Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Beyond sensitivity: mechanism-resolved error budgets for designing quantum sensors

Nima Leclerc

n1475@cornell.edu

Mitre (United States)

Marco Capelli

Quantum Brilliance

Kevin James Rietwyk

Quantum Brilliance

Mark Dong

Mitre (United States)

Dmitry Lyakh

Nvidia (United States)

Geoffrey Iwata

SandboxAQ

Brandon Rodenburg

Mitre (United States)

Sean Oliver

Mitre (United States)

Benedikt Kloss

Nvidia (United States)

Jin-Sung Kim

Nvidia (United States)

Stefan Bogdanovic

SandboxAQ

Yunheng Chen

Quantum Brilliance

Meysam Sharifzadeh Mirshekarloo

Quantum Brilliance

Cedric Weber

Quantum Brilliance

Marcus Doherty

Quantum Brilliance

Ethan Pratt

SandboxAQ

Joseph Hagmann

Mitre (United States)

Article

Keywords: quantum sensing, quantum metrology, open quantum systems, Fisher information, error attribution, nitrogen-vacancy centers

Posted Date: September 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10873461/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Beyond sensitivity: mechanism-resolved error budgets for designing quantum sensors

Nima Leclerc[1,*,†], Marco Capelli[2], Kevin James Rietwyk[2], Mark Dong[1], Dmitry Lyakh[3], Geoffrey Iwata[4], Brandon Rodenburg[1], Sean Oliver[1], Benedikt Kloss[3], Jin-Sung Kim[3], Stefan Bogdanovic[4], Yunheng Chen[2], Meysam Sharifzadeh Mirshekarloo[2], Cedric Weber[2], Marcus Doherty[2], Ethan Pratt[4], and Joseph Hagmann[1,*]

[1] *MITRE, 202 Burlington Rd, Bedford, Massachusetts 01730, USA*

[2] *Quantum Brilliance, Level 1, 477 Pitt Street, Haymarket, Sydney, New South Wales 2000, Australia*

[3] *NVIDIA, 2788 San Tomas Expressway, Santa Clara, California 95051, USA*

[4] *SandboxAQ, 780 High St, Palo Alto, California 94301, USA*

[†] *Present address: Diraq, Palo Alto, California 94306, USA*

[*] Correspondence: nl475@cornell.edu; jhagmann@mitre.org

Abstract

Quantum sensors are often specified by a headline sensitivity, yet applications also demand accuracy and reliability. The dominant limiter of one metric is often known, but no method resolves how interacting mechanisms combine into a signed, per-mechanism budget for each metric. We introduce a framework that computes a sensor's sensitivity, accuracy, and robustness from one open-system simulation and attributes each to its limiting mechanism. For a nitrogen-vacancy diamond ensemble the attribution inverts across metrics: dephasing limits sensitivity, the thermal ground-state shift limits accuracy, and optical leakage limits robustness. At identical sensitivity the recovered-field bias spans 8 to 1500 nT, so tuning to sensitivity alone can miss the accuracy target by two orders of magnitude. The same modeling transfers to a cesium optically pumped magnetometer recording a human magnetocardiogram. Acting as a digital twin, it predicts the gain from addressing each limiter, so sensors can be designed to the required metrics.

Keywords: quantum sensing; quantum metrology; open quantum systems; Fisher information; error attribution; nitrogen-vacancy centers

1 Introduction

Quantum sensors are moving from laboratory demonstrations toward deployed applications, each judged by many simultaneous measures of performance (MoPs) rather than a single figure of merit. Magnetometry is the most mature modality, with platforms that include nitrogen-vacancy (NV) ensembles in diamond [1, 2, 3, 4] and optically pumped magnetometers (OPMs) [5, 6]. Different applications stress different MoPs. Magnetic navigation requires accuracy that holds calibration under

motion [7]. Biomagnetic imaging of the heart [8, 9] and brain [13] requires low-frequency sensitivity. Operation outside a shielded laboratory additionally requires robustness against drifting backgrounds, with spatiotemporal imaging adding further trade-offs [14, 15, 16]. A single sensor must therefore meet several of these MoPs at once. Designs today are specialized to each application by expert judgment about which limiter dominates in that regime. Scoring all of the metrics together from one device model makes the cross-application trade-offs quantitative and lets a design be re-provisioned to a new requirement without rebuilding the analysis.

These figures of merit are not independent design knobs but functionals of the same physics. A magnetometer can be modeled as an open quantum system whose state $\rho(t)$ evolves under a Lindblad master equation, and its sensitivity, accuracy, and robustness are all set by how that evolution responds to the sensed field and to imperfections in the device. Sensitivity follows from the Fisher information: propagating the state together with its derivative with respect to the field returns the achievable field resolution from a single integration of the dynamics [17, 18]. Many imperfections enter the same evolution, including magnetic and control noise, laser and optical-modulation noise, temperature drift, and the finite spin-coherence times (T_2 and its inhomogeneous counterpart $T_2^\star$) that phenomenological models compress into single numbers. A physical mechanism that limits one metric need not limit another. Spin dephasing, for example, degrades sensitivity by blurring the field resolution while leaving the recovered value unbiased. The design question is therefore which mechanisms set each metric, not the value of any single figure such as the coherence time.

In current practice this identification is rarely made directly, with sensitivity tuned empirically on the bench and the other metrics assumed to follow. The quantitative tools available today each address one part of the problem. Filter-function theory characterizes the coherence response to stochastic dephasing [19, 20, 21]. The Cramér–Rao bound and its multiparameter extensions set the precision limit and quantify the cost of estimating several parameters at once [22, 23, 24]. Precision–accuracy trade-off results bound what any estimator can achieve [25, 26]. Forward simulators reproduce measured pulsed and continuous-wave optically detected magnetic resonance (ODMR) signals from a device model [27, 28]. Open-system estimation toolkits propagate the Fisher information directly from the master equation [29, 30, 31]. Each optimizes or bounds a single axis, most often the precision, and few attribute accuracy or robustness to specific mechanisms, or account for more than one metric from a common forward model of the device.

Mechanism-level error budgeting is not itself new. It is standard practice in every mature precision instrument. Gravitational-wave detectors maintain per-source noise budgets that decompose the measured displacement spectrum into quantum, thermal, seismic, and technical contributions [32], and optical atomic clocks publish detailed line-by-line systematic-shift budgets whose entries are individually measured or bounded and then summed [33]. These budgets are constructed to be additive, with each contribution isolated and characterized on its own. A quantum sensor instead reads its field through a nonlinear, dissipative optical cycle in which the error mechanisms interact within a single measurement, so an additive ledger does not by itself assign those interactions, the sensing analog of where Matthiessen's rule fails in charge transport [34]. Attributing a total among interacting contributors is a solved problem in other fields: Shapley allocation, the unique axiomatic partition

from cooperative game theory [35], underlies feature attribution in machine learning [36], extends variance-based global sensitivity analysis to correlated and interacting inputs [37], and has recently been imported into quantum science to attribute a circuit's output to its gates [38]. Likewise, the tangent-linear propagation at the core of our engine is the forward-mode counterpart of the adjoint construction that underpins gradient-based design in aerodynamics [39] and variational data assimilation in weather forecasting [40], fields in which simulation-grounded sensitivity replaced iteration with systematic optimization. Quantum sensing has the forward models to support the same transition, but no tool yet connects them to a multi-metric, per-mechanism budget.

Here we provide that missing tool: a single device-grounded model that evaluates a sensor's sensitivity, accuracy, and robustness (the stability of the sensitivity under drift of the operating parameters) from one open-system solve and assigns each a signed budget over the physical mechanisms that set it. The same construction extends to other figures of merit such as dynamic range. The tangent solve it uses is standard open-system machinery; the innovation of this work is to turn one such solve into a per-mechanism budget across all three metrics, one that resolves how the interacting mechanisms combine rather than assuming they add. The model combines three ingredients. It first computes all three metrics from a common open quantum system model. It then attributes each to individual physical mechanisms using Shapley values [35, 38], which are needed because the mechanisms interact and are not in general additive. Finally, it validates the resulting error budgets against measured devices spanning two sensing platforms, a nitrogen-vacancy (NV) diamond ensemble and a cesium optically pumped magnetometer array (Secs. 2.3, 2.4). The computational core is a standard open-system tangent-linear construction [29, 41, 42, 30], which we refer to as the *tangent master equation* (TME). For each coalition, a single tangent solve co-propagates the state $\rho(t)$ and its field derivative $\partial_b \rho(t)$ and jointly computes all three metrics, and GPU-accelerated batching over the 2^M coalitions makes the full attribution affordable [43, 44].

Applied across a design space and to the measured NV data, the model shows that error attribution is metric-dependent. For a specified set of experimental conditions, the dominant mechanism differs from one metric to the next, with spin dephasing limiting sensitivity, the thermal ground-state shift limiting accuracy, and optical leakage limiting robustness. Two sensors of equal sensitivity can therefore differ markedly in accuracy and calibration stability. Because each metric traces to a specific, measurable intermediate (readout contrast, optimized control [45], engineered materials [46, 47], or redesigned delivery [48]), the model identifies the dominant mechanism for each metric and predicts the metric gain from acting on it, in place of bench iteration. This makes designing to the application quantitative rather than aspirational. Although we develop the framework on magnetometry, nothing in its construction is specific to that modality: any sensor described by an open quantum system with an optical readout (an optical clock with its systematic-shift budget, an atom interferometer, a Rydberg electrometer) admits the same single-solve, per-mechanism, multi-metric accounting.

2 Results

2.1 A per-mechanism, multi-metric attribution model

The framework treats the sensor as an open quantum system and reads all three metrics from a single solve of its dynamics, then attributes each to individual mechanisms (Fig. 1). We develop it here on a nitrogen-vacancy (NV) diamond ensemble, and later show how the same open-system modeling extends to a physically unrelated optically pumped magnetometer built on different underlying physics. The device (Fig. 1a) is optically initialized and read out at 532 nm, driven by resonant microwave (MW) control, and reports a sensed field b through its spin-dependent fluorescence against an ambient-noise and temperature-drift environment. We instantiate it as a ten-level model whose density matrix $\rho(t)$ obeys the Lindblad master equation

$$\dot{\rho} = \mathcal{L}(t;\theta)[\rho] = -\mathrm{i}\,[H(t;\theta),\rho] + \sum_j \mathcal{D}[L_j(\theta)]\,[\rho], \qquad \text{where} \quad \mathcal{D}[L][\rho] = L\rho L^\dagger - \tfrac{1}{2}\{L^\dagger L,\rho\}. \tag{1}$$

Here $\mathcal{L}(t;\theta)$ is the Lindbladian superoperator, $H(t;\theta)$ the Hamiltonian, the $L_j(\theta)$ the jump (collapse) operators, $\mathcal{D}$ the dissipator, and $[\cdot,\cdot]$ and $\{\cdot,\cdot\}$ the commutator and anticommutator. The Hamiltonian and jump operators depend on the experimental parameters θ that the experiment sets and probes, and encode the dissipative channels of Fig. 1a: optical pumping, radiative and singlet decay, intersystem crossing [51, 52], phonon-driven orbital hopping [49], and dephasing (the full Hamiltonian and collapse set are given in Methods). Among these, optical leakage ε from an imperfect modulator enters as residual pump power on the readout channel, so a single measured extinction ratio lowers the contrast and shifts the photon baseline, effects that a phenomenological dephasing rate cannot capture.

The model draws on three input channels (Fig. 1b): the control pulse sequence, the stochastic environment synthesized from measured noise spectra into time traces, and the material-dependent device parameters such as strain and coherence. From these, one *tangent* solve co-propagates the state and its field derivative and returns all three metrics (Fig. 1b). The stochastic environment supports a full time-dependent-noise treatment. The sensitivity under $1/f$, Ornstein–Uhlenbeck, and white-noise baths is computed in the Supplementary Information (Supplementary Note 5), while the deterministic single-point results reported here do not use those traces.

We obtain the field derivative $G_b = \partial_b \rho$ by co-propagating it with the state, so one solve returns G_b to machine precision without re-solving the dynamics at perturbed fields. The tangent equations are given in Methods. Under Poisson photon counting with a detected-photon budget N_{det}, the readout probability $p = \mathrm{Tr}[M\rho(\tau)]$, for the readout (measurement) operator M, and its exact derivative $\partial_b p = \mathrm{Tr}[M\,G_b(\tau)]$ fix the Fisher information $\mathcal{F}$ about b and the shot-noise-limited sensitivity,

$$\mathcal{F}(\theta) = \frac{N_{\mathrm{det}}\,(\partial_b p)^2}{p} \text{ and } \eta_B(\theta) = \sqrt{T_{\mathrm{acq}}/\mathcal{F}(\theta)}, \tag{2}$$

with T_{acq} the total acquisition time, the smallest field resolvable in unit bandwidth [53, 54, 55]. The accuracy is set by the systematic bias $\beta = \langle\hat{b}\rangle - b$ left after inverting the readout through the forward-model calibration, which the root-mean-square error reduces to when the bias dominates. Robustness,

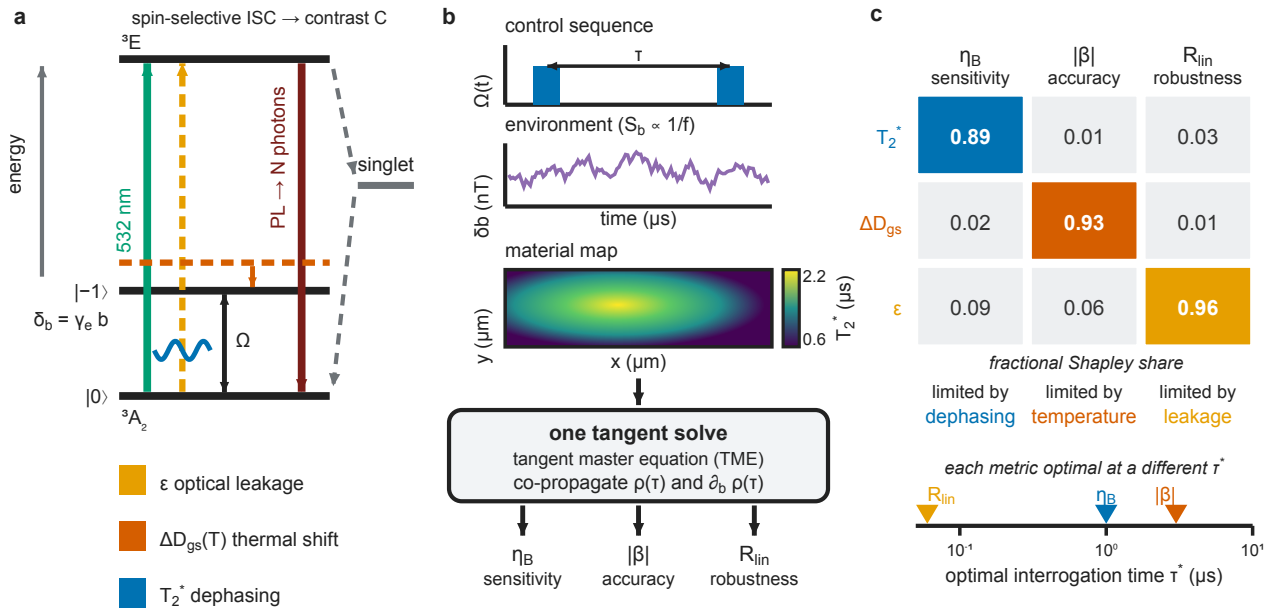


Figure 1: **Per-mechanism, multi-metric attribution from one open quantum system model.** A single open-system solve returns a sensor's three performance metrics, its sensitivity (η_B, the smallest field resolvable per unit bandwidth), accuracy (limited by the systematic bias β of the recovered field), and robustness (R_{lin}, the stability of the sensitivity under parameter drift), and attributes each to individual physical mechanisms. **(a)** Simplified level scheme of the nitrogen-vacancy (NV) center. Optical pumping at 532 nm and spin-selective intersystem crossing (ISC) through the singlet set the readout contrast C, the microwave drive Ω addresses the $|0\rangle \leftrightarrow |-1\rangle$ transition whose splitting responds to the sensed field as $\delta_b = \gamma_e b$, and photoluminescence returns N photons. The three error mechanisms carried through the paper are marked in color: optical (acousto-optic) leakage ε (orange) as residual pump on the readout channel, the thermal shift of the ground-state zero-field splitting $\Delta D_{\mathrm{gs}}(T)$ (vermillion, dashed level), and $T_2^\star$ dephasing (blue, wavy line). The same colors are used in Figs. 2–5. **(b)** The three input channels, the control pulse sequence $\Omega(t)$, the stochastic environment synthesized from measured noise spectra ($S_b \propto 1/f$ here), and the material maps of coherence and strain, enter a single tangent solve (tangent master equation, TME) that co-propagates the density matrix $\rho(\tau)$ and its field derivative $\partial_b \rho(\tau)$ at the interrogation time τ and returns all three metrics from one propagation. **(c)** Toggling each mechanism in that solve and combining the marginal effects with Shapley values gives a per-mechanism budget for every metric, reported as the fractional Shapley share. A different mechanism dominates each metric: sensitivity is limited by dephasing, accuracy by the temperature-driven ΔD_{gs} shift, and robustness by optical leakage. Each metric is also optimal at a different interrogation time $\tau^\star$ (lower strip).

as defined here, is the stability of the sensitivity η_B under drift of the operating parameters, summarized by a linearized second-moment index we derive in Supplementary Note 4,

$$R_{\mathrm{lin}}^2 = 1 + \sum_i \sigma_i^2 \left(\partial_{\theta_i} \log \eta_B\right)^2, \tag{3}$$

where σ_i is the assumed root-mean-square drift of parameter θ_i and $\partial_{\theta_i} \log \eta_B$ the fractional change it

produces. A value $R_{\mathrm{lin}} = 1$ marks drift that leaves the sensitivity untouched, while $R_{\mathrm{lin}} = 1.1$ costs 10% of the nominal sensitivity (Methods). To attribute each metric we recompute the solve with individual mechanisms toggled and combine the marginal effects with Shapley values [35], which distribute the interactions a leave-one-out accounting would leave unassigned, giving the per-mechanism budget of Fig. 1c.

We validate the engine end to end before applying it (Fig. 2). The co-propagated derivative agrees with central finite differencing to better than 10^{-4} (Fig. 2a), and the readout contrast is a prediction rather than an input, computed from the ten-level model so its fall with temperature and strain emerges without a fit (Methods). The calibration fringe of Fig. 2b makes the contrast-to-resolution link concrete under Poisson sampling: at the steep working point a high-contrast ensemble ($C = 0.46$) recovers the field to $\sigma_b = 32\,\mathrm{nT}$, where a low-contrast one ($C = 0.17$) reaches only 87 nT. Against measured data, the engine reproduces the coherence–density literature across nearly three decades of nitrogen concentration ($[N] = 0.08$ to $60\,\mathrm{ppm}$) [56]. Each $\eta_B(\tau)$ traces a U-shaped curve set by two competing effects. The accumulated precession phase grows with the interrogation time and sharpens the field response, while the readout contrast decays as $\exp(-\tau/T_2^\star)$ as coherence is lost. Sensitivity therefore improves with τ until the contrast decay overtakes the phase gain, giving a minimum at τ_{opt}, which tracks $T_2^\star([N])$ and moves to shorter τ as added nitrogen shortens the coherence ($T_2^\star \propto 1/[N]$, Fig. 2c-i). The floor of each curve first deepens as more nitrogen buys photons, then shifts back up at the highest $[N]$ as the conversion efficiency E_{conv} rolls off and the coherence penalty overtakes the photon gain, so more nitrogen eventually costs sensitivity. The envelope of these minima therefore flattens at long coherence, where E_{conv} is flat, and departs from the ideal $1/\sqrt{T_2^\star}$ scaling only where conversion rolls off at high $[N]$ (Fig. 2c-ii). This literature agreement is a consistency check under assumed collection prefactors, unlike the parameter-free derivative cross-check of Fig. 2a.

Panel c also carries a physical lesson. The achievable sensitivity is fixed by the charge-conversion efficiency, so beyond a point more coherence buys nothing, and the material property to engineer is E_{conv}, not the coherence time. Even for a single metric the limit traces to a specific measurable mechanism. All three metrics now come from one validated forward model, so any difference in what limits them is a property of the device.

2.2 Metric-dependent attribution

With the engine validated against internal Fisher-information cross-checks and the coherence–density literature (Fig. 2), we now apply it to the motivating hypothesis of this work and arrive at the central result (Fig. 3). Because all three metrics come from the same tangent solve, we can compare them on equal footing and ask directly whether one mechanism dominates all of them or the answer depends on the metric. Using our Shapley construction, we hold the device fixed, with the sensed field $b = 1\,\mathrm{nT}$ on the steepest part of the fringe, $T_2^\star = 1.5\,\mu\mathrm{s}$, $T_1 = 5\,\mathrm{ms}$, and a thermal drift $\Delta T = 0.1\,\mathrm{K}$, and sweep only the interrogation time from 0.05 to $3\,T_2^\star$ (Fig. 3a).

The sensed field b enters as a Ramsey phase $\phi = \gamma_e b\,\tau$, so the readout slope grows with τ while the fringe contrast decays. The sensitivity η_B therefore traces a U-shape, falling to a shallow minimum of about $24\,\mathrm{pT}/\sqrt{\mathrm{Hz}}$ near $\tau \approx 1\,\mu\mathrm{s}$, where one more unit of τ adds as much phase as it removes contrast,

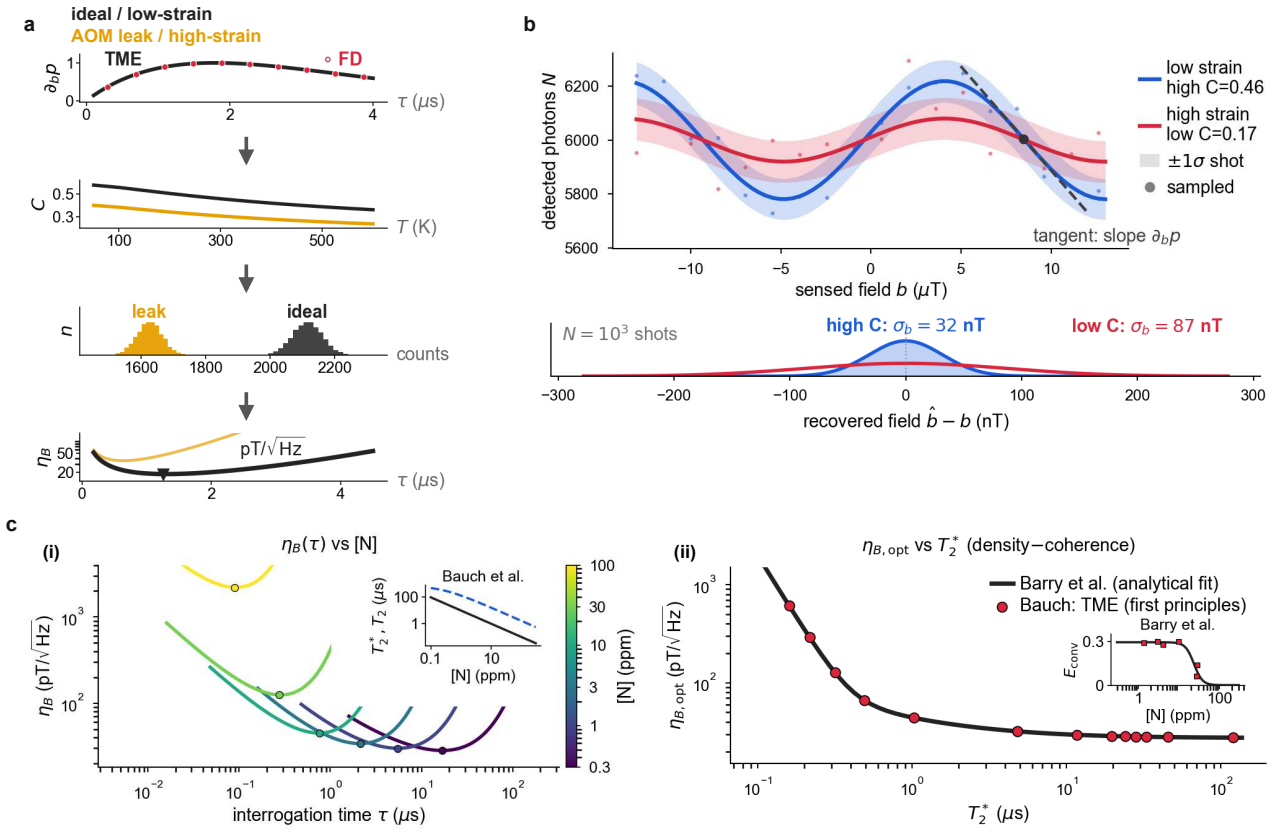


Figure 2: **Validation of the first-principles sensitivity engine.** The engine computes the shot-noise-limited sensitivity η_B (smallest field resolvable per unit bandwidth) as a function of the interrogation time τ. **(a)** The single-solve pipeline, from the field derivative $\partial_b \rho(\tau)$ of the density matrix (computed by the tangent master equation and checked against finite differencing, red circles; agreement $< 10^{-4}$) through the readout contrast C and the photon statistics to $\eta_B(\tau)$. Each stage is traced for an ideal, low-strain ensemble (black) and a degraded case combining high transverse strain with acousto-optic-modulator (AOM) leakage (orange). The AOM gates the pump laser on for initialization and readout and off during the dark sensing interval; imperfect extinction leaves a residual pump field ε in that interval, which re-polarizes the spin and lowers the readout contrast, so the leaky, strained case carries a smaller derivative, a lower contrast, and a worse $\eta_B(\tau)$ throughout. **(b)** The calibration fringe (detected photons N versus field b) for high- and low-contrast NV ensembles under Poisson shot noise, which fixes the field resolution σ_b. **(c)** The sensitivity–coherence relation versus nitrogen concentration $[N]$ [56], where $T_2^\star$ is the inhomogeneous (effective) spin-coherence time. *(c-i)* a family of U-shaped $\eta_B(\tau)$ curves, one per $[N]$. *(c-ii)* the optimal sensitivity $\eta_{B,\mathrm{opt}}$ versus $T_2^\star$, compared with the closed-form Ramsey analytic [1], flattening where the charge-conversion efficiency $E_{\mathrm{conv}} = [\mathrm{NV}^-]/[\mathrm{N}]$ is flat (inset). This literature agreement uses assumed collection prefactors (Methods) and is a consistency check, not a parameter-free prediction.

then climbing as the exponential contrast decay beats the linear phase gain (dashed curve, Fig. 3a). The solid curves in Fig. 3a are the Shapley fraction each mechanism supplies to the coherent (non-shot) part of η_B, obtained by toggling that mechanism in the solve, and they sum to one at every τ. Dephasing dominates, at about 89% across short and intermediate τ, because the contrast decay that

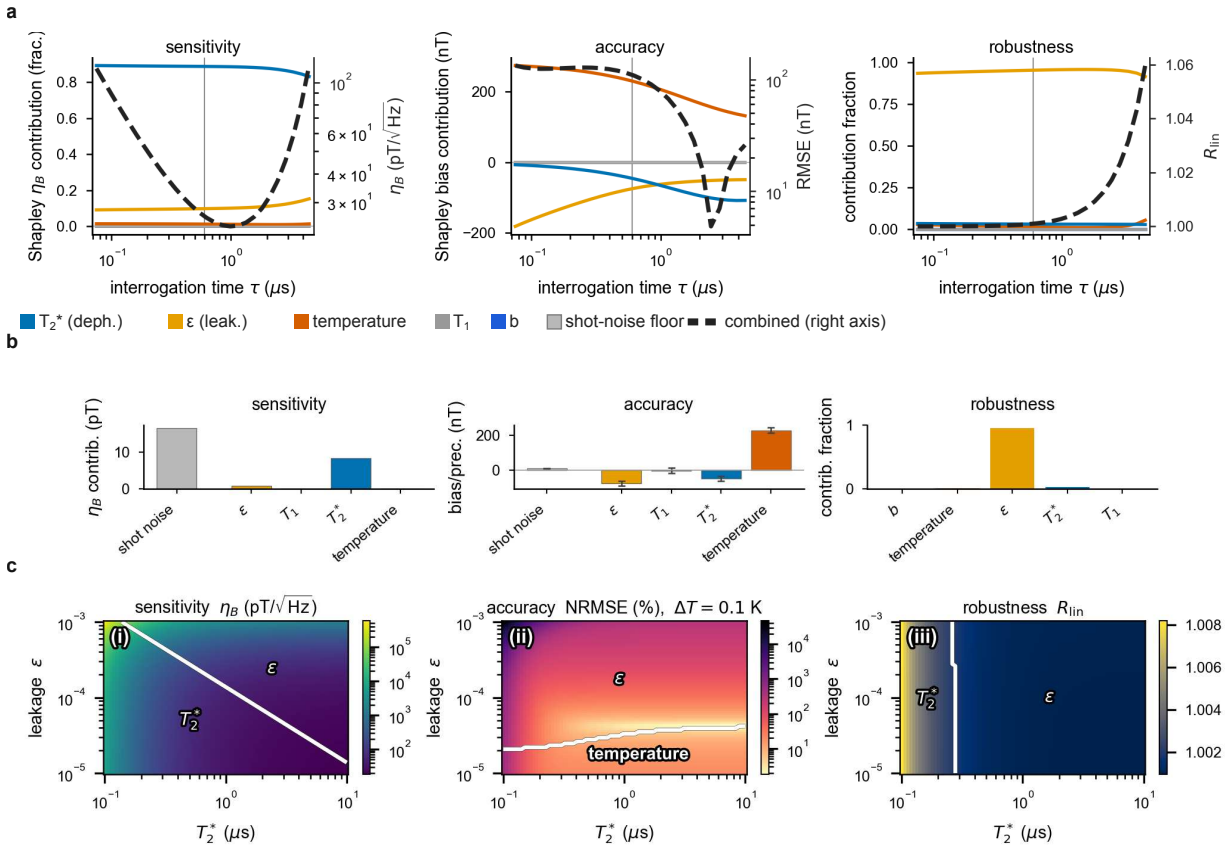


Figure 3: **Metric-dependent attribution.** For a fixed device, the per-mechanism budget of each metric, computed from one open-system solve: sensitivity (η_B), accuracy (the systematic bias, summarized by the root-mean-square error, RMSE), and robustness (the linearized drift index R_{lin}, equal to one when drift leaves the sensitivity untouched). **(a)** Per-mechanism contribution versus interrogation time τ (solid, left axis; Eq. 5), one panel per metric, with the mechanisms color-coded (dephasing $T_2^\star$ in blue, optical leakage ε in orange, thermal shift in vermillion, and longitudinal relaxation T_1 in gray). The shares are Shapley values for sensitivity and accuracy and the additive drift-sensitivity decomposition for robustness. They are fractional and sum to one for sensitivity and robustness, and are signed contributions (in nT) summing to the total bias for accuracy. The dashed curve is the combined metric on the right axis (the absolute η_B, RMSE, or R_{lin}), plotted on its own scale. The three metrics reach their optima at three different τ. **(b)** The budget at the operating τ (mechanism colors as in Fig. 1a), in which a different mechanism dominates each metric. The leftmost (gray) bar in the sensitivity and accuracy panels is the shot-noise baseline $v(\varnothing)$ against which the mechanism contributions are attributed, not itself a mechanism. The sensitivity and robustness shares are deterministic functionals (Cramér–Rao and finite-difference; no sampling distribution), whereas the accuracy shares carry a 500-realization shot-noise interval (Methods). **(c)** 80×80 maps over the coherence time $T_2^\star$ and the optical leakage ε of sensitivity, accuracy (reported as normalized root mean square error (NRMSE)), and robustness, with white curves marking the boundaries where the dominant limiter changes.

shapes the whole curve *is* the $T_2^\star$ dephasing. Its share slips to 83% at the longest τ as optical leakage

grows from 9 to 15%: the longer dark interval gives the residual pump more time to repolarize the spin and lift the photon baseline, so leakage costs a rising fraction of the sensitivity as τ increases. The thermal channel and T_1 never exceed 2%, since neither reshapes the fringe contrast that sets η_B.

Accuracy combines a systematic bias with the shot-noise floor. For it, Fig. 3a plots the *signed* per-mechanism contributions to the recovered-field bias (in nT), which sum to the total bias, with the resulting RMSE on the dashed axis. The dominant term is thermal. A 0.1 K offset moves the ground-state splitting by $\delta \approx -7.9\,\mathrm{kHz}$, an apparent field $\delta/\gamma_e \approx 282\,\mathrm{nT}$ that carries no τ dependence of its own, since the extra phase and the field inferred from it both scale with τ. The leakage contribution runs the other way, negative at short τ, so this and the still-large shot floor partly mask the thermal term, and the net RMSE sits below the full thermal value. At intermediate τ both fall away and the net bias approaches its full 282 nT. At long τ the linear calibration, built on a fringe slope that is itself washing out, over-corrects and pulls the recovered field low, canceling the thermal bias near $\tau \approx 2.5\,\mu\mathrm{s}$, where the RMSE dips to its few-nanoTesla shot floor (the sharp minimum of the dashed curve) before climbing again.

Robustness runs opposite to sensitivity because R_{lin} is built from the fractional sensitivity slopes $\partial_\theta \log \eta_B$, and those grow with τ. At short τ the spin has barely evolved, so η_B is set mainly by the photon budget and is nearly flat in the operating parameters, which keeps the slopes small and $R_{\mathrm{lin}} \approx 1.00$. As τ increases, η_B climbs the decoherence wall as $\mathrm{e}^{\tau/T_2^\star}$ and the leakage-driven contrast loss accumulates over the interrogation window, so a fixed drift in any channel is amplified and R_{lin} rises to about 1.06 by $3\,T_2^\star$ (dashed, Fig. 3a). The solid curves are each parameter's share of $R_{\mathrm{lin}}^2 - 1$, and leakage drift dominates throughout, at ∼96%. A small drift in the extinction ratio moves both the dark-window repolarization and the photon baseline, to which η_B is steeply sensitive. Because the leakage acts throughout the dark interval, a longer interrogation time gives its drift more opportunity to accumulate into the sensitivity, so the leakage share of R_{lin} grows with τ. Drifts in b, T, and T_1 change the sensitivity only weakly, and $T_2^\star$ drift matters only once τ is pushed onto the decoherence wall. Robustness is therefore best at the shortest τ, exactly where sensitivity is worst, so no single interrogation time is best for all three metrics.

At the operating τ, the per-source budget (Fig. 3b) gives each metric a different dominant origin. Coherence dominates sensitivity, temperature dominates accuracy, and leakage drift dominates robustness, all read off the one calculation. This ordering is stable across 500 shot-noise realizations and the parameter ranges above.

Physically, pure dephasing scrambles the phase symmetrically, lowering sensitivity without biasing the recovered field, whereas the thermal shift and the dissipative leakage displace the fringe and bias accuracy while barely touching sensitivity (the unbiased and biased Cramér–Rao bounds are derived in Supplementary Note 4). The thermal channel therefore maps an accuracy target directly onto a temperature-stability requirement ($\partial_T D_{\mathrm{gs}}/\gamma_e \approx 2.8\,\mu\mathrm{T/K}$, $1\,\mathrm{nT} \leftrightarrow 0.35\,\mathrm{mK}$).

Because every metric comes from the same batched solve, the framework can map out a whole design space at once rather than one operating point at a time. We propagate the $80 \times 80 = 6{,}400$-cell $(T_2^\star, \varepsilon)$ grid of Fig. 3c together as a single GPU-accelerated batch (spanning 0.1 to 10 µs and 10^{-5} to 10^{-3}, Methods) and attribute every cell, so each carries its own dominant limiter for each metric and

the white curves mark where that limiter changes. The three metrics partition the plane differently. Sensitivity is dephasing-limited over 57% of it, with its optimum ($18.4\,\mathrm{pT}/\sqrt{\mathrm{Hz}}$) at the longest-$T_2^\star$, lowest-$\varepsilon$ corner. Accuracy is temperature-limited below $\varepsilon \approx 3 \times 10^{-5}$ and leakage-limited above, its bias falling to $\sim$7 nT on the cancellation ridge. Robustness is leakage-drift-limited over 78% of the plane ($T_2^\star$-limited only below $\approx$0.27 µs). Read as a design tool, the map shows which mechanism to attack for a given target and where in the plane the limiting mechanism flips, so a specification can be met by moving to a favorable region of the design plane rather than by improving the whole device. (Supplementary Note 10 outlines, as an outlook, an adaptive closed-loop version of this targeting that would act on the identified mechanism at runtime.) The same batched solve extends beyond parameter grids to spatial degrees of freedom, and a per-pixel attribution map of a spatially varying device, computed as one batch across multiple GPUs, is demonstrated in the Supplementary Information (Supplementary Note 6).

No single operating point optimizes all three metrics. Along a single iso-η_B locus at $100\,\mathrm{pT}/\sqrt{\mathrm{Hz}}$ the accuracy bias ranges over 8 to 1500 nT, so a device tuned for sensitivity is not automatically accurate. This span is evaluated at the held $\Delta T = 0.1$ K drift and the assumed collection prefactors of the Methods, and scales linearly with the drift. The same analysis on five further design planes, sweeping τ, T_1, ΔT and b against $T_2^\star$, ε or one another, shows the same pattern, with accuracy alone changing its limiting mechanism from axis to axis. So no single mechanism dominates, and what limits a quantum sensor depends on which metric the application asks about. For each metric the per-mechanism budget names the single mechanism that limits it, so experimental effort can be directed at that mechanism, and at the measurable intermediate it acts through. We turn next to whether the same attribution survives on a fabricated device, where the mechanisms are measured rather than swept.

2.3 The error budget of a real device

So far the attribution has been demonstrated across a designed parameter space. We now apply it to a measured device. Measured data enter as model parameters θ and propagate to metrics and attributions. We exercise that pipeline starting from the NV charge state as a function of optical intensity P, reported as a power density (W/cm^2). With the illuminated area fixed, the beam power and the intensity are proportional, so we use the intensity P throughout this subsection.

Only NV^- carries spin contrast, while NV^0 adds background. We measured the $\mathrm{NV}^-/\mathrm{NV}^\mathrm{T}$ (where NV^T is the total NV defect concentration) charge fraction on two growth samples by confocal photoluminescence spectroscopy, decomposing each spectrum into its NV^- and NV^0 components at each laser intensity under 532 nm excitation (Supplementary Note 8). The measured fraction falls monotonically with optical intensity, from 0.59 to 0.085 on the 0.3 ppm diamond and from 0.73 to 0.25 on the 4.5 ppm diamond (Fig. 4a). The higher-nitrogen growth retains about three times more NV^- at full intensity [57]. The saturation curves we measured alongside set the detected-photon budget, which rises and then saturates and is several times brighter on the 4.5 ppm sample, with NV^- saturation intensities of 2.8 and $7.6 \times 10^6\,\mathrm{W/cm^2}$ (Fig. 4b). The effective readout contrast is the optical-cycle value computed from the ten-level Lindblad readout at each measured intensity ($C_{\mathrm{optical}} = 0.52$ at low

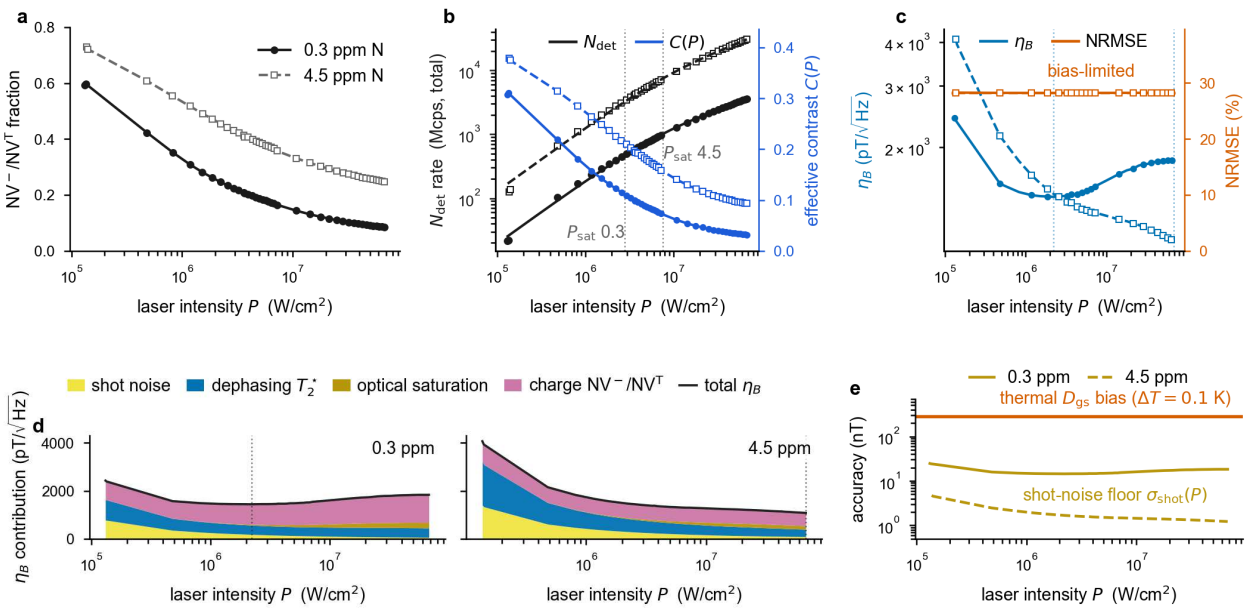


Figure 4: **Metric-dependent attribution on a measured operating knob.** The optical readout intensity P serves as a measured control knob on two diamonds of differing nitrogen content (0.3 and 4.5 ppm; solid and dashed curves). **(a)** the measured $\mathrm{NV}^-/\mathrm{NV}^{\mathrm{T}}$ charge fraction (the spin-active fraction of centers) versus laser intensity P. **(b)** the detected-photon budget $N_{\mathrm{det}}(P)$ (saturation intensities marked) and the effective readout contrast $C_{\mathrm{eff}}(P)$. **(c)** the sensitivity $\eta_B(P)$ (smallest field resolvable per unit bandwidth, left axis), with an interior optimum on the 0.3 ppm diamond, and the accuracy as a normalized root-mean-square error NRMSE(P) (right axis), flat because it is limited by a systematic bias rather than by photon shot noise. **(d)** the sensitivity budget, Shapley shares (Eq. 5) of charge-state dilution, optical saturation, and dephasing, stacked on the shot-noise floor. The floor-to-charge handoff underlies the interior optimum (optical leakage is excluded, as no $\varepsilon(P)$ was measured; Methods). **(e)** the accuracy decomposition. The thermal bias from the temperature-dependent ground-state splitting D_{gs} (a single line, identical for both diamonds because it is set by $\Delta D_{\mathrm{gs}}/\gamma_e$ alone, with γ_e the electron gyromagnetic ratio) sits far above the shot floor $\sigma_{\mathrm{shot}}(P)$, the only diamond-dependent term.

intensity, rolling off to 0.38 at optical saturation), diluted by the measured NV^- fraction, so it falls in step with the charge state (Fig. 4b).

We find that these two measured trends pull sensitivity in opposite directions. With $\eta_B \propto 1/(C_{\mathrm{eff}}\sqrt{N_{\mathrm{det}}})$, contrast and photon number are independent levers, and a high NV^- *fraction* does not directly imply a large photon *count*. At low intensity the contrast is high but the light is scarce: on the 4.5 ppm diamond the lowest intensity delivers so few photons (~ 125 per shot) that the fourfold contrast advantage cannot offset the 240-fold photon deficit, and the sensitivity is roughly four times worse. Raising the intensity adds photons faster than it costs contrast until the charge state begins to ionize. On the 0.3 ppm diamond, where we measured the NV^- fraction collapsing to 0.085, the contrast loss eventually overtakes the photon gain and $\eta_B(P)$ turns over at an interior optimum near $P^\star \approx 2.2 \times 10^6\,\mathrm{W/cm^2}$ (Fig. 4c). On the better-retained 4.5 ppm diamond the photons keep winning

and the optimum sits at the edge of the measured range.

Accuracy behaves differently on the same knob. The bias that limits it comes from the thermal $D_{\mathrm{gs}}(T)$ shift, about 282 nT for a 0.1 K stability specification, which is set by the interrogation point and not by photons. Contrast does enter the Poisson photon sampling, but the slope-based calibration inverts the readout by the same contrast, so it cancels from the recovered-field bias and survives only in the statistical shot floor $\sigma_{\mathrm{shot}} \sim 1/(C_{\mathrm{eff}}\sqrt{N})$. Once that floor is averaged below the bias, the accuracy is bias-limited and nearly flat across the full measured intensity range, almost three decades (Fig. 4c). The same first-principles contrast that dominates sensitivity leaves accuracy nearly unchanged. The shot-noise-limited regime, where the falling contrast does degrade accuracy at low photon number, is shown in Supplementary Note 4. Because the two diamonds have different charge-versus-intensity curves, the optimum $P^{\star}$ and the achievable sensitivity also differ by growth.

The per-mechanism budget makes this asymmetry quantitative on the same measured axis (Fig. 4d,e). For sensitivity we decompose $\eta_B(P)$ by the same Shapley construction over the three intensity-relevant mechanisms, the charge-state dilution, the optical-cycle saturation, and dephasing, stacked on the shot-noise floor at the measured photon budget. On the 0.3 ppm diamond the floor supplies 32% of η_B at the lowest intensity but only 3% at the highest, while the charge-state share grows from 33% to 63% and dominates at $P^{\star}$. Dephasing stays a roughly constant background set by [N] rather than by P. The handoff from photon scarcity to charge-state dilution is the mechanism behind the interior optimum.

The accuracy budget on the same axis needs no Shapley treatment. With a single bias source set against a single variance source, the thermal D_{gs} bias sits one to two orders of magnitude above the shot floor $\sigma_{\mathrm{shot}}(P)$ across all three decades of measured intensity; this dominance is why the normalized root-mean-square error (NRMSE) in Fig. 4c stays flat.

The metric-dependent asymmetry therefore holds on a real device driven by a measured control knob, not only across a designed parameter space. The same optical intensity that reshapes sensitivity leaves accuracy essentially flat, so tuning the readout for one metric does not tune it for the other. For experimental design this means the optical intensity should be set against the metric an application actually needs. Where sensitivity governs, push toward $P^{\star}$ to buy it. Where accuracy governs, our framework prescribes that design efforts should be directed at temperature stabilization, since the thermal bias sets the accuracy.

2.4 Cross-platform provisioning and design by specification

With the attribution established on the NV platform, a deployed sensor raises a complementary question. Beyond which mechanism limits each metric, we ask how far each layer of the stack is over- or under-provisioned for its application, and we answer it with the same open-system, tangent-Fisher modeling on a physically unrelated platform, with the three metrics realized in platform-specific form and read as provisioning budgets rather than a re-run of the per-mechanism attribution. The device is an unshielded magnetocardiography (MCG) system developed by SandboxAQ, containing a 26-channel scalar (total-field) cesium optically pumped magnetometer array [9]. This SandboxAQ MCG device has shown promise for detecting coronary artery disease (CAD) in clinical and emergency

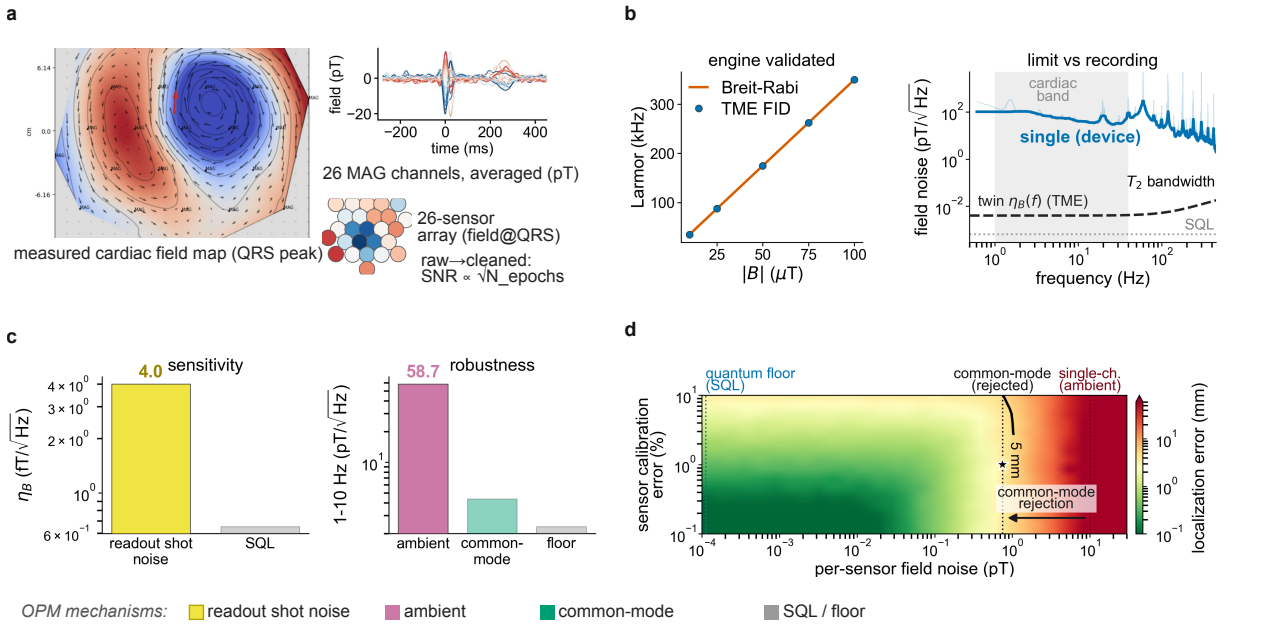


Figure 5: **One framework, two platforms: design by specification on a real scalar-OPM magnetocardiogram.** A 26-sensor scalar (total-field) cesium (Cs) optically pumped magnetometer (OPM) array, unshielded at the geomagnetic field, records a human magnetocardiogram. **(a)** The measured cardiac-dipole field map at the QRS peak (the largest deflection of the heartbeat), the averaged waveform, and the array geometry. **(b)** The model grounded in the recording: the Larmor frequency versus the exact Breit–Rabi prediction, and the frequency-resolved sensitivity $\eta_B(f)$ with its fundamental spin-projection (standard-quantum-limit) floor, against the measured field noise. **(c)** The sensitivity and robustness budgets read as provisioning margins. The sensitivity η_B is split into its spin-projection floor and the residual readout shot noise, and the low-band (1 to 10 Hz) robustness into the single-channel ambient noise, its residual after common-mode rejection, and the sensor floor. The quantum sensitivity floor lies orders of magnitude below the ambient, so the design margin is in rejecting the ambient rather than in the atomic floor. **(d)** Dipole-localization error versus per-sensor field noise and cross-sensor gain calibration, with the 5 mm contour of the arrhythmia-ablation requirement (the star marks the common-mode-rejection floor). Localization is limited by field noise, not by calibration.

settings [10, 11, 12]. A forward model with the cesium ground-state Breit–Rabi Hamiltonian [58] reproduces exact Breit–Rabi across the geomagnetic range ($\gamma \approx 3.5\,\mathrm{Hz/nT}$). The averaged trace ($\sim$ 340 beats) resolves the cardiac cycle, and at the QRS peak (representing the largest deflection of the heartbeat and the marker of ventricular depolarization) the array sees a clean current-dipole pattern (Fig. 5a). One clinical application of MCG is millimeter-scale localization of arrhythmogenic electrophysiological activity including ectopic foci or accessory pathways. As a related demonstration, we localize the net ventricular-depolarization current dipole at the QRS peak of the averaged beat, and adopt the few-millimeter target as one worked system-level metric.

On this platform we read the metrics as provisioning statements, asking what each layer of the stack must deliver for the application. The *sensitivity* we compute is a fundamental floor: the readout-

noise-limited TME field response $\eta_B(f)$ at the cardiac frequencies (Methods), flat across the cardiac band (Fig. 5b), with the spin-projection (standard-quantum-limit) floor of $\sim 1.1\,\mathrm{fT}/\sqrt{\mathrm{Hz}}$ beneath it. These are the limits set by photon and spin statistics alone. The single-channel noise of a real scalar OPM sits far above them, at the $\mathrm{pT}/\sqrt{\mathrm{Hz}}$ level (Fig. 5b), fixed by practical factors our first-principles model does not include, among them light shifts, spin-exchange broadening, and laser and electronic technical noise. We therefore take the *measured* field noise, not the fundamental floor, as the operating sensitivity that drives the localization analysis below; the intrinsic floor enters only as a design margin, roughly three orders of magnitude of headroom to the standard quantum limit, indicating that sensitivity is not the binding constraint for this application and that the atomic layer could be traded for cost, robustness, or a simpler source before it became the bottleneck (quoted at the assumed collection prefactors, Methods).

The *heading error* is the accuracy channel of a scalar magnetometer, the orientation-dependent bias by which the apparent $|B|$ shifts a few nT with the angle between the spin polarization and the field. On a moving or varying-orientation platform it is the dominant accuracy limiter, so which mechanism sets accuracy depends on the application. For this fixed array, every sensor is mounted near $\theta \approx 90°$, so the common part of the bias calibrates out. The array's accuracy is instead set by the 1% cross-sensor gain calibration of that localization analysis.

Robustness is set by the ambient field itself. Software common-mode rejection suppresses it $\sim 13\times$ to $\sim 4.4\,\mathrm{pT}/\sqrt{\mathrm{Hz}}$ by projecting out the leading principal component of the multi-sensor recording, which carries 99.9% of its variance and captures the spatially correlated ambient (Fig. 5c). Because this component is estimated from the data, the rejection does not presuppose cross-sensor gain calibration.

Cast as a specification, these metrics define what to build. Fitting an equivalent current dipole to the 26-channel QRS field under Monte-Carlo per-sensor field noise and gain error (Methods) maps the microscopic noise onto the clinical localization accuracy (Fig. 5d), which is limited by field noise, not by calibration. The error increases steeply from $\sim 5\,\mathrm{mm}$ at the $\sim 1\,\mathrm{pT}$ of effective per-sensor noise a 5 mm ablation target requires (a raw $\sim 6\,\mathrm{pT}/\sqrt{\mathrm{Hz}}$, after averaging the ~ 340 beats) to beyond $\sim 60\,\mathrm{mm}$ at the raw single-channel ambient. Cross-sensor gain enters only weakly, because the dipole position is fixed by the field pattern and averages down over the 26 sensors, so a realistic 1% gain spread leaves a $\sim 0.37\,\mathrm{mm}$ floor (Supplementary Note 7).

Meeting the target therefore comes down to rejecting the ambient. The raw single-channel ambient sits about tenfold above the required $\sim 6\,\mathrm{pT}/\sqrt{\mathrm{Hz}}$ floor, but the software common-mode rejection to $\sim 4.4\,\mathrm{pT}/\sqrt{\mathrm{Hz}}$ carries the device onto the 5 mm contour. The specification fixes the *required* floor by design while the *achieved* rejection is measured in place, so design effort belongs at the array and the macroscopic system, not the individual sensor's atomic floor.

3 Discussion

A quantum sensor is specified by its headline sensitivity, yet a device tuned to that number alone can miss the accuracy or robustness its application requires. Our central result is a quantitative, per-mechanism map that resolves this, tying each metric to the mechanism that sets it and predicting

the metric from it. For the NV-diamond ensemble and operating point studied here, spin dephasing governs sensitivity, the thermal ground-state shift governs accuracy, and optical leakage governs robustness, each with a signed share of its budget. Along a locus of identical nominal sensitivity, the accuracy bias spans 8 to 1500 nT at the assumed $\Delta T = 0.1$ K drift, a factor of $\sim$200 that scales with the drift, so two devices indistinguishable by their headline sensitivity can differ in accuracy by orders of magnitude. No single figure of merit therefore certifies a design, because which mechanism dominates depends on the metric, the device, and the operating environment, and is not a fixed property of the sensor.

The per-source error budgets maintained by mature precision instruments assume their contributions add. A quantum sensor reads its field through a nonlinear, dissipative optical cycle whose mechanisms interact, so its budget must instead be interaction-aware, which is what the Shapley decomposition supplies. Quantum metrology tools have taken a different route, compressing performance into a single scalar such as a coherence time, one Cramér–Rao floor [17, 18], or a closed-form sensitivity that presupposes which imperfection dominates [1]. Our framework is complementary, decomposing sensitivity, accuracy, and robustness into the device mechanisms that set each. Its outputs are of direct experimental relevance, because each metric traces to a measurable intermediate. An accuracy target translates into a temperature-stability specification, and a robustness target, dominated by optical leakage, selects a low-leakage acousto-optic or integrated-photonics switch [7]. Because the model is grounded in and updated from measured device data, it functions as a digital twin of the sensor, so each fabricated device maps to its own attributed budget, and the model predicts the gain from removing each limiter.

Because the figures of merit are differentiable, the trade-off surface is in principle navigable by gradient, so design can proceed by specification instead of by iteration. The mechanism-attributed landscape could seed optimization- or learning-based control with per-mechanism gradients in place of a scalar reward [59]. As a first-principles forward model, the twin can also generate physically accurate synthetic training data for those models, pairing the raw fluorescence and calibration traces with ground-truth mechanism labels that no measurement can supply, sampled cheaply across regimes where real data are scarce. We stop at attribution, however: the framework identifies the limiting mechanism and the lever to target for each metric, but we do not carry out the implied redesign or demonstrate the resulting metric improvement here. The attribution also sorts where each fix belongs. A share that lands on a parameter frozen at fabrication, the material, the control-line geometry, or the photonic switch, calls for a design-time change, whereas a share on a knob adjustable in operation, such as a pulse coefficient or a thermal setpoint, can be corrected online while the sensor runs. Supplementary Note 10 develops this runtime half into a closed-loop, self-adapting calibration scheme.

Looking forward, the most immediate next step is to close that loop experimentally, confirming a predicted metric gain on a device so that attribution becomes validated design. Making such forecasts trustworthy across regimes motivates the remaining refinements. The stochastic environment is modeled as Gaussian with literature-calibrated spectra, and extending the noise generators to strongly non-Gaussian processes such as random-telegraph noise, whose effects concentrate in the tails, would

guard against mis-ranked mechanisms. The attribution is also conditional on identifiability, since mechanisms with nearly degenerate signatures in the available observables are constrained only in sum, and richer readout channels would sharpen the partition.

In conclusion, one open quantum system model ties each performance metric of a quantum sensor to the physical mechanism that sets it from measured device data. The workflow is portable, as the transfer from an NV-diamond ensemble to a cesium optically pumped magnetometer array shows, and the same accounting applies to any sensor that is an open quantum system read out by photon counting, such as an optical clock, an atom interferometer, or a Rydberg electrometer. The practical implication is that a sensor can be designed against every metric its application needs at once, with the mechanism limiting each one identified at design time, before the device is built and characterized.

4 Methods

This section records the reproducibility detail for the construction of Sec. 2.1. Full derivations and parameter tables are in the Supplementary Information (SN1–SN10), and we do not repeat methods published elsewhere.

Open-system model of the NV center

The construction of Sec. 2.1 is instantiated for the NV center [50] as a ten-level open system (full Hamiltonian and collapse set in Supplementary Note 1): the spin-1 ground state ${}^3A_2$, the orbitally resolved spin-1 excited state ${}^3E$ ($\{E_x, E_y\}\otimes\{m_s{=}+1, 0, -1\}$) and a single metastable singlet (lumping ${}^1A_1/{}^1E$). The coherent part uses a rotating-wave Hamiltonian (in frequency units) addressing the $|0\rangle \leftrightarrow |{-1}\rangle$ transition, the sensed field entering as the effective detuning $\Delta_{\text{eff}}(t) = \big[D_{\text{gs}}(T) - \gamma_e(B_0 + b + \delta b(t))\big] - \omega_{\text{mw}} - \dot{\phi}(t)/2\pi$ with $\gamma_e \approx 28.03\,\text{GHz/T}$ ($g = 2.003$). The excited-state coherences are treated secularly, the solver imposing the fast-orbital-averaging limit that phonon-driven mixing justifies above $\sim$80 K. The thirty-one jump operators are built from first-principles rates (following the rate model in [49]): optical pumping ($\propto$ laser power), radiative decay, spin-selective inter-system crossing (the C_{3v} symmetry-adapted rates projected from the zero-field basis), metastable-singlet decay with a Bose–Einstein temperature-dependent lifetime and spin-dependent branching, phonon-driven orbital hopping [49] (evaluated as the full Debye integral, with a one-phonon and a two-phonon Raman term whose limiting forms scale as Δ^3 and T^5 and are related by detailed balance), longitudinal relaxation, and the Markovian dephasing $\gamma_\phi = 2/T_2^\star$ (disabled when synthesized field noise is the sole dephasing source). The splitting drifts with temperature in a cubic fashion $D_{\text{gs}}(T)$ [60] (local slope $\approx -79\,\text{kHz/K}$ at 300 K, an apparent field of $\approx 2.8\,\mu\text{T/K}$), anchored to its 0 K zero-field value 2.8777 GHz, with the slope evaluated at the 300 K reference (room-temperature value $\approx 2.870\,\text{GHz}$). Strain shifts map to $\{D_{\text{gs}}, E_\perp, \phi_E\}$ through the measured deformation-potential coefficients, and spin-state optical leakage ε is modeled as residual laser power on the optical-pumping channel, so it both lowers the contrast and shifts the readout baseline.

Tangent solve and pulses

The field-derivative co-propagation of Sec. 2.1 is the standard tangent-linear (forward-sensitivity) construction of open-system estimation [29], co-propagating the state and its field derivative $G_b = \partial_b \rho$ under

$$\dot{\rho} = \mathcal{L}(t;\theta)[\rho], \qquad \dot{G}_b = \mathcal{L}(t;\theta)[G_b] + \big(\partial_b \mathcal{L}(t;\theta)\big)[\rho], \tag{4}$$

as derived in Supplementary Note 1. Here it uses the bare Zeeman derivative $\partial_b H = (\gamma_e/2)\,\sigma_z^{(0,-1)}$ on the addressed qubit, with $\sigma_z^{(0,-1)} = |0\rangle\langle 0| - |-1\rangle\langle -1|$, so the $|0\rangle \leftrightarrow |-1\rangle$ detuning responds as $\gamma_e b$ (the field-independent form, so the Fisher information does not spuriously vanish at zero bias), with the spin readout $p_{\mathrm{spin}} = \mathrm{Tr}[M\rho]$ and its exact derivative $\mathrm{Tr}[M\,G_b]$ taken from a single solve. The two $\pi/2$ pulses of a Ramsey sequence are folded into the coherent initial state $(|0\rangle + |-1\rangle)/\sqrt{2}$ and the readout projector (ideal, instantaneous pulses). The free-precession interval is integrated with an adaptive time step, at least 30 per solve and capped by the fastest dissipative rate for numerical stability.

Optical readout and photon statistics

The spin projection is converted to the measured optical signal through the readout contrast C, computed from first principles by a separate ten-level optical-cycle solve (Supplementary Note 3) under continuous-wave pumping at an optical-pumping rate Γ_{opt} set by the laser power: the photoluminescence transient $\mathrm{PL}(t) = \Gamma_{\mathrm{rad}} \sum_k \rho_{kk}^{\mathrm{ES}}(t)$ is integrated over a 300 ns readout window for the spin prepared in m_s=0 versus m_s=±1, and $C = 1 - \int \mathrm{PL}_{m_s=\pm 1} / \int \mathrm{PL}_{m_s=0}$. The contrast is thus a first-principles function of temperature (through the phonon-driven intersystem-crossing and excited-state rates [49]), transverse strain (through excited-state orbital mixing), and laser power (through Γ_{opt} and its optical-saturation roll-off), with no parameter fitted to the readout it predicts; C is averaged over the strain angle (eight angles) and tabulated on a $(T, E_\perp)$ grid ($E_\perp \in [0.05, 2]$ GHz) at the operating pump rate for interpolation, with optional averaging over laser-intensity noise. The photon Fisher information of Sec. 2.1 is evaluated at the readout time. For the literature-grounded budget of Fig. 2c the strain-averaged table gives $\langle C \rangle = 0.46$ with $E_\perp$ ensemble (median 150 MHz, $\sigma_{\log} = 0.35$), reduced by a readout-fidelity/collection factor 0.30 to $C_{\mathrm{eff}} = 0.137$. The photon budget at $[N] = 10$ ppm is $N_{\mathrm{det}} = 4.04 \times 10^8$ detected photons per shot, and the sequence overhead (dead time) is 1 μs. The calibration fringe of Fig. 2b instead shows the bare optical contrast ($C = 0.46$ vs 0.17) so that the strain lever is visible directly.

Charge-state and saturation measurements

The $\mathrm{NV}^-/\mathrm{NV}^{\mathrm{T}}$ charge fraction (with NV^{T} the total NV concentration) and fluorescence saturation of Sec. 2.3 were measured by confocal photoluminescence spectroscopy on two diamond growths (the 0.3 and 4.5 ppm growths) under 532 nm excitation through a 100× objective. At each laser power density, spanning nearly three decades, the photoluminescence spectrum was decomposed by non-negative least squares into its NV^- and NV^0 components (Supplementary Note 8) to give the charge

fraction $f_{\mathrm{NV^-}} = [\mathrm{NV^-}]/[\mathrm{NV^T}]$, and the $\mathrm{NV^-}$ and total count rates were recorded (1 ms integration per point) to give the saturation curves. The measured $f_{\mathrm{NV^-}}(P)$ enters the model as the power-dependent effective contrast $C_{\mathrm{eff}}(P) = C_{\mathrm{optical}}(P)\, f_{\mathrm{NV^-}}(P)$, and the saturation curves fix the detected-photon budget $N_{\mathrm{det}}(P)$ (Sec. 2.3; Supplementary Note 3).

Stochastic terms and frequency response

The open quantum system model also admits time-dependent stochastic terms in $H_{\mathrm{env}}(t)$, synthesized from power spectral densities into time traces and injected as a Zeeman detuning or drive noise. The deterministic main-text results do not use these traces, and the full time-dependent-noise treatment, including the sensitivity under $1/f$, Ornstein–Uhlenbeck, and white-noise baths, is developed in Supplementary Note 5. A sensor's frequency response $\eta_B(f)$ is obtained by driving the model with a deterministic field oscillating at frequency f and summing the two signal quadratures of the Fisher information, the construction used for the OPM cardiac-band sensitivity of Sec. 2.4.

Estimation and attribution

The accuracy bias of Sec. 2.1 is computed by inverting the measured (Poisson-sampled, contrast-scaled) signal through a calibration generated at a reference operating point (so a thermal $D_{\mathrm{gs}}(T)$ shift surfaces as an apparent field $\approx -\Delta D_{\mathrm{gs}}/\gamma_e$ and leakage ε as a baseline offset), with field recovery by a linear, slope-based calibration inverse, and an instrument-grounded variant that re-runs the master equation to separate a removable calibration error from the irreducible bias. The robustness budget R_{lin} is assembled from centered finite-difference log-gradients $\partial_{\theta_i} \log \eta_B$, with the Monte-Carlo R_{env} used where the response is non-linear. The per-parameter contributions to $R_{\mathrm{lin}}^2 - 1$ add exactly by construction.

Per-mechanism attribution of sensitivity and accuracy uses *Shapley values*, the unique axiomatic allocation of a total among interacting contributors [35]. Its efficiency and uniqueness are set out in Supplementary Note 4. We treat the M error mechanisms as players: each is switched on or off in the open-system solve (leakage $\varepsilon \to 0$, $T_1 \to \infty$, dephasing off, a noise spectrum zeroed, or the calibration recovered for the thermal channel), and for a subset S of active mechanisms we evaluate the metric $v(S)$ from that solve (so $v(\varnothing)$ is the ideal device and $v(\mathcal{M})$ the full one). Mechanism j is assigned

$$\phi_j = \sum_{S \subseteq \mathcal{M} \setminus \{j\}} \frac{|S|!\,(M - |S| - 1)!}{M!} \big[v(S \cup \{j\}) - v(S) \big], \tag{5}$$

its marginal effect $v(S \cup \{j\}) - v(S)$ averaged over every order in which the mechanisms could be activated. The ϕ_j are the *unique* attribution that sums exactly to the total, $\sum_j \phi_j = v(\mathcal{M}) - v(\varnothing)$ (efficiency), gives identical mechanisms equal credit (symmetry), assigns zero to an inert mechanism (null player), and is linear. Being signed, a source that improves a metric receives a negative ϕ_j. This distributes the inter-mechanism interactions that a single leave-one-out difference, $v(\mathcal{M}) - v(\mathcal{M} \setminus \{j\})$, would leave in a non-additive residual. The two coincide when mechanisms act independently, the sensing analog of Matthiessen's rule for additive scattering resistivities in charge transport [34];

strongly interacting error processes admit no such additive decomposition. We find independence holds for sensitivity (leave-one-out and Shapley agree to within a few percent) but *not* for the accuracy bias, where a single leave-one-out ordering leaves ∼90% of the total in the residual, so the Shapley partition is required there. The interaction is physical, not a numerical artifact: the bias is recovered by inverting a nonlinear calibration, so two sources that each displace the operating point combine nonlinearly, the apparent-field offset from the thermal D_{gs} shift depends on where leakage has already moved the fringe baseline, and the bias ascribed to any one source by switching it off alone therefore depends on which others are still active. Averaging each mechanism's marginal effect over *all* activation orders (Eq. 5) is what removes that ordering dependence and returns a unique, order-independent share. This is the same cooperative-game construction recently imported into quantum science for related attribution tasks, assigning a quantum circuit's output to its individual gates in explainable quantum machine learning [38], with the error mechanisms here playing the role those gates play there. Sensitivity is attributed in η_B above the shot-noise floor $v(\varnothing)$ (the statistical limit at the detected-photon budget $N_{\mathrm{det}} = N_{\mathrm{eff}}\, n_{\mathrm{avg}}$, with N_{eff} the number of NV^- emitters and n_{avg} the mean detected photons per emitter per readout window), common to the Fisher information and to the Poisson sampling of the accuracy bias, so shot noise enters both metrics consistently. The 2^M coalition solves are evaluated at the representative operating point. For the design-space maps (Fig. 3c) they are repeated, together with the finite-difference robustness gradients, on an 80×80 $(T_2^\star, \varepsilon)$ grid, each cell colored by its dominant mechanism: $\arg\max_j |\phi_j|$ for sensitivity and accuracy, and the largest contribution for R_{lin}. On the accuracy surface the mapped quantity is the full root-mean-square error $\sqrt{\beta^2 + \sigma_b^2}$: the systematic bias β, carrying an assumed $\Delta T = 0.1\,\mathrm{K}$ thermal offset, added in quadrature with the shot-noise precision floor $\sigma_b = \eta_B/\sqrt{\tau + t_{\mathrm{dead}}}$ at the same detected-photon budget N_{det}, and normalized to a $1\,\mu\mathrm{T}$ reference field, so the coherence-washed low-$T_2^\star$ corner appears at its true shot-dominated error rather than at the bias alone.

The reported attribution shares carry two distinct uncertainties, quantified separately. The accuracy shares are the output of a photon-counting estimator, so their statistical uncertainty follows from resampling: holding the 2^M coalition solves fixed, we redraw the Poisson optical readout of each coalition (one $N_{\mathrm{det}} = 5 \times 10^6$-photon measurement per coalition) over 500 realizations, propagate each through Eq. 5, and report the 2.5 to 97.5 percentile interval of every share. Averaging over K repeated measurements narrows these by $\sqrt{K}$. Because the signed accuracy contributions partially cancel (the net bias is much smaller than the gross), we quote them as absolute apparent-field offsets rather than as percentages of the net. The sensitivity shares are a deterministic Cramér–Rao quantity and the robustness fractions a deterministic finite-difference functional, so neither carries shot-noise seed variance. Their uncertainty is parametric, reported by sweeping the dominant input, the measured $T_2^\star$ range for sensitivity and a factor-of-two range in the assumed leakage-drift budget for robustness. Across every realization and parameter draw the dominant mechanism for each metric is unchanged.

For the power-knob budget of Fig. 4d the players are the charge-state dilution (off: the measured NV^- fraction set to one at the measured photon count, so the counterfactual keeps N_{det}), the optical-cycle saturation (off: $C_{\mathrm{optical}}(P)$ held at its low-power value), and dephasing (off: $T_2^\star \to \infty$, the fringe re-solved), with the interrogation time and fringe bias held at the full-device operating values and

$v(\varnothing)$ the shot floor at the measured $N_{\mathrm{det}}(P)$. Optical leakage is excluded for lack of a measured $\varepsilon(P)$, and T_1 is held on throughout (negligible at these τ). Because the two contrast mechanisms enter multiplicatively, a leave-one-out accounting deviates from the Shapley shares by up to $\sim$20% of η_B on this axis, so the Shapley partition is used here as well. The stacked total reproduces the Fig. 4c curve to numerical precision. The accuracy decomposition of Fig. 4e is exact without it, the thermal bias β and the Poisson shot-noise floor σ_b combining as $\sqrt{\beta^2 + \sigma_b^2}$ (verified to 0.2%).

Numerical implementation

Both master equations are integrated in batched form: ρ is carried as a $[d, d, B]$ complex-128 tensor and the right-hand side applied by batched matrix products, with the batch axis $B = B_r \cdot B_\xi \cdot B_s$ spanning spatial/per-pixel realizations, noise realizations and parameter sweeps, so per-mechanism toggles of the dissipative channels and dense operating-point grids evaluate in one propagation (stochastic-trace toggles enter through the noise-realization axis instead). Three integrators are available: fixed-step RK4; a second-order Strang splitting whose Hamiltonian sub-step is an exact matrix exponential (Padé scaling-and-squaring); and an adaptive embedded Runge–Kutta reference. For Ramsey sensing the dynamics close on the $\{|0\rangle, |-1\rangle\}$ subspace, so a two-level reduction reproduces the ten-level free-precession result at a large speed-up and is used for the noise-averaged sweeps. The optical-cycle readout integrates the photoluminescence over 601 time steps across the 300 ns window, and the sensitivity is verified grid-independent (Supplementary Note 2). The benchmarks named in Sec. 2.1 are met to the following tolerances: the tangent derivative agrees with finite differencing to $< 10^{-4}$, and the NumPy CPU implementation is the bit-level reference for the GPU engine, with parity enforced by a dedicated CPU/GPU test suite (10^{-6}–10^{-7} tolerances).

GPU acceleration

The batched formulation runs on a single node of eight NVIDIA H100 GPUs on the cuQuantum [44] `cuDensityMat` backend, following the GPU-accelerated, differentiable open-system simulation paradigm of recent solvers [43]. Because per-mechanism rate toggles, noise realizations and design-space grids all map onto the batch axis $B = B_r \cdot B_\xi \cdot B_s$, an entire attribution or operating-point sweep is one large propagation rather than thousands of serial solves. For the metric-dependent attribution of Fig. 3, the 2^M mechanism coalitions and the full $80 \times 80 = 6{,}400$-point $(T_2^\star, \varepsilon)$ design-space grid populate this batch axis, so each panel is computed as a single batched propagation rather than cell by cell. Across representative design-space sweeps and noise-averaging benchmarks (up to 4096 realizations, a capability the main-text figures do not exercise) this runs 50–100$\times$ faster than the NumPy CPU reference. Kernel construction, the interaction-picture stepping, the GPU integrators, and the multi-GPU decomposition are in Supplementary Note 6, which also shows the per-pixel metric-dependent attribution resolved over a dense spatial grid at GPU scale, a spatial map not reproduced among the main-text figures.

Scalar optically-pumped magnetometer

For the cross-platform comparison the open-system, tangent-master-equation, Fisher and attribution infrastructure is instantiated for a scalar (total-field) cesium optically-pumped magnetometer (full model in Supplementary Note 7) at the geomagnetic field ($\sim 50\,\mu$T): a different Hamiltonian (the ground-state Breit–Rabi structure) and three incoherent relaxation channels, spin destruction, spin exchange, and optical pumping, written as mean-field (not Lindblad-jump) operators (SI), on the shared Strang-integrated solver. *Sensitivity* is the Cramér–Rao bound, the tangent master equation co-propagates ρ and $\partial_B\rho$, the field Fisher information is read from $\partial_B\langle F_y\rangle$, and η_B is optimized over the interrogation time. The standard-quantum-limit floor $1/(\gamma_{\mathrm{rad}}\sqrt{N_{\mathrm{at}}T_2})$ and the measured single-channel and common-mode-rejected floors (median Welch power spectral densities of the recording) are reported alongside. *Robustness* is the ambient field-noise floor in the cardiac band and its common-mode rejection, implemented as subtraction of the leading principal components of the multi-sensor recording. The calculated budgets follow directly from these: the sensitivity bars split η_B in quadrature into the spin-projection (SQL) floor and the residual photon-readout shot noise ($\eta_B^2 = \mathrm{SQL}^2 + \mathrm{shot}^2$, so the readout term is $\sqrt{\eta_B^2 - \mathrm{SQL}^2}$), and the robustness bars are the ambient noise in the low band (1 to 10 Hz), its residual after common-mode rejection, and the sensor floor. The design-by-specification panel (Fig. 5d) is a localization design space: the clinical dipole-localization error over per-sensor field noise and cross-sensor calibration, from an equivalent-current-dipole inverse fit to the recorded QRS field with additive per-sensor field noise and multiplicative per-sensor calibration error (Monte Carlo). The field-noise axis is the effective per-sensor noise ($\mathrm{ASD} \times \sqrt{\mathrm{BW}}/\sqrt{N_{\mathrm{epoch}}}$), marked at the spin-projection floor, the common-mode-rejection floor and the single-channel ambient. Here ASD is the per-sensor amplitude spectral density of the field noise ($\mathrm{T}/\sqrt{\mathrm{Hz}}$), BW the signal bandwidth over which it is integrated, and $N_{\mathrm{epoch}} \approx 340$ the number of averaged cardiac beats. The 5 mm contour sets the specification. Full model, rates, and the heading-error and localization derivations are in SN7.

Magnetocardiography recording

The magnetocardiogram of Sec. 2.4 was recorded with a 26-channel scalar (total-field) optically-pumped-magnetometer array operated unshielded from the geomagnetic field ($\sim 50\,\mu$T) from an author-participant (Ethics declarations). The array and its acquisition are described in Ref. [9]. The analyzed trace is the average of ~ 340 cardiac cycles, from which the QRS-peak current-dipole field is fit. Provenance and validation are in Supplementary Note 7.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to confidentiality, but are available from the corresponding author on reasonable request.

Code availability

Code implementing the tangent-master-equation solver and attribution pipeline is available from the corresponding authors upon reasonable request, subject to MITRE public release review. A public repository will be made available following completion of that process.

Ethics declarations

The human recording used in this study was a single, non-invasive, non-clinical technical measurement obtained from an author-participant and was not collected or used for diagnosis, treatment, or clinical decision-making. The participant provided written consent for the recording and for publication of the manuscript and the resulting processed MCG data. No raw identifiable data will be shared; processed MCG data may be made available from the corresponding author on reasonable request, subject to applicable privacy and data-governance requirements.

Acknowledgments

We thank Bonnie Marlow (MITRE) and Elica Kyoseva (NVIDIA) for high-level technical guidance on this work. This work is supported by the MITRE Independent Research and Development Program. Portions of this technical data were produced for the U. S. Government under Contract No. FA870225CB001 and W56KGU-18-D-0004, and is subject to the Rights in Technical Data-Noncommercial Items Clause DFARS 252.227-7013 (FEB 2014).

Approved for Public Release; Distribution Unlimited. Public Release Case Number 26-1803. © 2026 The MITRE Corporation. All rights reserved.

Competing interests

All authors declare no financial or non-financial competing interests.

Author contributions

N.L. developed the computational framework, including the Shapley attribution, the tangent master equation, and the digital-twin workflow. J.H. provided application-level direction. C.W., M.C., K.J.R., Y.C., M.S.M., and M.Doherty (Quantum Brilliance) contributed nitrogen-vacancy diamond expertise and the measured photoluminescence charge-state (NV^-) data. M.Dong contributed photonics and quantum-sensor-design expertise, including the treatment of optical-modulator imperfections. B.R. introduced the photon-shot-noise-limited derivation of the Fisher information. D.L., B.K., and J.-S.K. developed the GPU-accelerated framework and the multi-GPU parallelization. G.I., S.B., and E.P. contributed magnetocardiography expertise and the scalar-OPM array recording used to ground the cross-platform demonstration. S.O. contributed quantum-sensor expertise. N.L. wrote the manuscript

with input from all authors, and all authors reviewed and approved the final version. N.L. and J.H. are the corresponding authors.

References

[1] Barry, J. F. et al. Sensitivity optimization for NV-diamond magnetometry. *Rev. Mod. Phys.* **92,** 015004 (2020).

[2] Degen, C. L., Reinhard, F. & Cappellaro, P. Quantum sensing. *Rev. Mod. Phys.* **89,** 035002 (2017).

[3] Schirhagl, R., Chang, K., Loretz, M. & Degen, C. L. Nitrogen-vacancy centers in diamond: nanoscale sensors for physics and biology. *Annu. Rev. Phys. Chem.* **65,** 83 (2014).

[4] Rondin, L. et al. Magnetometry with nitrogen-vacancy defects in diamond. *Rep. Prog. Phys.* **77,** 056503 (2014).

[5] Budker, D. & Romalis, M. Optical magnetometry. *Nat. Phys.* **3,** 227 (2007).

[6] Kominis, I. K., Kornack, T. W., Allred, J. C. & Romalis, M. V. A subfemtotesla multichannel atomic magnetometer. *Nature* **422,** 596 (2003).

[7] Leclerc, N. et al. Miniature magnetometers for magnetic navigation using NV diamond on piezoelectric photonic circuits. In *2025 Conference on Lasers and Electro-Optics (CLEO)* (IEEE, 2025). IEEE Xplore 11190544.

[8] Jensen, K. et al. Magnetocardiography on an isolated animal heart with a room-temperature optically pumped magnetometer. *Sci. Rep.* **8,** 16218 (2018).

[9] Iwata, G. Z. et al. Bedside magnetocardiography with a scalar sensor array. *Sensors* **24,** 5402 (2024).

[10] Iwata, G. Z. et al. AI and quantum sensors: realization of a safe and effective unshielded bedside magnetocardiogram to detect ischemia in the emergency room. *Circulation* **150,** A4140246 (2024).

[11] Leroy, A. et al. Emergency department triage of patients with suspected acute coronary syndrome using unshielded magnetocardiography. *J. Am. Coll. Cardiol.* **87,** A1206 (2026).

[12] Leroy, A. et al. Machine learning-based detection of coronary artery disease using unshielded magnetocardiography. *J. Am. Coll. Cardiol.* **87,** A802 (2026).

[13] Boto, E. et al. Moving magnetoencephalography towards real-world applications with a wearable system. *Nature* **555,** 657 (2018).

[14] Schloss, J. M., Barry, J. F., Turner, M. J. & Walsworth, R. L. Simultaneous broadband vector magnetometry using solid-state spins. *Phys. Rev. Appl.* **10,** 034044 (2018).

[15] Levine, E. V. et al. Principles and techniques of the quantum diamond microscope. *Nanophotonics* **8,** 1945 (2019).

[16] Leclerc, N., Oliver, S. & Dong, M. Spatiotemporal arbitrary waveform quantum sensing. In *2025 Conference on Lasers and Electro-Optics (CLEO)* (IEEE, 2025). IEEE Xplore 11190516.

[17] Braunstein, S. L. & Caves, C. M. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **72,** 3439 (1994).

[18] Helstrom, C. W. *Quantum Detection and Estimation Theory* (Academic Press, 1976).

[19] Cywiński, Ł., Lutchyn, R. M., Nave, C. P. & Das Sarma, S. How to enhance dephasing time in superconducting qubits. *Phys. Rev. B* **77,** 174509 (2008).

[20] Biercuk, M. J., Doherty, A. C. & Uys, H. Dynamical decoupling sequence construction as a filter-design problem. *J. Phys. B* **44,** 154002 (2011).

[21] Álvarez, G. A. & Suter, D. Measuring the spectrum of colored noise by dynamical decoupling. *Phys. Rev. Lett.* **107,** 230501 (2011).

[22] Kaubruegger, R., Shankar, A., Vasilyev, D. V. & Zoller, P. Optimal and variational multiparameter quantum metrology and vector-field sensing. *PRX Quantum* **4,** 020333 (2023).

[23] Albarelli, F., Barbieri, M., Genoni, M. G. & Gianani, I. A perspective on multiparameter quantum metrology: from theoretical tools to applications in quantum imaging. *Phys. Lett. A* **384,** 126311 (2020).

[24] Liu, J., Yuan, H., Lu, X.-M. & Wang, X. Quantum Fisher information matrix and multiparameter estimation. *J. Phys. A* **53,** 023001 (2020).

[25] Song, C.-G. & Cai, Q.-y. Excessive precision compromises accuracy even with unlimited resources due to the trade-off in quantum metrology. *npj Quantum Inf.* **11,** 115 (2025).

[26] Rojkov, I., Layden, D., Cappellaro, P., Home, J. & Reiter, F. Bias in error-corrected quantum sensing. *Phys. Rev. Lett.* **128,** 140503 (2022).

[27] Tsunaki, L., Singh, A., Trofimov, S. & Naydenov, B. Digital twin simulations toolbox of the nitrogen-vacancy center in diamond. *Adv. Quantum Technol.* **9,** e00691 (2026).

[28] Pandey, S., Magapu, A., Anand, P., Khandelwal, A. & Girish Chandra, M. QDsiM: A noise-aware simulation toolkit for quantum diamond microscope. Preprint at arXiv:2601.11649 (2026).

[29] Zhang, M. et al. QuanEstimation: An open-source toolkit for quantum parameter estimation. *Phys. Rev. Res.* **4,** 043057 (2022).

[30] López-Pardo, V. & Rothkopf, A. Optimal observables for (non-)equilibrium quantum metrology from the master equation. Preprint at arXiv:2506.23600 (2025).

[31] Nakajima, S. & Utsumi, Y. Symmetric-logarithmic-derivative Fisher information for kinetic uncertainty relations. *Phys. Rev. E* **108,** 054136 (2023).

[32] Martynov, D. V. et al. Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Phys. Rev. D* **93,** 112004 (2016).

[33] Brewer, S. M. et al. $^{27}Al^{+}$ quantum-logic clock with a systematic uncertainty below 10^{-18}. *Phys. Rev. Lett.* **123,** 033201 (2019).

[34] Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Holt, Rinehart and Winston, 1976).

[35] Shapley, L. S. A value for n-person games. In *Contributions to the Theory of Games II* (eds Kuhn, H. W. & Tucker, A. W.) 307–317 (Princeton Univ. Press, 1953).

[36] Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (2017).

[37] Owen, A. B. Sobol' indices and Shapley value. *SIAM/ASA J. Uncertain. Quantif.* **2,** 245 (2014).

[38] Heese, R. et al. Explaining quantum circuits with Shapley values: towards explainable quantum machine learning. *Quantum Mach. Intell.* **7,** 27 (2025).

[39] Jameson, A. Aerodynamic design via control theory. *J. Sci. Comput.* **3,** 233 (1988).

[40] Courtier, P., Thépaut, J.-N. & Hollingsworth, A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120,** 1367 (1994).

[41] Lee, C.-W., Lee, J. H., Joo, J. & Seok, H. Quantum Fisher information of an optomechanical force sensor driven by a squeezed vacuum field. *Opt. Express* **30,** 25249 (2022).

[42] Qvarfort, S., Serafini, A., Barker, P. F. & Bose, S. Gravimetry through non-linear optomechanics. *Nat. Commun.* **9,** 3690 (2018).

[43] Guilmin, P., Bocquet, A., Genois, É., Weiss, D. & Gautier, R. dynamiqs: an open-source Python library for GPU-accelerated and differentiable simulation of quantum systems. GitHub https://github.com/dynamiqs/dynamiqs (2025).

[44] Bayraktar, H. et al. cuQuantum SDK: a high-performance library for accelerating quantum science. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)* 1050–1061 (IEEE, 2023).

[45] Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbrüggen, T. & Glaser, S. J. Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *J. Magn. Reson.* **172,** 296 (2005).

[46] Balasubramanian, G. et al. Ultralong spin coherence time in isotopically engineered diamond. *Nat. Mater.* **8,** 383 (2009).

[47] Edmonds, A. M. et al. Characterisation of CVD diamond with high concentrations of nitrogen for magnetic-field sensing applications. *Mater. Quantum Technol.* **1,** 025001 (2021).

[48] Stürner, F. M. et al. Integrated and portable magnetometer based on nitrogen-vacancy ensembles in diamond. *Adv. Quantum Technol.* **4,** 2000111 (2021).

[49] Ernst, S., Scheidegger, P. J., Diesch, S. & Degen, C. L. Modeling temperature-dependent population dynamics in the excited state of the nitrogen-vacancy center in diamond. *Phys. Rev. B* **108,** 085203 (2023).

[50] Doherty, M. W. et al. The nitrogen-vacancy colour centre in diamond. *Phys. Rep.* **528,** 1 (2013).

[51] Goldman, M. L. et al. Phonon-induced population dynamics and intersystem crossing in nitrogen-vacancy centers. *Phys. Rev. Lett.* **114,** 145502 (2015).

[52] Goldman, M. L. et al. State-selective intersystem crossing in nitrogen-vacancy centers. *Phys. Rev. B* **91,** 165201 (2015).

[53] Giovannetti, V., Lloyd, S. & Maccone, L. Advances in quantum metrology. *Nat. Photonics* **5,** 222 (2011).

[54] Paris, M. G. A. Quantum estimation for quantum technology. *Int. J. Quantum Inf.* **7,** 125 (2009).

[55] Tóth, G. & Apellaniz, I. Quantum metrology from a quantum information science perspective. *J. Phys. A* **47,** 424006 (2014).

[56] Bauch, E. et al. Decoherence of ensembles of nitrogen-vacancy centers in diamond. *Phys. Rev. B* **102,** 134210 (2020).

[57] Aslam, N., Waldherr, G., Neumann, P., Jelezko, F. & Wrachtrup, J. Photo-induced ionization dynamics of the nitrogen vacancy defect in diamond investigated by single-shot charge state detection. *New J. Phys.* **15,** 013064 (2013).

[58] Steck, D. A. Cesium D Line Data. http://steck.us/alkalidata.

[59] Leclerc, N., Miller, C. & Brawand, N. When does adaptation win? Scaling laws for meta-learning in quantum control. In *Proc. 43rd International Conference on Machine Learning (ICML)* (2026).

[60] Toyli, D. M. et al. Measurement and control of single nitrogen-vacancy center spins above 600 K. *Phys. Rev. X* **2,** 031001 (2012).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.pdf