6μs A2)



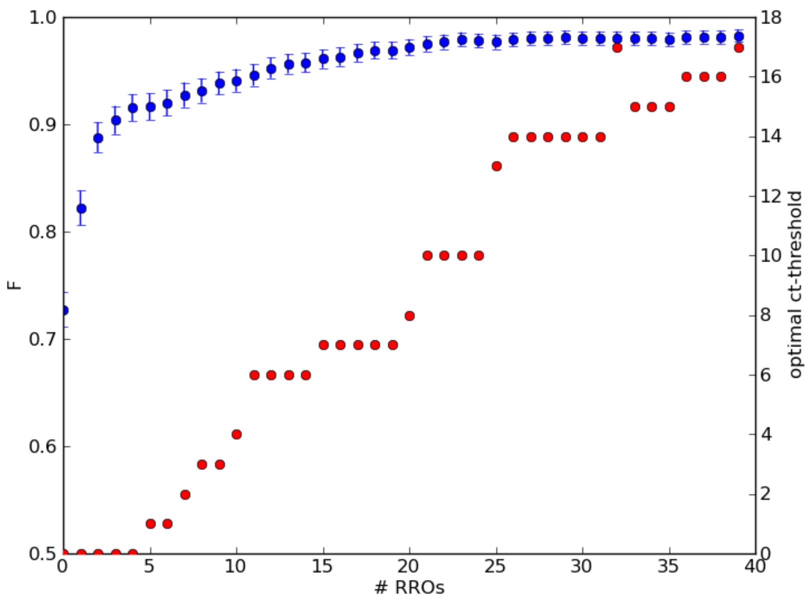


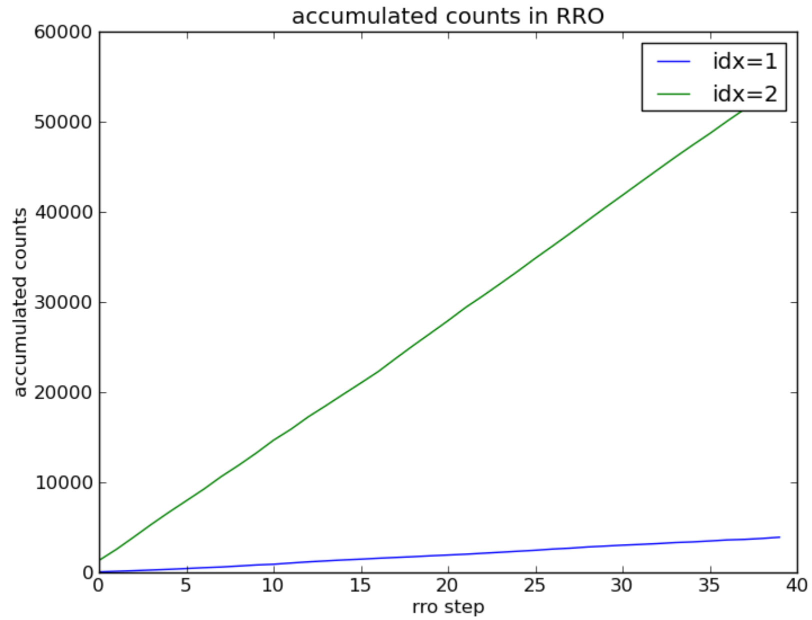
Fidelity 93.7%

Here, we were reading out on Ey! – seems to preferentially pump into mI=0

Repetitive readout

Init in mI=0, 40x RO (sel. Pi, 5μs Ex, 6μs A2)

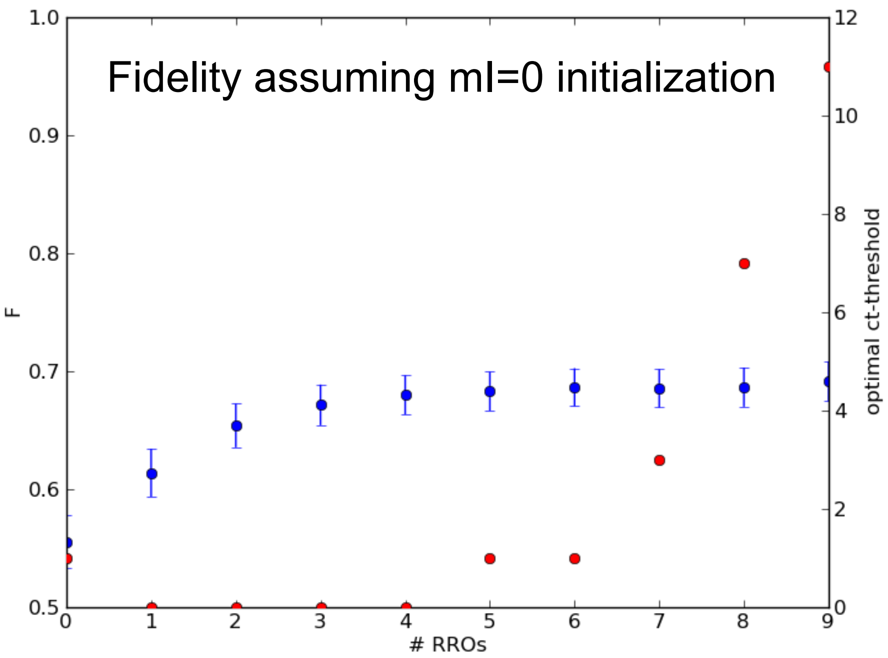


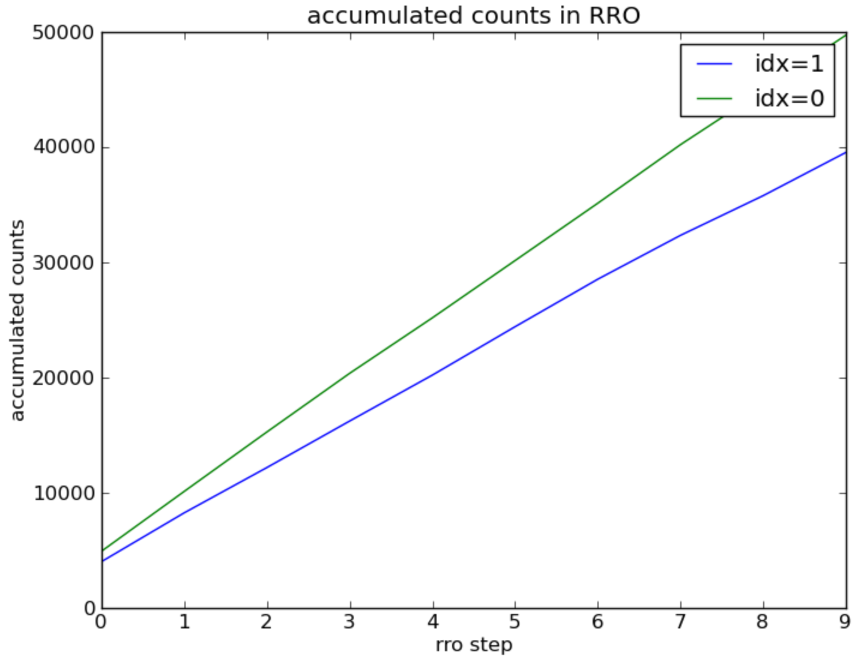


Fidelity 98.2%

Repetitive readout

No MBI (i.e. all MBI attempts with >=0 counts are valid)
10x RO (sel. Pi, 20μs Ex, 6μs A2)





Polarization in to mI=0 ?

notes:

• impossible to align to single-mode fiber: count rates are oscillating (mechanical vibrations?)

• resonant excitation:

looked for resonances in resonant counting mode (no laserscan due to B-field)

resonant counting with Matisse only, no MW:

3 resonances: 51.5 GHz (E‘)
54.4 GHz (Ey, resonance shows dip in center!)
56.6 GHz (Ex)

(AOMs cause a ~800 MHz difference in frequency reading between Newfocus and Matisse)

remaining resonances: resonant counting, Matisse locked on 54.4 GHz, stepping Newfocus

several resonances: 47.9 GHz E’ ?
50.7 GHz dip?
51.0 GHz peak?
52.9 GHz – Ey
55.8 GHz – Ex
57.1 GHz – A1
60.3 GHz – A2

resonant counting with one laser, no MW: Ex resonance has dip at center

17/12/11

notes:

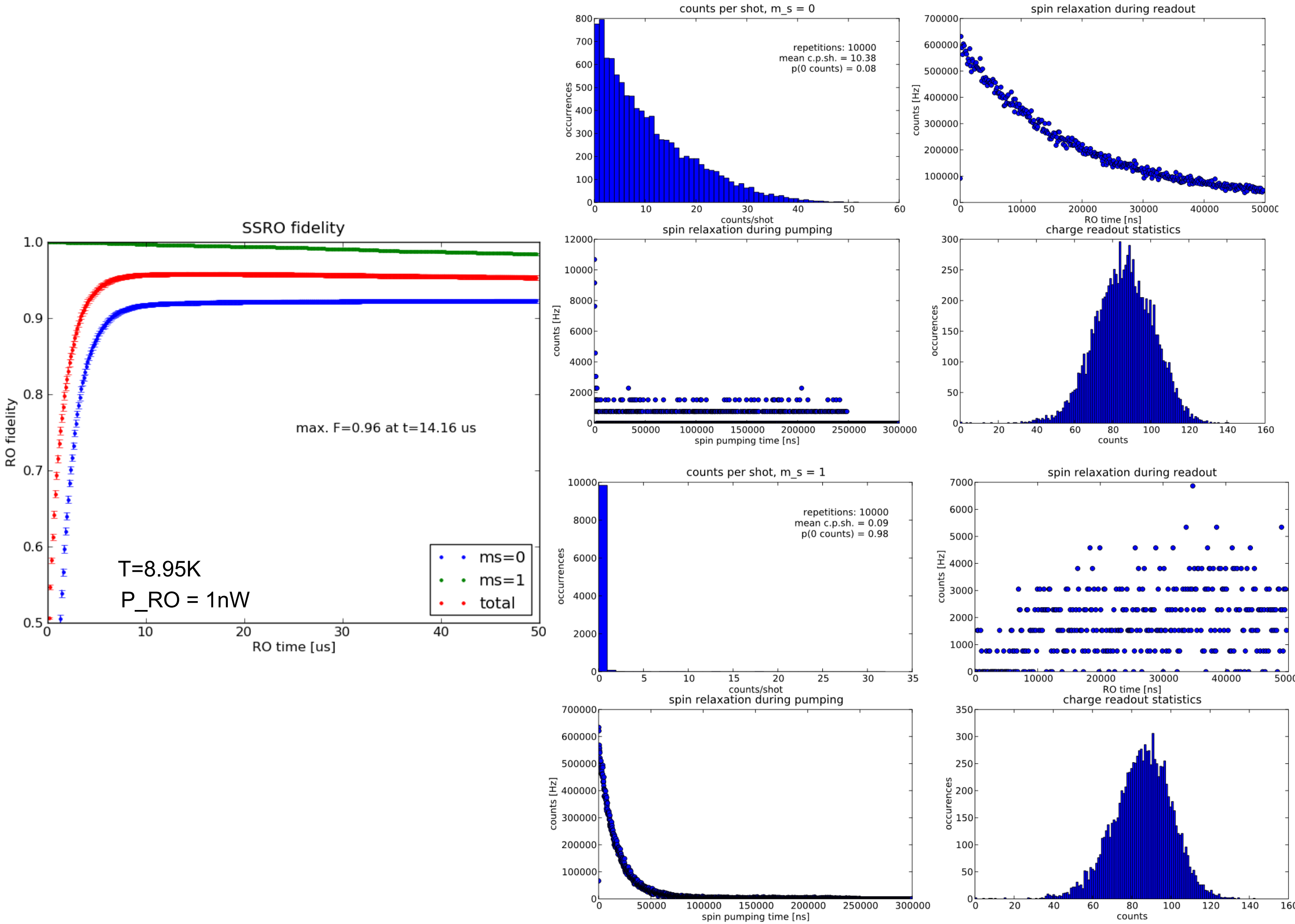
• resonant excitation (T=10.5K):
  looked for resonances in resonant counting mode (no laserscan due to B-field)
  resonant counting with Matisse only, no MW:
    51.6 GHz
    54.5 GHz (Ey)
    56.5 GHz (Ex)
  (AOMs cause a ~800 MHz difference in frequency reading between Newfocus and Matisse)
  remaining resonances: resonant counting, Matisse locked on 54.4 GHz, stepping Newfocus
  several resonances:
    47.6 GHz E' ?
    50.8 GHz
    52.9 GHz – Ey
    55.8 GHz – Ex
    57.1 GHz – A1
    60.1 GHz – A2

  resonant counting with one laser, no MW: Ex resonance has dip at center

• initialized Rabi: >=95% contrast at 1.1 MHz Rabi frequency; at 0.14 MHz only ~50% (short T2*?)

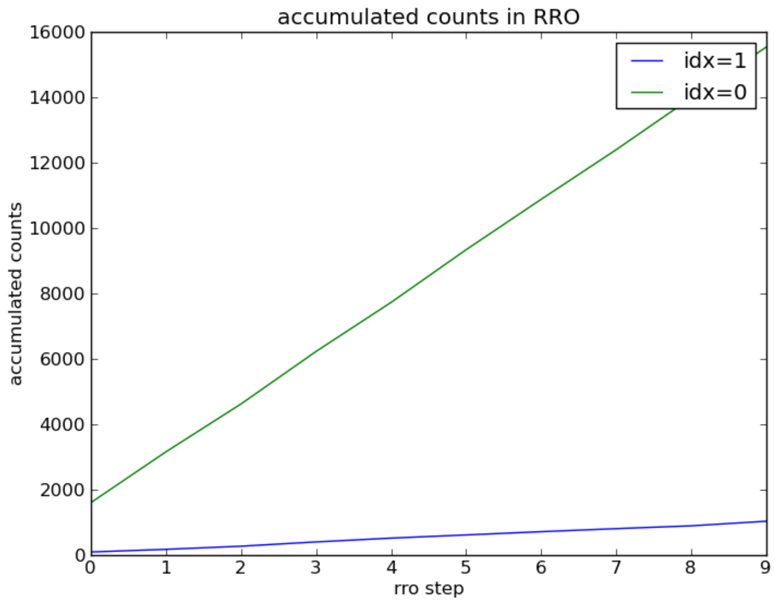
• MBI ~5..10x slower for mI=+-1 than for mI=0

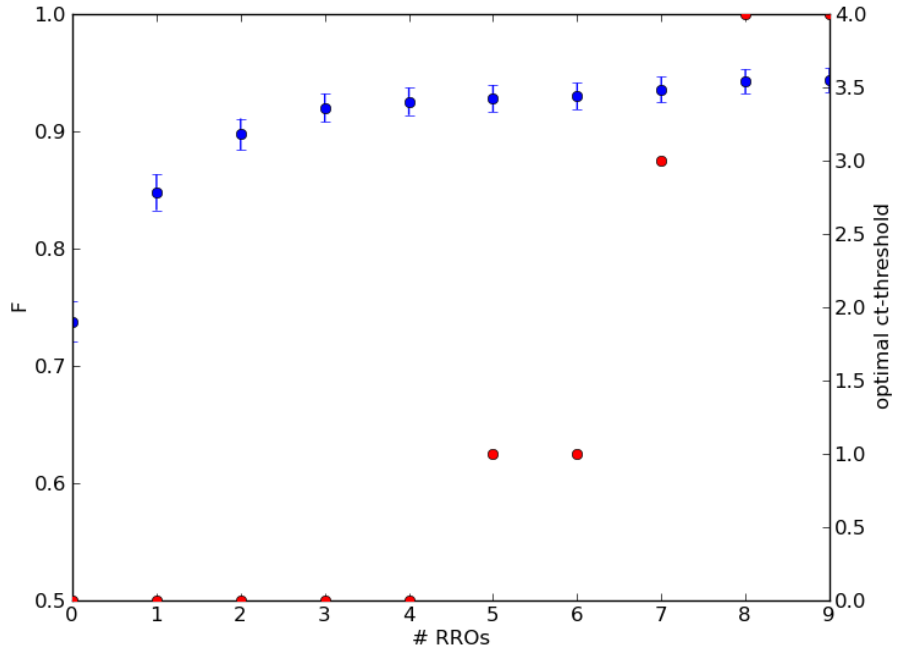
SIL2 – electron spin readout: now on Ex



Repetitive readout

Init in mI=0, 10x RO (sel. Pi, 10μs Ex, 6μs A2)

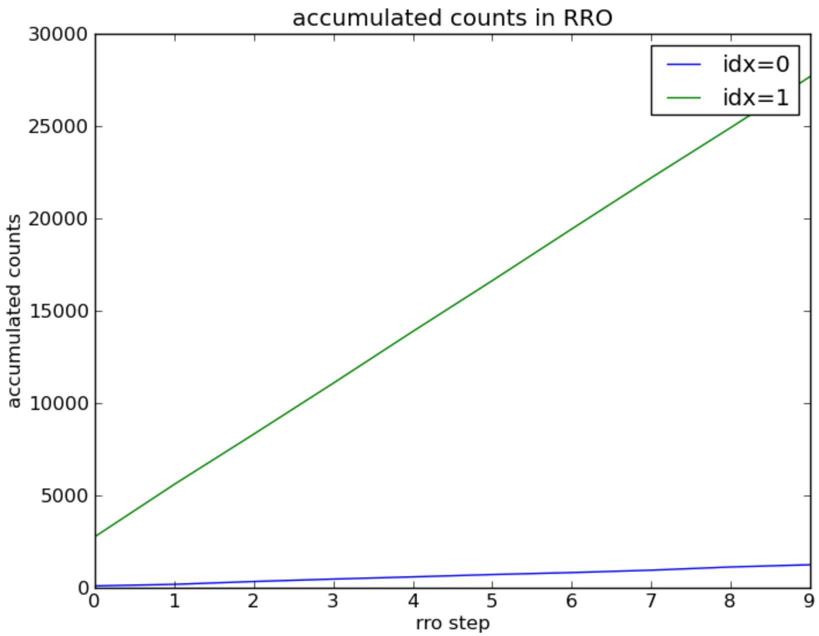


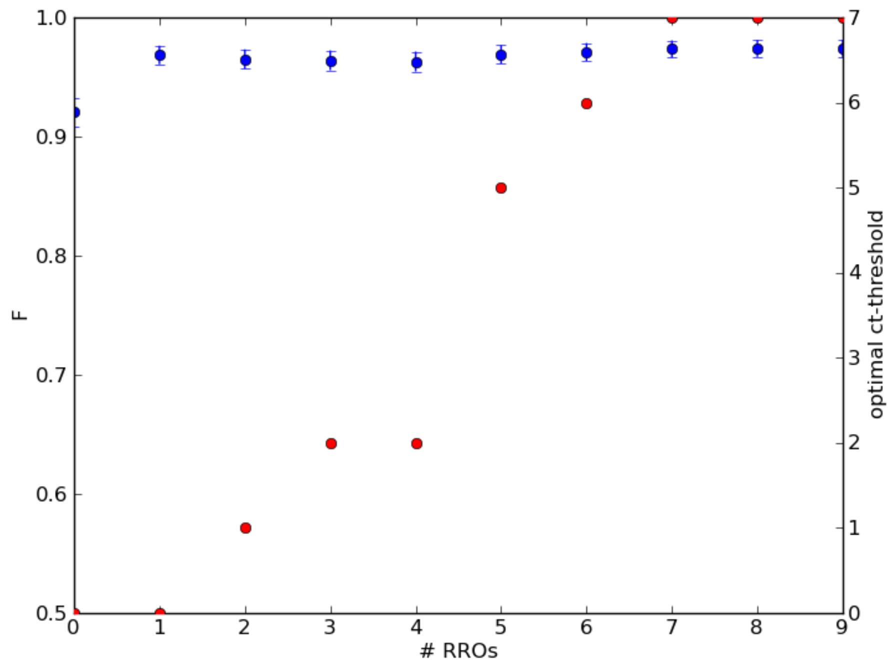


Fidelity 94.4%

Repetitive readout

Init in mI=-1, 10x RO (sel. Pi, 10μs Ex, 6μs A2)

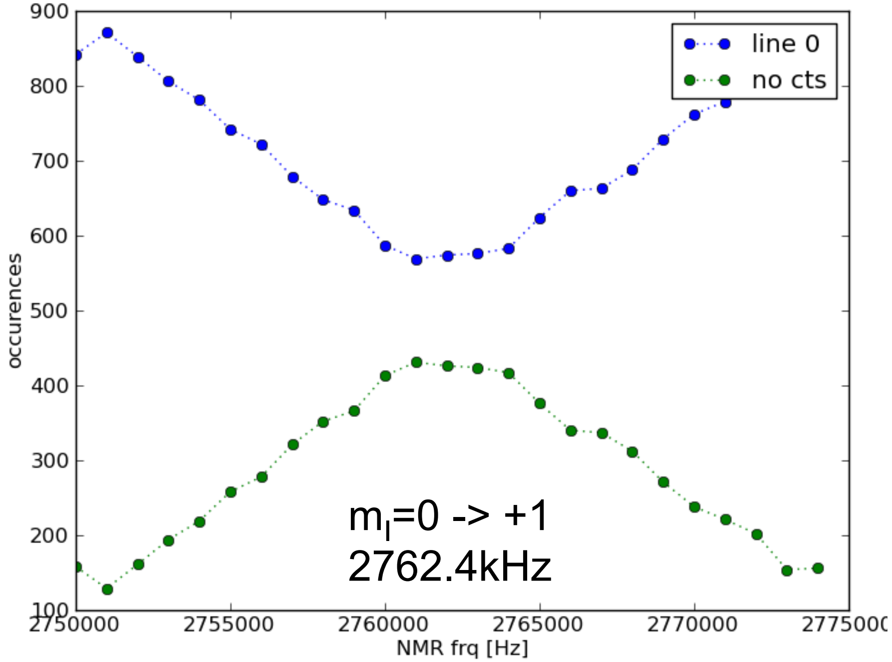


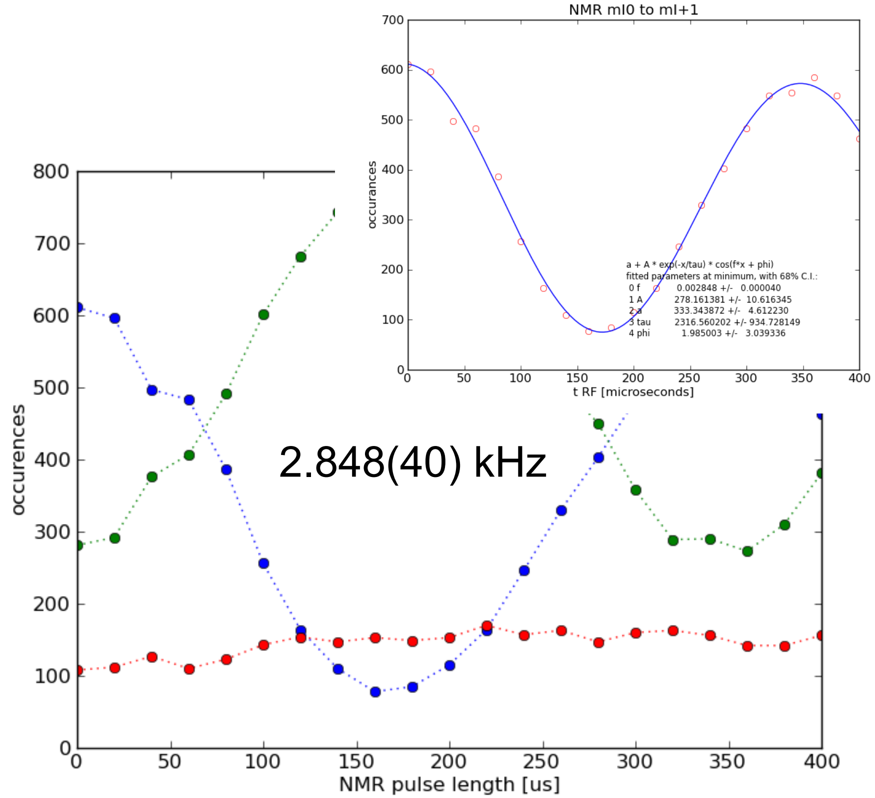


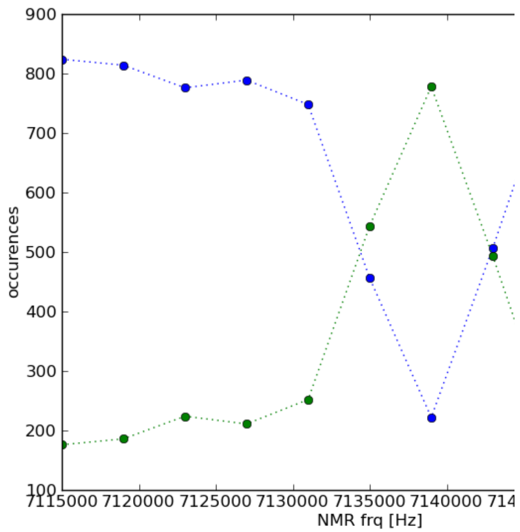
Fidelity 97.4%

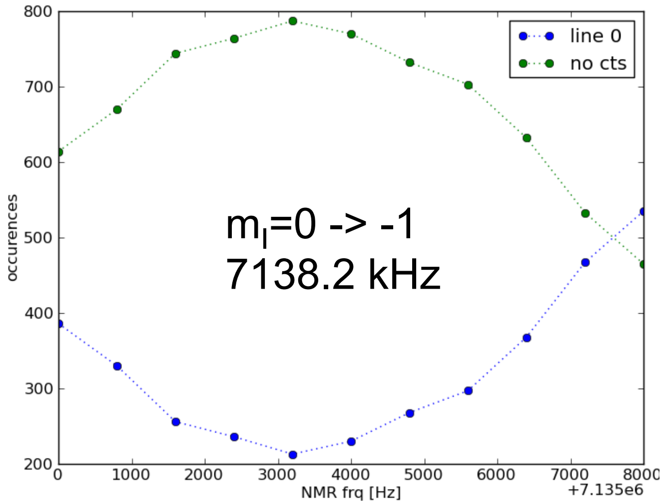
NMR

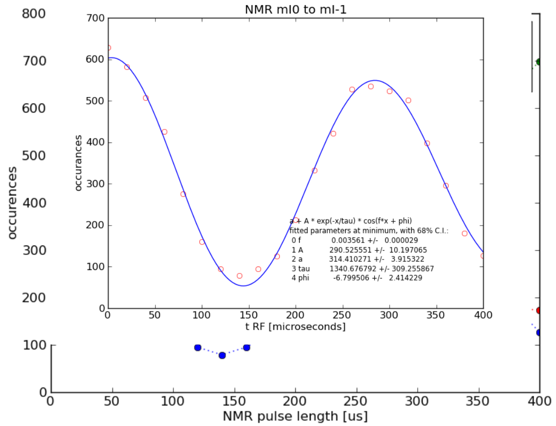
Initialized into ms=-1,mI=0





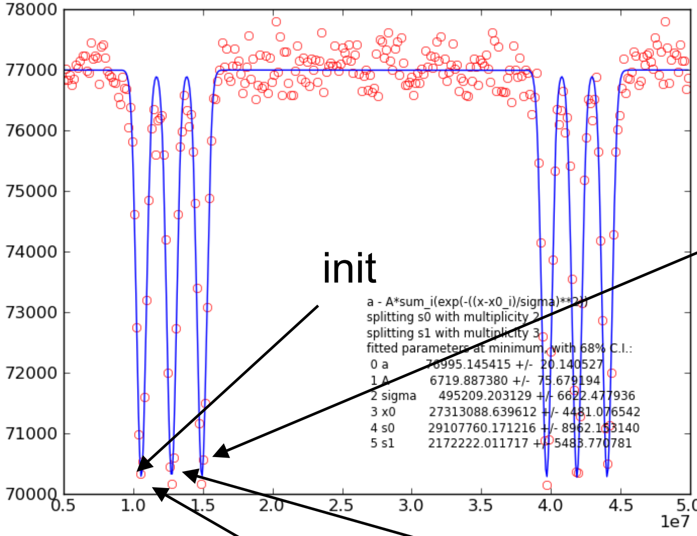


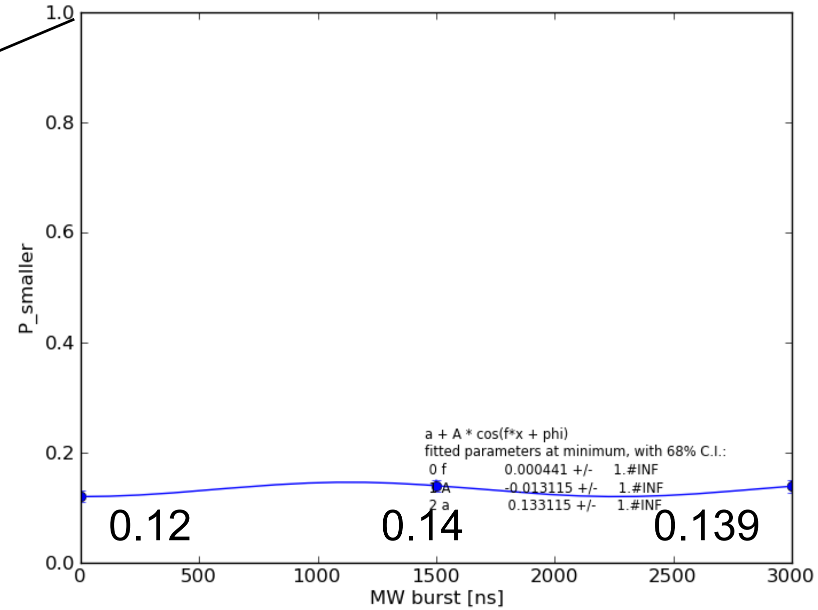




3.561(29) kHz

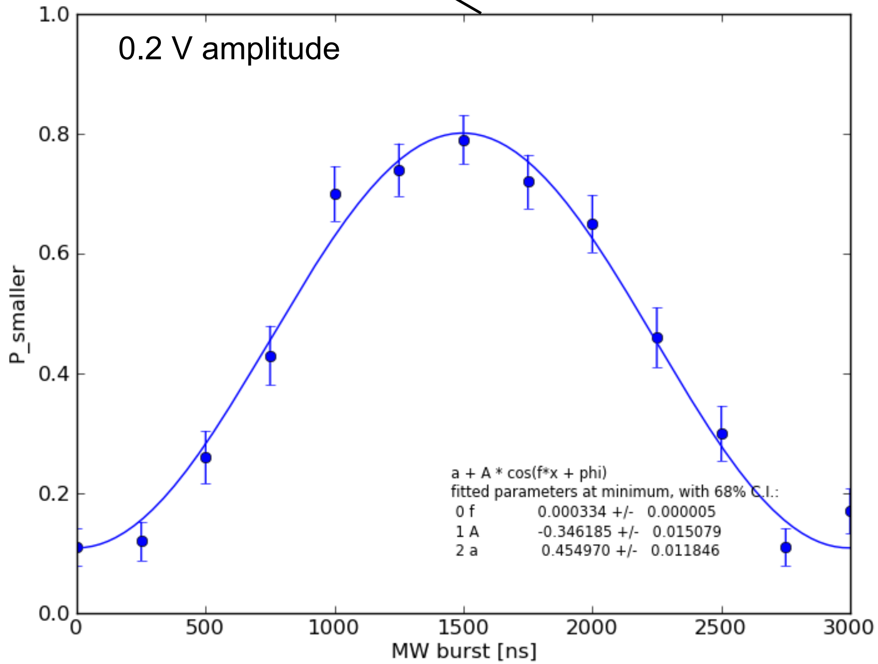
Slow MW pi pulse calibration



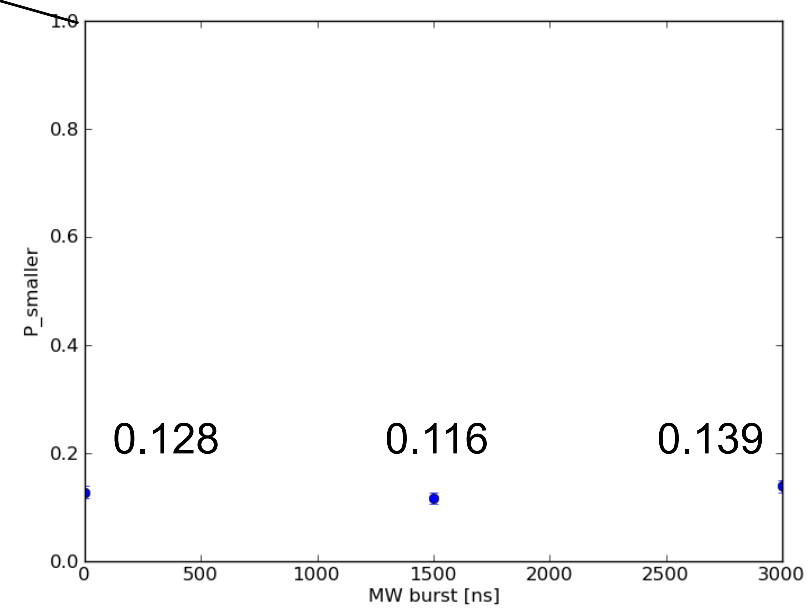


d:\measuring\data\20111219\181939_MBI_SIL2_MBI_Rabi

drive

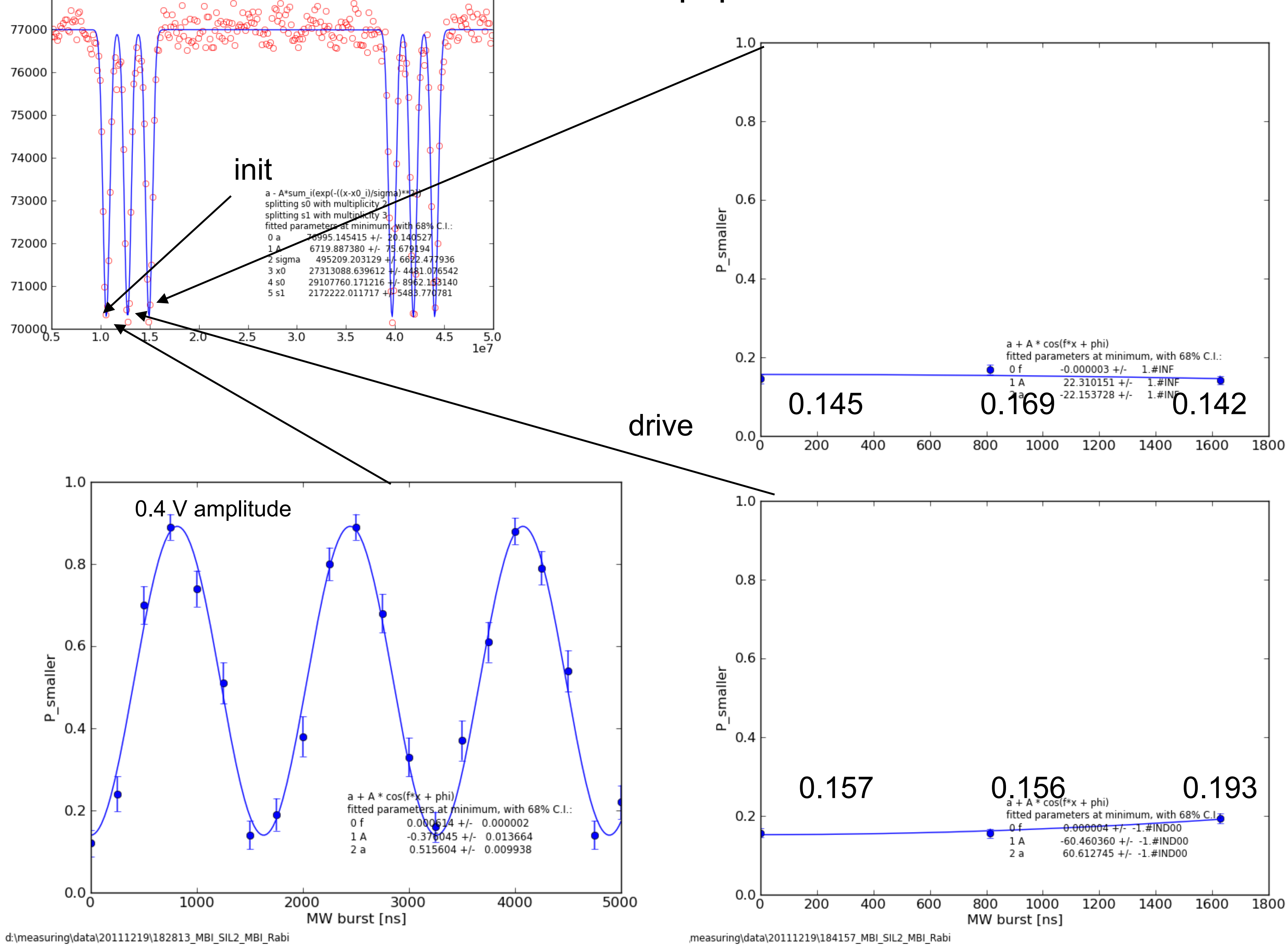


d:\measuring\data\20111219\175035_MBI_SIL2_MBI_Rabi

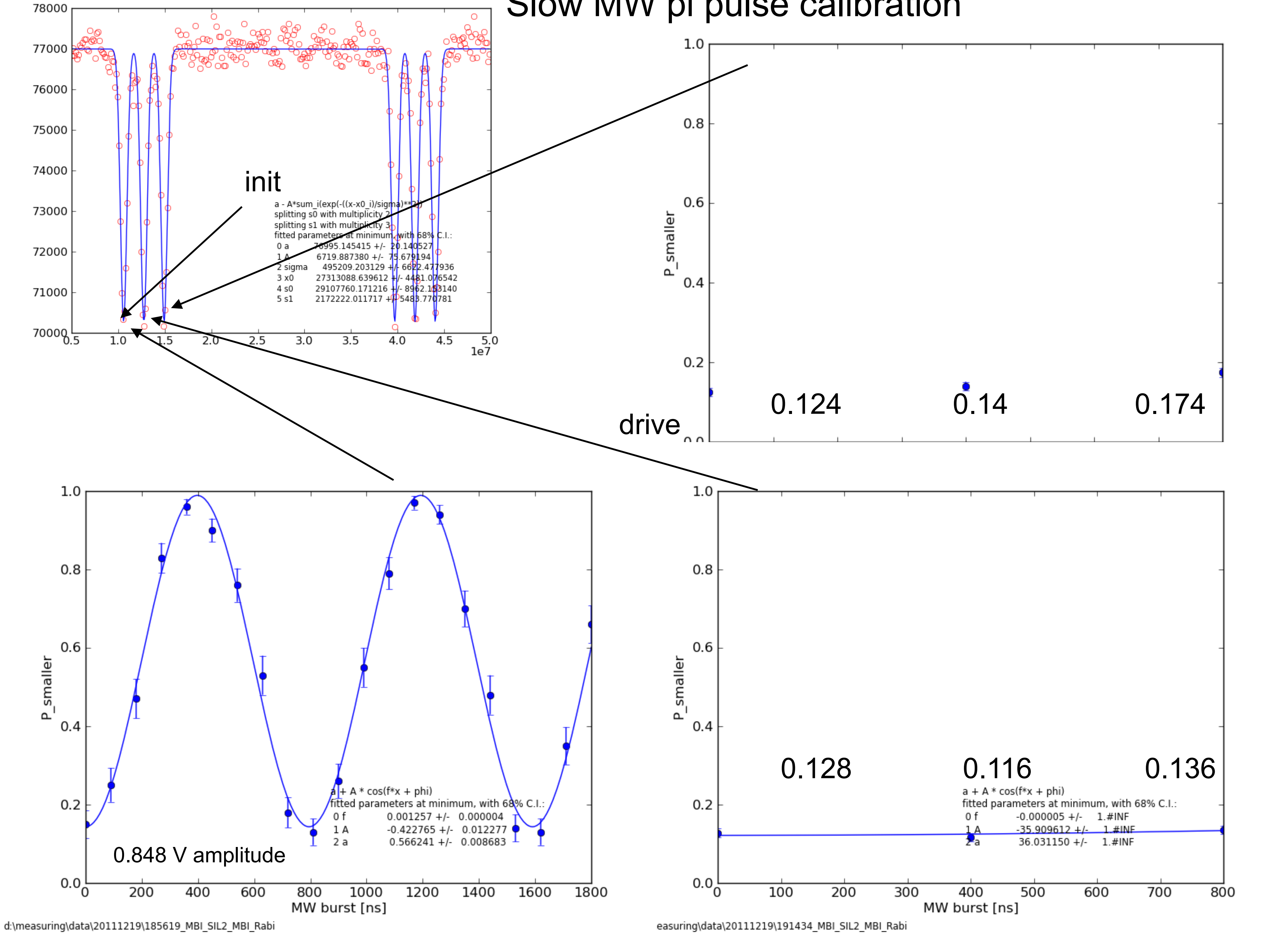


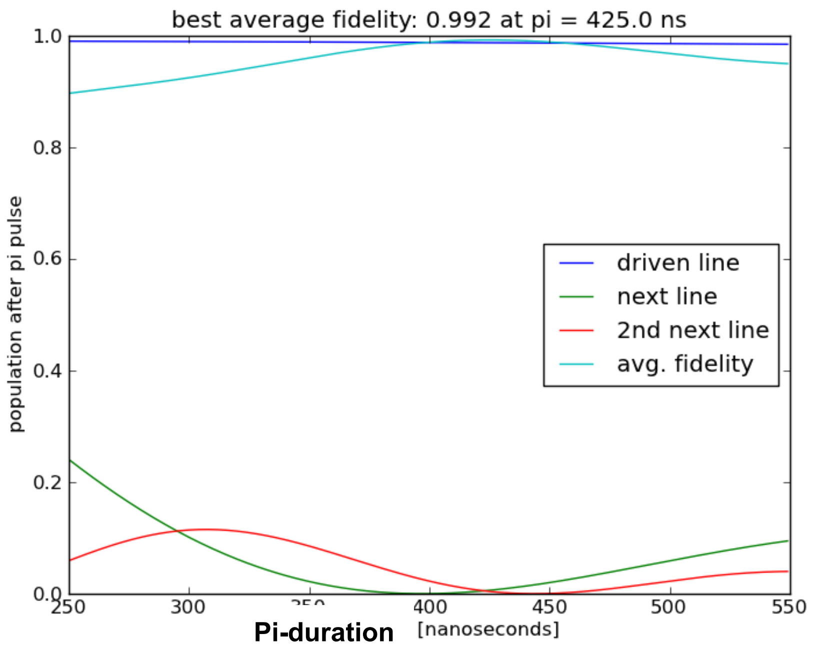
d:\measuring\data\20111219\181241_MBI_SIL2_MBI_Rabi

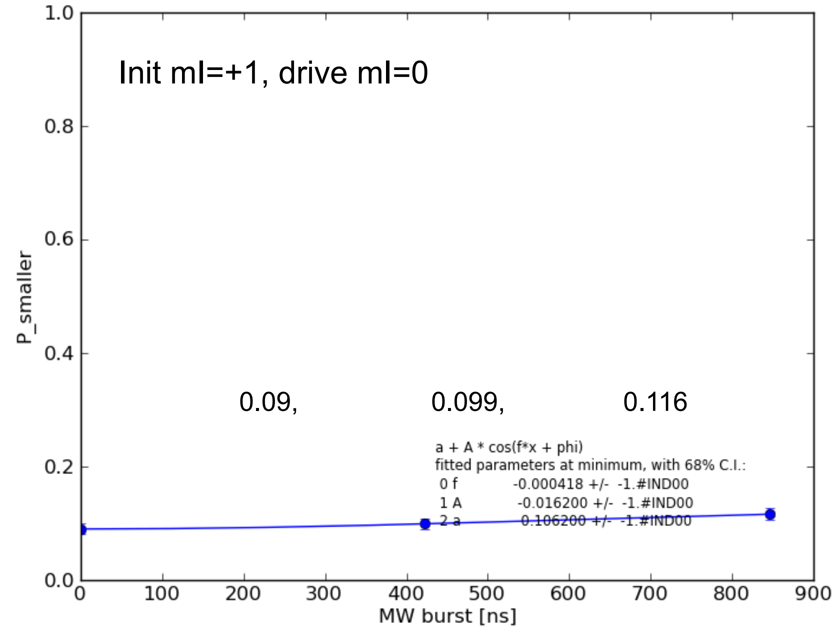
Slow MW pi pulse calibration



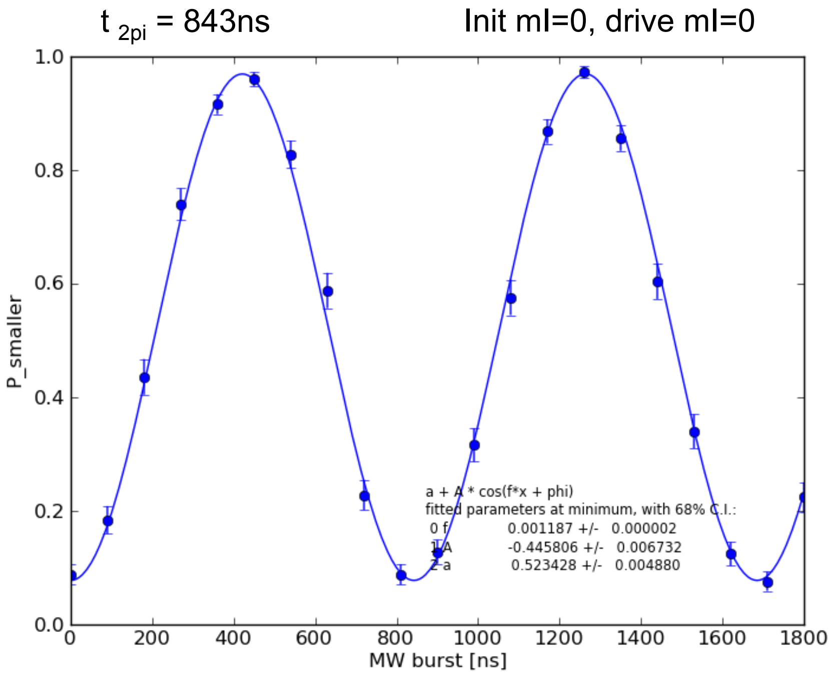
Slow MW pi pulse calibration



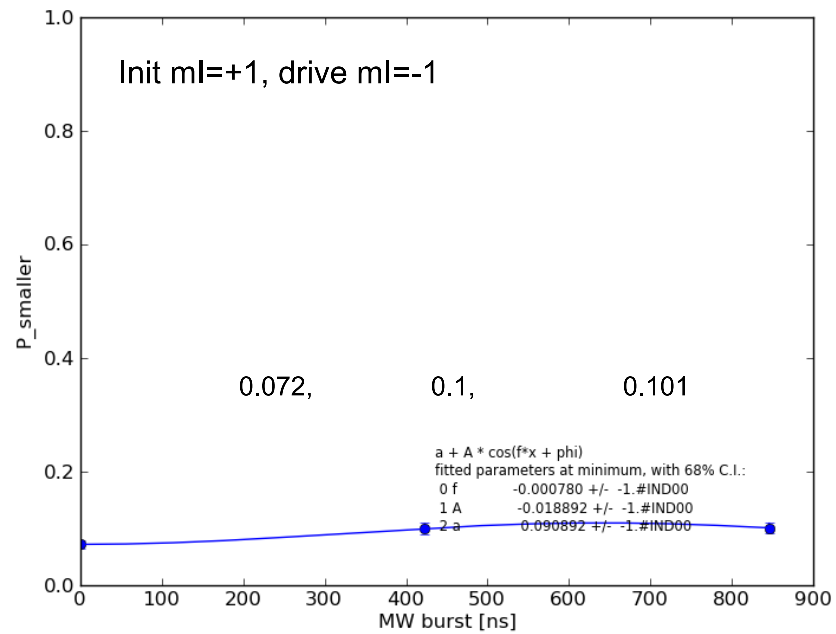




d:\measuring\data\20111219\204427_MBI_SIL2_MBI_Rabi

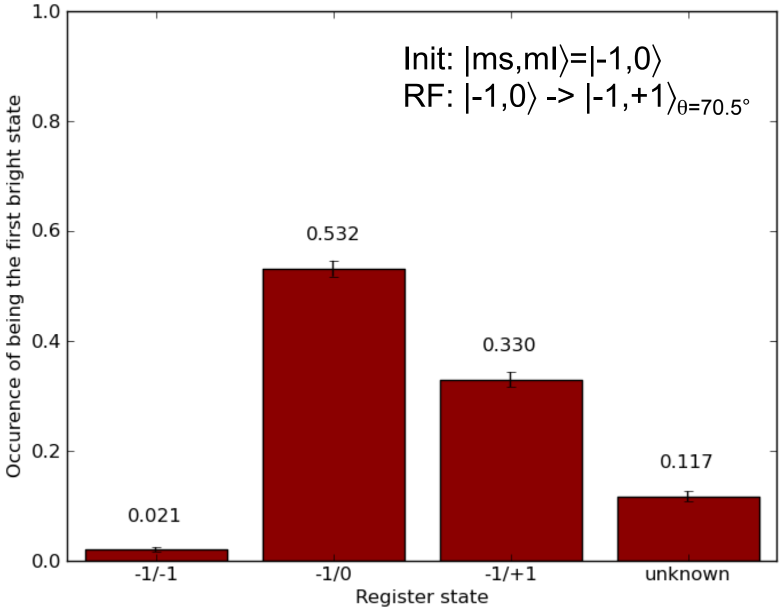


d:\measuring\data\20111219\201407_MBI_SIL2_MBI_Rabi

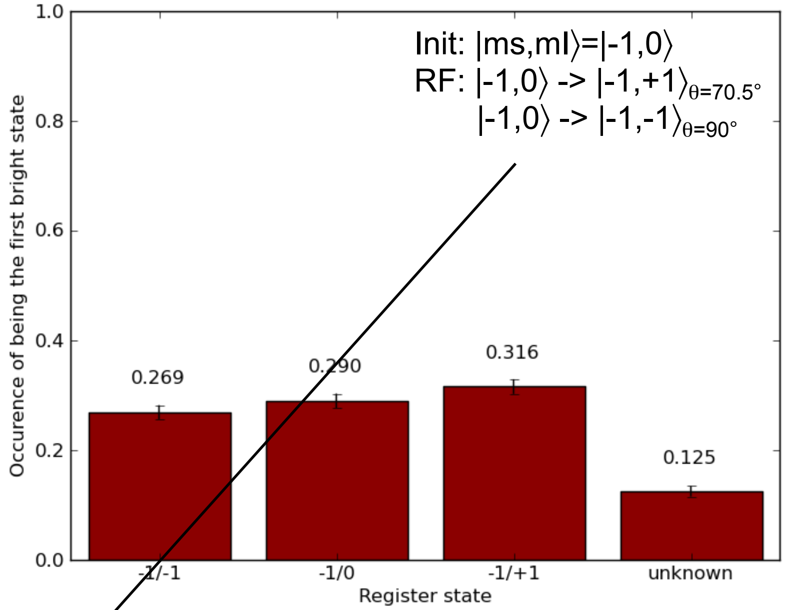


d:\measuring\data\20111219\203425_MBI_SIL2_MBI_Rabi

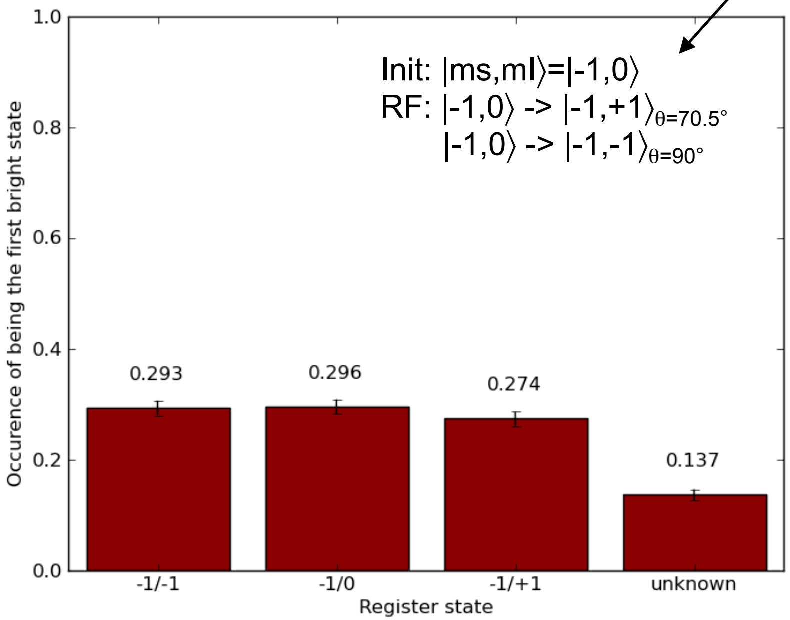
NMR: create state |I⟩



d:\measuring\data\20111219\233622_MBI_SIL2_MBI_NMR_prep_I_SRO



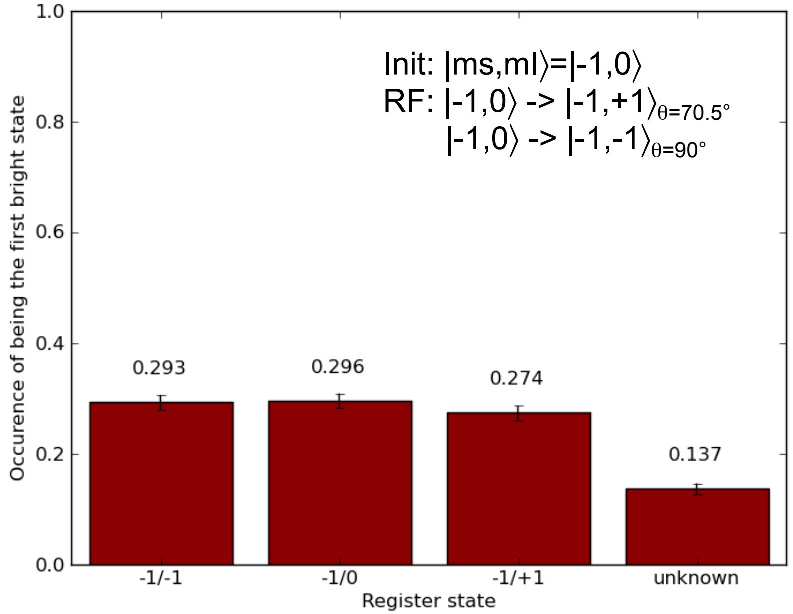
d:\measuring\data\20111219\234522_MBI_SIL2_MBI_NMR_prep_I_SRO



d:\measuring\data\20111219\235737_MBI_SIL2_MBI_NMR_prep_I_SRO

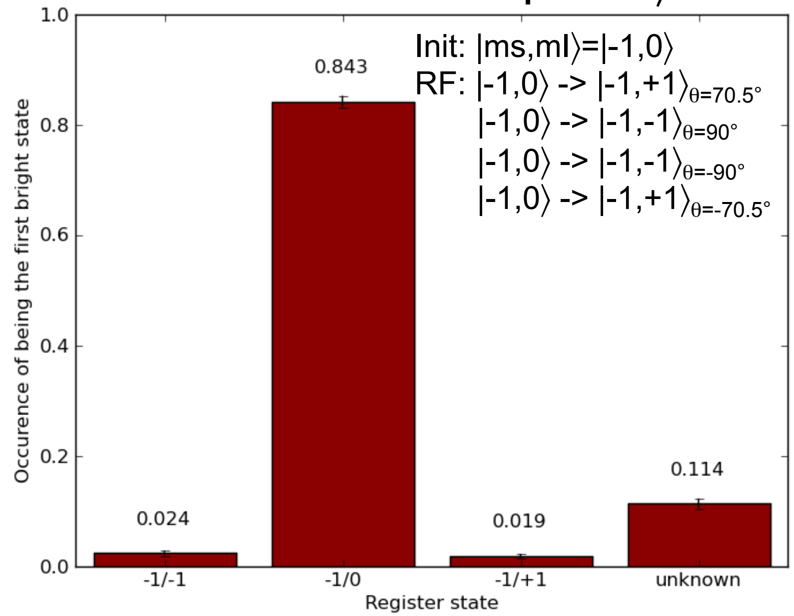
NMR1 duration / 1.06
NMR2 duration * 1.03

create state $|I\rangle$



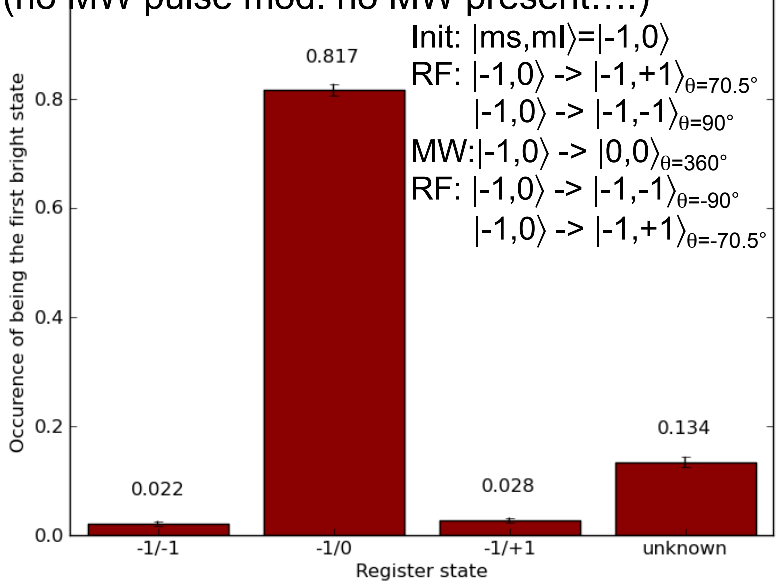
d:\measuring\data\20111219\235737_MBI_SIL2_MBI_NMR_prep_I_SRO

transform back in $|mI=0\rangle$



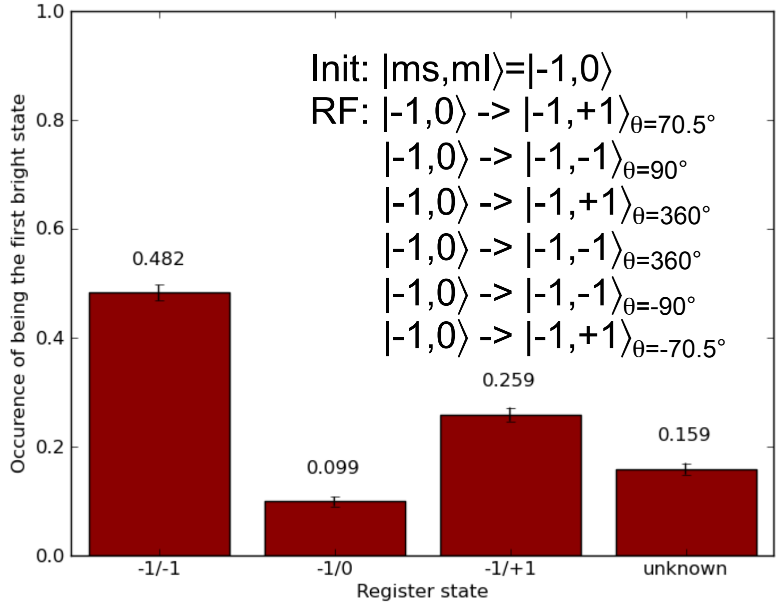
d:\measuring\data\20111220\000721_MBI_SIL2_MBI_NMR_prep_I_SRO

Flip phase of $|mI=0\rangle$ for readout in $|F\rangle$ (didn't work) (no MW pulse mod: no MW present….)



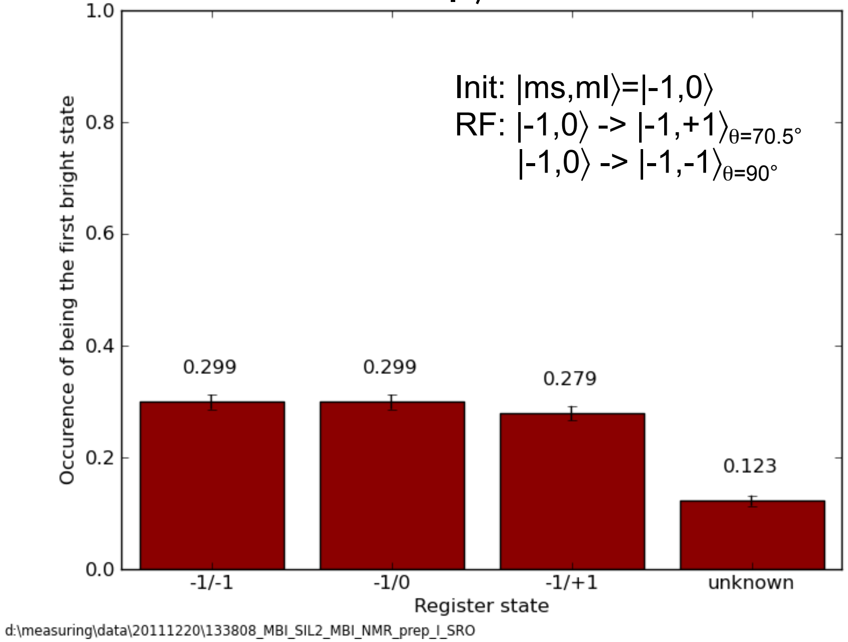
d:\measuring\data\20111220\002554_MBI_SIL2_MBI_NMR_prep_I_SRO

Flip phase of $|mI=0\rangle$ for readout in $|F\rangle$ (seems to work)

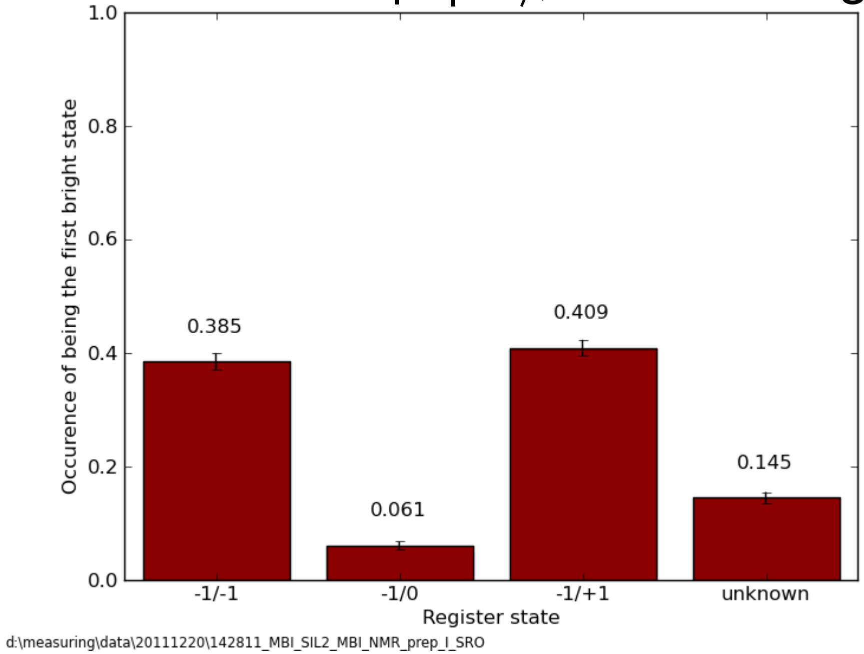


d:\measuring\data\20111220\003947_MBI_SIL2_MBI_NMR_prep_I_SRO

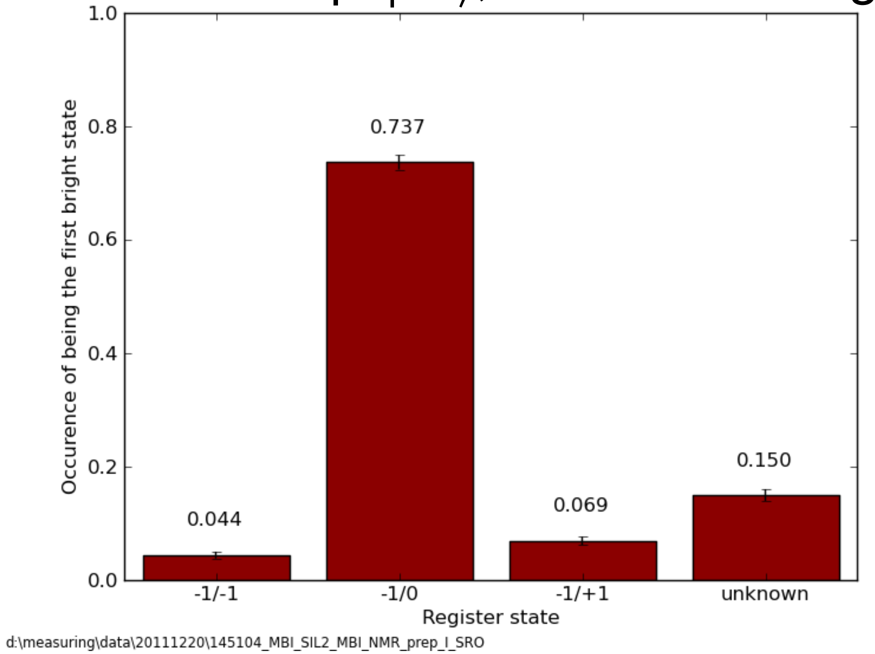
create state $|I\rangle$

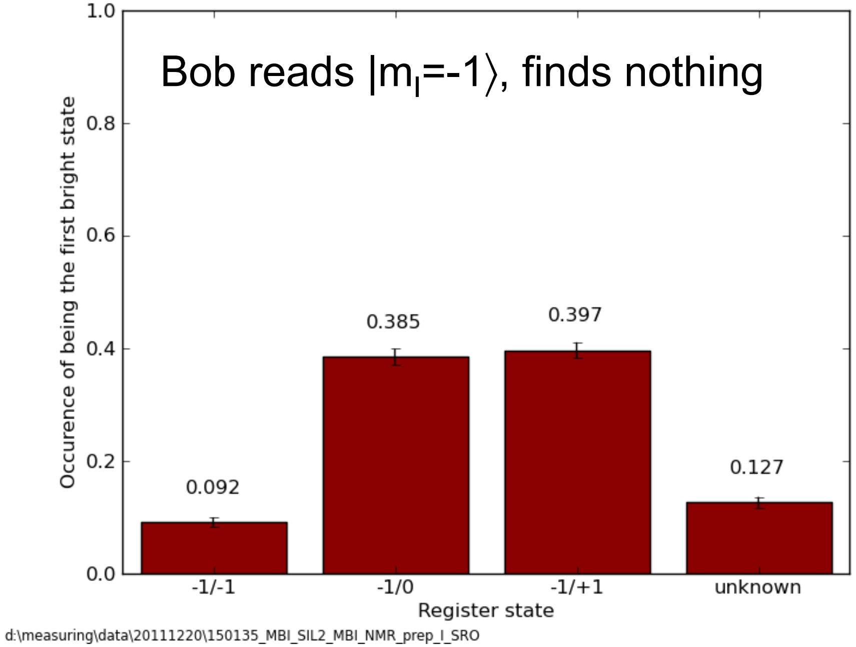


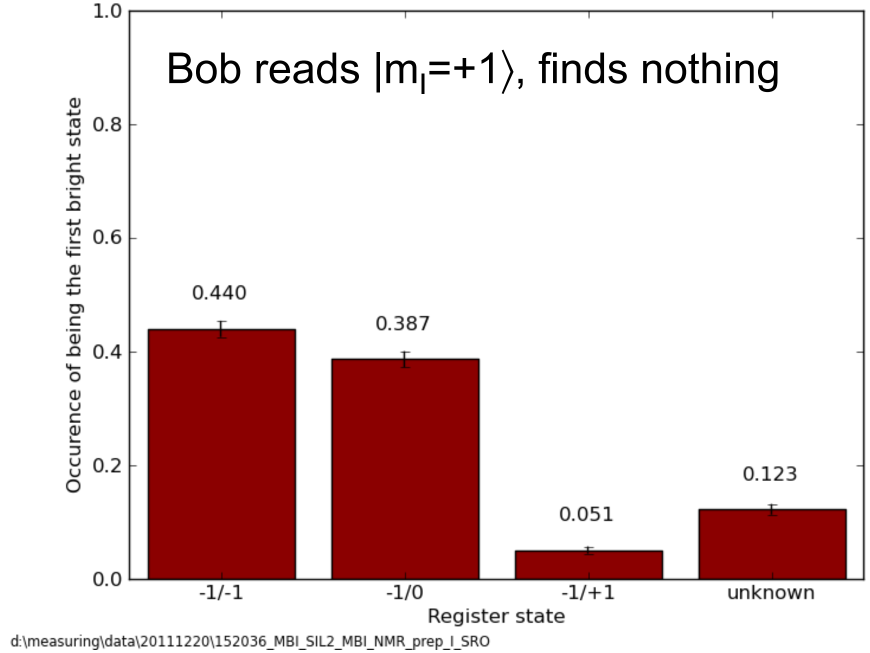
Bob reads $|m_I=0\rangle$, finds nothing

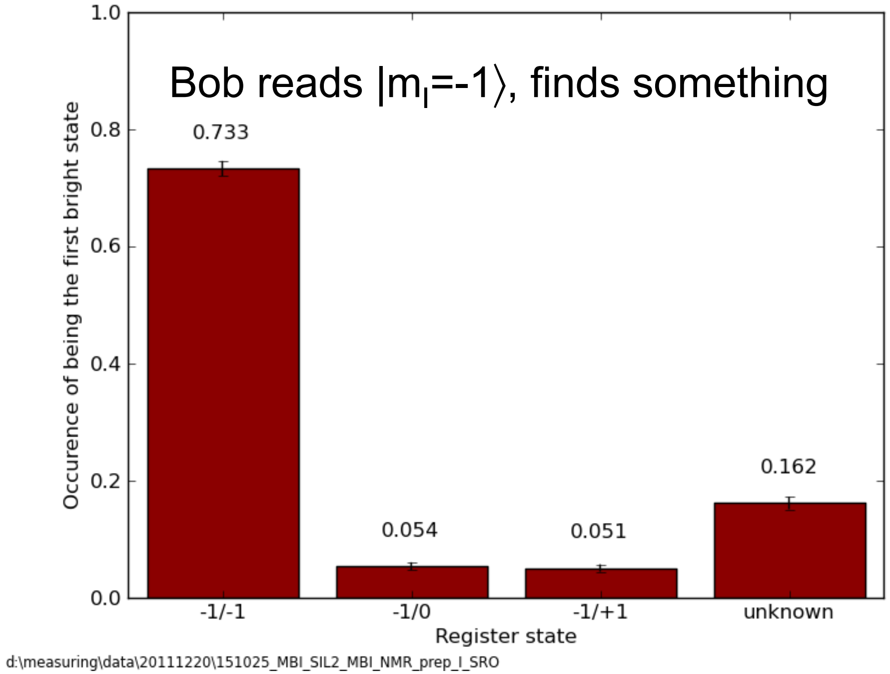


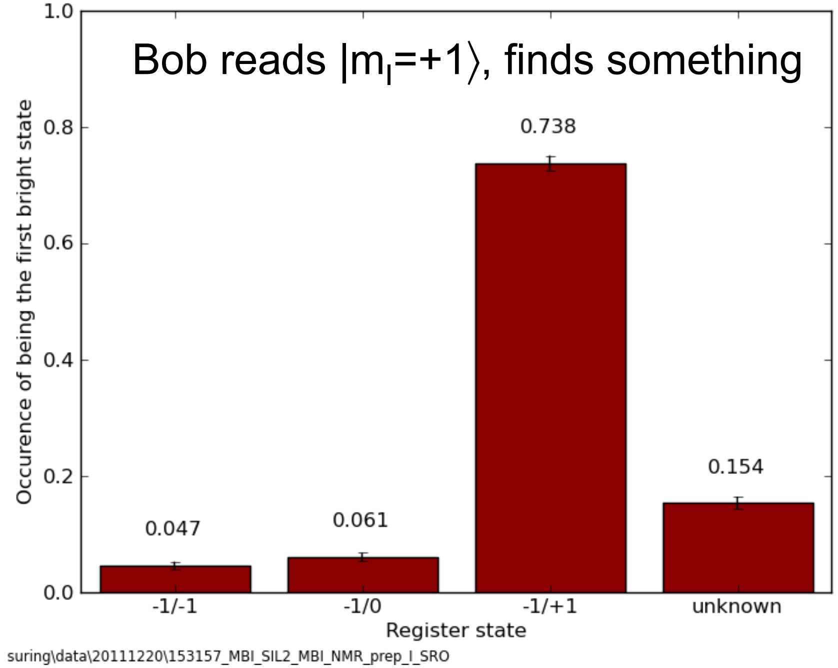
Bob reads $|m_I=0\rangle$, finds something

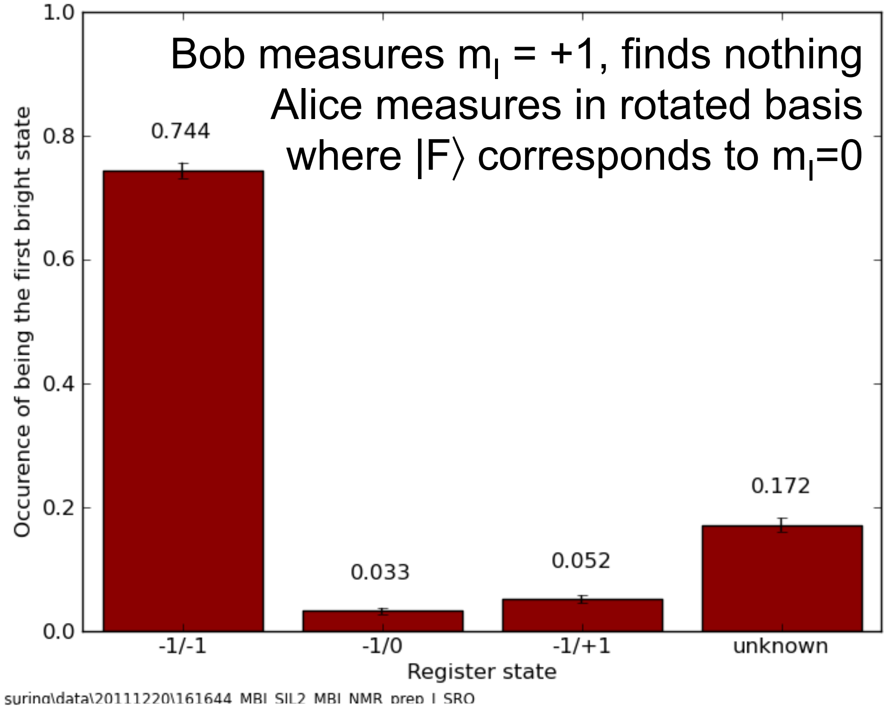


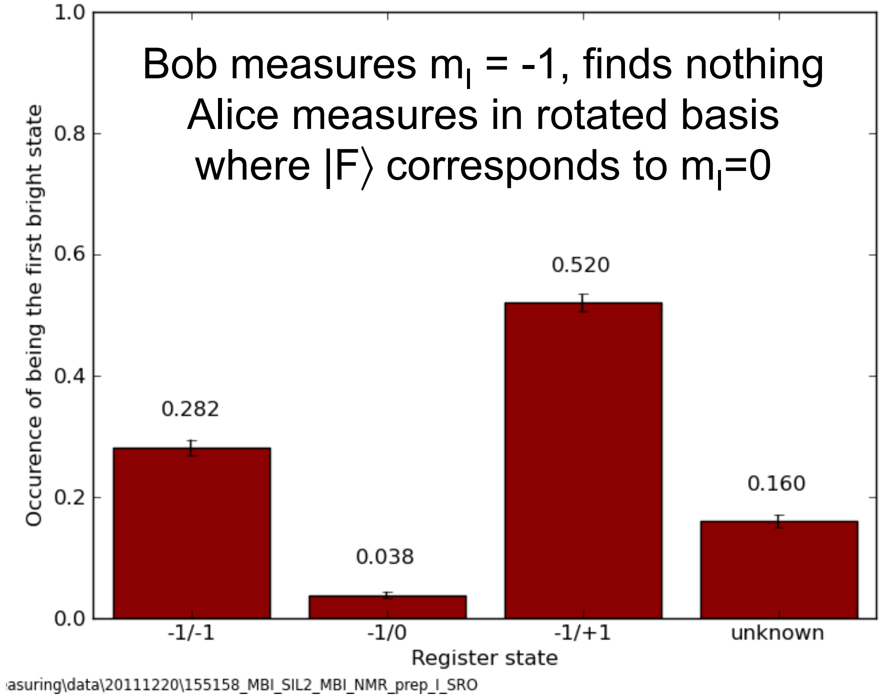

Bob reads |m_I=-1⟩, finds nothing
Occurence of being the first bright state
Register state
-1/-1
-1/0
-1/+1
unknown
0.092
0.385
0.397
0.127
0.0
0.2
0.4
0.6
0.8
1.0
d:\measuring\data\20111220\150135_MBI_SIL2_MBI_NMR_prep_I_SRO

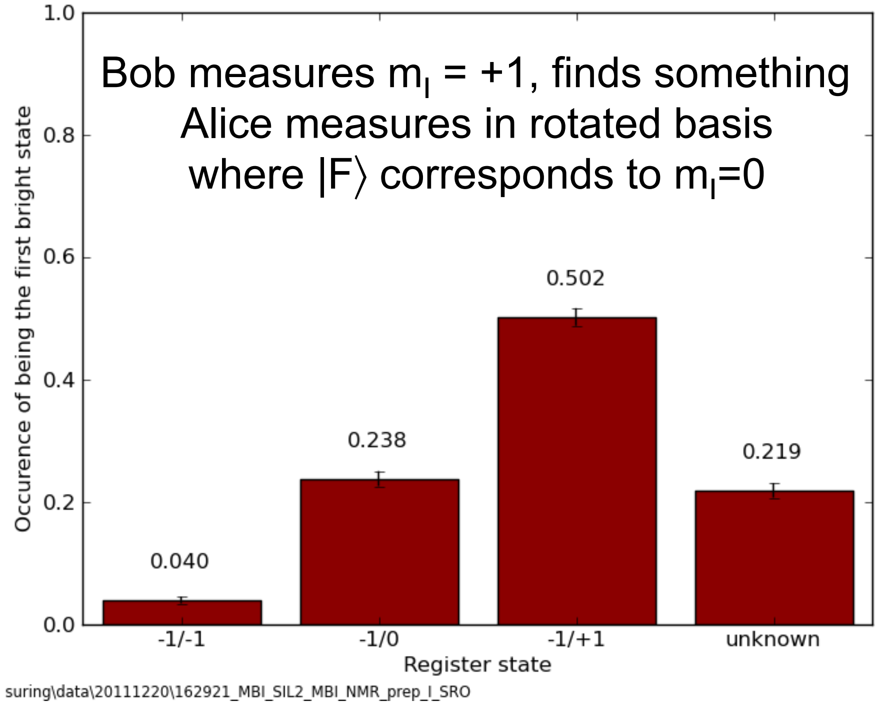

Bob reads |m_I=+1⟩, finds nothing
Occurence of being the first bright state
Register state
-1/-1
-1/0
-1/+1
unknown
0.440
0.387
0.051
0.123
0.0
0.2
0.4
0.6
0.8
1.0
d:\measuring\data\20111220\152036_MBI_SIL2_MBI_NMR_prep_I_SRO

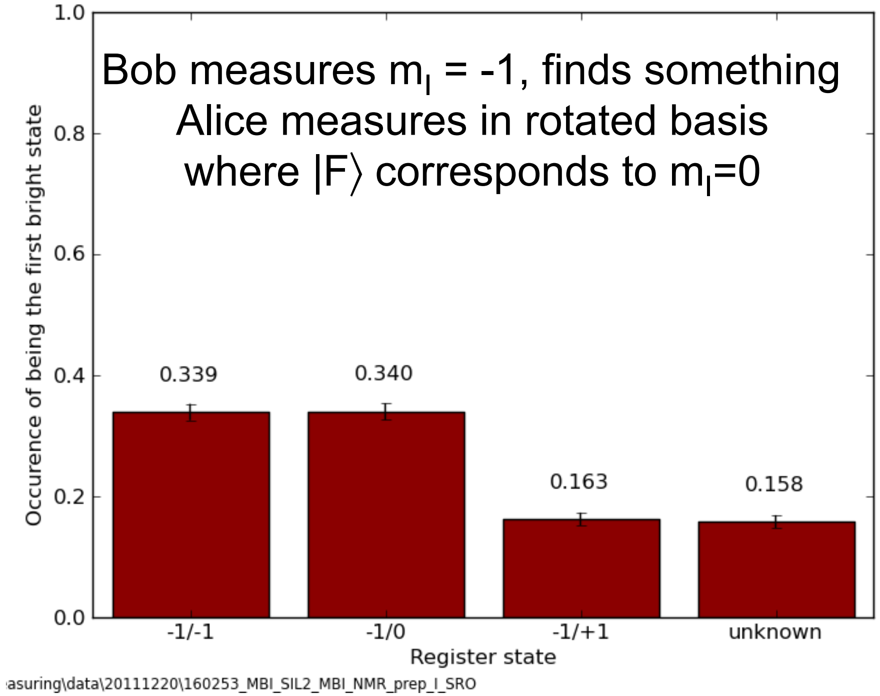

Bob reads |m_I=-1⟩, finds something
Occurence of being the first bright state
Register state
-1/-1
-1/0
-1/+1
unknown
0.733
0.054
0.051
0.162
0.0
0.2
0.4
0.6
0.8
1.0
d:\measuring\data\20111220\151025_MBI_SIL2_MBI_NMR_prep_I_SRO

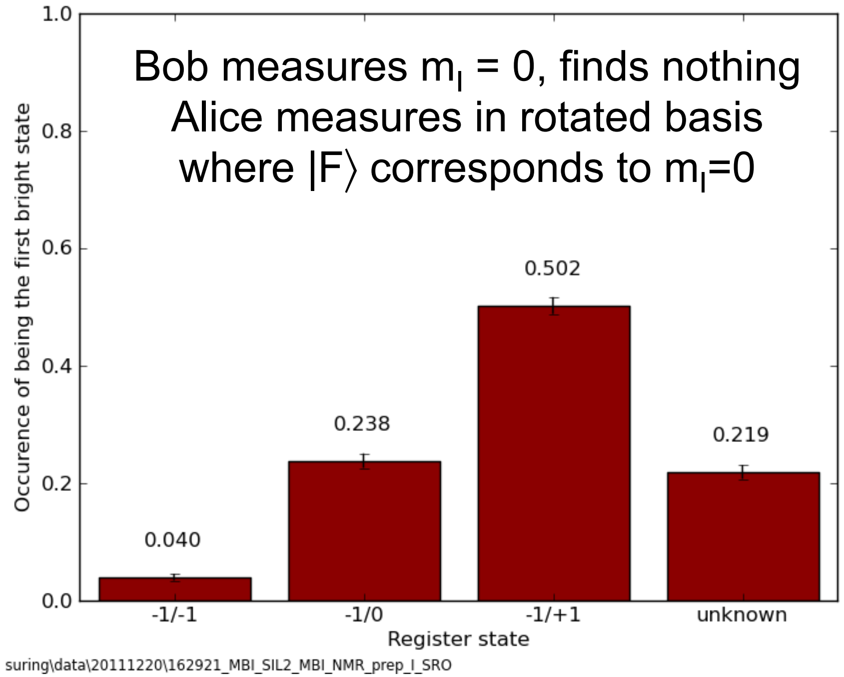

Bob reads |m_I=+1⟩, finds something
Occurence of being the first bright state
Register state
-1/-1
-1/0
-1/+1
unknown
0.047
0.061
0.738
0.154
0.0
0.2
0.4
0.6
0.8
1.0
suring\data\20111220\153157_MBI_SIL2_MBI_NMR_prep_I_SRO

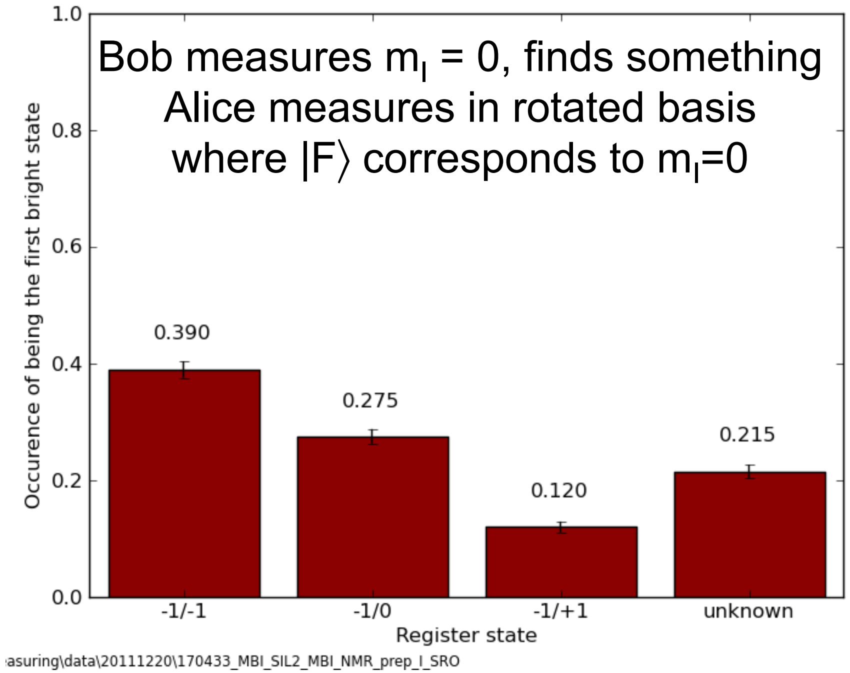

Bob measures m_I = +1, finds nothing
Alice measures in rotated basis
where $|F\rangle$ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.744
0.033
0.052
0.172
-1/-1
-1/0
-1/+1
unknown
Register state
suring\data\20111220\161644_MBI_SIL2_MBI_NMR_prep_I_SRO

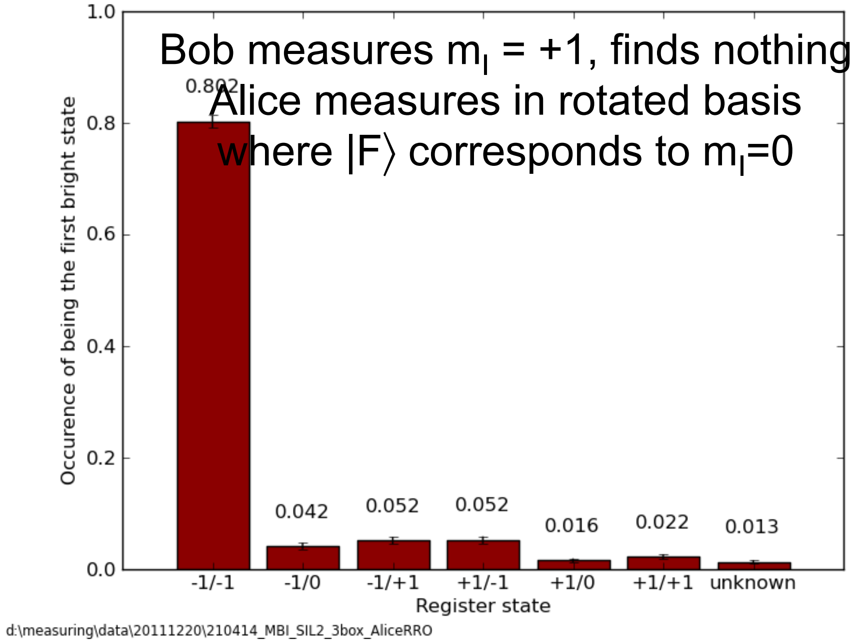

Bob measures m_I = -1, finds nothing
Alice measures in rotated basis
where $|F\rangle$ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.282
0.038
0.520
0.160
-1/-1
-1/0
-1/+1
unknown
Register state
asuring\data\20111220\155158_MBI_SIL2_MBI_NMR_prep_I_SRO

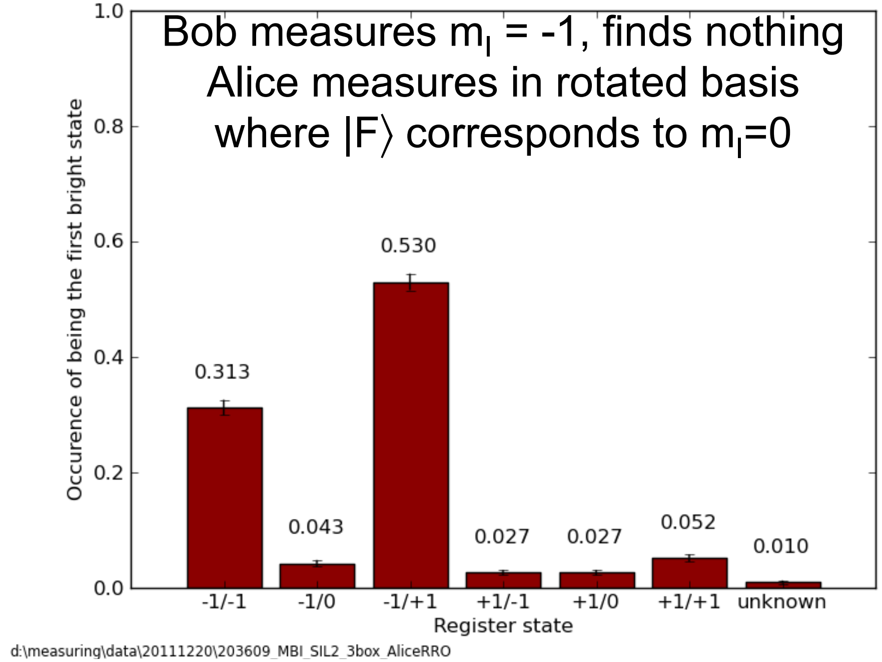

Bob measures m_I = +1, finds something
Alice measures in rotated basis
where $|F\rangle$ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.040
0.238
0.502
0.219
-1/-1
-1/0
-1/+1
unknown
Register state
suring\data\20111220\162921_MBI_SIL2_MBI_NMR_prep_I_SRO

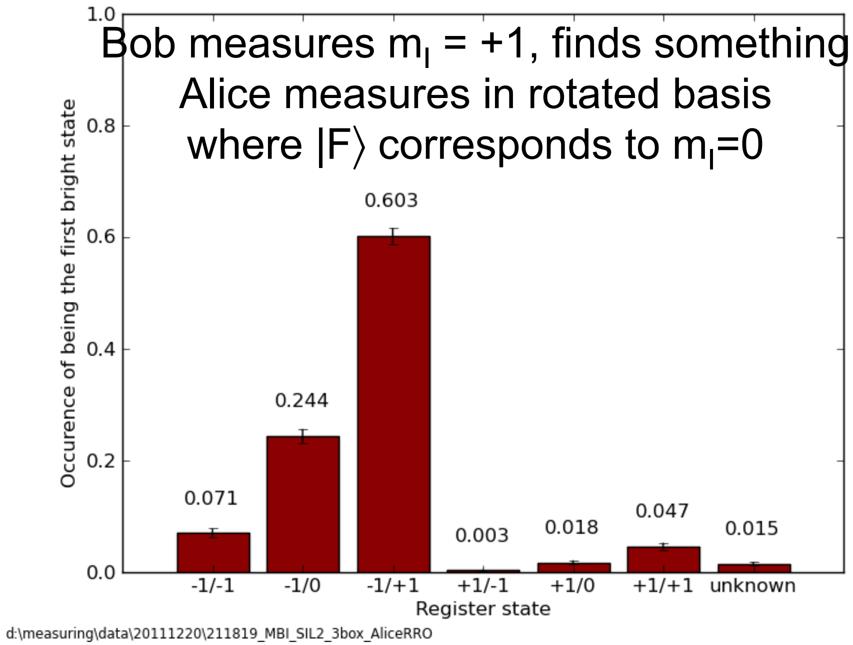

Bob measures m_I = -1, finds something
Alice measures in rotated basis
where $|F\rangle$ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.339
0.340
0.163
0.158
-1/-1
-1/0
-1/+1
unknown
Register state
asuring\data\20111220\160253_MBI_SIL2_MBI_NMR_prep_I_SRO

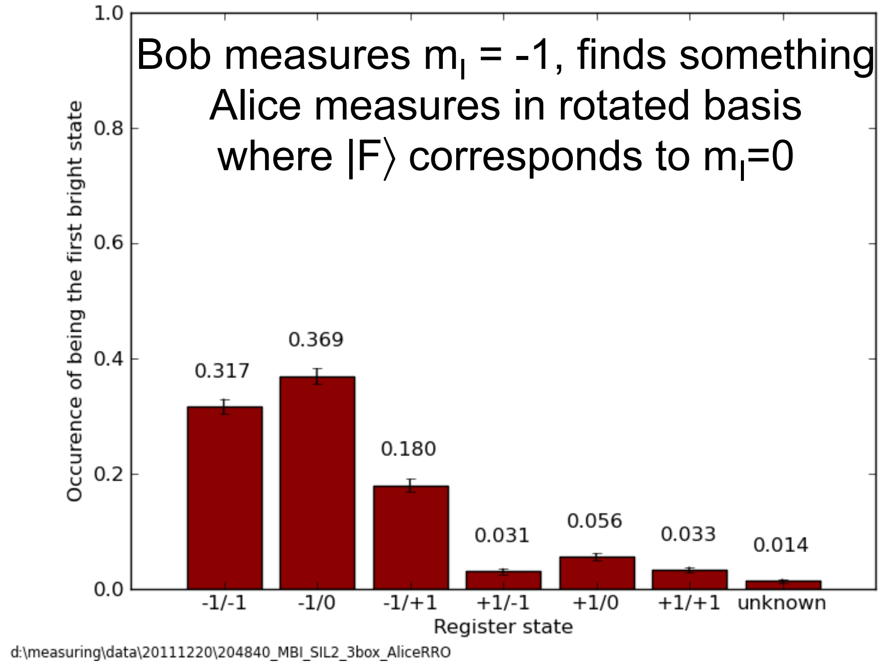

Bob measures m_I = 0, finds nothing
Alice measures in rotated basis
where $|F\rangle$ corresponds to m_I=0
Occurence of being the first bright state
Register state
1.0
0.8
0.6
0.4
0.2
0.0
0.040
0.238
0.502
0.219
-1/-1
-1/0
-1/+1
unknown
suring\data\20111220\162921_MBI_SIL2_MBI_NMR_prep_I_SRO

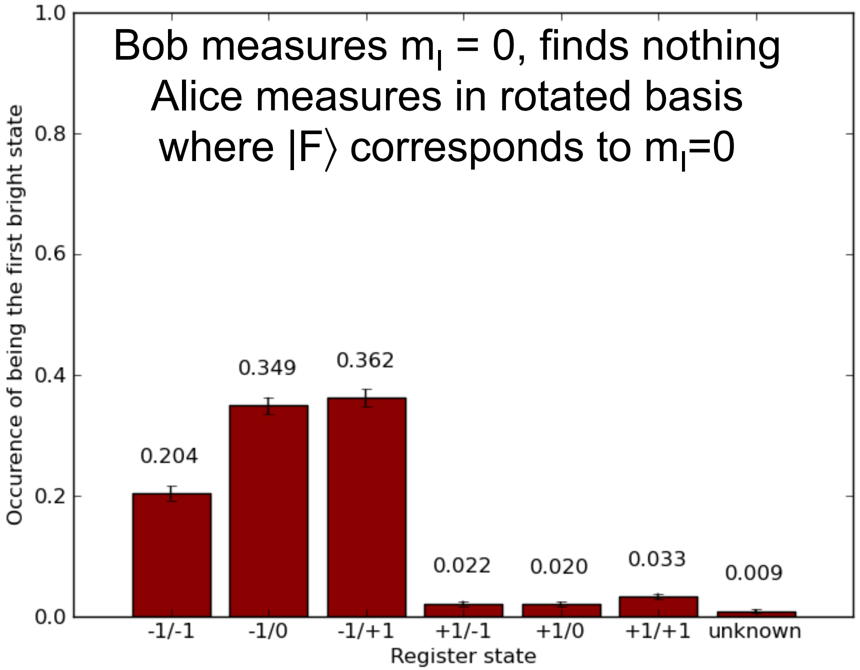

Bob measures m_I = 0, finds something
Alice measures in rotated basis
where $|F\rangle$ corresponds to m_I=0
Occurence of being the first bright state
Register state
1.0
0.8
0.6
0.4
0.2
0.0
0.390
0.275
0.120
0.215
-1/-1
-1/0
-1/+1
unknown
asuring\data\20111220\170433_MBI_SIL2_MBI_NMR_prep_I_SRO


Bob measures m_I = +1, finds nothing
Alice measures in rotated basis
where |F⟩ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.802
0.042
0.052
0.052
0.016
0.022
0.013
-1/-1
-1/0
-1/+1
+1/-1
+1/0
+1/+1
unknown
Register state
d:\measuring\data\20111220\210414_MBI_SIL2_3box_AliceRRO

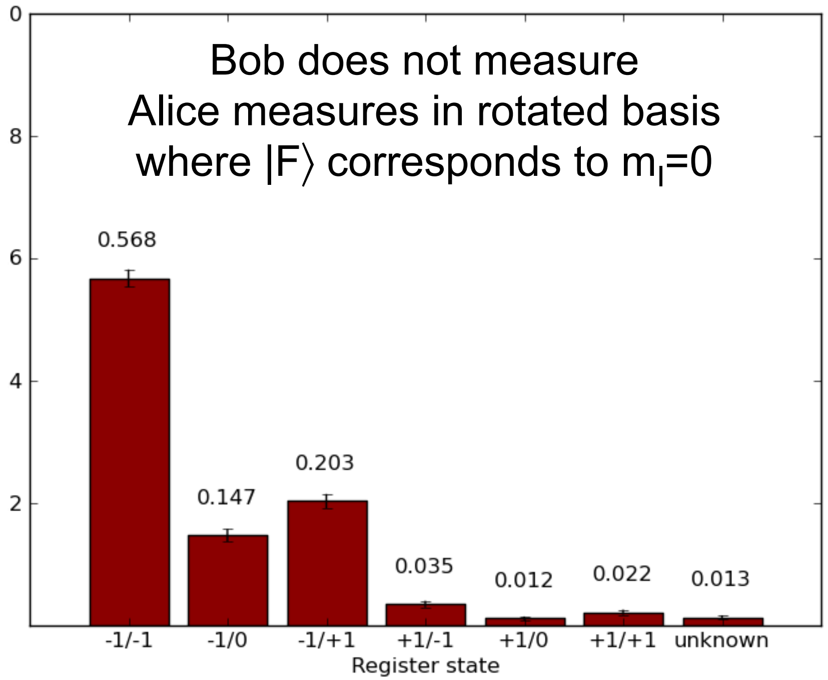

Bob measures m_I = -1, finds nothing
Alice measures in rotated basis
where |F⟩ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.313
0.043
0.530
0.027
0.027
0.052
0.010
-1/-1
-1/0
-1/+1
+1/-1
+1/0
+1/+1
unknown
Register state
d:\measuring\data\20111220\203609_MBI_SIL2_3box_AliceRRO


Bob measures m_I = +1, finds something
Alice measures in rotated basis
where |F⟩ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.071
0.244
0.603
0.003
0.018
0.047
0.015
-1/-1
-1/0
-1/+1
+1/-1
+1/0
+1/+1
unknown
Register state
d:\measuring\data\20111220\211819_MBI_SIL2_3box_AliceRRO

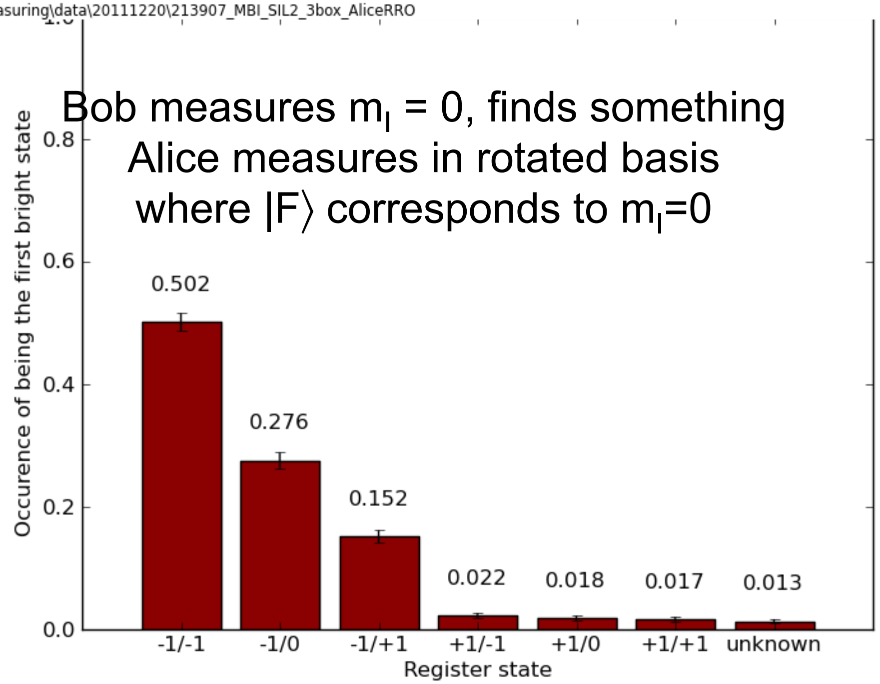

Bob measures m_I = -1, finds something
Alice measures in rotated basis
where |F⟩ corresponds to m_I=0
Occurence of being the first bright state
1.0
0.8
0.6
0.4
0.2
0.0
0.317
0.369
0.180
0.031
0.056
0.033
0.014
-1/-1
-1/0
-1/+1
+1/-1
+1/0
+1/+1
unknown
Register state
d:\measuring\data\20111220\204840_MBI_SIL2_3box_AliceRRO



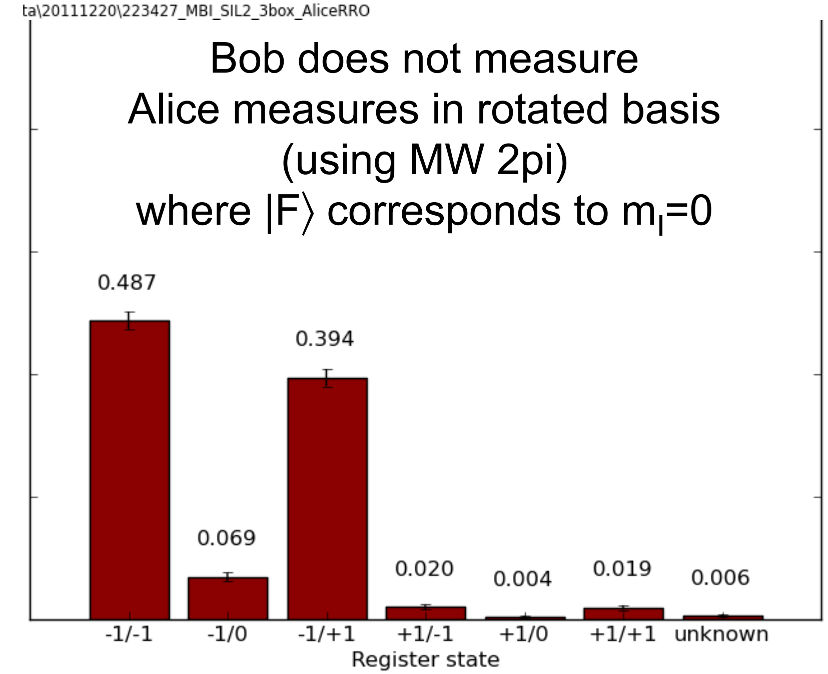
d:\measuring\data\20111220\213907_MBI_SIL2_3box_AliceRRO



ta\20111220\223427_MBI_SIL2_3box_AliceRRO



d:\measuring\data\20111220\215840_MBI_SIL2_3box_AliceRRO

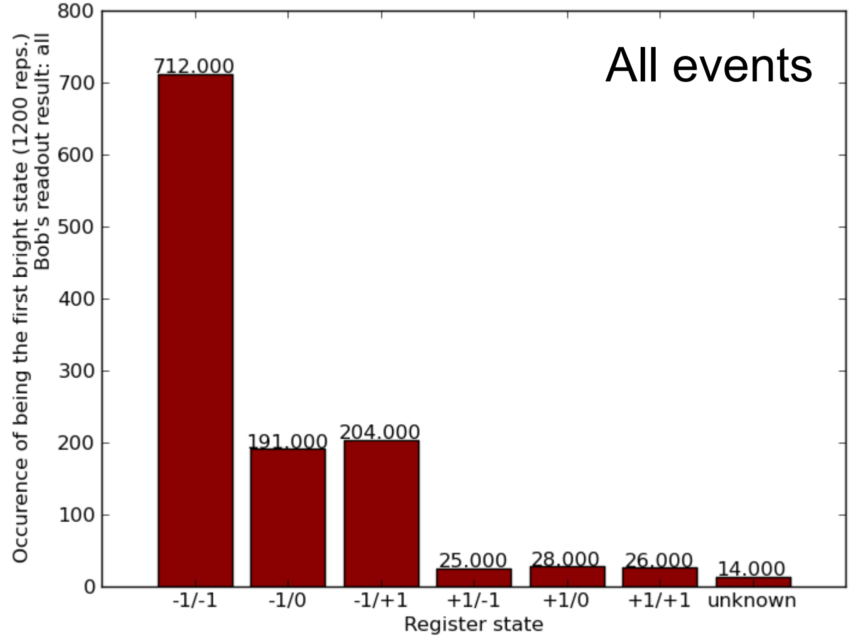


ta\20111220\224941_MBI_SIL2_3box_AliceRRO

This data is obtained without postselection (except MBI):

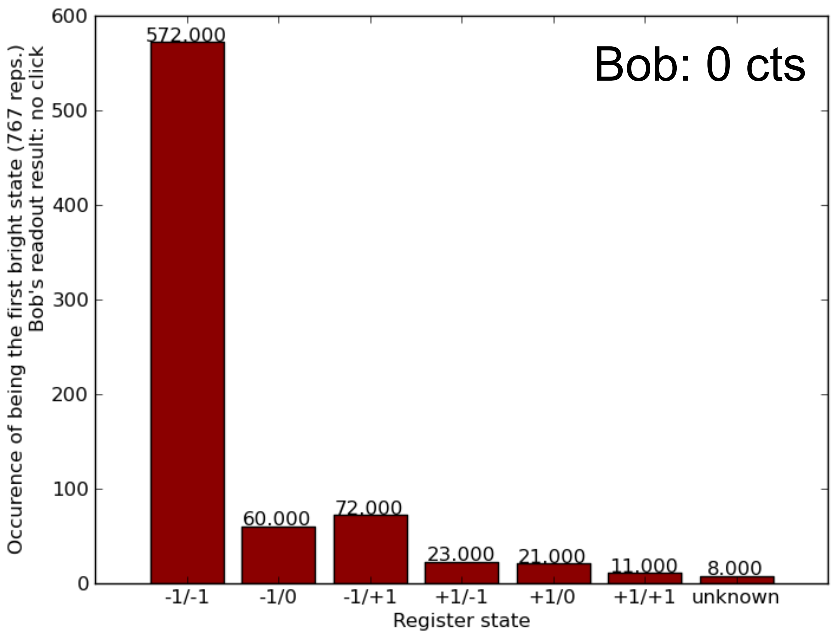
- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly close the box), if he found something

Bob measures mI = +1

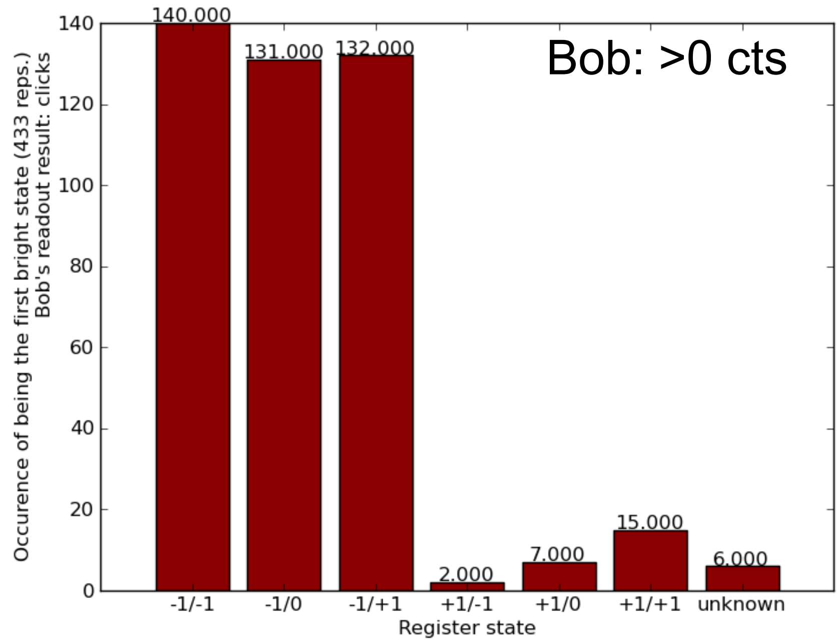


No MW in ‘bob_click’

d:\measuring\data\20111221\142428_MBI_SIL2_3box_AliceRRO



d:\measuring\data\20111221\142428_MBI_SIL2_3box_AliceRRO

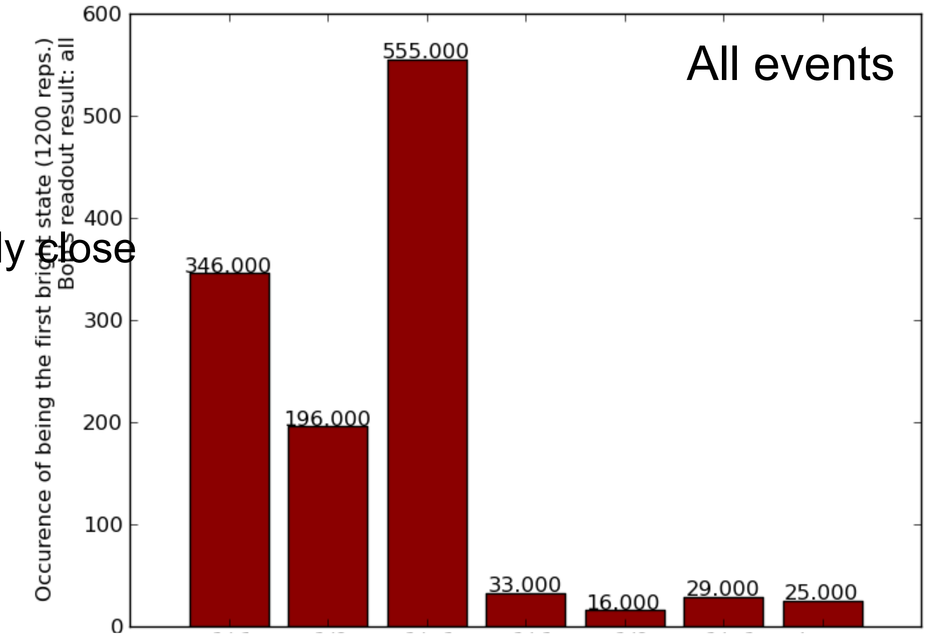


d:\measuring\data\20111221\142428_MBI_SIL2_3box_AliceRRO

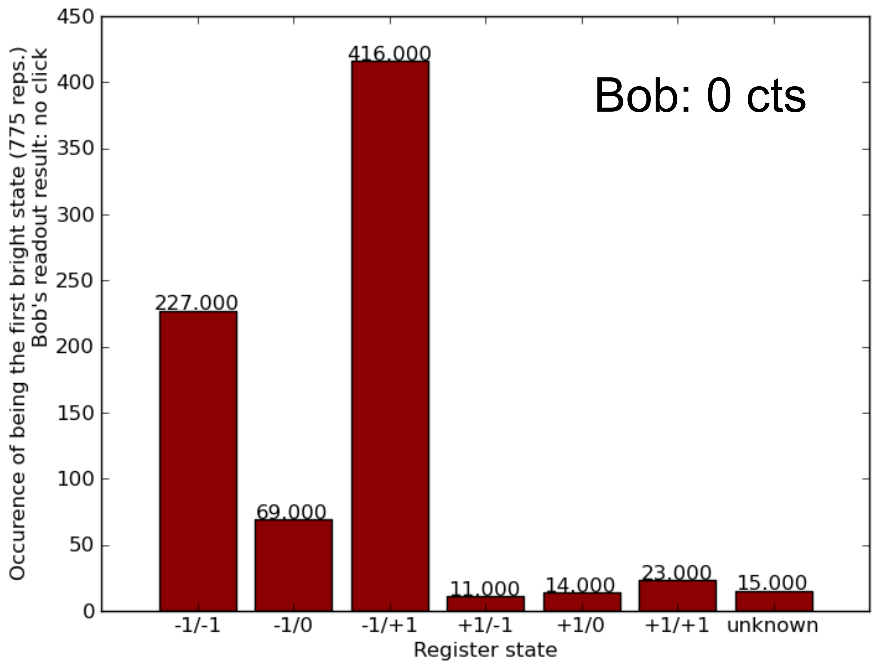
This data is obtained without postselection (except MBI):

- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly close the box), if he found something

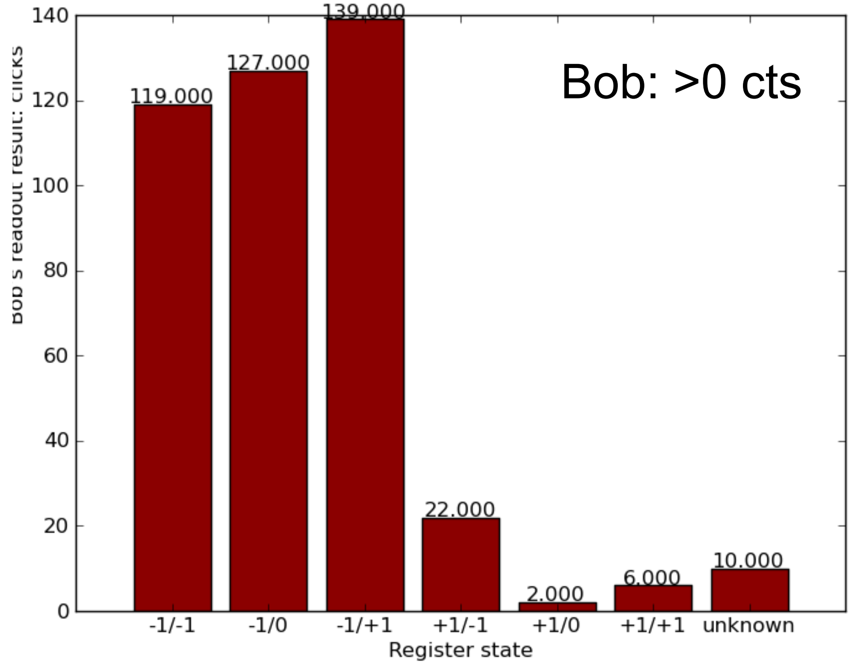
Bob measures mI = -1



No MW in ‘bob_click’



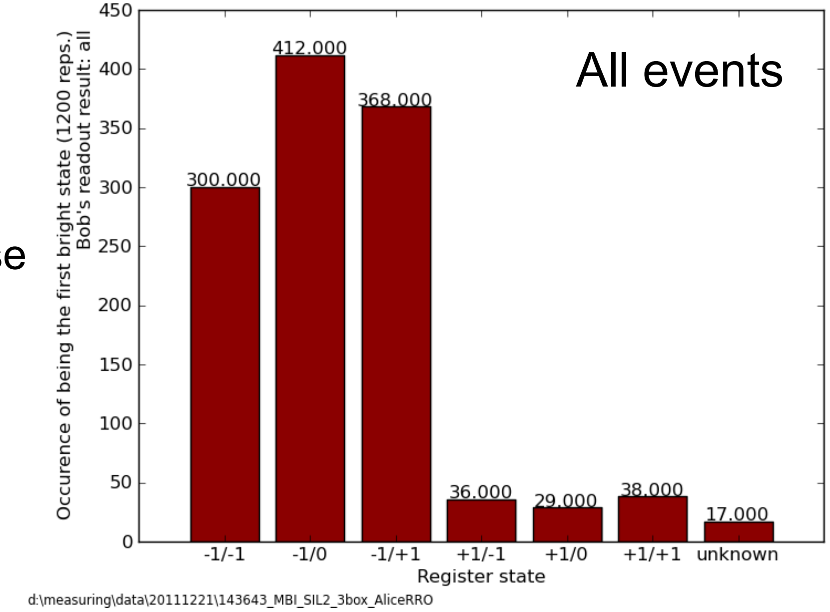
d:\measuring\data\20111221\141156_MBI_SIL2_3box_AliceRRO



uring\data\20111221\141156_MBI_SIL2_3box_AliceRRO

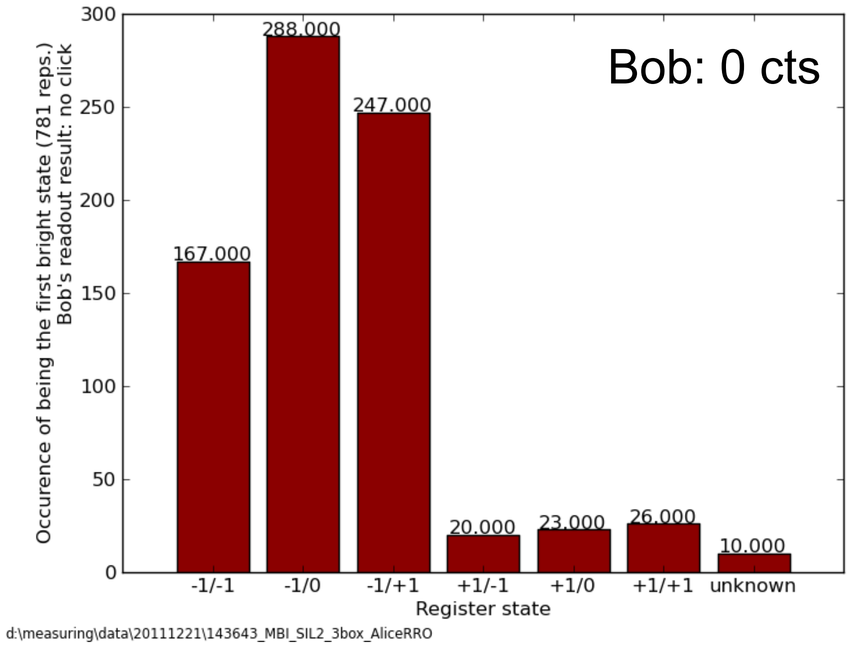
This data is obtained without postselection (except MBI):

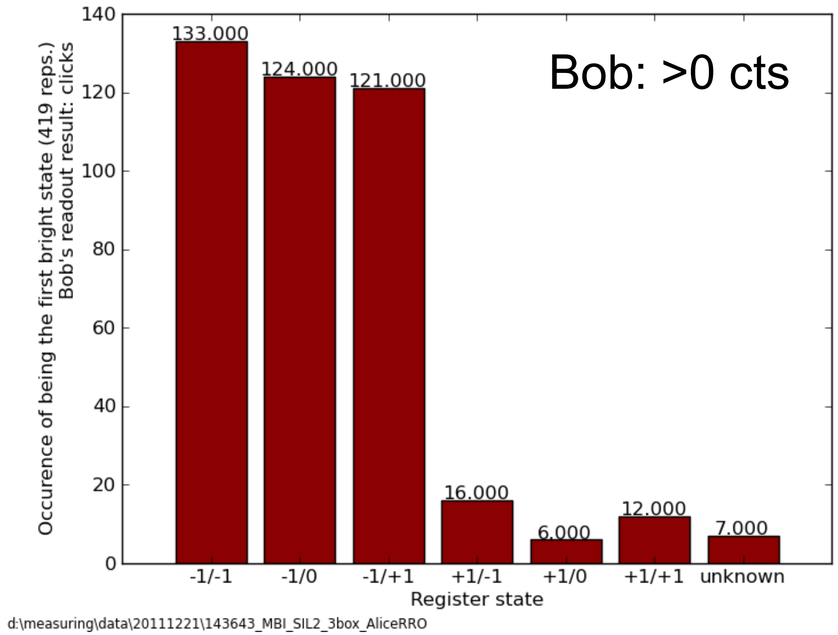
- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly close the box), if he found something



Bob measures mI = 0
(which he is not supposed to do!!!)

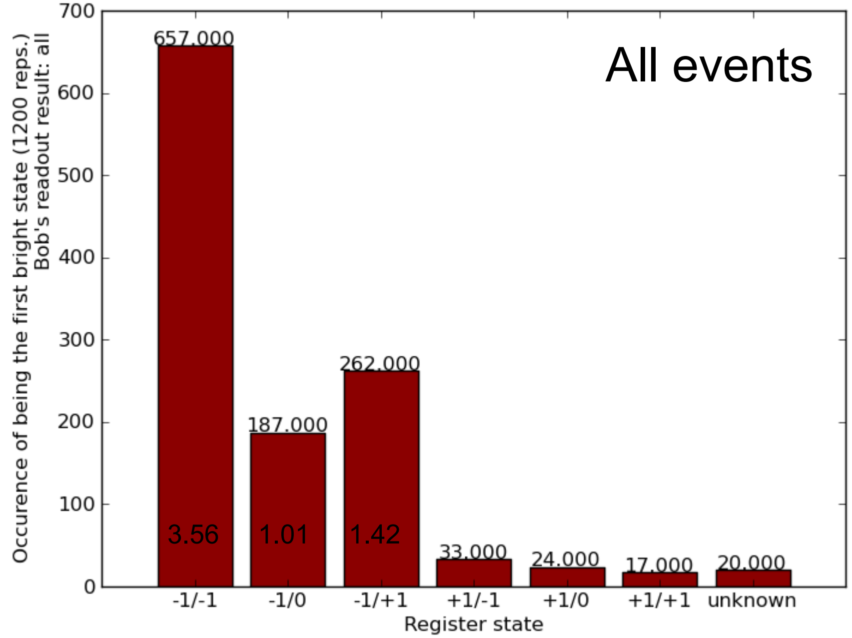
No MW in 'bob_click'





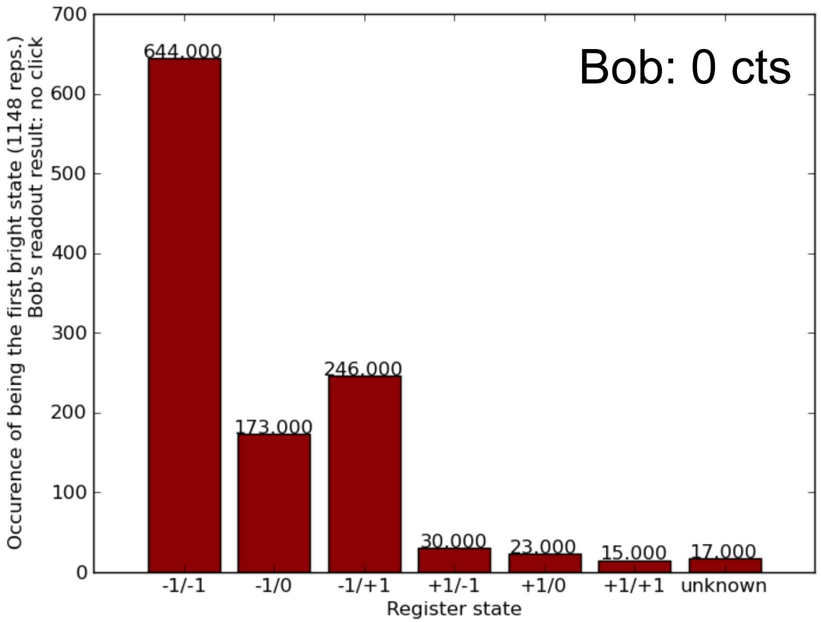
This data is obtained without postselection (except MBI):

- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly cl the box), if he found something

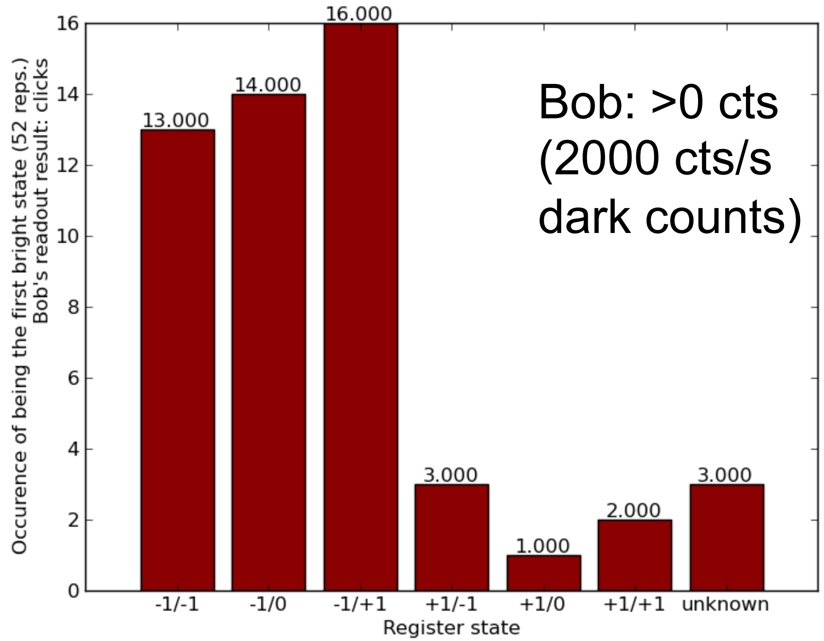


d:\measuring\data\20111221\145958_MBI_SIL2_3box_AliceRRO

Bob does not measure
(nominally: mI=-1, but 0 power)!



d:\measuring\data\20111221\145958_MBI_SIL2_3box_AliceRRO



d:\measuring\data\20111221\145958_MBI_SIL2_3box_AliceRRO

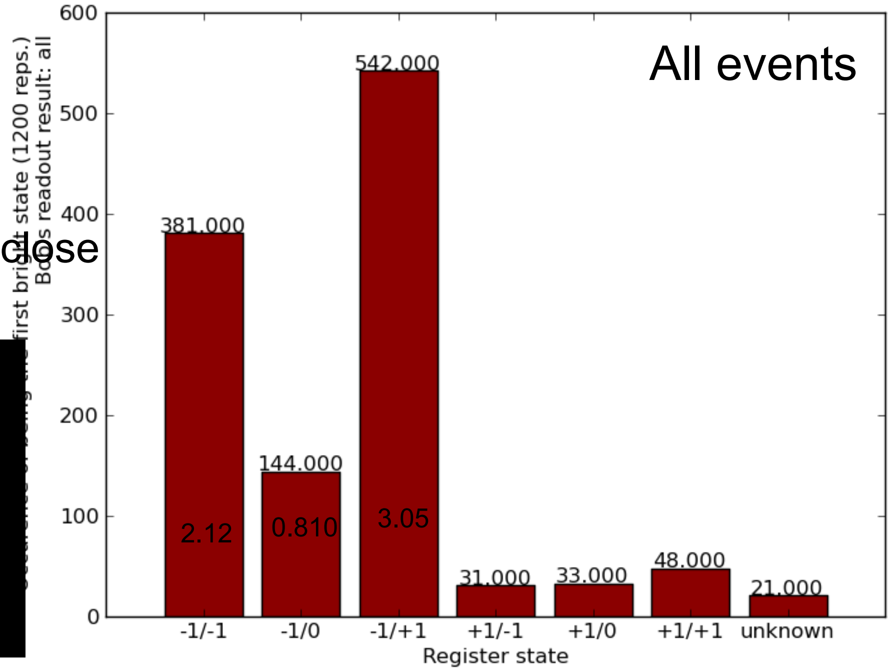
Good data starts here:

This data is obtained without postselection (except MBI):

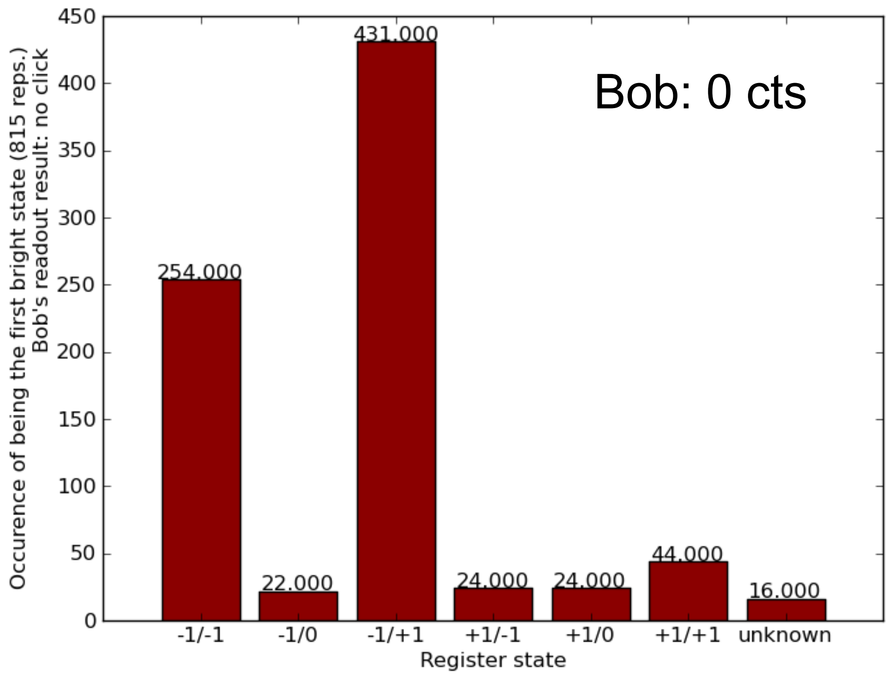
- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly close the box), if he found something

```
In [115]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 317
N(NB,A), lines 1 + 4: 61
N(B,NA), lines 1 + 4: 257
N(NB,NA), lines 1 + 4: 754
P(A) = 0.272138
P(B) = 0.413247
P( B| A) = 0.838624
P(!B| A) = 0.161376
```

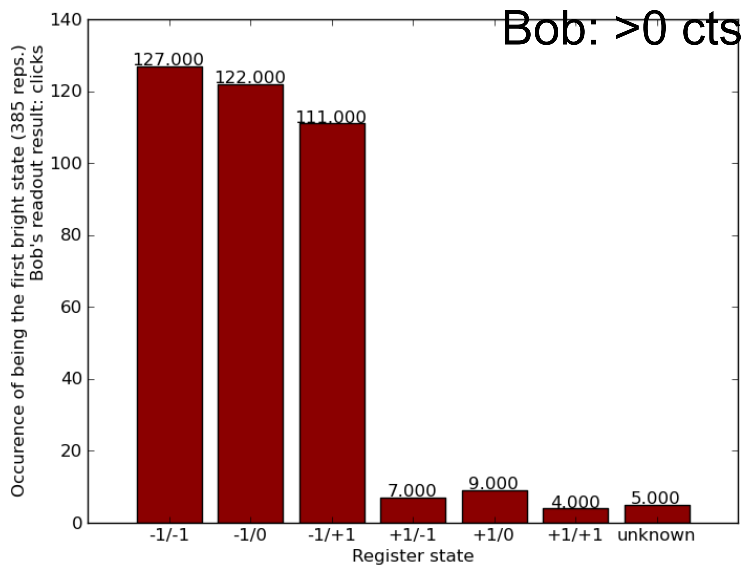
Bob measures ml=-1



d:\measuring\data\20111221\154727_MBI_SIL2_3box_AliceRRO



d:\measuring\data\20111221\154727_MBI_SIL2_3box_AliceRRO



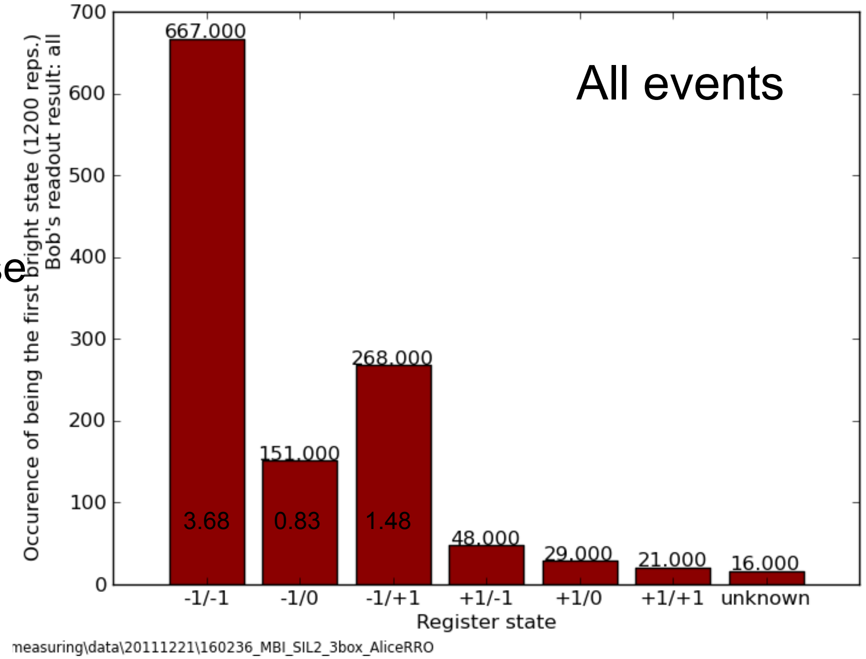
d:\measuring\data\20111221\154727_MBI_SIL2_3box_AliceRRO

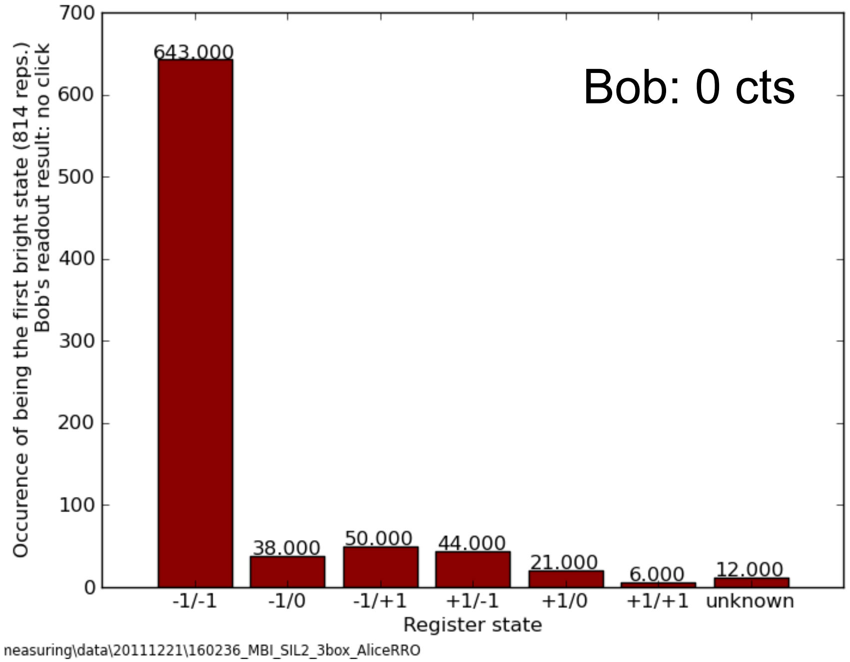
This data is obtained without postselection (except MBI):

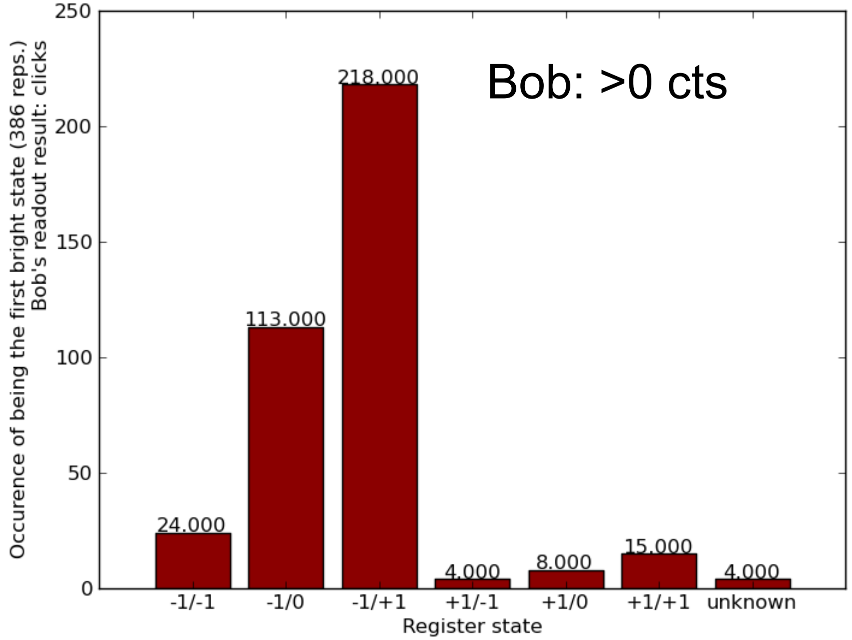
- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly close the box), if he found something

```
In [109]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 332
N(NB,A), lines 1 + 4: 82
N(B,NA), lines 1 + 4: 261
N(NB,NA), lines 1 + 4: 732
P(A) = 0.294243
P(B) = 0.421464
P( B| A) = 0.801932
P(!B| A) = 0.198068
```

Bob measures m_I=+1





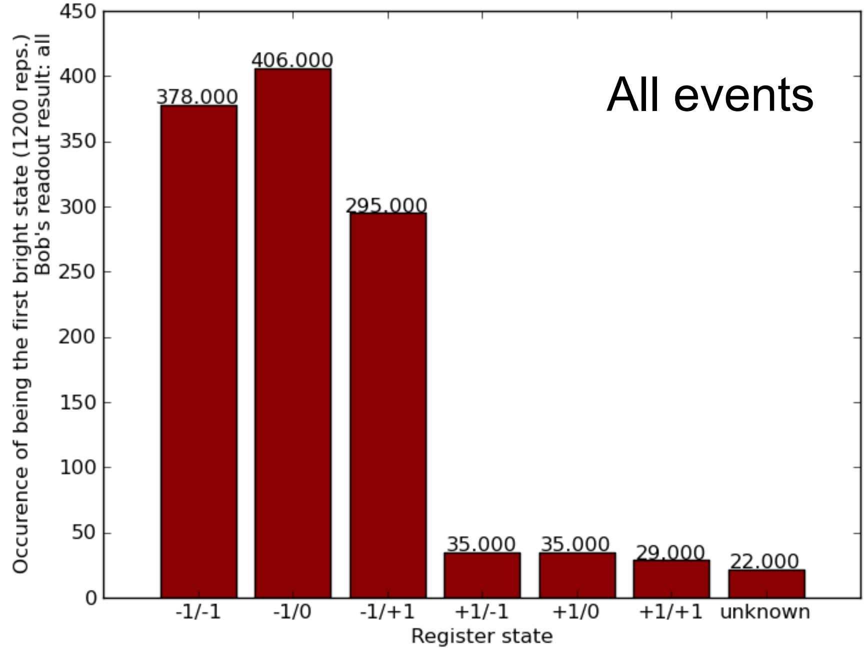


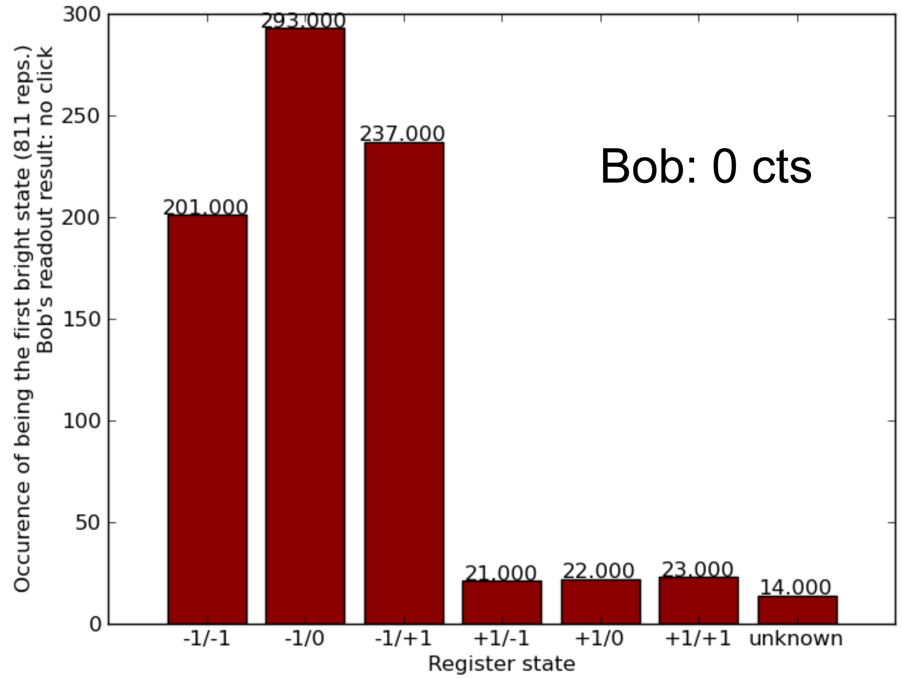
This data is obtained without postselection (except MBI):

- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly clo the box), if he found something

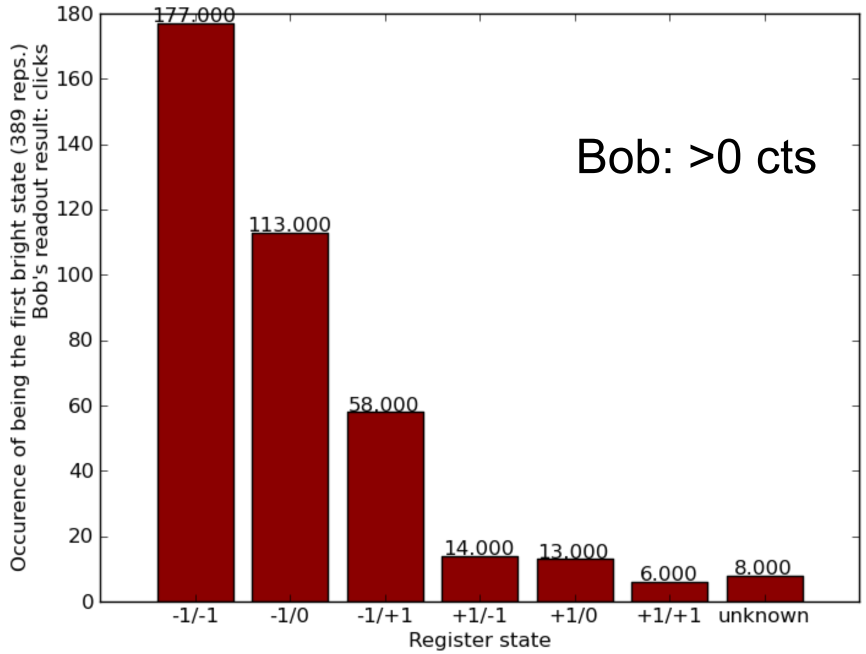
```
In [122]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 563
N(NB,A), lines 1 + 4: 313
N(B,NA), lines 1 + 4: 264
N(NB,NA), lines 1 + 4: 498
P(A) = 0.534799
P(B) = 0.504884
P( B| A) = 0.642694
P(!B| A) = 0.357306
```

Bob measures m_I=0 (cheating)





:\measuring\data\20111221\161527_MBI_SIL2_3box_AliceRRO



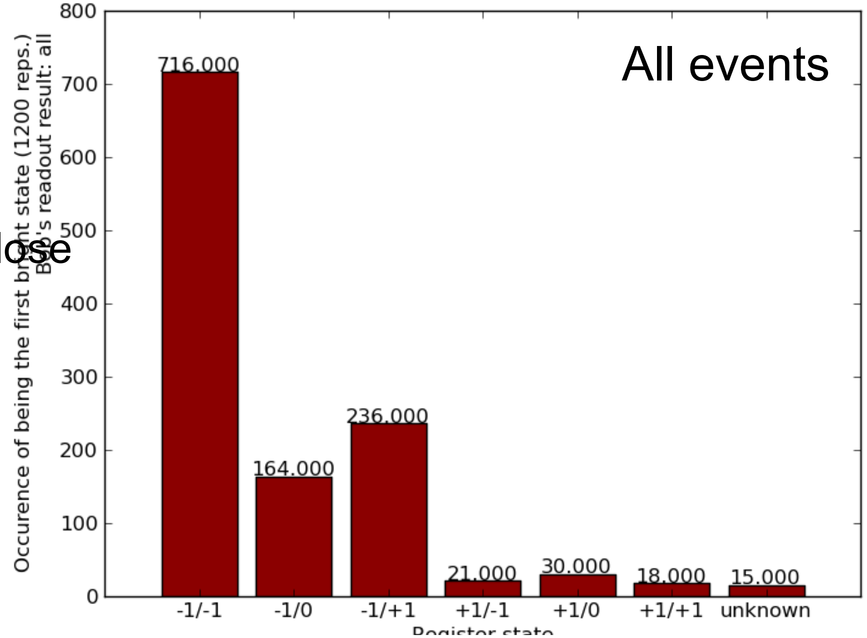
ıeasuring\data\20111221\161527_MBI_SIL2_3box_AliceRRO

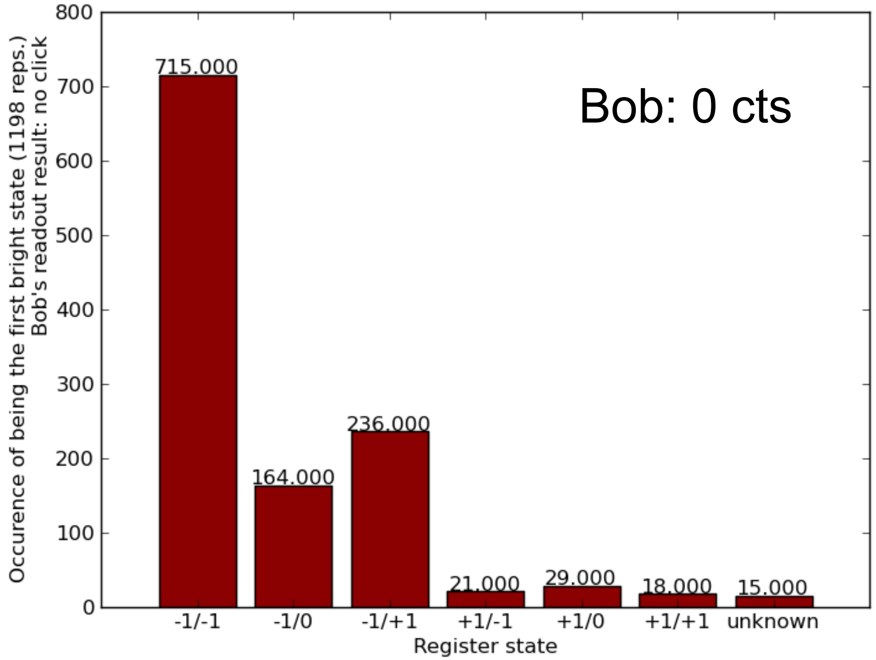
This data is obtained without postselection (except MBI):

- resonance check is applied before starting the sequence,
- bob only applies a repump pulse (to properly close the box), if he found something

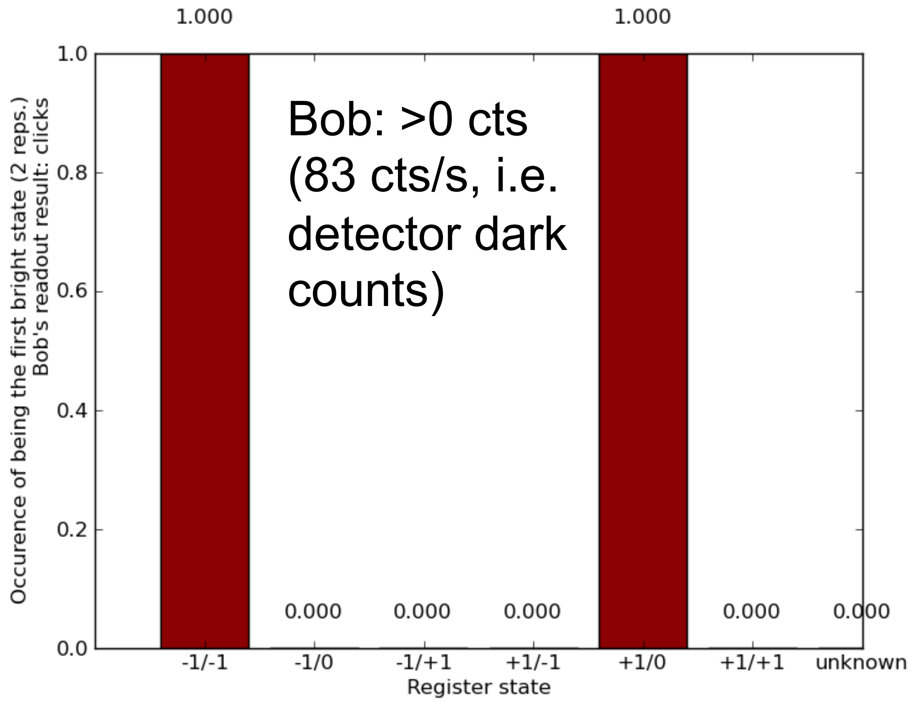
```
In [117]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 225
N(NB,A), lines 1 + 4: 223
N(B,NA), lines 1 + 4: 1
N(NB,NA), lines 1 + 4: 975
P(A) = 0.314607
P(B) = 0.158708
P( B| A) = 0.502232
P(!B| A) = 0.497768
```

Bob does not measure





d:\measuring\data\20111221\164624_MBI_SIL2_3box_AliceRRO



d:\measuring\data\20111221\164624_MBI_SIL2_3box_AliceRRO

$m_l = 1$

```
In [109]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 332
N(NB,A), lines 1 + 4: 82
N(B,NA), lines 1 + 4: 261
N(NB,NA), lines 1 + 4: 732
P(A) = 0.294243
P(B) = 0.421464
P( B| A) = 0.801932
P(!B| A) = 0.198068
```

Bob does not measure

```
In [117]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 225
N(NB,A), lines 1 + 4: 223
N(B,NA), lines 1 + 4: 1
N(NB,NA), lines 1 + 4: 975
P(A) = 0.314607
P(B) = 0.158708
P( B| A) = 0.502232
P(!B| A) = 0.497768
```

$m_l = -1$

```
In [115]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 317
N(NB,A), lines 1 + 4: 61
N(B,NA), lines 1 + 4: 257
N(NB,NA), lines 1 + 4: 754
P(A) = 0.272138
P(B) = 0.413247
P( B| A) = 0.838624
P(!B| A) = 0.161376
```

Bob does not measure

```
In [119]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 200
N(NB,A), lines 1 + 4: 200
N(B,NA), lines 1 + 4: 4
N(NB,NA), lines 1 + 4: 996
P(A) = 0.285714
P(B) = 0.145714
P( B| A) = 0.5
P(!B| A) = 0.5
```

$m_l = 0$ (Bob cheats)

```
In [122]: execfile('mbi-seqrostats-analysis_reduced.py')
N(B,A), lines 1 + 4: 563
N(NB,A), lines 1 + 4: 313
N(B,NA), lines 1 + 4: 264
N(NB,NA), lines 1 + 4: 498
P(A) = 0.534799
P(B) = 0.504884
P( B| A) = 0.642694
P(!B| A) = 0.357306
```

	Powerpoint	Slide	Datestamp	Dataset	Expt.	Description	Paper fig	ml = -1 Box 1	ml = 0 Box 3	ml = +1 Box 2	Undetermined	sigma Box 1	sigma Box 3	sigma Box 2	sigma Undetermined		ml = -1 Box 1	ml = 0 Box 3	ml = +1 Box 2	Undetermined	sigma Box 1	sigma Box 3	sigma Box 2	sigma Undetermined	
	2011-12-17-SIL2	18	20111220	133808	Bob reads ml = 0	Bob opens boxes for first time	3.a.i	0.299	0.299	0.279	0.123	0.0242	0.0242	0.0245	0.0270		**57.26**	**57.26**	**53.43**	**23.56**	**9.26**	**9.26**	**9.39**	**10.35**	
	2011-12-17-SIL2	18	20111220	142811	Bob reads ml = 0	Bob found nothing in 3, remeasures:		0.385	0.061	0.409	0.145	0.0226	0.0280	0.0222	0.0267		73.7	11.68	78.33	27.8	8.67	10.71	8.50	10.22	
	2011-12-17-SIL2	18	20111220	145104	Bob reads ml = 0	Bob found something in 3, remeasures:		0.044	0.737	0.069	0.150	0.0282	0.0148	0.0279	0.0266		8.43	141.14	13.21	29	10.81	5.67	10.67	10.19	
	2011-12-17-SIL2	19	20111220	150135	Bob reads ml = -1	Bob finds nothing in box 1, remeasures:		0.092	0.385	0.397	0.127	0.0275	0.0226	0.0224	0.0270		17.62	73.7	76.03	24.32	10.54	8.67	8.59	10.33	
	2011-12-17-SIL2	19	20111220	151025	Bob reads ml = -1	Bob finds something in box 1, remeasures:		0.733	0.054	0.051	0.162	0.0149	0.0281	0.0281	0.0264		140.38	10.34	9.77	31.02	5.71	10.75	10.77	10.12	
	2011-12-17-SIL2	19	20111220	152036	Bob reads ml = +1	Bob finds nothing in box 2, remeasures:	3.a.ii	0.440	0.387	0.051	0.123	0.0216	0.0226	0.0281	0.0270		**84.3**	**74.11**	**9.77**	**23.56**	**8.27**	**8.66**	**10.77**	**10.35**	***(ii)***
	2011-12-17-SIL2	19	20111220	153157	Bob reads ml = +1	Bob finds something in box 2, remeasures:	3.a.iii	0.047	0.061	0.738	0.154	0.0282	0.0280	0.0148	0.0266		***9.00***	***11.68***	***141.33***	***29.49***	***10.79***	***10.71***	***5.66***	***10.17***	***(iii)***

Fig 3b	M1|M1	¬M1|¬M1	M2|M2	¬M2|¬M2	M3|M3	¬M3|¬M3
	0.733	0.782	0.738	0.827	0.737	0.794
Change	*0.105*	*0.054*	*0.108*	*0.051*	*0.113*	*0.061*
	sigma	**sigma**	**sigma**	**sigma**	**sigma**	**sigma**
	0.015	0.010	0.015	0.008	0.015	0.009
	sigma (change)	**sigma (change)**	**sigma (change)**	**sigma (change)**	**sigma (change)**	**sigma (change)**
	0.027	*0.028*	*0.027*	*0.028*	*0.027*	*0.028*
	Bar length (pt)	**Bar length (pt)**	**Bar length (pt)**	**Bar length (pt)**	**Bar length (pt)**	**Bar length (pt)**
	140.4	*149.8*	*141.3*	*158.4*	*141.1*	*152.1*
	sigma (pt)	**sigma (pt)**	**sigma (pt)**	**sigma (pt)**	**sigma (pt)**	**sigma (pt)**
	5.7	*3.7*	*5.7*	*3.3*	*5.7*	*3.5*
	change (pt)	**change (pt)**	**change (pt)**	**change (pt)**	**change (pt)**	**change (pt)**
	20.1	*10.3*	*20.7*	*9.8*	*21.6*	*11.7*
	sigma change (pt)	**sigma change (pt)**	**sigma change (pt)**	**sigma change (pt)**	**sigma change (pt)**	**sigma change (pt)**
	5.2	*5.4*	*5.2*	*5.4*	*5.2*	*5.4*

	Powerpoint	Slide	Datestamp	Dataset	Expt.	Description	Paper fig	ml = -1 Box 1	ml = 0 Box 3	ml = +1 Box 2	ml = -1 Box 1'	ml = 0 Box 3'	ml = +1 Box 2'	Undetermined	sigma Box 1	sigma Box 3	sigma Box 2	sigma Undetermined		Bar Lengths (pt)							
	2011-12-17-SIL2	18	20111220	133808	Bob reads ml = 0	Bob opens boxes for first time	3.a.i	0.299	0.299	0.279				0.123	0.0242	0.0242	0.0245	0.0270		57.26	57.26	53.43	23.56	4.63	4.63	4.694	5.177
	2011-12-17-SIL2	18	20111220	142811	Bob reads ml = 0	Bob found nothing in 3, remeasures:		0.385	0.061	0.409				0.145	0.0226	0.0280	0.0222	0.0267		73.7	11.68	78.33	27.8	4.34	5.357	4.25	5.11
	2011-12-17-SIL2	18	20111220	145104	Bob reads ml = 0	Bob found something in 3, remeasures:		0.044	0.737	0.069				0.150	0.0282	0.0148	0.0279	0.0266		8.43	141.14	13.21	29	5.41	2.84	5.33	5.10
	2011-12-17-SIL2	19	20111220	150135	Bob reads ml = -1	Bob finds nothing in box 1, remeasures:		0.092	0.385	0.397				0.127	0.0275	0.0226	0.0224	0.0270		17.62	73.7	76.03	24.32	5.27	4.335	4.29	5.17
	2011-12-17-SIL2	19	20111220	151025	Bob reads ml = -1	Bob finds something in box 1, remeasures:		0.733	0.054	0.051				0.162	0.0149	0.0281	0.0281	0.0264		140.38	10.34	9.77	31.02	2.86	5.377	5.39	5.06
	2011-12-17-SIL2	19	20111220	152036	Bob reads ml = +1	Bob finds nothing in box 2, remeasures:	3.a.ii	0.440	0.387	0.051				0.123	0.0216	0.0226	0.0281	0.0270		84.3	74.11	9.77	23.56	4.14	4.328	5.39	5.18
	2011-12-17-SIL2	19	20111220	153157	Bob reads ml = +1	Bob finds something in box 2, remeasures:	3.a.iii	0.047	0.061	0.738				0.154	0.0282	0.0280	0.0148	0.0266		9.00	11.68	141.33	29.49	5.40	5.357	2.83	5.08