Multiparameter quantum metrology in the Heisenberg Limit regime: many repetition scenario vs. full optimization

Wojciech Górecki\textsuperscript{1} and Rafał Demkowicz-Dobrzański\textsuperscript{1}
\textsuperscript{1}Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

We discuss the Heisenberg limit in the multiparameter metrology within two different paradigms—the one, where the measurement is repeated many times (so the Cramér-Rao bound is guaranteed to be asymptotically saturable) and the second one, where all the resources are allocated into one experimental realization (analyzed with the minimax approach). We investigate the potential advantage of measuring all the parameter simultaneously compared to estimating them individually, while spending the same total amount of resources. We show that in general the existence of such an advantage, its magnitude and conditions under which it occurs depends on which of the two paradigms has been chosen. In particular, for the problem of magnetic field sensing using \( N \) entangled spin-\( 1/2 \), we show that the predictions based purely on the Cramér-Rao formalism may be overly pessimistic in this matter—the minimax approach reveals the superiority of measuring all the parameters jointly whereas the Cramér-Rao approach indicates lack of such an advantage.

I. INTRODUCTION

Quantum mechanics opens up new possibilities in metrology, enabling the use of coherence and entanglement to increase measurement precision \cite{1–9}. The most prominent example of this is the ability to overcome the shot noise-limit linear scaling of the estimation precision with the number of resources \( n \) used in measurement (which can be understood as number of photons, total energy, total time, etc.) and obtain a quadratic scaling, the so-called Heisenberg Scaling \cite{10–19}. Even if presence of decoherence makes the Heisenberg Scaling fragile and virtually impossible to preserve in the asymptotic limit \cite{20,21}, for many models the noise may be completely or partially canceled by applying a proper Quantum Error Correction protocols \cite{22–26}, which allows for the observation of the quadratic precision scaling in certain finite-resource regimes.

If an experiment aimed at estimation a parameter \( \theta \) is repeated \( k \) times and involves the use of \( n \) resources in each repetition, then provided the Heisenberg Scaling holds, the variance of the estimator will scale as:

\[
\Delta^2 \hat{\theta} \propto \frac{1}{k \cdot n^2}.
\]  

(1)

If \( k \) is sufficiently large, then the problem may be successfully analyzed with the use of the concept of Quantum Fisher Information (QFI) and the related Cramér-Rao (CR) bound (which is proven to be tight in the limit \( k \to \infty \)). However, as pointed out in \cite{27–32}, a subtle problem appears if one wants to discuss the best precision achievable when all the available resources \( N = n \cdot k \) are used optimally, which we will refer to as the actual Heisenberg Limit.

When inspecting Eq. (1) it is apparent that, when \( N = n \cdot k \) is kept fixed, one should accumulate as much resources as possible in a single repetition of an experiment and therefore increase \( n \) at the expense of smaller \( k \). Unfortunately, in general it is not clear what is the minimal number of repetition \( k \) needed to saturate the CR in practice (which may be different for various models), and hence the sole notion of the QFI does not provide a full understanding of the problem.

This case was broadly discussed for the problem of estimating phase-shift in the interferometer using an \( N \)-photon state (within the Bayesian \cite{33–35} or minimax \cite{27} formalism) as well as for general problem of single parameter unitary estimation \cite{32}. It was shown, that the optimal state is different than the one maximizing the QFI and that the final estimator’s variance is \( \pi^2 \) times larger then the one resulting from the QFI based analysis)—see III A for more discussion.

This implies that whenever the Heisenberg Scaling occurs, then in order to discuss the optimal measurement strategy, one needs to strictly define, which paradigm is under consideration. The one, where all resources \( N = n \cdot k \) may be used in the optimal way and accumulated in a single experiment’s realization (which in this work we analyse within minimax formalism, labeled by MM) or the second one, where the amount of resources used in single trial \( n \) is large but finite, and the whole experiment is repeated many times \( k \) (analysed within Cramér-Rao formalism, labeled by CR). In the latter case, the limit \( N \to \infty \) corresponds to \( k \to \infty \), \( n = \text{const} \). Only such a formulation allows us to apply the general argument about the asymptotical saturability of the CR bounds.

While the issues mentioned above appear now to be completely understood in a single-parameter estimation case, new questions and challenges arise when discussing multiparameter estimation models \cite{36–42}. In some situations, a properly designed multiparameter estimation protocol allows to reduce the total error in estimation when compared with a strategy where all the parameters are measured separately in independently prepared experiments \cite{43,44}. Heisenberg Scaling in multiparameter metrology has been discussed in the literature using both paradigms. Many repetition scenario has been considered in \cite{45–49}(using multiparameter quantum CR bound) and \cite{26,50–52} (using tighter variants of quantum CR bound), while single experiment scenario where the total amount of resources is limited has been analyzed mainly within the Bayesian paradigm for models with underlying group symmetry (covariant problems)—SU(2)/U(1) \cite{53,54}, SU(2) \cite{55–58}, SO(3) \cite{59}, SU(d) \cite{60}. The quantitative analysis of the relation between the results obtained within these two paradigms has started to be analyzed only very recently \cite{61}.

The most pressing question is whether a gain can be made by measuring all the parameters simultaneously instead of
separately, while consuming the same total amount of resources? That the necessity of splitting these resources between experiments focusing on estimating a given parameter in the separate case, will in general have different consequences in different paradigms [61]. If only the total amount of resources \( N \) is restricted, and \( p \) parameters are to be estimated, it is rather clear that one needs to spend \( \sim N/p \) resources per parameter. Since, we assume the quadratic scaling of precision with the amount of resource used, we may expect the scaling of the sum of variances to be:

\[
\sum_{i=1}^{P} \Delta^{2} \hat{\theta}_{i} \propto p \times \frac{1}{(N/p)^{2}} = \frac{p^{3}}{N^{2}}, \tag{2}
\]

In the many repetition scenario, different approaches to analyzing this problem may be found in modern literature. A quite common method is to compare the optimal cost obtainable with applying joint measurement with the one obtained by dividing amount of resources achievable in single trial \( n \) between all the parameters, with the total number of repetition of the whole experiment \( k \) kept unchanged [45, 46, 62–64], which leads to \( \sum_{i=1}^{P} \Delta^{2} \hat{\theta}_{i} \propto p \times \frac{1}{k} \times \frac{1}{(n/p)^{2}} = \frac{p^{3}}{kn^{2}}. \)

However, one should notice that there is no point in dividing \( n \), while we have \( k \gg n \) trials at our disposal. It is much more efficient to use \( \sim k/p \) trials for each parameter [65–67]. Only such formulation of the problem allows for a fair comparison between measuring parameters jointly vs. separately, and guarantees, that the eventual superiority of the first one comes directly from measuring parameters jointly, not from more efficient resources distribution. As a result we obtain:

\[
\sum_{i=1}^{P} \Delta^{2} \hat{\theta}_{i} \propto p \times \frac{1}{k/p} \times \frac{1}{n^{2}} = \frac{p^{2}}{kn^{2}}, \tag{3}
\]

which exhibits scaling with \( p \) with a different power than in Eq. (2).

The nontrivial question is, if the sum of variances for the optimal joint measurement will follow a similar scaling—it may turn out that the existence of the advantage depends on paradigm chosen. As recently shown in [61], for the multiphase estimation problem in a multiarm interferometer, this issue does not lead to divergent conclusions—even if the scaling of the total cost with the number of the parameters depends on the paradigm, the potential advantage obtainable by measuring all of the parameters jointly instead of separately is very similar in both paradigms. In this paper we analyze this issue more generally and show that result is not universal—it may happen that existence or absence of the advantage indeed depends on the paradigm chosen.

| \( \min \sum_{i=1}^{P} \Delta^{2} \hat{\theta}_{i} \) | model | 1 component | 2 components | 3 components | magnetic field estimation | 1 component, spatial distribution | multiphase estimation | scaling |
|---|---|---|---|---|---|---|---|---|
| Hamiltonian | \( \theta \sigma_{z}/2 \) | \( \sum_{i \in \{x,y\}} \theta_{i} \sigma_{x}/2 \) | \( \sum_{i \in \{x,y,z\}} \theta_{i} \sigma_{y}/2 \) | \( \sum_{i=1}^{P} \theta_{i} \sigma_{i}^{(1)}/2 \) | \( \sum_{i=1}^{P} \theta_{i} \sigma_{i} \) | \( \sum_{i=1}^{P} \theta_{i} |i\rangle \langle i| \otimes \sigma_{i}^{(1)}/2 \) | \( \sum_{i=1}^{P} \theta_{i} |i\rangle \langle i| \) | \( \frac{p^{3}}{N^{2}} \) |
| \( N = kn \) | separately | \( \pi^{2} \) | \( 8\pi^{2} \leq ? \) | \( 2\pi^{2} \leq ? \) | \( 4\pi^{2} \leq ? \) | \( \pi^{2}p^{2} \) | \( \pi^{2}p^{2} \) | \( \pi^{2}p^{2} \) | \( \pi^{2} \) |
| \( n \) | jointly | \( - \) | \( 4 \) (2) | \( 9 \) (3) | \( p \) | \( p \) | \( p^{2} \) | \( \frac{p^{2}}{4} \) | \( \frac{p^{2}}{4} \) |
| \( k \) reps. | separately | \( - \) | \( 9 \) (3) | \( 9 \) (3) | \( p \) | \( p \) | \( p^{2} \) | \( \frac{p^{2}}{4} \) | \( \frac{p^{2}}{4} \) |

The examples are intended to present to the reader in a simple way the diversity and the complicity of the relationship between optimal results obtainable within separate and joint strategies in both discussed paradigms, therefore most of them are easily solvable with basic algebra or are based on analyzing existing results [53–55, 57]. However, in IV B we present also the original results about optimal joint measurement within the minimax approach for spatially distributed magnetic field sensing (which is a specific example with commuting evolution generators). The end results are summarized in table I and section V.
II. PROBLEM FORMULATION

Let $E_\theta$ be a quantum channel depending on a vector of unknown parameters $\theta = [\theta_1, \ldots, \theta_p]^T$. The aim is to estimate the values of $\theta$ by sending through the channel some initial state $E_\theta(p_{in}) = \rho_0$, performing the measurement $\{M_x\}$ (satisfying $\int dx M_x = 1$) on the $\rho_0$ (leading to probability distribution of the result $x$ given by $p_\theta(x) = \text{Tr}(M_x \rho_0)$) and assigning to outputs of the measurement proper values of estimators $\hat{\theta}_i(x)$.

For such a strategy we define the estimators covariance matrix by:

$$\Sigma = \int dx p_\theta(x) (\hat{\theta}(x) - \theta)(\hat{\theta}(x) - \theta)^T,$$  \hspace{1cm} (4)

and the aim is to minimize its trace, i.e. the sum of squared deviations of the estimator from the true value (for simplicity we will refer to them as variances, implicitly assuming that the estimators will likely be unbiased and their expectation value will coincide with the true value of the parameter) of all parameters:

$$\Delta^2 \hat{\theta} := \sum_{i=1}^{p} \Delta^2 \hat{\theta}_i = \text{Tr}(\Sigma).$$ \hspace{1cm} (5)

We will compare the value of the above variance at some reference point $\theta = \theta_0$, keeping the condition that the measurement should works well also in some small neighbourhood of this point (two alternative methods how to formalize this condition will be given later).

In the literature, a more general cost function is sometimes considered, where the covariance matrix is additionally multiplied by a positive semidefinite weight matrix under the trace (which makes also non-diagonal elements of the covariant matrix important). Note, however, that such a general case can be reduced to the above after proper reparametrization. Indeed, if for some parametrization $\theta'$ the cost is given by $\text{Tr}(W \Sigma')$ (with $W \geq 0$), one may take $A$ satisfying $W = A^T A$ and then $\text{Tr}(W \Sigma') = \text{Tr}(A^T A \Sigma') = \text{Tr}(A \Sigma' A^T) = \text{Tr}(\Sigma)$ (where in last step we applied $\theta = A \theta'$). Therefore, for simplicity of further formulas we will only consider the cost of the form as given in Eq. (5).

We will consider multiple use of the channel. It may be seen as the action of $N$ gates in parallel on an arbitrary entangled $N$-probe state:

$$\rho_{N, \theta} = E_\theta \otimes_N \rho_{in},$$ \hspace{1cm} (6)

or, even more general, as a general adaptive scheme, where we apply $N$ sequential usage of the channel, where arbitrary large ancilla and arbitrary unitary controls between the actions of the channel are allowed $[1, 23, 68]$:

$$\rho_{N, \theta} = V_N \circ (E_\theta \otimes I) \cdots V_1 \circ (E_\theta \otimes I)(\rho_{in})$$ \hspace{1cm} (7)

(where $V_i \circ \rho$ is a shortcut for $V_i \rho V_i^\dagger$). Note, that the first one may be simulated by the latter. In such a formulation, the amount of resources corresponds to the number of uses of the channel.

The potential advantage of measuring $p$ parameters jointly vs. separately was extensively discussed in the literature with different approaches, where alternatively the amount of resources used in a single trial $[45, 46, 62–64]$ or the total number of trials $[65–67]$ was divided between parameters in separate strategy; another analysis free of the resource allocation problem was also performed $[43, 44]$ (see Appx. A for broader discussion). In this paper, we would like to focus on the situation, when the same task (i.e. estimation of the set of parameters with a given cost function) is performed with a joint or a separate strategy, when the same amount of resources is used in the end. The problem is schematically shown in Fig. 1, where Bob send to Alice $N$ copies of a quantum gate depending on unknown parameters $\theta$ (where each copy can only be used once) and expect from Alice that she will send him back estimated values of parameters $\hat{\theta}$, in a way that the total cost is minimized.

The Alice may alternatively use $N$ gates to measure all the parameters jointly or separately. In the first case, labeled as JNT, the minimal achievable cost is given by:

$$\Delta^2 \hat{\theta}_{\text{JNT}} = \min_{N \text{-protocol}} \text{Tr}(\Sigma),$$ \hspace{1cm} (8)

where by minimization over “N-protocol” we understand the minimization over the choice of the initial state $\rho_{in}$, the unitaries $V_i$ acting between $N$ usage of the gate $E_\theta$, the measurement $\{M_x\}$ and the estimator $\hat{\theta}(x)$.

In the second case, labeled as SEP, Alice needs to divide all $N$ gates between her “Minions”, sending to each of them $N_i$ gates and ordering to focus on the measurement of a single parameter $\theta_i$. Note, that we consider the case when all the parameters have some fixed (unknown) values which do not fluctuate themselves. Hence, in the separate scenario, different parameters are estimated from independent measurements and the resulting covariance matrix will always be diagonal (assuming, that all estimators are unbiased):

$$\Sigma = \text{diag}(\Delta^2 \hat{\theta}_1, \ldots, \Delta^2 \hat{\theta}_p).$$ \hspace{1cm} (9)

The minimal achievable cost for such a strategy is given by:

$$\Delta^2 \hat{\theta}_{\text{SEP}} = \min_{\{N_i\}} \left( \sum_{i=1}^{p} \min_{N_i \text{-protocol}} \Delta^2 \hat{\theta}_i \right).$$ \hspace{1cm} (10)

However, this strategy may by further optimized (while retaining the key feature that each of the Minions measures only a single parameter). Indeed, Alice may demand from them to measure arbitrary linearly independent combinations of original parameters $\theta_i = [A^{-1} \theta]_i$ (where $A$ is invertible matrix):

$$\Sigma' = \text{diag}(\Delta^2 \hat{\theta}_1', \ldots, \Delta^2 \hat{\theta}_p'), \quad \Sigma = A \Sigma' A^T.$$ \hspace{1cm} (11)

To distinguish this strategy from the one, where parametrization is fixed, we will label it as SEP+. Note that the resulting covariance matrix is diagonal in $\theta'$ parametrization, but not necessarily in the initial one. The resulting cost is given by:

$$\sum_{i=1}^{p} \Delta^2 \hat{\theta}_i = \text{Tr}(\Sigma) = \text{Tr}(A \Sigma' A^T) = \sum_{i=1}^{p} [A^T A]_{ii} \Delta^2 \hat{\theta}_i,$$ \hspace{1cm} (12)
Minions are allowed to measure mutually linearly independent combinations \( \tilde{\theta} = [\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_p] \) with \( \min \sum_{i=1}^{p} \Delta^2 \tilde{\theta}_i \). Therefore, in general we have:

\[
\Delta^2 \tilde{\theta}_{\text{JNT}} \leq \Delta^2 \tilde{\theta}_{\text{SEP}+} \leq \Delta^2 \tilde{\theta}_{\text{SEP}}. \tag{14}
\]
as for each inequality the right-hand side may be seen as the special case of the one on the left side.

We now want to analyze these strategies further using the two paradigms—the one when the constraint is imposed on the number of gates used in a single trial \( n \), and where the number of trials \( k \gg n \), and the second one where only the total number of gates \( N \) is limited.

### A. Resources distribution in separate strategies

Let us start with some general analysis of optimal resource distribution in separate strategies, focussing on the limit of large \( N \). Assume that for each parameter \( \theta_i \), the minimal variance obtainable with the use of \( N_i \) gates (where \( \sum_{i=1}^{p} N_i = N \) scales, in the leading terms, like \( N_i^{-\alpha} \):

\[
\Delta^2 \tilde{\theta}_{\text{SEP}} = \min_{\{N_i\}} \sum_{i=1}^{p} N_i \min_{\text{protocol}} \Delta^2 \tilde{\theta}_i = \min_{\{N_i\}} \sum_{i=1}^{p} \frac{c_i}{N_i^\alpha} + o(N_i^{-\alpha}), \tag{15}
\]

where all \( c_i \) have finite positive values. Neglecting term \( o(N_i^{-\alpha}) \), by applying the standard Lagrange multipliers method, we obtain the optimal resources redistribution to be:

\[
N_i = N \frac{1}{\sum_j c_j^{1/(\alpha+1)}} \tag{16}
\]

which leads to

\[
\Delta^2 \tilde{\theta}_{\text{SEP}} = \frac{1}{N^\alpha} \left( \sum_{i=1}^{p} c_i^{1/(\alpha+1)} \right)^{\alpha+1} + o(N^{-\alpha}). \tag{17}
\]

Further in this paper, we focus only on the asymptotic behavior and for simplicity of the formulas we will omit the term \( o(N^{-\alpha}) \), using the sign ‘\( \sim \)’ instead. Introducing \( \bar{c}_\alpha \) for the proper power mean \( \bar{c}_\alpha = \left( \frac{1}{p} \sum_{i=1}^{p} c_i^{1/(\alpha+1)} \right)^{\alpha+1} \)

we get:

\[
\Delta^2 \tilde{\theta}_{\text{SEP}} \approx \frac{\bar{c}_\alpha}{N^\alpha}, \tag{18}
\]

so (for models in which \( \bar{c}_\alpha \) does not scales with \( p \) we see a \( p^{\alpha+1} \) scaling of the cost with the number of parameters involved.

Moreover, the optimal SEP and SEP+ strategies may be bounded from above by:

\[
\Delta^2 \tilde{\theta}_{\text{SEP}+} \leq \Delta^2 \tilde{\theta}_{\text{SEP}} \ll p^\alpha \Delta^2 \tilde{\theta}_{\text{JNT}}, \tag{19}
\]
as one may always consider a suboptimal separate strategy, where each of the Minions performs a measurement corresponding to the optimal JNT protocol (with \( N/p \) gates), but sends to Alice only the estimated value corresponding to a single parameter \( \theta_i \).

FIG. 1. Bob sends to Alice \( N \) quantum gates that depend on \( p \) unknown parameters \( \theta_i \). Her goal is to send him back their estimated values \( \tilde{\theta}_i \), in order to minimize the total quadratic cost. In [JNT] she is allowed perform full optimization and measure all the parameters jointly. In [SEP] she divides the gates between her Minions, where each of them is told to measure a single parameter \( \theta_i \). In [SEP+] the Minions are allowed to measure mutually linearly independent combinations \( \theta_i' = [A^{-1} \theta_i] \), and send the result back to Alice, who reconstructs the initial parameters in post-processing. Note, that [SEP+] may be seen as a special case of [JNT], while [SEP] as a special case of [SEP+].

and therefore:

\[
\Delta^2 \tilde{\theta}_{\text{SEP}+} = \min_{A,\{N_i\}} \left( \sum_{i=1}^{p} \{[A^T A]_{ii} \min_{\text{protocol}} \Delta^2 \tilde{\theta}_i \} \right). \tag{13}
\]

Note, that for orthogonal transformations \( A = O \) where \( O^T O = I \), no additional term appears in Eq. (13). However, it may happen that the optimal separate strategy indeed requires a non-orthogonal transformations (see Appx. B for an example).
B. Many repetition scenario (n gates, k trials): 

Cramer-Rao like bounds

Consider now the situation where the number of gates used in a single trial is restricted to some large but finite number $n$. Then total amount of resources is divided into $k$ trials, satisfying $N = nk$, and the limit $N \rightarrow \infty$ corresponds to $k \rightarrow \infty, n = \text{const}$. Therefore, in the context of the discussion from the previous section, the scaling of the cost with the total amount of resources $N$ is always linear here (as it scales linearly with $k$), no matter how it depends on $n$.

Let us briefly remind the foundations of the CR bound approach. The CR bound is based on the idea of local unbiasedness. We consider only estimators satisfying:

$$\int dx p_0(x) \hat{\theta}_i(x) = \theta_{0i}, \quad \int dx \frac{d p_0(x)}{d \theta_i} |_{\theta = \theta_0} \hat{\theta}_j(x) = \delta_{ij}. \quad (20)$$

For a given output state $\rho^\theta$ the covariance matrix is bounded by Quantum CR inequality:

$$\Sigma \geq \frac{1}{k} F^{-1}, \quad F_{ij} = \text{Tr} \left[ \rho^\theta \frac{1}{2} (L_i L_j + L_j L_i) \right], \quad (21)$$

where the matrix inequality means that $\Sigma - \frac{1}{k} F^{-1}$ is positive semidefinite and $L_i$ are the symmetric logarithmic derivatives satisfying $\frac{d p_0}{d \theta_i} |_{\theta = \theta_0} = \frac{1}{2} (L_i \rho^\theta + \rho^\theta L_i)$. This leads to

$$\text{Tr}(\Sigma) \geq \frac{1}{k} \text{Tr}(F^{-1}). \quad (22)$$

Since measurements optimal for different parameters might be mutually incompatible, the above inequality is asymptotically saturable for large $k$ if and only if [44]:

$$\text{Im} \text{Tr}(\rho^\theta L_i L_j) = 0. \quad (23)$$

More precisely, if Eq. (23) is satisfied and additionally $\rho^\theta$ is pure, then there exists a local measurement (performed on a single copy of $\rho^\theta$) and an estimator depending on $k$ measurement results $\hat{\theta}(x_1, x_2, ..., x_k)$, which is asymptotically unbiased and saturate Eq. (22). However, if $\rho^\theta$ is the mixed state, in may in general happen that a collective measurement on all $k$ copies of the state $\rho^\theta$ is required in order to saturate the CR bound [50, 52]. Therefore, the minimal achievable cost obtainable with $k$ trials, where in each trial $n$ gates are used, may be bounded from below by the RHS of Eq. (22) minimized over all feasible output states $\rho_{n, \theta}$ (i.e. states that can be obtained when optimizing the protocol over $\rho_m$ and all $V_1, ..., V_n$ in Eq. (7)):

$$\Delta^2 \hat{\theta}_{\text{CR}} \geq \frac{1}{k} \min_{\rho_{n, \theta}} \text{Tr}(F^{-1}), \quad (24)$$

where the inequality is tight iff Eq. (23) holds—in fact, in all examples discussed in this paper this will be the case.

Now let us consider a separate strategy and let $\rho_{n, \theta}$ be the output state designed to estimate the value of parameter $\theta_i$. Note, that in principle it may by inefficient or even impossible to fully isolate the dependence of the state on the remaining parameters $\theta_{j \neq i}$—their variations may affect the measurement results, and they cannot be omitted in the analysis of the optimal separate protocol. That means that the remaining parameters should in general be treated as nuisance parameters [69, 70]. Denoting by $F_i$ the QFI matrix corresponding to $\rho_{n, \theta}$, the minimal variance of estimating parameter $\theta_i$ using $k$ repetitions of an experiment is bounded by:

$$\Delta^2 \hat{\theta}_i \geq \frac{1}{k_i} [F_i^{-1}]_{ii}. \quad (25)$$

Strictly speaking, it is not necessary for the whole matrix $F_i$ to be invertible—it is enough if $\lim_{\epsilon \to 0} \left[ (F_i + \epsilon \mathds{1})^{-1} \right]_{ii}$ converges to a finite value. For example, if $F_i$ has a block-diagonal structure, it is enough if only the block containing $[\cdot]_{ii}$ element is invertible.

Note that in general $[F_i^{-1}]_{ii} \geq [F_i]_{ii}^{-1}$, where the RHS of the inequality corresponds to a direct application of the single-parameter estimation theorem—neglecting the role of nuisance parameters. Moreover, Eq. (25) is always saturable for large $k_i$ if collective measurements allowed [52, Section 2.7], and therefore, Eq. (17) (with $\alpha = 1$) takes the form:

$$\Delta^2 \hat{\theta}_{\text{SEP}} \approx \frac{1}{k} \left( \sum_{i=1}^{P} \min_{\rho_{n, \theta}} \left[ \min_{F_i} [F_i^{-1}]_{ii} \right]^2 \right). \quad (26)$$

After optimization over the parametrization is performed we get:

$$\Delta^2 \hat{\theta}_{\text{SEP}} \approx \frac{1}{k} \min_{A} \left( \sum_{i=1}^{P} \sqrt{[A^T A]_{ii} \min_{\rho_{n, \theta}} [A^{-1} F_i^{-1} A^{-1 T}]_{ii}} \right), \quad (27)$$

where the rule for the transformation of the QFI matrix $F_i = A^T F_i A$ has been used.

Using Eq. (14) and Eq. (19) we can write:

$$\Delta^2 \hat{\theta}_{\text{INT}} \leq \Delta^2 \hat{\theta}_{\text{SEP}} \leq \Delta^2 \hat{\theta}_{\text{SEP}} \lesssim p \cdot \Delta^2 \hat{\theta}_{\text{CR}}. \quad (28)$$

C. Heisenberg Limit (total N gates): asymptotic local minimax bound

The methods discussed in the previous section cannot be used in a situation, where only the total amount of gates $N$ is constrained, since we can not invoke the general CR saturability arguments which require a many repetition scenario. In the single parameter case the problem has been discussed within the MM [27] and the Bayesian [32] approach.

In this paper we will follow the first one, as it is conceptually and technically simpler to apply in the case of estimating many parameters should in general be treated as nuisance parameters [50, 52]. Therefore, in the context of the discussion from the previous section, the scaling of the cost with the total amount of resources $N$ is always linear here (as it scales linearly with $k$), no matter how it depends on $n$.
consider a strategy, which minimizes the cost in the most pessimistic scenario (we always choose the point in \( \Theta \), where the strategy works the worst). Here, unlike in the previously discussed approach, only a single realization of the measurement is considered, and any measurement outcome is directly related with a given value of the estimator \( \tilde{\theta} \). Due to this fact, and for the simplicity of notation, we may label the measurement’s outcomes by \( \tilde{\theta} \), so that the formula for the MM bound takes the form:

\[
\inf_{\rho_0, M_0} \sup_{\theta \in \Theta(0, \delta)} \int d\tilde{\theta} \text{Tr}(M_0 \rho_0)(\tilde{\theta} - \theta)^2.
\]  \( (29) \)

This value, however, depends on the size of \( \Theta \). In order to get rid of this dependence, and hence be able to compare the results with the CR based approach (which is effectively a single-point estimation approach), we do the following construction. Let \( \{\rho_{N, \theta}, M_{N, \theta}\} \) be a sequence of output states (for an \( N \)-gate protocol) and the corresponding measurements. Then, assuming that for a large \( N \) the corresponding cost scales like \( 1/N^\alpha \) (where \( \alpha = 1 \) corresponds to the standard scaling, while \( \alpha = 2 \) to the Heisenberg Scaling), define [27, Section 5]:

\[
C^{(\alpha)}[\{\rho_{N, \theta}, M_{N, \theta}\}] = 
\lim_{\delta \to 0} \lim_{N \to \infty} N^\alpha \sup_{\theta \in \Theta(0, \delta)} \int d\tilde{\theta} \text{Tr}(M_{N, \tilde{\theta}} \rho_{N, \tilde{\theta}})(\tilde{\theta} - \theta)^2.
\]  \( (30) \)

Note, that the order of taking the limits matters (as for the opposite order the trivial constant estimator pointing \( \theta_0 \) independent on the measurement results would lead to a zero cost). Intuitively, such an order approximates a situation, where one consider a \( \delta \)-independent measurement strategy and for each \( \delta \) checks its validity only for \( N \) which is much larger than the inverse of \( \delta \). Finally, taking the limit \( \delta \to 0 \) makes the results independent of the fact that the estimation around different points in \( \Theta \) may be in principle harder than around \( \theta_0 \) (i.e. even the value of the maximal QFI may depend on the value of \( \theta \), as for example in [46]).

It was shown in [27, Section 5] that if for a given channel the Heisenberg Scaling is not achievable (\( \alpha = 1 \)), then indeed

\[
\inf_{\{\rho_{N, \theta}, M_{N, \theta}\}} C^{(1)} = \lim_{N \to \infty} \inf_{\rho_{N, \theta}} \frac{N}{F}.
\]  \( (31) \)

Therefore the MM (in the limit \( \delta \to 0 \)) and the CR based approaches return consistent results. However, if the channel estimation problem admits the Heisenberg Scaling (\( \alpha = 2 \)) then

\[
\inf_{\{\rho_{N, \theta}, M_{N, \theta}\}} C^{(2)} \geq \lim_{N \to \infty} \inf_{\rho_{N, \theta}} \frac{N^2}{F},
\]  \( (32) \)

where the inequality is not tight in general (see the next section). In this paper we are focusing only on the case \( \alpha = 2 \), and hence in what follows for a more compact notation we will drop the upper index.

For a general multiparameter estimation problem, let us define \( \Theta(0, \delta) = [\theta_{0,1} - \delta/2, \theta_{0,1} + \delta/2] \times \ldots \times [\theta_{0,p} - \delta/2, \theta_{0,p} + \delta/2] \) and

\[
C_i[\{\rho_{N, \theta}, M_{N, \theta}\}] = \lim_{\delta \to 0} \lim_{N \to \infty} N^2 \sup_{\theta \in \Theta(0, \delta)} \int d\tilde{\theta}_i \text{Tr}(M_{N, \tilde{\theta}_i} \rho_{N, \tilde{\theta}_i})(\tilde{\theta}_i - \theta_i)^2.
\]  \( (33) \)

Then, for a large \( N \), Eq. (17) (with \( \alpha = 2 \)) up to the leading term in \( N \) takes the form:

\[
\Delta^2 \theta^\text{MM}_{\text{SEP}} \simeq \frac{1}{N^2} \left( \sum_{i=1}^N \inf_{\{\rho_{N, \theta}, M_{N, \theta}\}} C_i \right)^3.
\]  \( (34) \)

After the optimization over reparametrizations it reads:

\[
\Delta^2 \theta^\text{MM}_{\text{SEP}+} \simeq \frac{1}{N^2} \min_{A} \left( \sum_{i=1}^N \inf_{\{\rho_{N, \theta}, M_{N, \theta}\}} C_i \right)^3
\]  \( (35) \)

where \( C_i \) is given by Eq. (33), after making the substitution \( \theta_i \to \theta_i = [A^{-1}\theta_i]_i \). Analogously, for the joint estimation case:

\[
C^\text{INT}[\{\rho_{N, \theta}, M_{N, \theta}\}] = \lim_{\delta \to 0} \lim_{N \to \infty} N^2 \sup_{\theta \in \Theta(0, \delta)} \int d\tilde{\theta} \text{Tr}(M_{N, \tilde{\theta}} \rho_{N, \tilde{\theta}}) \text{Tr}(\Sigma)
\]  \( (36) \)

and

\[
\Delta^2 \theta^\text{MM}_{\text{INT}} \simeq \frac{1}{N^2} \inf_{\{\rho_{N, \theta}, M_{N, \theta}\}} C^\text{INT}.
\]  \( (37) \)

Finally, using Eq. (14) and Eq. (19) we can write:

\[
\Delta^2 \theta^\text{MM}_{\text{INT}} \leq \Delta^2 \theta^\text{MM}_{\text{SEP}+} \leq \Delta^2 \theta^\text{MM}_{\text{SEP}} \lesssim p^2 \cdot \Delta^2 \theta^\text{MM}_{\text{INT}}.
\]  \( (38) \)

### III. HEISENBERG LIMIT BOUND

Further on we will be interested in the models where Heisenberg Scaling occurs for all the estimated parameter. This is the case for noiseless unitary evolution, where the parameters enter into the evolution as multipliers of the evolution generators:

\[
U_\theta = e^{i \theta A},
\]  \( (39) \)

with \( A = [A_1, \ldots, A_p]^T \) where all \( A_i \) are mutually linearly independent.

Note, that in presence of noise, Heisenberg Scaling may not be achieved in general [20, 72]. Still, for certain noise models, a proper quantum error correction protocol may be used to isolate the part of signal which is undisturbed by the noise and effectively obtain a purely unitary evolution [23, 25, 26, 73] of the form Eq. (39). Hence, our discussion here, will be relevant for such models as well.
A. Single parameter staurable lower bound

As a reference point for further considerations, let us recall a paradigmatic estimation model—a single phase estimation problem in a two arm interferometer. In this case, the parameter encoding channel acts on the two-mode single photon states space spanned by \( |0\rangle, |1\rangle \), and is represented by:

\[
U_\theta = e^{i\theta |1\rangle\langle 1|}.
\]  
(40)

For simplicity, let us focus on parallel strategies Eq. (6) first. The output state \( |\psi^N_\theta\rangle = U^\otimes N_\theta |\psi_{in}\rangle \), which maximizes the QFI is the famous \( |000n\rangle \) state:

\[
|\psi^N_\theta\rangle = \frac{1}{\sqrt{2}}(|0\rangle^\otimes n + e^{i\theta}|1\rangle^\otimes n),
\]  
(41)

for which \( F = n^2 \) and CR the bound may be saturated with protective measurement onto states \( |\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle^\otimes n \pm i|1\rangle^\otimes n) \). In fact, it may be also saturated by a standard photon counting measurement and an estimator based on the parity of detected number of photons [74].

Note, however, that the state Eq. (41) is unable to distinguish between the phases that differ by a multiple of \( 2\pi/n \) [27, Section 5]. Therefore it only allows for estimation of the parameter in a small region of \( \theta \in [\theta_0 - \pi/n, \theta_0 + \pi/n] \). While many repetition scenario is under consideration, this issue does not generate a serious problem, as even starting with the unknown phase, for \( k \gg n \) one may always spend the first \( \sqrt{k} \) trials to find such small region (for example by using the product states \( \frac{1}{\sqrt{2}}(|0\rangle^\otimes n + |1\rangle^\otimes n) \) in each repetition) and next in remaining \( k - \sqrt{k} \) trials use \( |000n\rangle \) states to finally achieve precision \( 1/kn^2 \) (up to the leading term in \( k \)). However, it is clear that an analogous single \( |N00N\rangle \) state cannot be used to obtain the fundamental Heisenberg Limit—if inserted into Eq. (30) it would lead to \( C = \infty \).

Calculation of the minimal obtainable value of \( C \) by a direct minimization of Eq. (30) is a hard problem. Fortunately, the task may be significantly simplified by taking into account a symmetry of the problem.

First note, that for such a channel the two phases which differ by a factor \( 2\pi \) should be regarded as equivalent. Therefore, we consider a periodic cost function of the form:

\[
\text{cost}(\theta, \tilde{\theta}) = 4\sin^2 \left( \frac{\theta - \tilde{\theta}}{2} \right),
\]  
(42)

which reflects this property whereas for small difference may be well approximated by \( \approx (\theta - \tilde{\theta})^2 \).

Next, as argued in [27], the local asymptotic MM cost Eq. (30) for this problem is exactly the same as the minimal obtainable cost for a completely unknown phase \( \theta \in \Theta = [0, 2\pi) \) multiplied by \( N^2 \) (in the limit \( N \to \infty \)). Intuitively, the reason for this is that for any finite \( \delta \), when we start from a completely unknown phase we always spend at the beginning \( \sqrt{N} \) gates to discriminate the region of size \( \delta \), where the true value of \( \theta \) lies (with the probability of the error decreasing as exponentially fast with \( N \)), and use the remaining \( N - \sqrt{N} \) to estimate the value inside this region.

Note, that while at a first glance the above construction seems to require adaptiveness (as in the second step we use the information from the first one), in may in fact be performed also within the parallel scheme (see details in Appx. C). See also [32, 61, 75] for further discussion about optimization of region discrimination in the first part of the above construction, as well as a general discussion about this bound for finite \( N \).

The state which is optimal for measuring a completely unknown phase is the \([|N\rangle\rangle \) state [33–35]:

\[
|\psi^N_\theta\rangle = \sum_{m=0}^N e^{im\theta} \frac{\sqrt{2}}{\sqrt{N+2}} \sin\left( \frac{(m+1)\pi}{N+2} \right) |N-m\rangle_0 |m\rangle_1,
\]  
(43)

where \(|N-m\rangle_0 |m\rangle_1\) is a fully symmetric state with \( m \) photons in the sensing arm \( |1\rangle \) and \( N - m \) is the reference arm \(|0\rangle\). The corresponding mean cost obtainable with applying covariant measurement is:

\[
\forall_{\theta} \int d\tilde{\theta} \text{Tr}(M_{N,\tilde{\theta}}\rho_{N,\theta})\text{cost}(\tilde{\theta}, \theta) = 2 \left( 1 - \cos \left( \frac{\pi}{N+2} \right) \right),
\]  
(44)

and therefore the constant which multiplies the leading term \( 1/N^2 \) equals:

\[
\inf_{\{M_{N,\theta}, \rho_{N,\theta}\}} C = \lim_{N \to \infty} N^2 2 \left( 1 - \cos \left( \frac{\pi}{N+2} \right) \right) = \pi^2. \tag{45}
\]

The natural question arises: is there a simple interpretation of this \( \pi^2 \) factor discrepancy between the minimal achievable variance and the inverse of maximal Fisher information? In fact, one may indeed use the \(|000n\rangle \) states even when estimating a completely unknown phase, provided one divides all the available resources \( N \) into \( M \) subsets, each of size \( m_i \) and then use \( M \) states of the form \( \frac{1}{\sqrt{2}}(|0\rangle^\otimes m_i + |1\rangle^\otimes m_i) \) for estimation. It was shown numerically, that for the optimal distribution of \( m_i \), the overhead factor \( \pi^2 \) is indeed recovered [76, 77]. Some suboptimal strategies have also been demonstrated experimentally, revealing only a slightly bigger variance [78, 79].

The above result may be directly generalized for an arbitrary quantum channel of the form \( U_\theta = e^{i\theta A} \) [32]. Denoting by \( \lambda[A] \) the difference between the maximal and the minimal eigenvalues of the operator \( A \), for optimal usage of \( N \) quantum gates we have:

\[
\inf_{\{\rho_{n,\theta}, M_{N,\theta}\}} C = \frac{\pi^2}{\lambda^2[A]} \Rightarrow \Delta^2 \tilde{\theta}_{\text{MM}} \approx \frac{\pi^2}{N^2 \lambda^2[A]}.
\]  
(46)

For comparison, in the scenario, where \( k \to \infty \) repetitions is considered (with usage \( n \) gates in each of them), minimal obtainable cost is given by:

\[
\max_{\rho_{n,\theta}} F = n^2 \lambda^2 \Rightarrow \Delta^2 \tilde{\theta}_{\text{CR}} \approx \frac{1}{kn^2 \lambda^2[A]}.
\]  
(47)

In both scenarios, this optimal precision is obtainable already in the parallel scheme and it cannot be beaten by any adaptive protocol that involves additional action of \( V_t \) operations in between [32].
B. Multi-parameter unitary estimation lower bound

Consider a general problem of local unitary channel estimation:

\[ U_\theta = e^{i\theta \Lambda}, \]  

(48)

with linearly independent generators \( \Lambda_i \) acting on a \( d \) dimensional space. First note that for such a formulated problem the Heisenberg Scaling is indeed achievable. This was shown for the most general \( SU(d) \) estimation problem for any \( d \) using a parallel scheme—in both many repetition scenario [80] (where the exact fundamental bound has been derived and proven to be saturable if entanglement with ancilla allowed) and single repetition scenario [60] (where optimal scaling has been proven to be \( \propto 1/N^2 \) with an exemplary state satisfying this scaling). As for any finite dimensional space, the problem states in Eq. (48) may be seen as estimation of some subset of \( SU(d) \) generators, the statement is proven.

Let us first argue, that from the point of view of estimation of any single parameter \( \theta_i \), the existence of an additional part of the generator \( \sum_j \theta_j \Lambda_j \) cannot help in estimation, i.e. it cannot decrease the minimal achievable cost. Note, that a single gate \( U_\theta \) (where all generator acts jointly) may be arbitrarily well approximated (for a sufficiently large \( l \)) by:

\[ U_\theta = e^{i\theta \Lambda} \approx (e^{i\theta_i \frac{\Lambda_i}{l}} e^{i\sum_i \theta_i \frac{\Lambda_i}{l}})^l, \]

(49)

(more precisely, the above approximation is exact in the limit \( l \to \infty \) due to the Trotter formula [81]). Therefore an \( N \)-fold action of \( U_\theta \) may be seen as \( l \cdot N \) action of \( U_{\theta_i} = e^{i\theta_i \frac{\Lambda_i}{l}} \), with unitary controls \( V = e^{i\sum_j \theta_j \frac{\Lambda_j}{l}} \) in between. After such a procedure the product \( \lambda[\Lambda_i/l] \cdot (IN) = \lambda[\Lambda_i]N \) remains unchanged. Therefore, both the asymptotic value of the bound and the rate of its convergence remain the same. Consequently, for each \( \theta_i \) we have:

\[ \inf_{\{\rho_{N,\theta,M_{\theta,\theta}}\}} C_i \geq \frac{\pi^2}{\lambda^2[A_i]}, \]

(50)

Note, however, that the saturability of the bound is not guaranteed. Let us now bound from below the minimal achievable cost by assuming the most optimistic scenario—the one where not only the existence of the other part of the Hamiltonian \( \sum_j \theta_j \Lambda_j \) does not disturb the sensing of \( \theta_i \), but also that there exist a single input state and a single measurement which are simultaneously optimal for sensing of all the parameters. Then, for any finite \( N \) we have:

\[ \min_{\rho_{N,\theta,M_{\theta,\theta}}} \text{Tr}(\Sigma) = \min_{\rho_{N,\theta,M_{\theta,\theta}}} \sum_{i=1}^p \Delta^2 \hat{\theta}_i \geq \sum_{i=1}^p \inf_{\rho_{N,\theta,M_{\theta,\theta}}} \Delta^2 \hat{\theta}_i, \]

(51)

and therefore, taking the asymptotic limit \( N \to \infty \), we can write:

\[ \inf_{\{\rho_{N,\theta,M_{\theta,\theta}}\}} C_{\text{INT}} \geq \sum_{i=1}^p \inf_{\rho_{N,\theta,M_{\theta,\theta}}} C_i \geq \sum_{i=1}^p \frac{\pi^2}{\lambda^2[A_i]}, \]

(52)

Notice, that we have some freedom in choosing the parametrization in Eq. (48). Indeed, all the steps remain valid after an application of any orthogonal rotation in the parameter space \( \theta' = O^{-1} \theta, \Lambda' = O^T \Lambda \), which does not change the local cost function. Note, however, that the restriction to orthogonal transformations \( O^T O = I \) is crucial here, as for a more general one \( A \) the off-diagonal elements may appear in the formula for the cost \( \text{Tr}(A^T \Delta \Sigma) \), which would make it impossible to bound it using Eq. (51). Therefore, we can tighten the bound resulting from Eq. (52) and write:

\[ \Delta^2 \hat{\theta}_{\text{INT}}^\text{MM} \geq \max_{O} \frac{1}{N^2} \sum_{i=1}^p \frac{\pi^2}{\lambda^2[(O^T \Lambda)_i]}. \]

(53)

An analogous bound may be derived for the trace of the inverse of the QFI (see also [26]):

\[ \text{Tr}(F^{-1}) = \sum_{i=1}^p |F^{-1}|_{ii} \geq \sum_{i=1}^p |F_{i-i}|_{ii} \geq \sum_{i=1}^p |F_{i-i}|_{ii} = \sum_{i=1}^p \frac{1}{\lambda^2[A_i]} \]

(54)

so

\[ \text{Tr}(F^{-1}) \geq \max_{O} \frac{1}{n^2} \sum_{i=1}^p \frac{1}{\lambda^2[(O^T \Lambda)_i]} \]

(55)

and:

\[ \Delta^2 \hat{\theta}_{\text{INT}}^\text{CR} \geq \max_{O} \frac{1}{kn^2} \sum_{i=1}^p \frac{1}{\lambda^2[(O^T \Lambda)_i]}, \]

(56)

where the same \( O \) maximizes both Eq. (53) and Eq. (56).

C. Relation between optimal global- and local-minimax costs

It is worth mentioning, that in the cases, where the local estimation problem may be extended to the covariant group estimation problem, the local minimax cost is the same as the one for estimating a completely unknown element of the group. Let us formalize it.

Let \( U_{\theta} \) (where \( g \in G \)) be a unitary representation of compact group \( G \) (so \( U_{\theta_1} U_{\theta_2} = U_{\theta_1+\theta_2} \)) and consider the cost function invariant with respect to the action of this group \( \forall g,h,h \in G \text{cost}(h g, h \hat{g}) = \text{cost}(g, \hat{g}) \). Let \( \theta = [\theta_1, \ldots, \theta_p] \) be a local apametrization around neutral element of the group \( e \in G \) such that \( U_{\theta_0} = e^{i\theta_0 \Lambda} \). Then assuming that \( \text{cost}(e, g_{\theta}) = \theta^2 + o(\theta \hat{\theta}) \), the asymptotic minimax cost for \( \text{cost}(g, \hat{g}) \) for \( g \in G \) is the same as the local one Eq. (36); see Appx. C for more details. The reasoning is based on the same idea as the one performed for single parameter case [27] (reminded here in section III A).

Moreover, in [82] it was shown that for covariant estimation the optimal results may be obtained within a parallel scheme (without the necessity of involving adaptiveness), which implies that also in the local minimax approach there is no advantage in applying adaptive strategy (in contrast to the results obtainable within CR formalism, which will be discussed in IV C).
D. Separate strategy lower bound

Finally, we would like to derive a simple bound for the minimal cost obtainable by the SEP+ strategy, which will allow us for a quick assessment of potential benefits due to a rotation in the parameter space. We have:

\[
\Delta^2 \theta_{\text{SEP+}}^{\text{MM}} \geq \min_{A,Nj} \sum_{j=1}^{p} \frac{\pi^2}{N^2} \frac{[A^T A]_{jj}}{\lambda^2([A^T A]_{jj})} \\
\geq \min_{A,Nj} \sum_{j=1}^{p} \frac{\pi^2}{N^2} \left( \min_i \frac{[A^T A]_{ii}}{\lambda^2([A^T A]_{ii})} \right) = \frac{p^3 \pi^2}{N^2} \min_{A,i} \frac{[A^T A]_{ii}}{\lambda^2([A^T A]_{ii})},
\]

(57)

Note, that the last term in above inequality depends only on the \(i^{th}\) column of \(A\). Therefore, minimization over both \(A\) and \(i\) is equivalent to minimization over a single vector \(a\):

\[
\min_{A,i} \frac{[A^T A]_{ii}}{\lambda^2([A^T A]_{ii})} = \min_{\alpha} \frac{|\alpha|^2}{\lambda^2(\alpha^2)} = \min_{\alpha:|\alpha|^2=1} \frac{1}{\lambda^2(\alpha^2)} = \frac{1}{\max_{\alpha:|\alpha|^2=1} \lambda^2(\alpha^2)},
\]

(58)

hence

\[
\Delta^2 \theta_{\text{SEP+}}^{\text{MM}} \geq \frac{p^3 \pi^2}{N^2} \frac{1}{\max_{\alpha:|\alpha|^2=1} \lambda^2(\alpha^2)}.
\]

(59)

Similarly, in the multiple repetition scenario we could write (see also [26]):

\[
\Delta^2 \theta_{\text{SEP+}}^{\text{CR}} \geq \frac{p^2 \pi^2}{N^2} \frac{1}{\max_{\alpha:|\alpha|^2=1} \lambda^2(\alpha^2)}.
\]

(60)

Thanks to the bounds derived in this section, we will be able to get an insight into the benefits of joint vs. separate strategies, even if we will not always be able to obtain a rigorous solution for the optimal achievable cost—see the next section.

IV. EXAMPLES

We will focus here on models which are inspired by various magnetic field sensing problems, but which are representative for a wide range of multiparameter unitary estimation problems, see Fig. 2. As shown in [61], if all the generators mutually commute \(V_i[A_i, A_j] = 0\), there is no asymptotic advantage (for large \(N\)) in using a general adaptive strategy when compared to the parallel one. Therefore when analyzing the first two examples, we will focus on the parallel scheme only. For the last one, where the generators do not commute, both strategies will be discussed.

![Figure 2](image-url)

**FIG. 2.** Various models of magnetic field sensing by spin-1/2 atoms are discussed. In (a) and (b) the spatial distribution of magnetic field oriented in the \(z\) direction is to be estimated. In (a) the atoms are uniformly distributed in \(p\) points and one has the freedom to choose their spin orientations in the optimal way, while in (b) for every single atom both the position and the spin orientation may be chosen arbitrary. In (c) the magnetic field in single point is measured, but multi-component estimation is discussed.

A. Spatially distributed magnetic field sensing—fixed atoms positions

Consider the problem of spatially distributed magnetic field sensing (which is directed along the \(z\) axis). The field is sensed by spin-1/2 atoms allocated in \(p\) spatially separated places.

Before moving on to the general solution of such a problem, we would like first to discuss it with an additional constraint imposed. Namely, we assume that the spatial distribution of sensing atoms is uniform and fixed, i.e. in each of \(p\) points there are exactly \(N\) atoms and one has only the freedom in choosing their spins orientations (this form of the problem was discussed in the many repetition scenario in [26]). Here, by an elementary amount of resources we understand a single layer of \(p\)-atoms, and the corresponding Hilbert space is the \(2^p\)-dimensional one spanned by the vectors of the form:

\[
|s\rangle = |s_1, s_2, ..., s_p\rangle, \quad \text{with} \quad s_i \in \{-1, +1\}.
\]

(61)

Then the elementary quantum gate is:

\[
U_\theta = e^{-i\theta A}, \quad \Lambda_i = \mathbb{1} \otimes s_i^{-1} \otimes \frac{1}{2} \sigma_z \otimes \mathbb{1}^{p-i},
\]

(62)

so \(\Lambda_i |s\rangle = \frac{1}{2} s_i |s\rangle\). Note that here the number of quantum
gates \(N\) is equal to number of \(p\)-atoms layers and in fact correspond to usage of \(N \cdot p\) atoms.

Let us consider first the SEP strategy (without optimization over reparametrizations). Since \(\forall \lambda \in \Lambda\) one gets respectively: for the CR approach (using the analogues of \([n00n]\) Eq. (41) states):

\[
\Delta^2 \theta_{\text{SEP}}^{\text{CR}} \simeq p \times \frac{1}{k/p} \times \frac{1}{n^2} = \frac{p^2}{kn^2},
\]

and for the MM approach (with the use of \([\text{SIN}]\) Eq. (43) states):

\[
\Delta^2 \theta_{\text{SEP}}^{\text{MM}} \simeq p \times \frac{\pi^2}{(N/p)^2} = \frac{p^2 \pi^2}{N^2}.
\]

However, looking at Eq. (59) and Eq. (60) one may see that there is a significant potential for improvement in SEP+, as the value of \(\lambda[a\Lambda]\) will be maximal for \(a = 1/\sqrt{p}[1, \ldots, 1]\) and equal to \(\sqrt{p}\). Below we show a concrete reparametrization for which the mentioned bounds may be saturated. For simplicity, let us restrict to the case \(p = 2^r\) for some natural \(r\).

In order to optimize the separate strategy one needs to find a parametrization for which each parameter may be sensed by all the atoms simultaneously. Therefore, instead of measuring the magnetic field point by point in \(p\) positions, one may decompose the field into a proper components by the Walsh–Hadamard transformation \(\theta' = O^{-1} \theta\) with:

\[
O_{ij} = \frac{1}{\sqrt{p}} \prod_{k=0}^{r-1} (-1)^{i_k s_k}, \quad \text{where} \quad i = \sum_{k=0}^{r-1} i_k 2^k,
\]

or, equivalently:

\[
O = \frac{1}{\sqrt{p}} \left[ \begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array} \right] \otimes^p .
\]

After the application of such a transformation, all the generators \([O^T \Lambda]_i\) remain diagonal in the basis \(\{|s\}\) and moreover:

\[
\langle s | [O^T \Lambda]_i | s \rangle = \frac{1}{2\sqrt{p}} \sum_{j=1}^p O_{ij} s_j.
\]

From that, indeed, \(\forall \lambda ([O^T \Lambda]_i) = \sqrt{p}\) and, moreover, as the eigenvectors of \(\lambda_i\) with minimal and maximal eigenvalues take the form:

\[
|\lambda_i^+\rangle = |+\rangle_{i_1} \otimes |+\rangle_{i_2} \ldots \otimes |+\rangle_{i_p},
\]

\[
|\lambda_i^-\rangle = |-\rangle_{i_1} \otimes |-\rangle_{i_2} \ldots \otimes |-\rangle_{i_p}
\]

the remaining generators act on them trivially:

\[
\Lambda_j |\lambda_{i\pm}\rangle = \pm \delta_{ij} \frac{1}{2} |\lambda_{i\pm}\rangle.
\]

Hence, when focusing on the estimation of a given parameter \(\theta_i\), there are no disturbance issues related with the presence of the other parameters. Therefore:

\[
\Delta^2 \theta_{\text{SEP}}^{\text{CR}} \simeq \frac{p}{kn^2}, \quad \Delta^2 \theta_{\text{SEP}}^{\text{MM}} \simeq \frac{p^2 \pi^2}{N^2}.
\]

We see that, thanks to the application of a proper reparametrization, we have decreased the cost obtainable in a separate strategy by a factor of \(p\). Note, that in order to use Eq. (65) we have assumed \(p = 2^r\). However, if this is not satisfied, one can still obtain qualitatively similar results, e.g. applying parameters transformation from \([83]\).

Going back to the initial parametrization, let us now discuss a joint strategy. As each of the parameters is associated with a different atom, all of them may be measured without disturbing the measurement outcomes of the remaining ones. More formally, as the Hilbert space corresponding to the single layer of atoms has the characteristic structure \(\mathcal{H} = (\mathbb{C}^2)^\otimes N\), then, for \(N\) layers it may be written in the form \(\mathcal{H}^{\otimes N} = ((\mathbb{C}^2)^\otimes p)^{\otimes N}\). Hence, we may use the \([n00n]^{\otimes p}\) state (in the CR approach) or \([\text{SIN}]^{\otimes p}\) state (in the MM approach), which yields:

\[
\Delta^2 \theta_{\text{NT}}^{\text{CR}} \simeq \frac{p}{kn^2}, \quad \Delta^2 \theta_{\text{NT}}^{\text{MM}} \simeq \frac{p^2 \pi^2}{N^2},
\]

and which satures the bound Eq. (53) (which is the tightest for the original parametrization, i.e. with \(O = \mathds{1}\)).

Comparing Eq. (70) with Eq. (71), we see that for this model the existence of the advantage of measuring parameters jointly depends on the chosen paradigm—in the many repetition scenario there is no advantage, while for the fully optimal usage of all resources the advantage increases linearly with the number of parameters.

**B. Spatially distributed magnetic field sensing—arbitrary spatial distribution of atoms**

Let us now consider the same problem with the full freedom in the distribution of the atoms in both space \((p\) positions\) and spin orientations. The single atom Hilbert space will, therefore, be spanned by:

\[
|i, s\rangle, \quad \text{with} \quad i \in \{1, 2, \ldots, p\}, \quad s \in \{-1, +1\}.
\]

The corresponding single quantum gate has the form:

\[
U_{\theta} = e^{i\theta \Lambda}, \quad \Lambda_i = |i\rangle \langle i| \otimes \frac{1}{2} \sigma_z = \frac{1}{2} |i, +\rangle \langle i, +| - |i, -\rangle \langle i, -|,
\]

so \(\Lambda_i |j, s\rangle = \delta_{ij} \frac{1}{2} |j, s\rangle\). Note that here, unlike in the previous example, the number of quantum gates is equal exactly to the number of atoms used.

Similarly as in Eq. (43), without loss we may restrict to the fully symmetric space, so that any \(N\)-atomic output state may be written as:

\[
|\psi^N_{\theta}\rangle = \sum_{|m| = N} e^{\frac{i}{2} \sum \theta_{i}(m_{i+} - m_{i-})} c_m |m\rangle,
\]

where \(m = [m_{i+}, m_{i-}, \ldots, m_{p-}]\) and \(|m| = \sum_{i=1}^p (m_{i+} + m_{i-})\). In this case \(\forall \lambda |\Lambda| = 1\)—their non-zero eigenspaces are mutually orthogonal and hence cannot be increased by taking
any linear combination $a\Delta$. Therefore, there is no advantage of the SEP+ over the SEP protocol, and we simply get:

$$\Delta^2 \theta_{\text{SEP}}^{\text{CR}} \simeq \frac{p^2}{k n^2}$$  \hspace{1cm} (75)

for the state:

$$|\psi_{\theta}^{\text{SEP}}\rangle = \frac{1}{\sqrt{2}} \left( e^{i \theta_1/2} |n\rangle |0\rangle + e^{-i \theta_1/2} |0\rangle |n\rangle \right),$$  \hspace{1cm} (76)

while in case of full optimization paradigm we get

$$\Delta^2 \theta_{\text{SEP}}^{\text{MM}} \simeq \frac{p^3 \pi^2}{N^2}$$  \hspace{1cm} (77)

for the state:

$$|\psi_{\theta}^{\text{SEP}}\rangle = \sum_{m=0}^{N/p} \sqrt{2} e^{i (2m - N/p) / 2\theta} \sin \left( \frac{(m + 1) \pi}{N/p} \right) |m\rangle |N/p - m\rangle.$$  \hspace{1cm} (78)

where $|m\rangle |N/p - m\rangle$ denotes a state with $N/p$ atoms in the $i$-th position, with $m$ of them being oriented up, and $N/p - m$ down.

Much more interesting aspects may be observed when analyzing the joint strategy. First, let us look what insight may be obtained from the bounds Eq. (56) and Eq. (53). To do so, we will use again the Walsh-Hadamard transformation. However, note that in this case the effect is completely opposite to the one observed in the previous example—application of this transformation decreases all $\lambda_i \{O^T \Lambda_i\} = 1/\sqrt{p}$—that is, in the case where one has the full freedom of distributing atoms in space, measuring combinations of magnetic fields at various points, instead of measuring the field point by point, is an inefficient separate strategy. Still, thanks to this observation, such a transformation may be used to tighten the bounds Eq. (56), Eq. (53). Since $\sum_{i=1}^{p} 1/\lambda^2 \{O^T \Lambda_i\} = p^2$, it gives:

$$\Delta^2 \theta_{\text{NT}}^{\text{CR}} \gtrsim \frac{p^2}{k n^2}, \quad \Delta^2 \theta_{\text{NT}}^{\text{MM}} \gtrsim \frac{p^2}{N^2}.$$  \hspace{1cm} (79)

In the CR case, the above bound may be saturated using the state:

$$|\psi_{\theta}^{\text{CR}}\rangle = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} \frac{1}{2} \left( e^{i\theta_1/2} |n\rangle |0\rangle + e^{-i \theta_1/2} |0\rangle |n\rangle \right),$$  \hspace{1cm} (80)

which is simply an equally weighted superposition of Eq. (76) and for which:

$$F_{ij} = 4 \text{Re} \left[ \langle \psi_i | \psi_j \rangle - \langle \psi_i | \psi \rangle \langle \psi | \psi_j \rangle \right] \Rightarrow F_{ij} = \frac{\delta_{ij}}{p},$$  \hspace{1cm} (81)

and consequently

$$\text{Tr}(F^{-1}) = \frac{p^2}{n^2} \Rightarrow \Delta^2 \theta_{\text{NT}}^{\text{CR}} \simeq \frac{p^2}{k n^2}.$$  \hspace{1cm} (82)

To analyze the joint strategy within the MM approach, we will use the Fourier analysis, originally applied to the single parameter estimation problem in [75], later generalized for another group estimation problems [59], and very recently applied for multiphase estimation in multiarm interferometer with constraining for the total number of photons [61]. Note that from the point of view of the discussed problem, only the differences between the number of atoms oriented up and down $\Delta n_i = (m_{i+} - m_{i-})/2$ matter. Therefore, for a given state Eq. (74) we define normalized states $|\Delta m\rangle$ and coefficients $c_{\Delta m}$ satisfying:

$$\forall \Delta m c_{\Delta m} |\Delta m\rangle = \sum_{m : \Delta n_i = 2m} \Delta m \langle m | \Delta m\rangle.$$  \hspace{1cm} (83)

Then Eq. (74) may be rewritten in the form:

$$|\psi_{\theta}^{N}\rangle = \sum_{|\Delta m| \leq N/2} \sum_{m : \Delta n_i = 2m} c_{\Delta m} |\Delta m\rangle.$$  \hspace{1cm} (84)

Next, for $N$ large enough, we may replace discrete variables by continuous ones $\Delta m / N \to \mu_i \in [-1/2, 1/2]$ to get:

$$|\psi_{f, \theta}^{N}\rangle = \int d\mu \cos \theta \mu f(\mu) |\mu\rangle.$$  \hspace{1cm} (85)

As argued in [61], the optimal measurement in the asymptotic limit will be the covariant one:

$$|\chi_{\theta}\rangle = \frac{1}{\sqrt{(2\pi/N)^p}} \int d\mu e^{i N \mu \theta} |\mu\rangle.$$  \hspace{1cm} (86)

For technical reasons, in further calculations we will treat the function $f(\mu)$ appearing in Eq. (85) as the one defined on the whole $\mathbb{R}^p$, but equal zero everywhere outside of $\{ |\mu_1, \ldots, \mu_p| \mid \sum_i |\mu_i| \leq 1/2 \}$ (which allows us to perform the standard Fourier transform of this function). The mean value of the quadratic cost is given by:

$$\int_{\mathbb{R}^p} d\theta \ |\langle \chi_{\theta} | \psi_{f, \theta}^{N}\rangle|^2 \langle \theta - \theta \rangle^2 = \frac{1}{N^2} \int_{\mathbb{R}^p} d\theta \ |\langle f(\theta) | \theta \rangle|^2 \hat{\theta}^2,$$  \hspace{1cm} (87)

where $\hat{\theta}$ is the Fourier transform of $\theta$ and we dropped the irrelevant dependence on $\theta$.

Going back to the $\mu$-representation and performing the minimization over $\theta$ we get:

$$\Delta^2 \theta_{\text{NT}}^{\text{MM}} \simeq \frac{1}{N^2} \min_{f} \int_{\mu_1, \ldots, \mu_p} d\mu f(\mu) \sum_{i=1}^{p} \hat{\theta}^2 \mu_i^2$$

$$\Rightarrow \int_{\mu_1, \ldots, \mu_p} d\mu |f(\mu)|^2 = 1,$$  \hspace{1cm} (88)

$$f(\mu) = 0 \quad \text{for} \mu \text{ on the boundary} \sum_i |\mu_i| = 1/2.$$
of a $p$-dimensional simplex. The analytical solutions are known only for $p = 1, 2$ (see Appx. D). For higher number of parameters we will derive a lower bound. As $f(\mu) = 0$ everywhere outside of $\sum_i |\mu_i| \leq 1/2$, the mean value of $\sum_i |\mu_i|$ is trivially smaller or equal $1/2$. Next, thanks to the symmetry, we may assume without loss for generality, that the function $f(\mu)$ minimizing the above is fully symmetric under the exchange of variables $\mu_i$ and therefore all the mean values of $|\mu_i|$ are equal and $\leq 1/(2p)$ (see also supplementary materials of [61] for a broader discussion of this argument). The minimal sum of variances may be therefore be bounded as:

$$\Delta^2 \theta_{\text{MM}} \geq p \times \min_g \int_{-\infty}^{\infty} d\mu g^*(\mu) \left( -\frac{\partial^2}{\partial^2 \mu} \right) g(\mu) \tag{89}$$

with

$$\int_{-\infty}^{\infty} d\mu |g(\mu)|^2 = 1, \quad \int_{-\infty}^{\infty} d\mu |g(\mu)|^2 |\mu| = \frac{1}{2p}. \tag{90}$$

Moreover, from the symmetry of the problem, the solution will be symmetric with respect to the 0 point (which, assuming differentiability, implies $\frac{\partial g}{\partial \mu} \big|_{\mu=0} = 0$). Therefore, the problem is equivalent to:

$$\min_g \int_0^{\infty} d\mu g^*(\mu) \left( -\frac{\partial^2}{\partial^2 \mu} \right) g(\mu), \quad \text{with} \quad \frac{\partial g}{\partial \mu} \big|_{\mu=0} = 0,$n$$

$$\int_0^{\infty} d\mu |g(\mu)|^2 = 1, \quad \int_0^{\infty} d\mu |g(\mu)|^2 |\mu| = \frac{1}{2p}. \tag{91}$$

which may be solved using the Lagrange multipliers method. The solution in terms of the Airy function yields the final bound (see Appx. D for detail derivation):

$$\Delta^2 \theta_{\text{MM}} \geq 0.63 p^3 N^2. \tag{92}$$

We are unable to proof the tightness of the above bound. However, we are able to point out an exemplary state for which the cost closely approaches the bound. Consider the largest possible $p$-dimensional ball inside the simplex $\sum_i |\mu_i| \leq 1/2$ and then as $f(\mu)$ choose the function which minimizes the kinetic energy inside this ball with the boundary condition $f(\mu) = 0$ on the border and outside of the ball. The cost corresponding to this construction, the bound Eq. (92), the values of analytical solution of Eq. (88) for $p = 1, 2$ as well as $\Delta^2 \theta_{\text{SEP}}^{\text{MM}}$ are plotted together in Fig. 3—details of the calculations may be found in Appx. D. For large $p$, the total cost corresponding to the described strategy leads to $p^3/N^2$, and we finally get:

$$\Delta^2 \theta_{\text{INT}}^{\text{MM}} \simeq \frac{cp^3}{N^2}, \quad 0.63 \leq c_1 \leq 1. \tag{93}$$

When comparing to Eq. (79), we see that in the MM scenario not only the bound is not tight, by it even fails to properly predict the scaling of the cost with the number of parameter $p$.

To summarize, for the problem of estimation of the spatially distributed single component magnetic field, with full freedom in choosing both the position and orientation of the atoms, for large $p$ we obtain:

$$\Delta^2 \theta_{\text{SEP}}^{\text{CR}} \simeq \frac{p^2}{kn^2}, \quad \Delta^2 \theta_{\text{SEP}}^{\text{MM}} \simeq \frac{\pi^2 p^3}{N^2},$$n

$$\Delta^2 \theta_{\text{CR}}^{\text{INT}} \simeq \frac{p^3}{kn^2}, \quad \Delta^2 \theta_{\text{MM}}^{\text{INT}} \simeq \frac{\pi^2 p^3}{N^2}. \tag{94}$$

If one wants to compare these results with the previous example (where the positions of the atoms were fixed), one should bare in mind, that in the previous example one gate corresponded to $p$ atoms, not one. Therefore, in order to make the comparison fair, one should rewrite Eq. (70), Eq. (71) in terms of the atoms used $N_a = p \cdot N$, $n_a = p \cdot n$, which gives:

$$\Delta^2 \theta_{\text{SEP}}^{\text{CR}} \simeq \frac{p^3}{kn_n^2}, \quad \Delta^2 \theta_{\text{SEP}}^{\text{MM}} \simeq \frac{\pi^2 p^3}{N_{a_n}^2},$$n

$$\Delta^2 \theta_{\text{CR}}^{\text{INT}} \simeq \frac{p^3}{kn_n^2}, \quad \Delta^2 \theta_{\text{MM}}^{\text{INT}} \simeq \frac{\pi^2 p^3}{N_{a_n}^2}. \tag{95}$$

Then it is clear, that for the same amount of atoms used, all the costs from Eq. (95) are larger than the corresponding ones from Eq. (94), as in Eq. (95) fewer degrees of freedom are allowed.

Finally, one may notice that the problem of estimating magnetic field in $p$ points by $N$ atoms (with full freedom in choosing their spin and position) is equivalent to a slightly modified problem of multi-phase estimation discussed in [61]. Indeed, treating the points in space as arms of an interferometer and atoms as photons, we may think about this problem in terms of a $2p$-arm interferometer with phase shifts $\pm \frac{1}{2} \theta_i$ in each arm. Therefore, it makes sense to compare Eq. (94) with the analogue costs obtained in [61] for the problem of estimating $p$ unknown phase-shifts in a $(p + 1)$-arm interferometer (where
the one arm is the reference arm)—both version are schematically presented in Fig. 4. In the latter case, the single photon Hilbert space is spanned by \{\{0\}, \{1\}, \ldots, \{p\}\} (where \{0\} corresponds to the reference arm, and \{i\} to the sensing arms), and the quantum gate is \(U_\theta = \exp(i\sum_{i=1}^p \theta_i |i\rangle \langle i|)\). For large \(p\):

\[
\begin{align*}
\Delta^2 \theta_{\text{SEP}}^\text{CR} &\simeq \frac{p^2}{k N^2}, \\
\Delta^2 \theta_{\text{SEP}}^\text{MM} &\simeq \frac{\pi^2 p^3}{N^2}, \\
\Delta^2 \theta_{\text{INT}}^\text{CR} &\simeq \frac{p^2}{4k N^2}, \\
\Delta^2 \theta_{\text{INT}}^\text{MM} &\simeq \frac{c_2 p^3}{N^2}.
\end{align*}
\]

(96)

Comparing Eq. (94) with Eq. (96), we may notice significant differences. While in the \((p + 1)\)-arm interferometer, the problem in both the CR and the MM approach, reveals a constant advantage of JNT over SEP strategies of a similar order (around 5), for Eq. (94) no advantage is observed in the CR approach, while the one in the MM approach is even larger (around 10).

This effect, however, may be easily understood after analyzing the structure of the states optimal for measuring parameter separately—Eq. (76) and Eq. (78). While in the case studied in [61], in order to estimate any of the parameters one needed to use the same reference arm, in the presently studied model each phase may be measured using two dedicated arms with phase shifts \(\pm \theta_i/2\) and \(\pm \theta_i/2\). Consequently, the corresponding \(n00n\) states Eq. (76) are mutually orthogonal—the resources “consumed” by one parameter cannot be used to estimate others. On the other hand, looking at [SIN] state Eq. (78) we see that the greatest weights are attached to the vectors with relatively small differences \(m_{i+} - m_{i-} = 0\). As there is no problem in distributing photons (or atoms) in such a way that this difference is small for all \(i\), such a component may be used in estimating all the parameters simultaneously, which is responsible for the significant advantage of the joint strategy in this case.

### C. Multicomponent magnetic field estimation

As the last problem let us discuss the canonical example of estimation of parameters associated with non-commuting generators. More specifically, we focus on the problem of estimating the three components of a magnetic field vector \(\theta = [\theta_1, \theta_2, \theta_3]^T\) in a given point in space using spin-1/2 atoms. Single atom Hilbert space is therefore simply a qubit space and the corresponding quantum gate reads:

\[
U_\theta = e^{i\theta\sigma/2},
\]

(97)

where \(\sigma = [\sigma_1, \sigma_2, \sigma_3]^T\) is the vector of Pauli matrices. Unlike in the previously discussed examples, here the minimal achievable cost depends on the actual values of the parameters \(\theta\) [46]. For simplicity, let us focus on estimation around point \(\theta_0 = [0, 0, 0]^T\).

Note that for any normalized vector \(a\), the operator \(a \cdot \sigma\) has the same eigenvalues. Hence, invoking from Eq. (59), Eq. (60) we see that a rotation in the parameter space cannot improve the precision in a separate protocol—there is no advantage in SEP+.

Let us start by discussing the many repetition scenario. In a separate strategy, each component may be measured with a proper \(n00n\) states \(\frac{1}{\sqrt{2}}(|+\rangle^{n}_{x,y,z} + |−\rangle^{n}_{x,y,z})\), which leads to:

\[
\Delta^2 \theta_{\text{SEP}}^\text{CR} \simeq 3 \times \frac{1}{k N^2} \times \frac{1}{n^2} \approx \frac{9}{k n^2}.
\]

(98)

For the joint strategy in turns out that, unlike in the previous discussed examples, application of the adaptive scheme with ancillas allows to beat the performance of the optimal parallel strategy. We will, therefore, discuss these two strategies independently.

The minimal trace of inverse of the QFI achievable for parallel scheme may be found analytically [36] (for \(n \geq 6\): \(\text{Tr}(F^{-1}_{\text{parallel}}) = \frac{9}{n(n+2)}\) (and condition Eq. (23) is satisfied), so

\[
\Delta^2 \theta_{\text{INT}}^\text{CRparallel} \simeq \frac{9}{k n(n + 2)} \approx \frac{9}{k n^2},
\]

(99)

which for large \(n\) is almost the same as for the separate strategy. Therefore, the advantage offered by joint measurement in the parallel strategy disappears with increasing \(n\).

In contrast to the above, it was shown in [45], that for the adaptive ancilla-assisted sequential scheme utilizing as an input state

\[
|\psi_{\text{in}}\rangle = \frac{1}{\sqrt{2}}(|+\rangle_z |0\rangle_A + |−\rangle_z |1\rangle_A)
\]

(100)

(where \(|0\rangle_A, |1\rangle_A\) belongs to ancillary system) and acting on in by the gate \(n\) times, one may obtain \(\text{Tr}(F^{-1}_{\text{adaptive}}) = \frac{n}{\pi^2}\) (satisfying Eq. (23)), and hence

\[
\Delta^2 \theta_{\text{INTadaptive}} \simeq \frac{3}{k n^2},
\]

(101)

which saturates Eq. (53). Therefore, we see that in the many repetition scenario, the possibility of acting sequentially is
crucially needed to take the advantage from the joint estimation approach as it allows to decrease the final variance by a factor 3, compared to the optimal separate or joint parallel strategy.

Consider now the fully optimal usage of $N$ gates. Note that, unlike the previously discussed examples, here the existence of unknown parameters $\theta_{j\neq i}$ may significantly impede estimation of $\theta_i$, as it may be impossible to find an initial state, for which the evolution would depend only on $\theta_i$. As the consequence, we can only write the lower bound:

$$\Delta^2\theta_{\text{SEP}} \gtrsim 3 \times \frac{\pi^2}{(N/3)^2} = \frac{27\pi^2}{N^2}. \quad (102)$$

Still, it is not obvious how to obtain such a precision while measuring parameters separately.

To calculate the asymptotically optimal cost in joint estimation we use the fact, that the problem may be extended to the covariant one, which allows us to use the reasoning from III C, stating that optimal asymptotical local minimax cost is the same as the global one, obtainable within the parallel scheme. In [55, 57], the estimation of a completely unknown element of $SU(2)$ within the parallel strategy was discussed with the covariant cost

$$e(\theta, \tilde{\theta}) = 6 - 2\text{Tr}(U^{(1)}_{\theta}U^{(-1)}_{\tilde{\theta}}) |\theta|, |\tilde{\theta}| \ll 1 \approx 2|\theta - \tilde{\theta}|^2, \quad (103)$$

where $U^{(1)}_{\theta}$ is a rotation matrix of a spin-1 particle. It was shown that the asymptotic minimal cost is $e(\theta, \tilde{\theta}) \simeq 8\pi^2/N^2$, which is achievable using initial state:

$$|\psi_{\text{m}}\rangle = \sqrt{\frac{2}{N/2 + 1}} \sum_{j=0}^{N/2-1} \sin\left(\frac{(j+1)\pi}{N/2+1}\right) \left(\sum_{\alpha=1}^{2j+1} |j\alpha, m_j = \alpha\rangle \right), \quad (104)$$

where $|j\alpha, m_j = \alpha\rangle$ are states with a well defined total angular momentum $j$ and its projection onto the $z$ direction, while $\alpha$ numerates different subspaces corresponding to equivalent irreducible representations of $SU(2)$. Therefore, including the factor 2 that appears in Eq. (103) we get:

$$\Delta^2\theta_{\text{MM}} \simeq \frac{4\pi^2}{N^2}, \quad (105)$$

which shows a significant advantage over the optimal separate strategy. Based on the discussion in III C and [82], this result, which is already obtainable within a parallel scheme, cannot be improved by applying any adaptive strategy. This is in stark contrast to many repetition scenario, where a parallel joint strategy offers no asymptotical advantage with respect to a separate one, while adaptiveness allows decreasing the cost by a factor of 3. It is also worth noting that the result Eq. (105) may be also obtained for parallel strategies with the usage of Fourier analysis, as shown in [59, Section 12].

Analogous reasoning may be performed in a situation, when one component of the magnetic field is known to be zero, and only the two remaining components are being estimated. For the separate strategy we get:

$$\Delta^2\theta_{\text{SEP}}^{\text{CR}} \approx \frac{4}{k^2h^2}, \quad \Delta^2\theta_{\text{SEP}}^{\text{MM}} \gtrsim \frac{8\pi^2}{N^2}, \quad (106)$$

for the joint CR:

$$\Delta^2\theta_{\text{INT}^{\text{PAR}} \text{parallel}}^{\text{CR}} \approx \frac{4}{k^2n(n + 2)}, \quad \Delta^2\theta_{\text{INT}^{\text{Adaptive}}}^{\text{CR}} \approx \frac{2}{k^2h^2}, \quad (107)$$

while for the joint MM:

$$\Delta^2\theta_{\text{INT}^{\text{MM}}} \approx \frac{4\pi^2}{N^2} \approx \frac{2.34\pi^2}{N^2}, \quad (108)$$

where the parallel strategy obtaining above was found in [53, 54] and $\xi \approx 2.4048$ is first zero of the Bessel function $J_0(x)$. Eq. (108), similarly like Eq. (105), is valid for both parallel and sequential adaptive strategies.

If, on the other hand, the direction of the magnetic field is known and only the length of the magnetic vector is to be estimated, the problem is equivalent to single phase estimation problem discussed before.

A natural extension of all the above considerations would be to combine all the examples, and consider the most general problem of estimating all the 3 components of a spatially distributed magnetic field. Based on the analysis performed we expect no improvement in the CR approach and some constant improvement in the MM approach. The strict analysis of this problem, however, is beyond of scope of this paper.

V. CONCLUSIONS

The examples studied, demonstrate that, unlike in the single unitary parameter case, in a multiparameter estimation problem there is no simple correspondence between the results obtained within the many repetition paradigms and the one, where all the resources are accumulated in single experimental realization. In the case, where the total amount of resources is limited (no matter how large), the analysis based only on the QFI is not sufficient to draw not only quantitative but even qualitative conclusions. The presented examples showed, that such an approach tend to overrate the performance of SEP/SEP+ strategies and this opens up the possibility that certain joint estimation metrological strategies may offer a significant advantage, even if this is not apparent in a formalism based on the QFI.

ACKNOWLEDGMENTS

We thank Howard M. Wiseman for many fruitful discussions. This work was supported by the National Science Center (Poland) Grant No. 2020/37/B/ST2/02134. Wojciech Górecki was also supported by the Foundation for Polish Science (FNP) via the START scholarship.
Appendix A: Different approach to the identification of the advantage of joint estimation protocols

Alternatively to the approaches described in the main text, the issue of the potential gain coming from measuring multiple parameters simultaneously versus measuring them separately may be explored abstracting from the problem of optimal division of resources, but instead by analyzing the so called probe incompatibility [43, 44]. In this approach the minimal cost achievable in a joint strategy is compared with the one coming from measuring each parameter individually, but with the assumption, that in the latter case for each parameter one spends the same amount of resources as in the whole join strategy (so effectively in the separate strategy $p$ times more resources are consumed).

While in the current paper the potential superiority of joint measurement is discussed for a particular cost function (defined by the chosen parametrization for which it is equal to identity), one may instead try to look at the problem as the feature of the channel itself. To do so, in [43] the following quantity was introduced:  

$$J^* = \max_{\{w_i\}} \left( \min_{(n,k)\text{-protocol}} \frac{\text{Tr}(W\Sigma')}{\sum_i \min_{(n,k)\text{-protocol}} w_i^T \Sigma' w_i} \right),$$  

(A1)

where $W = \sum_i w_i w_i^T$. Note that when compared to the original notation [43], we have added a prime sign to $\Sigma'$ and $F'$, as we reserve the un-primed parametrization for the case where the cost matrix is the identity.

In order to understand this approach, in the context of Fig. 1, let us perform a reparametrization $\theta = A\theta'$ with the transformation matrix $A = [w_1, ..., w_p]^T$. Then, for any fixed set $\{w_i\}$ we have:

$$\frac{\min_{(n,k)\text{-protocol}} \text{Tr}(W\Sigma')}{\sum_i \min_{(n,k)\text{-protocol}} w_i^T \Sigma' w_i} = \frac{\min_{(n,k)\text{-protocol}} \text{Tr}(\Sigma)}{\sum_i \min_{(n,k)\text{-protocol}} \Sigma_{ii}} \equiv \frac{\Delta^2 \theta_{\text{CR}}}{\Delta^2 \theta_i}$$  

(A2)

so for a fixed $\{w_i\}$ it corresponds to a situation, where Bob alternatively sends $n \cdot k$ gates to Alice (nominator) or the same amount of gates directly to each of the Minions (denominator). In this sense the denominator is similar to the SEP strategy (but with omitted problem or resources distribution). Note, however, that the maximization over $\{w_i\}$ in Eq. (A1) does not correspond to SEP+ (as here also the value in the nominator changes while for the reparametrization of $A$ in SEP+ it does not)—it should be rather seen as maximization over all possible Bob’s initial parametrizations.

Appendix B: Necessity of a non-orthogonal transformation in order to obtain the minimal SEP+ cost—example

Here we discuss an exemplary two-parameter estimation problem, for which in order to obtain the minimal cost in SEP+ protocol we need to apply a non-orthogonal transformation $A$ in the parameter space. We will focus on the many repetition paradigm and the CR formalism. Consider a unitary channel:

$$U_\theta = e^{i\theta A}, \quad A_1 = \frac{1}{2} \text{diag} (+\alpha, -\alpha, +\beta, -\beta), \quad A_2 = \frac{1}{2} \text{diag} (+\beta, -\beta, +\alpha, -\alpha),$$  

(B1)

where $0 < \beta < \alpha$ and $\text{diag}(\ldots)$ is a diagonal matrix acting on the Hilbert space spanned by $|1\rangle, |2\rangle, |3\rangle, |4\rangle$, so

$$\theta \cdot A = \begin{bmatrix} +\frac{1}{2}(\alpha\theta_1 + \beta\theta_2) & 0 & 0 & 0 \\ 0 & -\frac{1}{2}(\alpha\theta_1 + \beta\theta_2) & 0 & 0 \\ 0 & 0 & +\frac{1}{2}(\beta\theta_1 + \alpha\theta_2) & 0 \\ 0 & 0 & 0 & -\frac{1}{2}(\beta\theta_1 + \alpha\theta_2) \end{bmatrix}. \quad \text{(B2)}$$

Since the generators mutually commute, the Fisher information matrix for the input $|\psi\rangle$ is given by:

$$[F]_{ij} = 4 \left( \langle \psi | \Lambda_i \Lambda_j | \psi \rangle - \langle \psi | \Lambda_i | \psi \rangle \langle \psi | \Lambda_j | \psi \rangle \right)$$  

(B3)

and the corresponding CR bound is saturable (as $\text{Im}(A_1 A_2) = 0$). By a direct calculation one can see that the cost of joint estimation $\Delta^2 \theta_{\text{CR}}$ is minimized for the state $|\psi\rangle = \frac{1}{2}(|1\rangle + |2\rangle + |3\rangle + |4\rangle)$ and that for this sate:

$$\Delta^2 \theta_{\text{CR}} \approx \frac{1}{k} \text{Tr}(F^{-1}) = \frac{1}{k} \left( \frac{2}{(\alpha - \beta)^2} + \frac{2}{(\alpha + \beta)^2} \right). \quad \text{(B4)}$$
In order to calculate \( \Delta^2 \theta_{\text{SEP}^+}^{\text{CR}} \), instead of performing a direct optimization given in Eq. (27), we just use the fact that \( \Delta^2 \theta_{\text{SEP}^+}^{\text{CR}} \geq \Delta^2 \theta_{\text{INT}}^{\text{CR}} \) and show a particular transformation \( A \) for which this bound is saturated. Let us choose as new parameters \( \theta_1' = \alpha \theta_1 + \beta \theta_2, \theta_2' = \beta \theta_1 + \alpha \theta_2, \) so

\[
A^{-1} = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix} \Rightarrow A = \frac{1}{\alpha^2 - \beta^2} \begin{bmatrix} \alpha & -\beta \\ -\beta & \alpha \end{bmatrix}.
\] (B5)

After such a transformation the differences between extreme eigenvalues of new generators are both equal to:

\[
\Delta^2 \theta_{\text{SEP}^+}^{\text{CR}} = \frac{1}{k} \sum_{i=1}^{p} \sqrt{[A^2]_{ii}}^2 = \frac{1}{k} \left( \frac{2}{(\alpha - \beta)^2} + \frac{2}{(\alpha + \beta)^2} \right),
\] (B6)

which is indeed equal to Eq. (B4). Finally, let us show that this results is not achievable, if one restricted just orthogonal transformations \( O \). Let

\[
O^{-1} = [\mathbf{o}_1, \mathbf{o}_2]^T = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}.
\] (B7)

Then \( [O^T O]_{ii} = 1 \) and \( [O^{-1} F_i^{-1} O^{-1^T}]_{ii} = \mathbf{o}_1^T F_i^{-1} \mathbf{o}_i \) so:

\[
\Delta^2 \theta_{\text{SEP}^+, \text{ortho}}^{\text{CR}} = \frac{1}{k} \min_{\varphi} \left( \sum_{k=1}^2 \sqrt{\min_{|\psi_k|} [F_i]_{k}^{-1} [\mathbf{o}_i]} \right)^2.
\] (B8)

Both Eq. (B6) and Eq. (B8) are compared in Fig. 5 for different ratios between \( \alpha \) and \( \beta \). One may see that for the ration around 1/2 a significant advantage due to application of the non-orthogonal transformation may be observed.

**Appendix C: Relation between the optimal global group invariant cost and the local quadratic minimax cost in the limit of large \( N \)**

In this appendix, we formalize the reasoning from section III C.

**Theorem.** Consider a quantum channel \( U_g, g \in G \), which corresponds to a unitary representation of a compact group \( G \) in some Hilbert space \( U_{g_1} U_{g_2} = U_{g_1 g_2} \), and the cost function is invariant with respect to the action of the group \( \forall g, \tilde{g}, h \in G \text{cost}(hg, h\tilde{g}) = \text{cost}(g, \tilde{g}) \). Let \( \theta = [\theta_1, \ldots, \theta_p] \) be a local parametrization around some \( g_0 \in G \). Let \( G_\delta \subset G \) be the subset of \( G \) containing all \( g_0 \) such that \( \forall g, \theta_1 \in [-\delta/2, +\delta/2]. \) Then, for the most general adaptive scheme, the local asymptotic minimax cost is the same as the global asymptotic minimax cost:

\[
\inf_{\{M, \delta, \rho_{N,g}\}} \lim_{\delta \to 0} \lim_{N \to \infty} N^2 \sup_{g \in G_\delta} \int d\tilde{g} \text{Tr}(M_{N, \delta} \rho_{N, g}) \text{cost}(g, \tilde{g}) = \inf_{\{M, \delta, \rho_{N,g}\}} \lim_{N \to \infty} N^2 \sup_{g \in G} \int d\tilde{g} \text{Tr}(M_{N, \delta} \rho_{N, g}) \text{cost}(g, \tilde{g})
\] (C1)
Moreover, as obtaining optimal global asymptotic minimax cost was proven not to require adaptiveness [82], the above equation remains valid when restricting the RHS to parallel strategies. It also implies, that there is no asymptotic advantage in applying adaptive strategy also in the local case.

**Proof.** Let us introduce the notation for the minimax cost with a finite \( \delta, N \):

\[
\text{minimax}(G_\delta, N) = \inf_{M_{N,\theta}} \sup_{g \in G_\delta} \int d\tilde{g} \text{Tr}(M_{\tilde{g}} \rho_{N,g}) \text{cost}(g, \tilde{g}). \tag{C2}
\]

We would like to prove that in the limit of large \( N \), the value of \( \delta \) has no impact on the final cost. It is clear that:

\[
\forall \delta \text{minimax}(G_\delta, N) \leq \text{minimax}(G, N). \tag{C3}
\]

To bound the cost for finite \( \delta \) from below, we use the following construction [27, 61]. Having at our disposal \( N \) gates in total we may at first perform \( \sqrt{N} \) independent measurements, to find an approximated value \( g_{\text{est}} \). Due to the central limit theorem, the probability, that \( g \not\in g_{\text{est}} G_\delta \) (where \( G_\delta \) is the set \( G_\delta \) shifted by the action of \( g_{\text{est}} \)) decreases exponentially [80],

\[p_{\text{err}}(\sqrt{N}) \propto e^{-\sqrt{N}}.\]

Next, we spend the remaining \( N - \sqrt{N} \) gates to perform estimation around the point \( g_{\text{est}} \). Since from the point of view of the initial problem of estimating an unknown \( g \) with \( N \) gates, such a procedure might be suboptimal, we have:

\[
\text{minimax}(G, N) \leq p_{\text{err}}(\sqrt{N}) c_{\text{max}} + (1 - p_{\text{err}}(\sqrt{N})) \text{maximize}(g_{\text{est}} G_\delta, N - \sqrt{N}), \tag{C4}
\]

where \( c_{\text{max}} = \max_{g, \delta} \text{cost}(g, \tilde{g}) \). Moreover, due to the symmetry of the whole problem, the RHS does not depend on \( g_{\text{est}} \). After application of \( \lim_{N \to \infty} N^2 \) to both sides, and use of \( \lim_{N \to \infty} \left( \frac{N - \sqrt{N}}{N} \right)^2 = 1 \), we obtain:

\[
\forall \delta \lim_{N \to \infty} N^2 \text{minimax}(G, N) \leq \lim_{N \to \infty} N^2(G_\delta, N - \sqrt{N}) = \lim_{N \to \infty} N^2(G_\delta, N), \tag{C5}
\]

which, together with Eq. (C3) gives:

\[
\forall \delta \lim_{N \to \infty} N^2 \text{minimax}(G, N) = \lim_{N \to \infty} N^2 \text{minimax}(G, N). \tag{C6}
\]

Finally, since in general for any functional family \( \lim_{x \to x_0} \inf_y f_x(y) \leq \lim_{x \to x_0} \inf_{x_0} f_x(y) \), the LHS of Eq. (C1) may be bounded from below by \( \lim_{\delta \to 0} \lim_{N \to \infty} 2^2 \text{minimax}(G_\delta, N) = \lim_{N \to \infty} N^2 \text{minimax}(G, N) \), while the RHS is exactly equal to \( \lim_{N \to \infty} N^2 \text{minimax}(G, N) \). This ends the proof. \( \square \)

It is worth to note, that if all the elements of the group representation commute \( [U_{g_1}, U_{g_2}] = 0 \), then in order to implement the strategy Eq. (C4) one does not need adaptiveness and the procedure may be performed within the parallel scheme. Indeed, in such a case there is a single state optimal for local measurements around an arbitrary point \( g_{\text{est}} \), as rotating the state is equivalent to rotating the measurement in the opposite direction \( \text{Tr}(MU_{g_{\text{est}}} U_{g_1}^{-1} U_{g_0}^{-1}) = \text{Tr}(U_{g_{\text{est}}}^{-1} MU_{g_0} U_{g_1}) \)—so the knowledge of the value of \( g_{\text{est}} \) is not needed at the level of state preparation. This is, however, no longer true if \( [U_{g_1}, U_{g_2}] \neq 0 \). Still, regardless of this sub-optimal strategy used in the proof, a fully optimal strategy does not require adaptiveness and may be performed within parallel scheme mass [82], so theorem remains valid even when restricted to parallel strategies.

**Quadratic cost approximation**

As shown in [84, Section 9] [85, Section II-D], in local estimation, analyzed within many repetition scenario, even if one considers general cost function, in the limit \( k \to \infty \) it may be well approximated by the quadratic term \( \text{Tr}(W \Sigma) = (\theta - \theta)^T W (\theta - \theta) \) (where \( W \) is the Hessian of the cost function). However, in principle, it is not so clear in the single repetition approach, as while in the many repetition scenario, due to the central limit theorem, all probabilities converge to the gaussian ones (with exponentially decreasing tails), in the single-shot case the strategy minimizing the cost may lead to much slower decreasing tails [75].

More formally, let \( W \) be the Hessian of the cost function around \( g_0 \), i.e. \( W_{ij} = \partial_i \partial_j \text{cost}(g_0, g_0) \). Then for the particular strategy \( \{ M_{N, \tilde{g}, \rho_{N, \theta}} \} \) which is known to lead to the cost \( \propto \frac{1}{\sqrt{N}} \), it is not clear if in the point \( \theta = [0, ..., 0] \):

\[
\lim_{N \to \infty} N^2 \int d\tilde{\theta} \text{Tr}(M_{N, \tilde{g}, \rho_{N, \theta}}) \text{cost}(g_0, \tilde{g}) \frac{2}{N^2} \int d\tilde{\theta} \text{Tr}(M_{N, \tilde{g}, \rho_{N, \theta}}) \sum_{ij} W_{ij} \tilde{\theta}_i \tilde{\theta}_j \tag{C7}
\]

as in principle it may happen that \( \text{Tr}(M_{N, \tilde{g}, \rho_{N, \theta}}) \) decrease like \( \frac{1}{\sqrt{N}} \) for large \( \theta \) (making also higher derivatives of \( \text{cost}(g_0, \tilde{g}) \) not negligible). Below we prove, that for the strategy minimizing LHS of above, such a tail always decreases fast enough (because if it were otherwise, one could use some small part of gates \( \sim \sqrt{N} \) to cut this tail).
Theorem. For the optimal (both adaptive or parallel) global strategy \{M_{N,g}^{opt}, \rho_{N,g}^{opt}\}, invariant under the group action, for the cost \(\text{cost}(g_0, g_\theta) = \sum_{ij} W_{ij} \hat{\theta}_i \hat{\theta}_j + o\left(\hat{\theta}_i \hat{\theta}_j\right)\) (where \(W > 0\)):

\[
\lim_{N \to \infty} N^2 \int d\hat{\theta} \text{Tr}(M_{N,g}^{opt} \rho_{N,g}^{opt}) \text{cost}(g_0, g_\hat{\theta}) = \lim_{N \to \infty} N^2 \int d\hat{\theta} \text{Tr}(M_{N,g_0}^{opt} \rho_{N,g_0}^{opt}) \sum_{ij} W_{ij} \hat{\theta}_i \hat{\theta}_j
\]  

\[\text{(C8)}\]

Proof. For any \(g\) we split the mean cost integral into two parts:

\[
\forall g \in G \int d\hat{\theta} \text{Tr}(M_{N,g}^{opt} \rho_{N,g}^{opt}) \text{cost}(g_0, g_\hat{\theta}) = \int_{G_\delta} d\hat{\theta} \text{Tr}(M_{N,g}^{opt} \rho_{N,g}^{opt}) \text{cost}(g_0, g_\hat{\theta}) + \int_{G \setminus G_\delta} d\hat{\theta} \text{Tr}(M_{N,g}^{opt} \rho_{N,g}^{opt}) \text{cost}(g_0, g_\hat{\theta}).
\]  

\[\text{(C9)}\]

Next, we prove by contradiction that \(\forall \delta > 0 \lim_{N \to \infty} N^2 C_2(\delta, N) = 0\).

Assume, that \(\lim_{N \to \infty} N^2 C_2(\delta, N) > 0\). For a given \(\delta\) let us choose finite neighborhoods of \(g_0\), namely \(G_{\Delta_1}, G_{\Delta_2}\), such that \(G_{\Delta_1} \subset G_{\Delta_2}\) and \(\forall g \in G_{\Delta_1}, G_{\Delta_2} \subset gG_\delta\), satisfying \(\max_{g \in G_{\Delta_1}} \text{cost}(g_0, g) < \min_{g \in G_{\Delta_1}, \bar{g} \in G \setminus G_{\Delta_2}} \text{cost}(g_0, g)\) (see left part of Fig. 6).

Next, similarly as in the previous proof, one may at first spend \(\sqrt{N}\) gates to find \(g_{est}\), such that the probability that the true value of \(g\) lays outside of \(g_{est}G_{\Delta_1}\), decreases exponentially \(p_{err}(\sqrt{N}) \propto e^{-\sqrt{N}}\). Then one uses the remaining \(N - \sqrt{N}\) gates to perform the mentioned \(M_{N,g_0}^{opt} \rho_{N,g_0}^{opt}\) with the following correction: each time when \(g\) points outside of \(g_{est}G_{\Delta_2}\), one is forced to estimate \(g = g_{est}\). The total cost for such a constructed strategy may be bounded from above:

\[
\leq p_{err}(\sqrt{N})c_{\max} + (1 - p_{err}(\sqrt{N}))(C_1(N - \sqrt{N}, \delta) + C_\Delta(N - \sqrt{N}, \delta))
\]  

\[\text{(C10)}\]

where \(c_{\max} = \max_{g, \bar{g} \in G} \text{cost}(g_0, g_\bar{g})\) and

\[
C_\Delta(N - \sqrt{N}, \delta) = \left(\int_{G \setminus G_\delta} d\hat{\theta} \text{Tr}(M_{N,g}^{opt} \rho_{N,g}^{opt})\right) \max_{g \in G_{\Delta_1}} \text{cost}(g_0, g).
\]  

\[\text{(C11)}\]

Since

\[
C_2(N - \sqrt{N}, \delta) \geq \left(\int_{G \setminus G_\delta} d\hat{\theta} \text{Tr}(M_{N,g}^{opt} \rho_{N,g}^{opt})\right) \min_{g \in G_{\Delta_1}, \bar{g} \in G \setminus G_{\Delta_2}} \text{cost}(g_0, g_\bar{g})
\]  

\[\text{(C12)}\]

we have \(\lim_{N \to \infty} N^2 \left(C_2(N - \sqrt{N}, \delta) - C_\Delta(N - \sqrt{N}, \delta)\right) > 0\). It means that cutting the tail indeed decreases the cost, which leads to a contradiction with the assumption about the optimality of \(\{M_{N,g}^{opt}, \rho_{N,g}^{opt}\}\). Now, as from \(C(g_0, g_\theta) = W_{ij} \hat{\theta}_i \hat{\theta}_j + o(\hat{\theta}_i \hat{\theta}_j)\) we have \(\forall \varepsilon \exists \delta_{\varepsilon, \delta/2} \leq \hat{\theta}_i \leq \delta/2, |C(g_0, g_\theta) - W_{ij} \hat{\theta}_i \hat{\theta}_j| \leq \varepsilon\), the statement is proven. \(\square\)

From the above, if \(W_{ij} = \delta_{ij}\), then for the optimal strategies the cost is equivalent to \(\Delta^2 \hat{\theta}\) discussed in the main paper. Below we show, that this is indeed the case for examples discussed in section IV.C.

In [55] the authors considered the problem of transmitting a reference frame by sending \(N\) spin-1/2 atoms, which is equivalent to the estimation of a completely unknown element of SU(2). As the figure of merit they chose the error function given as:

\[
e(g, \bar{g}) = 6 - 2\text{Tr}(U_g^{(1)} U_{\bar{g}}^{(1)})
\]  

\[\text{(C13)}\]

where \(U_g^{(1)}\) is the rotation matrix of the spin-1 particle. By a direct calculation

\[
\text{Tr}(U_g^{(1)} U_{\bar{g}}^{(1)}) = 1 + 2\cos(|\theta|),
\]  

\[\text{(C14)}\]

so

\[
e(g_0, g_\theta) = 4(1 - \cos(|\theta|)) = 8\sin^2(|\theta|/2).
\]  

\[\text{(C15)}\]

Therefore indeed we have:

\[
\text{cost}(g_0, g_\theta) = \frac{1}{2}e(0, g_\theta) = 4\sin^2(|\theta|/2) = \hat{\theta}^2 + o(\hat{\theta}_i \hat{\theta}_j).
\]  

\[\text{(C16)}\]
the solution may be found using the standard Lagrange multiplier method,

\[ f(\theta) \text{ is parameterized by the angle } 45°. \]

For the problem Eq. (91):

\[ g(\mu) = e^{i(\theta_1/2 + 2\theta_2/2)} \]

The solution is:

\[ F(\hat{n}, \hat{n}) = (1 + \hat{n} \cdot \hat{n})/2, \]

where we have:

\[ F(\hat{n}_0, \hat{n}_\theta) = (1 + \cos(|\theta|))/2 = 1 - \sin^2(|\theta|)/2, \]

so indeed:

\[ \text{cost}(\hat{n}_0, \hat{n}_\theta) = 4(1 - F(\hat{n}_0, \hat{n}_\theta)) = 4\sin^2(|\theta|)/2 = \hat{\theta}^2 + o(\hat{\theta}_i\hat{\theta}_j). \]

### Appendix D: Derivation of the bound and an exemplary state for estimation of magnetic field in \( p \) points of space

#### 1. Analytical solutions for \( p = 1, 2 \)

For \( p = 1 \) the simplex \( \sum_{i=1}^{p} |\mu_i| \leq 1/2 \) is simply the line \( \mu_1 \in [-1/2, +1/2] \), so the optimal solution is \( f(\mu_1) = \sqrt{2} \cos(\pi \mu_1) \) with the cost \( \pi^2/N^2 \). For \( p = 2 \) the simplex \( \sum_{i=1}^{p} |\mu_i| \leq 1/2 \) takes the form of a square with side \( \sqrt{2}/2 \) rotated by the angle 45° relative to the coordinate axes, which allows for an effective coordinates separation; therefore the solution is:

\[ f(\mu_1, \mu_2) = 2\cos(\sqrt{2}\pi(\mu_1 + \mu_2)/\sqrt{2}) \cos(\sqrt{2}\pi(\mu_1 - \mu_2)/\sqrt{2}) \]

with the corresponding cost \( \Delta^2 \theta^\text{MM} = 2 \times (\sqrt{2})^2 \pi^2/N^2 = 4\pi^2/N^2 \).

#### 2. Derivation of the bound

For the problem Eq. (91):

\[ \Delta^2 \theta^\text{MM} \geq \frac{p}{N^2} \min_{g} \int_{0}^{+\infty} d\mu g^*(\mu) \left( -\frac{\partial^2}{\partial \mu^2} \right) g(\mu), \]  

with \( \frac{\partial g}{\partial \mu} \big|_{\mu=0} = 0, \]

\[ \int_{0}^{+\infty} d\mu |g(\mu)|^2 = 1, \]

\[ \int_{0}^{+\infty} d\mu |g(\mu)|^2 |\mu| = \frac{1}{2\mu}, \]

the solution may be found using the standard Lagrange multiplier method,

\[ -\frac{\partial^2}{\partial \mu^2} g(\mu) + g(\mu)(\lambda_1 + \mu \lambda_2) = 0 \Rightarrow g(\mu) \propto Ai \left( \lambda_2^{1/3} (\lambda_1 \lambda_2^{-1/3} + \mu) \right), \]
where $\text{Ai}(\cdot)$ is the Airy function of the first kind. The condition $\frac{\partial f}{\partial \mu} \bigg|_{\mu=0} = 0$ implies $\lambda_1 \lambda_2^{1/3} = A_0' \approx -1.019$, where $A_0'$ is the first zero of derivative of $\text{Ai}(\cdot)$. From Eq. (D1) and Eq. (D2) we have:

$$\Delta^2 \theta_{\text{INT}}^{\text{MM}} \geq \frac{p^3}{N^2} \left( \int_0^{+\infty} d\mu |\partial_\mu \text{Ai}(A_0' + \mu)|^2 \right) \left( \int_0^{+\infty} d\mu |\text{Ai}(A_0' + \mu)|^2 \right) \approx 0.63 p^3. \tag{D5}$$

To get the value of $\lambda_2$ we use Eq. (D3) and Eq. (D2):

$$\frac{1}{2p} = \frac{\int_0^{+\infty} d\mu |\text{Ai}(\lambda_2^{1/3} (A_0' + \mu))|^2 \mu}{\int_0^{+\infty} d\mu |\text{Ai}(\lambda_2^{1/3} (A_0' + \mu))|^2} = \frac{\int_0^{+\infty} d\mu |\text{Ai}(A_0' + \mu)|^2 \mu}{\int_0^{+\infty} d\mu |\text{Ai}(A_0' + \mu)|^2} \cdot \lambda_2^{-1/3}, \tag{D6}$$

so finally:

$$\Delta^2 \theta_{\text{INT}}^{\text{MM}} \geq \frac{p^3}{N^2} \left( \int_0^{+\infty} d\mu |\partial_\mu \text{Ai}(A_0' + \mu)|^2 \right) \left( \int_0^{+\infty} d\mu |\text{Ai}(A_0' + \mu)|^2 \right) \approx 0.63 p^3. \tag{D7}$$

3. Exemplary state

Finally, let us present a suboptimal, but an explicit analytical solution of the initial problem Eq. (88), which shows a significant advantage compared with the optimal SEP protocol. Namely, we choose the largest possible $p$-dimensional ball inside the simplex $\sum_i |\mu_i| \leq 1/2$ and then take as the $f(\mu)$ the function which minimizes the kinetic energy inside this ball with a boundary condition $f(\mu) = 0$ on the border and outside of the ball. The Laplacian for spherical coordinated is given as:

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{p - 1}{r} \frac{\partial f}{\partial r} + \text{angular part}, \tag{D8}$$

where the exact form of the angular part is irrelevant for the discussion. The corresponding eigenstates are the ones of the form:

$$f(r) \propto r^{(2-p)/2} J_{p/2-1}(\sqrt{E} r) \Rightarrow -\Delta f(r) = Ef(r), \tag{D9}$$

where $J_{\alpha}(\cdot)$ is the Bessel function of the first kind. As the radius of the biggest ball inside the simplex $R$ satisfies:

$$R^2 = \sum_{i=1}^P (1/2p)^2 = \frac{1}{4p}, \tag{D10}$$

and taking into account the boundary condition $f(R) = 0$ we have:

$$\sqrt{E} \frac{1}{2\sqrt{p}} = j_{p/2-1,1} \Rightarrow E = p(2j_{p/2-1,1})^2, \tag{D11}$$

where $j_{p/2-1,1}$ is the zero of the Bessel function $J_{p/2-1}(x)$. Since for large $p$ we have $j_{p/2-1,1} \approx p/2$, we get:

$$E \approx p^3, \tag{D12}$$

which yields the joint cost $\frac{p^3}{2\pi}$. 