Direct Randomized Benchmarking, Gate Set Tomography and Error Analysis on mixed-species trapped ion qubits

Alessio Ruggeri
aruggeri@student.ethz.ch

August 2021

MSc Physics Master Thesis
Supervised by Tanja Behrle, Alfredo Ricci Vásquez and Jonathan Home
Trapped Ion Quantum Information Group

ETH zürich

Abstract

Trapped ions are a promising architecture thanks to which scalable quantum information processing could be achieved. There is a consensus that, in order to reach this goal, experimental noise has to be reduced down to a threshold of 10^{-4}[18; 31] per logic operation. Identifying and quantifying errors is therefore pivotal.

This dissertation proposes two algorithms whose aim is noise characterisation: Direct Randomized Benchmarking and Gate Set Tomography. These have been executed on Segtrap, a 3D segmented ion trap able to control $^{9}Be^{+}$ and $^{40}Ca^{+}$. This work applies the aforementioned algorithms to the single Beryllium ion and to the single Calcium ion cases and offers a method of interpreting Direct Randomized Benchmarking data, inspired by Ball et al. [3], based on Γ distributions.

Acknowledgments

Without Alfredo Ricci Vásquez and Tanja Behrle the realization of this project would have not been possible. Alfredo, you have been an inspiration and a guide through the intricate theory of Randomized Benchmarking. Tanja, you have been an attentive and knowledgeable leader in the B20 lab, shedding light on any component of Segtrap. I believe I learnt a lot from the both of you. Many thanks. The theoretical contribution of Ivan Rojkov was a very valuable one: without him theoretical particularities of Direct Randomized Benchmarking would have been ungraspable. Martin Stadler must also be mentioned for his hardware and software support, as well as for his efforts on making experimental sequences longer. Thanks also to everyone in the TIQI group who offered help when troubles arose.

Introduction

Evaluating the performances of a quantum processing setup is pivotal if scalable quantum information processing is to be achieved and if an error threshold of 10^{-4} is to be obtained. Quantifying how well quantum operations can be executed, and therefore how noise affects the executions of quantum algorithms, can be the first step in order to recognize noise sources and mitigate them. Different algorithm classes, whose aim is describing the aforementioned performances, are at disposal. Two of these classes are tomographic algorithms [27] and Randomized Benchmarking protocols [3; 34]. The computational time of the former scales exponentially with the number of qubits involved, while the latter is efficient [21]. This difference reflects another dissimilarity: tomographic algorithms are able to fully describe quantum logic operations, while Randomized Benchmarking protocols can only offer average quantities. In both cases however, *representations* play a crucial role. A representation, roughly speaking, is the way of explicitly writing down element by element any operation belonging to a certain gate set. In general, for any gate set there are more equivalent representations that produce the same physical observables. For the Randomized Benchmarking case, it was believed that the results of such algorithms were in "direct mathematical relation" with the notion of *average gate infidelity*. This is however not true. The numerical value of gate infidelity depends on the representation; therefore it is not a physical quantity. Randomized Benchmarking results are instead physical, because they are measured by experiments [32]. Thus, the physical meaning of these results is less trivial than just an estimation of the average gate infidelity. As far as tomographic routines are concerned, a representation is assumed, for example, when a process is characterized with respect to a set of measurements and state preparations. Measurement and states are fixed (hence a representation for them is provided) and as a consequence assumed noiseless. The process is described with respect to them but the error on the process tomography will be at least as great as the error on the knowledge one has about states and measurements [27]. *Calibration errors* of states and measurements affect process tomography.

Executing Randomized Benchmarking and tomographic experiments on the same setup is interesting. It can be observed whether these different classes of algorithms provide consistent results and how they are connected. Executing them an a mixed species ion trap is even more interesting. An advantage of having different ion species is spectral separation: performing quantum operations on an ion of one species does not perturb the information stored in a different species ion [22]. Single qubit experiments, multi qubit experiments and parallel single qubit experiments with more ions trapped can lead to a characterisation of the same operations, with and without the presence of other ions, and of cross talk noise [34].

Thesis outline

Section 1 introduces the theoretical foundations of Randomized Benchmarking and highlights the differences between and the motivations of two algorithms, Clifford Randomized Benchmarking and Direct Randomized Benchmarking, while Section 2 summarises the features and theory of Gate Set Tomography, a *calibration-free* tomographic routine. This dissertation focuses on Direct Randomized Benchmarking rather than Clifford Randomized Benchmarking because the former benchmarks a setup by exploiting logic operations that are natively available to the setup itself. This is not the case for the latter. In Section 3, an analyses of simulated DRB experiments attempts to devise a theoretical model that, under the assumption of coherent noise, should help evaluating the experimental DRB results by providing a noise correlation length estimation. Section 4 recapitulates the main characteristics of the 3D segmented ion trap whose noise this project attempted to characterize: Segtrap. Section 5 illustrates the necessary code to perform Gate Set Tomography and Randomized Benchmarking experiments on Segtrap. This code is also offered to the whole TIQI group as a set of abstract routines implementable on any setup belonging to the group. Section 6 illustrates the collected data and the results that can be drawn from them. Direct Randomized Benchmarking and Gate Set Tomography experiments on the single Calcium and on the single Beryllium ion were carried out.

Contents

1 Randomized Benchmarking: Theory

Randomized Benchmarking (RB) is a protocol that was devised [17] with the aim of estimating the average gate fidelity of a quantum processing system in a scalable way, decoupling state preparation and measurement (SPAM) and gate errors, and as a comparison mean among different setups. Later studies exhibited proof of RB efficient scalability in the number of qubits [16; 21]. Recently, another way of performing RB was developed, which also considers another scaling problem: *gate compilation*. The two algorithms are respectively called Clifford Randomized Benchmarking (CRB) and Direct Randomized Benchmarking (DRB)[1] [34]. Because of DRB recentness, its basis is not as solid and broad as than the CRB one. The goal of this section is to summarize some theoretical results of the last 15 years about RB.

1.1 Clifford Randomized Benchmarking

1.1.1 Mathematical preliminaries

CRB has been broadly used for around a decade [52; 5]. Its core concepts are *average fidelity* and *process twirling*. The average fidelity of a CPTP (noise) process Λ is defined as follows [30]:

$$F(\Lambda) \equiv \int \mathrm{d}\psi \, \langle\psi| \, \Lambda(\psi) \, |\psi\rangle \tag{1}$$

where $\mathrm{d}\psi$ is the Fubini measure [10] and $|\psi\rangle$ a pure state. F can in general be estimated tomographically, but this escalates exponentially with the number of qubits, whilst here a more efficient approach is sought. To achieve this objective the *twirled process* Λ_T can be taken into account:

$$\Lambda_T(\rho) \equiv \int \mathrm{d}U \, U^\dagger \Lambda \left(U \rho U^\dagger\right) U \tag{2}$$

with ρ some density operator, U unitary and $\mathrm{d}U$ the Haar measure[2]. It can be shown [30] that

$$\boxed{F(\Lambda_T) = F(\Lambda)} \tag{3}$$

and also that Λ_T is a depolarizing channel [10],

$$\boxed{\Lambda_T(\rho) = \frac{1-p}{d}\mathbb{1} + p\rho} \tag{4}$$

with d the dimension of the Hilbert space. By direct substitution one finds

$$\boxed{\boxed{F(\Lambda) = \frac{1-p}{d} + p}} \tag{5}$$

with $1-p$ the depolarisation probability. The "mechanism" that is taking shape is the following: in order to asses a quantum system performances, the noise affecting the system has to be twirled. A disadvantage of this procedure is the impossibility of applying all the unitaries for the twirling, since they constitute a continuous set. An auxiliary tool is needed to solve this issue: *unitary 2-designs*. There are more definitions for these objects. One of them is particularly interesting given that it is an operational one [6]. Be $\mathcal{H} = \mathbb{C}^d$. A unitary 2-design is a set of unitaries $\{U_k\}_k$ on $\mathcal{H}$ s.t., $\forall\, X$, A, B linear on $\mathcal{H}$,

$$\frac{1}{K}\sum_{k=1}^{K} U_k^\dagger A U_k X U_k^\dagger B U_k = \int \mathrm{d}U U^\dagger A U X U^\dagger B U. \tag{6}$$

In this manner it can be seen that, given the suitable set $\{U_k\}_k$ and the Kraus decomposition of Λ,

$$\boxed{\Lambda_T(\rho) = \frac{1}{K}\sum_{k=1}^{K} U_k^\dagger \Lambda \left(U_k \rho U_k^\dagger\right) U_k}. \tag{7}$$

[1] Another known protocol is Interleaved Randomized Benchmarking, where a specific gate of interest is applied between every Clifford gate [11]. In general, variations of the RB protocol are known [17; 16; 47; 12].

[2] Without noise , i.e. with $\Lambda = \mathbb{1}$, $\Lambda_T(\rho) = \rho$.

Does such a set exist? At least one exists, and it is the Clifford group [6]. The Clifford group on n qubits C_n is the group of unitaries mapping the Pauli group to itself under conjugation [48]:

$$C_n \equiv \left\{ U \in \mathcal{U}\left(2^n\right) \mid U\mathcal{P}_n U^\dagger = \mathcal{P}_n \right\}. \tag{8}$$

Moreover, circuits with only Clifford gates can be simulated efficiently by a classical computer [13] and are invertible by a single Clifford gate which can be found in polynomial time [1]. The idea of CRB is to twirl the identity multiple times with Clifford circuits, which are equal to $\mathbb{1}$ when noiseless, on some relevant quantum system and measure the survival probability of a predefined fiducial state, usually $|0\rangle^{\otimes n}$.

1.1.2 The protocol and its results

This is the CRB protocol [3; 21].

1. Choose some lengths $J_1 < J_2 < \cdots < J_{\text{last}}$.
2. Choose an initial state $|\psi\rangle$ (usually $|0\rangle^{\otimes n}$).
3. $\forall J_j$:
   (a) Sample K_j sequences $S_{j,k}$ of lengths $J_j - 1$ from C_n. The J_jth gate in each sequence inverts the whole circuit such that under no noise every $S_{j,k} = \mathbb{1}$. K_j should be large enough to guarantee a meaningful statistical sample.[3]
   (b) $\forall k \in [1, K_j]$:
      i. Apply the sampled circuit and measure the survival probability of the chosen state after every circuit, i.e. $F(\tilde{S}_{k,j}, |\psi\rangle) \equiv \langle\psi| \tilde{S}_{j,k} |\psi\rangle$, and store it ($\tilde{S}_{k,j}$ is the noisy $S_{k,j}$).
   (c) Average the survival probabilities over k and store the result.

Given this paradigm, an actual experimental sequence would have different error channels,

$$\tilde{S}_{j,k} = \bigcirc_{i=0}^{J_j} \left(\Lambda_{j,k,i} \circ C_{j,k,i}\right). \tag{9}$$

By expanding each error channel $\Lambda_{j,k,i}$ at first order as $\Lambda_j^{\text{average}} + \delta\Lambda_{j,k,i}$, it was shown that the survival probability F has this behaviour [21]:

$$F(J_j, \psi) = A_1(J_j)p^{J_j} + B_1(J_j) + C_1(J_j - 1)\left[q(J_j) - p^2\right] p^{J_j - 2}. \tag{10}$$

A_1, B_1, C_1 and q are functions of the state preparation $|\psi\rangle$ and measurement, while p is the same of Eq. (4). All things considered, the noise affecting logic operations is characterized by a single parameter p, which indicates the average impact of the noise in terms of survival probability.

On the other hand, what can be often read in the literature, given the experience of the author of this report, is just the zeroth order model.[4] This means that every $\delta\Lambda = 0$. In other words, the noise is independent of the gate and it is time uncorrelated. Although this hypothesis is rather unphysical, this way of proceeding is the most used one. In this scenario

$$F(J_j, \psi) = A_0 p^{J_j} + B_0. \tag{11}$$

The interesting feature is that also here SPAM errors influence only the values of A_0 and B_0 without modifying the exponential decaying rate, decoupling SPAM and gate noise. In more detail, the average gate set infidelity (AGI) ε can be formalised. The relevant gate set is here the Clifford group C_n. Therefore

$$\varepsilon \equiv \left\langle 1 - \int \mathrm{d}\psi \, \mathrm{Tr}\left[\tilde{C}(|\psi\rangle\langle\psi|)\hat{C}(|\psi\rangle\langle\psi|)\right] \right\rangle_{\hat{C} \in C_n} \tag{12}$$

where $\tilde{C}$ is a noisy Clifford gate. Under the same assumption $\delta\Lambda = 0$, every $\tilde{C} = \Lambda\hat{C}$ and Eq. (12) reduces to $\varepsilon = 1 - F(\Lambda)$ as defined in Eq. (1). As a consequence, an estimator of ε called *the RB number* r can be defined as follows

$$r \equiv 1 - F(\Lambda) = \frac{(d-1)(1-p)}{d}, \tag{13}$$

[3] In an RB experiment, all K_js are usually set equal.

[4] An argument on the feasibility of neglecting higher order terms can be found in the selected reference [21]. It involves the diamond and trace norm. Other reasons are also present here [32].

with d the Hilbert space dimension. Although experimentally $\delta\Lambda \neq 0$ is most likely the case, it safe to use the 0th order model anyway for the behaviour of $F(J_j, \psi)$, i.e. Eq. (11). It was in fact exhibited that "the exact behavior of randomized benchmarking under general gate-dependent noise converges exponentially to a true exponential decay of exactly the same form as that predicted by previous analysis for gate-independent noise" (Wallman, [49], page 1). **However it is misleading and wrong to consider r to be the AGI with $\delta\Lambda \neq 0$.** A quantity which depends on the representation (or gauge), ε, and a representation invariant quantity, $F(J_j, \psi)$, are compared [32].

Given a gate set with unitaries U_w, state preparations ρ_x and measurements $E_{y,z}$[5], changing gauge means replacing

$$U_w \to M U_w M^{-1} \tag{14}$$

$$\rho_x \to M \rho_x \tag{15}$$

$$E_{y,z} \to E_{y,z} M^{-1} \tag{16}$$

for some invertible linear map M. All the physical properties of every gate set do not change under a gauge transformation. The RB number r defined as

$$r \equiv \frac{(d-1)(1-p)}{d} \tag{17}$$

with p from Eq. (11) is gauge invariant, ε is not [6]. One might advance the hypothesis that, with $||\delta\Lambda||^H_{1\to 1} \ll 1$[7], r is nonetheless a good estimator of ε. This is again a mistake. It was shown [32] that even for rather small overrotations the ratio ε/r diverges. A gauge in which $\Lambda = \Lambda_{j,k,i}\, \forall j,k,i$ can always be found, but in general the gate set is not CPTP and thus nonphysical. The physical meaning of r is non trivial[8]. To this level of tractation it can be said that r is a function of the eigenvalues, which are gauge invariant, of the noise and of the operations in the gate set. Please have a look at the selected reference [32] fore more details.

As an additional note, it is also wrong to use r to estimate error thresholds. CRB makes average properties of the noise evident, while thresholds are about the worst case scenario. "[..] many realistic noise processes admit a linear relation between the average error rate (which is accessible experimentally) and the worst-case error (which is the relevant figure of merit for fault tolerance). The exceptions to this rule are highly coherent errors, where the worst-case error scales proportionally to the square root of the average error rate" (Kueng et al., [19], page 5). In the last case, discrepancies can amount to order of magnitudes.

1.1.3 Survival probability errors and distributions

A prominent question is: How many Clifford circuits for a given J_j must be run in order to obtain sensible results for r? Höeffding's inequality is helpful [21]. Let us use the stochastic variable

$$\Sigma_{K_j}(J_j, \psi) \equiv \frac{1}{K_j} \sum_{k=1}^{K_j} F(\tilde{S}_{k,j}, |\psi\rangle). \tag{18}$$

Every symbol has already been defined in Section 1.1.2. Höeffding's inequality states

$$\mathbb{P}\left(\mid \Sigma_{K_j}(J_j, \psi) - F(J_j, \psi) \mid \geqslant \epsilon\right) \leqslant 2e^{-2K_j\epsilon^2/(b-a)^2}. \tag{19}$$

ε, defined in Eq. (12), shall not be confused with ϵ, the accuracy of the estimate. As it can be seen, the confidence δ is

$$2e^{-2K_j\epsilon^2/(b-a)^2} \equiv \delta \tag{20}$$

where $[a, b]$ is the range of $F(\tilde{S}_{k,j}, |\psi\rangle)$ for varying k. Of course this relation can be inverted in order to answer the aforementioned question:

$$K_j = \frac{\ln\left(\frac{2}{\delta}\right)(b-a)^2}{2\epsilon^2}. \tag{21}$$

[5] The first index for the basis, the second one for the eigenvector.

[6] In particular, the p in Eq. (5) and Eq. (11) are respectively gauge dependent and invariant.

[7] Given a linear superoperator $\mathcal{R} : L(\mathbb{C}^m) \to L(\mathbb{C}^n)$, $\|\mathcal{R}\|^H_{1\to 1} \equiv \max_{A:A=A^\dagger, \|A\|_1 \leqslant 1} \|\mathcal{R}(A)\|_1$, "where $\|\|_1$ on superoperators is defined to be the ∞ norm induced by the trace norm $\|\|_1$ on $L(\mathbb{C}^m)$ and $L(\mathbb{C}^n)$." (Magesan et al., [21], page 2)

[8] Neither is trying to bound the average error rate given r [46].

This expression seems independent of J_j, but $b-a$ and ϵ implicitly depend on it. $b-a$ is an indicator of how much the survival probabilities fluctuate, and this is crescent with J_j without diverging [21], while ϵ can be relaxed for big J_j given the exponential decay. In a reasonable regime, it was shown that [21]

$$\boxed{K_j \sim 10^4 - 10^5}. \tag{22}$$

Although this is actually independent of n and therefore efficient, so many sequences are hardly ever used for an RB experiment. Some [11; 16; 50] have shown that K_j can be orders of magnitude smaller in the presence of only Markovian noise[9,10]. This hypothesis does not however really work when it comes to experiments [33; 3].

By not disregarding time correlated noise, one obtains the following analysis [3], which treats only coherent errors and was experimentally verified [23]. Given CRB sequences as in Eq. (9) and only 1 qubit[11], each $\Lambda_{j,k,i}$ is set to be unitary and of the form $\exp\left\{-i\delta_{j,k,i}\hat{Z}\right\}$, with $\hat{Z}$ the Pauli Z gate[12]. $\delta_{j,k,i}$ is Gaussian distributed and three different correlation schemes are identified, imposing null correlations among different js and ks. For a (j,k) there can be:

1. Markovian noise. All δ_i are Gaussian i.i.d. with 0 mean and standard deviation σ.
2. DC noise. $\delta_i = \delta \,\forall\, i \in [0, J_j]$. δ is Gaussian distributed with 0 mean and standard deviation σ.
3. Block correlated noise. As the previous case, but δ is sampled again from the same Gaussian distribution after a certain correlation length cc.

Given these assumptions it is proven that [13]

$$1 - \langle \mathrm{PDF}[F(J_j)] \rangle = O(J_j\sigma^2) + O(J_j^2\sigma^4), \tag{23}$$

where PDF stands for probability density function. In particular, disregarding $O(\sigma^4)$ corrections, $1-\mathrm{PDF}[F(J_j)]$ is Γ distributed:

$$\Gamma_{\alpha,\beta}(x) \equiv \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \approx 1 - \mathrm{PDF}[F(J_j)] \tag{24}$$

with $\Gamma(\alpha)$ Euler's gamma function, $\alpha,\ \beta > 0$.[14] For $J_j\sigma^2 \lesssim 1$, $J, cc \gg 1$ one arrives at Table 1. $\mathrm{PDF}[F(J_j)]$ can be numerically probed by the distribution of the trace fidelity. The trace fidelity is defined as

$$\langle F(J_j) \rangle_{k,\eta} \equiv \left\langle \left| \frac{1}{2} \mathrm{Tr}\left(S_{k,j}^\dagger \tilde{S}_{k,j,\eta} \right) \right|^2 \right\rangle_{k,\eta} = \frac{1}{4\tilde{n}K_j} \sum_{k=1,\eta=1}^{K_j,\tilde{n}} \left| \mathrm{Tr}\left(\tilde{S}_{k,j,\eta} \right) \right|^2 \tag{25}$$

and is the proxy thanks to which $F(J_j)$ can be estimated [3]. $\tilde{n}$ is the number of noise realizations. Averaging over $\tilde{n}$ is a way of keeping in mind that, experimentally, $\tilde{S}_{k,j}$ is sampled multiple times and the noise process disturbing the sequence does not have to be same for each sampled point, even though it is assumed to maintain its DC, Markovian or block correlated characteristics. By observing the distribution of $\mathrm{PDF}[F(J_j)]$, the noise correlation length can be inferred. This model gives rise to K_js that "increase rapidly as confidence bounds tighten. These sample sizes are much larger than typically employed in experimental settings [..]" (Ball et al., [3], page 8). Quantitative expressions for K_j error bars can be found in the appendix of Ball et al. [3][15].

1.2 Direct Randomized Benchmarking

CRB has some disadvantages. The RB number defined in Eq. (17) is an indication of the error per Clifford gate. Clifford gates, however, are usually not natively available to an experimental setup, especially with $n > 1$

[9] The $1/f$ noise model was also investigated. The interested physicist shall read the reference [11].
[10] K_j can be smaller also by disregarding noise, but with the addition of prior information [14].
[11] This framework can be extended to more qubits [36].
[12] The same reference [3] provides a generalization to errors along X, Y and Z.
[13] For a general treatment the initial state is dropped.
[14] $\mathbb{E}[\Gamma_{\alpha,\beta}] = \alpha/\beta$ and $\mathbb{V}[\Gamma_{\alpha,\beta}] = \alpha/\beta^2$.
[15] Interesting methods involving engineered noise and dynamically corrected gates had also the aim to estimate noise correlation lengths [9].

	Markovian noise	DC noise	Block-correlated noise
α	$\frac{3}{2}\tilde{n}$	$\frac{3}{2}$	$\frac{3}{2}J_j/(cc-1)$
β	$\frac{2}{3}J_j\sigma^2/\tilde{n}$	$\frac{2}{3}J_j\sigma^2$	$\frac{2}{3}(cc-1)\sigma^2$
$\mathbb{E}$	$1-J_j\sigma^2+\frac{2}{3}J_j^2\sigma^4$	$1-J_j\sigma^2$	$1-J_j\sigma^2$
$\mathbb{M}$	$1-J_j\sigma^2\left(1-\frac{2}{3\tilde{n}}\right)+\frac{2}{3}J_j^2\sigma^4$	$1-\frac{1}{3}J_j\sigma^2$	$1-J_j\sigma^2\left(1-\frac{2}{3}\frac{cc-1}{J_j}\right)$
$\mathbb{V}$	$\frac{2}{3}J_j^2\sigma^4/\tilde{n}$	$\frac{2}{3}J_j^2\sigma^4$	$\frac{2}{3}J_j(cc-1)\sigma^4$
$\mathbb{S}$	$-2\sqrt{2/3\tilde{n}}$	$-2\sqrt{2/3}$	$-2\sqrt{2(cc-1)/3J_j}$

Table 1: Asymptotic α, β, expected value, mode, variance, skew for PDF$[F(J_j)]$, for different noises. Taken from [3].

qubits[16], and they have to be compiled in $O(n^2/\log n)$ native gates [1]. Therefore, the noise of a Clifford gate is the result of the noise processes affecting the native gates with which the Clifford gate is compiled, thence errors become gate dependent in CRB. On top of this, the more qubits are involved, the noisier a compiled Clifford element will result. All things considered, CRB requires compilation, which becomes increasingly difficult with n, and does not reflect the native properties of the setup. Because of this, the Direct Randomized Benchmarking protocol was developed [34]. A relevant drawback affects DRB though: the theory supporting the protocol is very limited, also because of DRB novelty. Some of this paper work goes in the direction of trying to extend some CRB results to DRB.

1.2.1 The protocol and its results

It is required that the native gate set generates the Clifford group C_n for n qubits. The notion of *layer* will be used: a layer is a depth-1 circuit of native gates[17]. These are the steps for DRB [34].

1. Choose some lengths $J_1 < J_2 < \cdots < J_{\text{last}}$. The initial state shall always be $|0\rangle^{\otimes n}$
2. $\forall\, J_j$:
   (a) Choose a $K_j \gg 1$.
      i. Sample uniformly a $C_0 \in C_n$. $C_0\,|0\rangle^{\otimes n}$ is effectively a stabilizer state preparation.
      ii. Sample a J_j-layer circuit U_{J_j}[18].
      iii. Implement a circuit U_s such that $U_s U_{J_j} C_0\,|0\rangle^{\otimes n} = |s\rangle$ is a known state in the computational basis. For convenience reasons, $|s\rangle$ will be in this work set to $|0\rangle^{\otimes n}$.
      iv. Measure experimentally ${}^{\otimes n}\langle 0|\,\tilde{U}_0\tilde{U}_{J_j}\tilde{C}_0\,|0\rangle^{\otimes n}$ and store the result.
   (b) Average the previous quantity over K_j to calculate $F(J_j)$.

Two differences in regards to CRB are already to be noticed. First, $\tilde{U}_0\tilde{U}_{J_j}\tilde{C}_0$ does not have to and is not sampled to be equal to $\mathbb{1}$. Second, the effective length of a sequence is greater than K_j given C_0 and U_0. Notwithstanding these points, from experimental evidence $F(J_j)$ decays exponentially:

$$F(J_j) = A + Bp^{J_j}. \tag{26}$$

[16] The order of C_n is $2^{n^2+n+2}\left(2^n-1\right)\prod_{j=1}^{n-1}\left(4^j-1\right)$ [25].
[17] In other words, in a layer all gates are executed in parallel.
[18] The protocol accepts any probability distribution for native gate sampling. Here uniformity is assumed for the sake of simplicity.

The interpretation of these parameters does not change: A and B depend on SPAM errors and p on gate errors. Arguments comprising only stochastic errors [34] suggest adopting the error rate

$$r \equiv \frac{4^n - 1}{4^n}(1 - p). \tag{27}$$

"[..] Linking r to a formal notion of gate error rate is subtler with coherent errors, [..], as will be discussed in future work" (Proctor et al., [34], page 5). In the meantime, this r shall be called *the DRB number*.

1.2.2 Survival probability errors and distributions

Can the results in Section 1.1.3 be extended to DRB? At the time of writing no scientific article supporting such claims is known. Not everything can be treated in a master project, worse luck. As a consequence, part of this is focus just on understanding whether, for a single qubit subjected to coherent noise, the infidelity measured via DRB would be Γ distributed as well. An additional effort was also spent to directly fit a model resembling the one in Table 1.

To begin with, a DRB sequence structure is much richer than a CRB one. For an arbitrary native gate set generating the Clifford group C_n not much can be said yet. One way to prove the validity of Eq. (26) for a chosen gate set and arbitrary noise process would be to construct the approximate or exact 2-designs [6; 15; 24]. However, the nature of C_0, U_{J_j} and U_0 complicate the tractation.

An alternative way to analyze the problem is to restrict our choice of native gate set to be a subset of C_n. Then some of the steps in Ball et al. [3] will remain valid (see Appendix D for more details). In particular, native Clifford gates conserve the random walk framework in DRB. The coherent errors $\exp(i\sigma Z)$ expanded at first order in σ, when combined with the Clifford gates, gives rise to a sequence of effective errors which can be seen as a random walk in the XYZ Pauli space [3; 9; 23]. Nonetheless, the different sequence structure of DRB has not allowed so far to exhibit a formal proof that, at first order $\Gamma_{\alpha,\beta}(x) \approx 1 - \mathrm{PDF}[F(J_j)]$. Going in details, the main obstacle is that an ideal CRB sequence is effectively a unit operation i.e. $S_{j,k} = \mathbb{1}$, while for DRB the unique constraint is $\left|\langle 0| U_0 U_{J_j} C_0 |0\rangle\right| = 1$. This last quantity is the one measured by a DRB experiment and the one considered in simulations. For CRB on the other hand, the same sequence permits to measure the trace of the whole circuit: this is achieved by just changing initial state. While the relevant quantity for CRB was mentioned in Eq. (25), here

$$\langle F(J_j)\rangle_{k,\eta} \equiv \frac{1}{\tilde{n}K_j} \sum_{k=1,\eta=1}^{K_j,\tilde{n}} \left|\langle 0| \tilde{U}_0 \tilde{U}_{J_j} \tilde{C}_0 |0\rangle\right|^2 \tag{28}$$

must be dealt with[19]. Notwithstanding this, simulations (Section 3.2) and experiments (Section 6) show that this quantity is Γ distributed, although it was not possible to obtain a thorough model like the one presented in Table 1 by Ball et al. [3].

[19] Nothing changes by choosing $|s\rangle \neq |0\rangle$. The whole sequence will be in general different and only its $0s$ element will be measured.

2 Gate Set Tomography: Theory

Gate Set Tomography (GST) is a protocol whose aim is the characterization of state preparation, measurements and gates of a quantum computing setup. In contrast to RB and in analogy to tomographic routines, GST main task is to achieve a complete gate set (or setup) description and not just computing average quantities[20]. In addition, GST is *calibration free*. This and many other points will be explained in the rest of this section, which is mostly a summary of Nielsen et al. [27]. Given the complexity of the topic, some results will be provided here without proof. Please read the original article in case a better insight is desired.

2.1 Preliminaries

Given a Hilbert space $\mathcal{H}$ of dimension d, states can be written as $d \times d$ complex matrices acting on $\mathcal{H}$. The set of these matrices is a vector space $\mathcal{B}(\mathcal{H})$ called *Hilbert-Schmidt space*. Such a space is equipped with the inner product $\langle A, B\rangle \equiv \mathrm{Tr}\left(A^\dagger B\right)$. Therefore B can be represented by a d^2-column vector $|B\rangle\rangle$, the *superket*, and $A^\dagger$ by a d^2-row vector $\langle\langle A|$, the *superbra*, and $\langle\langle A|B\rangle\rangle \equiv \mathrm{Tr}\left(A^\dagger B\right)$. Given the hermiticity of states, only the d^2-real subspace of $\mathcal{B}(\mathcal{H})$ that trivializes the $\dagger$ operation is relevant. Therefore a d^2-real representation for physical states is admitted. It follows that physical superoperators for $\mathcal{H}$ (or operators on $\mathcal{B}(\mathcal{H})$) have a $d^2 \times d^2$-real matrix representation[21], although not every superoperator is a *quantum channel*, because they must be completely positive and trace preserving (CPTP). A gate set $\mathcal{G}$ is an object of the following kind

$$\mathcal{G} = \left\{ \left\{ |\rho^{(i)}\rangle\rangle \right\}_{i=1}^{N_\rho} ; \{G_i\}_{i=1}^{N_\mathrm{G}} ; \left\{ \langle\langle E_i^{(m)}| \right\}_{m=1,i=1}^{N_\mathrm{M}, N_\mathrm{E}^{(m)}} \right\}. \tag{29}$$

N_ρ state preparations $|\rho^{(i)}\rangle\rangle$ are present; there are also the measurement outcomes $\langle\langle E_i^{(m)}|$: m indexes the N_M measurement and i the $N_\mathrm{E}^{(m)}$ outcomes for the measurement m; lastly come N_G logic operations G_i, which have to be intended as layers (depth-1 circuits). They act on superkets as

$$\text{operation} : |\rho\rangle\rangle \to G_{\text{operation}}|\rho\rangle\rangle. \tag{30}$$

As similarly mentioned in the previous sections, given an invertible superoperator M, an equivalent gate set (or representation) is provided by

$$\begin{aligned} \langle\langle E_i^{(m)}| &\to \langle\langle E_i^{(m)}|M^{-1} \\ |\rho^{(i)}\rangle\rangle &\to M|\rho^{(i)}\rangle\rangle \\ G_i &\to MG_iM^{-1}. \end{aligned} \tag{31}$$

It can be noticed immediately that physical observables do not depend on M. As in electromagnetism, additional degrees of freedom underlying the mathematical formalization of the physical nature emerge. They are purely mathematical and named *gauge degrees of freedom*.

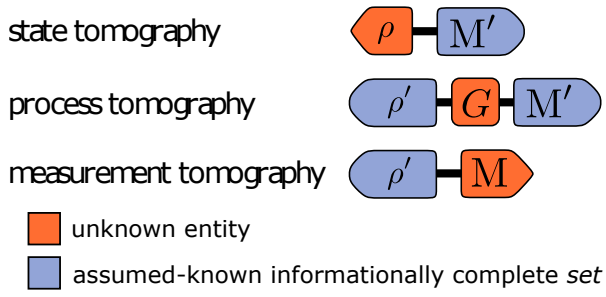


Figure 1: Previous tomographic protocols, from [27].

Precedent tomographic algorithms such as state tomography, process tomography and measurement tomography fulfilled the purpose of characterizing state preparations, gates and measurement respectively assuming complete knowledge of measurements, measurement and states and measurements. *Calibration errors* affect these protocols. In other words, the tomographic errors are at least as large as the errors on the fiducial states and measurements. GST offers a calibration free tomography, which considers each element of a gate set relative to all the others. As better clarified in Section 2.2, GST results describe a gate set up to a gauge transformation. Selecting the most appropriate gauge is an addition to the GST protocol (see Section 2.3.2).

[20] Some experimental studies comparing GST and RB [4; 7].
[21] Of course there are other representations: the operator-sum representation is one of the most known.

2.2 The GST protocol step by step

2.2.1 Linear GST

Linear GST (LGST) is a propedeutic algorithm to GST. It is already calibration free and is the first step of the more advanced Long-sequence GST (Section 2.2.2). This protocol assumes:

1. No finite sample error in estimated probabilities.
2. The existence of informationally complete sets of fiducial states and measurement effects[22] $|\rho'_i\rangle\rangle$ and $\langle\langle E'_j|$. These sets do not have to be the same of Eq. (29).
3. The cardinality of these two sets is d^2, i.e. they are not overcomplete.

These assumptions will be relaxed later on. One can start with the idea of applying each gate G_k to all the states and project this with every effect. The matrix

$$[P_k]_{i,j} = \langle\langle E'_i|G_k|\rho'_j\rangle\rangle \tag{32}$$

is a measurable object. One can define

$$A \equiv \begin{pmatrix} \langle\langle E'_1| \\ \langle\langle E'_2| \\ \vdots \\ \langle\langle E'_{d^2}| \end{pmatrix}, \qquad B = \left(|\rho'_1\rangle\rangle, |\rho'_2\rangle\rangle, \cdots, |\rho'_{d^2}\rangle\rangle\right), \qquad \tilde{\mathbb{1}} \equiv AB \tag{33}$$

where only $\tilde{\mathbb{1}}$ is measurable. It follows that

$$P_k = AG_kB, \qquad \tilde{\mathbb{1}}_{i,j} = \langle\langle E'_i|\rho'_j\rangle\rangle. \tag{34}$$

Straightforwardly one observes that

$$\boxed{G_k = B\tilde{\mathbb{1}}^{-1}P_kB^{-1}.} \tag{35}$$

This is it! G_k was characterized with only measurable quantities, up to an unmeasurable gauge transformation B, which can be fully specified only when a representation is chosen. $\langle\langle E_l^{(m)}|$ and $|\rho^{(l)}\rangle\rangle$ are yet to be treated. The following objects can be defined:

$$\left[\vec{R}^{(l)}\right]_j = \langle\langle E'_j \mid \rho^{(l)}\rangle\rangle, \qquad \left[\vec{Q}_l^{(m)\,T}\right]_j = \langle\langle E_l^{(m)} \mid \rho'_j\rangle\rangle. \tag{36}$$

These are also measurable. They can be rewritten as

$$\vec{R}^{(l)} = A|\rho^{(l)}\rangle\rangle, \qquad \vec{Q}_l^{(m)\,T} = \langle\langle E_l^{(m)}|B. \tag{37}$$

This leads to

$$\boxed{|\rho^{(l)}\rangle\rangle = B\tilde{\mathbb{1}}^{-1}\vec{R}^{(l)}, \qquad \langle\langle E_l^{(m)}| = \vec{Q}_l^{(m)T}B^{-1}.} \tag{38}$$

Exactly like Eq. (35), the states and the measurements are fully specified with only measurable quantities, up to an unmeasurable gauge transformation B! The physical properties of the gate set are now fully determined.

Now it is time to relax some of the hypotheses. Overcomplete sets are not a problem; on the contrary they allow for redundancy. In this case A, B and $\mathbb{1}$ are not invertible but they are full rank and admit a pseudo inverse[23]. Having informationally incomplete state and measurement sets is usually not problematic. Typically experimental setups possess only one POVM and one state preparation. Nonetheless, additional states and measurement effects can be obtained given the native ones and circuits compiled in the native logic operations:

$$\langle\langle E_{\text{added}}| = \langle\langle E_{\text{native}}|G^{-1}, \qquad |\rho_{\text{added}}\rangle\rangle = F|\rho_{\text{native}}\rangle\rangle. \tag{39}$$

[22] An effect set is informationally complete if any state $|\zeta\rangle\rangle$ can be fully described by $\{p_j\}_j \equiv \{\langle\langle E_j|\zeta\rangle\rangle\}_j$. The analogous definition holds for state preparations.

[23] Read [27] for more details.

These G^{-1} and F are called in the literature *fiducial circuits*[24]. They decrease the degrees of freedom of LGST but not its structure. The last obstacle is finite sample errors. LGST minimizes by construction the squared differences between predicted probabilities $p_j^{(\text{gate set})}$ and observed frequencies $f_j^{(\text{observed})}$ for the outcomes of LGST circuits, i.e.

$$\text{sum of squared differences} = \sum_j \left(p_j^{(\text{gate set})} - f_j^{(\text{observed})} \right)^2. \tag{40}$$

Nonetheless, least squares optimization is not justified. One should take into account maximum likelihood estimation (MLE) methods. The likelihood function $\mathcal{L}$ is the probability of collecting the acquired data given the validity of the model (in this case the gate set):

$$\mathcal{L}(\text{gate set}) = \Pr(\text{data}|\text{gate set}). \tag{41}$$

Another issue of finite sampling is that

$$f^{\text{observed}} = p^{\text{gate set}} \pm O\left(\frac{1}{\sqrt{N}}\right) \tag{42}$$

if the outcome of every LGST circuit is measured N times. This is a very poor scaling: if an uncertainty threshold of 10^{-4} is required, 10^8 measurements are entailed.

2.2.2 Long-sequence GST

The LGST finite sampling scaling inconvenience can be solved by circuits with depth > 1 in which gates are repeated: $\Pr = \langle\langle E|G_k G_k G_k G_k|\rho\rangle\rangle$. Circuits of depth $O(L)$ enable $O(1/L)$ estimation for the gate set parameters, which are amplified by deeper circuits. Later on in Section 2.2.3 it will become clearer what these parameters are. The Long-sequence GST protocol will be now made explicit as in [27].

1. Prepare a state $|\rho'_k\rangle\rangle$, i.e. prepare a native state followed by a fiducial circuit.
2. Perform p repetitions of a short circuit g called *germ*.
3. Perform a measurement $\langle\langle E_i'^{(m)}|$ by performing a fiducial circuit and then a native POVM measurement.

The germ repetitions are meant to amplify errors. For example, if an angle can be measured up to ϵ precision and G_x overrotates by θ, with this protocol θ can be measured within $\pm\epsilon/p$ precision. Not every error is amplified by rerunning a single gate, like axis misalignments, but any physical gate error can be amplified. SPAM errors cannot, since SPAM operations appear only once in every circuit. How are the germs to be selected, in order to amplify every possible error? It can be demonstrated that, SPAM errors excluded, each physical error lies in a subspace of the gate set parameters; gauge parameters find themselves out of this subspace. Thence this subspace has to be spanned by the germs. Achieving this requirement yields an *amplificationally complete* set of germs. A more detailed tractation is found in the appendix of the original article [27].

After the germs, the powers p have to be chosen. In the original article the authors argue that a logarithmic spacing is a good choice. The same can be said for the circuits depths L. Thus both p and L will be 1, 2, 4, 8, 16, ...[25] Where to stop though? "GST experiments are tailored to a hardware's capabilities." (Nielsen et al., [27], page 19). If every gate has a decoherence rate η, there is little to gain from probing circuits with $L > O(1/\eta)$.

The last part of the Long-sequence GST algorithm deals with finding the optimal physical parameters for $\mathcal{G}$, obtainable by maximising the loglikelihood, which is a very complicated function. GST developers, after dealing with this maximisation problem for years, devised the following algorithm. Given a GST experiment, let s index the circuits to be run, N_s be the times s is run, m_s the number of all possible outcomes of circuit s, N_{s,β_s} the number of times the outcome β_s was observed for s. The loglikelihood is

$$\log \mathcal{L} = \sum_s \log \mathcal{L}_s = \sum_{s,\beta_s} N_{s,\beta_s} \log\left(p_{s,\beta_s}\right) = \sum_{s,\beta_s} N_s f_{s,\beta_s} \log\left(p_{s,\beta_s}\right) \tag{43}$$

[24] Despite the name, no representation for them is fixed.
[25] In practice any other base is ok.

with $f_{s,\beta_s} \equiv N_{s,\beta_s}/N_s$ and p_{s,β_s} is the probability predicted by $\mathcal{G}$ of measuring β_s for circuit s. Let $\vec{\theta}$ be the vector of parameters to be optimized, $\mathcal{D}$ the data set, Truncate$(\mathcal{D}, L)$ the data subset collected from circuits no longer than L and Argmin$(S, \mathcal{G}, \mathcal{D}, \vec{\theta}_1)$ the function yielding the $\vec{\theta}$ for which a statistical function $S(\mathcal{G}(\vec{\theta}), \mathcal{D})$ is minimal and whose optimizer is seeded at $\vec{\theta}_1$. The first seed $\vec{\theta}_0$ is provided by LGST and $\mathcal{D}_0$ is the full data set. Then the optimization algorithm for Long-sequence GST is, as stated in [27],

$$\begin{aligned}
&\vec{\theta} \leftarrow \vec{\theta}_0 \\
&\textbf{for } L \in 1, 2, 4, 8, 16 \ldots \textbf{do} \\
&\quad \mathcal{D} \leftarrow \text{Truncate}\,(\mathcal{D}_0, L) \\
&\quad \vec{\theta} \leftarrow \text{Argmin}(\chi^2, \mathcal{G}, \mathcal{D}, \vec{\theta}) \\
&\textbf{end for} \\
&\vec{\theta} \leftarrow \text{Argmin}(-\log \mathcal{L}, \mathcal{G}, \mathcal{D}_0, \vec{\theta}).
\end{aligned} \tag{44}$$

The last $\vec{\theta}$ is the vector with the best gate set parameters. This structure comes from the authors' experience. They observed that χ^2 defined like

$$\chi^2 = \sum_{s,\beta_s} N_s \frac{\left(p_{s,\beta_s} - f_{s,\beta_s}\right)^2}{p_{s,\beta_s}} \tag{45}$$

is a good enough proxy for the likelihood, despite being slightly biased[26]. Optimizing χ^2 at every stage except the last one speeds up the calculations, because χ^2 is a simpler function than $\log \mathcal{L}$. What is more, χ^2 is more well-behaved than $\log \mathcal{L}$. This helps avoiding local maxima at every step of the algorithm. In the end $\log \mathcal{L}$ is maximized[27] to find the globally best parameters. The next section tries to clarify what is meant with "parameters".

2.2.3 Advanced Long-sequence GST

Another pivotal point is the role of gate set parameters and gate set models[28]. A gate set $\mathcal{G}$ as defined in Eq. (29) is a point in a real matrix linear space $\mathcal{M}$ with dimension

$$N_\text{e} = d^4 N_\text{G} + d^2 \left(N_\rho + \sum_{m=1}^{N_\text{M}} N_\text{E}^{(m)} \right), \tag{46}$$

with N_G the number of available gates, N_ρ the number of states and $\sum_{m=1}^{N_\text{M}} N_\text{E}^{(m)}$ the number of measurement results, where the summation runs over the N_M measurements. N_e accounts for the entries in each matrix (gate), vector (state) and dual vector (measurement result) in $\mathcal{G}$. The parameter space is linear, will be called $\mathcal{P}$ and has dimension $N_\text{p} \leq N_\text{e}$. A gate set model is a map $W : \mathcal{P} \to \mathcal{M}$. Long-sequence GST finds a $\vec{\theta} \in \mathcal{P}$ which is the optimal set of parameters. From the optimal gate set $W(\vec{\theta}) = \mathcal{G}$ physical quantities are computed. For any model $N_\text{p} = N_\text{p}^\text{gauge} + N_\text{p}^\text{nongauge}$ holds. These two quantities are respectively the number of gauge and physical parameters. Gauge degrees of freedom trace orbits in the parameter space when keeping fixed nongauge d.o.f.s.

The simplest case for a gate set model is the *fully parametrized model*. This has $\mathcal{P} = \mathcal{M}$ and $W(x) = x$. Trivially $N_\text{p} = N_\text{e}$ for this model. In most cases it has d^2 gauge d.o.f.s [27]. More physical models are trace preserving (TP). This requirement, together with the representation for $\mathcal{G}$ introduced right before Eq. (29), forces the first row of every superoperator to be [1, 0, ... 0] and the first element of every state preparation to be $1/\sqrt{d}$ [27]. Once these constraints are provided,

$$\text{for the TP model} \qquad N_\text{p} = N_\text{G} d^2 \left(d^2 - 1\right) + N_\rho \left(d^2 - 1\right) + \left(\sum_{m=1}^{N_\text{M}} N_\text{E}^{(m)} \right) d^2 \tag{47}$$

[26] An attentive reader can identify the necessity of regularizing χ^2 for small p_{s,β_s}. More details in [27].
[27] The exact procedure can be found in the original article.
[28] The original article also treats an additional algorithm to reduce overcompleteness and speed up a GST experiment. This will be skipped in this work.

where $d^2 - d$ of them are gauge parameters, since all the allowed TP gauge transformations are TP matrices (a full rank $d^2 \times d^2$ matrix has d^2 d.o.f.s, the eigenvectors, and $-d$ comes from fixing the first column). The TP model causes GST no complications; on the contrary complete positivity (CP) does. The CP constraint is non linear and it interacts in an unfortunate way with gauge freedom, because non-unitary gauge transformations do not preserve CP [27]. It can be exhibited that the CPTP model and the TP one have the same N_p. The same relation applies for the CP and the full one.

CP also complicates MLE and error bar calculations. As a rule of thumb, both CPTP and TP models should be included when analysing GST data. If they provide consistent results, the CPTP one should be used, because more physical. If not, either the CPTP optimisation did not find the global maximum or non-Markovian noise affected the experiment.

2.3 GST estimates analysis

2.3.1 Goodness of the fit

GST can describe and fit gate set models with Markovian noise. On this basis, χ^2 statistics allows the goodness of the fit to be evaluated. Given a GST experiment with N_exp different circuits for a quantum setup with only 1 native POVM, which features only 2 possible outcomes, the GST data will have all in all N_exp independently measured frequencies. With more complicated setups, the number of independent degrees of freedom in the dataset is $N_\mathrm{o} \geq N_\mathrm{exp}$. Calling

$$k \equiv N_\mathrm{o} - N_\mathrm{p}^\mathrm{nongauge}, \tag{48}$$

under Markovian noise one has

$$2\left(\log \mathcal{L}_\mathrm{max} - \log \mathcal{L}\right) \sim \chi^2_k. \tag{49}$$

This is Wilk's theorem [27]. χ^2_k is the χ^2 distribution with k d.o.f. $N_\mathrm{p}^\mathrm{nongauge}$ is the number of physical parameters for the gate set model selected by the tomographer. $\log \mathcal{L}$ is the loglikelihood of this model, while $\log \mathcal{L}_\mathrm{max}$ is the loglikelihood of a *maximal model* with N_o parameters. The maximal model always fits perfectly the data. If the Markovian noise hypothesis is verified, it makes sense to consider

$$N_\sigma \equiv \frac{2\left(\log \mathcal{L}_\mathrm{max} - \log \mathcal{L}\right) - k}{\sqrt{2k}}. \tag{50}$$

N_σ estimates of how many sigmas $2\left(\log \mathcal{L}_\mathrm{max} - \log \mathcal{L}\right)$ deviates from χ^2_k. $N_\sigma \gg 1$ is indication of non-markvianities having occurred. These can for instance be due to slow drifts and leakages. The same paradigm can be applied to $\log \mathcal{L}_s$ to check for model violations relative to a single circuit.

2.3.2 Gauge optimization

Certain gauge-variant metrics, like fidelity and diamond distance, are very important in the scientific dialectics. Thus, calculating the gauge that makes the GST fit the most similar to the ideal gate set is a necessity. Forcing the fitted model to have the same appearance of the ideal one may sound unfaithful, but gauge transformations cannot change any physical observables nor eliminate physical errors, (they can add and remove unphysical ones though). GST minimizes the Frobenius distance between fit and ideal model, in order to find to best gauge. Infidelity and diamond distance are not utilized by GST because not positively definite and expensive to compute respectively. Given a matrix A, the Frobenius norm is [51]

$$|A|_F \equiv \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}, \tag{51}$$

with $a_{i,j}$ entries of A. Better said, GST makes use of a weight sum of Frobenius distances between the target $\mathcal{G}$ and the estimation $\hat{\mathcal{G}}$,

$$f(\hat{\mathcal{G}}, \mathcal{G}) = \sum_{i=1}^{N_\rho} \alpha_i \left|\hat{\rho}^{(i)} - \rho^{(i)}\right|_F^2 + \sum_{i=1}^{N_\mathrm{G}} \beta_i \left|\hat{G}_i - G_i\right|_F^2 + \sum_{m=1}^{N_\mathrm{M}} \sum_{i=1}^{N_\mathrm{E}^{(m)}} \gamma_{m,i} \left|\hat{E}_i^{(m)} - E_i^{(m)}\right|_F^2, \tag{52}$$

because SPAM errors cannot be amplified and SPAM estimates are less accurate, $O(1/\sqrt{N})$ compared to $O(1/L\sqrt{N})$ for gates. Gauge optimization consists in different steps where the weight α_i, β_i and $\gamma_{m,i}$ are changed and subsequent optimizations are run. The interested physicist shall read the original aricle [27] for more details.

2.3.3 Error bars

In order to assign uncertainties to GST results, the authors adopted a confidence region approach presupposing *local asymptotic normality* (LAN). LAN means that $\mathcal{L}$ for the collected data is Gaussian and that the likelihood for other possible data sets that were not sampled is also Gaussian with the same covariance matrix. Under these hypotheses, every region with confidence α is an ellipsoid centered around the point found with the MLE. Changing α means rescaling the ellipsoid. From this error bars are assigned to the gate set parameters. The confidence region can be identified in two ways:

1. by calculating the Hessian matrix of $\log \mathcal{L}$ at the MLE and with this circumscribe the confidence region;
2. by simulating further data sets, with either the MLE gate set or directly with the experimental data, and using their scatter to define the confidence region. This method is called *bootstrapping*.

Either way has to deal with gauge d.o.f.s while calculating error bars. It was decided to fix the gauge for the region of interest before assigning error bars. This excludes the possibility of very large error bars due to gauge variations in the imaginary situation where all gauge invariant properties are known exactly. In this scenario in fact, all the measurable characteristics of a gate set are certain and it would not be sensible to have error bars.

3 Direct Randomized Benchmarking: Simulations

The purpose of these simulations was to study the infidelity distributions of 1 qubit DRB experiments. DRB circuits were sampled with the library pyGSTi[29] [29] and Qiskit [2] was utilized to track the state evolution of such sequences, affected by coherent noise as defined in Section 1.1.3. Numerical fits were computed thanks to the SciPy library [8]. All computational variables were of the `float64` datatype. The values for J_j were [8, 16, 32, 64, 128]; the correlation length cc (analogous to the one in Section 1.1.3) was allowed to be [1, 30, 90, 300] and [10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}] for σ. 300 is just a number > 128, the maximum sequence length, so to have DC noise. $cc = 1$ means Markovian noise. Depending on the value of J_j, any $cc \neq 1, 300$ gives rise to either block correlated or DC noise. For each truple (J_j,cc, σ) $K = 500$ circuits were sampled, i.e. $K_j = K\,\forall j$, and $\tilde{n} = 50$ noise realizations were taken into account The chosen gate set included $+\pi$ and $+\pi/2$ rotations along X, Y and Z. All of the following data are available in the repository[38].

3.1 Exponential decay

To begin with, the law

$$F(J_j) = A + Ap^{J_j}. \tag{53}$$

was considered. This differs from Eq. (26). It was observed that with the additional degree of freedom B the quality of the fit degraded. The choice of eliminating one degree of freedom (d.o.f.) is also justified by the absence of SPAM errors in the simulations: state preparation and measurement were simulated error less and deviations from the expected value 0.5 were expected to be small. To remove one d.o.f., one should fix either A or B to 0.5, and just fit for p and either B or A. Now one question arises: Shall B or A be set constant? This is arbitrary. In order to reduce the arbitrariness of the choice but still have one fewer d.o.f., these parameters were fixed to be equal[30]. For the weighted fit, it was established for each $F(J_j)$ to have the uncertainty $d_{F(J_j)} = \frac{\sigma_K}{\sqrt{K}}$, with σ_K the standard deviation of the K-fold distribution (1 data point per circuit) at length J_j. This model was fit with the function `scipy.optimize.curve_fit`, which performs least squares optimization. Now the main results shall be here summarized.

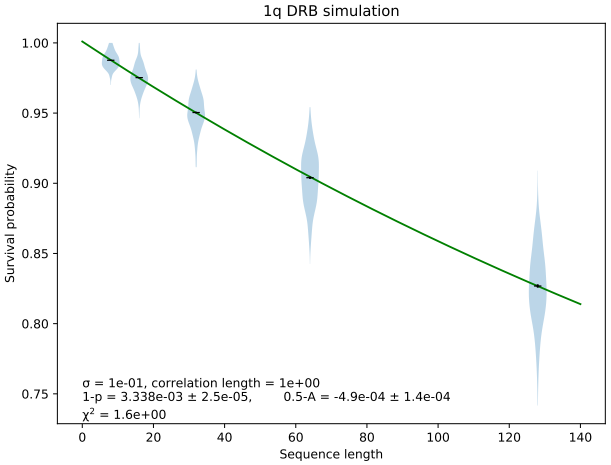


Figure 2: Simulated DRB decay, example 1.

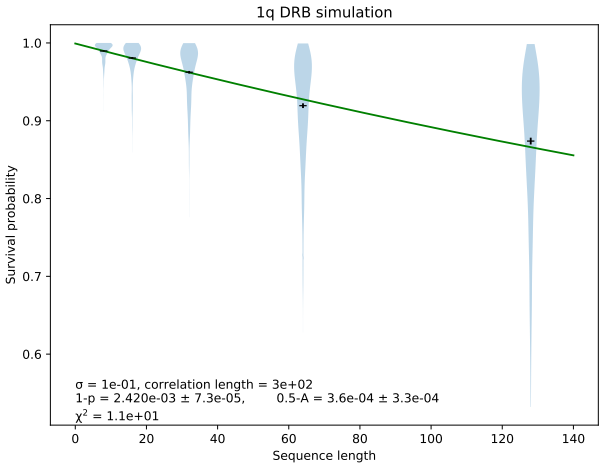


Figure 3: Simulated DRB decay, example 2.

The Figs. 2 and 3 portray the law in Eq. (53) for the same $\sigma = 0.1$ but with different cc, 1 and 300 respectively, i.e. Markovian and DC noise. The DC violin plot[31] features long tails. These are due to noise correlations. Moreover, the tails become longer (for a given σ) for greater cc. This general behaviour can be noted from all the plots in Appendix A. For CRB this is a well known trend [3; 36].

$$\chi^2 \equiv \sum_{\text{data points}} \frac{(\text{fit} - \text{data})^2}{\text{variance}} \tag{54}$$

is another quantity that increases with cc according to the simulations (see Appendix A). For $cc = 1$ $\chi^2 =$

[29] As its name says, this library allows also GST circuits sampling.
[30] All of this was of course arbitrary.
[31] If not familiar, a violin plot provides an indication (relative density if you want) of how many points were measured to be of a specific value.

$O(1)$, while for $cc = 300$ $\chi^2 = O(10)$ with 3 d.o.f. As it will be understood in the next section, the less cc is, the more similar the underling distributions of the shadowed areas in the violin plots are to Gaussians. More figures can be found in Table 2. Another aspect emerging from these data is SPAM error: 0.5-A is ≈ 3 sigmas away from 0 on average. The state preparation and the measurements were simulated errorless, though. This shows that certain error contributions do not vary with J_j. This stems from the coherent errors on C_0 and U_0. For 1 qubit and given the chosen gate set, they are compiled into at most 4 gates each.

sigma	cc	1-p	0.5-A	chi squared
1×10^{-1}	1	$3.338\times10^{-3} \pm 2.5\times10^{-5}$	$-4.9\times10^{-4} \pm 1.4\times10^{-4}$	1.6
1×10^{-1}	30	$2.741\times10^{-3} \pm 5.3\times10^{-5}$	$-5.2\times10^{-4} \pm 3.0\times10^{-4}$	2.7
1×10^{-1}	90	$2.518\times10^{-3} \pm 6.7\times10^{-5}$	$6\times10^{-5} \pm 3.3\times10^{-4}$	4.5
1×10^{-1}	300	$2.420\times10^{-3} \pm 7.3\times10^{-5}$	$3.6\times10^{-4} \pm 3.3\times10^{-4}$	11.0
1×10^{-2}	1	$3.334\times10^{-5} \pm 2.2\times10^{-7}$	$-4.0\times10^{-6} \pm 1.4\times10^{-6}$	0.24
1×10^{-2}	30	$2.967\times10^{-5} \pm 5.8\times10^{-7}$	$-8.9\times10^{-6} \pm 3.1\times10^{-6}$	6.4
1×10^{-2}	90	$3.014\times10^{-5} \pm 9.0\times10^{-7}$	$-8.7\times10^{-6} \pm 3.9\times10^{-6}$	8.9
1×10^{-2}	300	$3.041\times10^{-5} \pm 9.9\times10^{-7}$	$-1.09\times10^{-5} \pm 3.7\times10^{-6}$	8.0
1×10^{-3}	1	$3.359\times10^{-7} \pm 2.3\times10^{-9}$	$-4.7\times10^{-8} \pm 1.4\times10^{-8}$	3.2
1×10^{-3}	30	$2.973\times10^{-7} \pm 6.0\times10^{-9}$	$-8.4\times10^{-8} \pm 3.2\times10^{-8}$	4.1
1×10^{-3}	90	$3.029\times10^{-7} \pm 8.9\times10^{-9}$	$-9.4\times10^{-8} \pm 3.8\times10^{-8}$	6.6
1×10^{-3}	300	$3.07\times10^{-7} \pm 1.0\times10^{-8}$	$-1.17\times10^{-7} \pm 3.8\times10^{-8}$	9.4
1×10^{-4}	1	$3.339\times10^{-9} \pm 2.2\times10^{-11}$	$-4.4\times10^{-10} \pm 1.4\times10^{-10}$	0.63
1×10^{-4}	30	$2.999\times10^{-9} \pm 6.0\times10^{-11}$	$-1.00\times10^{-9} \pm 3.3\times10^{-10}$	5.7
1×10^{-4}	90	$3.009\times10^{-9} \pm 8.9\times10^{-11}$	$-9.3\times10^{-10} \pm 3.7\times10^{-10}$	9.3
1×10^{-4}	300	$3.05\times10^{-9} \pm 1.0\times10^{-10}$	$-1.17\times10^{-9} \pm 3.8\times10^{-10}$	9.3

Table 2: Parameters for the simulated DRB decays. All of the related plots can be found in Appendix A.

3.2 Gamma distributions

Instead of looking at the whole decay as in Fig. 2, one could focus on the different shaded areas of a violin plot. It is of interest whether the measured infidelities $(1 - F(J_j))$ are Γ distributed as in CRB experiments. Numerically there is strong evidence of such feature, as displayed by all the graphs in Appendix B. With the help of `scipy.stats.gamma.fit` which uses maximum likelihood estimation methods, $\forall J_j$, cc, σ the respective distribution, originating from the K infidelities, was fit by a Γ distribution. The model has an additional parameter with respect to Eq. (24):

$$\Gamma_{\alpha,\beta,x_0}(x) \equiv \frac{\beta^\alpha}{\Gamma(\alpha)}(x - x_0)^{\alpha-1}e^{-\beta(x-x_0)} \approx 1 - \mathrm{PDF}[F(J_j)]. \tag{55}$$

The offset $x_0 \approx 0$ but $\neq 0$ serves for this work the only purpose of avoiding singular behaviours of the already introduced `scipy.stats.gamma.fit`. In many cases, varying x_0 manually helped with finding a good fit. Other actions to improve the fits were setting the starting values for the optimisation according to Footnote 14, i.e.

$$\alpha = \frac{\mathbb{E}\left[\Gamma_{\alpha,\beta}\right]^2}{\mathbb{V}\left[\Gamma_{\alpha,\beta}\right]}, \tag{56}$$

$$\beta = \frac{\mathbb{E}\left[\Gamma_{\alpha,\beta}\right]}{\mathbb{V}\left[\Gamma_{\alpha,\beta}\right]}. \tag{57}$$

A interesting trend can be observed. An instance of it is now shown by the Gamma distributions for $\sigma = 10^{-4}$ and $cc = 30$ (Figs. 4 to 8). For $J_j < cc$ the distribution is highly skewed and monotonic. For $J_j > cc$ the first derivative changes sign, a maximum appears and this moves away from 0 with increasing J_j. In particular, DC noise features skewed distributions and Markovian noise rather symmetric ones. For $J_j \approx cc$ a "transition" between the two trends can be observed. From the plots in Appendix B it can be inferred that this holds for any pair (σ,cc), except (10^{-1},300) and (10^{-1},90)[32]. The reason can be that in this case $J_j\sigma^2 \gtrsim 1$ and the first order

[32]Since no $J_j \approx 90$, the exact transition around this value cannot be observed, but the general trend in regards to the distribution peak can.

expansion leading to Eq. (24) does not hold anymore. Similar facts for CRB were reported before [9; 3; 36], but they have not been used so far to estimate cc. By noticing when a Gamma distribution changes shape for different J_js, cc can be estimated. If J_j are such that $J_{i+1} = 2J_i$, the estimate will be at worst a factor of 2 distant, provided that $J_i\sigma^2 \lesssim 1$.

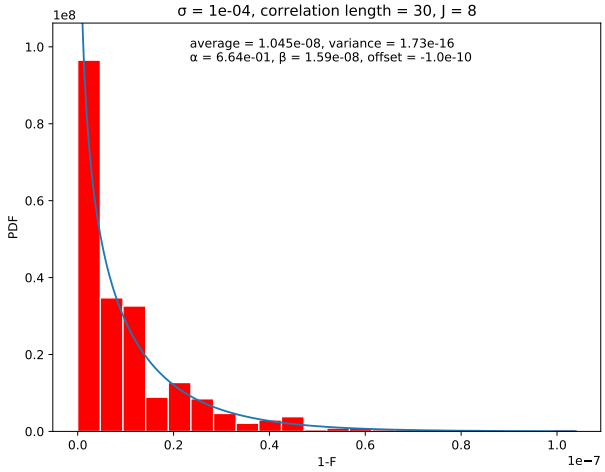


Figure 4: Gamma distribution. J = 8

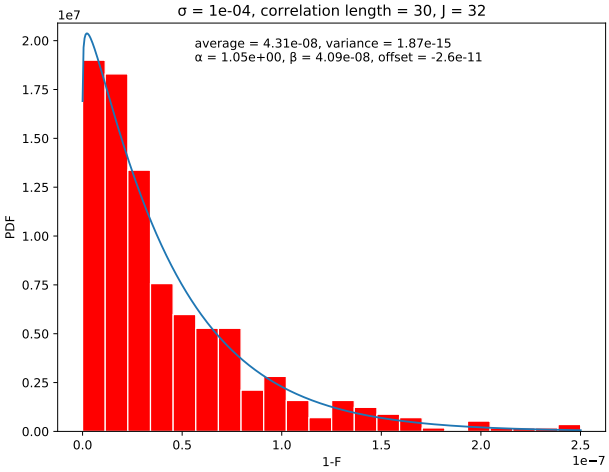


Figure 6: Gamma distribution. J = 32

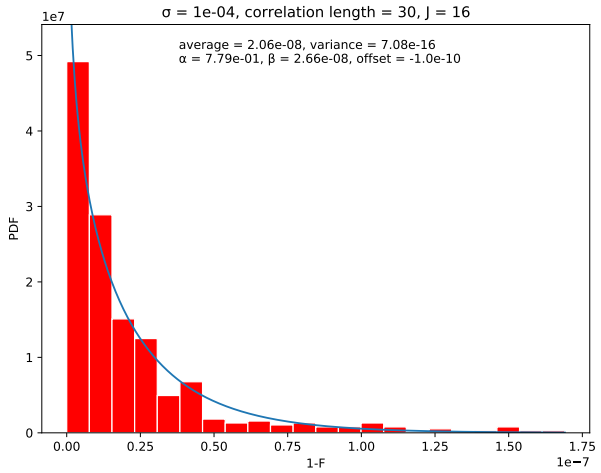


Figure 5: Gamma distribution. J = 16

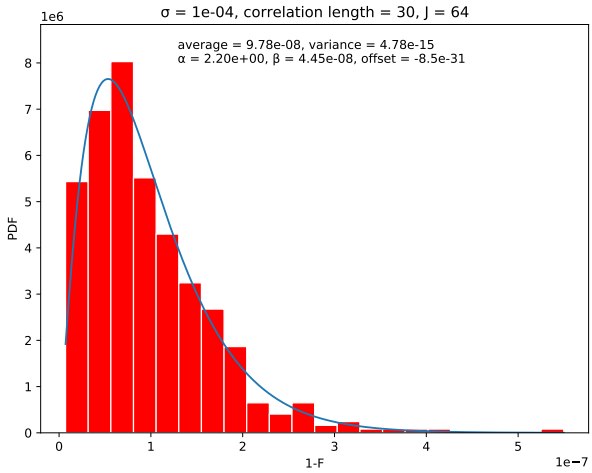


Figure 7: Gamma distribution. J = 64

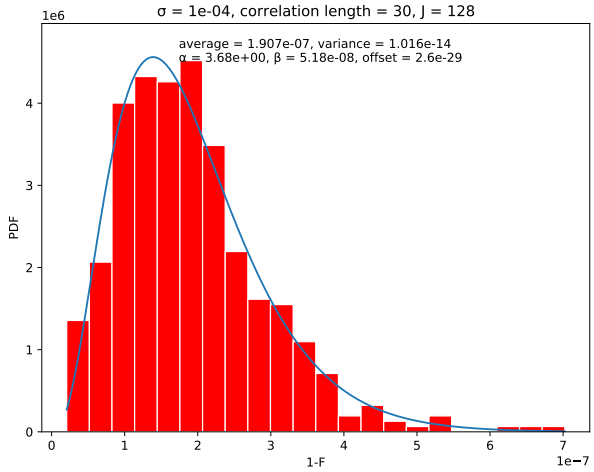


Figure 8: Gamma distribution. J = 128.

These plots show how the behaviour of the Γ distribution changes when J_j, from being less than cc, becomes larger. All of these histograms have $\sigma = 10^{-4}$, $cc = 30$ and the same number of data points: 500. Plots with other values of σ and cc can be found in Appendix B.

This "transition" will be experimentally used in this work in Section 6 to estimate noise correlation lengths. This is rather qualitative however. According to Ball et al. [3], information on cc can be deduced from the variance of the distributions. Therefore, a more quantitative model that given cc, σ and J_j could predict the average and variance of $F(J_j)$ distributions is sought. The model for $\mathbb{E}[1-\Gamma_{\alpha,\beta}] = \mathbb{E}[F(J_j)]$ is the following.

$$\mathbb{E}[F(J_j)] = \begin{cases} A - B_1 J_j \sigma^\nu + C_1 J_j^2 \sigma^{2\nu} \text{ for } cc = 1 \text{ or Markovian noise} \\ A - B_2 J_j \sigma^\nu + C_2 J_j^2 \sigma^{2\nu} \text{ for } cc < J_j \text{ or block correlated noise} \\ A - B_3 J_j \sigma^\nu + C_3 J_j^2 \sigma^{2\nu} \text{ for } cc \geq J_j \text{ or DC noise} \end{cases} \tag{58}$$

A, B_1, B_2, B_3, C_1, C_2, C_3 and ν are free parameters to be fit. `scipy.optimize.curve_fit` was used. The uncertanty for each $F(J_j)$ is $d_{F(J_j)} = \frac{\sigma_K}{\sqrt{K}}$, with σ_K the standard deviation of the K-fold distribution. The results are the following.

-	A	B_1	B_2	B_3	C_1	C_2	C_3	ν
value	1.0000000000	1.59×10^{-1}	1.41×10^{-1}	1.33×10^{-1}	1.9×10^{-2}	2.4×10^{-2}	2.7×10^{-2}	1.994
$\pm$	2×10^{-10}	1×10^{-3}	2×10^{-3}	1×10^{-3}	1×10^{-3}	2×10^{-3}	3×10^{-3}	1×10^{-3}

Table 3: $\mathbb{E}[F(J_j)]$ fit parameters summary. $\chi^2 = 169$

It was expected and desired for ν to be approximately 2, as previously reported in Table 1. It is also comforting that $A = 1$: without errors, i.e. $\sigma = 0$, the measured fidelity would always be 1. The first difference between this model and the one of Table 1 is the presence of second order terms in $O(J\sigma^2)$. Numerically this is preferred: χ^2 decreases of 1 order of magnitude by adding second order terms. On top of that, different C_i and B_i parameters among the 3 noise schemes are preferred to a model with the same B and C. At first order this represents something new when compared to the CRB case. With 72 d.o.f. and a $\chi^2 = 169$, on average every d.o.f. is 1.5 sigmas away from the fit model. Although the theoretical explanation for these values is yet to be unveiled[33], this is a satisfying result.

As $\mathbb{E}[F(J_j)]$ was treated, now it is the time of $\mathbb{V}[F(J_j)]$. A possible model is

$$\mathbb{V}[F(J_j)] = \begin{cases} A_1 (J_j - B_1)(J_j - C_1)\sigma^\nu \text{ for } cc = 1 \text{ or Markovian noise} \\ A_2 (J_j - B_2)(cc - C_2)\sigma^\nu \text{ for } cc < J_j \text{ or block correlated noise} \\ A_3 (J_j - B_3)(J_j - C_1)\sigma^\nu \text{ for } cc \geq J_j \text{ or DC noise} \end{cases} \tag{59}$$

[33] Carrying out simulations where $\tilde{n}$ is let vary is an idea.

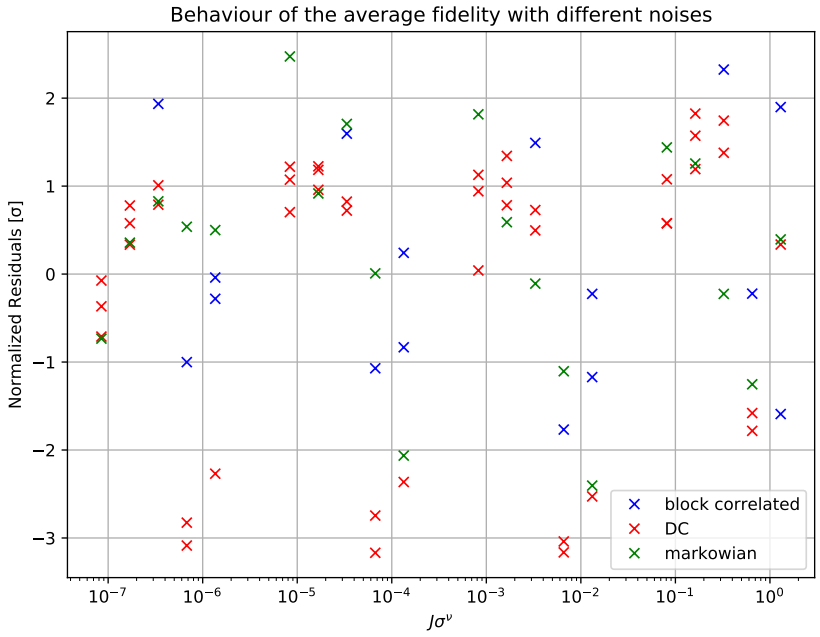


Figure 9: $\mathbb{E}[F(J_j)]$. Plot of the normalized residuals, i.e. (data-fit)/uncertainty. Average distance from 0: 1.5σ.

Table 1 suggests a dependence on *cc* for the block correlated scenario. Because of this the second line of Eq. (59) is of the displayed form. Here the data points are the variances of each distribution. Given a variance Δ, its uncertainty for the weighted fit was chosen to be $\sqrt{\frac{2}{K-1}}\Delta$. More information is available in Appendix C. This is a good choice for normal distributions though, not gamma. Nothing better was found at the time of the data analysis, worse luck. The results are the following.

-	A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	C_3	ν
value	2.1×10^{-4}	4.9×10^{-3}	6.8×10^{-3}	1.6	1	-4.0×10^{1}	-1.6×10^{2}	-4.3×10^{1}	4.8	3.925
$\pm$	2×10^{-5}	4×10^{-4}	3×10^{-4}	4×10^{-1}	2	4	3×10^{1}	7	1×10^{-1}	3×10^{-3}

Table 4: $\mathbb{V}[F(J_j)]$ fit parameters summary. $\chi^2 = 1.3 \times 10^3$

In this case the fit is not as appealing as before. The residuals plot (Fig. 10) makes manifest the unsuitability of this model in regards to DC and block correlated noise. Markovian noise data points are the only ones within 5 sigmas from 0. The chosen uncertainty works in fact better with Gaussian distributions, and gamma distributions subjected to Markovian noise are the most similar to Gaussins. Edmundus et al. [9] have shown that for $\tilde{n} \sim 100$, $\mathbb{V}[F(J_j)]$ converges, while for smaller values it is susceptible to fluctuations of order of magnitudes and comparable with its own value. Other simulations could be run with this $\tilde{n}$. A greater value of $\tilde{n}$ might diminish the overall χ^2 as well. On top of this issue, some parameters of this fit are much greater than 1. This is hard to justify. The only positive note is that $\nu \approx 4$ as for the CRB analogue. Given the failure of the variance model, a quantitative approach for the noise correlation length estimation has not been found.

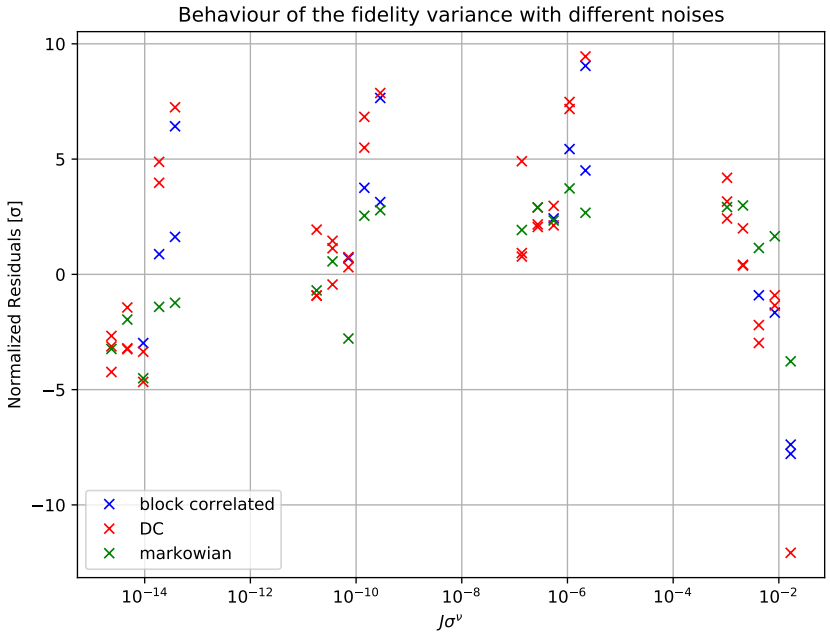


Figure 10: $\mathbb{V}[F(J_j)]$. Plot of the normalized residuals, i.e. (data-fit)/uncertainty. Average distance from 0: 4.3σ.

4 The experimental setup

The experimental work of this Master's thesis was focused on the *Segtrap* setup. This is a 3D segmented ion trap, which was designed to work with beryllium $^{9}Be^{+}$ and $^{40}Ca^{+}$ ions. The long term aim of the setup is to adopt the *Quantum Charged-Coupled Device (QCCD)* approach. This involves designated regions for cooling, gate implementation and storage between which ions are shuttled back and forth to implement an arbitrary algorithm. Different ion species are used in order to have the spectral separation needed to avoid perturbing the quantum state of an ion when carrying out operations on the other one and to realize sympathetic cooling, with which one ion is cooled by the other one of a different species. The 15 segments of the trap enable shuttling of ions along the trap axis. Quantum operations, be they state initialization, quantum gates, state readout or cooling, are driven by laser light while the ions are trapped in a potential well. All of the information in this section and much more can be found in Matteo Marinelli's PhD thesis [22].

4.1 Ions electronic levels and related operations

An experimental run can be divided in 4 parts: cooling of the motional modes, state preparation, qubit manipulation and state detection. While the qubit manipulation differs from experiment to experiment, the other 3 parts are executed for any experiment on Segtrap with minor differences.

The theme of cooling motional modes is here only introduced (see [22] for the complete tractation). There are 4 cooling "stages": far detuned Doppler cooling or *precooling*, standard Doppler cooling, EIT cooling and sideband cooling. Precooling has a large capture range. In this way large motional excitation caused by gas collisions or laser frequency drifts can be recovered. Then standard Doppler cooling is applied to further reduce the motion close to the Doppler limit. Then it is the turn of EIT cooling (applied only on Ca on the radial motional mode with lowest frequency) which reduces the average excitation quantum number $\bar{n}$ to $\bar{n} \approx 1$. Eventually it is the turn of sideband cooling, applied on the axial motional mode, which cools the ion to the ground state. With mixed species ions, sideband cooling is applied on the two axial modes.

4.1.1 Calcium ion

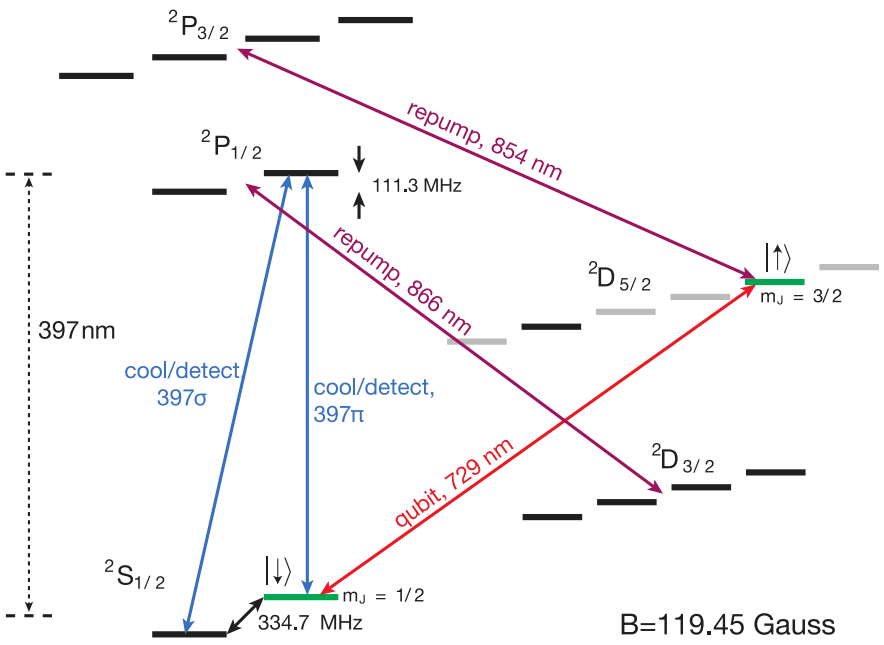


Figure 11: $^{40}Ca^+$ levels. In green the ones used as qubit levels, in black the accessible levels. The solid purple, blue and red lines represent the lasers used for the ion manipulation. Taken from [22].

Once a magnetic field of 119.45 Gauss is applied the relevant electronic structure of $^{40}Ca^+$ becomes the one portrayed in Fig. 11. The chosen qubit transition is the $|S_{1/2}, m_J = 1/2\rangle \leftrightarrow |D_{5/2}, m_J = +3/2\rangle$ quadrupole allowed transition[34]. This is also called the *optical qubit transition* and it is driven by a laser with wavelength 729.35 nm. The coherence time of the optical qubit is approximately 1.7 ms. This value stems from how much the transition frequency varies with the magnetic field, i.e. 1.12 MHz/G [22]. The $S_{1/2} \leftrightarrow P_{1/2}$ transition at 396.96 nm is used to prepare and readout the state and cool the motional modes. The $P_{1/2}$ levels can however also decay to the metastable $D_{3/2}$ manifold. For this reason, repumping at 866.45 nm is necessary. The $D_{5/2}$ levels are also repumped to the ground state exploiting the short lived $P_{3/2}$ manifold. The $P_{3/2} \leftrightarrow D_{5/2}$ transition is driven by a 854.44 nm laser.

In regards to state preparation, population in the $|S_{1/2}, m_J = -1/2\rangle$ state is unwanted. This is depopulated by driving the $|S_{1/2}, m_J = -1/2\rangle \leftrightarrow |P_{1/2}, m_J = 1/2\rangle$ with 397 nm σ_+-polarized laser light which does not couple the $|S_{1/2}, m_J = +1/2\rangle$ state. After the drive the population will decay back to both ground state levels. Continuing cyclically, after some μs $|S_{1/2}, m_J = -1/2\rangle$ will be depopulated.

As far as state readout is concerned, the 397 nm transition is driven and if the qubit is in the bright state $|\downarrow\rangle$ photons will be measured by a photomultiplier tube (PMT). The $|\uparrow\rangle$ is instead a dark state and fewer photons will be measured. According to the number of measured photons the qubit states can be discriminated.

[34] $|S_{1/2}, m_J = 1/2\rangle \leftrightarrow |D_{5/2}, m_J = -1/2\rangle$ is another possible transition.

4.1.2 Beryllium ion

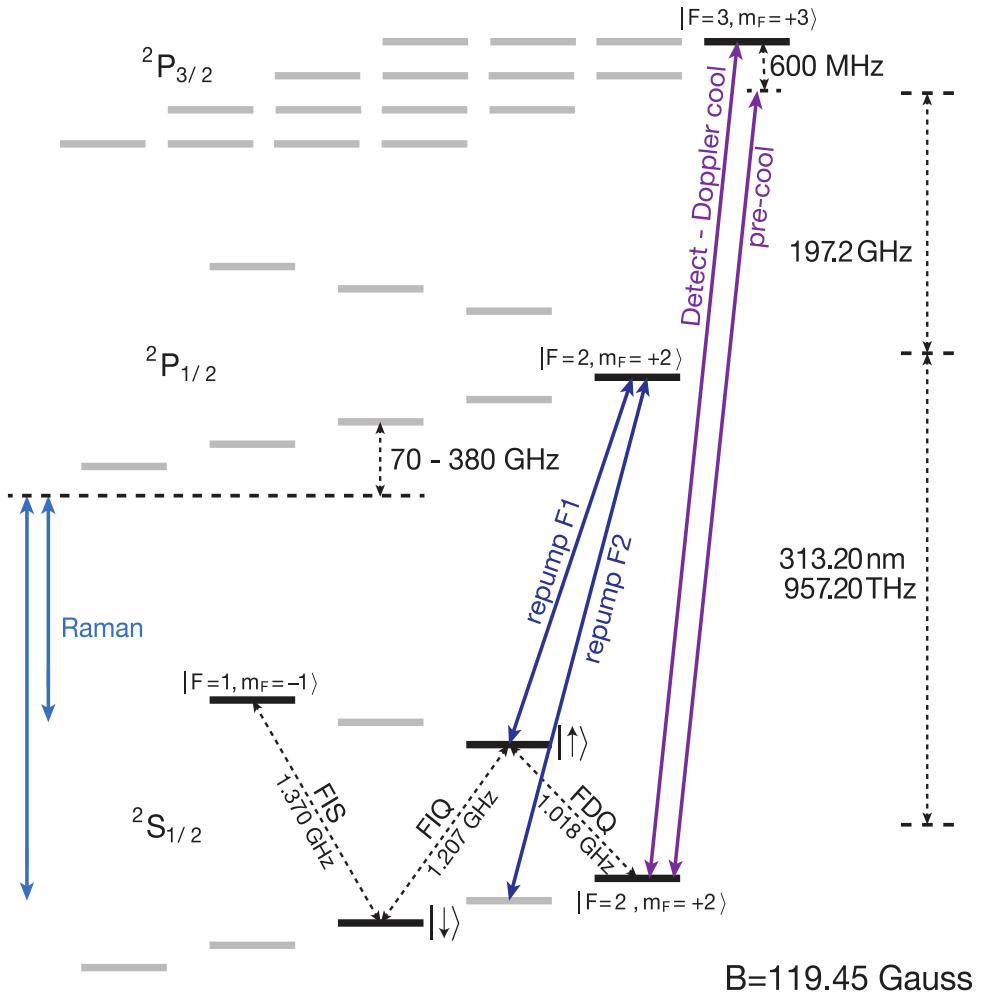


Figure 12: $^9Be^+$ levels. $|\downarrow\rangle$ and $|\uparrow\rangle$ are the qubit levels. The solid lines represent the lasers used for the ion manipulation. Taken from [22].

Beryllium has nuclear spin 3/2 while Calcium does not have a nuclear spin. Because of this, Be features hyperfine structures. The degeneracy is removed once a magnetic field of 119.45 Gauss is applied. Two-photon Raman transitions couple the $S_{1/2}, F = 1$ and $S_{1/2}, F = 2$ manifolds via the $P_{1/2}$ states. The detuning of the Raman lasers from the $P_{1/2}$ is ∼230 GHz [22]. With the chosen magnetic field B, the $|F = 2, m_F = 0\rangle \leftrightarrow |F = 1, m_F = 1\rangle$ transition is at first order independent of B. Therefore it is called *field independent qubit (FIQ)*. Its coherence time is ∼ 4s. The light used for any transition has wavelength around 313 nm. Co-propagating and perpendicular beams are available. The former are used to drive motion insensitive qubit transitions, the latter are employed for motion sensitive transitions, e.g. sidebands.

When performing state readout, the probability of the the dark state $|\downarrow\rangle$ to be measured as bright is decreased by driving the FIS (field independent qubit shelving) transition as in Fig. 12, while the FDQ (field dependent qubit) transition is driven to transfer the population from the bright state to $\left|S_{1/2}, F = 2, m_F = 2\right\rangle$. Then $\left|S_{1/2}, F = 2, m_F = 2\right\rangle \leftrightarrow \left|P_{3/2}, F = 3, m_F = 3\right\rangle$ is driven with σ_+-polarized light. Photons emitted during the decay can be measured with a PMT. Like with Ca, the state is distinguished depending on the number of measured photons. Possible error sources are the following ones. Polarization impurities allows the decay to any state of the $S_{1/2}$ manifold, not only $\left|S_{1/2}, F = 2, m_F = 2\right\rangle$. In addition off resonant pumping of the dark states can populate $\left|S_{1/2}, F = 2, m_F = 2\right\rangle$.

The $|\uparrow\rangle$ state is prepared by repumping from both $\left|S_{1/2}, F = 2, m_F = 1\right\rangle$ and $\left|S_{1/2}, F = 1, m_F = 1\right\rangle$ to the state $\left|P_{1/2}, F = 2, m_F = 2\right\rangle$. This enables $\left|S_{1/2}, F = 2, m_F = 2\right\rangle$ to be repopulated. Then the FDQ transition is driven.

4.2 The control system and limitations

Complex experiments like Segtrap need an adequate control system. The TIQI group developed over the years the *M-ACTION* control system. It consists of a hardware and a software part. The relevant aspects for this project are here listed. Please refer to Negnevitsky's thesis [26] for more information beyond that.

The hardware's core component is the Zedboard. This board has, among other parts, a Field Programmable Gate Array (FPGA) and a Xilinux Zynq-700 0 chip composed of two ARM CPUs which control the experiment. The Zedboard is connected to the DDS boards (direct-digital synthesizer) which generate the signal for the AOMs (acousto-optic modulator), which enable the experimental physicist to obtain the desired laser pulses. Other devices belonging to the M-ACTION system are the AWG (arbitrary waveform generator) board, administering the voltages on the DC electrodes of the trap, and the DEATHs (Direct Ethernet-Adjustable Transport Hardware), controlling the time dependent voltages.

In regards to the software part, with the Xilinix software development kit (SDK) the user can code in C++ the experiments to be run. This code controls the M-ACTION system. On top of this, there is also the C++ GUI (Graphical User Interface) called Ionizer2. Ionizer2 communicates with Zedboard, the DEATHs and other devices. It queries the Zedboard, on which experiments and parameters are compiled, for the available experiments and for digital outputs results. With these data it creates a page for every experiment at disposal. On these pages the user can set parameters, run experiments and in general run scans. A scan is a sequence of experiments in which one or also two parameters are changed sequentially. During a scan, Ionizer2 saves its results obained from the Zedboard for every data point, i.e. for every value of the parameter(s). In addition, Ionizer2 exposes a scripting API. Python scripts can be launched from the dedicated Ionizer page. Thanks to the group efforts, the API permits the scripts to reproduce any action executed manually by a TIQI group member. Scripting allows scheduling tasks and running calibrations.

4.2.1 Some preexisting SDK APIs

The SDK provides different APIs with different levels of abstraction. Some of the important characteristics are mentioned here. A more detailed description is found in Matteo Marinelli's thesis [22]. The `experiment` class is the object that executes an experimental sequence as defined in Section 4.1. This class interacts with the `remote parameters`. These are experiment-specific, like frequencies and pulse times. They can be set via the GUI or the SDK and are compiled onto the Zedboard.

An additional software element developed by Matteo Marinelli and Tanja Behrle are the `crystal` classes. A crystal object has information on what ions are trapped and on the relevant `remote parameters`. The `crystal` classes encapsulate sets of operations that are often carried out during an experiment, like detection and readout, adding abstraction and allowing users without a low level knowledge of the M-ACTION system to code and perform experiments. `crystal` objects contain `transition` objects as well. "A transition object groups together the main pulses [e.g. π and $\pi/2$ pulses] that can be run on a particular qubit transition, together with the remote parameters that the user controls" (Marinelli, [22], page 54). Phase accumulators were coded by Matteo, too. They belong to the `crystal` classes and their task is to ensure phase coherence between DDSs (and therefore light pulses) and qubits. Phase accumulators calculate and add a phase offset to the relevant DDSs after each applied pulse. As an example, for the Calcium case, considering the Stark shift as well, the phase accumulated by the qubit, after a pulse of time t_p and a waiting time t_w before the next pulse, is

$$\phi_q = \omega_d t_p + \omega_0 t_w, \tag{60}$$

where ω_0 is the transition frequency without the drive, or bare frequency, and ω_d the one with the drive, or dressed frequency. The phase accumulators make sure that the relevant DDSs have the same phases of the qubits after $t_p + t_w$ time.

4.2.2 Sequence length limitation

The way the `crystal` classes and the phase accumulators were coded, in combination with the limited DDSs memory, limits the number of pulses that can be run for a single qubit manipulation to 35-40. By keeping track of the Stark shift phase correction the pulses have different phases and therefore the already allocated DDS memory cannot be reused to instantiate other pulses with the same characteristics. With Beryllium one can run ~ 40 pulses, with Calcium ~ 35. Ca is also EIT cooled, Be is not. For this reason less DDS memory is at disposal and therefore fewer pulses can be performed on Ca when compared to Be. These numbers limited the potential of the RB and GST experiments executed for this thesis. RB and GST experiments would benefit from longer sequences. Martin Stadler has eased the limitation by modifying the code that instantiates the pulses. The new code updates the relevant parameters after a pulse has been executed, without the need of instantiating another one. Now a sequence can be some hundreds of pulses long. This new software was however not yet adopted by or tested on Segtrap due to time constraints.

5 The Randomized Benchmarking Framework

One of the main tasks of this project was to code and develop software which could be used by any TIQI subgroup, as well as more specific code for Segtrap, with the goal of designing and executing RB and GST experiments. The following sections will clarify RBF (Randomized Benchmarking Framework) motivations and aspects. Alfredo Ricci Vásquez was an active contributor to the RBF, especially in its design and early stage coding. The RBF documentation is available at the dedicated TIQI WIKI page [45]. The RBF supports at the moment any single qubit gate and 2-qubit gates which act symmetrically on the qubits, i.e. GATE(q1,q2)=GATE(q2,q1), like the MS gate. At the time of writing, a missing requirement of the Randomized Benchmarking Framework (RBF) is the possibility of generating and performing 2-qubit GST experiments.

5.1 RB implementation

Randomized Benchmarking is the first class of algorithms that has been implemented in this framework. The library pyGSTi [29] enables the physicist to design CRB or DRB experiments provided the available gates, the sets of lengths J_j, as defined in Sections 1.1.2 and 1.2.1, and the number of randomizations K independent of the index j. The output of this library is the collection of CRB or DRB circuits, compiled in the native gates of the system, which the benchmarker will apply to the relevant quantum system with the aim of running an RB experiment. pyGSTi supports an arbitrary number of qubits n.

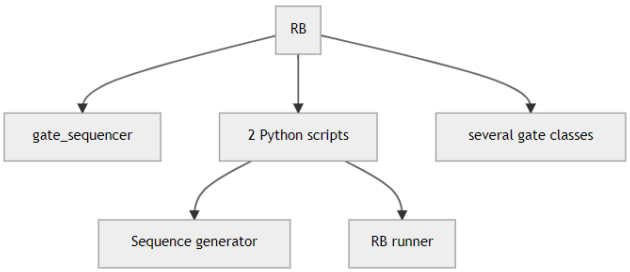


Figure 13: RBF software components for RB. Taken from the TIQI RBF repo [45].

5.1.1 A human-readable framework

Solutions that guarantee transparency in regards to what circuits are to be and are being executed and user-friendliness were looked for. JSON files were identified to be an attractive solution. They allow to store simple data structures in the form `"parameter_name" : parameter_value`. The value can be a boolean, a number, a string, a list or another JSON data structure. JSON files can be modified by any text editor [35].

Two python scripts, which can be found in the TIQI repo [37], were coded. The first one is `sequence_generator.py`. It is run by the user and needs a JSON file as input, in which the J_js, K, the available gates, the choice between DRB or CRB and other parameters are declared. The script makes use of pyGSTi and as output it provides a JSON file in which the CRB or DRB sequences, compiled in the native gate set, are stored. If not already available, `sequence_generator.py` creates a `Quantum Processor` object. This is used by pyGSTi to sample the RB circuits. The `Quantum Processor` can be reused for other RB experiments with the same native gates. `run_randomized_benchmarking.py` is the second script. This is executed by Ionizer. It necessitates the JSON file with the RB sequences and an additional JSON setting file. This JSON file allows to define the value of any parameter that appears in the relevant Ionizer experimental page[35], like the qubit transition, the cooling type, shaping or not the pulses, the number of shots per data point etc. `run_randomized_benchmarking.py` starts from the randomization 0 and ends with the randomization $K-1$. Given $i \in \{0, 1, .., K-1\}$, the script loads the i-th circuit $\forall\, J_j$ in Ionizer, Ionizer performs a scan over J_j and `run_randomized_benchmarking.py` appends the results of the i-th randomization onto an established text file. Other experiments start by running the shortest sequences and finish with the longest ones. This approach is here rejected. In this scenario any time dependence of the experimental errors accumulate on the longest sequences, leading to worse data for them, i.e. a bias. In the RBF case instead, long and short sequences are performed through all the experiment. In this way, slow noise contributions will not appear as a "parasiticly" additional J_j-dependence worsening the experimental results of circuits with high values of J_j. `run_randomized_benchmarking.py` enables the physicist to stop the experiment at any point and to resume it by inputting the desired value of i, while keeping the already collected data thanks to the usage of the same text file. This is very convenient if the SDK crashes (for example if the DDSs memory is depleted) or if major experimental issues require RB to come temporarily to a halt.

[35] What page this is will become explicit in the next lines.

The experimental page on which `run_randomized_benchmarking.py` loads RB circuits relates to an `experiment` object which must inherit from the class `gate_sequencer` . This class is available in the `ionpulse_sdk_core` [44]. When an experiment extends `gate_sequencer`, its page in Ionizer exposes one `rp_string` per qubit involved in the RB experiment. The `rp_string` boxes are filled in by `run_randomized_benchmarking.py` with the RB circuits, in JSON format. When Ionizer runs a scan over J_j, the loaded circuits are carried out on the setup. The RapidJSON library [20], also available in the `ionpulse_sdk_core`, is exploited by `gate_sequencer` to translate the JSON content of the `rp_string`s into logic gates, which are then executed. `gate_sequencer` does not directly call the functions generating the pulses: this can be very different among the different setups; by implementing one single solution, `gate_sequencer` would not be general anymore. For this reason, `gate_sequencer` features an abstract function which has to be overwritten by the `experiment` object that inherits from the sequencer. This function, depending on the setup, will use the relevant SDK to generate the actual pulses.

5.1.2 The case of Segtrap

In the specific case of Segtrap, the explicit way for the experiments inheriting from `gate_sequencer` to generate the light pulses had to be defined. The new `gate` classes were added to the APIs developed by Tanja and Matteo (see Section 4.2.1). These `gate`s were defined for the Calcium, Beryllium and BerylliumCalcium `crystal` classes and can be found in the SDK repository [39]. They make use of the pulse functions exposed by the `crystal` classes and run the desired pulse. They are grouped by a `Native_Gate_Set` object. This object instantiates `gate`s along the X, Y and Z[36] axes with rotational angles π and $\pi/2$ once the user has selected the qubit transition, whether to keep track of the Stark shift phase and other parameters. An RB experiment (a derived class of the `gate_sequencer`)[37] possesses a `Native_Gate_Set` object. The latter enables the former to access the `gate`s and generates the pulses. The `gate` classes approach is thought to be a valuable one: logic gates can now be used on Segtrap for other non-RB experiments. Other TIQI members working with different setups do not necessarily have to code their `gate`s, because their RB experiment could directly call their pulse functions at disposal in their SDK. They do however have to code their RB experiment class.

5.2 GST implementation

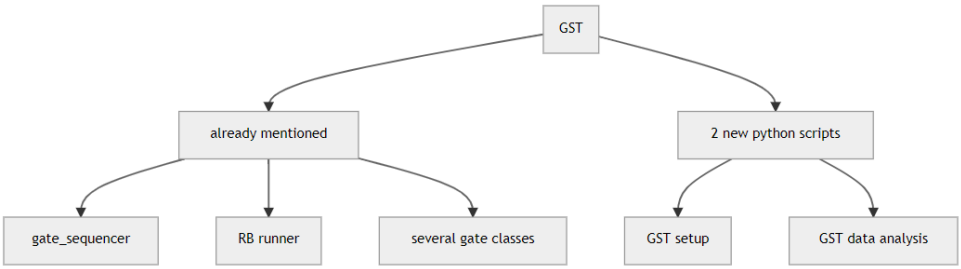


Figure 14: RBF software components for GST. Taken from the TIQI RBF repo [45].

GST is also at the TIQI group's disposal via the RBF. The software already coded for RB is not to be thrown away. In fact, the `gate_sequencer`, `run_randomized_benchmarking.py` and the `gate` classes are used for the GST case as well and with the same objectives. In addition, 2 further python scripts were coded. `gst_setup.py` performs the GST "preprocessing": it uses pyGSTi to generate the GST circuits which are saved in a JSON file just like `sequence_generator`. It also saves a folder which is used in postprocessing. This folder and the generated sequences can be reused for any setup if the gate set is the same. `gst_setup.py` requires a JSON file with the allowed lengths for the GST sequences and other parameters. The JSON file containing the GST sequences is structured in a way that reminds the RB structure, i.e. K randomizations for every J_j. "K randomizations for the length J_j" does not make any sense when talking about GST, because there are no randomizations in the RB sense. Rather different germs are performed more times (see Section 2.2.2). All of this only means rearranging the GST sequences in order to "trick" `run_randomized_benchmarking.py`, i.e. to

[36] Z gates do not need any actual pulse: they are phase shifts applied to subsequent gates.
[37] An RB experiment also inherits from the ion-related `frame_CoolDet` class.

provide it with a suitably structured file, and to exploit it to scan GST sequences with `gate_sequencer`. K is chosen to be

$$K = \left\lceil \frac{\#\text{ all circuits}}{\#\text{ all lengths}} \right\rceil . \tag{61}$$

GST circuits are order per length by pyGSTi. The effective k-th randomization groups the circuits with the enumeration decided by pyGSTi like

$$k\text{-th randomization} \equiv \{k, \quad K+k, \quad 2K+k, \quad \cdots, \quad (\#\text{ all lengths} - 1) \times K + k\} . \tag{62}$$

In this way every randomization has circuits that come from every part of the circuit collection generated by pyGSTi, not only from the beginning or the end, which means circuits with (potentially) different lengths. If (# all circuits/# all lengths) is not an integer, additional circuits with only SPAM operations are added such that every randomization has the same number of circuits. `gst_analysis.py` rearranges then the results saved by `run_randomized_benchmarking.py` and analyzes the GST data. In the settings JSON file one can specify what gate set models (TP, CPTP , etc.) and what gauge optimization procedures to use, as well as whether adding error bars at 95% confidence level with the Hessian matrix method described in Section 2.3.3[38]. `gst_danalysis.py` generates than a report in HTML form with all the results in the aforementioned folder.

5.3 2-qubit GST

The RBF does not offer yet the possibility of designing a 2-qubit GST experiment. The reason will be now explained.

Let us assume that each qubit can be addressed by a set of gates whose cardinality is Ξ. If a setup can generate and handle contemporaneous pulses, the 2-qubit gate set will be of cardinality $\Xi^2 + \#$ 2 qubit gates. If not, the executable gates are the 2 qubit ones, $\mathbb{1} \otimes G_{\text{single qubit}}$ and $G_{\text{single qubit}} \otimes \mathbb{1}$. The resulting gate set has cardinality $2\Xi + \#$ 2 qubit gates. At the time of writing, pyGSTi does not seem to support a gate set with more than 26 logic gates. This fact can limit its usefulness in the former case. A reply from the pyGSTi creators is awaited. In addition, Luca Huber and Alfredo Ricci Vásquez are putting effort in developing simultaneous pulses for the SDK. It is not yet clear whether the 2-qubit GST should be able to work in both cases or just in the latter one.

More information is required about the pyGSTi library and further reasoning is needed for a successful 2-qubit GST implementation in the RBF. The available time for this master project brought the 2-qubit GST implementation to a halt.

[38] This is preferred to bootstrap because better documented [28]

6 DRB and GST experimental data on Segtrap

On the Segtrap setup, DRB and GST data sets were collected for the single Calcium ion and for the single Beryllium ion employing the RBF. Tanja Behrle was an indispensable help in the laboratory. It was attempted to also collect DRB data for the 2 qubit case with Beryllium and Calcium both loaded. Unfortunately experimental issues led to relatively high ion loss rate and made this unfeasible. 2 qubit data would have been interesting in order to compare the setup performances for the same operations with and without cross-talk noise [34], as well as benchhmarking the 2 qubit MS gate. DRB was preferred over CRB because it benchmarks with the natively available gates. Had more time been at disposal, CRB experiments would have been also executed for the sake of undestanding whether DRB and CRB produce equivalent results. When not specified differently, the gate set used comprises $+\pi/2$ and $+\pi$ rotations along the logical axis X, Y and Z.

6.0.1 Calibrations for both ions

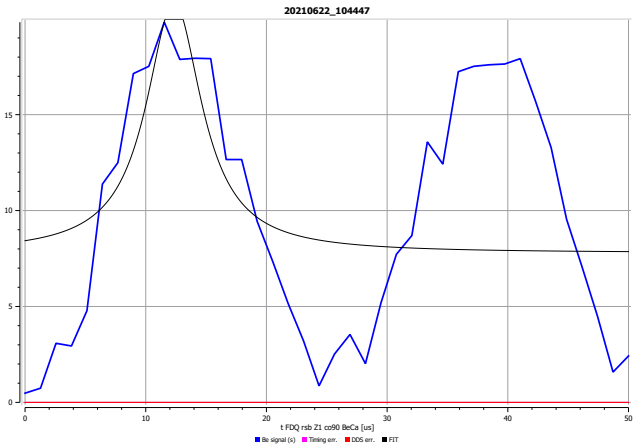


Figure 15: Time scan of the FDQ red sideband at 90° for Be^+. The transition flops between bright and dark. The solid black line can be ignored. Photon counts on the Y axis.

Before running the actual experiment, several calibrations had to be performed. Although these differ between the Ca and the Be case, some had to be executed for both, like *micromotion compensation*, before the actual experiments started. Fabrication defects and electrodes misalignments cause the time dependent voltages (or *rf potential*) in the trap to have a not perfect orientation. This leads the ions to move around the point of equilibrium and give rise to what is called *micromotion*. Micromotion can increase the chance of losing ions, and with more than one ion, frequency and directions of normal modes are modified [22]. In order to compensate for the micromotion, Calcium was loaded and the voltages on the relevant electrodes, the *shim electrodes*, were tuned. For each of the two radial directions, the respective motional frequencies were scanned. Subsequently, the *shim voltage* α was sequentially changed, and the shim voltage β was scanned with the aim of having a bright ion. α and β are the voltages related to the shim electrodes. From the optimal βs, the new voltages to apply to the shim electrodes can be calculated using pre-existing routines.

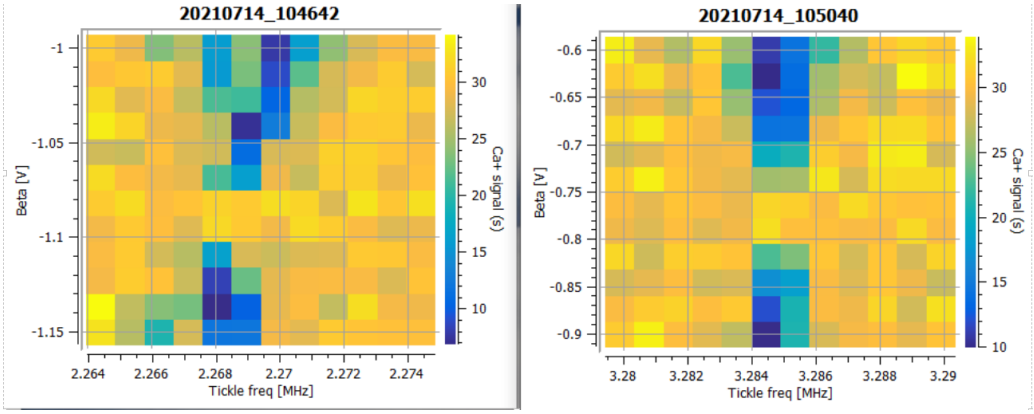


Figure 16: Example of plots for the micromotion compensation. The blue areas with few photon counts (dark Ca^+) are caused by the motional resonances.: the ion is excited and therefore appears as dark. With the optimal β Ca^+ is bright.

Another calibration regards *thresholding*. When measuring a qubit, a PMT collects a certain number of photons. The probability of detecting n photons follows a superposition of two Poisson distributions [22] if errors like leakage, off reasonant pumping and impure polarization are not present. The already coded routines allow to

keep these non-idealities into consideration and they produce a threshold (≈ 4 counts for Beryllium and 6 for Calcium). When fewer photons than the threshold are measured, the state is considered dark, otherwise bright.

Sideband cooling also had to be tuned. It was made certain that the 729 red sideband and the FDQ with perpendicular beams red sideband transitions could flop properly. In this way it is assured that the transitions can be driven when executing sideband cooling. This was achieved by calibrating the motional frequencies, by calibrating the frequency and π time of 729 and FDQ 90° red sibands and by tuning the intensities of the relevant laser light.

6.1 Beryllium ion

DRB and GST data were collected in the same instance for the single Beryllium ion. The FIQ transition was used as qubit transition. In addition to the aforementioned calibrations, the FDQ carrier with co-propagating beams was calibrated. The π time was tuned too, obtaining a value of $\sim 6\mu$s. After this it was the time for the FIQ bare frequency, the FIQ π and $\pi/2$ times, with values of ≈ 10 and 5 μs, at a fixed equal laser amplitudes for both. By keeping the amplitudes equal and fixed, the Stark shift phase correction (see Eq. (60)) can be applied to both π and $\pi/2$ pulses. Subsequently the FIS π time was also calibrated, resulting to be $\approx 40\mu$s. As explained in Section 4.1.2, the FDQ and FIS transitions play a role in SPAM operations. In the end the FIQ Stark shift frequency was calibrated. Frequencies are calibrated via Ramsey experiments [22].

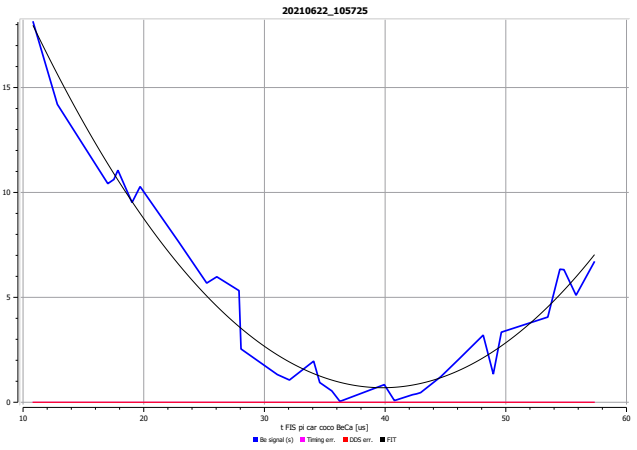


Figure 17: Time scan of the FIS π time. The state is dark after $\approx 40\mu s$.

Then the DRB and GST experiments were executed, making use of the Randomized Benchmarking Framework (see Section 5). During the experiments, FDQ π time, FIQ bare frequency, π time and $\pi/2$ time calibrations were performed every $\sim$10 minutes. The laser amplitudes for the FIQ pulses were not varied and the same Stark shift differential frequency correction was used for both experiments.

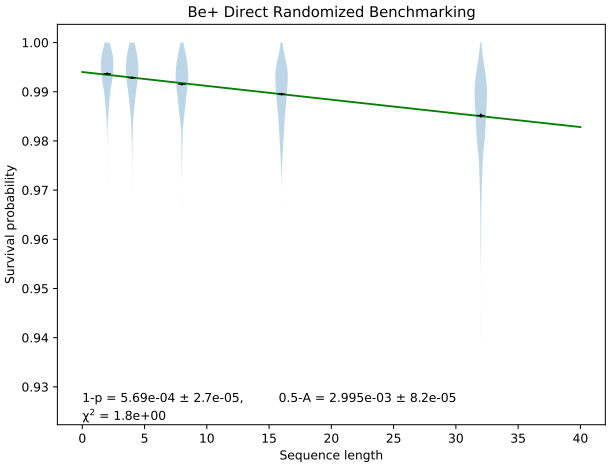


Figure 18: Exponential DRB decay for Be^+.

6.1.1 DRB data and results

The possible lengths J_j were [2, 4, 8, 16, 32]. For each of them, $K = 500$ circuits were run. Each one of these K circuits was executed 500 times (also known as 500 shots). The raw data is found in its repository [40]. The law

$$F(J_j) = A + Ap^{J_j} \tag{63}$$

was fitted. As already stated in Section 3.1, the law $F(J_j) = A + Bp^{J_j}$ is rejected: with the additional degree of freedom `scipy.optimize.curve_fit` cannot provide a sensible fit. Like in Section 3.1, in order to work out a weighted fit, for each J_j the average survival probability was assigned the uncertainty $\frac{\sigma_K}{\sqrt{K}}$, with σ_K the standard deviation of the K-fold distribution for the length J_j. As shown by Fig. 18, the fitted parameters are:

$1-p$	$0.5-A$	χ^2
$5.69 \times 10^{-4} \pm 2.7 \times 10^{-5}$	$2.995 \times 10^{-3} \pm 8.2 \times 10^{-5}$	1.8

Table 5: Fit results for Eq. (63) for Be^+. The data are shown in Fig. 18.

When compared to the simulations in Section 3.1, the Beryllium data show a low value of χ^2. This suggests that non-markovianities have a short correlation length. The violin plot (Fig. 18) shows however tails. Given these two points, one can expect a noise correlation length cc greater than one but comparable to 1. cc can be better estimated by observing the infidelities distributions for each J_j (Figs. 19 to 23). Just like in Section 3.2, `scipy.stats.gamma.fit` was employed to fit the gamma distributions. Table 6 puts together the fitted parameters for Figs. 19 to 23.

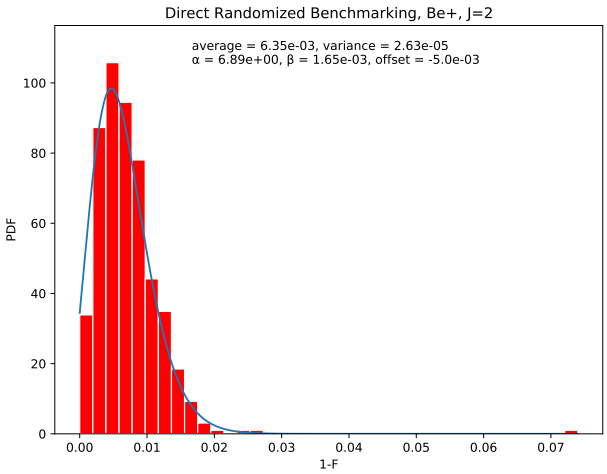


Figure 19: Distribution of $1 - F(2)$, Be^+.

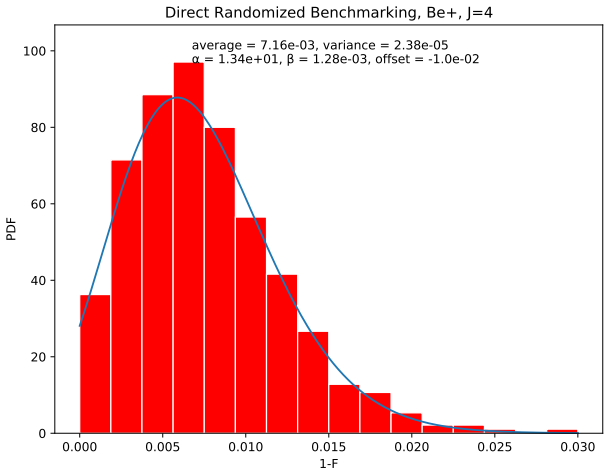


Figure 20: Distribution of $1 - F(4)$, Be^+.

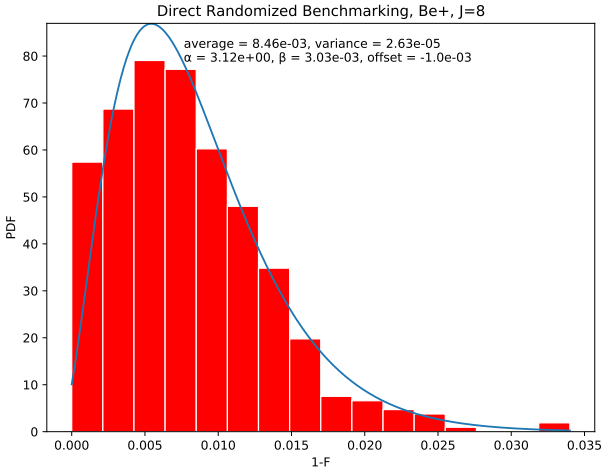


Figure 21: Distribution of $1 - F(8)$, Be^+.

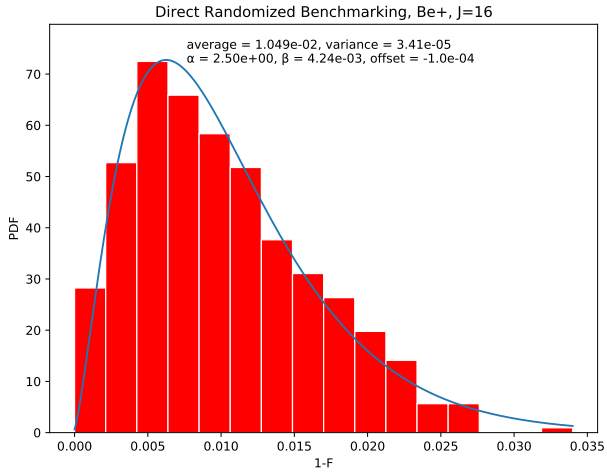


Figure 22: Distribution of $1 - F(16)$, Be^+.

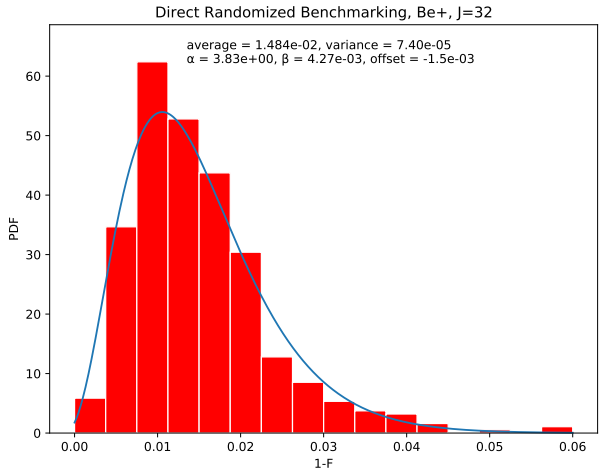


Figure 23: Distribution of $1 - F(32)$, Be^+.

Figs. 19 to 23 show, for each value of J_j, 1-the survival probability produced by the $K = 500$ DRB circuits. These distributions are gamma distributions.

Recalling the qualitative estimation method in Section 3.2 based on the shape of the gamma distributions, it

J	average	variance	alpha	beta	offset
2	6.35×10^{-3}	2.63×10^{-5}	6.89×10^{0}	1.65×10^{-3}	-5.0×10^{-3}
4	7.16×10^{-3}	2.38×10^{-5}	1.34×10^{1}	1.28×10^{-3}	-1.0×10^{-2}
8	8.46×10^{-3}	2.63×10^{-5}	3.12×10^{0}	3.03×10^{-3}	-1.0×10^{-3}
16	1.049×10^{-2}	3.41×10^{-5}	2.50×10^{0}	4.24×10^{-3}	-1.0×10^{-4}
32	1.484×10^{-2}	7.40×10^{-5}	3.83×10^{0}	4.27×10^{-3}	-1.5×10^{-3}

Table 6: Summary table with the parameters of Figs. 19 to 23.

can be observed that none of the Beryllium data distributions have either a peak in 0 or are monotonic like Figs. 4 and 5. This is indication of the noise not being DC for $J_j = 2$ already, but rather block-correlated. From this a bound on the noise correlation time can be obtained. Firstly, the effective average sequence length for $F(J_j) = 2$ has to be calculated. The DRB algorithm in Section 1.2.1 has to be kept in mind. A sequence is composed by 3 circuits: U_s, U_{J_j} and C_0. For $J_j = 2$, U_j has depth 2 while U_s and C_0 depths do not depend on J_j. With the chosen gate set, where $+\pi$ and $+\pi/2$ rotations along X, Y and Z are available, C_0 is compiled in at most 4 native gates [52] and U_s in at most 2. The last result can be deduced by imagining a block sphere, with the allowed states and rotations. Assuming that on average only half of these gates is used to compile U_s and C_0, there are three additional gates. Therefore the average DRB sequence length for $J_j = 2$ is 5. Now the average noise correlation time can be bounded. With depth-5 sequences the noise does not appear as DC noise. Therefore the noise correlation length $cc \leq 4$. The average time duration of 4 gates has to be calculated. π and $\pi/2$ Z rotations are just phase modifications for successive gates: they need no time to be executed. X and Y π pulses need approximately 10μs, while X and Y $\pi/2$ pulses 5μs. All things considered, on average a gate needs 5μs. This means that the noise correlation time with a Beryllium ion was $\leq 5\mu\text{s} \times 4$ gates $= 20\mu s$. This can be caused by magnetic fields or laser frequency fluctuations.

This estimation, based on the simulations in Section 3.2, is valid under three assumptions.

1. SPAM operations are noiseless.
2. The gates are subjected only to coherent noise.
3. Gate errors occur only along one axes.

The first hypothesis is not valid as the value of $0.5 - A$ in Table 5 shows that $0.5 - A$ is not compatible with 0 (within 36σ): this is indication of SPAM errors. RB experiments are designed to decouple SPAM and gate errors. The former manifest themselves with values of A away from the ideal 0.5. A different value of A, keeping p fixed, changes the gamma distribution offsets, as defined in Eq. (55). It is not believed that (small) SPAM errors can transform a gamma function from monotonic to not monotonic. Thus this should not invalidate the previous analysis. As far as the second assumption is concerned, GST data provide evidence of the presence of only coherent errors (see next section). The third element of the list could be the weak point of this analysis. The simulations of Section 3 took into account merely coherent errors along Z while they can occur along any axes. Ball et al. [3] outlined the theory for this scenario: $1 - F(J_j)$ is distributed like a sum of 3 gamma distributions for CRB experiments. The Be^+ DRB data can be fitted by just one gamma distribution. It might have surprisingly happened that most of the errors affected only one axis or that the same errors affected more axes. This is just a speculation, though.

6.1.2 GST data and results

The GST analysis report is found in the GIT repository [41]. By downloading this folder, `main.html` can be opened and the results visualized. The raw data is found in the superior directory. The file is named `gst_segtrap_be.txt`. Only parts of this very detailed report will are shown in this work. The GST data was analysed using the CPTP and TP gate set models (see Section 2.2.3) and different gauge optimization algorithms[39] with different SPAM weights (see Section 2.3.2). The GST circuit formal depths were established to be 1, 2, 4, 8, 16 and 32. The actual depth can vary from circuit to circuit, depending on the SPAM fiducial circuits (see Section 2.2.1), which increase the depth, and on the depth of the germ which is repeated[40]. An

[39] Their results can be selected in the report [41].
[40] If the germ has depth 3 and the length considered is 16, the germ will be repeated 5 times, accounting for a depth of 15.

amplificationally complete set of germs up to depth 6 were considered (this is the default option).

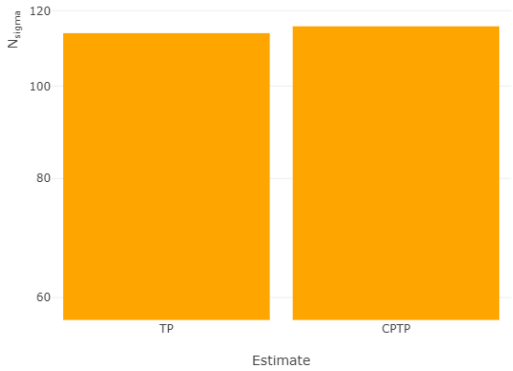


Figure 24: Model violation summary for the TP and CPTP gate set models. N_σ was defined in Eq. (50). Its high values are indication of non markovianities. Taken from [41].

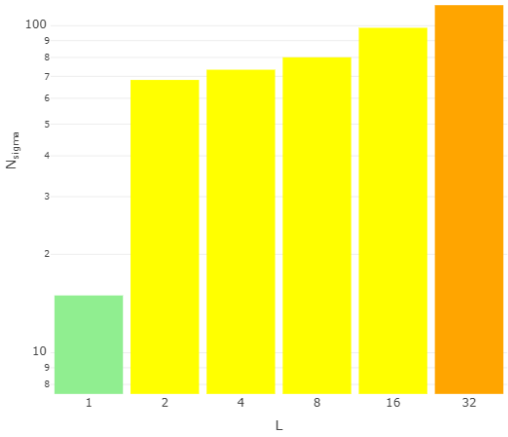


Figure 25: Model violation including circuits up to formal depth L for the CPTP gate set model. A different MLE optimal model is considered for any column of the graph. L = 1 and 2 contribute the most to N_σ. The green - red scale stands for small - great model deviations. Taken from [41].

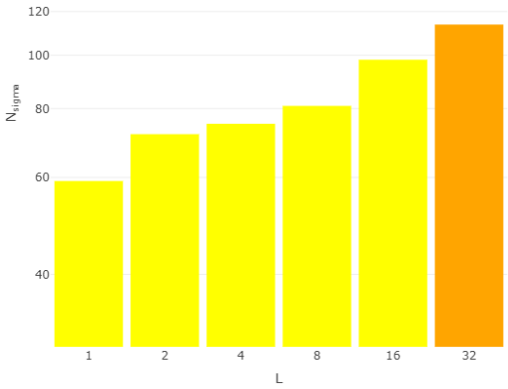


Figure 26: Model violation including circuits up to formal depth L for the TP gate set model. A different MLE optimal model is considered for any column of the graph. L = 1 and 2 contribute the most to N_σ. Taken from [41].

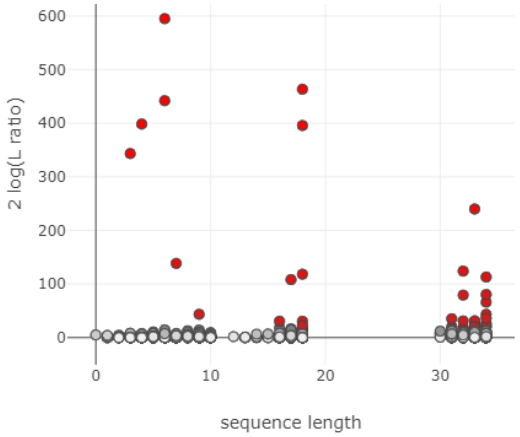


Figure 27: Model violation for any single circuit for the CPTP model. On the X axis the actual length (not the formal one) of the circuits. On the Y axis the equivalent of Eq. (50) for 1 d.o.f. (up to a a factor $\sqrt{2}$). Model discrepancies (non markovianities) are greater for shorter circuits. Red means highly non markovian. Taken from [41].

pyGSTi assumes Markovian noise. Deviations from this assumption are quantified by N_σ as defined in Eq. (50). Figs. 24 to 26 reveal that non markovianities are present, both in the CPTP and TP model, since $N_\sigma \gg 1$. No significant differences can be observed, in the whole report, between the CPTP and TP gate set model results. Therefore, from now on, only the CPTP gate set model will be dealt with because more physical than the TP one. According to Fig. 25 most of the non markovianities stem from circuits with short formal length. The fact that non markovianities are more relevant for short GST circuits is confirmed by Fig. 27. This plot shows the N_σ for each single performed circuit (up to a $\sqrt{2}$ factor) as a function of their actual length. Greater discrepancies occur for shorter lengths, i.e. non markovianities are more "relevant" for shorter circuits, which indicates a short noise correlation time, in accordance to the DRB data. The element by element characterization of every gate is provided by the pyGSTi report. This dissertation reports of other quantities, e.g. Table 7[41]. Leaving out Z

[41] Gauge dependent quantities, like the ones of Table 7, depend on the gauge and therefore also on the gauge optimization

rotations, for which no pulses is run, underrotations affect X and Y gates. The relative error amounts to 1-4 parts in 1000. Compared to the measured π and $\pi/2$ times, this is equal to an error of order 10 ns in time calibration. Z gates are subjected instead to relative overrotations of order 10^{-4}. Z gates are less noisy because no actual pulses is involved, but only an AOM phase modification. Rotation axes have only one coefficient equal

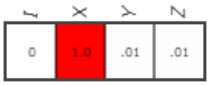
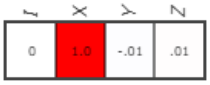
gate	rotation angle [π]	rotation axis	average gate infidelity
$X(+\pi)$	0.998757		0.00125 ± 0.005829
$Y(+\pi)$	0.996638		0.001082 ± 0.002774
$Z(+\pi)$	1.00023		0.000105 ± 0.009126
$X(+\pi/2)$	0.498118		0.000406 ± 0.003332
$Y(+\pi/2)$	0.498985		0.000597 ± 0.00163
$Z(+\pi/2)$	0.500212		0.000058 ± 0.002454

Table 7: The rotation angles, the rotation axes coefficients and the average gate infidelities are shown for the Beryllium GST data. The rotation axes coefficients are presented rounded.. 3m3 stands for 0.003. The pictures in the second last column are taken from the pyGSTi report [41]. No error bars are available for the first two columns and they are not reported for the third one. The error bars for the fourth column are of 95% confidence level. They are not reliable because non markovianities were observed.

to 1 in the noiseless case with the others equal to 0. Despite the unwanted coefficients being of order 10^{-2}, the error on the rotational axes should be considered of order 10^{-4}, because the relevant normalization is the sum of the squared coefficients. The average gate infidelities, defined in analogy to Eq. (12) but without averaging over more gates, are of order $10^{-3} - 10^{-4}$. These cannot be directly compared to $1 - p$, obtained via the DRB experiment. The former are gauge variant, the latter gauge independent. Table 8 illustrates Segtrap SPAM performances with Beryllium. The state preparation has 0.035% probability of preparing a dark state; the probability of a dark state being measured as bright is 0.014% and 0.38% vice versa. The provided uncertainties are unreliable: they are calculated based on the local asymptotic normality assumption (see Section 2.3.3). The pyGSTI report includes information on the type of errors affecting the gates. The error generator formalism is adopted, i.e. a noise gate $\tilde{G}$ and the ideal one G are related in the following manner:

$$\tilde{G} = e^{\mathbb{L}} G. \tag{64}$$

$\mathbb{L}$ is the error generator; it modifies the gate after its application. The report also offers a decomposition, written in the Pauli basis, of $\mathbb{L}$ in unitary errors, stochastic errors[42] and affine errors[43]. Listing all of this information in these pages would overload them. An example is provided for the $X(\pi)$ gate. Fig. 28 indicates that the $X(\pi)$

procedure used by pyGSTi. Several were used in the analysis but they do not show significant variations. From now on the *standard gauge optimization* data is presented.

[42] These are the ones appearing in the Kraus decomposition of a noise process.

[43] An affine error has the following action on a state ρ: $\rho \to \mathrm{Tr}(\rho) B_i$, with B_i some vector.

operation	probabilities
state preparation	$\begin{pmatrix} 0.999648 \\ 0.000352 \end{pmatrix} \pm \begin{pmatrix} 0.00588 \\ 0.00588 \end{pmatrix}$
bright state measurement	$\begin{pmatrix} 0.996212 \\ 0.000139 \end{pmatrix} \pm \begin{pmatrix} 0.011254 \\ 0.001552 \end{pmatrix}$
dark state measurement	$\begin{pmatrix} 0.003788 \\ 0.999861 \end{pmatrix} \pm \begin{pmatrix} 0.011254 \\ 0.001552 \end{pmatrix}$

Table 8: SPAM characterization. Error bars are of 95% confidence level. Like for Table 7, they are not meaningful. The last two rows of the table are linearly dependent: each of them is the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ − the other.

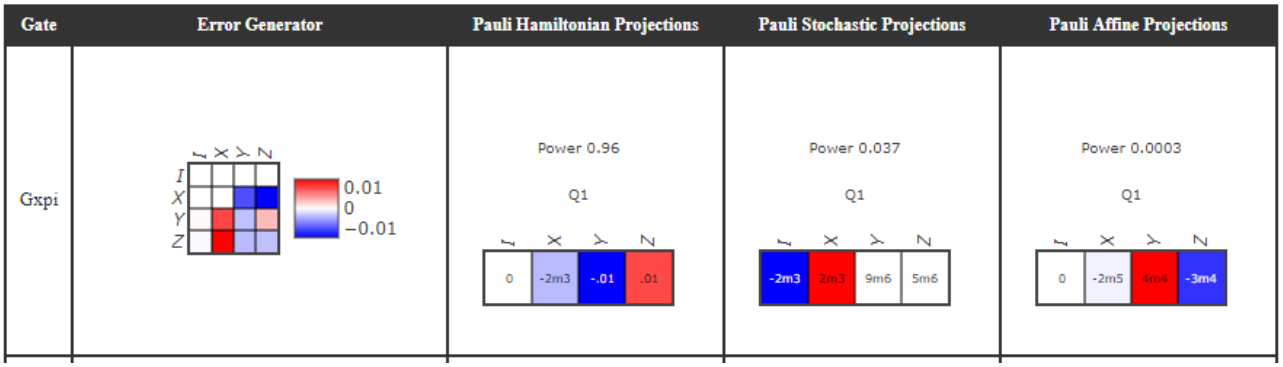


Figure 28: Error generator and its decomposition for the $X(\pi)$. Gxpi stands for $X(\pi)$. A blue-red scale is used to show how negative-positive a number is. From left to right, the columns present the unitary, stochastic and affine part of the error generator. To the best of this author's understanding, *Power* is just a multiplicative factor of the related coefficients. From the powers and the coefficients it can be observed that non unitary errors are negligible. In addition, X, Y and Z axes are all affected by coherent errors.

gate is almost exclusively affected by coherent errors and that more axes are subjected to them[44]. The original report highlights that this is the case for all the gates used: coherent noise is the principal source of noise and it manifest along not just one axis. When these data are related to the hypotheses underlying the noise analysis method employed for the DRB data, GST data confirms that noise is coherent and indicates that it does not occur just for one axes. It is somewhat surprising that the DRB infidelity distributions are explainable by just one Γ distribution and not a sum of them.

[44]Error bars, not shown here, are available for these values as well. As already mentioned, they are not considered credible.

6.2 Calcium ion

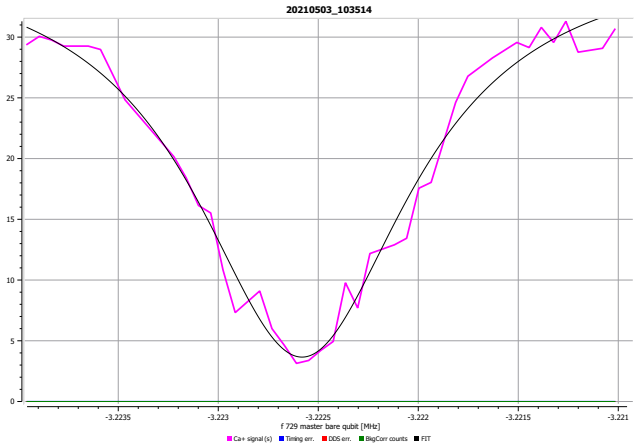


Figure 29: 729 bare frequency calibration. On the Y axis photon counts. The resonance with few photon counts indicates the qubit frequency.

For Calcium GST and DRB data were not collected in the same occasion. This is not ideal, since the two data sets portray the setup in two different experimental conditions. This is what was possible due to time limitations and to the GST absence in the RBF when the DRB experiment was performed. For both experiments, the initial calibrations included the 729 bare and dressed frequency. For the GST experiment, the π and $\pi/2$ times were also calibrated, with values around 3 and 1.5 μs respectively, keeping fixed and equal the laser intensities. For the DRB experiment instead, the intensities were calibrated and the times kept fixed at around 3 and 1.5 μs. This was not ideal. The Stark shift phase correction worked in proper conditions for the GST experiment and not for the DRB one. During the former, pulse times where periodically calibrated and the fixed intensities allowed the same dressed frequency to be kept and therefore the Stark shift correction, too. During the latter, pulse intensities periodically changed due to calibrations. Therefore also the dressed frequency and the phase accumulated with the drive on varied, since the Stark shift was calibrated only at the beginning. The relative intensity changes throughout the experiment amounted to $\sim$0.01. Having observed changes of only this order of magnitude, it is believed that, even though not optimal, the Stark shift phase tracking was effective. For both experiments, the 729 bare frequency was also periodically updated. Shaped pulses were used, in order to prevent unwanted states excitations, and it was kept track of the Stark shift phase. During both experiments, calibrations were repeated every $\sim$3 minutes.

Before the execution of the GST experiment, jumps of hundreds of MHz were observed for the 729 laser frequency. After optimizing the incoming power to the 729 cavity, this problem was mitigated.

6.2.1 DRB data and results

The set of lengths and number of shots are the same of Section 6.1.1. With Ca^+ fewer randomizations could be run: $K = 184$. The far detuned cooling time had been set 10 times longer than needed: this slowed down the experiment overall and fewer data could be sampled. In addition, some randomizations could not be performed because their longest circuit made the SDK crash for the reasons explained in Section 4.2.2. The raw data is available in the GIT repository [42]. The same software routines of Section 6.1.1 and theoretical model (Eq. (63)) have been employed for this data.

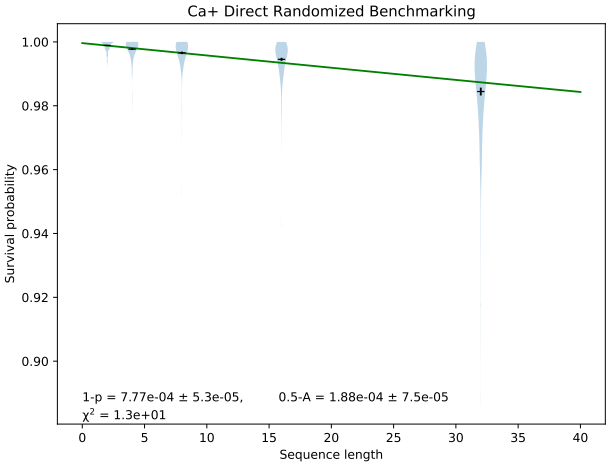


Figure 30: Exponential DRB decay for Ca^+.

The data are portrayed in Fig. 30. Long tails are visible and their length increases with J_j. This, together with $\chi^2 = 13$, is a sign of a long noise correlation length. Comparing Tables 5 and 9 it is deduced that Calcium SPAM errors are smaller than the Beryllium case, but for the latter $1 - p$ is smaller. Therefore less SPAM noise but more gate noise affects Calcium than Beryllium. This could be caused by the wrong tracking of the Stark shift phase.

$1-p$	$0.5-A$	χ^2
$7.77\times10^{-4}\pm5.3\times10^{-5}$	$1.88\times10^{-4}\pm7.5\times10^{-5}$	13

Table 9: Fit results for Eq. (63) for Ca^+. The data are shown in Fig. 30.

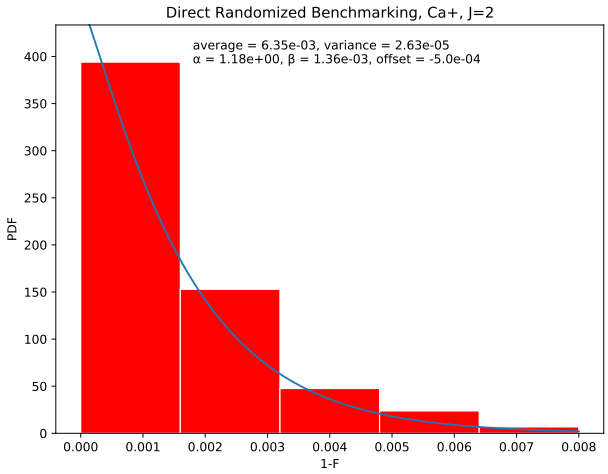


Figure 31: Distribution of $1-F(2)$, Ca^+.

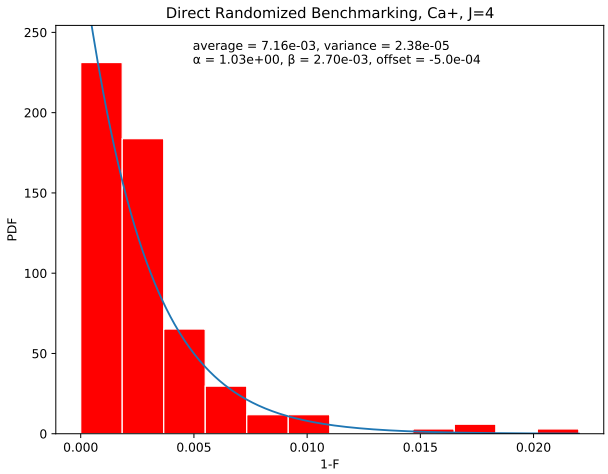


Figure 32: Distribution of $1-F(4)$, Ca^+.

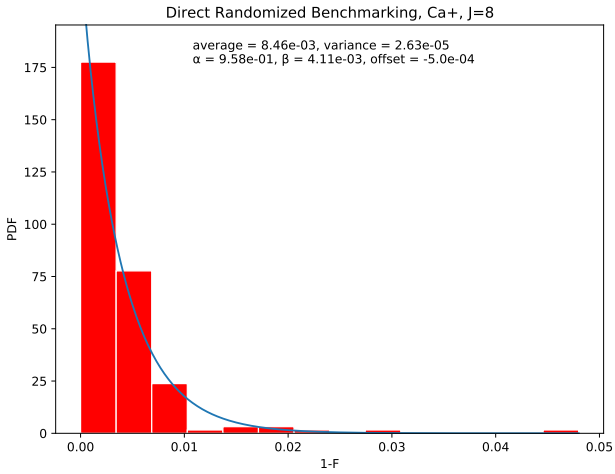


Figure 33: Distribution of $1-F(8)$, Ca^+.

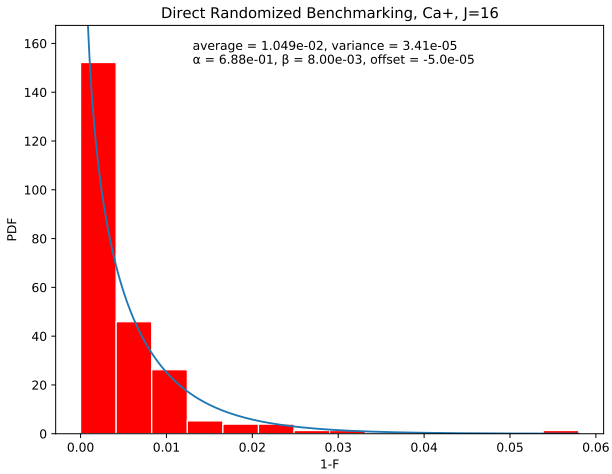


Figure 34: Distribution of $1-F(16)$, Ca^+.

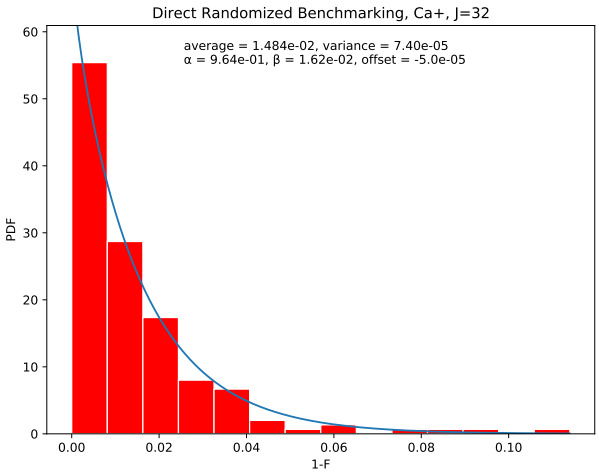


Figure 35: Distribution of $1 - F(32)$, Ca^+.

Figs. 31 to 35 show, for each value of J_j, 1-the survival probability produced by the $K = 184$ DRB circuits. These distributions are gamma distributions.

J	average	variance	alpha	beta	offset
2	6.35×10^{-3}	2.63×10^{-5}	1.18×10^{0}	1.36×10^{-3}	-5.0×10^{-4}
4	7.16×10^{-3}	2.38×10^{-5}	1.03×10^{0}	2.70×10^{-3}	-5.0×10^{-4}
8	8.46×10^{-3}	2.63×10^{-5}	9.58×10^{-1}	4.11×10^{-3}	-5.0×10^{-4}
16	1.049×10^{-2}	3.41×10^{-5}	6.88×10^{-1}	8.00×10^{-3}	-5.0×10^{-5}
32	1.484×10^{-2}	7.40×10^{-5}	9.64×10^{-1}	1.62×10^{-2}	-5.0×10^{-5}

Table 10: Summary table with the parameters of Figs. 31 to 35.

In accordance to the previous aurgumentations and to the simulations of Section 3.2, the highly skewed Γ distributions portrayed by Figs. 31 to 35 indicate DC noise. Thus, the noise correlation length is greater than 32. Adding a correction of 3 gates as for Beryllium, the effective circuit length for $J_j = 32$ becomes 35. Considering the π and $\pi/2$ times, the noise correlation time can be bounded to be > 35 gates $\times\, 1.5\mu s = 52.5\mu s$. Because both Calcium and Beryllium ions experience the same magnetic field, the different noise correlation times for Be and Ca could be due to laser frequency shifts. As for Beryllium, one should verify with GST whether the assumptions 2 and 3 of Section 6.1.1 are verified.

6.2.2 GST data and results

The same gate set models, germs and gauge optimizations of Section 6.1.2 were included in this analysis. The set of formal lengths was however different: [1, 2, 4, 8, 16, 24]. During the DRB experiment it was noticed that some circuits with $J_j = 32$ could not be performed because too long (see Section 4.2.2). GST does not admit an incomplete data set as the basis of its analysis. Therefore it was made sure that every circuit was short enough to be run. The report is available at [43] and the raw data file `gst_segtrap_ca.txt` in the superior.

Fig. 38 shows interestingly that non markovianities are highly less relevant for the model including circuits with L up to 4 than for L up to 1 for the TP model. The same plot fot the CPTP model, Fig. 37 does not confirm this phenomenon. This anomaly is thought to be caused by the numerical loglikelihood maximization for the TP gate set model getting stuck in a local maximum. Any other data in the report does not vary significantly between the TP and CPTP models. The results of the CPTP one will be from now on shown because more physical and because Fig. 38 does not seem reasonable. Fig. 39 evidences that longer circuits manifest greater non markovianities, confirming the DRB results proposing DC noise. This trend is the opposite of the one noticed in Fig. 27 where the noise had a shorter correlation length.

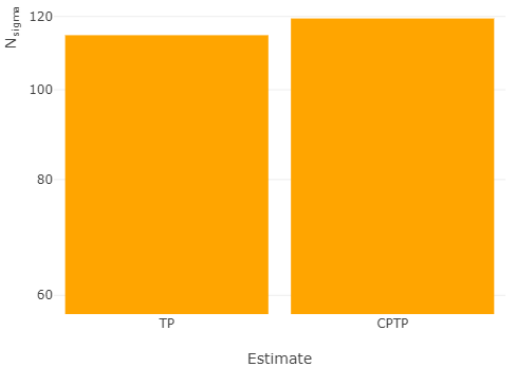


Figure 36: Model violation summary for the TP and CPTP gate set models. N_σ was defined in Eq. (50). Its high values are indication of non markovianities. Taken from [43].

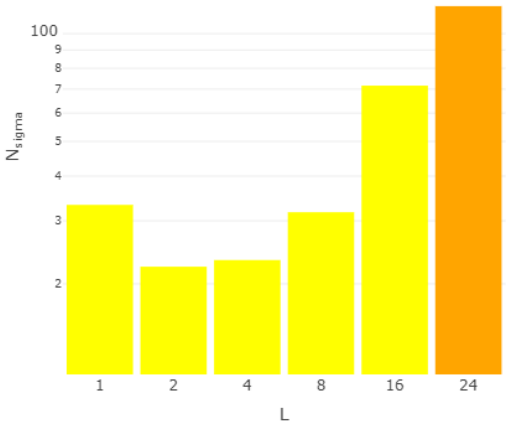


Figure 37: Model violation including circuits up to formal depth L for the CPTP gate set model. A different MLE optimal model is considered for any column of the graph. Taken from [43].

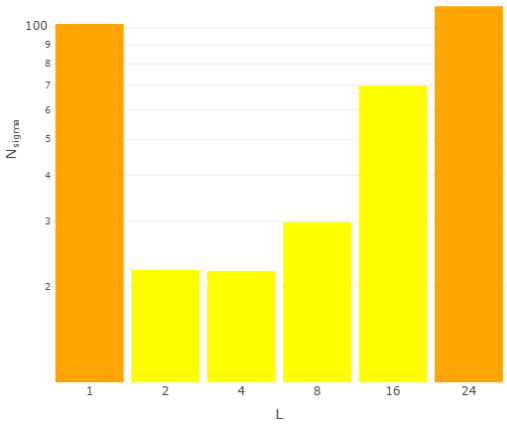


Figure 38: Model violation including circuits up to formal depth L for the TP gate set model. A different MLE optimal model is considered for any column of the graph. Taken from [43]. Non markovianities appear to be less relevant for models including short and medium lengths.

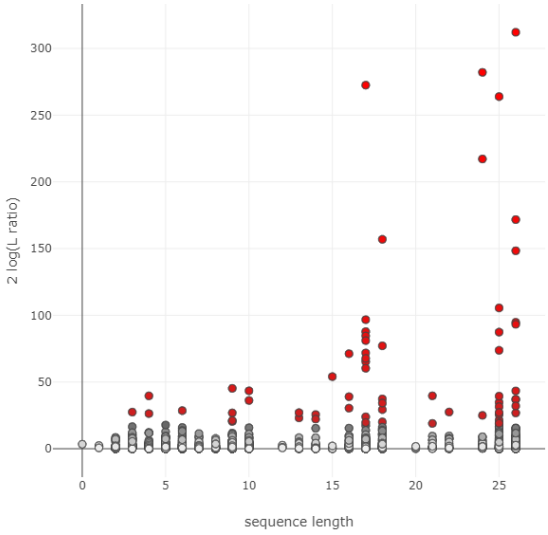


Figure 39: Model violation for any single circuit for the CPTP model. On the X axis the actual length (not the formal one) of the circuits. On the Y axis the equivalent of Eq. (50) for 1 d.o.f. (up to a a factor $\sqrt{2}$). Model discrepancies (non markovianities) are greater for longer circuits. Red means highly non markovian.

Compared to Table 7 for Beryllium, Table 12 displays similar rotation axis deviations from the noiseless case and similar infidelities but with higher uncertainties[45]. The interesting aspect of Table 12 is that Segtrap underrotates gates with Calcium even more than with Beryllium, with a relative error on the angles that reaches the 1% level. This could be an effect of the 729 laser frequency drifts observed in the laboratory. These greater underrotations also support that $1 - p_{\mathrm{Ca}} > 1 - p_{\mathrm{Be}}$ for DRB, which is a sign of better performance of Segtrap when performing logic gates on Beryllium. When it comes to SPAM operations, Segtrap executes them more reliably on Calcium than Beryllium (see Table 11): the probability of preparing the dark state instead of the bright one is 10^{-4}, of a bright state being measured as dark 2×10^{-3}, of a dark state being measured as bright 1.5×10^{-4}. This is also supported by DRB data, where $|0.5 - A|_{\mathrm{Be}} > |0.5 - A|_{\mathrm{Ca}}$. The error generator

[45] As for the Beryllium GST experiment, the data reported was obtained with the standard gauge optimization.

operation	probabilities
state preparation	$\begin{pmatrix} 0.999904 \\ 0.000096 \end{pmatrix} \pm \begin{pmatrix} 0.002923 \\ 0.002923 \end{pmatrix}$
bright state measurement	$\begin{pmatrix} 0.998081 \\ 0.000152 \end{pmatrix} \pm \begin{pmatrix} 0.009458 \\ 0.002353 \end{pmatrix}$
dark state measurement	$\begin{pmatrix} 0.001919 \\ 0.999848 \end{pmatrix} \pm \begin{pmatrix} 0.009458 \\ 0.002353 \end{pmatrix}$

Table 11: SPAM characterization. The last two rows of the table are linearly dependent: each of them is the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ – the other. Error bars are not reliable.

table reports that noise is coherent for all gates (an example is Fig. 40) but it affects more than a single axis, supporting hypothesis 2. but not 3. of Section 6.1.1.

For this reason, it is surprising that the two DRB data sets can be explained by single Γ distributions. A further DRB data set on Ca^+, mentioned in Appendix E and collected before the RBF was coded, is believed to show sums of Γ distributions. Not having kept track of the Stark shift phase through this additional DRB experiment could be the explanation of these new distributions.

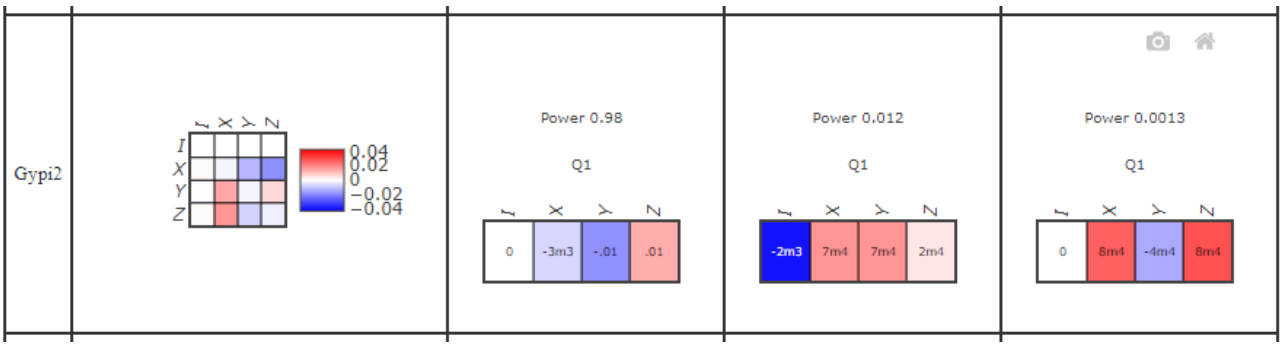


Figure 40: Error generator and its decomposition for the $Y(\pi/2)$. Gypi2 stands for $Y(\pi/2)$. From the powers and the coefficients it can be observed that non unitary errors are negligible. In addition, X, Y and Z axes are all affected by coherent errors.

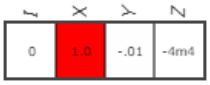
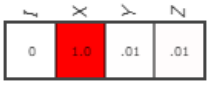
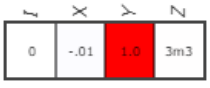
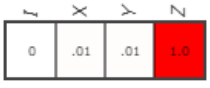
gate	rotation angle $[\pi]$	rotation axis	average gate infidelity
$X(+\pi)$	0.986758		0.001406 ± 0.01832
$Y(+\pi)$	0.988196		0.001236 ± 0.010094
$Z(+\pi)$	1.000947		0.000006 ± 0.0007
$X(+\pi/2)$	0.493296		0.000554 ± 0.00953
$Y(+\pi/2)$	0.494253		0.001113 ± 0.01642
$Z(+\pi/2)$	0.499887		0.000074 ± 0.001609

Table 12: The rotation angles, the rotation axes coefficients and the average gate infidelities are shown for the Calcium GST data. The rotation axes coefficients are presented rounded. Xm3 stands for X$\times 10^{-3}$. The pictures in the second last column are taken from the pyGSTi report [43]. No error bars are provided for the first two columns and they are not reported for the third one. The error bars are of 95% confidence level. They should not be trusted because non markovianities were observed.

7 Conlusions, outlooks and remarks for the future

A qualitative method was developed to infer the noise correlation length from 1 qubit DRB experiments subjected to coherent noise along one single axis. Numerical simulations showed different shapes for the Γ distributions describing the infidelities of DRB experiments. On the basis of these distributions shapes, i.e. whether they exhibit a monotonic behaviour or not, the noise correlation time can be bounded and/or estimated. With the simulated data it was also attempted to devise a more quantitative model, with the aim of extrapolating noise intensity and correlation length from the statistical properties of the Γ distributions of future real experiments. This was unsuccessful: the best fit linking variance of the distributions, noise intensity and correlation length is distant more than 4σ from the simulated data. This idea was inspired by the work of Ball et al. [3], which treated similar topics but for CRB. A theoretical extension of Ball's results from CRB to DRB was not possible, because the two protocols involve different, even though similar, measurable fidelities.

A set of software routines, the RBF, was coded. This allows any subgroup of the TIQI group to design and perform GST and RB experiments, except 2 qubit GST experiments, given its complex structure. The RBF is based on the pyGSTi [29] and rapidJSON [20] libraries, the `crystal` classes and the Ionizer2 APIs [22]. pyGSTi allows to design GST and RB experiments and generates their sequences, which can be very long. Therefore being aware of what circuits are run and what their expected outcomes are is crucial. JSON files were chosen for the clarity they offer. User-RBF communication and communication between different parts of the RBF are based in part on JSON files.

DRB and GST data sets were collected for the single Calcium ion and for the single Beryllium ion on Segtrap. Due to experimental problems leading to a frequent ion loss and time limitations, no 2 qubit DRB experiment with Beryllium and Calcium trapped was executed. The selected qubit transitions were respectively $\left|2^2S_{1/2}, F=1, m_F=1\right\rangle \leftrightarrow \left|2^2S_{1/2}, F=2, m_F=0\right\rangle$ and $\left|4^2S_{1/2}, m=\frac{1}{2}\right\rangle \leftrightarrow \left|3^2D_{5/2}, m=\frac{3}{2}\right\rangle$ for Beryllium and Calcium. Limitations of the M-ACTION system limited the maximum depth of an experimental sequence to be less than 40. This limitation was eased by end of this project's time but not yet implemented. It was noticed that Γ distributions can explain both Calcium and Beryllium DRB data. From these, the noise correlation time was bounded to, for the two respective cases, $> 52.5\mu$s and $\leq 20\mu$s. These values can be attributed to laser frequency shifts and magnetic field fluctuations. In addition, DRB data show that SPAM operations are performed better when Calcium is trapped and logic operations with Beryllium instead. The state survival probability per native logic gate p, with value 1 in the noiseless case, is such that $7.77 \times 10^{-4} \pm 5.3 \times 10^{-5} = 1 - p_{\mathrm{Ca}} > 1 - p_{\mathrm{Be}} = 5.69 \times 10^{-4} \pm 2.7 \times 10^{-5}$. On the other hand, considering the SPAM operation indicator A (with value 0.5 in the noiseless case), $2.995 \times 10^{-3} \pm 8.2 \times 10^{-5} = 0.5 - A_{\mathrm{Be}} > 0.5 - A_{\mathrm{Ca}} = 1.88 \times 10^{-4} \pm 7.5 \times 10^{-5}$. GST data confirms the previous remarks on the noise correlation lengths by reporting that non markovian noise leads to greater discrepancies with respect to a noiseless scenario on short sequences for Beryllium and on long sequences for Calcium. GST indicates that logic gates are affected by coherent noise that occurs along the X, Y and Z logical axes. It is believed that this contradicts the possibility of describing DRB data with single Γ distributions, which stem out of coherent errors occurring along only one axis. GST supports as well the fact that logic operations are better performed with Beryllium and SPAM operations with Calcium. Relative errors on rotation angles up to 10^{-2} were observed with Calcium and 10^{-3} with Beryllium. SPAM operations errors are for both cases of order $10^{-3} - 10^{-4}$ but, as already declared, less relevant on Calcium.

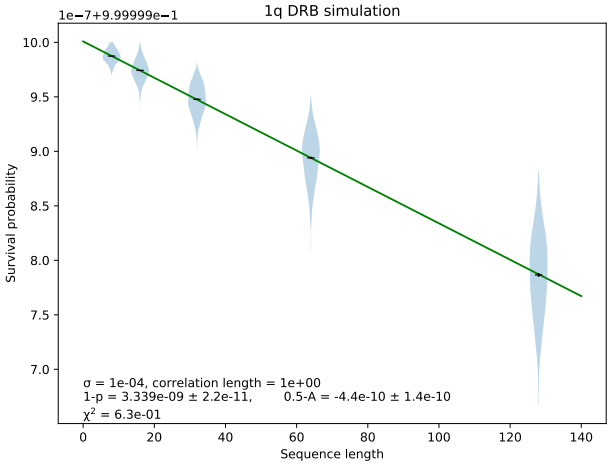
In order to continue with noise characterisation, it is important to add into the RBF the possibility of designing 2 qubit GST experiments. Together with 2 qubit DRB and CRB experiments, they would make feasible to benchmark and characterize cross talk noise and the MS gate. Of primary relevance is also implementing the new pulses coded my Martin Stadler, such that longer experimental sequences are exploitable. On the other hand, in order to solidify DRB theoretical basis more simulations and/or more analytical efforts should be put in order to develop a quantitative model able to link Γ distributions statistical moments with noise intensity and correlation length.

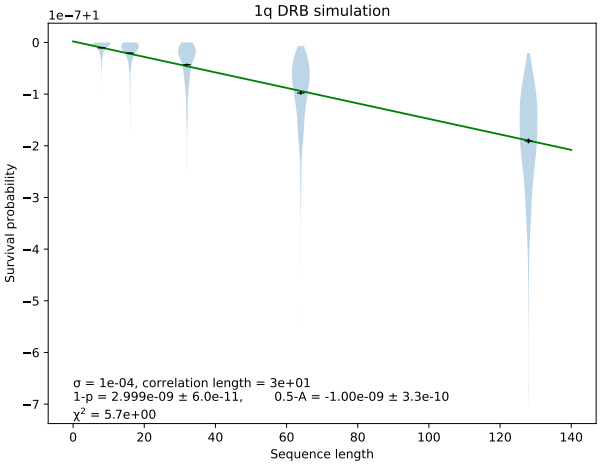
Appendix

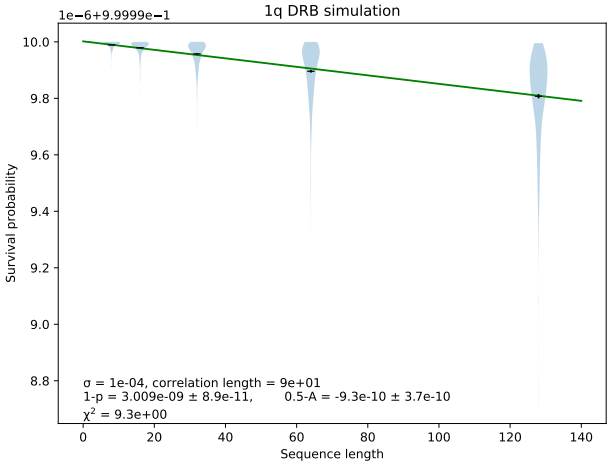
A DRB simulations. Exponential decay graphs

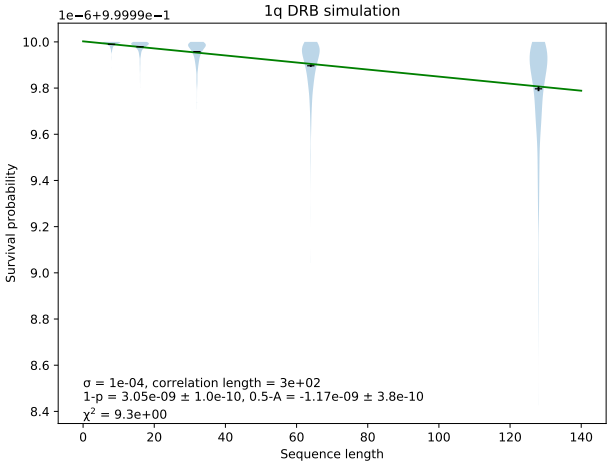
The graphs show 1-qubit DRB experiment simulations data. Noise correlation length and intensity σ vary.

$\sigma = 10^{-4}$

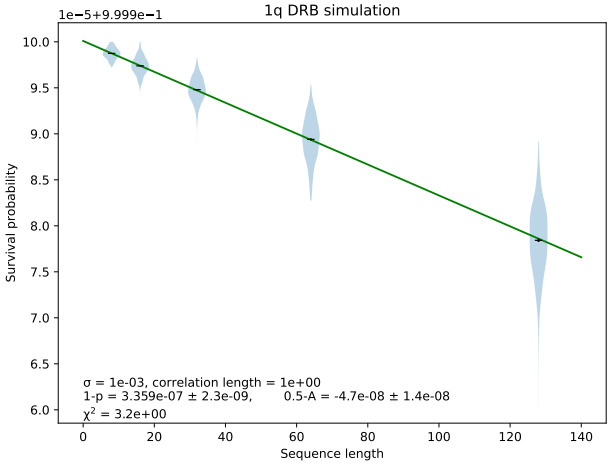


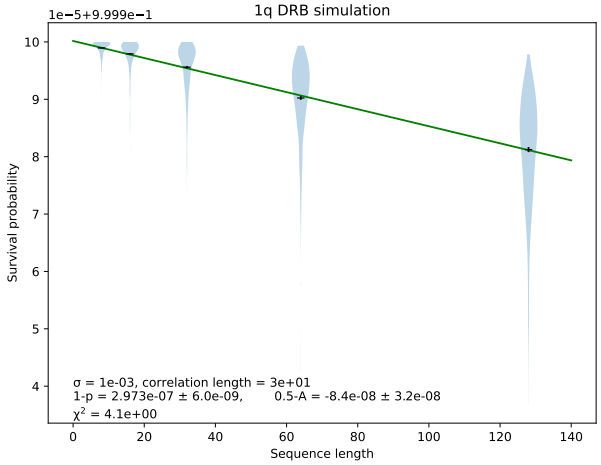


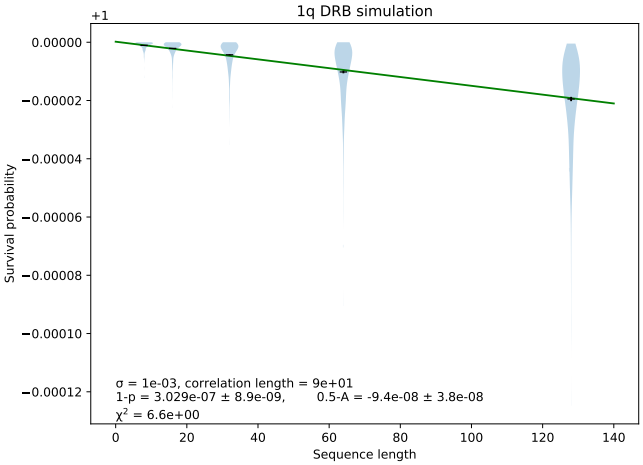


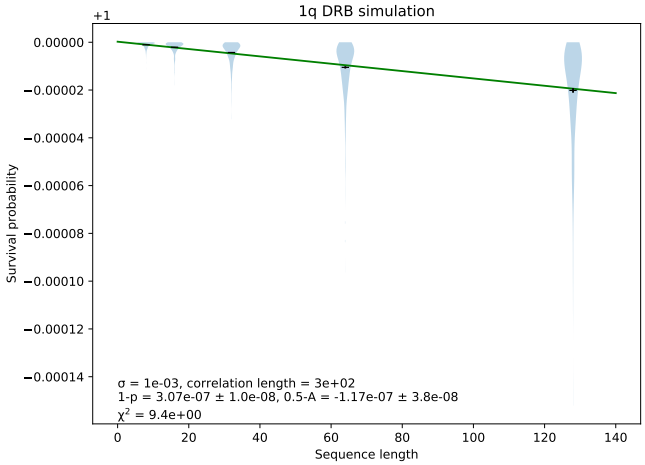


$\sigma = 10^{-3}$

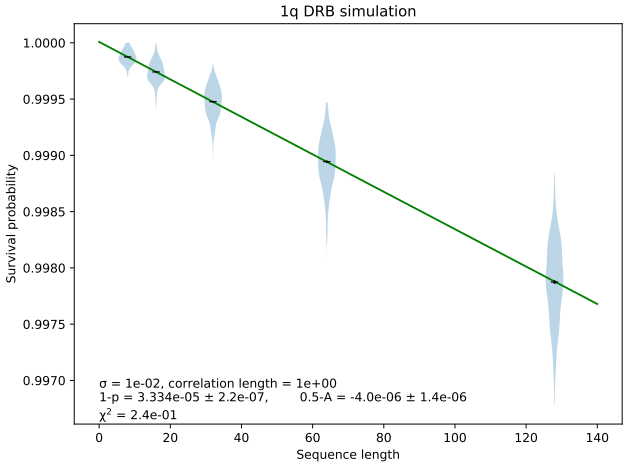


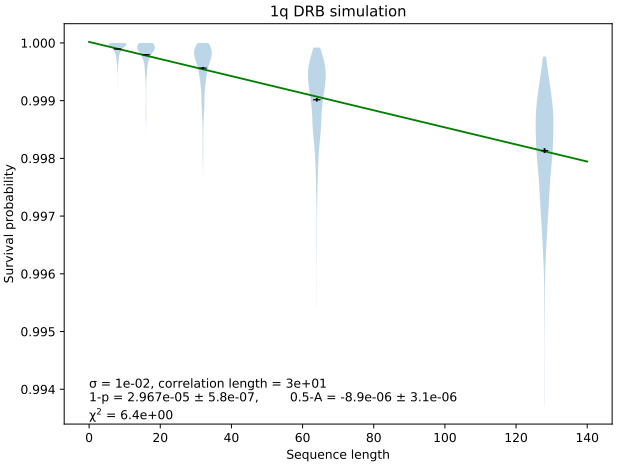


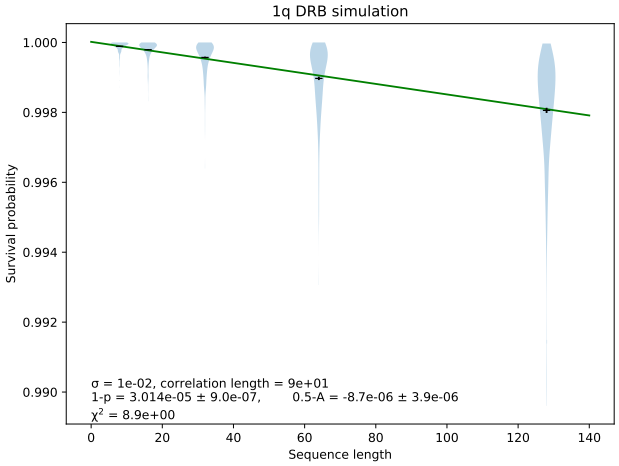

1q DRB simulation
+1
Survival probability
Sequence length
σ = 1e-03, correlation length = 9e+01
1-p = 3.029e-07 ± 8.9e-09, 0.5-A = -9.4e-08 ± 3.8e-08
χ² = 6.6e+00

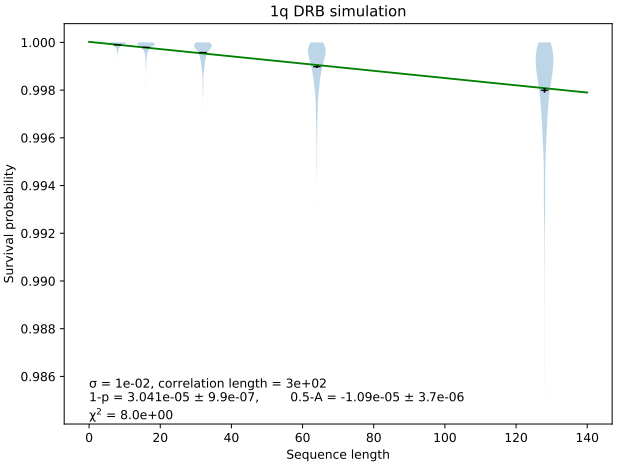

1q DRB simulation
+1
Survival probability
Sequence length
σ = 1e-03, correlation length = 3e+02
1-p = 3.07e-07 ± 1.0e-08, 0.5-A = -1.17e-07 ± 3.8e-08
χ² = 9.4e+00

$$\sigma = 10^{-2}$$

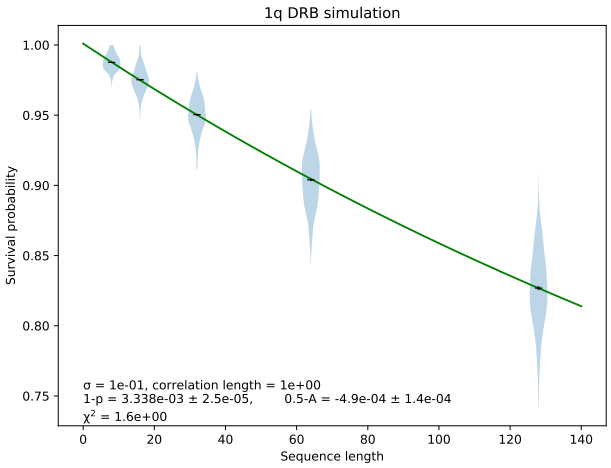

1q DRB simulation
Survival probability
Sequence length
σ = 1e-02, correlation length = 1e+00
1-p = 3.334e-05 ± 2.2e-07, 0.5-A = -4.0e-06 ± 1.4e-06
χ² = 2.4e-01

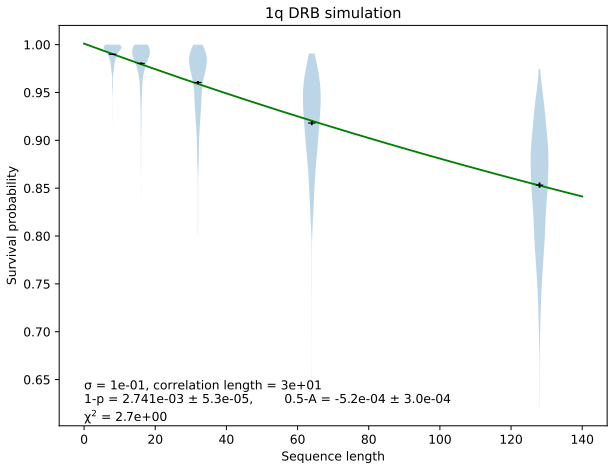

1q DRB simulation
Survival probability
Sequence length
σ = 1e-02, correlation length = 3e+01
1-p = 2.967e-05 ± 5.8e-07, 0.5-A = -8.9e-06 ± 3.1e-06
χ² = 6.4e+00

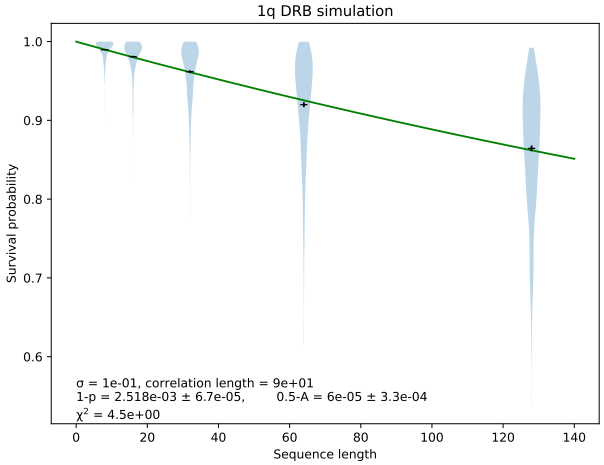

1q DRB simulation
Survival probability
Sequence length
σ = 1e-02, correlation length = 9e+01
1-p = 3.014e-05 ± 9.0e-07, 0.5-A = -8.7e-06 ± 3.9e-06
χ² = 8.9e+00

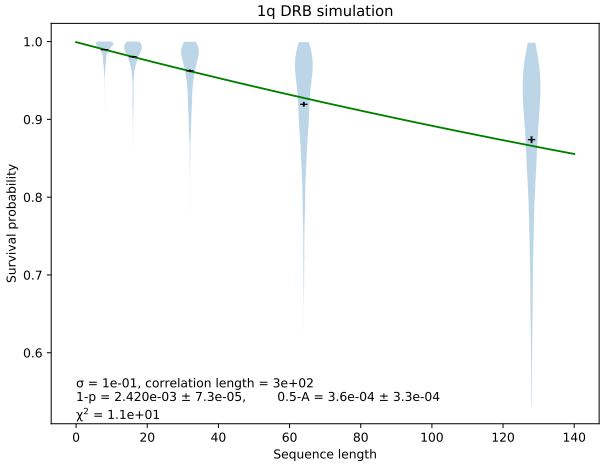

1q DRB simulation
Survival probability
Sequence length
σ = 1e-02, correlation length = 3e+02
1-p = 3.041e-05 ± 9.9e-07, 0.5-A = -1.09e-05 ± 3.7e-06
χ² = 8.0e+00

$\sigma = 10^{-1}$



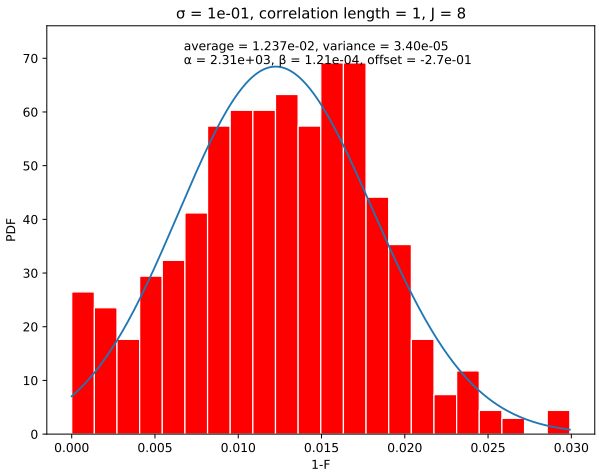


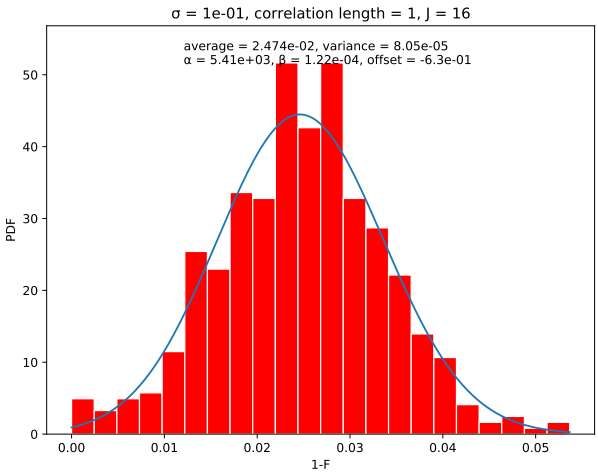


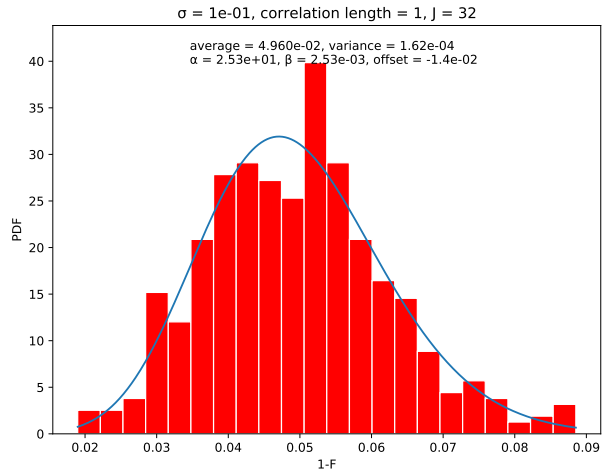


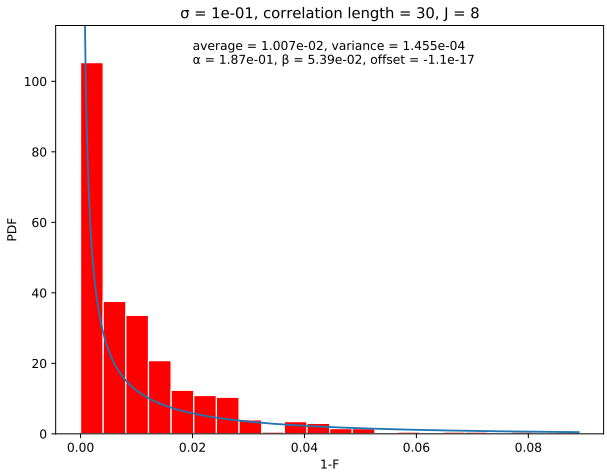
B DRB simulations. Gamma distribution graphs

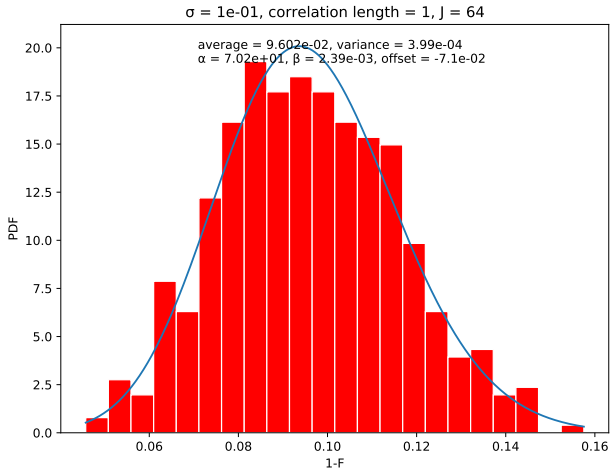
To be read per column. For summary table with all the fitted parameters see Table 13.

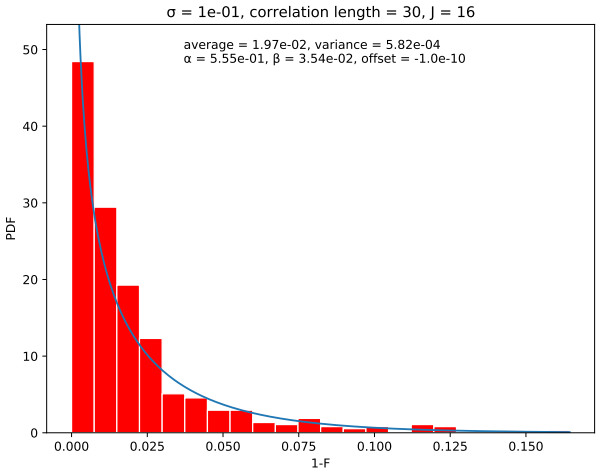


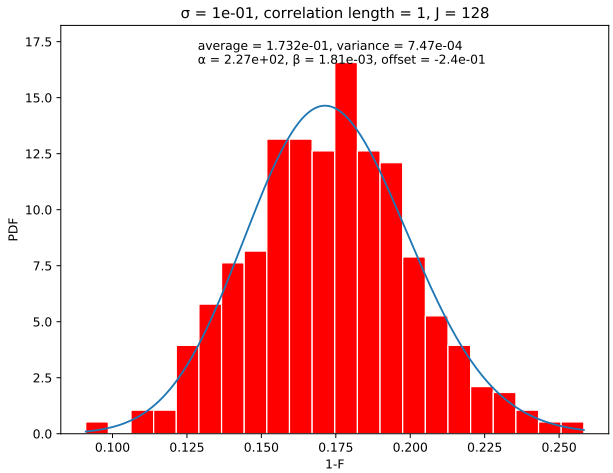


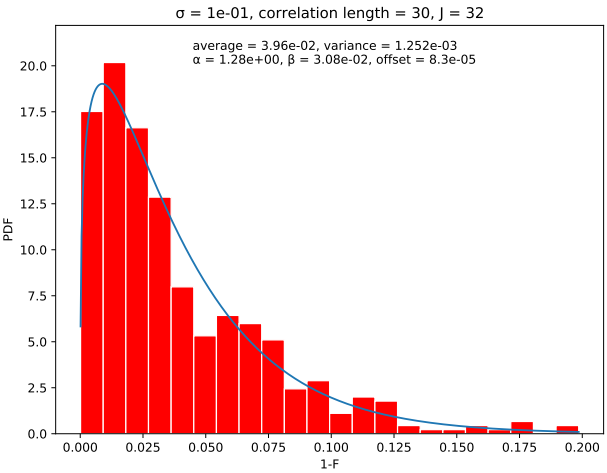

σ = 1e-01, correlation length = 1, J = 32
average = 4.960e-02, variance = 1.62e-04
α = 2.53e+01, β = 2.53e-03, offset = -1.4e-02
PDF
1-F

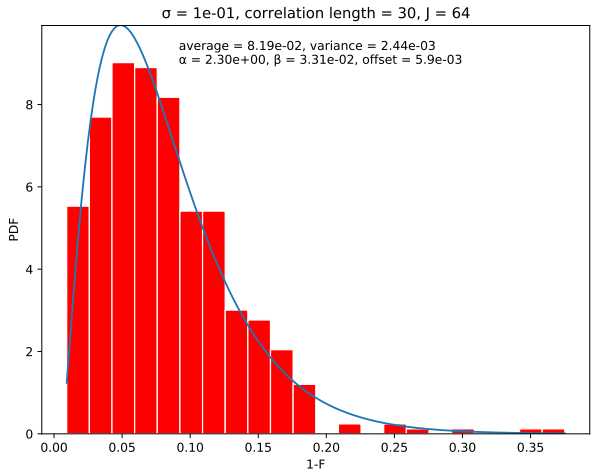

σ = 1e-01, correlation length = 30, J = 8
average = 1.007e-02, variance = 1.455e-04
α = 1.87e-01, β = 5.39e-02, offset = -1.1e-17
PDF
1-F

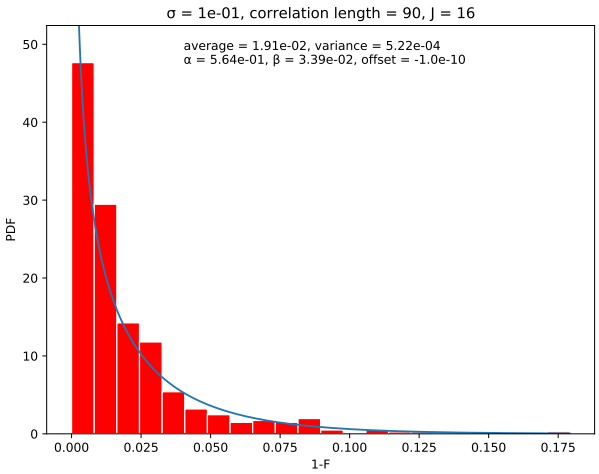

σ = 1e-01, correlation length = 1, J = 64
average = 9.602e-02, variance = 3.99e-04
α = 7.02e+01, β = 2.39e-03, offset = -7.1e-02
PDF
1-F

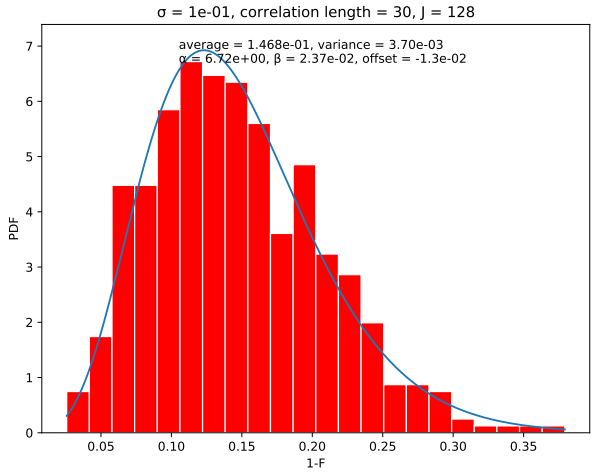

σ = 1e-01, correlation length = 30, J = 16
average = 1.97e-02, variance = 5.82e-04
α = 5.55e-01, β = 3.54e-02, offset = -1.0e-10
PDF
1-F

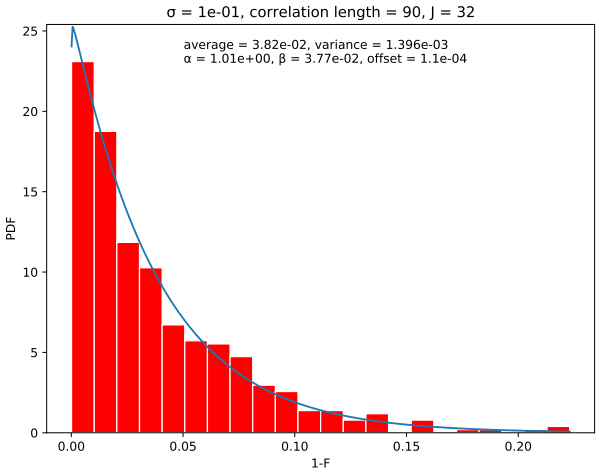

σ = 1e-01, correlation length = 1, J = 128
average = 1.732e-01, variance = 7.47e-04
α = 2.27e+02, β = 1.81e-03, offset = -2.4e-01
PDF
1-F

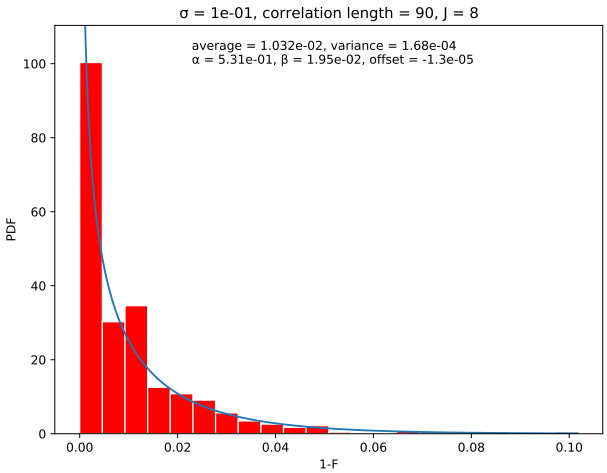

σ = 1e-01, correlation length = 30, J = 32
average = 3.96e-02, variance = 1.252e-03
α = 1.28e+00, β = 3.08e-02, offset = 8.3e-05
PDF
1-F

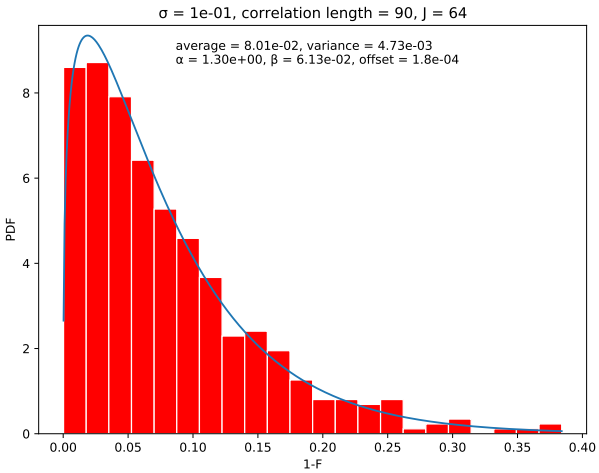

σ = 1e-01, correlation length = 30, J = 64
average = 8.19e-02, variance = 2.44e-03
α = 2.30e+00, β = 3.31e-02, offset = 5.9e-03
PDF
1-F

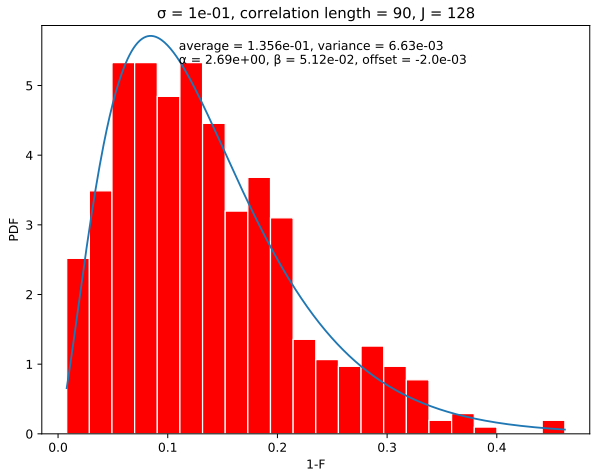

σ = 1e-01, correlation length = 90, J = 16
average = 1.91e-02, variance = 5.22e-04
α = 5.64e-01, β = 3.39e-02, offset = -1.0e-10
PDF
1-F

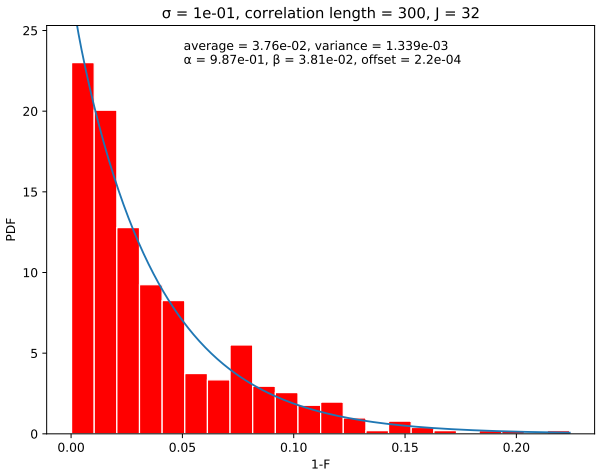

σ = 1e-01, correlation length = 30, J = 128
average = 1.468e-01, variance = 3.70e-03
α = 6.72e+00, β = 2.37e-02, offset = -1.3e-02
PDF
1-F

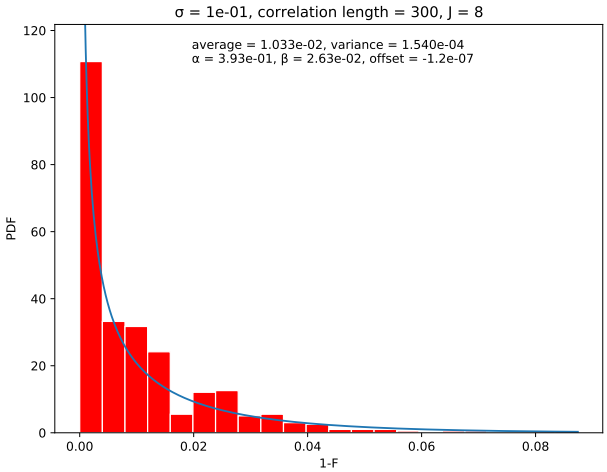

σ = 1e-01, correlation length = 90, J = 32
average = 3.82e-02, variance = 1.396e-03
α = 1.01e+00, β = 3.77e-02, offset = 1.1e-04
PDF
1-F

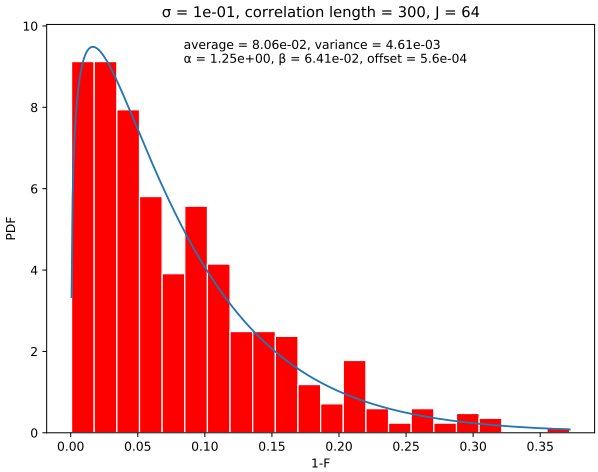

σ = 1e-01, correlation length = 90, J = 8
average = 1.032e-02, variance = 1.68e-04
α = 5.31e-01, β = 1.95e-02, offset = -1.3e-05
PDF
1-F

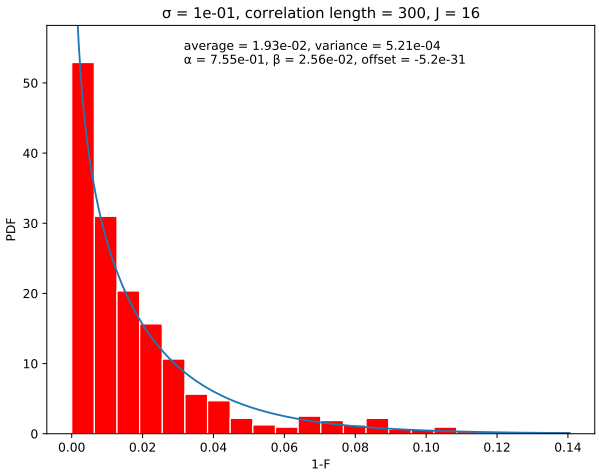

σ = 1e-01, correlation length = 90, J = 64
average = 8.01e-02, variance = 4.73e-03
α = 1.30e+00, β = 6.13e-02, offset = 1.8e-04
PDF
1-F

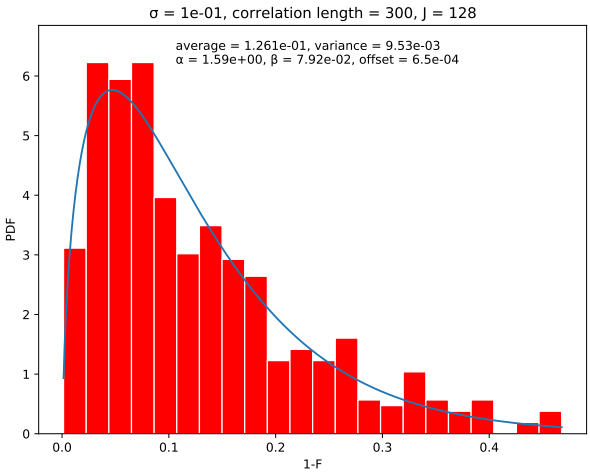

σ = 1e-01, correlation length = 90, J = 128
average = 1.356e-01, variance = 6.63e-03
α = 2.69e+00, β = 5.12e-02, offset = -2.0e-03
PDF
1-F

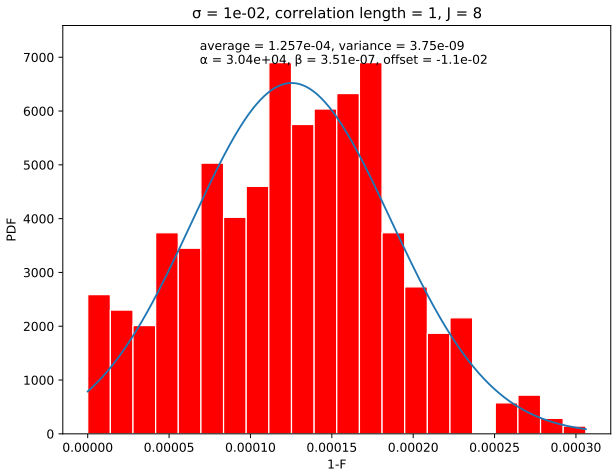

σ = 1e-01, correlation length = 300, J = 32
average = 3.76e-02, variance = 1.339e-03
α = 9.87e-01, β = 3.81e-02, offset = 2.2e-04
PDF
1-F

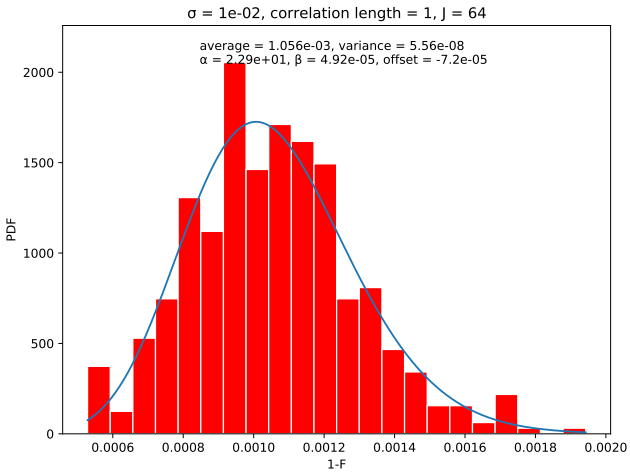

σ = 1e-01, correlation length = 300, J = 8
average = 1.033e-02, variance = 1.540e-04
α = 3.93e-01, β = 2.63e-02, offset = -1.2e-07
PDF
1-F

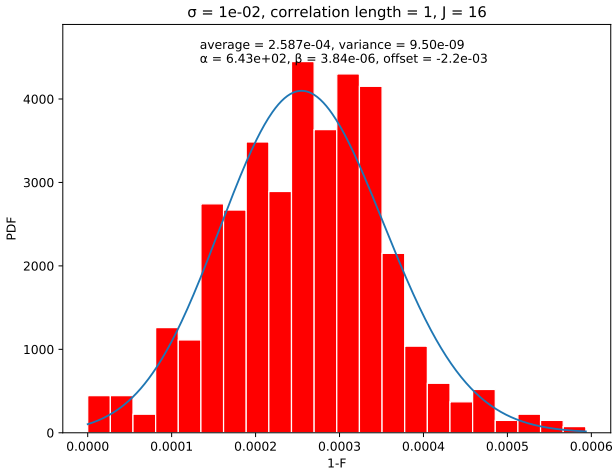

σ = 1e-01, correlation length = 300, J = 64
average = 8.06e-02, variance = 4.61e-03
α = 1.25e+00, β = 6.41e-02, offset = 5.6e-04
PDF
1-F

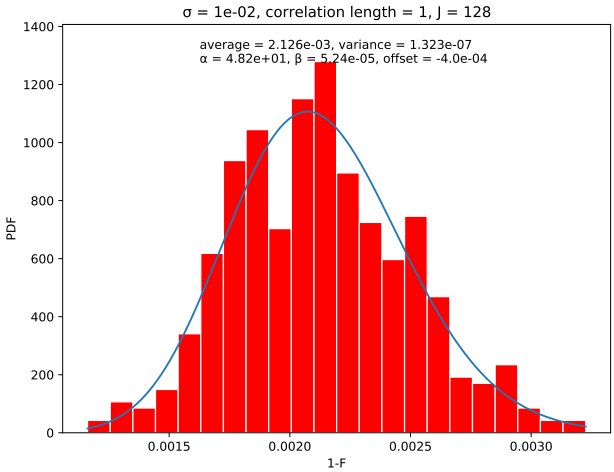

σ = 1e-01, correlation length = 300, J = 16
average = 1.93e-02, variance = 5.21e-04
α = 7.55e-01, β = 2.56e-02, offset = -5.2e-31
PDF
1-F

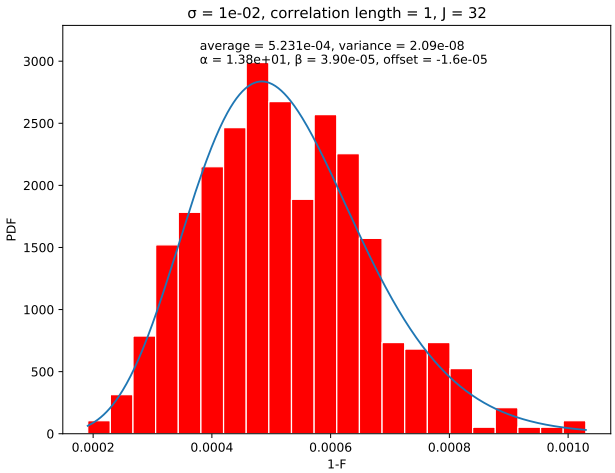

σ = 1e-01, correlation length = 300, J = 128
average = 1.261e-01, variance = 9.53e-03
α = 1.59e+00, β = 7.92e-02, offset = 6.5e-04
PDF
1-F

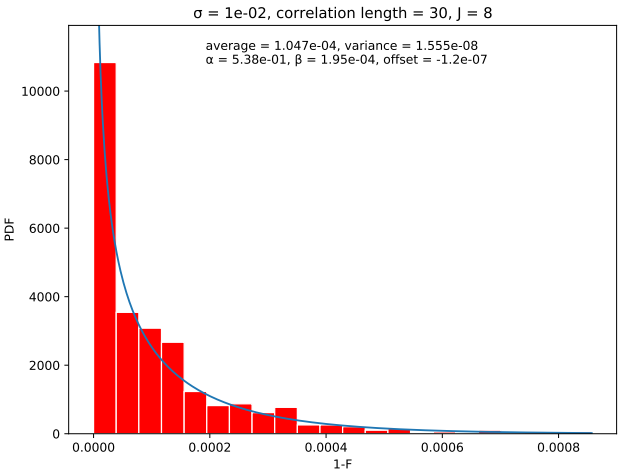

σ = 1e-02, correlation length = 1, J = 8
average = 1.257e-04, variance = 3.75e-09
α = 3.04e+04, β = 3.51e-07, offset = -1.1e-02
PDF
1-F

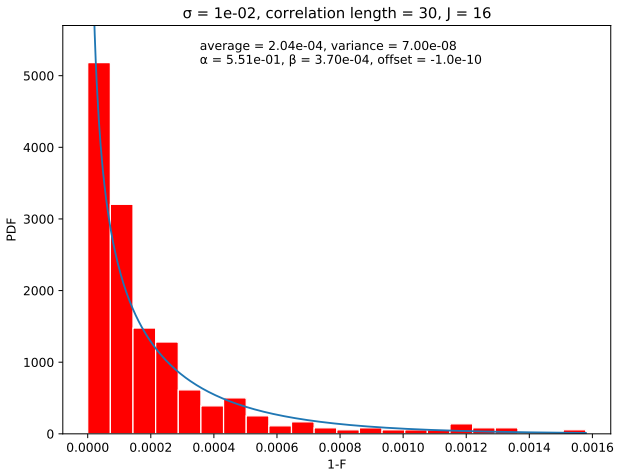

σ = 1e-02, correlation length = 1, J = 64
average = 1.056e-03, variance = 5.56e-08
α = 2.29e+01, β = 4.92e-05, offset = -7.2e-05
PDF
1-F

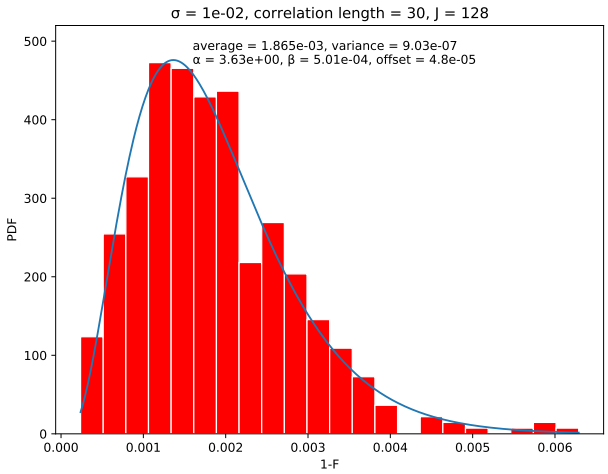

σ = 1e-02, correlation length = 1, J = 16
average = 2.587e-04, variance = 9.50e-09
α = 6.43e+02, β = 3.84e-06, offset = -2.2e-03
PDF
1-F

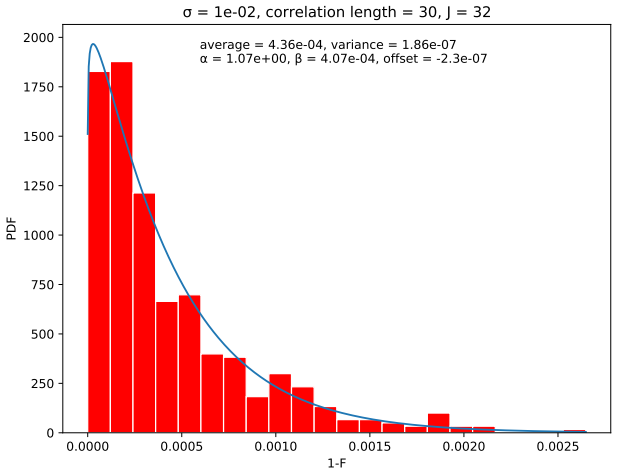

σ = 1e-02, correlation length = 1, J = 128
average = 2.126e-03, variance = 1.323e-07
α = 4.82e+01, β = 5.24e-05, offset = -4.0e-04
PDF
1-F

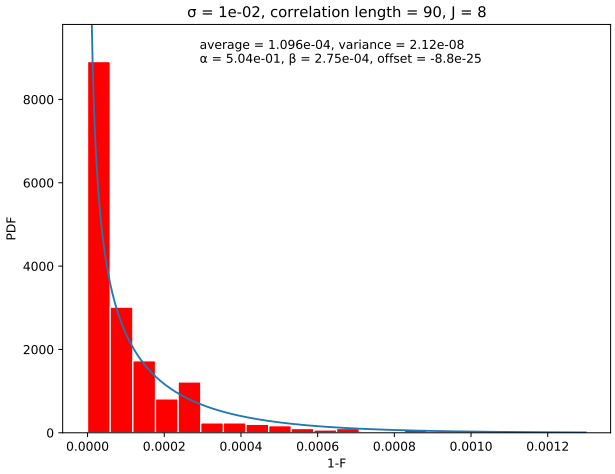

σ = 1e-02, correlation length = 1, J = 32
average = 5.231e-04, variance = 2.09e-08
α = 1.38e+01, β = 3.90e-05, offset = -1.6e-05
PDF
1-F

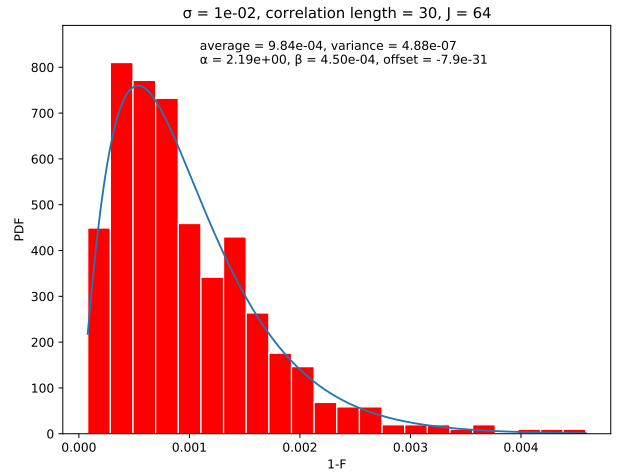

σ = 1e-02, correlation length = 30, J = 8
average = 1.047e-04, variance = 1.555e-08
α = 5.38e-01, β = 1.95e-04, offset = -1.2e-07
PDF
1-F

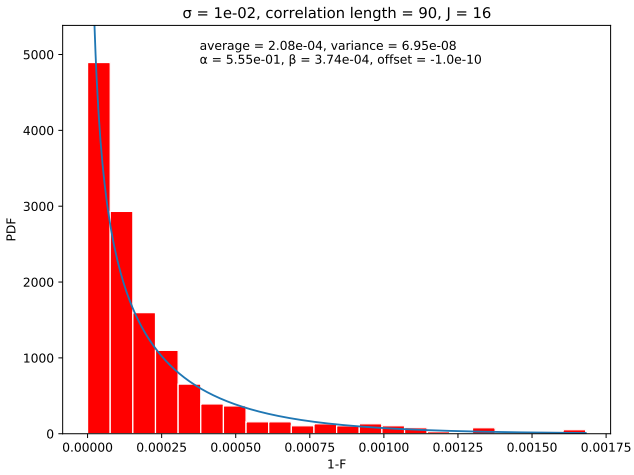

σ = 1e-02, correlation length = 30, J = 16
average = 2.04e-04, variance = 7.00e-08
α = 5.51e-01, β = 3.70e-04, offset = -1.0e-10
PDF
1-F

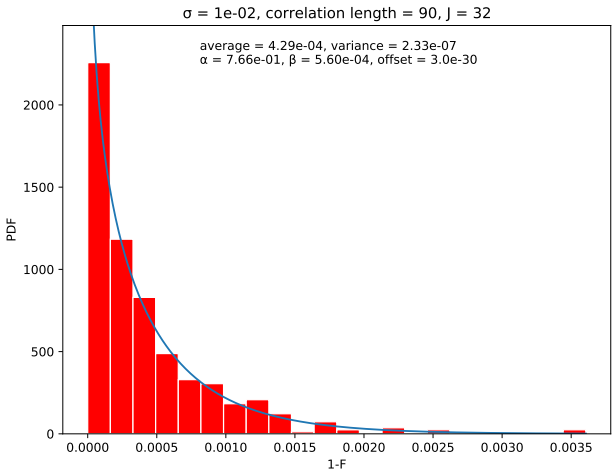

σ = 1e-02, correlation length = 30, J = 128
average = 1.865e-03, variance = 9.03e-07
α = 3.63e+00, β = 5.01e-04, offset = 4.8e-05
PDF
1-F

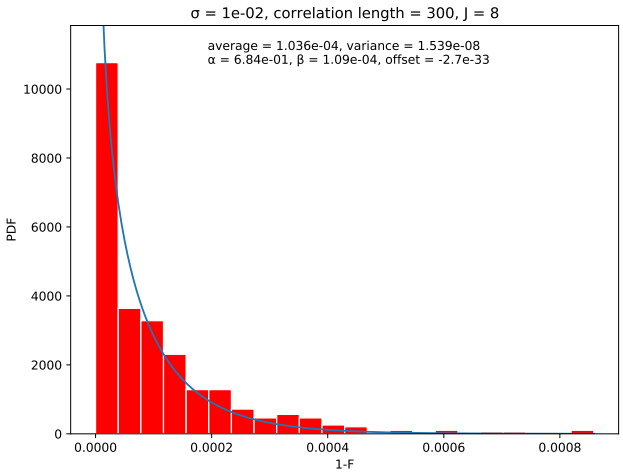

σ = 1e-02, correlation length = 30, J = 32
average = 4.36e-04, variance = 1.86e-07
α = 1.07e+00, β = 4.07e-04, offset = -2.3e-07
PDF
1-F

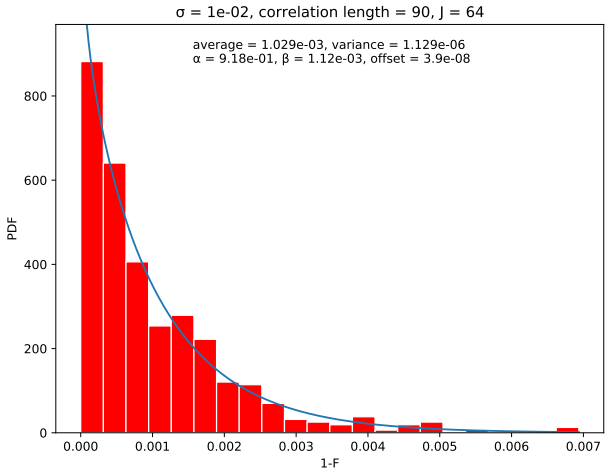

σ = 1e-02, correlation length = 90, J = 8
average = 1.096e-04, variance = 2.12e-08
α = 5.04e-01, β = 2.75e-04, offset = -8.8e-25
PDF
1-F

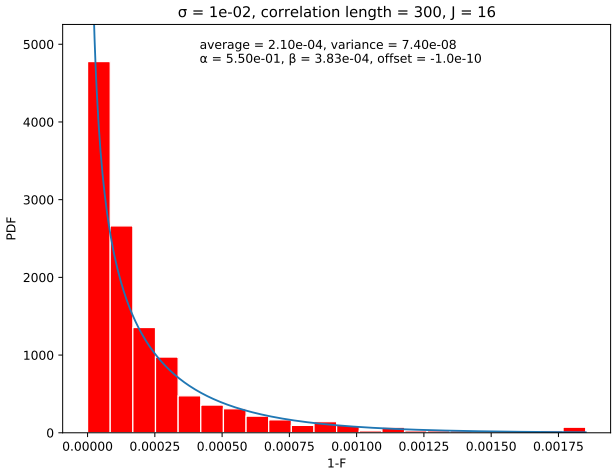

σ = 1e-02, correlation length = 30, J = 64
average = 9.84e-04, variance = 4.88e-07
α = 2.19e+00, β = 4.50e-04, offset = -7.9e-31
PDF
1-F

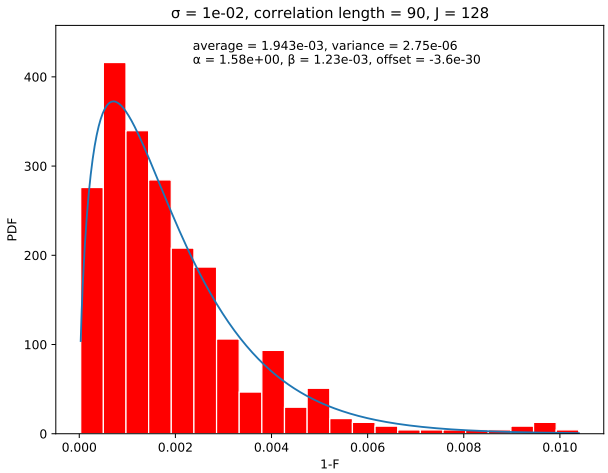

σ = 1e-02, correlation length = 90, J = 16
average = 2.08e-04, variance = 6.95e-08
α = 5.55e-01, β = 3.74e-04, offset = -1.0e-10
PDF
1-F

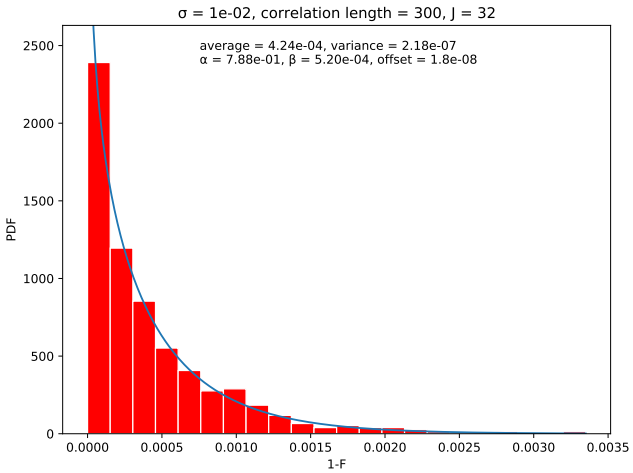

σ = 1e-02, correlation length = 90, J = 32
average = 4.29e-04, variance = 2.33e-07
α = 7.66e-01, β = 5.60e-04, offset = 3.0e-30
PDF
1-F

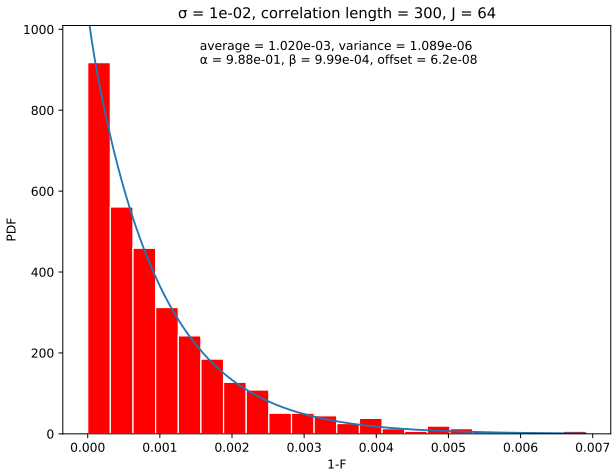

σ = 1e-02, correlation length = 300, J = 8
average = 1.036e-04, variance = 1.539e-08
α = 6.84e-01, β = 1.09e-04, offset = -2.7e-33
PDF
1-F

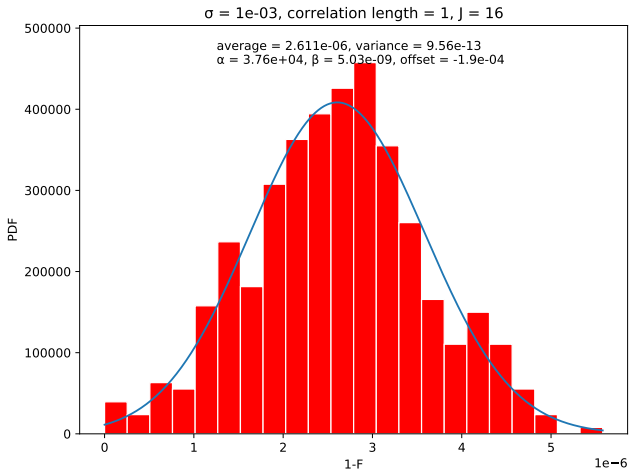

σ = 1e-02, correlation length = 90, J = 64
average = 1.029e-03, variance = 1.129e-06
α = 9.18e-01, β = 1.12e-03, offset = 3.9e-08
PDF
1-F

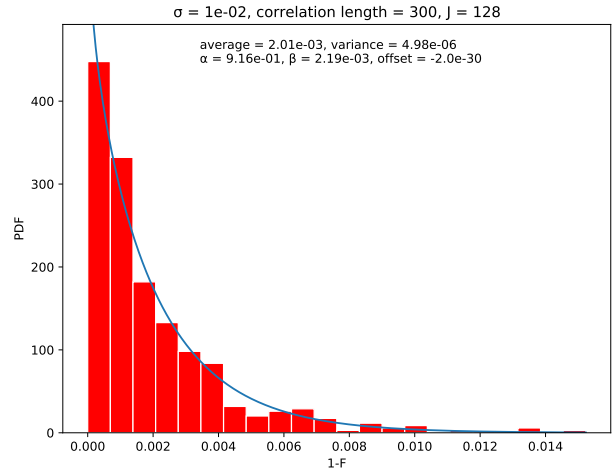

σ = 1e-02, correlation length = 300, J = 16
average = 2.10e-04, variance = 7.40e-08
α = 5.50e-01, β = 3.83e-04, offset = -1.0e-10
PDF
1-F

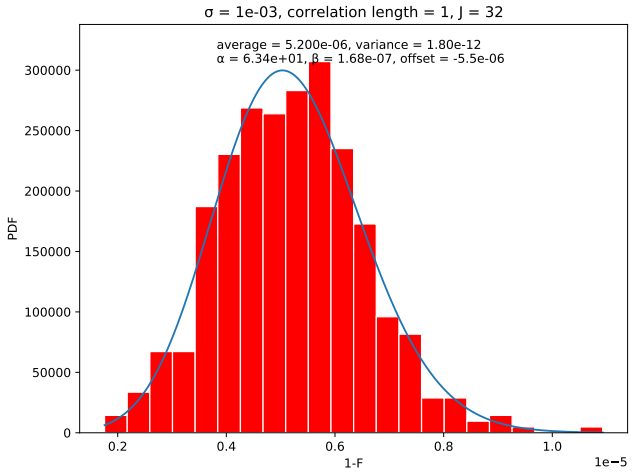

σ = 1e-02, correlation length = 90, J = 128
average = 1.943e-03, variance = 2.75e-06
α = 1.58e+00, β = 1.23e-03, offset = -3.6e-30
PDF
1-F

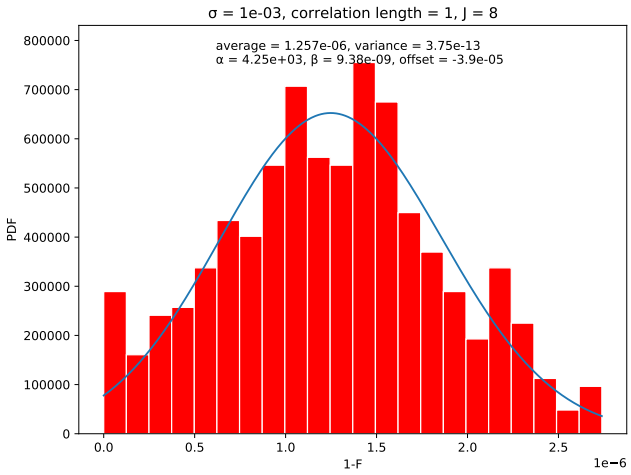

σ = 1e-02, correlation length = 300, J = 32
average = 4.24e-04, variance = 2.18e-07
α = 7.88e-01, β = 5.20e-04, offset = 1.8e-08
PDF
1-F

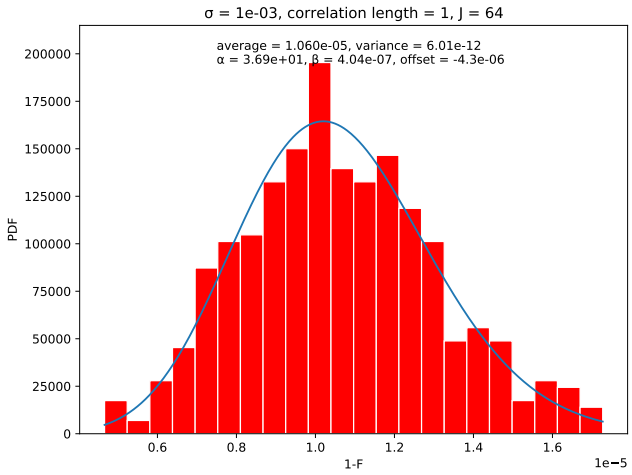

σ = 1e-02, correlation length = 300, J = 64
average = 1.020e-03, variance = 1.089e-06
α = 9.88e-01, β = 9.99e-04, offset = 6.2e-08
PDF
1-F

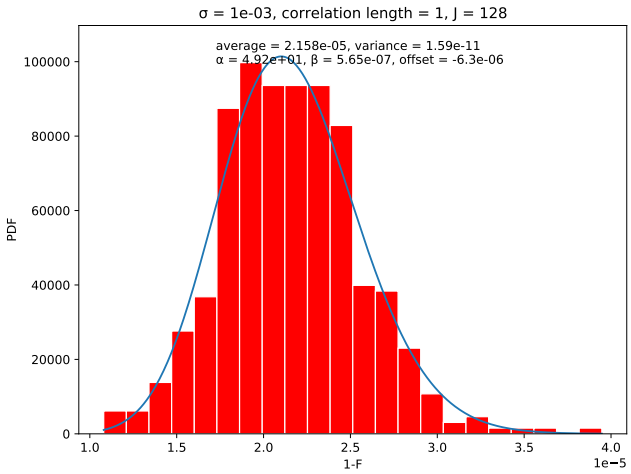

σ = 1e-03, correlation length = 1, J = 16
average = 2.611e-06, variance = 9.56e-13
α = 3.76e+04, β = 5.03e-09, offset = -1.9e-04
PDF
1-F
1e−6

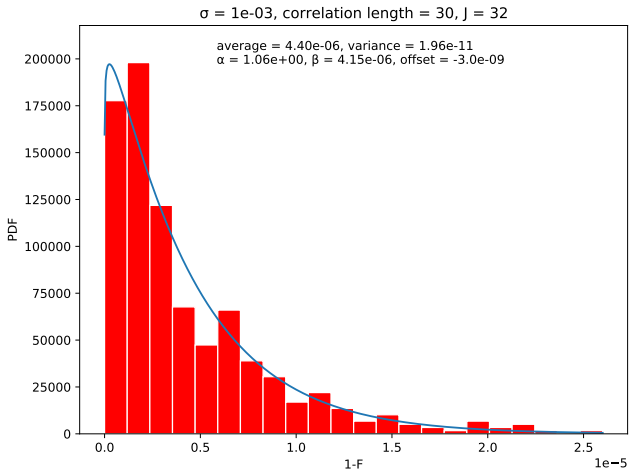

σ = 1e-02, correlation length = 300, J = 128
average = 2.01e-03, variance = 4.98e-06
α = 9.16e-01, β = 2.19e-03, offset = -2.0e-30
PDF
1-F

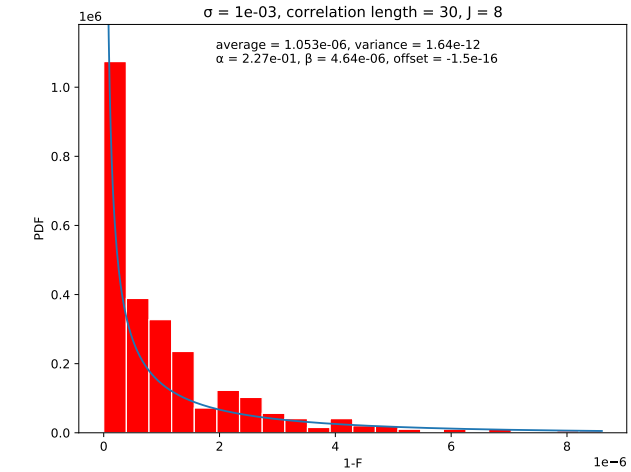

σ = 1e-03, correlation length = 1, J = 32
average = 5.200e-06, variance = 1.80e-12
α = 6.34e+01, β = 1.68e-07, offset = -5.5e-06
PDF
1-F
1e−5

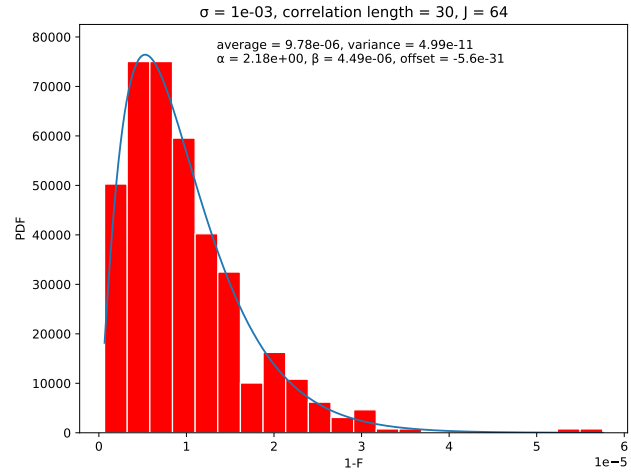

σ = 1e-03, correlation length = 1, J = 8
average = 1.257e-06, variance = 3.75e-13
α = 4.25e+03, β = 9.38e-09, offset = -3.9e-05
PDF
1-F
1e−6

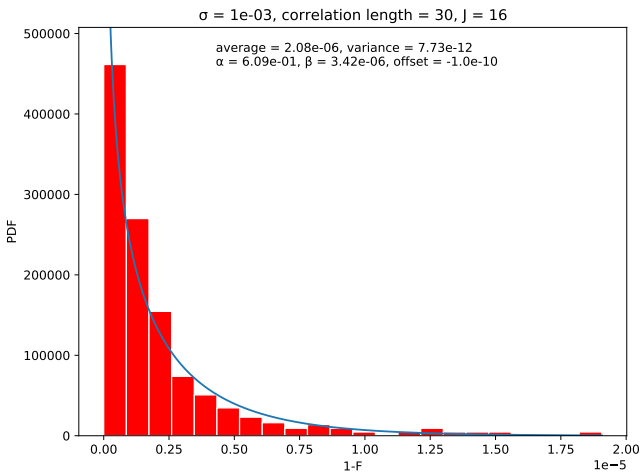

σ = 1e-03, correlation length = 1, J = 64
average = 1.060e-05, variance = 6.01e-12
α = 3.69e+01, β = 4.04e-07, offset = -4.3e-06
PDF
1-F
1e−5

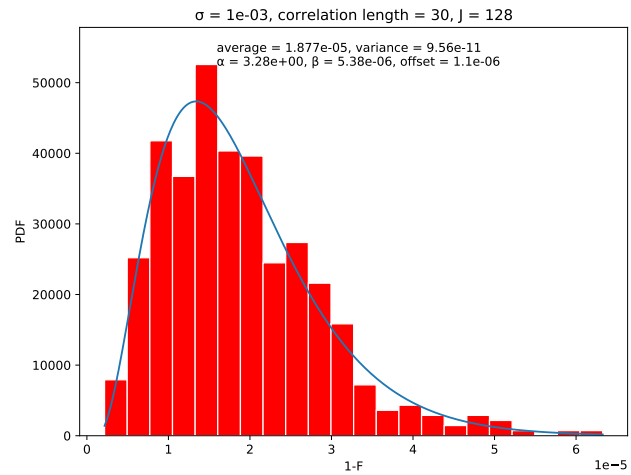

σ = 1e-03, correlation length = 1, J = 128
average = 2.158e-05, variance = 1.59e-11
α = 4.92e+01, β = 5.65e-07, offset = -6.3e-06
PDF
1-F
1e−5

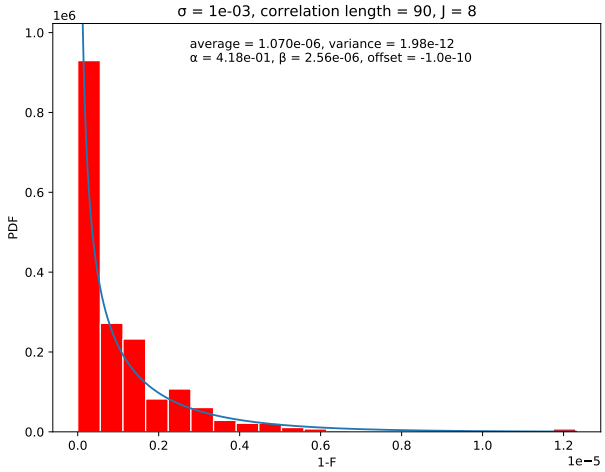

σ = 1e-03, correlation length = 30, J = 32
average = 4.40e-06, variance = 1.96e-11
α = 1.06e+00, β = 4.15e-06, offset = -3.0e-09
PDF
1-F
1e−5

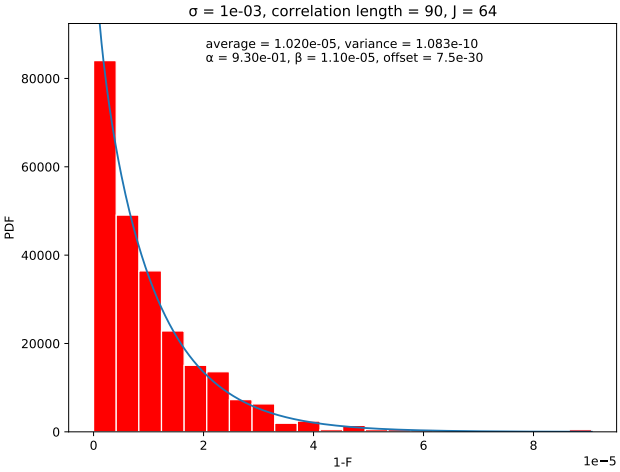

σ = 1e-03, correlation length = 30, J = 8
average = 1.053e-06, variance = 1.64e-12
α = 2.27e-01, β = 4.64e-06, offset = -1.5e-16
1e6
PDF
1-F
1e−6

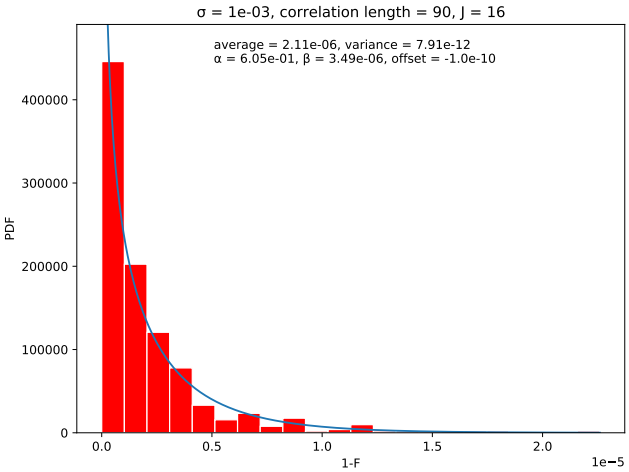

σ = 1e-03, correlation length = 30, J = 64
average = 9.78e-06, variance = 4.99e-11
α = 2.18e+00, β = 4.49e-06, offset = -5.6e-31
PDF
1-F
1e−5

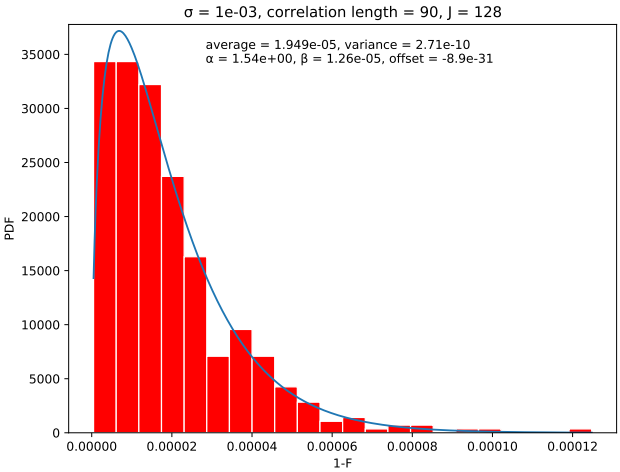

σ = 1e-03, correlation length = 30, J = 16
average = 2.08e-06, variance = 7.73e-12
α = 6.09e-01, β = 3.42e-06, offset = -1.0e-10
PDF
1-F
1e−5

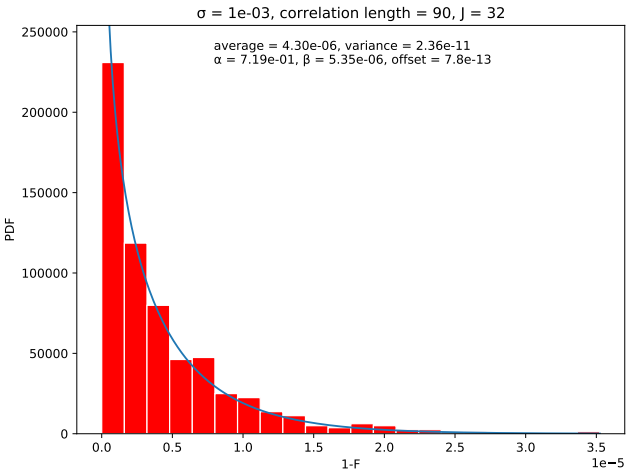

σ = 1e-03, correlation length = 30, J = 128
average = 1.877e-05, variance = 9.56e-11
α = 3.28e+00, β = 5.38e-06, offset = 1.1e-06
PDF
1-F
1e−5

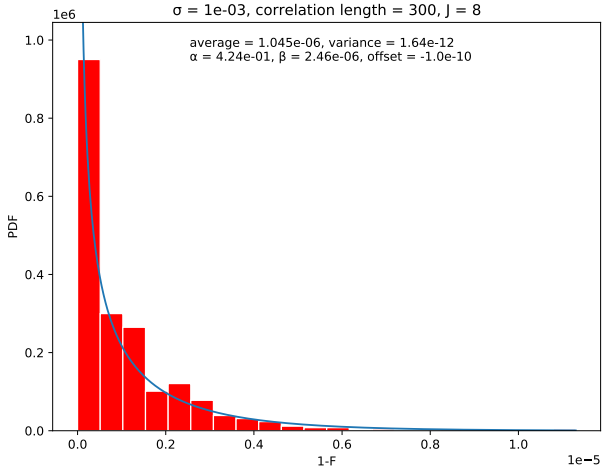

σ = 1e-03, correlation length = 90, J = 8
average = 1.070e-06, variance = 1.98e-12
α = 4.18e-01, β = 2.56e-06, offset = -1.0e-10
1e6
PDF
1-F
1e−5

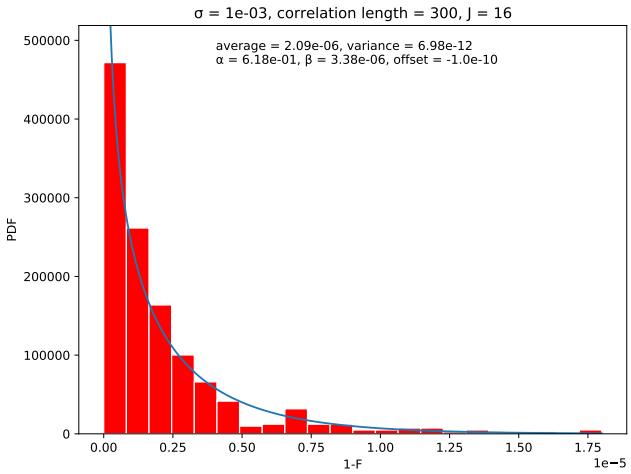

σ = 1e-03, correlation length = 90, J = 64
average = 1.020e-05, variance = 1.083e-10
α = 9.30e-01, β = 1.10e-05, offset = 7.5e-30
PDF
1-F
1e−5

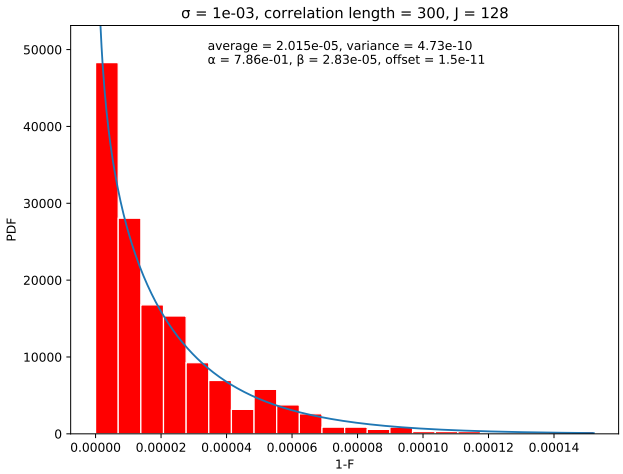

σ = 1e-03, correlation length = 90, J = 16
average = 2.11e-06, variance = 7.91e-12
α = 6.05e-01, β = 3.49e-06, offset = -1.0e-10
PDF
1-F
1e−5

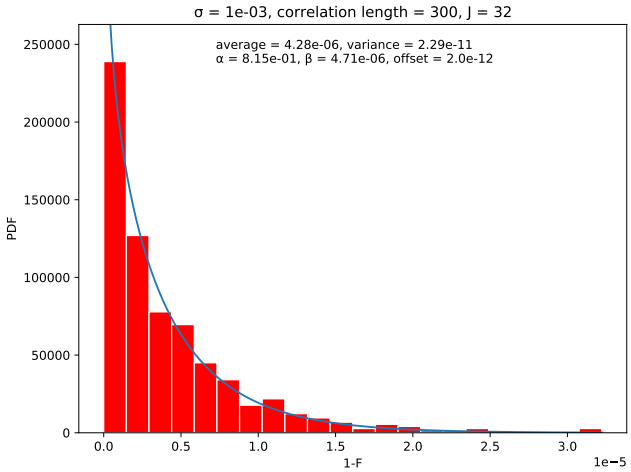

σ = 1e-03, correlation length = 90, J = 128
average = 1.949e-05, variance = 2.71e-10
α = 1.54e+00, β = 1.26e-05, offset = -8.9e-31
PDF
1-F

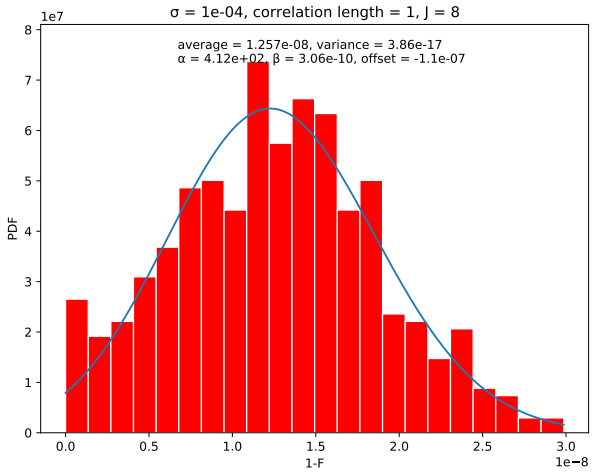

σ = 1e-03, correlation length = 90, J = 32
average = 4.30e-06, variance = 2.36e-11
α = 7.19e-01, β = 5.35e-06, offset = 7.8e-13
PDF
1-F
1e−5

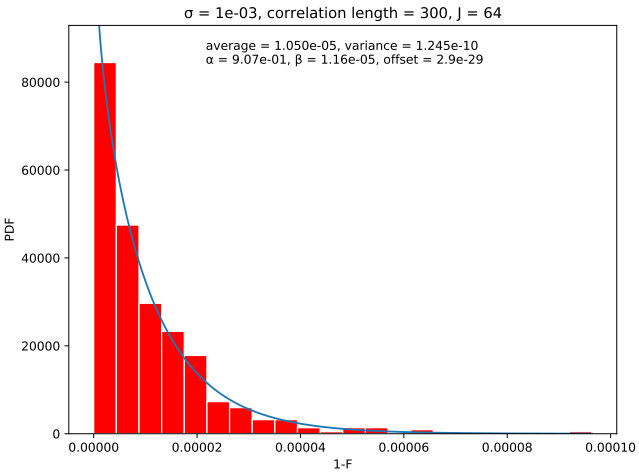

σ = 1e-03, correlation length = 300, J = 8
average = 1.045e-06, variance = 1.64e-12
α = 4.24e-01, β = 2.46e-06, offset = -1.0e-10
1e6
PDF
1-F
1e−5

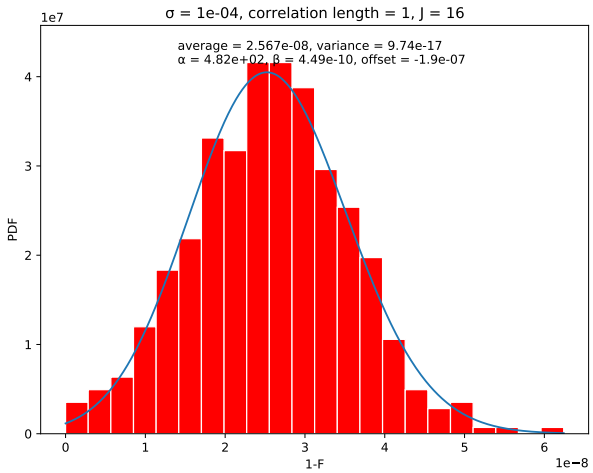

σ = 1e-03, correlation length = 300, J = 16
average = 2.09e-06, variance = 6.98e-12
α = 6.18e-01, β = 3.38e-06, offset = -1.0e-10
PDF
1-F
1e−5

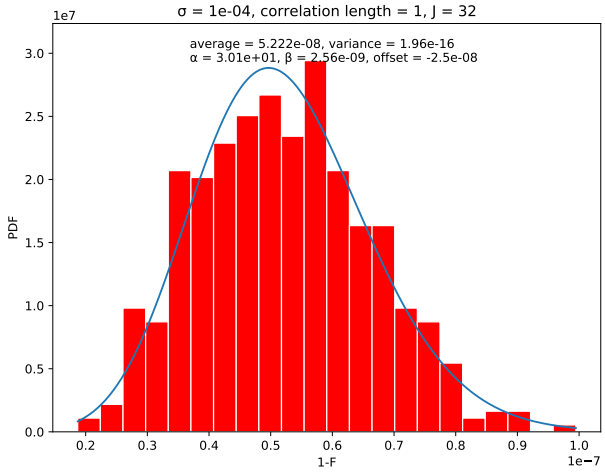

σ = 1e-03, correlation length = 300, J = 128
average = 2.015e-05, variance = 4.73e-10
α = 7.86e-01, β = 2.83e-05, offset = 1.5e-11
PDF
1-F

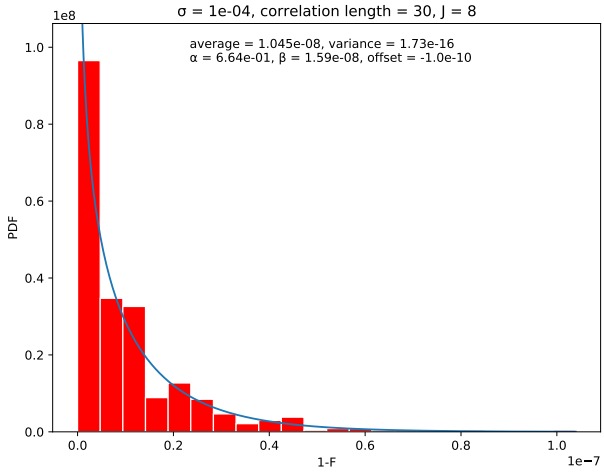

σ = 1e-03, correlation length = 300, J = 32
average = 4.28e-06, variance = 2.29e-11
α = 8.15e-01, β = 4.71e-06, offset = 2.0e-12
PDF
1-F
1e−5

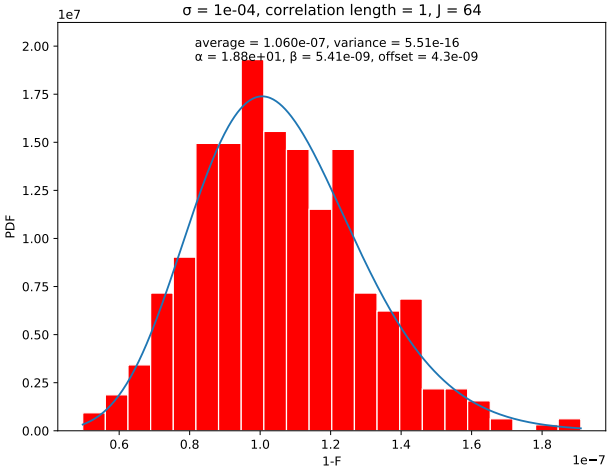

σ = 1e-04, correlation length = 1, J = 8
1e7
average = 1.257e-08, variance = 3.86e-17
α = 4.12e+02, β = 3.06e-10, offset = -1.1e-07
PDF
1-F
1e−8

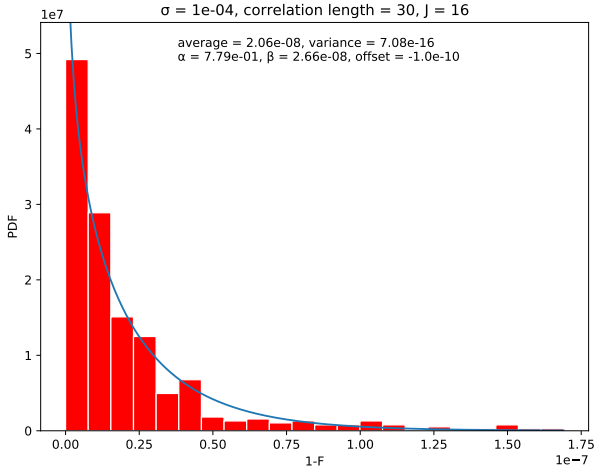

σ = 1e-03, correlation length = 300, J = 64
average = 1.050e-05, variance = 1.245e-10
α = 9.07e-01, β = 1.16e-05, offset = 2.9e-29
PDF
1-F

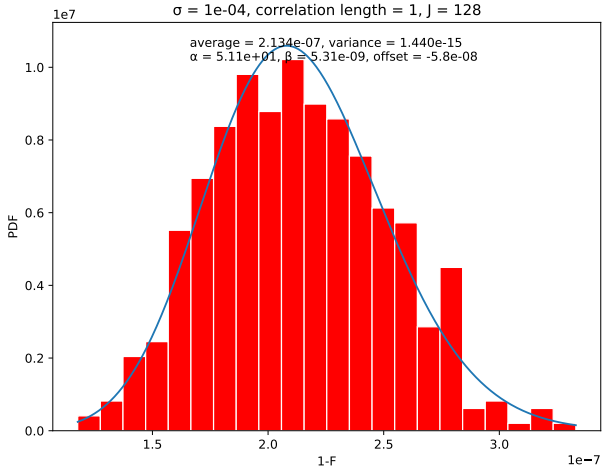

σ = 1e-04, correlation length = 1, J = 16
1e7
average = 2.567e-08, variance = 9.74e-17
α = 4.82e+02, β = 4.49e-10, offset = -1.9e-07
PDF
1-F
1e−8

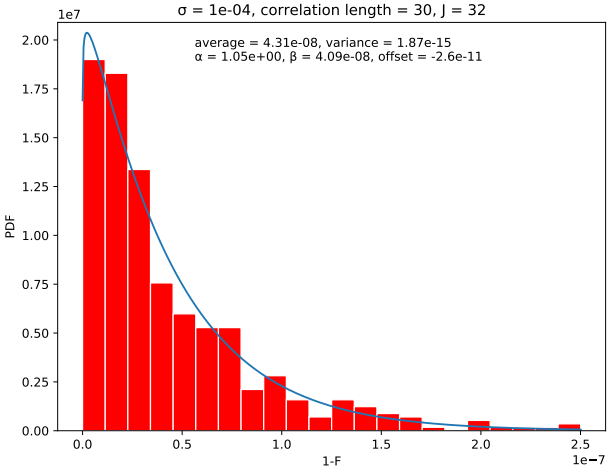

σ = 1e-04, correlation length = 1, J = 32
average = 5.222e-08, variance = 1.96e-16
α = 3.01e+01, β = 2.56e-09, offset = -2.5e-08
1e7
PDF
1-F
1e−7

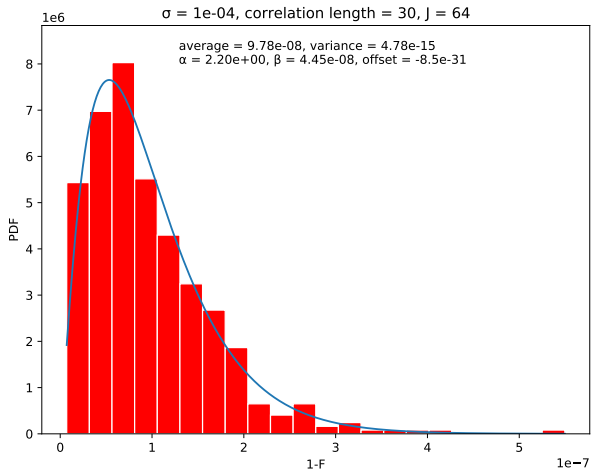

σ = 1e-04, correlation length = 30, J = 8
average = 1.045e-08, variance = 1.73e-16
α = 6.64e-01, β = 1.59e-08, offset = -1.0e-10
1e8
PDF
1-F
1e−7

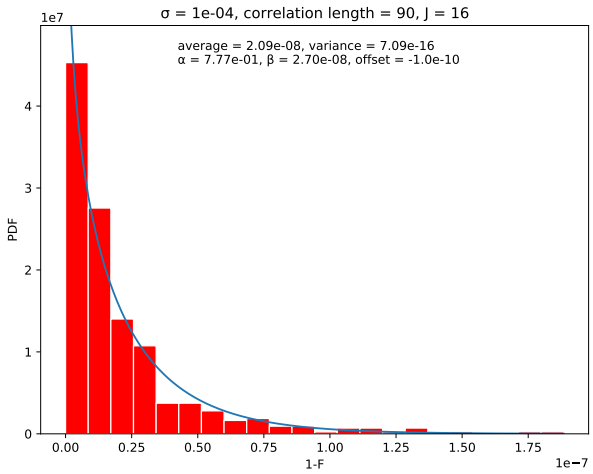

σ = 1e-04, correlation length = 1, J = 64
average = 1.060e-07, variance = 5.51e-16
α = 1.88e+01, β = 5.41e-09, offset = 4.3e-09
1e7
PDF
1-F
1e−7

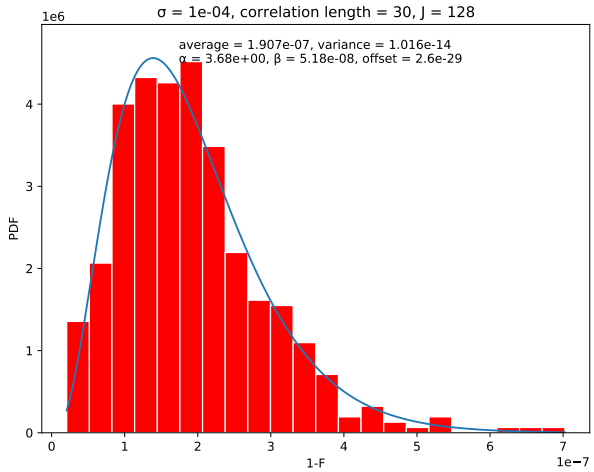

σ = 1e-04, correlation length = 30, J = 16
average = 2.06e-08, variance = 7.08e-16
α = 7.79e-01, β = 2.66e-08, offset = -1.0e-10
1e7
PDF
1-F
1e−7

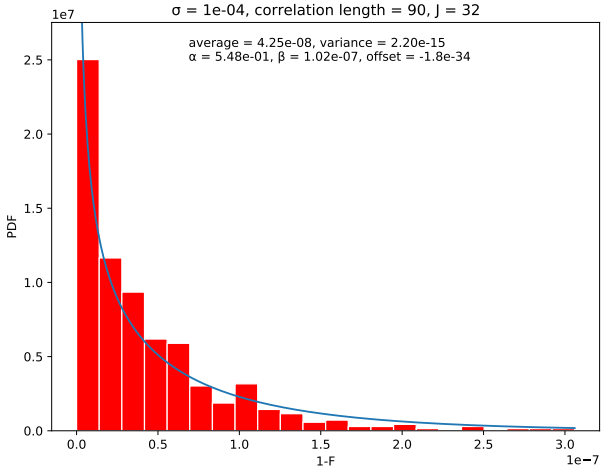

σ = 1e-04, correlation length = 1, J = 128
average = 2.134e-07, variance = 1.440e-15
α = 5.11e+01, β = 5.31e-09, offset = -5.8e-08
1e7
PDF
1-F
1e−7

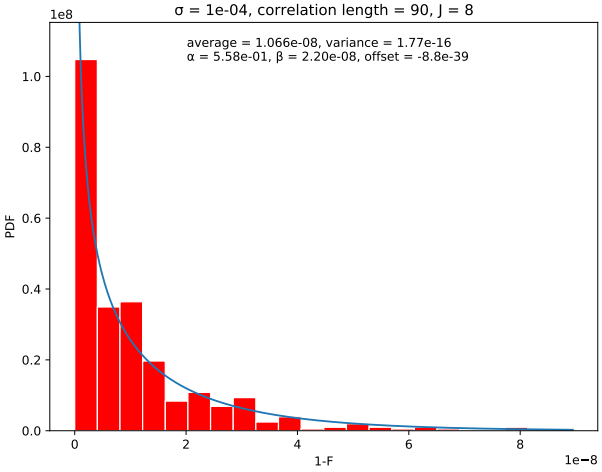

σ = 1e-04, correlation length = 30, J = 32
average = 4.31e-08, variance = 1.87e-15
α = 1.05e+00, β = 4.09e-08, offset = -2.6e-11
1e7
PDF
1-F
1e−7

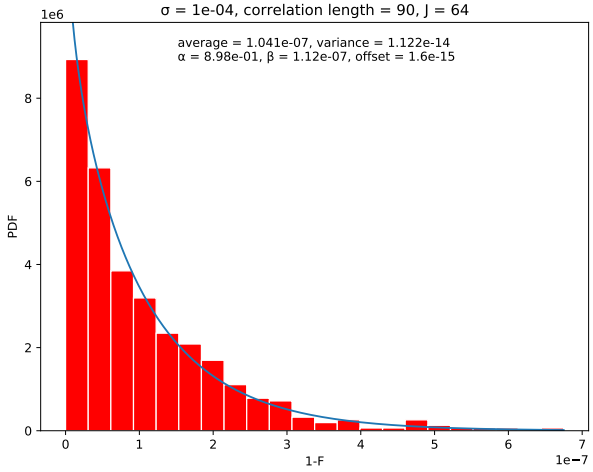

σ = 1e-04, correlation length = 30, J = 64
average = 9.78e-08, variance = 4.78e-15
α = 2.20e+00, β = 4.45e-08, offset = -8.5e-31
1e6
PDF
1-F
1e−7

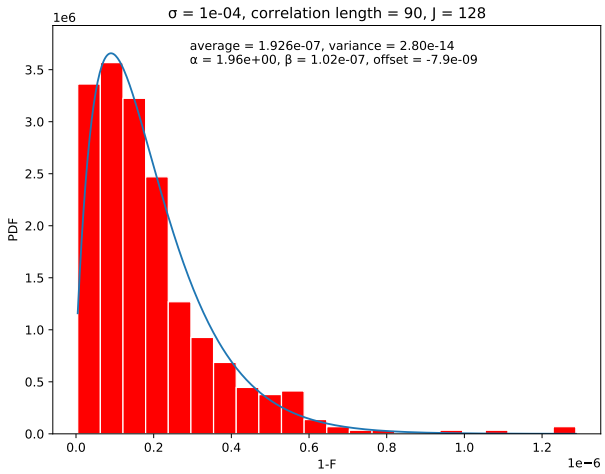

σ = 1e-04, correlation length = 90, J = 16
average = 2.09e-08, variance = 7.09e-16
α = 7.77e-01, β = 2.70e-08, offset = -1.0e-10
1e7
PDF
1-F
1e−7

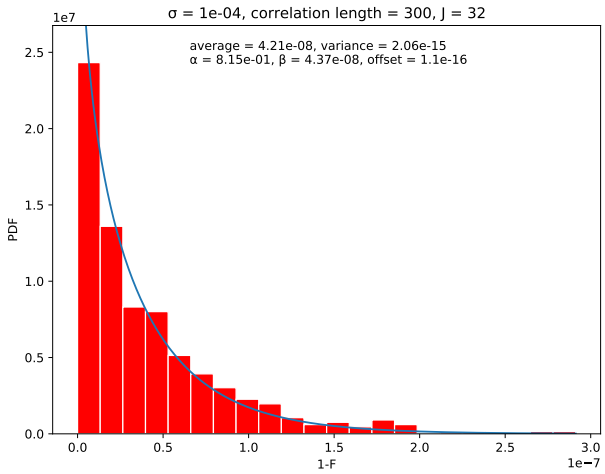

σ = 1e-04, correlation length = 30, J = 128
average = 1.907e-07, variance = 1.016e-14
α = 3.68e+00, β = 5.18e-08, offset = 2.6e-29
1e6
PDF
1-F
1e−7

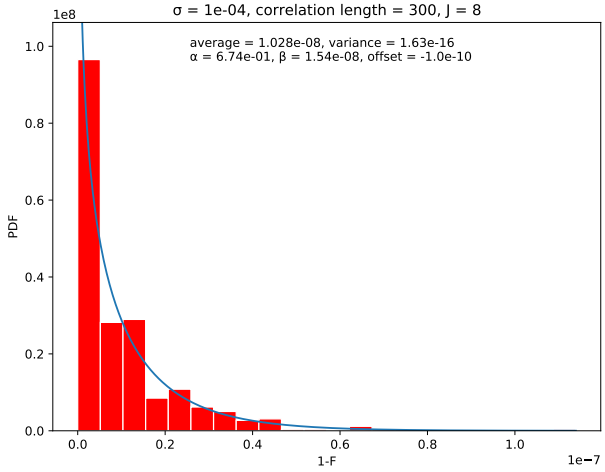

σ = 1e-04, correlation length = 90, J = 32
average = 4.25e-08, variance = 2.20e-15
α = 5.48e-01, β = 1.02e-07, offset = -1.8e-34
1e7
PDF
1-F
1e−7

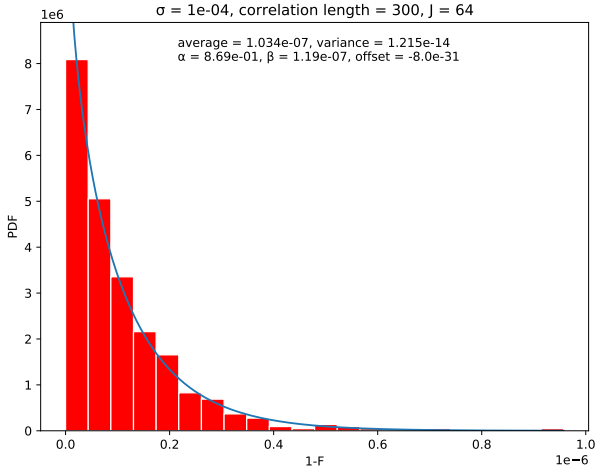

σ = 1e-04, correlation length = 90, J = 8
average = 1.066e-08, variance = 1.77e-16
α = 5.58e-01, β = 2.20e-08, offset = -8.8e-39
1e8
PDF
1-F
1e−8

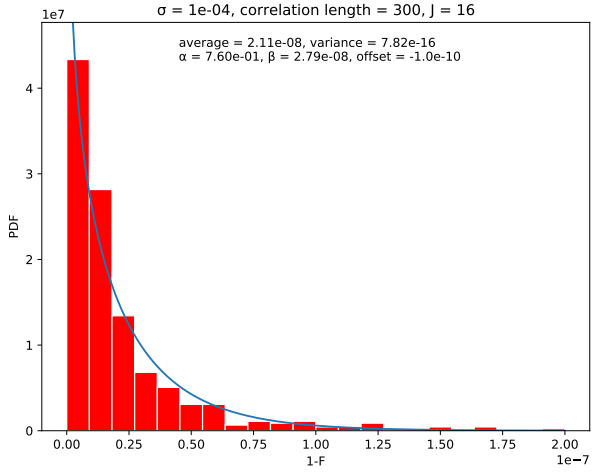

σ = 1e-04, correlation length = 90, J = 64
average = 1.041e-07, variance = 1.122e-14
α = 8.98e-01, β = 1.12e-07, offset = 1.6e-15
1e6
PDF
1-F
1e−7

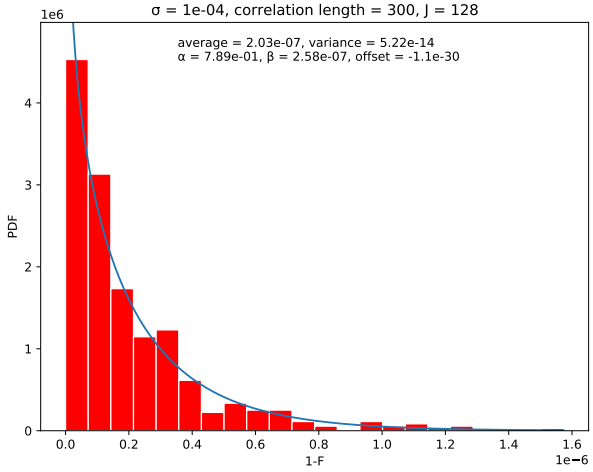

σ = 1e-04, correlation length = 90, J = 128
average = 1.926e-07, variance = 2.80e-14
α = 1.96e+00, β = 1.02e-07, offset = -7.9e-09
1e6
PDF
1-F
1e−6


σ = 1e-04, correlation length = 300, J = 32
average = 4.21e-08, variance = 2.06e-15
α = 8.15e-01, β = 4.37e-08, offset = 1.1e-16
1e7
PDF
1-F
1e−7


σ = 1e-04, correlation length = 300, J = 8
average = 1.028e-08, variance = 1.63e-16
α = 6.74e-01, β = 1.54e-08, offset = -1.0e-10
1e8
PDF
1-F
1e−7


σ = 1e-04, correlation length = 300, J = 64
average = 1.034e-07, variance = 1.215e-14
α = 8.69e-01, β = 1.19e-07, offset = -8.0e-31
1e6
PDF
1-F
1e−6


σ = 1e-04, correlation length = 300, J = 16
average = 2.11e-08, variance = 7.82e-16
α = 7.60e-01, β = 2.79e-08, offset = -1.0e-10
1e7
PDF
1-F
1e−7


σ = 1e-04, correlation length = 300, J = 128
average = 2.03e-07, variance = 5.22e-14
α = 7.89e-01, β = 2.58e-07, offset = -1.1e-30
1e6
PDF
1-F
1e−6

B.1 Summary table

Table 13: This table shows the key parameters for each of the previous 80 plots. cc stands for correlation length and J for sequence length.

sigma	cc	J	average	variance	alpha	beta	offset
1×10^{-1}	1	8	1.237×10^{-2}	3.40×10^{-5}	2.31×10^{3}	1.21×10^{-4}	-2.7×10^{-1}
1×10^{-1}	1	16	2.474×10^{-2}	8.05×10^{-5}	5.41×10^{3}	1.22×10^{-4}	-6.3×10^{-1}
1×10^{-1}	1	32	4.960×10^{-2}	1.62×10^{-4}	2.53×10^{1}	2.53×10^{-3}	-1.4×10^{-2}
1×10^{-1}	1	64	9.602×10^{-2}	3.99×10^{-4}	7.02×10^{1}	2.39×10^{-3}	-7.1×10^{-2}
1×10^{-1}	1	128	1.732×10^{-1}	7.47×10^{-4}	2.27×10^{2}	1.81×10^{-3}	-2.4×10^{-1}
1×10^{-1}	30	8	1.007×10^{-2}	1.455×10^{-4}	1.87×10^{-1}	5.39×10^{-2}	-1.1×10^{-17}
1×10^{-1}	30	16	1.97×10^{-2}	5.82×10^{-4}	5.55×10^{-1}	3.54×10^{-2}	-1.0×10^{-10}
1×10^{-1}	30	32	3.96×10^{-2}	1.252×10^{-3}	1.28×10^{0}	3.08×10^{-2}	8.3×10^{-5}
1×10^{-1}	30	64	8.19×10^{-2}	2.44×10^{-3}	2.30×10^{0}	3.31×10^{-2}	5.9×10^{-3}
1×10^{-1}	30	128	1.468×10^{-1}	3.70×10^{-3}	6.72×10^{0}	2.37×10^{-2}	-1.3×10^{-2}
1×10^{-1}	90	8	1.032×10^{-2}	1.68×10^{-4}	5.31×10^{-1}	1.95×10^{-2}	-1.3×10^{-5}
1×10^{-1}	90	16	1.91×10^{-2}	5.22×10^{-4}	5.64×10^{-1}	3.39×10^{-2}	-1.0×10^{-10}
1×10^{-1}	90	32	3.82×10^{-2}	1.396×10^{-3}	1.01×10^{0}	3.77×10^{-2}	1.1×10^{-4}
1×10^{-1}	90	64	8.01×10^{-2}	4.73×10^{-3}	1.30×10^{0}	6.13×10^{-2}	1.8×10^{-4}
1×10^{-1}	90	128	1.356×10^{-1}	6.63×10^{-3}	2.69×10^{0}	5.12×10^{-2}	-2.0×10^{-3}
1×10^{-1}	300	8	1.033×10^{-2}	1.540×10^{-4}	3.93×10^{-1}	2.63×10^{-2}	-1.2×10^{-7}
1×10^{-1}	300	16	1.93×10^{-2}	5.21×10^{-4}	7.55×10^{-1}	2.56×10^{-2}	-5.2×10^{-31}
1×10^{-1}	300	32	3.76×10^{-2}	1.339×10^{-3}	9.87×10^{-1}	3.81×10^{-2}	2.2×10^{-4}
1×10^{-1}	300	64	8.06×10^{-2}	4.61×10^{-3}	1.25×10^{0}	6.41×10^{-2}	5.6×10^{-4}
1×10^{-1}	300	128	1.261×10^{-1}	9.53×10^{-3}	1.59×10^{0}	7.92×10^{-2}	6.5×10^{-4}
1×10^{-2}	1	8	1.257×10^{-4}	3.75×10^{-9}	3.04×10^{4}	3.51×10^{-7}	-1.1×10^{-2}
1×10^{-2}	1	16	2.587×10^{-4}	9.50×10^{-9}	6.43×10^{2}	3.84×10^{-6}	-2.2×10^{-3}
1×10^{-2}	1	32	5.231×10^{-4}	2.09×10^{-8}	1.38×10^{1}	3.90×10^{-5}	-1.6×10^{-5}
1×10^{-2}	1	64	1.056×10^{-3}	5.56×10^{-8}	2.29×10^{1}	4.92×10^{-5}	-7.2×10^{-5}
1×10^{-2}	1	128	2.126×10^{-3}	1.323×10^{-7}	4.82×10^{1}	5.24×10^{-5}	-4.0×10^{-4}
1×10^{-2}	30	8	1.047×10^{-4}	1.555×10^{-8}	5.38×10^{-1}	1.95×10^{-4}	-1.2×10^{-7}
1×10^{-2}	30	16	2.04×10^{-4}	7.00×10^{-8}	5.51×10^{-1}	3.70×10^{-4}	-1.0×10^{-10}
1×10^{-2}	30	32	4.36×10^{-4}	1.86×10^{-7}	1.07×10^{0}	4.07×10^{-4}	-2.3×10^{-7}
1×10^{-2}	30	64	9.84×10^{-4}	4.88×10^{-7}	2.19×10^{0}	4.50×10^{-4}	-7.9×10^{-31}
1×10^{-2}	30	128	1.865×10^{-3}	9.03×10^{-7}	3.63×10^{0}	5.01×10^{-4}	4.8×10^{-5}
1×10^{-2}	90	8	1.096×10^{-4}	2.12×10^{-8}	5.04×10^{-1}	2.75×10^{-4}	-8.8×10^{-25}
1×10^{-2}	90	16	2.08×10^{-4}	6.95×10^{-8}	5.55×10^{-1}	3.74×10^{-4}	-1.0×10^{-10}
1×10^{-2}	90	32	4.29×10^{-4}	2.33×10^{-7}	7.66×10^{-1}	5.60×10^{-4}	3.0×10^{-30}
1×10^{-2}	90	64	1.029×10^{-3}	1.129×10^{-6}	9.18×10^{-1}	1.12×10^{-3}	3.9×10^{-8}
1×10^{-2}	90	128	1.943×10^{-3}	2.75×10^{-6}	1.58×10^{0}	1.23×10^{-3}	-3.6×10^{-30}
1×10^{-2}	300	8	1.036×10^{-4}	1.539×10^{-8}	6.84×10^{-1}	1.09×10^{-4}	-2.7×10^{-33}
1×10^{-2}	300	16	2.10×10^{-4}	7.40×10^{-8}	5.50×10^{-1}	3.83×10^{-4}	-1.0×10^{-10}
1×10^{-2}	300	32	4.24×10^{-4}	2.18×10^{-7}	7.88×10^{-1}	5.20×10^{-4}	1.8×10^{-8}
1×10^{-2}	300	64	1.020×10^{-3}	1.089×10^{-6}	9.88×10^{-1}	9.99×10^{-4}	6.2×10^{-8}
1×10^{-2}	300	128	2.01×10^{-3}	4.98×10^{-6}	9.16×10^{-1}	2.19×10^{-3}	-2.0×10^{-30}
1×10^{-3}	1	8	1.257×10^{-6}	3.75×10^{-13}	4.25×10^{3}	9.38×10^{-9}	-3.9×10^{-5}
1×10^{-3}	1	16	2.611×10^{-6}	9.56×10^{-13}	3.76×10^{4}	5.03×10^{-9}	-1.9×10^{-4}
1×10^{-3}	1	32	5.200×10^{-6}	1.80×10^{-12}	6.34×10^{1}	1.68×10^{-7}	-5.5×10^{-6}
1×10^{-3}	1	64	1.060×10^{-5}	6.01×10^{-12}	3.69×10^{1}	4.04×10^{-7}	-4.3×10^{-6}
1×10^{-3}	1	128	2.158×10^{-5}	1.59×10^{-11}	4.92×10^{1}	5.65×10^{-7}	-6.3×10^{-6}
1×10^{-3}	30	8	1.053×10^{-6}	1.64×10^{-12}	2.27×10^{-1}	4.64×10^{-6}	-1.5×10^{-16}
1×10^{-3}	30	16	2.08×10^{-6}	7.73×10^{-12}	6.09×10^{-1}	3.42×10^{-6}	-1.0×10^{-10}
1×10^{-3}	30	32	4.40×10^{-6}	1.96×10^{-11}	1.06×10^{0}	4.15×10^{-6}	-3.0×10^{-9}
1×10^{-3}	30	64	9.78×10^{-6}	4.99×10^{-11}	2.18×10^{0}	4.49×10^{-6}	-5.6×10^{-31}

sigma	cc	J	average	variance	alpha	beta	offset
1×10^{-3}	30	128	1.877×10^{-5}	9.56×10^{-11}	3.28×10^{0}	5.38×10^{-6}	1.1×10^{-6}
1×10^{-3}	90	8	1.070×10^{-6}	1.98×10^{-12}	4.18×10^{-1}	2.56×10^{-6}	-1.0×10^{-10}
1×10^{-3}	90	16	2.11×10^{-6}	7.91×10^{-12}	6.05×10^{-1}	3.49×10^{-6}	-1.0×10^{-10}
1×10^{-3}	90	32	4.30×10^{-6}	2.36×10^{-11}	7.19×10^{-1}	5.35×10^{-6}	7.8×10^{-13}
1×10^{-3}	90	64	1.020×10^{-5}	1.083×10^{-10}	9.30×10^{-1}	1.10×10^{-5}	7.5×10^{-30}
1×10^{-3}	90	128	1.949×10^{-5}	2.71×10^{-10}	1.54×10^{0}	1.26×10^{-5}	-8.9×10^{-31}
1×10^{-3}	300	8	1.045×10^{-6}	1.64×10^{-12}	4.24×10^{-1}	2.46×10^{-6}	-1.0×10^{-10}
1×10^{-3}	300	16	2.09×10^{-6}	6.98×10^{-12}	6.18×10^{-1}	3.38×10^{-6}	-1.0×10^{-10}
1×10^{-3}	300	32	4.28×10^{-6}	2.29×10^{-11}	8.15×10^{-1}	4.71×10^{-6}	2.0×10^{-12}
1×10^{-3}	300	64	1.050×10^{-5}	1.245×10^{-10}	9.07×10^{-1}	1.16×10^{-5}	2.9×10^{-29}
1×10^{-3}	300	128	2.015×10^{-5}	4.73×10^{-10}	7.86×10^{-1}	2.83×10^{-5}	1.5×10^{-11}
1×10^{-4}	1	8	1.257×10^{-8}	3.86×10^{-17}	4.12×10^{2}	3.06×10^{-10}	-1.1×10^{-7}
1×10^{-4}	1	16	2.567×10^{-8}	9.74×10^{-17}	4.82×10^{2}	4.49×10^{-10}	-1.9×10^{-7}
1×10^{-4}	1	32	5.222×10^{-8}	1.96×10^{-16}	3.01×10^{1}	2.56×10^{-9}	-2.5×10^{-8}
1×10^{-4}	1	64	1.060×10^{-7}	5.51×10^{-16}	1.88×10^{1}	5.41×10^{-9}	4.3×10^{-9}
1×10^{-4}	1	128	2.134×10^{-7}	1.440×10^{-15}	5.11×10^{1}	5.31×10^{-9}	-5.8×10^{-8}
1×10^{-4}	30	8	1.045×10^{-8}	1.73×10^{-16}	6.64×10^{-1}	1.59×10^{-8}	-1.0×10^{-10}
1×10^{-4}	30	16	2.06×10^{-8}	7.08×10^{-16}	7.79×10^{-1}	2.66×10^{-8}	-1.0×10^{-10}
1×10^{-4}	30	32	4.31×10^{-8}	1.87×10^{-15}	1.05×10^{0}	4.09×10^{-8}	-2.6×10^{-11}
1×10^{-4}	30	64	9.78×10^{-8}	4.78×10^{-15}	2.20×10^{0}	4.45×10^{-8}	-8.5×10^{-31}
1×10^{-4}	30	128	1.907×10^{-7}	1.016×10^{-14}	3.68×10^{0}	5.18×10^{-8}	2.6×10^{-29}
1×10^{-4}	90	8	1.066×10^{-8}	1.77×10^{-16}	5.58×10^{-1}	2.20×10^{-8}	-8.8×10^{-39}
1×10^{-4}	90	16	2.09×10^{-8}	7.09×10^{-16}	7.77×10^{-1}	2.70×10^{-8}	-1.0×10^{-10}
1×10^{-4}	90	32	4.25×10^{-8}	2.20×10^{-15}	5.48×10^{-1}	1.02×10^{-7}	-1.8×10^{-34}
1×10^{-4}	90	64	1.041×10^{-7}	1.122×10^{-14}	8.98×10^{-1}	1.12×10^{-7}	1.6×10^{-15}
1×10^{-4}	90	128	1.926×10^{-7}	2.80×10^{-14}	1.96×10^{0}	1.02×10^{-7}	-7.9×10^{-9}
1×10^{-4}	300	8	1.028×10^{-8}	1.63×10^{-16}	6.74×10^{-1}	1.54×10^{-8}	-1.0×10^{-10}
1×10^{-4}	300	16	2.11×10^{-8}	7.82×10^{-16}	7.60×10^{-1}	2.79×10^{-8}	-1.0×10^{-10}
1×10^{-4}	300	32	4.21×10^{-8}	2.06×10^{-15}	8.15×10^{-1}	4.37×10^{-8}	1.1×10^{-16}
1×10^{-4}	300	64	1.034×10^{-7}	1.215×10^{-14}	8.69×10^{-1}	1.19×10^{-7}	-8.0×10^{-31}
1×10^{-4}	300	128	2.03×10^{-7}	5.22×10^{-14}	7.89×10^{-1}	2.58×10^{-7}	-1.1×10^{-30}

C Error of the variance estimator

K occurrences x_i (i from 1 to K) of the same Gaussian random variable x with average and variance $\mathbb{E}[x]$ and $\mathbb{V}[x]$ shall be considered. Every occurrence is independent of the other, i.e. all the x_i are i.i.d. Gaussian random variables. The variance estimator is

$$\tilde{\mathbb{V}}[x] = \frac{1}{K-1}\sum_{i=0}^{K}\left(\tilde{\mathbb{E}}[x] - x_i\right)^2. \tag{65}$$

with $\tilde{\mathbb{E}}[x]$ the estimated expected value of x. It is quickly understood that

$$\tilde{\mathbb{V}}[x] = \frac{\mathbb{V}[x]}{K-1}\chi^2_{K-1} \tag{66}$$

with χ^2_{K-1} the chi squared random variable with $K-1$ degrees of freedom. From this follows that

$$\mathbb{E}\left[\tilde{\mathbb{V}}[x]\right] = \frac{\mathbb{V}[x]}{K-1}\mathbb{E}\left[\chi^2_{K-1}\right] = \mathbb{V}[x], \tag{67}$$

which means that the variance estimator converges to the distribution variance. Moreover,

$$\mathbb{V}\left[\tilde{\mathbb{V}}[x]\right] = \left(\frac{\mathbb{V}[x]}{K-1}\right)^2\mathbb{V}\left[\chi^2_{K-1}\right] = \frac{2}{K-1}\mathbb{V}[x]^2. \tag{68}$$

All in all, the standard deviation of the variance estimator is $\sqrt{2/(K-1)}\mathbb{V}[x]$ which can be estimated by $\sqrt{2/(K-1)}\tilde{\mathbb{V}}[x]$.

D Extension of noise correlation analysis to DRB protocols

Ivan Rojkov, who must be thanked for his contribution, extended some of the results for Clifford Randomized Benchmarking reported in Ball et al. [3] to the Direct Randomized Benchmarking formalism.

One must first mention that Ball et al. [3] work primarily analyses the effects of temporal noise correlations on single qubit Clifford Randomized Benchmarking experiments and outlines some results for more qubits, whereas the DRB pioneer work of Proctor et al. 2019 [34] covers mainly benchmarking of multiqubit devices. However, the general approach of the former generalises straightforwardly to any CRB protocol and can also be, to some extent, applied to the new DRB paradigm.

In this protocol, presented in Section 1.2.1, one considers a set of native gates that has the only requirement to generate the Clifford group C_n. Even though this degree of freedom exists, the authors mention that a natural choice for such a set is the subset $C_{DRB} \subset C_n$ of depth-one circuits, e.g. parallel one- and two-qubit gates [46]. This additional assumption implies that DRB sequences $U_0\, U_{J_j}\, C_0$ are, as in the CRB, composed exclusively of Clifford gates. One can thus replicate the analysis performed in Section C of [3] and conclude that coherent-noise-affected sequences $\tilde{U}_0 \tilde{U}_{J_j} \tilde{C}_0$ result in a random walk in the XYZ Pauli space. The main difference with the standard CRB is that for short benchmarking sequences the walks will be restricted to only certain regions of the Pauli space. It is believed that this regions would depend on the size of the native gate set C_{DRB} and the probability density function that is used to sample the layers U_{J_j}. Large sequences would however be boundless, since C_{DRB} is supposed to generate C_n, i.e. ultimately any Clifford gate could be randomly generated in the sampling process.

However, another major disparity exist between the CRB and the DRB: the expression of the trace fidelity that captures the overlap between an ideal and a noisy sequence. Indeed, in the CRB protocol it is written for a single sequence as:

$$F_{k,\eta}^{CRB} = \left| \frac{1}{2} \operatorname{Tr}\left(S_k^\dagger \tilde{S}_{k,\eta} \right) \right|^2 = \frac{1}{4} \left| \operatorname{Tr}\left(\tilde{S}_{k,\eta} \right) \right|^2 \tag{69}$$

where the second equality follows from $S_k \equiv \mathbb{1}$. The trace of the noisy sequence is then what has effectively the random walk behaviour in the Pauli space. For the DRB, the overlap fidelity takes the following form

$$\begin{aligned} F_{J_j}^{DRB} &= \left| \frac{1}{2} \operatorname{Tr}\left((U_0\, U_{J_j}\, C_0)^\dagger \tilde{U}_0 \tilde{U}_{J_j} \tilde{C}_0 \right) \right|^2 = \frac{1}{4} \left| \operatorname{Tr}\left(C_0^\dagger\, U^\dagger\, U_0^\dagger U_0 \tilde{U}_{J_j} C_0 \right) \right|^2 = \\ &= \frac{1}{4} \left| \operatorname{Tr}\left(U_{J_j}^\dagger \tilde{U}_{J_j} \right) \right|^2 \leq \frac{1}{4} \left| \operatorname{Tr}\left(U_{J_j}^\dagger \right) \right|^2 \left| \operatorname{Tr}\left(\tilde{U}_{J_j} \right) \right|^2 . \end{aligned} \tag{70}$$

Here, for the sake of simplicity, it was assumed that the intial random Clifford operation C_0 and the last unitary U_0 are noiseless[47]. Then the third equality follows from the cyclicity of the trace and the unitarity of the operations. Finally, the inequality is obtained from Cauchy-Schwarz. The quantity which would in this situation follow a random walk in the Pauli space is the trace of the noisy layers $\tilde{U}_{J_j}$.

The empirical fidelities given in Eqs. (25) and (28) are in fact estimators of the overlap fidelity given above. However, only in the CRB one can truly express their behaviour as random walks in the XYZ Pauli space. Indeed, in the DRB, it only helps us to determine an upper bound to the overlap fidelity which also has the additional difficulty of evaluating the trace of $U_{J_j}^\dagger$, since no way has been found yet to connect Eq. (70) to Eq. (28).

This brief investigation of the effect of correlated noise for DRB shows that this protocol despite being interesting in terms of scalability adds some subtlety in the analysis of the output. It would require further investigation to draw stronger conclusion on the relationship between the theoretical channel distances and the empirical data.

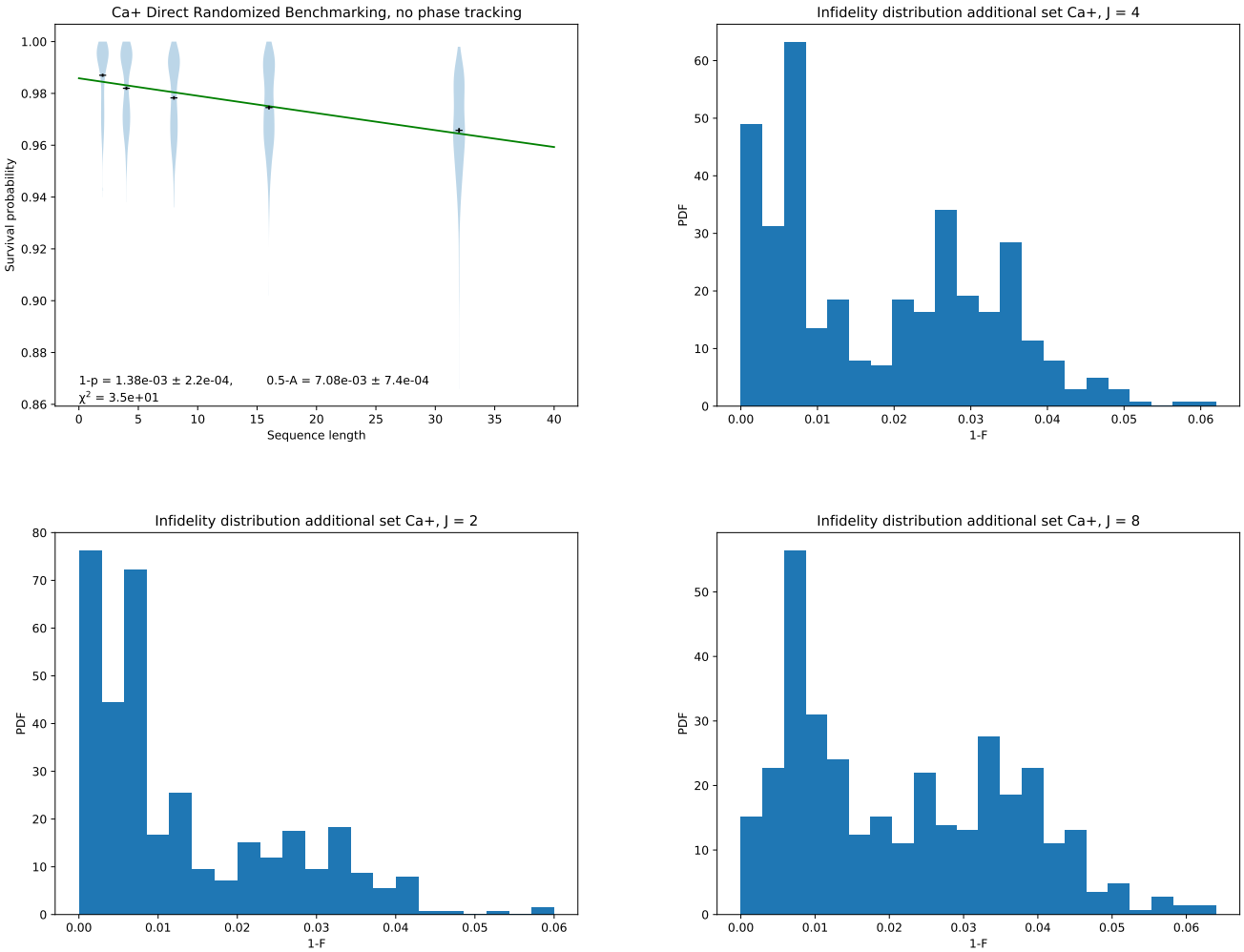
[46] This could be as well extended to higher order qubit gates.
[47] For noisy C_0 and U_0, a similar conclusion can be drawn if they are both elements of the Clifford group.

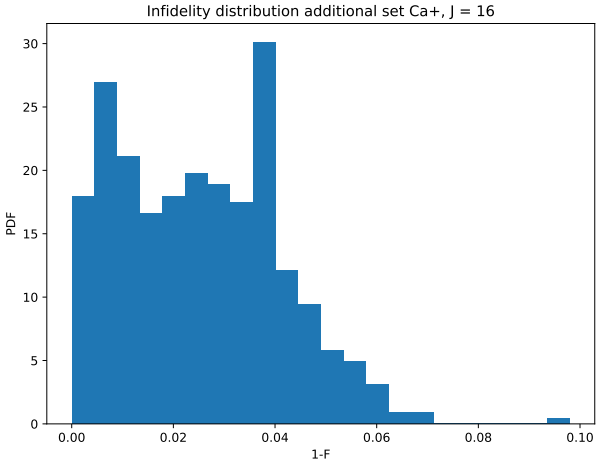
E Sums of gamma distributions in DRB experimental data

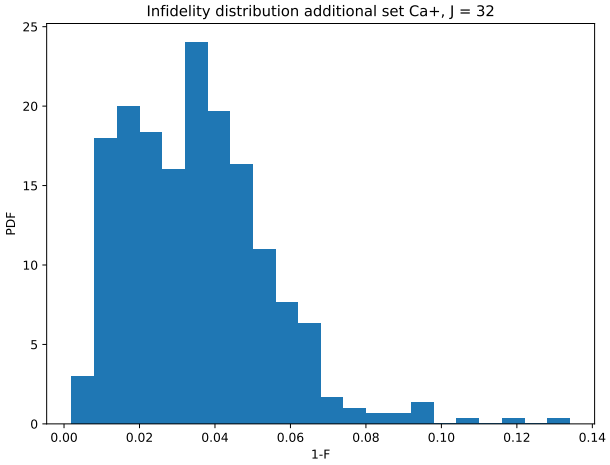
A DRB data set with the Calcium ion had already been sampled before the RBF was developed. The set of lengths was the same of the other DRB experiments in the main text and $K = 500$, but for $J_j = 2$ only 441 circuits were sampled due to experimental problems. For this experiment the gate set did not include Z gates and the Stark shift phase was not taken care of. These differences make this data set not suitable for comparison with the other data. Nonetheless these data are interesting because $\mathrm{PDF}(1 - F(J_j))$ cannot be fit by a single Γ distribution. With intuition one case see two Γ distributions in the following histograms. The speculation is that not considering the Stark shift phase led to additional coherent errors which led to an additional Γ distribution. It was attempted to produce a fit for them, but the double amount of parameters has not allowed this yet. Truth be told, this data set was neglected for moths because related to worse experimental procedures than what was presented in the main text and it was "rediscovered" only some days before the official end of this project and not much time was left to try and fit sums of Γ distributions.

$1 - p$	$0.5 - A$	χ^2
$1.38 \times 10^{-3} \pm 2.2 \times 10^{-4}$	$7.08 \times 10^{-3} \pm 7.4 \times 10^{-4}$	35

Table 14: When compared to Table 9 and Table 5, these fitted parameters are worse.




Infidelity distribution additional set Ca+, J = 16
PDF
1-F
0
5
10
15
20
25
30
0.00
0.02
0.04
0.06
0.08
0.10


Infidelity distribution additional set Ca+, J = 32
PDF
1-F
0
5
10
15
20
25
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14

References

[1] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. A*, 70:052328, Nov 2004. doi: 10.1103/PhysRevA.70.052328. URL https://link.aps.org/doi/10.1103/PhysRevA.70.052328.

[2] Gadi Aleksandrowicz et al. Qiskit: An open-source framework for quantum computing, 2019. URL https://zenodo.org/record/2562111. Last accessed on August 2021.

[3] Harrison Ball, Thomas M. Stace, Steven T. Flammia, and Michael J. Biercuk. Effect of noise correlations on randomized benchmarking. *Phys. Rev. A*, 93:022303, Feb 2016. doi: 10.1103/PhysRevA.93.022303. URL https://link.aps.org/doi/10.1103/PhysRevA.93.022303.

[4] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Kenneth Rudinger, Jonathan Mizrahi, Kevin Fortier, and Peter Maunz. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nature Communications*, 8(1), February 2017. doi: 10.1038/ncomms14485. URL https://doi.org/10.1038/ncomms14485.

[5] K. R. Brown, A. C. Wilson, Y. Colombe, C. Ospelkaus, A. M. Meier, E. Knill, D. Leibfried, and D. J. Wineland. Single-qubit-gate error below $\mathbf{10^{-4}}$ in a trapped ion. *Phys. Rev. A*, 84:030303, Sep 2011. doi: 10.1103/PhysRevA.84.030303. URL https://link.aps.org/doi/10.1103/PhysRevA.84.030303.

[6] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80:012304, Jul 2009. doi: 10.1103/PhysRevA.80.012304. URL https://link.aps.org/doi/10.1103/PhysRevA.80.012304.

[7] Juan P Dehollain, Juha T Muhonen, Robin Blume-Kohout, Kenneth M Rudinger, John King Gamble, Erik Nielsen, Arne Laucht, Stephanie Simmons, Rachpon Kalra, Andrew S Dzurak, and Andrea Morello. Optimization of a solid-state electron spin qubit using gate set tomography. *New Journal of Physics*, 18(10): 103018, oct 2016. doi: 10.1088/1367-2630/18/10/103018. URL https://doi.org/10.1088/1367-2630/18/10/103018.

[8] SciPy developers. Scipy.org, 2009. URL https://docs.scipy.org/doc/. Last accessed on August 2021.

[9] C. L. Edmunds, C. Hempel, R. J. Harris, V. Frey, T. M. Stace, and M. J. Biercuk. Dynamically corrected gates suppressing spatiotemporal error correlations as measured by randomized benchmarking. *Phys. Rev. Research*, 2:013156, Feb 2020. doi: 10.1103/PhysRevResearch.2.013156. URL https://link.aps.org/doi/10.1103/PhysRevResearch.2.013156.

[10] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347–S352, sep 2005. doi: 10.1088/1464-4266/7/10/021. URL https://doi.org/10.1088/1464-4266/7/10/021.

[11] Jeffrey M. Epstein, Andrew W. Cross, Easwar Magesan, and Jay M. Gambetta. Investigating the limits of randomized benchmarking protocols. *Phys. Rev. A*, 89:062321, Jun 2014. doi: 10.1103/PhysRevA.89.062321. URL https://link.aps.org/doi/10.1103/PhysRevA.89.062321.

[12] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland. Randomized benchmarking of multiqubit gates. *Phys. Rev. Lett.*, 108: 260503, Jun 2012. doi: 10.1103/PhysRevLett.108.260503. URL https://link.aps.org/doi/10.1103/PhysRevLett.108.260503.

[13] D. Gottesman. The heisenberg representation of quantum computers, 7 1998. URL https://arxiv.org/abs/quant-ph/9807006.

[14] Christopher Granade, Christopher Ferrie, and D G Cory. Accelerated randomized benchmarking. *New Journal of Physics*, 17(1):013042, jan 2015. doi: 10.1088/1367-2630/17/1/013042. URL https://doi.org/10.1088/1367-2630/17/1/013042.

[15] D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: On the structure of unitary designs. *Journal of Mathematical Physics*, 48(5):052104, 2007. doi: 10.1063/1.2716992. URL https://doi.org/10.1063/1.2716992.

[16] Jonas Helsen, Joel J. Wallman, Steven T. Flammia, and Stephanie Wehner. Multiqubit randomized benchmarking using few samples. *Phys. Rev. A*, 100:032304, Sep 2019. doi: 10.1103/PhysRevA.100.032304. URL https://link.aps.org/doi/10.1103/PhysRevA.100.032304.

[17] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, Jan 2008. doi: 10.1103/PhysRevA.77.012307. URL https://link.aps.org/doi/10.1103/PhysRevA.77.012307.

[18] Emanuel Knill. Quantum computing. *Nature*, 463(7280):441–443, January 2010. doi: 10.1038/463441a. URL https://doi.org/10.1038/463441a.

[19] Richard Kueng, David M. Long, Andrew C. Doherty, and Steven T. Flammia. Comparing experiments to the fault-tolerance threshold. *Phys. Rev. Lett.*, 117:170502, Oct 2016. doi: 10.1103/PhysRevLett.117.170502. URL https://link.aps.org/doi/10.1103/PhysRevLett.117.170502.

[20] THL A29 Limited. RapidJSON, 2021. URL https://github.com/Tencent/rapidjson. Last accessed on August 2021.

[21] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, Apr 2012. doi: 10.1103/PhysRevA.85.042311. URL https://link.aps.org/doi/10.1103/PhysRevA.85.042311.

[22] Matteo Marinelli. *Quantum information processing with mixed-species ion crystals*. PhD thesis, ETH Zurich, 2020. URL P:/tiqi/Literature/TIQI%20PhD%20Theses/Thesis_Marinelli.pdf.

[23] S. Mavadia, C. L. Edmunds, C. Hempel, H. Ball, F. Roy, T. M. Stace, and M. J. Biercuk. Experimental quantum verification in the presence of temporally correlated noise. *npj Quantum Information*, 4(1), February 2018. doi: 10.1038/s41534-017-0052-0. URL https://doi.org/10.1038/s41534-017-0052-0.

[24] Yoshifumi Nakata, Christoph Hirche, Ciara Morgan, and Andreas Winter. Unitary 2-designs from random x- and z-diagonal unitaries. *Journal of Mathematical Physics*, 58(5):052203, 2017. doi: 10.1063/1.4983266. URL https://doi.org/10.1063/1.4983266.

[25] Gabriele Nebe, E. M. Rains, and N. J. A. Sloane. The invariants of the clifford groups. *Designs, Codes and Cryptography*, 24(1):99–122, 2001. doi: 10.1023/a:1011233615437. URL https://doi.org/10.1023/a:1011233615437.

[26] Vlad Negnevitsky. *Feedback-stabilised quantum states in a mixed-species ion system*. PhD thesis, ETH Zurich, 2018. URL P:/tiqi/Literature/TIQI%20PhD%20Theses/Thesis_VNegnevitsky_20180928_final.pdf.

[27] Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout. Gate set tomography, 2020.

[28] Erik Nielsen, Kenneth Rudinger, John King Gamble, and Robin Blume-Kohout. pyGSTi: Report Generation Tutorial, 2020. URL https://github.com/pyGSTio/pyGSTi/blob/master/jupyter_notebooks/Tutorials/reporting/ReportGeneration.ipynb. Last accessed on August 2021.

[29] Erik Nielsen, Kenneth Rudinger, Timothy Proctor, Antonio Russo, Kevin Young, and Robin Blume-Kohout. Probing quantum processor performance with pyGSTi. *Quantum Science and Technology*, 5(4): 044002, jul 2020. doi: 10.1088/2058-9565/ab8aa4. URL https://doi.org/10.1088/2058-9565/ab8aa4.

[30] Michael A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249–252, 2002. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(02)01272-0. URL https://www.sciencedirect.com/science/article/pii/S0375960102012720.

[31] John Preskill. Reliable quantum computers. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):385–410, January 1998. doi: 10.1098/rspa.1998.0167. URL https://doi.org/10.1098/rspa.1998.0167.

[32] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-Kohout. What randomized benchmarking actually measures. *Phys. Rev. Lett.*, 119:130502, Sep 2017. doi: 10.1103/PhysRevLett.119.130502. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.130502.

[33] Timothy Proctor, Melissa Revelle, Erik Nielsen, Kenneth Rudinger, Daniel Lobser, Peter Maunz, Robin Blume-Kohout, and Kevin Young. Detecting and tracking drift in quantum information processors. *Nature Communications*, 11(1), October 2020. doi: 10.1038/s41467-020-19074-4. URL https://doi.org/10.1038/s41467-020-19074-4.

[34] Timothy J. Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. Direct randomized benchmarking for multiqubit devices. *Phys. Rev. Lett.*, 123:030503, Jul 2019. doi: 10.1103/PhysRevLett.123.030503. URL https://link.aps.org/doi/10.1103/PhysRevLett.123.030503.

[35] Sharpened Productions. .JSON File Extension, 2021. URL https://fileinfo.com/extension/json. Last accessed on August 2021.

[36] Alfredo Ricci Vásquez. Randomised benchmarking of ion-trap digital quantum simulators. Master's thesis, University of Edinburgh, 2021.

[37] Alfredo Ricci Vásquez and Alessio Ruggeri. Ionizer addons: randomized benchmarking, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/ionizer-addons/randomized-benchmarking.

[38] Alessio Ruggeri. Drb simulation data, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/ionizer-addons/randomized-benchmarking/-/tree/alessio_branch/jupyter%20notebooks/results_1q_v2.

[39] Alessio Ruggeri. Gate classes repo, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/segtrap/ionpulse_sw/-/tree/tb/tomography/src/gates.

[40] Alessio Ruggeri and Tanja Behrle. Beryllium DRB data repo, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/ionizer-addons/randomized-benchmarking/-/blob/alessio_branch/Results/be_sequencer.txt.

[41] Alessio Ruggeri and Tanja Behrle. Beryllium GST data repo, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/ionizer-addons/gate-set-tomograpahy/-/tree/master/segtrap_be.

[42] Alessio Ruggeri and Tanja Behrle. Calcium DRB data repo, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/ionizer-addons/randomized-benchmarking/-/blob/alessio_branch/Results/ca_sequencer.txt.

[43] Alessio Ruggeri and Tanja Behrle. Calcium GST data repo, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/ionizer-addons/gate-set-tomograpahy/-/tree/master/gst_segtrap_ca.

[44] Alessio Ruggeri and Alfredo Ricci Vásquez. gate_sequencer repo, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/ionpulse_sdk_core/-/commit/818434a7b23e43f1833238201cbb4d2f8b889438.

[45] Alessio Ruggeri and Alfredo Ricci Vásquez. Randomized benchmarking framework, 2021. URL https://gitlab.phys.ethz.ch/tiqi-projects/rb-framework.

[46] Yuval R Sanders, Joel J Wallman, and Barry C Sanders. Bounding quantum gate error rate based on reported average fidelity. *New Journal of Physics*, 18(1):012002, dec 2015. doi: 10.1088/1367-2630/18/1/012002. URL https://doi.org/10.1088/1367-2630/18/1/012002.

[47] Sarah Sheldon, Lev S. Bishop, Easwar Magesan, Stefan Filipp, Jerry M. Chow, and Jay M. Gambetta. Characterizing errors on qubit operations via iterative randomized benchmarking. *Phys. Rev. A*, 93:012301, Jan 2016. doi: 10.1103/PhysRevA.93.012301. URL https://link.aps.org/doi/10.1103/PhysRevA.93.012301.

[48] J Tolar. On clifford groups in quantum computing. *Journal of Physics: Conference Series*, 1071:012022, aug 2018. doi: 10.1088/1742-6596/1071/1/012022. URL https://doi.org/10.1088/1742-6596/1071/1/012022.

[49] Joel J. Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, January 2018. ISSN 2521-327X. doi: 10.22331/q-2018-01-29-47. URL https://doi.org/10.22331/q-2018-01-29-47.

[50] Joel J Wallman and Steven T Flammia. Randomized benchmarking with confidence. *New Journal of Physics*, 16(10):103032, oct 2014. doi: 10.1088/1367-2630/16/10/103032. URL https://doi.org/10.1088/1367-2630/16/10/103032.

[51] Eric W. Weisstein. Frobenius norm, a wolfram web resource, 2021. URL https://mathworld.wolfram.com/FrobeniusNorm.html. Last accessed on August 2021.

[52] T. Xia, M. Lichtman, K. Maller, A. W. Carr, M. J. Piotrowicz, L. Isenhower, and M. Saffman. Randomized benchmarking of single-qubit gates in a 2d array of neutral-atom qubits. *Phys. Rev. Lett.*, 114:100503, Mar 2015. doi: 10.1103/PhysRevLett.114.100503. URL https://link.aps.org/doi/10.1103/PhysRevLett.114.100503.