An analysis of the stationary operation of atomic clocks

Martin Fraas
Theoretische Physik, ETH Zürich, 8093 Zürich, Switzerland

October 17, 2013

Abstract
We develop an abstract model of atomic clocks that describes the full dynamics of repeated synchronization of a classical oscillator with a quantum reference. We prove existence of a stationary state of the model and study its dependence on the control scheme, the interrogation time and the stability of the oscillator. Control schemes are classified by autocorrelations of the frequency before and after the synchronization. This provides a one-parameter family of stationary states that differ in variance of the frequency, yet produce an equally good clock time. Formally, this is a consequence of a novel Cramer-Rao type inequality. We also derive an optimal interrogation time and show that it is determined by the balance between the dissipation of the oscillator and the information gained from the synchronization. Similar models have been studied in connection to the Dick effect.

1 Introduction
Advances in the microwave spectroscopy in 40’s and early 50’s led to the development of the first atomic clocks [2, Chapter 4]. The milestone was a Cesium atomic clock constructed by Essen and Parry in 1955, it was based on the spectroscopy method of Ramsey and it achieved an accuracy of 10µs per day – an order better than the best quartz clocks available at the time. This turned to be a beginning of a steady progress and 60 years later the best atomic clock [23] is by 7 orders better than that of Essen and Parry. Also the history repeats: Construction of such a precise clocks is possible only due to major advances in manipulation of trapped ions and in the laser spectroscopy.

It is easy to understand why a precise spectroscopy is a prerequisite for the construction of atomic clocks. The precision achievable in a single synchronization of the clocks oscillator with the quantum reference is determined by the precision of the spectroscopy method employed in the synchronization. This precision is well understood (at least on the fundamental level) and extensively studied within the quantum estimation theory. It is less clear how exactly the
spectroscopy precision determines the precision of the clock time. And it is the aim of this article to have a closer look upon this question. We aim to study the stationary operation of the atomic clock as achieved after many repeated synchronizations and to show how the precision of the clock is determined. The effects that we are going to describe are well understood on the engineering level and are referred to as the Dick effect [13].

We shortly describe how an atomic clock operates. A clock is any device that uses the periodic motion of an oscillator to measure time. It counts beats of the oscillator and displays them in convenient units. Provided that it counts correctly, accuracy of the clock is determined by properties of its oscillator, the most important being its mean frequency and its stability. To increase the stability one often employs a frequency reference through the following control scheme.

The frequency reference is a physical process with a stable characteristic frequency that can be observed repeatedly. The oscillator is tuned to this frequency and then continuously (or in discrete time intervals) synchronized with the reference. This stabilizes the oscillator given that the reference frequency is more stable than the oscillator itself. There is always some obstruction in observing the frequency reference that prevents perfect synchronization. We proceed by giving two examples.

A somewhat fairy tale example is a pendulum (clock) with adjustable length that is tuned to the frequency of a chosen pulsar. The pulsar is the frequency reference and at anytime we observe the pulsar in the sky we adjust the length of the pendulum to match the frequency of the pulsar.

The next example is less artificial and more relevant for our study. We describe a Cesium atomic clock: A voltage-controlled quartz crystal oscillator tuned to a frequency of transition between two hyperfine split levels of the Cesium atom, a ground state $|g\rangle$ and an excited state $|e\rangle$ separated by an energy gap $\omega_0$. The free Hamiltonian of the atom being

$$H_0 = \omega_0 \sigma_z / 2.$$  

A quartz crystal with frequency $\omega(t)$ operates an electro-magnetic field in an U-shaped cavity (see fig. 1) so that a Cesium atom passing through the cavity experiences an interaction Hamiltonian

$$H_i(t) = -\mu \vec{B}(t) \cdot \vec{\sigma}, \quad \vec{B}(t) = B(\sin \omega(t), \cos \omega(t), 0).$$

The strength of the magnetic field $B$ can be adjusted so that a Cesium atom entering the left part of the cavity in the ground state would exit the cavity in the Bloch superposition, e.g. $(|g\rangle + |e\rangle)/\sqrt{2}$. During the flight between the cavity ends (of duration $\tau$) the state acquires a relative phase

$$\int_0^\tau (\omega_0 - \omega(s))ds.$$  

Upon exiting the right side of the cavity the probability of finding the atom in the excited state is $P(e) = \cos^2(\int (\omega_0 - \omega(s))/2)$. Using a beam of Cesium atoms one can then adjust a voltage controlling the quartz crystal to achieve $P(e) = 1$.  

2
Figure 1: A control scheme of the Cesium atomic clock. A quartz crystal operates an electromagnetic field inside a cavity. A beam of Cesium atoms passes through the cavity into a detector and provides information on the frequency difference between the quartz crystal frequency and the Cesium atom reference frequency. This information is used in a loop to control the quartz crystal in order to make the difference zero.

The above synchronization procedure employs the Ramsey interferometry (see fig. 1) and represents a set-up that includes the preparation of a state, its evolution for time $\tau$ and a measurement, the evolution being described in a rotating frame by a Hamiltonian $H(t) = (\omega_0 - \omega(t))H$. Regardless of details of the experimental setup, it is clear that information about $\omega_0 - \omega(t)$ can be acquired given knowledge of $H$ and $\tau$. This is a standard starting point for an abstract model of an atomic clock.

We call any clock with a quantum mechanical frequency reference an atomic clock. It is a device that synchronizes a classical oscillator with a quantum frequency reference and its clock time is given by counting the former classical oscillations. The quantum mechanical nature of the reference makes perfect synchronization impossible due to the fundamental uncertainty in observing the reference. Furthermore, the observation disturbs the state of the frequency reference and as a consequence the reference needs to be reseted in between every use. The resulting entropy production has to be viewed as an additional cost of the synchronization.

If we had a perfect classical oscillator there would be no need for continuous synchronization with a reference. It is therefore clear that the stability of the clock’s oscillator should play a major role in determining the stationary operation of the atomic clock: A stable oscillator requires less synchronization than an unstable oscillator in order to achieve the same accuracy of the clock. By a standard diffusion-dissipation relation one guesses that an optimal operation should be characterized by a balance between the dissipation of the oscillator and the information gained by the synchronization. To rigorously establish such
Figure 2: Bloch representation of the state of a Cesium atom during the Ramsey interferometry. Top left: Before entering a cavity the atom is in the ground state. Top Right: In between cavity ends the state is on the equator and undergoes Bloch oscillations with frequency $\omega_0$. Bottom: The state before detection has an angle with the excited state proportional to the acquired relative phase $\int (\omega_0 - \omega(s))$.

A relation is one of the goals of this article.

There are many types of atomic clocks differing both in the classical oscillator and the quantum frequency reference. We do not aim to discuss any particular type and motivated by the example of the Cesium clock we establish the following abstract model.

We decompose the frequency of the oscillator, $\omega(t)$, into the reference frequency, $\omega_0$, and an instantaneous relative error $y(t)$,

$$\frac{\omega(t) - \omega_0}{\omega_0} = y(t).$$

In absence of synchronization the frequency of the oscillator (and hence the error) is described by a classical stochastic process. The quantum frequency reference is modeled by a $y(t)$-dependent quantum state $\rho$. At discrete times, a measurement of the quantum state provides an estimation of $y(t)$ and an adjustment of $\omega(t)$ towards the intended frequency $\omega_0$ is made according to the outcome.

In more details, the synchronization process consists of three steps: (We
describe it at \( t = 0 \)

1. The reference system is initialized to a state \( \rho_0 \) and evolves freely for an interrogation time \( \tau \),

\[
\rho_0 \rightarrow \rho(\bar{y}) = e^{-i\tau \omega_0 \bar{y} H} \rho_0 e^{i\tau \omega_0 \bar{y} H}, \quad \left( \bar{y} := \frac{1}{\tau} \int_0^\tau y(s) ds \right),
\]

where \( H \) is a known Hamiltonian.

2. Subsequent measurement provides an estimation, \( \hat{\bar{y}} \), of the time averaged frequency error, \( \bar{y} \). The frequency is then adjusted according to the outcome,

\[
y(\tau) \rightarrow y(\tau) - \hat{\bar{y}}. \tag{1}
\]

3. The procedure is repeated with a period \( T \geq \tau \). For simplicity we will consider only \( T = \tau \).

In due time we make this even more abstract and a clock model (see Section 5) would eventually be defined by:

- A stochastic evolution is absence of the synchronization, this is often referred to as the local oscillator noise. To ensure that synchronization with the quantum reference is the only way how to determine and control the frequency error we assume that this process is a martingale.

- Family of states \( \rho_T(\bar{y}) \). This generalizes and replaces Hamiltonian evolution from the first step above. In particular it incorporates an open system evolution (e.g. dephasing) of the quantum reference. The index \( T \) of the state is a reminder that the dependence on \( T \) is non-trivial for an open system evolution.

- The estimation strategy. It consists of a POVM measurement \( \Pi(\alpha) \) and an estimator \( \Phi \). The estimator describes the post-processing of a measurement outcome, upon an outcome \( \alpha \) a guess \( \Phi(\alpha) \) of the unknown \( \bar{y} \) is made.

We call the stochastic process \( y(t) \), the state of the clock.

Many generalizations of the model are possible. We mention the most prominent. \( T \neq \tau \): The case \( T > \tau \) would describe a clock with a dead time, like Cs fountain clock [32]. The case \( T < \tau \) describes an atomic clock with more then one atomic ensemble [14]. Another generalization is to consider a more elaborate control scheme, the control in Eq. (1) can depend on the history and the first derivative of the error. The size of the Dick effect depends and can be controlled by such improvements. Although this is very important for applications, the principle and consequences of the effect remains the same whichever model one consider.
The clock time within our model is given by
\[ t_{\text{clock}} = \frac{1}{\omega_0} \int_0^t \omega(s) \, ds \]
\[ = t + \int_0^t y(s) \, ds. \]
It follows that the accuracy of clock is determined by a relative frequency error \( y(t) \). In particular a high frequency reference is superior to a low frequency reference provided they are equal in all other aspects.

Our goal is to describe an atomic clock in a well defined mathematical manner and to pose several interesting problems related to its operation. Among the problems are:

1. Existence of a stationary state of the clock.
2. Properties of the stationary state, especially the variance of the associated clock time.
3. Entropy production of the clock.

In this article we first develop the model in its full generality and then focus on the first and the second question. We focus on a special family of unbiased clocks (see Section 5). This is a natural family of clocks for which the clock time is eventually unbiased, \( E[t_{\text{clock}}] = t \), independently of the initial conditions. This holds true if and only if the estimation strategy is a multiple (denoted by \( 1 - \zeta \)) of an unbiased estimation strategy. Whenever we want to stress a particular value of \( \zeta \) we speak about \( \zeta \)-unbiased clock.

We show that for an unbiased clock, the state \( y(t) \) is a supemartingale. It then follows from Doob’s martingale convergence theorem that an unbiased clock has an unbiased stationary state, i.e. a state for which \( y(t + nT) \), \( n \in \mathbb{N} \) is a stationary process and \( E[y(t)] = 0 \). The parameter \( \zeta \) effect the autocorrelation of \( y(t) \) through the adjustment \( 1 \) and is equivalently described by an exponentially decaying autocorrelation of the stationary state
\[ E[y(t + T)y(0)] = \zeta E[y(t)y(0)]. \]
The case \( \zeta = 0 \) (unbiased estimation) corresponds to a solution with no correlations over periods longer then \( T \).

The main problem is to describe properties of the stationary state and their dependence on the local oscillator noise, family of states and the parameter \( \zeta \). The stationary state can be computed explicitly for a Gaussian solvable model (see Sec. 6). Such model is expected to describe a central limit and equalities that hold true in this case turns into inequalities in the general case. Let \( \sigma^2 \) be a variance of the stationary state \( y(t) \) immediately after the synchronization.

Then we derive a bound
\[ \sigma^2 \geq \frac{1 - \zeta}{1 + \zeta} \frac{1}{F_T} + \sigma^2_{LO}(T) \frac{g(\zeta)}{1 - \zeta^2}, \]
where $F_T$ is the Fisher information of the family $\rho_T(\varphi)$ and $\sigma^2_{LO}(T)$ is a simplified version of Allan variance a quantity that describes the local oscillator noise. Here $\zeta \in (-1, 1)$ and $g(\zeta)$ is bounded in this region and non-vanishing at $\zeta = \pm 1$.

The clock time associated to such a solution has a diffusive behavior and satisfies a bound

$$\lim_{t \to \infty} \frac{\mathbb{E}[(t_{\text{clock}} - t)^2]}{t} \geq T \frac{1}{F_T} + T \sigma^2_{LO}(T) \frac{\beta}{(1 - |\zeta|)^2},$$

where $\beta$ is a constant associated to the noise. Note that the clock time error consists of two terms. The first term is universal and describes the quantum projection noise. The second depends on the control scheme and the local oscillator noise and corresponded to the Dick effect. In this view, the inequality can be considered a rigorous justification of the Dick formula. These two terms are also the most relevant for the current experiments [23].

One of the punchlines of this formula is that even if we disregard the local oscillator noise we need to optimize an expression $T F_T^{-1}$. In contrast, the single step optimization would aim to optimize just $F_T^{-1}$. To analyze this formula further we look at a phenomenological dependence $\sigma^2_{LO}(T) \sim T^\alpha$ and a Fisher information scaling for unitary evolution $F_T \sim T^2 N^{1+\varepsilon}$, where $N$ is a number of atoms and $\varepsilon$ is 0 for separable states and 1 for entangled states. It is easy to see that the optimal interrogation time $T$ satisfies a formula

$$\frac{1}{F_T} \sim \sigma^2_{LO}(T),$$

which justifies the intuition that the dissipation and the information obtained from the synchronization should be proportional. For this optimal time the clock time variance scales with the number of atoms like

$$\lim_{t \to \infty} \frac{\mathbb{E}[(t_{\text{clock}} - t)^2]}{t} \sim \frac{1}{N^{\frac{1+\alpha}{1+\alpha}}}.$$

This type of behavior was predicted in [31, Appendix A].

A comprehensive reference for atomic clocks are books by Riehle [29] and Audoin, Guinot [2]: in particular the operation of the Cesium atomic clock that we sketched in the introduction is explained there in all details. Foundations of quantum estimation theory are described in a monograph by Helstrom [22] or in one by Holevo [24]. Dick effect was discovered and studied in [13, 14] and there is also a collection of articles in a special issue of IEEE transactions devoted to this topic. We single out an article by Greenhall [20] who studies a model very similar to ours. Other models of atomic clocks were studied in [1, 5, 6]. Atomic clocks from the perspective of estimation theory were studied in works [9, 28], however none of these works computes the clock time.

An accuracy of synchronization between a classical oscillator and a quantum reference was first studied by Itano et. al. [26] who coined the name quantum projection noise. That an entanglement can improve the precision was suggested...
by Bollinger et. al. [4] and a role of dephasing has a very clear exposition in a paper by Huelga et. al. [25]. Further results on quantum estimation in open systems can be found in [27, 16] and references there in. A concise general view on a quantum metrology can be found in Giovannetti et. al. [19]. Furthermore a repeated interaction model (e.g. [8]) was an additional motivation to the author, the model has certain similarities: An atomic clock can be viewed as a repeated interaction of a chain of quantum systems with a classical system.

The article is organized as follows. In a preliminary Section 2 we recall the basic theory of stochastic processes. In Sections 3 and 4 we describe the classical and the quantum estimation theories. In particular we derive a novel version of the Cramer-Rao bound that emphasizes the role of correlations between an unknown and its estimation. Our model of an atomic clock is fully described in Section 5, where we also derive the aforementioned bounds. In Section 6 we give an example where all bounds are saturated and in Section 7 we discuss the optimization of the clock’s performance. We close our exposition with outlooks in Section 8.

The focus of our work is on an exposition of the model and a study of the aforementioned fundamental properties. In particular we do not aim to proof our statements under minimal conditions.

**Assumption 1** We assume that all functions appearing in the text are continuously differentiable in an appropriate space and that all probability distributions have a finite second moment.

## 2 Stochastic processes

We would consider a probability distribution $p(\theta)$ of a single real parameter $\theta$ or a joint probability distribution $p(\theta, \theta')$ of two real parameters $\theta, \theta'$. The former is a reduced probability distribution of the latter, $p(\theta) = \int p(\theta, \theta')d\theta'$. Furthermore, associated to the latter there is a conditional probability distribution of a single parameter,

$$p(\theta' | \theta) := \frac{p(\theta, \theta')}{\int p(\theta, \theta')d\theta'},$$

describing the probability of $\theta'$ given $\theta$.

For a probability distribution $p(\theta)$ we denote by $\mu(p)$, $\sigma^2(p)$ its mean and variance respectively,

$$\mu(p) := \int \theta p(\theta)d\theta, \quad \sigma^2(p) := \int (\theta - \mu)^2 p(\theta)d\theta.$$

The mean of a joint probability distribution $p(\theta, \theta')$ is the vector of means and its variance is a matrix of mutual covariances.

Conversely (with a slight abuse of notation), we will often consider pairs of random real-valued variables $\theta, \theta'$ on a probability space $\Omega, d\mu$. This induces

---

1To simplify the notation we never spell out sigma-algebra explicitly. Those who care should be always able to fill it from the context.
a joint probability distribution \( p(\theta, \theta') \) that reproduces expectations,

\[
E[f(\theta, \theta')] = \int f(\theta, \theta') p(\theta, \theta') d\theta d\theta'.
\]

The random variable \( \theta \) by itself has a probability distribution \( p(\theta) \). If random variables are specified only by prescribing their joint probability distribution, then their usage would be independent of a realization (as a function on a certain probability space).

Crucial for estimation theory (and our work) is a notion of conditional expectation. A conditional expectation of \( \theta \) given \( \theta' \) is a real valued random variable \( E[\theta|\theta'] \) on a probability space \( \{\Omega, d\mu\} \) given by

\[
E[\theta|\theta'](x) = \int \theta p(\theta|\theta'(x)) d\theta, \quad x \in \Omega.
\]

Conditional expectation is a unique random variable measurable with respect to a sigma algebra generated by \( \theta' \) (i.e. such that it is constant on the sets where \( \theta' \) is constant) that reproduces expectations,

\[
E[f(\theta') E[\theta|\theta']] = E[f(\theta') \theta].
\]

The most useful instance of this formula is \( f(x) = 1 \), a conditional expectation \( E[\theta|\theta'] \) has the same expectation as \( \theta \),

\[
E[\theta] = E[E[\theta|\theta']].
\]

A space of real valued random variables has a natural associated scalar product \( (\theta, \theta') := E[\theta \theta'] \). Random variables of finite variance equipped with this scalar product form a real Hilbert space. We refer to this scalar product whenever we speak about orthogonality of two random variables.

A stochastic process is a collection of random variables; we will use both stochastic processes, \( X_t \), in continuous time \( t > 0 \) and discrete processes \( X_n, n \in \mathbb{N} \). The first naturally describes frequency dependence on time, the second is a suitable description of measurements occurring in discrete time steps. We would also encounter integrated processes,

\[
\int_0^t X_s ds, \sum_{j=0}^n X_j.
\]

These processes naturally occur as a relation between clock time and instantaneous frequency.

Below we consider only discrete processes in details. The corresponding concepts for processes in real time should be clear.

Of main interest will be the mean and the variance of instantaneous frequency and the variance of the associated clock time. More generally we will frequently use quadratic quantities associated to the process \( X_n \). In particular its mean \( E[X_n] \) and autocovariance

\[
C(X_{n+h}, X_n) = E[(X_{n+h} - E[X_{n+h}]) (X_n - E[X_n])].
\]
For \( h = 0 \) autocovariance reduces to a variance of the process at time \( n \). Quadratic quantities of an integrated process might be computed in terms of integrated covariance. For completeness we give an explicit formula,

\[
C\left(\sum_{j=0}^{n+h} X_j, \sum_{j=0}^{n} X_j\right) = \sum_{j=0}^{n} C(X_j, X_j) + 2 \sum_{j=0}^{n} \sum_{k=1}^{n-j} C(X_{j+k}, X_j) \\
+ \sum_{j=0}^{n} \sum_{k=n-j+1}^{n-h-j} C(X_{j+k}, X_j).
\]

A process \( X_n \) is called stationary if a joint distribution of \( X_{n+h}, \ldots X_{n+j} \) is independent of \( h \). In particular its mean, variance and autocovariance are independent of \( n \), we denote \( \gamma(h) := C(X_{n+h}, X_n) \) and \( \gamma(0) = \sigma^2 \). The ratio \( \zeta(h) := \gamma(h)/\sigma^2 \) is known as a correlation function. The formulas for integrated stationary process simplifies by one summation, e.g.:

\[
C\left(\sum_{j=0}^{n} X_j, \sum_{j=0}^{n} X_j\right) = (n+1)\sigma^2 + 2 \sum_{h=1}^{n} (n - h + 1)\gamma(h).
\]

The formula implies that for a stationary process \( X_n \) with zero mean, \( \mathbb{E}[X_n] = 0 \), we have

\[
\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[\sum_{j=0}^{n} X_j\right]^2 = \sigma^2 + 2 \sum_{h=1}^{\infty} \gamma(h),
\]

provided the sum on the RHS converges. In fact, under somewhat more strict conditions on \( \gamma(h) \) the central limit theorem gives convergence of \( n^{-1/2} \sum_{j=0}^{n} X_j \) to a Gaussian random variable of zero mean and variance given by the RHS of Eq. (3).

After an initial stage, a frequency source approaches a process that can be described as a mixture of a stationary process and a drift (also called aging). If the latter is negligible, frequency source is a stationary process. Decay of the correlation function \( \zeta(h) \) roughly describes stability of the source.

A process \( \{X_n\} \) is a martingale if \( \mathbb{E}[X_{n+1}|X_n] = X_n \) and it is Markov if the future depends on the past only through the present, \( \mathbb{E}[X_{n+1}|X_j, j \leq n] = \mathbb{E}[X_{n+1}|X_n] \). The Markov property can be equivalently stated that past and future are independent given the present. This is the first part of following lemma.

**Lemma 2** Suppose that \( \{X_1, X_2, X_3\} \) is a Markov chain, then it holds

\[
\mathbb{E}[X_1X_3|X_2] = \mathbb{E}[X_1|X_2]\mathbb{E}[X_3|X_2].
\]

Furthermore when \( \mathbb{E}[X_3|X_2] = \zeta X_2 \) for some \( \zeta \in \mathbb{R} \) then

\[
\mathbb{E}[X_1X_3] = \zeta \mathbb{E}[X_1X_2].
\]

\(^2\)In standard notation this would be denoted by \( \rho(h) \) however we shall need \( \rho \) to denote a quantum state.
Proof: The first equation is the equivalent definition of Markov property as mentioned in the text above the lemma, see [15, Chapter II.6]. We prove the second part,

\[ E[X_1X_3] = E[E[X_1|X_2]E[X_3|X_2]] = E[E[X_1|X_2]ξX_2] = ζE[X_1X_2]. \]

In the first and last equality we used Eq. (2). □

Let \( ω_t \) be a real valued stochastic process describing a frequency. Then the stability of the frequency source is often described in terms of the standard Allan variance [29, Chapter 3]

\[ σ^2(τ) = \frac{1}{2τ} E\left[ \left( \int_0^τ y(s)ds - \int_0^{2τ} y(s)ds \right)^2 \right], \]

where \( y(t) = (ω_t − E[ω_t])/ω_0 \) is the relative frequency error and \( τ \) is an averaging time. It is important to note that Allan variance is a function of the averaging time, not a single number. For our purpose Allan variance is unnecessary complicated and throughout the text we will use a simplified quantity that neglects correlations

\[ σ^2(τ) := \frac{1}{τ} E\left[ \left( \int_0^τ y(s)ds \right)^2 \right]. \quad (4) \]

We end this section with examples of various stochastic processes appearing in the following sections.

Example 3 (Standard diffusion) White noise is a stationary process, \( X_t \), of uncorrelated random variables. They have a constant mean \( µ \) and autocorrelation function

\[ γ(h) := C(X_{t+h}, X_t) = Dδ(h). \]

The integral of white noise, \( B_t = \int_0^t X_s \), is a Brownian motion. Its mean and variance are given by formulas

\[ E[B_t] = µt, \quad C(B_{t+h}, B_t) = 2Dt. \quad (5) \]

A drift \( µ \) and a diffusion coefficient \( D \) are constants whose physical dimension \([\cdot]\) depends on the process. More precisely, \([µ] = [X_s], \ [D] = [X_s]^2\).

Brownian motion \( B_t \) is a continuous martingale.

Example 4 (Gaussian random process) A discrete process \( X_n \) is called Gaussian if the joint probability distribution of \( X_{n_1}, X_{n_2}, \ldots, X_{n_j} \) is a multivariate normal distribution for any \( j \)-tuple \( n_1, \ldots, n_j \).

The Gaussian process is completely determined by the means \( µ(X_n) \) and covariances \( C(X_n, X_m) \). A particular property of interest (see [11], Chapter...
It is that for stationary Gaussian processes with zero mean and variance \( \sigma^2 \) it holds that
\[
E[X_{n+h}X_n] = \zeta^h \sigma^2,
\]
where \( \zeta = E[X_{n+1}X_n]/\sigma^2 \).

**Example 5 (Exponentially decaying correlations)** A stationary discrete stochastic process has exponentially decaying correlations if for some \( |\zeta| < 1 \),
\[
E[X_{n+h}X_n] = \zeta^h \sigma^2.
\]

The variance of the associated integrated process can be compute explicitly by summing a geometric series. Note that the result is consistent with Eq. (3).

\[
\mathbb{E} \left( \sum_{n=0}^{N} X_n \right)^2 = (N + 1)\sigma^2 + 2N\sigma^2 \frac{\zeta}{1 - \zeta} \left( 1 + \frac{1}{N} \frac{\zeta(N-1)}{1 - \zeta} \right) = N\sigma^2 \frac{1 + \zeta}{1 - \zeta} + O(1). 
\]

### 3 Estimation theory

An estimation theory studies strategies how to estimate an unknown physical parameter \( \varphi \) based on a data collected from a single or multiple measurements. In the classical estimation theory there is usually a one to one correspondence between ideal measurement and the unknown. The problem is then to decrease a measurement error using large data sets.

Our exposition of the estimation theory would be directed towards application in atomic clocks. A reader can find a general reference in e.g. [12], [3].

We examine strategies to estimate a parameter \( \varphi \in \mathbb{R} \) based on a measurement outcome \( \alpha \). We assume that the space of outcomes, \( M \), is a probability space equipped with a measure \( d\mu \). Measurement outcomes \( \alpha(\varphi) \) are a family of random variables with probability distributions \( p(\alpha|\varphi) \); a conditional probability distribution of an outcome \( \alpha \) given \( \varphi \). The estimation strategy is then defined by an estimator \( \Phi \). Upon a measurement outcome \( \alpha \) a guess \( \Phi(\alpha) \) is made. \( \Phi \) is a function from the space of outcomes to real numbers.

In a Bayesian approach to the estimation theory \( \varphi \) is a random variable with a certain prior probability distribution \( q(\varphi) \). It is then convenient to view its estimation, \( \hat{\varphi} \), as a random variable.

**Definition 6 (Estimation)** Let \( \varphi \) be a real valued random variable on a probability space \( (\Omega, d\mu) \), \( \alpha(\cdot) \) be a family of measurement outcomes and \( \Phi \) an estimator. Then an estimation \( \hat{\varphi} \) of \( \varphi \) is a random variable
\[
\hat{\varphi} := \Phi \circ \alpha(\varphi).
\]
In explicit terms, this is a random variable

\( \hat{\varphi} : (\Omega \oplus \mathcal{M}, d\mu(x) \otimes p(\alpha|\varphi(x))d\alpha) \to \mathbb{R} \)

given by

\[ \hat{\varphi}(x, \alpha) = \Phi(\alpha). \]

It is common to denote the estimator \( \Phi \) and the estimation \( \hat{\varphi} \) by the same letter. This is indeed convenient if \( \varphi \) is fixed with a given prior distribution. However we will consider estimations of a chain of random variables based on a fixed estimator \( \Phi \). For that reason we prefer to stress in our notation that \( \hat{\varphi} \) depends on the random variable that is estimated while \( \Phi \) is a fixed function.

Unbiased estimators play a central role in the estimation theory.

**Definition 7 (Unbiased estimation)** Fix a family \( \alpha(\cdot) \). We say that an estimator \( \Phi \) is \( \zeta \)-biased if for all random variables \( \varphi \)

\[ E[\hat{\varphi}|\varphi] = (1 - \zeta)\varphi. \]

The estimator is unbiased if it is 0-biased. We also say that an estimator is conditionally unbiased if

\[ E[\varphi] = 0 \implies E[\hat{\varphi}] = 0. \]

A \( \zeta \)-biased estimation (for \( \zeta \neq 0 \)) is not a common concept, in fact a \( \zeta \)-biased estimator is proportional to an unbiased estimator; however the parameter \( \zeta \) will play an important role in our description of an atomic clock. Note also that \( \{ \varphi, \hat{\varphi} + \zeta \varphi \} \) is a martingale, \( E[\hat{\varphi} + \zeta \varphi|\varphi] = \varphi \). In particular we will often use that for \( \zeta \)-biased estimation

\[ E[\varphi(\hat{\varphi} - (1 - \zeta)\varphi)] = 0. \] (7)

The \( \zeta \)-biased property of estimation can be equivalently stated by referring only to the conditional probability distribution \( p(\hat{\varphi}|\varphi) \). Consequently we often say that \( \hat{\varphi} \) is a \( \zeta \)-biased estimation of \( \varphi \), i.e. this is an estimation of the unknown \( \varphi \) based on a \( \zeta \)-biased estimator.

The following lemma summarizes various useful statements about unbiased estimators.

**Lemma 8** Fix a family \( \alpha(\cdot) \) of measurement outcomes. For an estimator \( \Phi \) the following is equivalent

(i) \( \Phi \) is conditionally unbiased,

(ii) there exists \( \zeta \in \mathbb{R} \) such that \( \Phi \) is \( \zeta \)-biased estimator.

Suppose in addition that \( \zeta \neq 1 \). Then a \( \zeta \)-biased estimator \( \Phi \) has the form \( \Phi = (1 - \zeta)\Phi_0 \) where \( \Phi_0 \) is an unbiased estimator.
Proof: (i) $\implies$ (ii): Let $p(\hat{\varphi}|\varphi)$ be a conditional probability distribution of $\hat{\varphi}$ given $\varphi$ and let $q(\varphi)$ be a probability distribution of $\varphi$. Then (i) states that for all distributions $q(\varphi)$ with zero mean it holds
\[ \int \hat{\varphi} p(\hat{\varphi}|\varphi) q(\varphi) d\hat{\varphi} d\varphi = 0. \]
A standard variational argument implies
\[ \int \hat{\varphi} p(\hat{\varphi}|\varphi) d\hat{\varphi} = \zeta \varphi \]
for some $\zeta \in \mathbb{R}$. This is exactly (ii).

(ii) $\implies$ (i): For a random variable $\varphi$ with zero mean and $\zeta$-biased estimation $\hat{\varphi}$ it holds
\[ \mathbb{E}[\hat{\varphi}] = \mathbb{E}[\mathbb{E}[\hat{\varphi}|\varphi]] = (1 - \zeta) \mathbb{E}[\varphi] = 0. \]

When $\Phi$ is a $\zeta$-biased estimator and $\zeta \neq 1$ then $(1 - \zeta)^{-1}\Phi$ is clearly an unbiased estimator.

Most of the work in estimation theory is centered on minimizing certain cost of $\varphi - \hat{\varphi}$ not hitting zero. We discuss this in the following section.

3.1 A cost of the estimation

A cost of the estimation (i.e. a functional we aim to minimize) is given by
\[ \text{Cost} = \mathbb{E}[(\varphi - \hat{\varphi})^2] = \int (\varphi - \Phi(\alpha))^2 p(\alpha|\varphi) q(\varphi) d\alpha d\varphi, \tag{8} \]
where $q(\varphi)$ is a prior probability distribution of $\varphi$. The choice of the cost function is to a large extent arbitrary. The quadratic cost function is distinguished by its simplicity and a direct relation to variance, the quantity that is most suitable for a description of time precision.

It is well known how to optimize the cost, Eq. (8), with respect to the estimator $\Phi$ for a fixed prior distribution of the variable $\varphi$.

Lemma 9 (Optimal estimator) Fix a conditional probability distribution $p(\alpha|\varphi)$ and a prior distribution $q(\varphi)$. Then an estimator
\[ \Phi(\alpha) = \mathbb{E}[\varphi|\alpha] \]
\[ = \int \varphi p(\varphi|\alpha) d\varphi \]
minimizes the cost (8) with respect to the estimator $\Phi(\cdot)$.

Proof: We use the formula $\mathbb{E}[Z(\alpha)Y] = \mathbb{E}[Z(\alpha)\mathbb{E}[Y|\alpha]]$ twice to rewrite the cost as
\[ \mathbb{E}[(\Phi(\alpha) - \varphi)^2] = \mathbb{E}[\Phi(\alpha)^2 - 2\Phi(\alpha)\varphi + \varphi^2] \]
\[ = \mathbb{E}[\Phi(\alpha)^2 - 2\Phi(\alpha)\mathbb{E}[\varphi|\alpha] + \varphi^2] \]
\[ = \mathbb{E}[(\Phi(\alpha) - \mathbb{E}[\varphi|\alpha])^2] + \mathbb{E}[(\mathbb{E}[\varphi|\alpha] - \varphi)^2]. \]
The last expression is a sum of two squares, where the second is independent of Φ. Hence the minimum is achieved when the first square vanishes.

The explicit expression, Eq. (9) is often hard to analyze. This is the case when the conditional probability distribution \( p(\alpha|\varphi) \) has an analytical expression, however there is no such expression for the conditional probability distribution \( p(\varphi|\alpha) \). In such cases bounds of the cost from below are very useful. Of such bounds the most famous is the Cramer-Rao bound, a variant of which we present here. It bounds the cost from below using Fisher information. This is a point-wise quantity that (roughly speaking) measures how fast is a conditional probability distribution changes with the value of the condition.

The Fisher information, \( F(\varphi) \), associated to a family of measurement outcomes \( \alpha(\cdot) \) (or probability distributions \( p(\alpha|\varphi) \)) is given by

\[
F(\varphi) := \int \left( \frac{\partial}{\partial \varphi} \log p(\alpha|\varphi) \right)^2 p(\alpha|\varphi) d\alpha.
\] (10)

An important property of the Fisher information is that it decreases by processing of the information. The Fisher information associated to the family \( \Phi \circ \alpha(\cdot) \) is always less than equal\(^3\) to the Fisher information associated to the family \( \alpha(\cdot) \). This fact allows to simplify the notation in the proofs below: We average over the conditional probability distribution \( p(\hat{\varphi}|\varphi) \) and keep in mind that the Fisher information of Eq. (10) satisfies

\[
F(\varphi) \geq \int \left( \frac{\partial}{\partial \varphi} \log p(\hat{\varphi}|\varphi) \right)^2 p(\hat{\varphi}|\varphi) d\hat{\varphi}.
\] (11)

The original Cramer-Rao bound (that we present in an integrated version) is the following statement.

**Proposition 10** Suppose that \( \hat{\varphi} \) is an unbiased estimation (i.e. an estimation based on an unbiased estimator) of a random variable \( \varphi \). Then

\[
\mathbb{E}[ (\varphi - \hat{\varphi})^2 ] \geq \mathbb{E} \left[ \frac{1}{F(\varphi)} \right].
\] (12)

**Proof:** For an unbiased estimation a conditional probability, \( p(\hat{\varphi}|\varphi) \), of \( \hat{\varphi} \) given \( \varphi \) satisfies

\[
\int \hat{\varphi} p(\hat{\varphi}|\varphi) d\hat{\varphi} = \varphi.
\]

We differentiate the expression, subtract zero and use the Cauchy-Schwarz inequality

\[
1 = \left( \int (\hat{\varphi} - \varphi) \frac{\partial}{\partial \varphi} p(\hat{\varphi}|\varphi) d\hat{\varphi} \right)^2 \\
\leq F(\varphi) \int (\hat{\varphi} - \varphi)^2 p(\hat{\varphi}|\varphi) d\hat{\varphi}.
\] (13)

\(^3\)In fact the equality holds if and only if \( \Phi \circ \alpha \) is a sufficient statistics for \( \varphi \).
Dividing by $F(\varphi)$ gives a pointwise version of the inequality, Eq. \[12\] can be then obtained by applying $E[\cdot]$ to both sides. □

An immediate corollary is a bound for $\zeta$-biased estimation.

**Corollary 11** Suppose that $\hat{\varphi}$ is a $\zeta$-biased estimation of a random variable $\varphi$. Then

$$E[(\varphi - \hat{\varphi})^2] \geq (1 - \zeta)^2 E[\frac{1}{F(\varphi)}] + \zeta^2 E[\varphi^2]. \quad (14)$$

**Proof:** For $\zeta \neq 1$ an estimation $\hat{\varphi}/(1 - \zeta)$ is unbiased and the statement follows from

$$E[(\varphi - \hat{\varphi})^2] = (1 - \zeta)^2 E[\frac{1}{F(\varphi)}] + \zeta^2 E[\varphi^2]$$

where the equality in the first line follows from orthogonality of $(1 - \zeta)\varphi - \hat{\varphi}$ and $\varphi$, see Eq. \[7\].

The case $\zeta = 1$ is somewhat special. In view of $E[\varphi \hat{\varphi}] = 0$ it then holds

$$E[(\varphi - \hat{\varphi})^2] = E[\varphi^2] + E[\hat{\varphi}^2]$$

and we see that the optimal estimation is $\hat{\varphi} = 0$. □

Van Trees [30] proved a Cramer-Rao type bound for an arbitrary estimator. We give a version of this bound, Eq. \[17\], that generalizes Eq. \[14\] and which to the best of our knowledge is new. It recognizes a role of correlations between $\varphi$ and $\hat{\varphi}$ in the Cramer-Rao inequality.

An extension of Cramer-Rao inequality beyond unbiased estimators comes at the expense of a less natural averaging of the Fisher information or introduction of additional terms. We choose the former approach because it has the simplest proof and gives the nicest formulas; A note on the other approach would be given elsewhere. For a given probability distribution $q(\varphi)$ we introduce an average Fisher information

$$\tilde{F} = \int F(\varphi) \frac{\tilde{q}(\varphi)^2}{q(\varphi)} d\varphi, \quad (15)$$

where

$$\tilde{q}(\varphi) = \frac{\int_{0}^{\infty} s \cdot \mu(q) q(s) ds}{\sigma(q)^2}. \quad (16)$$

For simplicity of the exposition we assume in the following theorem that $\varphi$ has zero mean. This is also the only case we will use in the article.

**Theorem 12** Fix a family of measurement outcomes $\alpha(\cdot)$ and let $\hat{\varphi}$ be an estimation of a random variable $\varphi$ (of zero mean) with a prior probability distribution $q(\varphi)$. Denote $\zeta := E[(\varphi - \hat{\varphi})\varphi]/E[\varphi^2]$. Then it holds

$$E[(\varphi - \hat{\varphi})^2] \geq (1 - \zeta)^2 \frac{1}{F} + \zeta^2 E[\varphi^2], \quad (17)$$

4 And completely unimportant.
where $\tilde{F}$ is the average Fisher information, Eq. (15), associated to the family $\alpha(\cdot)$.

**Proof:** The definition of $\zeta$ implies that variables $(1 - \zeta)\varphi - \hat{\varphi}$ and $\varphi$ are orthogonal with respect to a natural scalar product. This suggests (and proves) a decomposition

$$
\mathbb{E}[(\varphi - \hat{\varphi})^2] = \mathbb{E}[(\varphi - (1 - \zeta)\varphi)^2] + \zeta^2\mathbb{E}[\varphi^2].
$$

(18)

Now we bound the first term on the RHS. Using the definition of $\zeta$ once again we have

$$
\int \hat{\varphi}p(\hat{\varphi}|\varphi)q(\varphi)d\varphi = (1 - \zeta)\mathbb{E}[\varphi^2].
$$

It follows by integration by parts that for any $a \in \mathbb{R}$ (the term proportional to $a$ is point-wise zero)

$$
\int (\hat{\varphi} - a\varphi)\frac{\partial}{\partial \varphi}p(\hat{\varphi}|\varphi)\int_{\varphi}^{\infty} sq(s)ds d\varphi = (1 - \zeta)\mathbb{E}[\varphi^2].
$$

This further implies (note a definition of $\tilde{q}$, Eq. (16))

$$
(1 - \zeta)^2 = \left(\int (\hat{\varphi} - a\varphi)\frac{\partial}{\partial \varphi}p(\hat{\varphi}|\varphi)\tilde{q}(\varphi)d\varphi\right)^2
\leq \mathbb{E}[(\hat{\varphi} - a\varphi)^2] \int F(\varphi)\tilde{q}(\varphi)^2q(\varphi)d\varphi
\leq \mathbb{E}[(\hat{\varphi} - a\varphi)^2]\tilde{F}.
$$

Inserting this into Eq. (18) proves the sought inequality.

The inequality, Eq. (17), naturally bridges between the classical Cramer-Rao inequality for an unbiased estimator and a global Cramer-Rao inequality. To see this note that minimizing over $\zeta$ on the right hand side gives us a clone of Van Trees inequality (see [18]),

$$
\mathbb{E}[(\varphi - \hat{\varphi})^2] \geq \inf_{\zeta} \left((1 - \zeta)^2\frac{1}{\tilde{F}} + \zeta^2\mathbb{E}[\varphi^2]\right) = \frac{1}{\tilde{F} + 1/\mathbb{E}[\varphi^2]}.
$$

On the other hand $\zeta = 0$ reproduces Eq. (12) up to different averaging of the Fisher information.

The special averaging of Theorem 12 is very suitable for Gaussian prior distributions.

**Example 13** Suppose that the prior distribution $q(\varphi)$ of a random variable $\varphi$ is Gaussian, then $\tilde{F}$ is an average Fisher information with respect to the distribution $q$, i.e.

$$
\tilde{F} = \mathbb{E}[F(\varphi)].
$$

17
Note however that by the Jensen inequality
\[ E \left[ \frac{1}{F(\varphi)} \right] \geq \frac{1}{E[F(\varphi)]} \]
and so if an estimator is \( \zeta \)-biased the Cramer-Rao inequality \([14]\) gives better bound than \([17]\) even in this case. Inequalities coincide only if we further assume that the Fisher information \( F(\varphi) \) is constant.

Proof: One can directly verify that \( \tilde{q}(\varphi) \) of Eq. \((16)\) associated to a Gaussian distribution \( q(\varphi) \) satisfies \( \tilde{q}(\varphi) = q(\varphi) \). \( \square \)

4 Quantum estimation theory

In contrast to the classical estimation theory, quantum measurements cannot distinguish between non-orthogonal states even in the ideal situation of no external noise. This gives a fundamental bound on estimation precision which is referred to as the Heisenberg limit. Unlike the classical case, the probability distribution of this intrinsic quantum measurement error is described by the theory itself.

Throughout the text we fix a Hilbert space \( \mathcal{H} \) of, possibly infinite, dimension \( N \). A state, \( \rho \), on \( \mathcal{H} \) is a positive operator of unit trace. A pure state is represented by a one dimensional projection, which we mostly denote by \( P \).

A POVM measurement is defined by operators \( \Pi(\alpha) \geq 0 \) that decompose the identity, \( \int \Pi(\alpha)d\alpha = 1 \). The outcome \( \alpha \), given the state \( \rho \), is a random variable with a probability distribution \( \text{tr}(\rho \Pi(\alpha)) \) with respect to a (possibly singular) measure \( d\alpha \). As before, the space of outcomes \( \mathcal{M} \) is a probability space.

We examine strategies to estimate a parameter \( \varphi \in \mathbb{R} \) of a quantum state \( \rho_\varphi \equiv \rho(\varphi) \), whose dependence on the parameter \( \varphi \) is known. The estimation strategy is defined by a POVM measurement \( \Pi(\alpha) \) and an estimator \( \Phi(\alpha) \). A POVM measurement \( \Pi(\alpha) \) induces a conditional probability distribution of measurement outcomes \( p(\alpha|\varphi) = \text{tr}(\rho \Pi(\alpha)) \) and hence for every fixed POVM measurement we obtain a well posed classical estimation problem.

We call a couple \( \{ \Pi(\alpha), \Phi \} \) an estimation strategy. Given such a strategy and a real valued random variable \( \varphi \) we defined (Definition \(6\)) its estimation \( \hat{\varphi} \). All associated definitions generalize in a straightforward manner, e.g. an estimation strategy is \( \zeta \)-biased if for all random variables \( \varphi \) it holds that \( E[\hat{\varphi}|\varphi] = (1 - \zeta)\varphi \).

Now let \( \{ \Pi(\alpha), \hat{\varphi} \} \) be an estimation strategy. Then a conditional probability distribution function \( p(\hat{\varphi}|\varphi) \) of \( \hat{\varphi} \) conditioned upon \( \varphi \) is given by (we assume that \( |\nabla \Phi| > 0 \) and use a coarea formula)
\[ p(\hat{\varphi}|\varphi) = \int_{\Phi^{-1}(\hat{\varphi})} \text{tr}(\rho(\varphi)\Pi(x)) |\nabla \Phi(x)|^{-1} d\nu(x), \tag{19} \]
where \( d\nu \) is the induced measure by \( d\alpha \) on the manifold \( \Phi^{-1}(\hat{\varphi}) \). In particular
we see that a POVM measurement with outcomes \(\hat{\varphi} \in \mathbb{R}\) given by
\[
\tilde{\Pi}(\hat{\varphi}) = \int_{\Phi^{-1}(\hat{\varphi})} \Pi(x)|\nabla \Phi(x)|^{-1}d\nu(x)
\]
and an identity estimator function is equivalent to the original pair \(\{\Pi(\alpha), \Phi\}\).

The equivalence of these two pairs can be also explained in a down to earth language: The label \(\alpha\) of the measurement outcome is a superficial quantity and we can always re-parameterize it so that the measurement outcome is the estimation itself. In particular \(\text{tr}(\tilde{\Pi}(\hat{\varphi})\rho(\varphi))\) is a conditional probability of \(\hat{\varphi}\) given \(\varphi\).

For an exposition of the subject it is more convenient to use the original pair \(\{\Pi, \Phi\}\); however inside proofs we sometimes assume that without loss of generality the estimator is the identity function.

The cost of the estimation is now given by, cf. Eq. (8),
\[
\text{Cost} = \mathbb{E}[(\varphi - \hat{\varphi})^2] = \int (\varphi - \Phi(\alpha))^2 \text{tr}(\Pi(\alpha)\rho(\varphi)) q(\varphi)d\alpha d\varphi,
\]
where \(q(\varphi)\) is a prior distribution of a random variable \(\varphi\).

For a fixed POVM measurement \(\Pi(\alpha)\) Lemma 9 describes optimization of the cost with respect to the estimator \(\Phi(\cdot)\). It is natural to further try to optimize the cost with respect to the measurement POVM \(\Pi(\alpha)\). This optimization problem can be also solved algebraically (see [22, Chapter 8.1.(d)] and also [11]) and once again the explicit expression is hard to analyze. Therefore we discuss bounds from below.

The quantum Cramer-Rao bound is a generalization of the classical one. It bounds the cost from below using the (quantum) Fisher information, which is a point-wise quantity that (roughly speaking) measures how fast a family of states \(\rho(\varphi)\) is changing at a given point.

The Fisher information \(F(\varphi)\) is given by the expression
\[
F(\varphi) = \text{tr}(\rho(\varphi)X(\varphi)^2),
\]
where \(X(\varphi)\) is a solution\(^5\) of an equation
\[
\frac{1}{2}\{X(\varphi), \rho(\varphi)\} = \hat{\rho}(\varphi), \quad (\cdot = \frac{d}{d\varphi}).
\]

When \(\rho(\varphi) \equiv P(\varphi)\) is a family of projections then \(X(\varphi) = \hat{P}(\varphi)\) and Fisher information is proportional to the Fubini-Study metric, \(F(\varphi) = 2\text{tr}(\hat{P}(\varphi)^2)\). Also note that a Fisher information \(F_\tau(\varphi)\) associated to a family \(\rho(\tau\varphi), (\tau \in \mathbb{R})\), satisfies
\[
F_\tau(\varphi) = \tau^2 F(\tau \varphi).
\]

\(^5\)The equation does not determine \(QX(\varphi)Q\), where \(Q\) is the orthogonal projection on \(\text{Ker}(\rho(\varphi))\), and this part of \(X(\varphi)\) can be chosen arbitrary.
In parallel to the classical case we define $\tilde{F}$ with respect to a probability distribution $q$ by Eq. (15).

Braunstein and Caves [7] give a connection between the classical Fisher information associated to a fix measurement $\Pi(\alpha)$ and the quantum Fisher information.

**Proposition 14** Consider a family of states $\rho(\phi)$ and POVM measurements $\Pi(\alpha)$. Let $F(\phi)$ be the quantum Fisher information associated to the family $\rho(\phi)$ and let $F_{\Pi}(\phi)$ be the Fisher information, Eq. (10), associated to the conditional probability distribution $p(\alpha|\phi) = \text{tr}(\Pi(\alpha)\rho(\phi))$.

Then it holds that

$$F(\phi) = \sup_{\Pi} F_{\Pi}(\phi),$$

where the supremum is taken over all POVM measurements $\Pi(\alpha)$.

**Proof:** Let $X$ be a hermitian operator and $A, B$ non-negative operators, then the following inequality holds true

$$|\text{tr}(X\{A,B\})|^2 \leq 4\text{tr}(AB)\text{tr}(XBX).$$

Note that $\{A,B\}$ does not need to be a positive operator and so we cannot apply the Cauchy-Schwarz inequality on the LHS immediately. However we can do it after expanding the anti-commutator and in this way we get the RHS.

We use the inequality for $X = X(\phi)$, $A = \rho(\phi)$ and $B = \Pi(\alpha)$, (we omit the arguments of the operators)

$$|\text{tr}(\Pi\rho)|^2 = |\text{tr}(\Pi\{X,\rho\})|^2 = |\text{tr}(X\{\Pi,\rho\})|^2 \leq 4\text{tr}(\Pi\rho)\text{tr}(X\rho\Pi).$$

Hence we have the following estimate for the classical Fisher information,

$$F_{\Pi}(\phi) = \int \frac{(\text{tr}(\Pi(\alpha)\dot{\rho}(\phi))^2}{\text{tr}(\Pi(\alpha)\rho(\phi))} d\alpha \
\leq \int 4\text{tr}(X(\phi)\rho(\phi)X(\phi)\Pi(\alpha)) d\alpha = F(\phi),$$

the last expression being the quantum Fisher information. Equality can be achieved by taking $\Pi(\alpha)$ as a spectral decomposition of $X(\phi)$. \(\Box\).

All versions of a quantum Cramer-Rao bound are then immediate corollaries. We present one as an example, which is a compilation of bounds (14) and (17).

**Theorem 15** Consider a family of states $\rho(\phi)$ and let $F(\phi)$ be the associated quantum Fisher information. Let $\hat{\phi}$ be an estimation of a random variable $\phi$ (of zero mean) with prior distribution $q(\phi)$ and denote $\zeta := E[(\phi - \hat{\phi})^2]/E[\phi^2]$. Then it holds

$$E[(\phi - \hat{\phi})^2] \geq (1 - \zeta)^2 \frac{1}{F} + \zeta^2 E[\phi^2],$$

(21)
where \( \bar{F} \) is an average Fisher information, Eq. [15].

If furthermore \( \hat{\varphi} \) is an unbiased estimation then the term \( 1/\bar{F} \) in the inequality can be replaced by a simple average \( \mathbb{E}[1/F(\varphi)] \).

**Example 16 (Hamiltonian family)** Let \( P(\varphi) = e^{-i\varphi H} Pe^{i\varphi H} \) be a family of pure states generated by a Hamiltonian \( H \). Then the Fisher information is constant and proportional to the variance of the energy,

\[
F(\varphi) = 4 \left( \text{tr}(H^2 P) - \text{tr}^2(HP) \right).
\]

The Cramer-Rao inequality then takes a form of the Heisenberg uncertainty relation.

### 5 A model of atomic clocks

In our model the clock is fully described by a relative frequency error \( y(t) \). It is a real valued random variable and the purpose of this section is to define the process \( \{y(t)\}_{t \geq 0} \) and to discuss its basic properties. Clock’s frequency and the clock time are then determined through

\[
y(t) = \frac{\omega(t) - \omega_0}{\omega_0}, \quad t_{\text{clock}} - t = \int_0^t y(s) \, ds.
\]

The model consists of various parameters/objects that determine the process; we list them here:

- Time between two consecutive synchronizations \( T \). It is also equal to the interrogation time.
- A Markovian stochastic process \( K_t \varphi \) that describes evolution of the error in absence of synchronization given an initial condition \( K_0 \varphi = \varphi \).
- A family of states \( \rho_T(\varphi) \) and an estimation strategy \( \{\Pi(\alpha), \Phi\} \). These objects describe the synchronization.

The adjustment of error after the synchronization is performed periodically at times \( nT \), \( n \in \mathbb{N} \). We denote

\[
y_n(t) := y(nT + t), \quad \text{for} \quad t \in [0, T),
\]

\[
\bar{y}_n := \frac{1}{T} \int_0^T y_n(s) \, ds. \quad (22)
\]

We also abbreviate \( y_n := y_n(0) = y(nT) \). The stochastic process \( y(t) \) is defined implicitly by an initial condition \( y_0 \) through equations

\[
y_n(t) = K_t y_n \quad \text{for} \quad t \in [0, T), \quad (23)
y_{n+1} = K_T y_n - \hat{\bar{y}}_n. \quad (24)
\]

The random variable \( \hat{\bar{y}}_n \) appearing in the last line is an estimation of \( \bar{y}_n \). The estimation is obtained by a measurement \( \Pi(\alpha) \) on a state \( \rho_T(\bar{y}_n) \).
Definition 17 We call a triple \( \{ \rho_T(\varphi), \Pi(\alpha), \Phi \} \) an atomic clock. The local oscillator noise \( K_t \) will be clear from the context. A solution \( y(t) \) of Eqs. (23), (24) is called a state of an atomic clock.

Eq. (23) describes the evolution in absence of synchronization, Eq. (24) describes jumps due to the synchronization and the corresponding adjustment of frequency. The latter equation defines a (sub)process \( y_n \). This is a Markovian process that encodes the synchronization and hence has a distinguished role.

Definition 18 (Stationary state of a clock) We say that \( y(t) \) describes a stationary state of an atomic clock if \( y_n \) is a stationary process.

A stationary state, \( y(t) \), of an atomic clock is \( T \) periodic, meaning that the joint probability distributions of \( y(t_1), \ldots, y(t_n) \) and \( y(t_1 + T), \ldots, y(t_n + T) \) are identical. This in particular implies that the averaged error \( \bar{y}_n \), Eq. (22), is then a stationary process.

We aim to study a situation when a clock time is unbiased, \( \mathbb{E}[t_{\text{clock}}] = t \). This is true if and only if the relative frequency error has a zero average. Consequently we say that a clock has an unbiased stationary state \( y(t) \) if \( \mathbb{E}[y(t)] = 0 \) for all \( t \geq 0 \).

Whether a given clock has an unbiased stationary state is not a robust statement. It is sensitive to the noise \( K_t \) and to the choice of estimation strategy. A natural question is under which conditions on \( K_t \) and \( \rho(\varphi) \) we can find an estimation strategy \( \{ \Pi, \Phi \} \) such that the clock has an unbiased stationary state. We do not know any general answer to that question and rather choose to assume more about the clock.

First, that the frequency error remains unbiased provided the initial error \( \varphi_0 \) is unbiased, i.e. \( \mathbb{E}[\varphi_0] = 0 \). In Lemma 8 we proved that this can happen only if the estimation strategy is \( \zeta \)-biased for some \( \zeta \in \mathbb{R} \). This guarantees that the subspace of unbiased random variables is an invariant subspace of the clock. To assure that it is also attractive we further require that \( |\zeta| < 1 \). This is a stability condition for the feedback loop.

Definition 19 (Unbiased clock) We say that a clock is unbiased if \( K_t \varphi \) is a martingale and the estimation strategy \( \{ \Pi, \Phi \} \) is \( \zeta \)-biased (with respect to the family \( \rho(\tau \varphi) \)) with \( |\zeta| < 1 \). If a value of \( \zeta \) is given we say that a clock is \( \zeta \)-unbiased.

We will see that a stochastic process \( y_n \) associated to an unbiased clock is a supermartingale. It then follows that an unbiased clock has a an unbiased stationary state.

It pays to examine a clock without noise. We do that in the following section. Afterwards we study general properties of a stationary state of an unbiased clock.
5.1 A clock without noise

It is rather surprising that an important feature of clock operation can be demonstrated in a trivial case $K_t = 1$. The stochastic process $y(t)$ simplifies significantly. The relative frequency error $y(t)$ is constant in the intervals $(nT, (n+1)T)$ and jumps on its boundary. Its value inside the interval was denoted by $y_n$ and the jump at the right side of the interval is $\delta y_n = \hat{y}_n$. Eq. (24) takes a form

$$y_{n+1} = y_n - \hat{y}_n.$$

The clock time associated to a state $\phi(t)$ is given by

$$t_{\text{clock}} - t = \int_0^t y(s) \, ds = T \sum_{n=0} y_n. \tag{25}$$

We claim that the variance of the clock time has a universal bound, although the variance of the frequency error can be arbitrary small.

**Theorem 20 (Unbiased clock without noise)** Suppose that $K_t = 1$ and that $C_\zeta = \{\rho_T(\varphi), \Pi(\varphi), \Phi_\zeta\}$ is a family of unbiased atomic clocks; $\Phi_\zeta = (1 - \zeta)\Phi_0$ being a $\zeta$-biased estimator. Let $y_\zeta(t)$ be a stationary state of the clock $C_\zeta$, then

$$E[y_\zeta^2] \geq E[\frac{1}{F_T(y_\zeta)}]\frac{1 - \zeta}{1 + \zeta}. \tag{26}$$

The variance of clock time associated to $y_\zeta(t)$ satisfies a $\zeta$ independent bound,

$$\lim_{t \to \infty} E[\frac{(t_{\text{clock}} - t)^2}{t}] \geq T E[\frac{1}{F_T(y_\zeta)}], \tag{27}$$

where $F_T(\varphi)$ is the Fisher information associated to the family $\rho_T(\varphi)$.

**Proof:** Fix $\zeta$ and denote $y_n := y_\zeta(Tn), n \in \mathbb{N}$. Then $y_n$ is a stationary process with zero mean and variance $\sigma^2 := E[y_n^2]$. The Cramer-Rao inequality, Eq. (14), then implies

$$\sigma^2 \geq (1 - \zeta)^2 E[\frac{1}{F_T(y_n)}] + \zeta^2 \sigma^2.$$

The inequality (26) follows by solving for $\sigma^2$.

We claim that $y_n$ is a Markov chain with exponentially decaying correlations

$$E[y_{n+h}y_n] = \zeta^h \sigma^2, \quad \zeta \geq 0.$$  

Then according to Example 5 the variance of the clock time satisfies

$$\lim_{t \to \infty} \frac{E[(t_{\text{clock}} - t)^2]}{t} = \sigma^2 \frac{1 + \zeta}{1 - \zeta}.$$
Plugging in inequality (26) one obtains the bound (27).

Exponential decay of correlations follows from the unbiasedness condition,
\[ \mathbb{E}[y_{n+h} | y_{n+h-1}] = \zeta y_{n+h-1}, \]
which by Lemma 2 implies that for \( h \geq 1 \),
\[ \mathbb{E}[y_{n+h} y_n] = \zeta \mathbb{E}[y_{n+h-1} y_n]. \]
\[ \Box \]

In Section 6 we will see an example where all bounds in the theorem are achieved. The moral to be taken is that there is a 1-parameter family of clocks whose stationary states differ in autocorrelations, however giving an equally good clock time.

We believe that the bound (27) should be valid without assuming that the clock is unbiased. Instead only a certain ergodicity assumption to prevent a trivial counterexample of no synchronization should be made. Let say an assumption that the transition map of the Markov process \( y_n \) has 1 as an eigenvalue isolated by a gap from the rest of its spectrum. We were not able to find a proof of such general statement.

To be more concrete, let \( y_n \) be an unbiased stationary state and denote \( \zeta := \mathbb{E}[y_{n+1} y_n] / \mathbb{E}[y_n^2] \). Then the inequality (26) holds true in view of Eq. (21), but we cannot anymore conclude (27) without additional assumptions. Eq. (27) holds if we (for example) assume a detailed balance of \( y_n \). We do not see any physical motivation for such assumption, however we include the proof in order to support the above conjecture.

In the reminder of the Section we fix an atomic clock \( \{ \rho_T(\varphi), \Pi(\varphi), \Phi \} \) and study the corresponding Markov process
\[ y_{n+1} = y_n - \hat{y}_n. \]  
(28)
We denote by \( A(y, y') \) the associated transfer matrix, i.e. \( A(y, y') = \text{Prob}(y_{n+1} = y | y_n = y') \). A probability distribution \( q \) of a stationary solution of the Eq. (28) then satisfies \( Aq = q \). Equivalently, when \( q \) is a solution of \( Aq = q \) and \( y_0 \) a random variable with that probability distribution, then \( y_n := (A^*)^n y_0 \) is a stationary solution of the Eq. (28). By \( A^* \) we denote the adjoint of \( A \) corresponding to a duality between probability distributions and random variables.

**Proposition 21** Let \( A(y, y') \) be a transfer matrix associated to the Eq. (28) and assume that \( A \) is reversible with respect to a (stationary) probability distribution \( q \) of zero mean. Let \( y_n \) be the associated stationary state. Assume moreover that 1 is a simple eigenvalue of \( A \) that is isolated from the rest of the spectra, i.e. \( \sigma(A) \setminus \{ 1 \} \subset B_R \) for some \( R < 1 \). Then the associated clock time satisfies a bound
\[ \lim_{t \to \infty} \frac{\mathbb{E}[(t_{\text{clock}} - t)^2]}{t} \geq T \frac{1}{\tilde{F}_T}, \]
where \( \tilde{F}_T \) is the averaged Fisher information of \( \rho(\varphi) \) with respect to a probability distribution \( q \), see Eq. (17).

---

A process \( \varphi_{n+1} = \varphi_n \) with initial conditions \( \varphi_0 = 0 \).
Proof: For two real valued random variables $X$, $Y$ we define a scalar product $(X, Y) = \int X(y)Y(y)q(y)dy$ and we denote the associated norm by $\| \cdot \|$. In terms of this product covariances of the stationary process $y_n$ are given by

$$E[y_{n+h}y_n] = ((A^*)^h y, y)$$

where $y$ is an identity function. Denote $\zeta := (A^* y, y)/\|y\|^2$ then the Cramer-Rao inequality, Eq. [21], implies (like in Theorem 20)

$$E[y_n^2] \geq \frac{1}{F_T} \frac{1 - \zeta}{1 + \zeta}. \quad \text{(29)}$$

The clock time variance can be expressed in terms of $A^*$ as

$$\lim_{t \to \infty} \frac{(t - t_{\text{clock}})^2}{t} = T \left( \|y\|^2 + 2 \sum_{h=1}^{\infty} ((A^*)^h y, y) \right)$$

$$= T \left( 2 \left( \frac{1}{1 - A^*} y, y \right) - \|y\|^2 \right),$$

where the summability is guaranteed by $y \in \text{Ran}(A^* - 1)$ and our spectral assumptions. Reversibility means that $A^*$ is hermitian, in particular we have the following Cauchy-Schwartz inequality

$$\|y\|^4 \leq \left( \frac{1}{1 - A^*} y, y \right) ((1 - A^*) y, y)$$

$$\leq \left( \frac{1}{1 - A} y, y \right) (1 - \zeta) \|y\|^2.$$

Plugging this inequality into the expression for the clock time and using Cramer-Rao inequality [29] leads to the inequality claimed in the proposition. \(\square\)

5.2 An unbiased clock

In this section we prove that the stochastic process $y_n$ associated to an unbiased clock is a supermartingale. It would then follow by Doob’s martingale convergence theorem that an unbiased clock has an unbiased stationary state. We finish the section by classifying these states and describing their properties.

Throughout the section we explicitly compute several quantities related to a state of an $\zeta$-biased clock. These computations are possible due to two relations

$$E[y(t)\tilde{y}_n] = (1 - \zeta)E[y(t)\bar{y}_n], \quad t \leq nT, \quad \text{(30)}$$

$$E[y_n(t)y_n(s)] = E[y_n(t)^2], \quad t \leq s. \quad \text{(31)}$$

The first equality follows from Lemma [2] applied to a random variables triple $\{y(t), \tilde{y}_n, \bar{y}_n\}$. The second is an immediate consequence of $y_n(t)$ being a martingale.
Furthermore to describe local oscillator noise we use a simplified Allan variance, Eq. (4),
\[ \sigma_{LO}^2(T) := \mathbb{E}[(\bar{y}_n - y_n)^2]. \]
We relate all other local oscillator quantities to \( \sigma_{LO}^2(T) \) with a help of noise
dependent constants \( \alpha, \beta \). These constants are defined on an appropriate place
below.

**Proposition 22** Let \( y(t) \) be a state of an unbiased clock then \( y_n \) is a super-
martingale.

**Proof:** We fix \( \zeta \) and assume that the clock is \( \zeta \)-unbiased. Since \( y_n \) is a Markov
process we need to prove that \( \mathbb{E}[y_{n+1}|y_n] \leq y_n \). The LHS can be computed
explicitly by observing that a triple of random variables \( \{y_n, \bar{y}_n, \hat{\bar{y}}_n\} \) has a
Markov property, so \( \mathbb{E}[\hat{\bar{y}}_n|y_n] = \mathbb{E}[\mathbb{E}[\hat{\bar{y}}_n|\bar{y}_n]|y_n]. \) It then holds
\[
\mathbb{E}[y_{n+1}|y_n] = \mathbb{E}[y_n(T) - \hat{\bar{y}}_n|y_n]
= y_n - (1 - \zeta)\mathbb{E}[\bar{y}_n|y_n]
= \zeta y_n.
\]
The stability condition \( |\zeta| < 1 \) implies the statement. \( \Box \).

For the next theorem we would need a technical assumption. We say that
the estimation is \( a \)-bounded if there exists a constant \( C \) such that
\[ \mathbb{E}[(\varphi - \hat{\varphi})^2] \leq C + a\mathbb{E}[\varphi^2]. \]

**Theorem 23** Let \( \{\rho_T(\varphi), \Pi(\alpha), \Phi\} \) be an unbiased clock and suppose that the
estimation is \( a \)-bounded with \( a < 1 \). Assume also that the increments of \( K_t \)
have bounded variation, \( \mathbb{E}[(K_t y - y)^2] < \infty, \) for \( t \in [0, T] \). Then the clock has
an unbiased stationary state.

**Proof:** Let \( y(t) \) be a state of the \( (\zeta \)-unbiased) clock. By Proposition 22 \( y_n \)
is a supermartingale. The existence of the stationary state follows form Doob’s
martingale convergence theorem provided we can bound the variance of \( y_n \). By
a straightforward computation using (30) and (31) we have
\[
\mathbb{E}[y_{n+1}^2] = \mathbb{E}[(\bar{y}_n - \hat{\bar{y}}_n)^2] + \mathbb{E}[(y_n(T) - \bar{y}_n)(y_n(T) - (1 - 2\zeta)\bar{y}_n)].
\]
By the assumptions this implies
\[ \mathbb{E}[y_{n+1}^2] \leq \text{Const} + a\mathbb{E}[y_n^2]. \]
It then follows that \( \mathbb{E}[y_n^2] \leq \text{Const}/(1 - a) \) for \( n \) large enough.

That the stationary state is unbiased follows from an equation \( \mathbb{E}[y_{n+1}] = \zeta \mathbb{E}[y_n]. \) \( \Box \).

Next we present a version of Dick formula. This formula is traditionally
derived and discussed in a frequency domain, however for our purposes time
domain is more natural.

26
Proposition 24 Let \( y(t) \) be a stationary state of a \( \zeta \)-unbiased clock and denote \( \sigma^2 = \mathbb{E}[y_n^2] \). Then for the clock time it holds
\[
\lim_{t \to \infty} \frac{(t_{\text{clock}} - t)^2}{t} = T \left( \frac{\sigma^2 1 + \zeta}{1 - \zeta} + \sigma_{\text{LO}}^2(T) \frac{1 + \alpha + \zeta}{1 - \zeta} \right),
\]
where \( \alpha \) is defined through an equation
\[
\frac{1}{T} \int_0^T \mathbb{E}[(y_n(s) - y_n)^2] = \sigma_{\text{LO}}^2(T) \frac{\alpha + 2}{2}.
\]

Remark 25 The parameter \( \alpha \) is defined so that it would be consistent with an additive local oscillator noise whose Allan variance \( \sigma_{\text{LO}}^2(T) \sim T^\alpha \).

Proof of the proposition: According to Eq. (3) we have
\[
\lim_{t \to \infty} \frac{(t_{\text{clock}} - t)^2}{t} = T \left( \mathbb{E}[\hat{y}_n^2] + 2 \sum_{h=1}^{\infty} \mathbb{E}[\hat{y}_{n+h}\bar{y}_n] \right).
\]
We can express the quantities on the RHS using Eqs. (30), (31). For the first term we have
\[
\mathbb{E}[\hat{y}_n^2] = \mathbb{E}[(\bar{y}_n - y_n)^2] + \mathbb{E}[y_n^2] = \sigma_{\text{LO}}^2(T) + \sigma^2.
\]
In the second term the correlations decay exponentially, i.e. for \( h > 1 \) it holds
\[
\mathbb{E}[\hat{y}_{n+h}\bar{y}_n] = \zeta \mathbb{E}[\hat{y}_{n+h-1}\bar{y}_n].
\]
And the \( h = 1 \) term can be expressed as
\[
\mathbb{E}[\hat{y}_{n+1}\bar{y}_n] = \mathbb{E}[(y_n(T) - \hat{y}_n)\bar{y}_n] = \frac{1}{T} \int_0^T \mathbb{E}[y_n^2(s)]ds - (1 - \zeta)\mathbb{E}[y_n^2] = \zeta \sigma^2 + \frac{\alpha + 2}{2} \sigma_{\text{LO}}^2(T) - (1 - \zeta)\sigma_{\text{LO}}^2(T).
\]
After summing the geometric series and adding all the terms one gets Eq. (32).
\[\square\]

We are ready to prove our main theorem that describes stationary states of unbiased clocks.

Theorem 26 Suppose that \( C_\zeta = \{\rho_T(\varphi), \Pi(\alpha), \Phi_\zeta\} \) is an \( \zeta \)-unbiased clock and let \( F_T(\varphi) \) be a Fisher information associated to the family \( \rho_T(\varphi) \). Let \( y(t) \) be a stationary state of the clock, then its variance \( \sigma^2 = \mathbb{E}[y_n^2] \) satisfies an inequality
\[
\sigma^2 \geq \frac{1}{F_T} \frac{1 - \zeta}{1 + \zeta} + \sigma_{\text{LO}}^2(T) \frac{\zeta^2 + \alpha \zeta + \beta - 1 - \alpha}{1 - \zeta^2},
\]
(33)
and for the associated clock time variance we have a bound
\[
\lim_{t \to \infty} \frac{E[(t_{\text{clock}} - t)^2]}{t} \geq T \frac{1}{F_T} + T\sigma^2_{LO}(T) \frac{\beta}{(1 - \zeta)^2},
\] (34)

where \(\alpha\) was defined in the Proposition 24 and
\[
\beta\sigma^2_{LO}(T) = E[(K_T y_n - y_n)^2].
\]

Above \(1/F_T\) is a shorthand for \(E[1/F_T(\bar{y}_n)]\).

**Proof:** We follow proof of Theorem 20 (the case \(K_t = 1\)) only the details are more involved.

We have
\[
E[y^2_{n+1}] = E[(y_n(T) - \hat{y}_n)^2]
\]
\[
= E[(\bar{y}_n - \hat{y}_n)^2] + E[(y_n(T) - \bar{y}_n)(y_n(T) - (1 - 2\zeta)\bar{y}_n)]
\]
\[
= E[(\bar{y}_n - \hat{y}_n)^2] + \sigma^2_{LO}(T)(-1 - \alpha + \beta + \zeta\alpha).
\]

Using the Cramer-Rao inequality (14) on the RHS we then have
\[
\sigma^2 \geq \frac{(1 - \zeta)^2}{F_T} + \zeta^2(\sigma^2_{LO}(T) + \sigma^2) + \sigma^2_{LO}(T)(-1 - \alpha + \beta + \zeta\alpha).
\]

The first inequality of the Theorem follows by solving for \(\sigma^2\). The inequality for the clock time follows by substituting Eq. (33) into the formula for the clock time Eq. 32. \(\square\)

**Example 27 (Additive noise)** For a square integrable function \(f(s)\) on the interval \([0, T]\) a stochastic process \(K_t \varphi = \varphi + \int_0^t f(s)dW_s\) satisfies all the requirements of the above section. In particular a power law ansatz for \(f(s)\) gives \(\sigma^2_{LO}(T) = DT^\alpha\), where \(D\) is a constant and \(\alpha\) is a parameter consistent with that appearing in Proposition 24, i.e.
\[
\frac{1}{T} \int_0^T E[(K_s\varphi - \varphi)^2] = \sigma^2_{LO}(T) \frac{\alpha + 2}{2}.
\]

Parameter \(\beta\) of Theorem 26 is then given by
\[
\beta = \frac{1}{2}(\alpha + 2)(\alpha + 1).
\]

6 Gaussian families

Here we aim to illustrate our concepts on a simple solvable example. To easy the notation we put \(T = 1\). We also assume that the local oscillator noise is a Brownian motion, \(K_t \varphi = \varphi + DW_t\) (generalization to any noise of the type of Example 27 is straightforward).
Let \( \rho(\varphi) = |\psi(\varphi)\rangle \langle \psi(\varphi)| \) be a family of Gauss states on a real line,

\[
\psi_\varphi(x) = \langle x|\psi(\varphi)\rangle = \frac{F^{1/4}}{(2\pi)^{1/4}} \exp \left( -\frac{F}{4} (x - \varphi)^2 \right).
\]

Our notation highlights the Fisher information. It is a matter of simple computation to find that the Fisher information, \( F(\varphi) \), of \( \rho(\varphi) \) is indeed constant and equal to \( F \).

We consider an estimation strategy \( \{\Pi(\alpha), \Phi_\zeta\} \), where \( \Pi(\alpha) \) is an orthogonal decomposition of a position operator, \( X \), on a line,

\[
X = \int \alpha \Pi(\alpha) d\alpha, \quad \Pi(\alpha) = \delta(x - \alpha)
\]

and \( \Phi_\zeta(\alpha) = (1 - \zeta)\alpha \). The conditional probability distribution of an estimate \( \hat{\varphi} \) given parameter \( \varphi \) is then, c.f. Eq. (19),

\[
p(\hat{\varphi}|\varphi) = \frac{1}{1 - \zeta} \text{tr} \left( \Pi \left( \frac{1 - \zeta}{1 - \zeta} \hat{\varphi} \right) \rho(\varphi) \right) = \frac{1}{\sqrt{2\pi(1 - \zeta)}} \exp \left( -\frac{1}{2} \frac{F}{(1 - \zeta)^2} \left( \hat{\varphi} - (1 - \zeta)\varphi \right)^2 \right).
\]

We see that \( p(\hat{\varphi}|\varphi) \) is a Gaussian kernel. For \( \zeta = 0 \) it is a symmetric heat kernel and hence unbiased. In general the estimator is multiple of unbiased estimator and the estimation strategy \( \{\Pi, \Phi_\zeta\} \) is \( \zeta \)-biased. This can be also checked by direct integration of \( \hat{\varphi} \) with respect to the kernel.

Now we consider a clock \( \{\rho(\varphi), \Pi(\alpha), \Phi_\zeta\} \) and its state \( y(t) \). We claim that for such a clock the Markov chain \( y_n \) is Gaussian. We show this directly by computing the transition map \( T(y, y') = p(y_{n+1} = y|y_n = y') \).

This map can be computed by considering a joint probability distribution \( p(\bar{y}_n, y_n(T)|y_n) \). It is a binomial Gaussian distribution with mean \( \mu = (y_n, y_n) \) and a covariance matrix independent of \( y_n \), whose elements might be computed in a standard way, for example using Eqs. (30), (31). The transition map is then given by (recall that \( y_{n+1} = y_n(T) - \bar{y}_n \))

\[
T(y, y') = \int p(\bar{y}_n = x - y, y_n(T) = x|y_n = y') dx
\]

\[
= \int p(\bar{y}_n = x - y|\bar{y}_n = z)p(y_n = z, y_n(T) = x|y_n = y') dx dz.
\]

An integral of Gaussian kernels is itself a Gaussian, proving the claim that \( y_n \) is a Gaussian process.

The transition map can also be computed explicitly. One can either compute the involved Gaussian integrals or read the outcome from the computation in the proof of Theorem 26. Either way one arrives at

\[
T(y, y') = \frac{1}{\sqrt{2\pi s}} \exp \left( -\frac{1}{2s^2} (y - \zeta y')^2 \right),
\]

\[
s^2 = \frac{(1 - \zeta)^2}{F} + \zeta^2 2D + \frac{2}{3} D(1 + \zeta - 2\zeta^2).
\]
Gaussian states are determined by their mean, $\mu$, and variance, $\sigma^2$. If we represent them by a column vector $(\mu, \sigma^2)^T$ then $T$ is an affine operation

$$T\left(\begin{array}{c} \mu \\ \sigma^2 \end{array}\right) = \left(\begin{array}{c} \zeta \mu \\ \zeta^2 \sigma^2 + s^2 \end{array}\right).$$

It is then easy to determine a stationary Gaussian distribution, it has zero mean and a variance satisfying equation $\sigma^2 = \zeta^2 \sigma^2 + s^2$. This gives

$$\sigma^2 = \frac{1 - \zeta}{1 + \zeta} \frac{1}{F} + \frac{2}{3(1 - \zeta^2)} D(1 + \zeta + \zeta^2).$$

This is exactly the RHS of the bound $\text{(33)}$. Saturating this bound it also saturates the bound for the clock time. We summarize (to compare with Theorem $\text{26}$ put $\alpha = 1, \beta = 3$):

**Theorem 28** Let $C_\zeta = (\rho(\varphi), \Pi, \Phi_\zeta)$ be a Gaussian clock described above. Then $C_\zeta$ possesses a Gaussian stationary state $y_\zeta(t)$ with variance given by

$$\sigma^2 = \frac{1 - \zeta}{1 + \zeta} \frac{1}{F} + \sigma^2_{LO}(T) \frac{1 + \zeta + \zeta^2}{1 - \zeta^2}.$$ 

The associated clock time has a standard diffusive behavior

$$\lim_{t \to \infty} \frac{\mathbb{E}[(t_{\text{clock}} - t)^2]}{t} = T \frac{1}{F} + T \sigma^2_{LO}(T) \frac{3}{(1 - \zeta)^2}.$$ 

**Remark 29** There is a reason for saturation of bounds: In the Gaussian case the Cramer-Rao bound, Eq. $\text{(14)}$, is saturated, because condition for equality in Cauchy-Schwarz in Eq. $\text{(13)}$ is met. In fact this proves Theorem $\text{28}$ without any computation, however we believe that the explicit computations that were presented in this section complement a rather abstract approach of previous sections.

### 7 Optimization of an atomic clock operation

A practical reason to study atomic clocks is an optimization of their performance. We have seen in the last section that bounds $\text{(33)}, \text{(34)}$ are attainable and now we proceed to optimize the error of clock time as it appears in the bounds. To proceed with such optimization we need to know how $F_T$ and $\sigma^2_{LO}(T)$ depends on the interrogation time. We choose to study a Hamiltonian evolution (see Example $\text{16}$),

$$F_T = 4T^2 \frac{\Delta^2 E}{\omega_0^2},$$

where $\Delta^2 E$ is the variance of energy. And we assume a phenomenological ansatz for the local oscillator noise, $\sigma^2_{LO}(T) = DT^\alpha$. 

30
With respect to the parameter $\zeta$ the clock time variance does not possess a minimizer. It is formally minimized by $\zeta = -1$, however there is no associated stationary solution as can be seen from the bound on the variance of the stationary state that has a blow up at this value of $\zeta$. This implies that the value of $\zeta$ needs to be chosen independently, for example by considering mixing times of the clock. In the following we fix value of $\zeta$ and minimize with respect to the interrogation time $T$.

Minimizing the RHS of Eq. (34) is then straightforward and we find that the minimum satisfies an equation

$$
\frac{1}{F_T} = (\alpha + 1)\sigma_{LO}^2(T) \frac{\beta}{(1 - \zeta)^2}
$$

and the corresponding clock time variance is given by

$$
\lim_{t \to \infty} \frac{\mathbb{E}[(t_{\text{clock}} - t)^2]}{t} = 2 \left( \frac{\omega_0^2}{4\Delta^2 E} \right)^{\frac{\alpha + 1}{\alpha + 2}} \left( \frac{\beta D}{(1 - \zeta)^2} \right)^{\frac{1}{\alpha + 2}}.
$$

It is interesting to apply this formula to the experimentally relevant case of system of $N$ spins and a Hamiltonian

$$
H = \sigma_z^{(1)} + \sigma_z^{(2)} + \cdots + \sigma_z^{(N)},
$$

although this is out of our framework (see the next Section for the discussion of this important issue). The energy variance depends on the initial state, for a separable initial state it scales like $N$, while for an GHZ state

$$
|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle\cdots|0\rangle + |1\rangle|1\rangle\cdots|1\rangle), \quad (\text{GHZ}|H^2|\text{GHZ}\rangle = N^2.
$$

It follows that the time clock variance scales like $N^{(1+\epsilon)\frac{\alpha + 1}{\alpha + 2}}$, where $0 \leq \epsilon \leq 1$ is a parameter that depends on the initial state.

8 Outlooks

Here we want to outline further problems that are in our opinion both important and attractive.

Large system limit

Consider a family of states of $N$ spins, e.g.

$$
\rho(\varphi) = e^{-iT\varphi H}|\text{N spins}\rangle \langle\text{N spins}| e^{iT\varphi H},
$$

where $H = \sigma_z^{(1)} + \sigma_z^{(2)} + \cdots + \sigma_z^{(N)}$. Such a family is $2\pi/T$ periodic in $\varphi$ and there cannot be any stationary state of the clock associated to this family. Indeed, the stationary probability distribution of such state would have the same periodicity and hence it wouldn’t be normalizable.
When the initial state is given by \( M \leq N \) independent copies of the same state
\[
\rho(\varphi) = \rho^{(1)}(\varphi) \otimes \rho^{(1)}(\varphi) \otimes \cdots \otimes \rho^{(1)}(\varphi),
\]
then in the large \( M \) limit this state can be represented in the vicinity of \( \varphi = 0 \) by a Gaussian state with the Fisher information equal to the Fisher information of \( \rho(0) \), see [21, 17]. This is an instance of the quantum central limit theorem. In particular, with respect to the estimation theory, the example in Section 6 is generic in the large \( M \) limit. However to draw any conclusions about the stationary state of the clock in this limit one needs to understand the relevant time scale after which the periodicity mentioned in the first paragraph exhibits itself.

To understand this problem is very important in determination how much entanglement in the initial state is appropriate.

**Entropy production**

An atomic clock perpetually measures the quantum reference standard in order to synchronize the clock. This measurement causes entropy production on the quantum system, which can be related to a heat by the Landauer principle.

To compute this entropy production during the stationary operation was the original motivation of the author when approaching the subject of atomic clocks. However the subject turned out to be much wider and this particular question had to be postponed.

**Acknowledgements.** The author appreciates a help of N. Crawford with various aspects of the probability theory and discussions with G. M. Graf, V. Beaud and A. Zarkh. My special thanks belongs to R. Demkowicz-Dobrzanski and Y. Avron for many discussions during visits at their institutions.

**References**

[1] Axel André, AS Sørensen, and MD Lukin. Stability of atomic clocks based on entangled atoms. *Physical review letters*, 92(23):230801, 2004.

[2] Claude Audoin and Bernard Guinot. *The measurement of time: time, frequency and the atomic clock*. Cambridge University Press, 2001.

[3] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985.

[4] JJ Bollinger, Wayne M Itano, DJ Wineland, and DJ Heinzen. Optimal frequency measurements with maximally correlated states. *Physical Review A*, 54(6):4649–4652, 1996.

[5] Johannes Borregaard and Anders S Sørensen. Efficient atomic clocks operated with several atomic ensembles. *arXiv preprint arXiv:1304.5944*, 2013.
[6] Johannes Borregaard and Anders S Sørensen. Near heisenberg limited atomic clocks in the presence of decoherence. *arXiv preprint arXiv:1304.2288*, 2013.

[7] Samuel L Braunstein and Carlton M Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439–3443, 1994.

[8] Laurent Bruneau, Alain Joye, and Marco Merkli. Repeated and continuous interactions in open quantum systems. In *Annales Henri Poincare*, volume 10, pages 1251–1284. Springer, 2010.

[9] V Bužek, Radoslav Derka, and Serge Massar. Optimal quantum clocks. *Physical review letters*, 82(10):2207, 1999.

[10] David Roxbee Cox and Hilton David Miller. *The theory of stochastic processes*, volume 134. Chapman & Hall/CRC, 1977.

[11] Rafał Demkowicz-Dobrzański. Optimal phase estimation with arbitrary a priori knowledge. *Physical Review A*, 83(6):061802, 2011.

[12] Ralph Deutsch. *Estimation theory*. Prentice-Hall New York, 1965.

[13] G John Dick. Local oscillator induced instabilities in trapped ion frequency standards. Technical report, DTIC Document, 1987.

[14] G John Dick, John D Prestage, Charles A Greenhall, and Lute Maleki. Local oscillator induced degradation of medium-term stability in passive atomic frequency standards. Technical report, DTIC Document, 1990.

[15] Joseph L Doob. *Stochastic processes*, volume 101. New York, 1953.

[16] BM Escher, RL de Matos Filho, and L Davidovich. General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology. *Nature Physics*, 7(5):406–411, 2011.

[17] Richard D Gill and Mădălîn Guţă. On asymptotic quantum statistical inference. *arXiv preprint arXiv:1112.2078*, 2011.

[18] Richard D Gill and Boris Y Levit. Applications of the van trees inequality: a bayesian cramér-rao bound. *Bernoulli*, pages 59–79, 1995.

[19] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, Jan 2006.

[20] Charles A Greenhall. A derivation of the long-term degradation of a pulsed atomic frequency standard from a control-loop model. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 45(4):895–898, 1998.
[21] Mădălina Guță and Anna Jenčová. Local asymptotic normality in quantum statistics. *Communications in Mathematical Physics*, 276(2):341–379, 2007.

[22] C.W. Helstrom. *Quantum Detection and Estimation Theory*. Mathematics in Science and Engineering Series. Acad. Press, 1976.

[23] N. Hinkley, J. A. Sherman, N. B. Phillips, M. Schioppo, N. D. Lenke, K. Beloy, M. Pizzocaro, C. W. Oates, and A. D. Ludlow. An atomic clock with 1018 instability. *Science*, 341(6151):1215–1218, 2013.

[24] A.S. Holevo. *Statistical Structure of Quantum Theory*. Lecture Notes in Physics Monographs. Springer, 2001.

[25] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac. Improvement of frequency standards with quantum entanglement. *Phys. Rev. Lett.*, 79:3865–3868, Nov 1997.

[26] W. M. Itano, J. C. Bergquist, J. J. Bollinger, J. M. Gilligan, D. J. Heinzen, F. L. Moore, M. G. Raizen, and D. J. Wineland. Quantum projection noise: Population fluctuations in two-level systems. *Phys. Rev. A*, 47:3554–3570, May 1993.

[27] Jan Kolodynski and Rafal Demkowicz-Dobrzanski. Phase estimation without a priori phase knowledge in the presence of loss. *Physical Review A*, 82(5):53804, 2010.

[28] Michael Mullan and Emanuel Knill. Improving quantum clocks via semidefinite programming. *Quantum Information & Computation*, 12(7-8):553–574, 2012.

[29] F. Riehle. *Frequency Standards*. Wiley, 2006.

[30] Harry L Van Trees. Detection, estimation, and modulation: Part i, 1968.

[31] DJ Wineland, C Monroe, WM Itano, D Leibfried, BE King, and DM Meekhof. Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *Journal of Research of the National Institute of Standards and Technology*, 103(3), 1998.

[32] R Wynands and S Weyers. Atomic fountain clocks. *Metrologia*, 42(3):S64, 2005.