Koterle et al. *EPJ Quantum Technology* (2024) 11:11
https://doi.org/10.1140/epjqt/s40507-024-00221-5

EPJ Quantum Technology
a SpringerOpen Journal

EPJ.org

RESEARCH Open Access

Mbit/s-range alkali vapour spin noise quantum random number generators

Matija Koterle[1,2*], Samo Beguš[3], Jure Pirman[1,2], Tadej Mežnaršič[1,2], Katja Gosar[1,2], Erik Zupanič[1], Rok Žitko[1,2*] and Peter Jeglič[1,2*]

*Correspondence:
matija@koterle.com; rok.zitko@ijs.si; peter.jeglic@ijs.si
[1]Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia
[2]Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia
Full list of author information is available at the end of the article

Abstract

Spin noise based quantum random number generators first appeared in 2008 and have since then garnered little further interest, in part because their bit rate is limited by the transverse relaxation time T_2 which for coated alkali vapour cells is typically in the kbit/s range. Here we present two advances. The first is an improved bit generation protocol that allows generating bits at rates exceeding $1/T_2$ with only a minor increase of serial correlations. The second is a significant reduction of the time T_2 itself by removing the coating, increasing the vapour temperature and introducing a magnetic-field gradient. In this way we managed to increase the bit generation rate to 1.04 Mbit/s. We analyse the quality of the generated random bits using entropy estimation and we discuss the extraction methods to obtain high-entropy bitstreams. We accurately predict the entropy output of the device backed with a stochastic model and numerical simulations.

Keywords: Quantum random number generator; Entropy extraction; Stochastic modelling; Spin noise spectroscopy

1 Introduction

The conceptual challenge in designing a random number generator (RNG) is to guarantee the true randomness of its output. Traditionally, the device was put through a number of tests [1–3] to ensure a satisfactory degree of randomness. Passing those tests is a necessary, but not a sufficient condition for true randomness. In fact, there has recently been a strong shift in recommendations away from empirical testing and towards theoretical modelling of the entropy generating physical process [4, 5]. With *a priori* knowledge of the system behaviour, the device can be guaranteed to produce cryptographically secure random numbers with near perfect entropy, so long as the model assumptions are shown to hold experimentally. Quantum systems, which are naturally probabilistic, are perfect candidates in the construction of a RNG, because the physical origin of randomness (measurement processes) are usually well defined, thus suitable for modelling.

First quantum random number generators (QRNGs) were based on the timing of radioactive decay [6], while nowadays photonic systems are more common [7]. Standard approaches are the 50/50 beam splitter configuration, timing photon detection events,

Springer

or using photon counts [7], with the fastest QRNGs based on laser phase noise fluctuations surpassing bitrates of tens of Gbit/s [7]. There also exist a number of non-optical approaches, for example avalanche detection in semiconductors [8] or quantum tunneling effect of electrons across p-n junctions [9].

In this work, we focus on the implementation known as the spin noise QRNG [10]. This QRNG approach is rooted in the experimental technique of spin noise spectroscopy (SNS), used in studying the relaxation properties of alkali metal [11] or semiconductor systems [12] in a nonperturbative way. When a sample is placed in a transverse magnetic field, a laser beam along a longitudinal axis can be used to probe fluctuations of the spin polarization of the sample which are imprinted on the polarization of the laser beam through Faraday rotation. A spin noise spectrum with a width connected to the transverse relaxation time T_2 [13] can be measured by averaging many spectra. Spin noise RNGs generate random numbers from the random spin polarization fluctuations of a sample [10]. Correlations present in the output bits are dependent on the T_2 of the measured system, requiring short relaxation times to achieve high bitrates. On the other hand, a shorter T_2 leads to a wider spin-noise spectrum [13], thus lowering the signal-to-noise ratio of the signal. An example are semiconductor systems with typical values of T_2 ranging from ns to ps [14, 15], however, such systems are currently challenging to use for a spin noise QRNG due to a poor signal-to-noise ratio.

Since being first explored in Ref. [10], little further research has been done on the spin noise QRNGs. In this work we propose to improve random number generation in a Cs gas cell by reducing T_2 using three modifications. First, an uncoated cell is used in order to increase the dephasing due to wall collisions. Second, heating of the cell increases the number of collisions between atoms per unit of time, leading to increased dephasing. Third, a magnetic gradient, that can be tuned continuously, is applied along the beam propagation axis to additionally increase the relaxation rate in a controlled manner. This leads to improvements of a factor of 50 over previous works.

The original bit generation scheme presented in Ref. [10] works by checking when the random spin noise signal crosses a threshold. In our work we present an approach that looks at the timing of the signal's fluctuations. This alternative protocol generates a larger quantity of random bits per second, leading to bitrates that surpass $1/T_2$. In this way, more entropy can be extracted from the quantum system, although some bits have to be rejected as part of the extraction phase to rid the bitstream of minimal serial correlations. It can also be added that this is an approach for the generation of random bits from a stochastic signal and is not limited to this specific type of QRNG.

2 Experimental work

The experimental setup is outlined in Fig. 1(a). It consists of a Toptica TA pro 852 nm laser source blue-detuned by 1 GHz from the Cs $D2$ line. The beam is Gaussian with the $1/e^2$ width of approximately 1 mm. The laser light is linearly polarized before entering a magnetically shielded cylindrical uncoated Thorlabs Cs glass reference cell with a diameter of 19 mm and a length of 75 mm (GC19075-CS). The input polarization is set by a $\lambda/2$ wave-plate after the linear polarizer. The laser beam passes through the cell and is split on a polarizing beam splitter (PBS). We attenuate one of the beams by a factor of 10 by using an OD1 attenuator and measure their intensities with photodiodes.

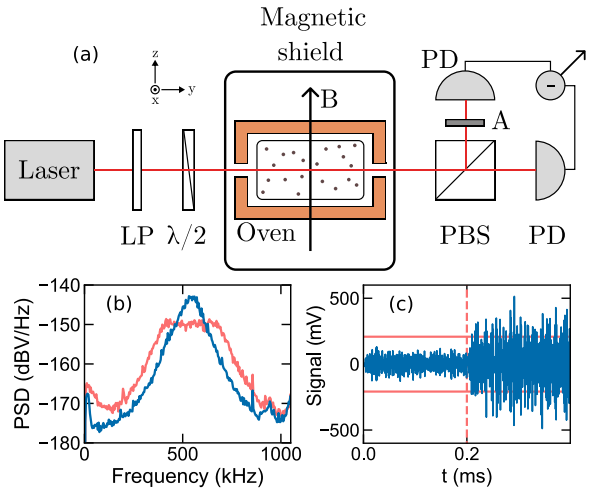

Figure 1 (**a**) Schematic representation of the experimental configuration. A laser beam is linearly polarized and sent through a magnetically shielded uncoated heated Cs cell in a magnetic field. The beam is split on a PBS according to the orientation of the λ/2 plate and recorded by a balanced polarimeter, after which the signal is sent to post processing. (**b**) Examples of recorded spin noise power spectral densities (PSD) at $\nu_L = 510$ kHz in the absence of a magnetic field gradient (blue) and in a high gradient (red) ($\mathrm{d}B_z/\mathrm{d}y \approx 1.3$ mT/m). (**c**) Polarimeter output signal; the red vertical dashed line separates the cases where the spin noise is absent (left) and when it is present (right). The red horizontal lines correspond to $\pm 5\sigma_{bg}$ where σ_{bg} is the standard deviation of photon shot noise and electronic noise of the measurement system. The QCNR is $\sigma_s/\sigma_{bg} \approx 3.16$ where σ_s corresponds to the standard deviation of the signal with spin noise. LP, linear polarizer; λ/2, half-wave plate; B, magnetic field; PBS, polarizing beam splitter; A, attenuator; PD, photodiode

By using the λ/2 wave-plate we balance the beam intensities after the PBS so that the intensity is equal on both photodiodes. In the case where one of the beam lines is attenuated, a balanced signal will correspond to the case where one beam line exiting the PBS is significantly higher in power than the other. This allows us to use higher laser powers in the sample while at the same time eliminating any non-linear effects of the photodetectors due to high incident intensity [16]. We use this setup to ensure SNR saturation of the spin noise signal, although higher powers lead to power broadening [17].

The Cs cell is placed in a magnetic field $B = 0.129$ mT perpendicular to the laser propagation axis, corresponding to a Larmor frequency of approximately $\nu_L = 450$ kHz, which defines the center of the spin noise spectrum. When the polarization has a non-zero component along the beam propagation axis, $\langle s_y \rangle \neq 0$, the polarization angle θ (vertical in the xz plane before entering the λ/2 waveplate) of the beam changes according to $\theta = \mathcal{N}\theta_0\langle s_y \rangle$; this is known as Faraday rotation (here $\mathcal{N}$ is the number of atoms in the volume of the beam and θ_0 is the rotation per atom). Since fluctuations of $\langle s_y \rangle$ are random, θ also changes randomly with time. This is then mapped to fluctuations of amplitude by use of the PBS, and measured by a balanced polarimeter. The balanced polarimeter subtracts two photodetector signals, and feeds its output into a SR650 filter unit, which amplifies the polarimeter signal by 20–30 dB. The filter output is sampled using a Digilent Analog Discovery Pro 3000 Series at a sample rate of 100 MHz.

An example of the captured spin noise spectrum is shown in Fig. 1(b). The recorded spectrum is comprised of the Lorentzian spin noise signal, and a flat background (−170 to −180 dBV/Hz) which is a sum of photon shot noise and electronic noise. The background as seen in Fig 1(b) primarily consists of photon shot noise as in our experiments electronic

noise is orders of magnitude lower. In our experiments the ratio $\sigma_s/\sigma_{bg} \approx 2-4$ represents the quantum to classical noise ratio (QCNR) where σ_s and σ_{bg} are the standard deviations of the signal with and without spin noise present, respectively.

We note that it is crucial that before any bit generation occurs the data is band-pass filtered to eliminate any noise arising from unwanted sources. For this purpose one may use, for example, a Butterworth filter of order 20 centered at ν_L. The width of the bandpass filter is to be chosen such that the entire spectrum above the photon shot noise floor is captured. This filtering step is necessary as other spectral components of the polarimeter signal arise from sources outside the measurement system (e.g. radio signals, 50 Hz power line hum etc.)

2.1 Transverse relaxation time tuning

The overarching idea of this section is to introduce improvements towards a faster dephasing rate T_2^{-1} in different experimental conditions in order to reduce bit correlations. Relaxation time estimation was primarily done by computation of the autocorrelation function $C(t) = 1/Z \sum_{i=1}^{Z} s(t_i)s(t_{i+i})$ of the signal $s(t)$, where Z is the length of $s(t)$. This can be calculated efficiently using the Wiener-Khinchin theorem:

$$C(t) = \mathcal{F}^{-1}\left(\left|\mathcal{F}\left(s(t)\right)\right|^2\right) \propto \mathrm{e}^{-t/T_2}\cos(\omega_L t), \tag{1}$$

where $\omega_L = 2\pi\nu_L$, $\mathcal{F}(f(t))$ denotes the Fourier transform (FT) of $f(t)$, and $\mathcal{F}^{-1}(f(\omega))$ its inverse. As can be seen in Fig. 1(b), the spectrum of spin noise is Lorentzian [18]. This implies that in this case $C(t)$, the inverse FT of a Lorentzian, is an exponentially decaying trigonometric function. The decay of the autocorrelation function coincides with the transverse relaxation time T_2, which can consequently be extracted. This is done by fitting an exponentially decaying trigonometric function to a numerically computed $C(t)$.

In coated Cs cells the largest contributor to spin dephasing seem to be wall collisions [20, 21], which randomize the valence electron spin, and, due to hyperfine interaction, the nuclear spin. This relaxation is directly proportional to the number density of atoms in the cell n. The lack of a coating in our cell increases the dephasing from the usual rates of 20 Hz [20] or even 0.01 Hz [22] for coated cells to approximately 0.28 MHz at 75° C, as seen from Fig. 2(a).

In this case, the spin noise spectrum is broadened to such an extent that the peak is below the shot noise level at room temperature. Due to this, we heat the cell in order to increase the number density of Cs atoms n, as the signal scales with $\sqrt{n}$ (see the Appendix). We install our cell in a ceramic oven which is additionally thermally isolated from the surrounding environment with a 5 mm layer of glass wool. This setup allows us to reach temperatures of up to 140° C. The heating element, powered by a DC current, consists of a twisted wire to minimize stray magnetic fields. The constant heating is done throughout experiments, keeping the Cs cell at a stable 140° C; we observed no detrimental effect of the current on the spin noise spectrum.

The second largest contribution to dephasing is spin-exchange with a dephasing rate $1/T_2 = \sigma_{\mathrm{SE}}\bar{\nu}n$ [23], where σ_{SE} is the cross section for a spin-exchange collision, and $\bar{\nu}$ is the relative velocity of Cs atoms. This dephasing rate depends on the temperature through $\bar{\nu}$ and n, both of which increase with temperature, making the dephasing quicker. This is clearly seen in Fig. 2(a). It is unclear to what degree spin exchange plays a role in comparison to wall collisions.

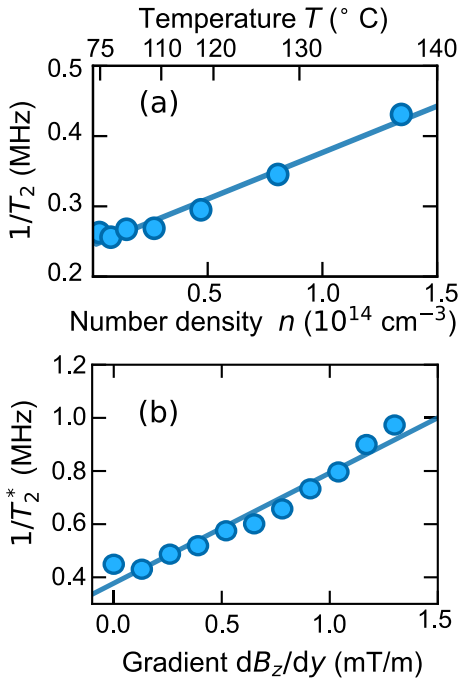

Figure 2 Dephasing rate effects. (**a**) Dependence of the relaxation rate $1/T_2$ on the temperature and number density in the cell. Heating increases the dephasing rate due to increased wall and spin exchange collisions through the relative velocity and number density of Cs atoms (described in the text). The number density n at every temperature is determined from data in Ref. [19]. (**b**) Dependence of the relaxation rate $1/T_2^*$ on magnetic field gradient. The added gradient effectively dephases the system at a faster rate $(T_2^*)^{-1}$. This was done at $T = 140°$ C. The solid lines correspond to linear fits, the error bars are inside the markers

In order to further increase $1/T_2$ we install an additional gradient generating coil on top of the heating oven. This allows us to continuously change T_2 by an additional factor. When a large gradient is present in the sample, it is no longer true that the time correlator of the system will be an exponentially decaying trigonometric function. The spin noise spectrum spreads from a Lorentzian [24] as seen in Fig. 1(b) and can be thought of a sum of peaks at varying Larmor frequencies. The time correlation function will then be

$$C(t) \propto \int_{\omega} f(\omega)\, \mathrm{e}^{-t/T_2(\omega)} \cos(\omega t)\, \mathrm{d}\omega, \tag{2}$$

where $f(\omega)$ represents the spectral profile of the spin fluctuations. Due to this, the system decoheres at a faster effective rate T_2^* which can still be determined by using the Wiener-Khinchin theorem. By assuming the system precesses at an average frequency $\bar{\omega}_L$, we can perform the same fitting procedure as $C(t) \propto \mathrm{e}^{-t/T_2^*} \cos(\bar{\omega}_L t)$. The faster decay is shown in Fig. 2(b), where we observe a change by a factor of 2. A stronger gradient could be achieved by either moving the coil closer to the cell, applying a higher current to the coils, or altering the coil geometry. This allows continuous changes to the T_2, provided the entire spin noise spectrum remains in the positive frequency domain. Since increasing the $1/T_2$ spreads the signal in the frequency domain this worsens the SNR, as shown on Fig. 1(b).

2.2 Bit generation

We use the digitized polarimeter signal to generate a bitstream of N integers valued either 0 or 1 (bits). A perfect random bit has equal probability to be 0 or 1 and is not correlated to any other bit; such bits are identically and independently distributed (IID), as well as uniformly distributed. In this case, an entropy source will have a non-biased output $B = 0$, where we define the bias as

$$B = \frac{N(1) - N(0)}{N(1) + N(0)}, \tag{3}$$

where $N(i)$ represents the number of bits valued i in the entire bitstream. We now present two different methods of generating bits from a digitized spin noise signal – here dubbed protocols. We first define a threshold Σ that the signal has to cross before any random numbers can be generated. This is done to ensure robustness of the QRNG (i.e. the generated noise comes from the quantum system in question and not from electronic or shot noise).

To fix a suitable threshold Σ we must find the variance of the signal in the absence of spin noise fluctuations. To do this, we shift the spin noise spectrum outside of the working frequency range by applying a high magnetic field ($\nu_L > 5$ MHz), so that only the electronic and photon shot noise remain. The time signal is then measured until 10^6 samples are collected, from which we calculate the variance σ^2. The threshold Σ is chosen as a multiple of σ, usually $\Sigma = 5\sigma$. Finally, the spin noise is shifted back to $\nu_L \approx 450$ kHz and the bit generation can proceed. An example of this is pictured on Fig. 1(c).

2.2.1 Protocol 1: *threshold hitting*

The threshold hitting protocol is a simple approach of bit generation that was explored in Ref. [10]. One generates bits according to whether a threshold was crossed in the positive or the negative direction. An additional waiting step is implemented. The generation algorithm is as follows:

1. Wait until a threshold Σ ($-\Sigma$) is exceeded in the positive (negative) direction.
2. Record a 1 (0) if the amplitude is higher (lower) than ($-$)Σ.
3. Wait MT_2, where $M \in \mathbb{R}$.
4. Back to 1.

The third step is crucial to avoid bit correlations, as the time correlator for this Ornstein-Uhlenbeck [25] process is $C(t) \propto e^{-t/T_2}\cos(\omega_L t)$. Therefore waiting MT_2 between bit detections exponentially removes correlations from the bitstream. Usually $M = 10$ to limit correlations to $e^{-10} \approx 10^{-5}$. If we define an event as one execution of the algorithm above, we can say that one event generates close to 1 bit of entropy (this is further discussed later). Using this protocol we achieve bitrates of up to 50 kHz, however this approach is heavily limited by the T_2.

2.2.2 Protocol 2: *times above threshold*

Here the random variable in question is the time the signal spends above (or below) the threshold Σ (or $-\Sigma$). This time is given by two crossings (events), and we will show it gives more bits of entropy per event than protocol 1. The protocol is given as follows:

1. Wait until a threshold Σ or $-\Sigma$ is exceeded at t_1.
2. Wait until the same threshold is crossed a second time at t_2.

3. Record the time above (under) threshold as $t = t_2 - t_1$.
4. Back to step 1.

Instead of looking at the variance of the signal in time the entropy comes from the phase jitter of the signal, analogous to a temporal mode optical QRNG [7]. This approach is limited in principle by the inverse of the sampling rate $\delta t = 10$ ns of our digitizer, since phase information is lost between the samples.[1] This means that the times above threshold will have an uncertainty up to one sample period (10 ns) which is not produced by the quantum system. This non-quantum source of randomness should be excluded in the extraction process.

This protocol generates a non-uniform distribution as shown in Fig. 3(a). To generate bits from this distribution we represent each integer value using 8 bits and take L least significant bits (LSB). This is because each successive bit is distributed more unevenly, as shown on Fig. 3(b). Major imbalances in the output bitstream should be avoided in order to retain an acceptably low bias, therefore last three bits are discarded (additionally, the first LSB should be discarded due to its non-quantum nature as noted in the previous paragraph). If we now instead define an event as one execution of the second algorithm, an event now generates at most L bits of entropy. The blocks of length $L - 1$ are joined into a bitstream. However, the blocks can be expected to remain correlated and therefore

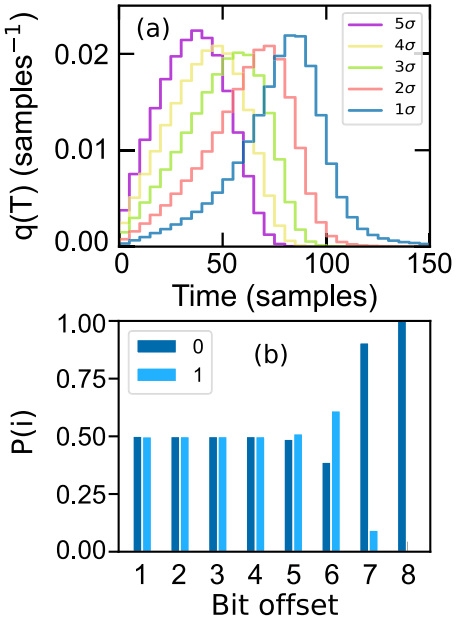

Figure 3 Protocol 2 (**a**) Examples of distributions of times above threshold for different thresholds. The discerning features are a zero-probability at $t = 0$ due to band-pass filtering, a peak at $t > 0$ due to Larmor precession, and the short-time tail that evolves from concave to convex for increasing threshold values. (**b**) Distribution of bits generated from random times above a threshold $\Sigma = 1.44\sigma$. The dark bars represent the probability for a bit to equal 0, and the light the probability for a bit to be equal to 1. One can see that each successive bit has a higher bias, so L least significant bits are taken from each integer in order to avoid large biases in the output bitstream

[1] We additionally performed measurements of the electronic jitter and found a value of approximately 10 ns, exactly corresponding to the inverse of the sampling rate.

produce a sequence that is not ideally random. The next section deals with tackling this issue.

3 Entropy

The randomness of a RNG is quantified by its entropy output. Generally, instead of Shannon entropy, the more conservative min-entropy is used, giving the lower bound of randomness. In this use case, we are interested in maximizing the min-entropy generated per bit,

$$\frac{H_{\min}}{N} = -\frac{1}{N}\log\Big(\max_{x\in\{0,1\}^N} \mathrm{Prob}(X = x)\Big), \tag{4}$$

where N is the length of the bitstream. Empirically, this can be estimated using a variety of tests. In our case, we use the methods given in the NIST SP800-90B special publication [26], which defines entropy estimation procedures for IID (identically and independently distributed) and non-IID sources. Entropy estimation for IID sources of random numbers is much simpler compared to estimation for non-IID sources, and the claim that the source is IID can also be tested, as defined in Ref. [26]. Using the IID entropy estimation techniques on a non-IID source will overestimate the entropy, while non-IID testing on IID sources will underestimate the entropy. This is hinted at in Table 1. In any case, if the estimated min-entropy is not satisfactory (sufficiently close to 1 bit of entropy per bit) then an extraction algorithm must be used on the output of the entropy source.

For protocol 1 with high waiting times MT_2 between bit generations, the resulting bitstreams can have sufficiently low correlation, and when extra care is put in the balancing the bias can be $B \approx 10^{-5}$–10^{-3}, which makes it possible to use protocol 1 with no extraction.[2] Protocol 2 typically has a larger bias $B \approx 10^{-2}$ in the output bitstream than protocol 1. Futhermore, converting non-uniformly distributed integers into series of bits introduces short range correlations on the order of L (this is discussed later).

The extraction method we use is the Toeplitz hashing algorithm, which works in the following way. Take a randomly generated Toeplitz matrix $\mathcal{T}$ with dimensions $\dim(\mathcal{T}) = (a, b)$. Now take b bits from the bitstream to generate a vector $\boldsymbol{x}$ of length b and multiply it $\boldsymbol{y} = \mathcal{T}\boldsymbol{x}$ to generate a new vector $\boldsymbol{y}$ of length a. This can be repeated using all the bits from the bitstream to generate a compressed bitstream with more entropy per bit if $b > a$.

The dimensions of $\mathcal{T}$ are chosen according to how much compression b/a is needed. In our case, we generate random bits for 10 s using protocol 2 ($L = 7$) at a rate of 1.41 Mbit/s.

Table 1 Comparison of SP800-90B testing for protocols 1 and 2 (no extraction). From protocol 1 we see that the non-IID entropy is underestimated if the IID assumption is true. On the contrary, we can see from protocol 2 that the min-entropy is greatly overestimated in IID testing if the IID claim is false; non-IID testing has to be done to get an accurate min-entropy initial claim before extraction. Protocol 1 entropy is estimated from one sample of size 0.4×10^6 bits, where for protocol 2, 10 samples of 1.4×10^6 bits were used

Protocol	IID assumption	IID entropy	Non-IID entropy
1	True	0.984	0.723
2	False	0.9276 ± 0.0008	0.750 ± 0.005

[2] We note that to achieve higher bitrates using protocol 1 it is more practical to use a lower M value and a higher compression ratio.

Table 2 The result of compressing a 1.41 Mbit/s bitstream with differently sized Toeplitz extractors $\mathcal{T}$ with compression ratios CR. The bistream was generated using protocol 2 with $L = 6$ and with the first LSB bit removed, which gives a biased output with correlations present on the order of L bits. The min-entropy per bit was determined after extraction by splitting the bitstream in 10 equally sized samples of approximately 10^6 bits and then using either non-IID or IID tests, depending on whether the IID assumption holds for the given bitstream

dim(T)	CR	Non-IID entropy	IID entropy
(921,1024)	1.11	0.86 ± 0.02	/
(757,1024)	1.35	/	0.9954 ± 0.0009
(682,1024)	1.5	/	0.995 ± 0.001

This produces approximately 14.1 Mb of random bits with an estimated min-entropy per bit $H_{\min}/N = 0.703 \pm 0.001$ using non-IID tests on 10 samples of 10^6 bits. We additionally remove the first least significant (LSB) bit due to possible classical contributions which can significantly impact the min-entropy of the bitstream. To minimize this effect we find that it is also more suitable to take $L = 6$ least significant bits. This leads to an estimated $H_{\min}/N = 0.750 \pm 0.005$. Since blocks of L bits are correlated, we choose $b = 1024 \gg L$. The second dimension is then given by $a = \lfloor bH_{\min}/N \rfloor = 768$.

A Toeplitz matrix of this size will guarantee that the output bitstream is practically IID and unbiased. An example of this claim is shown in Table 2. We compress the above mentioned bitstream with differently sized Toeplitz matrices. Using insufficient compression ratios leads to failure of the IID tests. Since the compression ratio for $a = 682$ exceeds approximately $1/0.75 = 1.333$ the bitstream passes all IID tests and the output entropy per bit is estimated as $H_{\min} = 0.995 \pm 0.001$. Performing the extraction with a slighly lower compression ratio of 1.35 makes it possible to reach the final bitrate of 1.04 Mbit/s of highly entropic random numbers. We note that this bitrate is higher than $1/T_2 = 500$ kHz in this particular test where no magnetic field gradient was applied.

On top of entropy estimation, batteries of statistical tests are usually done on the output bitstreams in order to find statistical shortcomings of the entropy source. Some standard testing suites we used are dieharder [1], TestU01 [2], and the NIST [3] tests, which were all required to pass (and they did pass). We do not put much emphasis on such statistical testing, as it is always possible to successfully pass all standard tests using sufficient compression even with badly flawed generators.

4 Stochastic modelling

We describe the spin state of the alkali vapour using a density matrix ρ which evolves randomly in time. In the absence of coherent excitations, ρ is diagonal, since all off-diagonal elements decay due to decoherence effects. In magnetic fields typically used, the alkali gas remains close to unpolarized due to thermal effects ($kT \gg g\mu_B B$, where g is the g-factor and μ_B is the Bohr magneton) and, to a good approximation, the density matrix is maximally entropic, $\rho \propto \mathbb{1}$.

The time evolution of ρ, however, is random. The mechanism of randomness generation are collisions. Two processes can be distinguished. The first is spin exchange when two Cs atoms collide, where the total spin S^2 is conserved. Atoms with an initial angular momentum state $|m\rangle$ evolve to $|m+j\rangle$ with $j = 0, \pm 1$ during a collision [10, 27]. Collisions with $j = 0$ cause fluctuations of diagonal elements of ρ, whereas collisions with $j = \pm 1$ cause relaxation. There exist other kinds of collisions between atoms, for example spin-destruction

collisions, but their cross sections are typically orders of magnitude lower than those of spin-exchange [28].

The second major randomness generating process are wall collisions. Reference vapour cells are coated internally (e.g. paraffin) such that the polarization of an atom persists through as many collisions as possible. However, due to lack of a cell coating in our experiment, an atom's polarization is essentially randomized upon collision with the cell wall surface [20]. The interaction between the atoms of the wall and the valence alkali-metal spin is comprised of a dipole-dipole interaction and a spin orbit-type coupling, and it is a well-understood quantum process [29]. Because of the large reduction in T_2 after the coating is removed, we believe that this randomness generating process dominates over Cs-Cs atomic collisions.

To model the evolution of the system it is easier to consider a stochastic picture rather than to perform quantum-mechanical calculations. Consider a system of alkali metal atoms with a transverse relaxation rate T_2 in a magnetic field along the z axis with a Larmor frequency $\omega_L = 2\pi\nu_L$. Then the time evolution of the spin expectation value $\langle s\rangle = (\langle s_x\rangle, \langle s_y\rangle)^T$ is given by a stochastic differential equation [23]

$$\mathrm{d}\langle s\rangle = -D\langle s\rangle\,\mathrm{d}t + F\,\mathrm{d}\eta, \tag{5}$$

where the matrices D and F are defined as

$$D = \begin{pmatrix} 1/T_2 & -\omega_L \\ \omega_L & 1/T_2 \end{pmatrix}, \qquad F = \frac{1}{2\sqrt{T_2}}\mathbb{1}, \tag{6}$$

and $\mathrm{d}\eta = \binom{\mathrm{d}\eta_x}{\mathrm{d}\eta_y}$ is a two-dimensional Wiener process. This alternative model is exact when the alkali gas is unpolarized, which is only approximately fulfilled given that $kT/g\mu_B B \approx 10^4$ for typical values $B = 10^{-2}$ T and $T = 140\,^\circ$C in our experiments. Knowing all the system parameters, we can calculate the entropy per generated bit. To do this we calculate stochastic properties of $\langle s\rangle$, apply the appropriate protocol, and then calculate what is expected on the output of the polarimeter.

4.1 Protocol 1

To find the entropy per bit in protocol 1 we must know the amplitude distribution of the signal. We rewrite Eq. (5) as a Langevin equation for the variable $s = \langle s_x\rangle + i\langle s_y\rangle$:

$$\dot{s} = -i\omega_L s - \frac{1}{T_2}s + f(t) \tag{7}$$

with the fluctuations $f(t)$ defined by $\langle f(t)\rangle = 0$ and $\langle f(t)f(t')\rangle = \frac{1}{4T_2}\delta(t-t')$. From this the time evolution of the variances of s_x and s_y are found to be the same as a univariate Ornstein-Uhlenbeck process:

$$\sigma^2_{s_y}(t) = \frac{1}{8}\left(1 - \mathrm{e}^{-\frac{2t}{T_2}}\right) \underset{t\to\infty}{=} \frac{1}{8}. \tag{8}$$

We consider the stationary case $t \to \infty$, as the sample is approximately unpolarized. Alternatively, one can solve the stationary Fokker-Planck equation to obtain the same result.

In order to predict the min-entropy we must know the variance $(\sigma')^2 = 4N\theta_0\sigma_{s_y}^2$ (see the Appendix for the derivation) of the signal at the output of the polarimeter.

If we suppose that there are no correlations within the bistream (large waiting time M), then the min-entropy of the bitstream depends only on the bias μ of the signal, which is the mean of the signal's amplitude distribution. This bias is the result of imperfect balancing or drifting of the measurement system. One event generates $H_{\min} = -\log_2(\max_{x=\{0,1\}} \mathrm{P}(X = x))$ bits of entropy, where as shown in the Appendix,

$$\max_{x=\{0,1\}} \mathrm{P}(X = x) = \frac{1}{2} + \left| \frac{(\mu/\sigma')}{\sqrt{2\pi} - 2(\Sigma/\sigma')} \right| + \mathcal{O}(\mu^3), \tag{9}$$

where $(\sigma')^2$ is the variance of the polarimeter signal. For example, if $\mu/\sigma' = 10^{-4}$, and $\Sigma/\sigma' = 1.4$, then per event we generate a minimum of $-\log_2(\mathrm{P}(X = 1)) \approx 0.9990$ bits of entropy. This holds when successive bits are uncorrelated, i.e., as M approaches infinity. In reality, serial correlations are present to some small degree.

4.2 Protocol 2

The problem of calculating the distribution of times above a threshold q for a Markov process has been studied since the early 70s [30]. The challenge of evaluating q depends upon the complexity of the infinitezimal propagator $\mathcal{A}$ that drives the stochastic process [31]. Although the problem is solved in Ref. [30] for the univariate Ornstein-Uhlenbeck process, the bivariate case presented in this paper leads to a non-trivial differential equation. We opted to numerically simulate this distribution using the stochastic model in Eq. (5) to generate times above the different thresholds as shown in Fig. 4(a).[3]

The generated distributions can be used to estimate the entropy production of the experimental system. We start with an experimentally observed distribution q with threshold Σ binned in n bins. We then compute a distribution q' from a simulated signal by using the same experimental parameters $\omega_L = 2\pi\nu_L$, T_2, Σ, again using n bins. Because of uncertainties of the experimental parameters, imperfections of the integrator, and deviations from the quantum model, the two distributions might not match perfectly. An example of the overlap between experimental and simulated histograms is demonstrated in Fig. 4(b).

Now, the min-entropies $H_{\min}(q) = -\log_2 \max_t q(t)$ are computed. The simulated curve has a min-entropy of $H_{\min} = 5.56$ bits, while the experimental $H_{\min} = 5.55$ bits. In our case we generate $L\,(H_{\min}/N)$ bits of entropy from each block of L bits. If we consider the entirety of the experimental distribution we must take all $L = 8$ LSB, and the estimated min-entropy is found to be $H_{\min}/N = 0.4735 \pm 0.0007$ (using 10 samples of 2.2×10^6 bits). One event thus generates $L(H_{\min}/N) = 3.788$ bits of entropy, the rest is lost due to correlations.

This is clearly the first advantage over protocol 1, where each event (i.e., one execution of the protocol 1 algorithm) generates at most 1 bit of entropy. By using a more sophisticated extraction protocol, we can extract more entropy from the system per event as established above.

Another way to improve the bitrate is to increase the Larmor frequency ν_L while also keeping $\nu_L\delta t$ constant. This makes for more events per second with adequate compensation by sampling faster, which is required to keep the distribution as flat (maximally

[3] The numerical integration was done by a fixed step-size explicit 3-stage Milstein method for an Ito problem with strong and weak order 1.0. The used parameters were $dt = 10$ ns, $\omega_L = 2\pi \times 450$ kHz, and $T_2 = 2\,\mu$s. The 10^8 samples of the time signal are computed according to Eq. (5) and then band-pass filtered around ω_L.

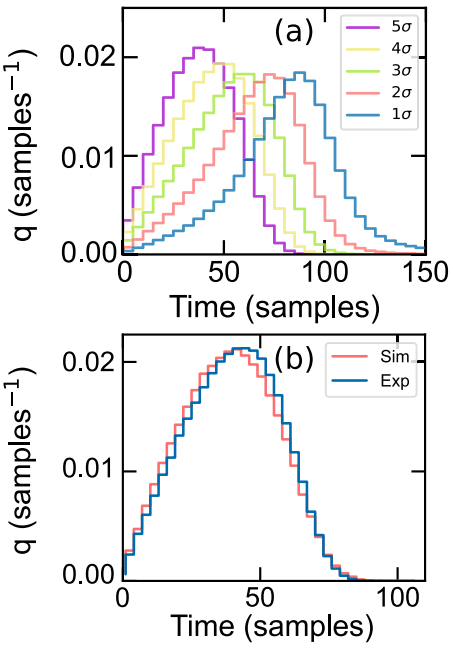

Figure 4 Simulations of protocol 2. (**a**) Distributions of times above threshold for a simulated time signal at different threshold values Σ. Here, a sample denotes a timestep of the integrator.[3] This captures the experimental results, although the experimental and simulated distributions do not overlap completely due to uncertainty of the experimental parameters. (**b**) The simulated signal can be used to calculate the distribution q (using the same $\omega_L = 450$ kHz, $T_2 = 2$ µs, and threshold $\Sigma = 5\sigma$), which can be used to estimate the entropy production of the experimental system

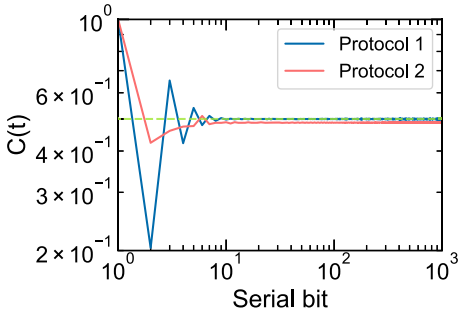

Figure 5 Decay of serial bit correlations for both protocols. The dashed green line represents the case where $C(t) = 1/Z \sum_{i=1}^{Z} X_i X_{i+t} = 0.5$. Protocol 1 (with $M = 0$) has exponentially decaying correlations present in the output bitstream that decays close to 0.5, based on the bias. Protocol 2 has a short-term correlation of length L (here $L = 5$). An uptick at the 5th serial bit is seen for protocol 2, as the times above threshold remain correlated, although, as shown, this correlation decays faster than that of protocol 1. Due to the larger biases present in protocol 2, it can also be seen that the autocorrelation does not decay as close to 0.5 as that of protocol 1

entropic) as possible. However, these events would be more correlated, so a higher compression ratio would have to be used in the extraction process, thus it is hard to predict the limit to this approach.

The second reason why protocol 2 is superior to protocol 1 can be seen from the serial correlation of the generated bits. This is shown on Fig. 5, where bit autocorrelations $C(t)$ were computed using the Wiener-Khinchin theorem for bits generated using both protocols (for a fair comparison $M = 0$ has to be taken) on the same experimental run.

For an unbiased IID source, each bit would have an autocorrelation $C(t)$ of exactly 0.5 to the previous bit. In this case, this is not true until we perform extraction, where the extraction ratio required is a direct consequence of the magnitude of correlations present. What

Fig. 5 suggests is that the correlation between the bits generated using protocol 2 decays faster than that of the bits generated using protocol 1 (with $M = 0$). Consequently, a lower extraction ratio is required in order to have a satisfactory min-entropy using protocol 2.

5 Conclusion

We show an improvement of three orders of magnitude of the bit rate in spin noise QRNGs by using several advances. First, we reduced T_2 by using a heated uncoated cell. The relaxation rate is further continuously tunable using a gradient inducing coil. This allows a bit generation rate of up to 50 kHz using previously known methods (protocol 1). Secondly, we used a more efficient entropy extraction scheme (protocol 2). We show that bits generated using different protocols have correlations that decay at different rates, in this case making protocol 2 superior to protocol 1, achieving bitrates of 1.04 Mb/s.

A spin noise QRNG could also be realized in solid-state systems. We propose two possible candidates for further research. The first is heavily doped (close to the metal-insulator transition) GaAs heterostrocture, because data suggests a relatively high $T_2 \approx$ ns at room temperatures [14], while the band gap can be engineered by changing the heterostructure geometry. Another alternative could be a *n*-doped CdTe quantum well, which can have up to an order of magnitude slower relaxation rates at room temperatures [14].

By lowering the T_2 to increase the bit rate or min-entropy, the signal-to-noise ratio (SNR) is worsened as the spectrum broadens. The SNR can be improved using more involved measurement systems (e.g., heterodyne detection improves SNR in low-shot-noise scenarios [32]). The limiting factor in spin noise QRNG type experiments is photon shot noise, as the electronic noise of the measurement electronics is orders of magnitude lower. Shot noise squeezing could also improve SNR [33] by tackling this issue, although there are known limits to this approach [34]. It seems that improvement in experimental techniques is still required to capture live spin noise signals at faster relaxation rates without lengthy time averaging. If, one day, semiconductor based spin noise QRNGs are realised, the advances to bit generation and T_2 tuning presented in this work directly apply.

Appendix: Protocol 1 entropy estimate

To model the entropy exiting the polarimeter we have to propagate the probability distribution function of the amplitudes, as given by Eq. (8), through the measurement system. We pick the propagation axis for the laser y. At any given time, the Faraday rotation θ of the laser is proportional to the polarization $\langle s_y \rangle$ along the axis of propagation

$$\theta = \mathcal{N}\theta_0 \langle s_y \rangle, \qquad \theta_0 = \frac{\lambda^2 \gamma}{2\pi A \delta}, \tag{10}$$

where $\mathcal{N}$ is the number of atoms in the laser beam volume in the cell, θ_0 the Faraday rotation due to one spin, λ the wavelength of the light, A the laser cross section, γ the radiative width of the transition, and δ the detuning from the transition line. Upon exiting the glass cell the laser light, imprinted with randomly fluctuating polarization, is split on the PBS into two components given by the vector $\boldsymbol{J}$

$$\boldsymbol{J} = \begin{bmatrix} \cos(\theta + \pi/4) \\ \sin(\theta + \pi/4) \end{bmatrix}. \tag{11}$$

The phase shift of $\pi/4$ is due to the balancing of the polarimeter. Although $\mathcal{N} = 10^{12}$, and $\theta_0 \sim 10$ nrad in our experiments, $\langle s_y \rangle$ remains 0. On the other hand, fluctuations of the Faraday rotation angle are given by $\sigma_\theta = \sqrt{\mathcal{N}}\theta_0\sigma_{s_y}$ where $\sigma_{s_y} = 1/8$ are fluctuations of the spin polarization in the y direction as shown in Eq. (8). As $\sqrt{\mathcal{N}}\theta_0 \approx 10^{-3}$, the fluctuations σ_θ remain small, which allows us to calculate the polarimeter output signal S by Taylor expansion

$$S = \cos(\theta + \pi/4) - \sin(\theta + \pi/4) \approx -2\theta + \mathcal{O}\left(\theta^3\right). \tag{12}$$

This shows that the variance of the polarimeter signal, $(\sigma')^2$, is the sum of two completely correlated processes, therefore $(\sigma')^2 = 4\mathcal{N}\theta_0^2\sigma_{s_y}^2$. This shows the distribution of amplitudes of the signal outputted from the polarimeter remains Gaussian, given that θ are small. Any amplification of the signal (in the polarimeter, the SR650, or the Digilent Analog Discovery Pro) additionally spreads this Gaussian, however, we do not elaborate on this further.

In protocol 1 with $M \to \infty$ the outputted bits are uncorrelated. The only deviation from ideal random numbers is then due to an imbalanced signal. Consider that the mean of the signal μ drifts in time away from 0. At any given time, the probability the signal is above or below a threshold is easily calculated

$$\mathrm{P}(S > \Sigma) = 1 - \mathrm{Prob}(S \leq \Sigma) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\Sigma - \mu}{\sqrt{2}\sigma'}\right), \tag{13}$$

and similarly $\mathrm{P}(S < -\Sigma) = \frac{1}{2}\,\mathrm{erfc}(\frac{\Sigma+\mu}{\sqrt{2}\sigma'})$. These two probabilities correspond to the probability to generate bit $X \in \{0, 1\}$ at the output, given that $|S| > \Sigma$. For example, $\mathrm{P}(X = 1) = \mathrm{P}(S > \Sigma \mid |S| > \Sigma)$. This allows us to predict μ from the measured bias using Bayes' theorem

$$B = \frac{\mathrm{P}(X = 1) - \mathrm{P}(X = 0)}{\mathrm{P}(X = 1) + \mathrm{P}(X = 0)} = \frac{\mathrm{P}(S > \Sigma) - \mathrm{P}(S < -\Sigma)}{\mathrm{P}(S > \Sigma) + \mathrm{P}(S < -\Sigma)}. \tag{14}$$

This also enables us to make a min-entropy claim for the output of the QRNG with protocol 1 (and a large M) by expanding $\mathrm{P}(X = 1)$ or $\mathrm{P}(X = 0)$ as a Taylor series around $\Sigma \pm \mu$. To second order in μ, the probability to generate a bit $X = 1$ is

$$\mathrm{P}(X = 1) = \frac{\mathrm{P}(S > \Sigma)}{\mathrm{P}(|S| > \Sigma)} \approx \frac{1}{2} + \frac{(\mu/\sigma')}{|\sqrt{2\pi} - 2(\Sigma/\sigma')|}. \tag{15}$$

Alternatively, for $\mathrm{P}(X = 0)$ one has to make the substitution $\mu \to -\mu$. In this approximation the bias equals

$$B \approx \frac{\sqrt{2}(\mu/\sigma')}{|\sqrt{\pi} - \sqrt{2}(\Sigma/\sigma')|}. \tag{16}$$

It is important to note the shortcomings of this analytical method. The issue with such an expansion is that when $\Sigma \gg \sigma$ the approximations cease to hold well, as the derivative of $P(X = 1)$, D, becomes underestimated to the extreme case where it is the inverse of the correct value. This can partly be mended by taking the absolute value of the derivative of

$P(X = 1)$ after the expansion. For orders higher than $\mathcal{O}(\mu)$, the derivative of $P(X = 1)$ is generally also a function of μ. In general we can write

$$\mathrm{P}(X = 1) = \frac{1}{2} + \left|D(\mu, \Sigma, \sigma)\right|\mu. \tag{17}$$

Additionally, as the values of μ in protocol 1 are small, we can also perform a Taylor expansion of $D(\mu, \Sigma, \sigma)$ up to first order to obtain $D(\mu, \Sigma, \sigma) = D(\Sigma, \sigma) + \mathcal{O}(\mu^2)$. Taking a higher order expansion of $P(X = 1)$ in Eq. (15) allows a better estimation of the gradient $D(\mu, \Sigma, \sigma)$, and leads to a better approximation. In practice, however, it is simpler and more accurate to do this entropy evaluation numerically.

Acknowledgements
Not applicable.

Funding
This work was supported by the Government Office for the Protection of Classified Information (Research Project No. V1-2119), the Slovenian Research Agency (Research Project No. V1-2119 and Research Core Fundings No. P1-0125, No. P1-0099 and No. P1-0416), the Slovenian Research Program P2-0250 and MORS.

Data availability
The data is publically available on the Zenodo repository (https://doi.org/10.5281/zenodo.7934348). There are two datasets published – the first dataset is recorded in the presence of a large magnetic gradient while the second is recorded in the absence of a gradient. The second dataset (Cs14) was used as the example in this work.

Declarations

Competing interests
The authors declare no competing interests.

Author contributions
P.J., E.Z., and R.Ž. proposed the idea for the work. M.K., S.B., T.M., K.G., and P.J. carried out the experimental measurements. S.B. designed the custom electronics that were used. J.P. suggested and worked on protocol 2. M.K., and P.J. wrote the required code for simulations. M.K. wrote code for post processing of the data. M.K., and R.Ž. worked on entropy estimation and extraction. All authors analyzed and interpreted the results. M.K. carried out writing the manuscript. All authors reviewed and approved the final manuscript.

Author details
[1] Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia. [2] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia. [3] Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, SI-1000 Ljubljana, Slovenia.

Received: 16 June 2023 Accepted: 7 February 2024 Published online: 19 February 2024

References

1. Brown R. Dieharder: a random Number Test Suite, version 3.31.1. 2004. https://webhome.phy.duke.edu/~rgb/General/dieharder.php.
2. McCullough BD. J Appl Econ. 2006;21:677–82.
3. Rukhin A, et al. A statistical test suite for Random and Pseudorandom Number Generators for cryptographic applications, NIST Special Publication 800-22, Revision 1a. 2010. http://csrc.nist.gov/rng/.
4. Peter M, Schindler W. A proposal for functionality classes for Random Number Generators, version 2.35. 2022.
5. Fischer V. Design of secture TRNGs for cryptography – past, present, and future. In: Workshop Wr0ng2017. Paris. 2017.
6. Schmidt H. J Appl Phys. 1970;41:462.
7. Herrero-Collantes M, Garcia-Escartin JC. Rev Mod Phys. 2017;89:015004.
8. Stipcevic M, Rogina BM. Rev Sci Instrum. 2007;78:045104.
9. Zhou H, Li J, Zhang W, Long G-L. Phys Rev Appl. 2019;11:034060.
10. Katsoprinakis GE, Polis M, Tavernarakis A, Dellis AT, Kominis IK. Phys Rev A. 2008;77:054101.
11. Mihaila B, Crooker SA, Rickel DG, Blagoev KB, Littlewood PB, Smith DL. Phys Rev A. 2006;74:043819.
12. Müller GM, Oestreich M, Römer M, Hübner J. Physica E. 2018;43:569.
13. Sinitsyn NA, Pershin YV. Rep Prog Phys. 2016;79:106501.
14. Harmon NJ, Putikka WO, Joynt R. Phys Rev B. 2010;81:085320.
15. Oertel S, Hübner J, Oestreich M. Appl Phys Lett. 2008;93:132112.
16. Glasenapp P, Greilich A, Ryzhov II, Zapasskii VS, Yakovlev DR, Kozlov GG, Bayer M. Phys Rev B. 2013;88:165314.
17. Fomin AA, Petrov MY, Kozlov GG, Vershovskii AK, Glazov MM, Zapasskii VS. Phys Rev A. 2021;103:042820.
18. Katsoprinakis GE, Dellis AT, Kominis IK. Phys Rev A. 2007;75:042502.
19. Steck D. Cesium D line data. 2010.

20. Graf MT, Kimball DF, Rochester SM, Kerner K, Wong C, Budker D, Alexandrov EB, Balabas MV, Yashchuk VV. Phys Rev A. 2005;72:023401.
21. Budker D, Hollberg L, Kimball DF, Kitching J, Pustelny S, Yashchuk VV. Phys Rev A. 2005;71:012903.
22. Balabas MV, Karaulanov T, Ledbetter MP, Budker D. Phys Rev Lett. 2010;105:070801.
23. Katsoprinakis G. Spin noise, decoherence and magnetic effects in alkali atoms and biomolecules. Doctoral dissertation, University of Crete, Department of Physics Fo.R.T.H., Institute of Electronic Structure and Lasers. 2010.
24. Levitt MH. Spin dynamics: basics of nuclear magnetic resonance. New York: Wiley; 2008.
25. Gillespie DT. Am J Phys. 1996;64:225.
26. Turan MS, Barker E, Kelsey J, McKay KA, Baish ML, Boyle M. NIST Special Publication SB800-90B. https://doi.org/10.6028/NIST.SP.800-90B.
27. Appelt S, Ben-Amar Baranga A, Erickson CJ, Young AR, Happer W. Phys Rev A. 1998;58:1412.
28. Bhaskar ND, Pietras J, Camparo J, Happer W, Liran J. Phys Rev Lett. 1980;44:930.
29. Bouchiat MA, Brossel J. Phys Rev. 1966;147:41.
30. Stone LD, Belkin B, Snyder MA. J Math Anal Appl. 1970;30:448.
31. Oksendal B. Stochastic differential equations. Berlin: Springer; 2000.
32. Cronenberger S, Scalbert D. Rev Sci Instrum. 2016;87:093111.
33. Bai L, Zhang L, Yang Y, Chang R, Qin Y, He J, Wen X, Wang J. Opt Express. 2022;30:1925.
34. Lucivero VG, Dimic A, Kong J, Jiménez-Martínez R, Mitchell MW. Phys Rev A. 2017;95:041803(R).

Publisher's Note