Quantum Metrology: How Close Can We Get to the Ideal Heisenberg Limit Under Realistic Noise?

Deepinder Sidhu*
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County, MD 21250, USA
sidhu@umbc.edu

July 18, 2026

Abstract

Quantum metrology promises Heisenberg-limited precision, yet every practical platform exhibits a breakdown of this scaling as resources increase. We present a compact, platform-independent framework in which the total sensing resource is decomposed as $N = ML$, where M denotes the number of probes interrogated in parallel and L denotes the coherent sequential interrogation depth. Under realistic decoherence, sequential phase accumulation incurs an exponential noise penalty that imposes a finite bound on usable coherent depth, thereby terminating ideal Heisenberg scaling.

We show that the resulting *Heisenberg-saturation knee* is not a plotting artifact or a channel-specific bound, but the direct consequence of constrained optimization under a fixed coherence budget. For fixed total resource N, contour plots in the (M, L) plane reveal a saddle-point structure at which achievable precision is minimized and beyond which additional resources cannot be converted into coherent phase sensitivity. This geometric interpretation yields a concrete design rule: quantum advantage is maximized not by increasing N indiscriminately, but by optimally allocating resources between parallelization and coherent depth.

This perspective elevates resource partitioning to a foundational principle for scalable quantum sensing and timing, applicable uniformly across optical, atomic, solid-state, and superconducting platforms.

1 Introduction: Why Heisenberg Scaling Always Breaks Down

Heisenberg scaling, $\Delta\phi \propto 1/N$, represents the ideal precision limit of quantum metrology under perfectly coherent resource use [1]. However, no practical sensing platform sustains this behavior as resources increase. Across optical, atomic, solid-state, and superconducting systems, achievable precision inevitably deviates from Heisenberg scaling and bends toward the standard quantum limit (SQL), $\Delta\phi \propto 1/\sqrt{N}$.

*This work is partially supported by Cyberspace Analytics Corporation.

In this work, the term *noise* is used in a broad, operational sense. It includes all physical processes that reduce quantum-state distinguishability and therefore degrade achievable precision, including decoherence (dephasing, amplitude damping, depolarization), loss (photon or particle loss), environmental coupling (Markovian and non-Markovian channels), and operational imperfections that limit coherence length or interrogation depth. All such effects are captured at the level of quantum Fisher information through an effective architecture-level coherence retention function.

We model coherence retention through a single effective function (specified by the noise model in Section 2)

$$\Lambda(L) \in (0, 1], \tag{1}$$

where L denotes coherent sequential interrogation depth. The specific functional form of $\Lambda(L)$ follows from the adopted noise model (Section 2), and will be instantiated for Markovian pure dephasing.

The central challenge is therefore not merely the presence of noise, but understanding *how* noise reshapes scaling laws and determines the regime in which quantum advantage can be retained. This work provides a minimal, platform-independent answer.

By expressing the total sensing resource as $N = ML$, where M denotes spatial (parallel) depth and L denotes coherent sequential depth, we show that increasing L is fundamentally bounded by coherence loss. This produces a universal *Heisenberg-saturation knee*: beyond a noise-determined optimum, additional resources cannot be converted into coherent phase sensitivity. Once this bound is reached, further performance improvement requires reallocating resources into parallelization M, or more generally, into coherence-matched resource partitions.

For optical interferometers, γ incorporates loss and phase diffusion; for atomic clocks it captures collective dephasing; for NV-center sensors it reflects control noise and inhomogeneous broadening. The same effective description applies across all of them.

1.1 Interpreting Precision: $\Delta\phi$ and the Role of $R(N)$

In this framework, each sensor is treated as a phase-estimation device: an unknown parameter ϕ is encoded through an interrogation process, and the effective quantum Fisher information F_Q^{eff} quantifies how distinguishable the resulting probe states remain under realistic noise.

Throughout this work, all precision bounds are evaluated under a common interrogation-time convention: either the sensing duration is fixed across architectures or absorbed consistently into the definition of the total resource N. Under this assumption, time-dependent prefactors do not affect relative scaling comparisons.

From quantum estimation theory, the achievable precision satisfies the quantum Cramér–Rao bound

$$\Delta\phi_{\text{ach}} \geq \frac{1}{\sqrt{F_Q^{\text{eff}}}}. \tag{2}$$

We assume that this bound can be saturated up to an $\mathcal{O}(1)$ efficiency factor common to all architectures; such constants are therefore suppressed when discussing scaling behavior.

To compare performance against standard limits, we introduce a *precision-based retention function*

$$R(N) \equiv \frac{\Delta\phi_{\text{SQL}}(N)}{\Delta\phi_{\text{ach}}(N)}, \tag{3}$$

where $\Delta\phi_{\mathrm{SQL}}(N) \propto 1/\sqrt{N}$. Under consistent normalization,

$$R(N) = 1 \quad \text{(SQL)}, \qquad R(N) \propto \sqrt{N} \quad \text{(ideal Heisenberg scaling)}.$$

Thus, in all physically meaningful regimes,

$$1 \;\le\; R(N) \;\le\; \sqrt{N}. \tag{4}$$

In the noise-limited models considered here, $R(N)$ initially increases with N before saturating once coherence loss dominates. The saturation knee marks the boundary beyond which additional resources no longer yield coherent gain.

1.2 Observation on Scale Factors and Repetitions

All precision expressions in this work refer to a single interrogation or architecture-level realization. Numerical prefactors associated with generator normalization, estimator efficiency, and sensing duration are treated as $\mathcal{O}(1)$ and suppressed when analyzing scaling behavior.

An exception is the number of statistically independent repetitions, u_w. With u_w repetitions, the bound becomes

$$\Delta\phi \geq \frac{1}{\sqrt{u_w F_Q}}, \tag{5}$$

introducing a uniform factor $u_w^{-1/2}$. This factor does not affect relative scaling, saturation knees, or optimal resource allocation, and is therefore omitted from all analytic comparisons and figures. All results should thus be interpreted as single-shot or per-architecture baselines.

2 Core Concept: A Finite Coherence Budget

In our framework, we describe the total resource as

$$N = ML, \tag{6}$$

where M denotes the number of probes interrogated in parallel and L denotes the number of coherent sequential phase-imprinting interactions. Importantly, L represents coherent phase accumulation rather than classical statistical repetition.

Realistic noise reduces coherence as L increases. Following Escher *et al.* and Demkowicz-Dobrzański *et al.* [2, 3, 4], we define the effective QFI (for given M and L) as

$$F_Q^{\mathrm{eff}}(L, M) = (LM)^2\, \Lambda(L)^2, \tag{7}$$

Architecture-level optimization. The architecture-level design problem can therefore be written as

$$\max_{L,M\geq 1} F_Q^{\mathrm{eff}}(L, M) \quad \text{s.t.} \quad C(L, M) \leq C_{\max}. \tag{8}$$

with a representative coherence factor We model coherent depth-dependent noise using a standard Markovian pure-dephasing channel. Under independent dephasing with rate γ per sequential layer, the

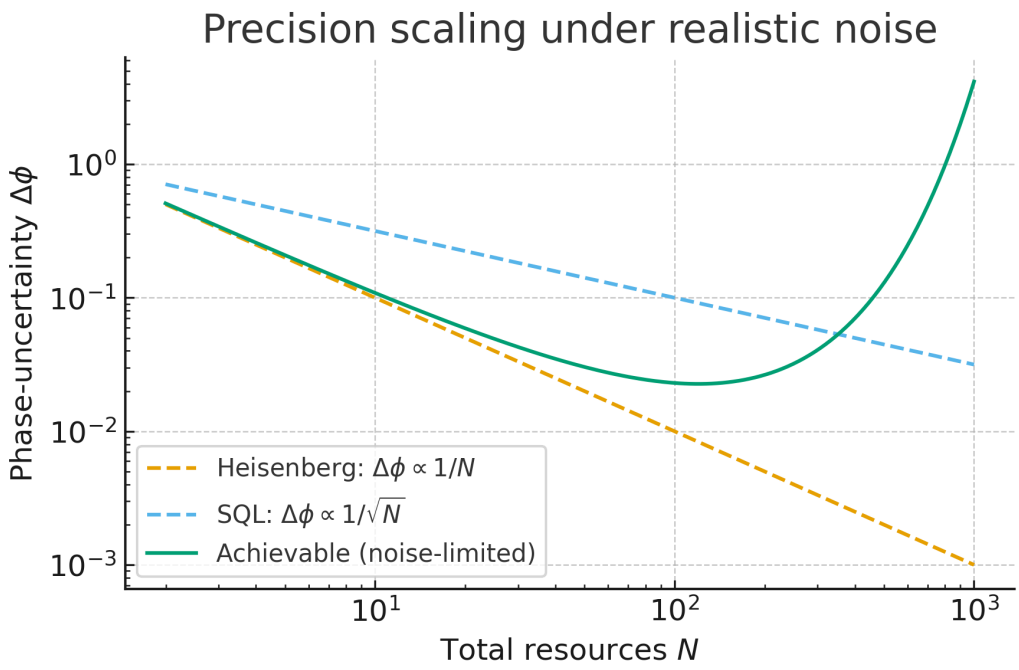

Figure 1: **Precision scaling under realistic noise.** Comparison of ideal Heisenberg scaling ($\propto 1/N$), SQL ($\propto 1/\sqrt{N}$), and achievable noise-limited precision. The minimum marks the *saturation knee* where coherence loss overtakes resource gain. Curves are analytic predictions, not fits.

coherence factor after L sequential operations is

$$\Lambda(L) = e^{-\gamma L}. \tag{9}$$

Equivalently, one may write $F_Q^{\text{eff}}(M, L{:}N{=}ML)$ to emphasize that the L-dependence is physical even when the total budget N is fixed.

The achievable precision scales as

$$\Delta\phi_{\text{ach}}(M, L) = \frac{1}{ML\sqrt{\Lambda(L)}} = \frac{e^{\gamma L/2}}{LM}. \tag{10}$$

This explicitly shows the noise penalty: when $\gamma = 0$, $\Delta\phi_{\text{ach}}(M, L) = 1/(ML) = 1/N$ and (for fixed M) $\Delta\phi$ continues improving as $L \to \infty$; when $\gamma > 0$, the exponential factor terminates Heisenberg scaling by enforcing a finite optimal L.

Figure 1 illustrates this analytic prediction using $\gamma = 0.02$, showing the initial Heisenberg-like improvement, the minimum at $L^* \approx 2/\gamma$, and the eventual rise due to decoherence.

Key interpretation: Noise does not merely "bend" the curve; it imposes a *finite usable coherent depth* L. Therefore the Heisenberg improvement associated with increasing coherent phase accumulation *terminates* at the knee. Once L is bounded, any remaining scaling benefit must come from allocating resources into M (or, more generally, into coherence-matched partitions), which is the architectural motivation behind resource-partitioning paradigms such as DQRPA.

2.1 Interpretation of the saturation knee

The minimum observed in Fig. 1 is not an empirical crossover between scaling laws, but a direct consequence of constrained optimization in the (M, L) resource space. For a fixed total resource $N = ML$, the achievable precision $\Delta\phi(M, L) = e^{\gamma L/2}/(ML)$ defines a two-dimensional landscape in which increasing the coherent interrogation depth L simultaneously enhances phase accumulation and accelerates coherence loss. The saturation knee corresponds to a constrained saddle point of this landscape: along the constraint curve $ML = N$, $\Delta\phi$ attains a minimum at $L^* = 2/\gamma$, while orthogonal variations in (M, L) do not produce a true global minimum. Physically, this saddle point marks the noise-induced termination of Heisenberg scaling, beyond which additional coherent depth yields diminishing— and ultimately negative—returns despite increasing total resources. The knee therefore reflects an optimization boundary imposed by finite coherence, not a plotting artifact or a failure of quantum estimation theory.

2.2 Relationship to prior "knee" or transition plots

In contrast to prior treatments where the knee emerges as an empirical crossover or as a consequence of channel-specific bounds, here the knee follows directly from constrained optimization of the sensing architecture itself. Knee-shaped precision curves have appeared in noisy metrology studies, typically as manifestations of channel-specific information loss, finite optimal sensing time, or optimized interrogation protocols in a particular physical model. It is therefore reasonable to call the minimum in Fig. 1 a "knee" (as is common in the literature). The distinction here is that the knee is explained as the outcome of a *resource-partitioned model* and an *explicit optimization*: under a fixed budget $N = ML$, increasing coherent depth L provides an ideal L^2 amplification, but noise adds an exponential penalty $e^{-\gamma L}$ that enforces a finite optimum. In this view, the knee is a **noise-induced termination of Heisenberg scaling**, not a plotting artifact and not merely an empirical crossover between $1/N$ and $1/\sqrt{N}$ trends. It is an actionable design marker indicating the maximum coherent depth a platform can use before coherence loss dominates.

3 Parameter Encoding and the QFI Origin of N- and L-Scaling

This paper treats all sensing tasks as estimation of a phase-like parameter ϕ encoded into a probe state by a unitary transformation

$$U_\phi = e^{-i\phi G}, \tag{11}$$

where G is the generator determined by the platform's probe–signal coupling (e.g., a number operator in interferometry, a collective spin component in magnetometry, or an effective Hamiltonian in time or frequency estimation).

This section establishes the quantum Fisher information (QFI) origin of both the familiar N-scaling (parallel resources) and the L-scaling (coherent sequential depth), thereby setting the stage for understanding why their combination ultimately fails under realistic noise.

3.1 Pure-state quantum Fisher information

For a pure state family $|\psi_\phi\rangle = U_\phi |\psi\rangle$, the quantum Fisher information is given by

$$F_Q(\phi) = 4\left(\langle\partial_\phi\psi_\phi|\partial_\phi\psi_\phi\rangle - |\langle\psi_\phi|\partial_\phi\psi_\phi\rangle|^2\right).$$

Using $\partial_\phi |\psi_\phi\rangle = -iG|\psi_\phi\rangle$, one obtains the standard pure-state identity

$$F_Q = 4\left(\langle G^2\rangle - \langle G\rangle^2\right) = 4\,\mathrm{Var}(G), \tag{12}$$

which is independent of ϕ for this encoding model, since unitary evolution preserves generator moments.

3.2 SQL scaling from additivity

For N probes with collective generator $G = \sum_{k=1}^{N} g^{(k)}$, an uncorrelated product state exhibits vanishing cross-covariances between distinct probes. Consequently,

$$\mathrm{Var}(G) = \sum_{k=1}^{N} \mathrm{Var}(g^{(k)}) \propto N,$$

and by Eq. (12), the QFI scales linearly with N: $F_Q \propto N$. This yields the standard quantum limit $\Delta\phi_{\mathrm{SQL}} \propto 1/\sqrt{N}$.

3.3 Heisenberg scaling from maximal covariance

Entanglement enables the creation of $O(N^2)$ covariances among the individual generators $g^{(k)}$, producing $\mathrm{Var}(G) \propto N^2$. A canonical example is an N-qubit GHZ state under the collective generator $G = J_z = \frac{1}{2}\sum_k \sigma_z^{(k)}$, where the state is a superposition of J_z eigenvalues $\pm N/2$. In this case, $\mathrm{Var}(J_z) = (N/2)^2$, yielding $F_Q \propto N^2$ and the ideal Heisenberg scaling $\Delta\phi_{\mathrm{HL}} \propto 1/N$.

3.4 Sequential depth: L coherent applications

If the phase imprinting occurs coherently over L identical steps, or over a duration proportional to L, the effective unitary becomes

$$U_\phi^{(L)} = \left(e^{-i\phi G}\right)^L = e^{-i\phi(LG)}. \tag{13}$$

The effective generator is therefore $G_L = LG$, and the corresponding QFI is

$$F_Q^{(L)} = 4\,\mathrm{Var}(G_L) = 4\,\mathrm{Var}(LG) = L^2 F_Q. \tag{14}$$

This quadratic enhancement in L is the ideal origin of sequential Heisenberg scaling. In realistic platforms, however, coherent accumulation is accompanied by a coherence retention factor $\Lambda(L)$ that decays with increasing L. The competition between the ideal L^2 growth and coherence loss under $\Lambda(L)$ produces a universal saturation knee, which is shown in later sections to arise from constrained optimization in the (M, L) resource plane rather than from channel-specific pathologies.

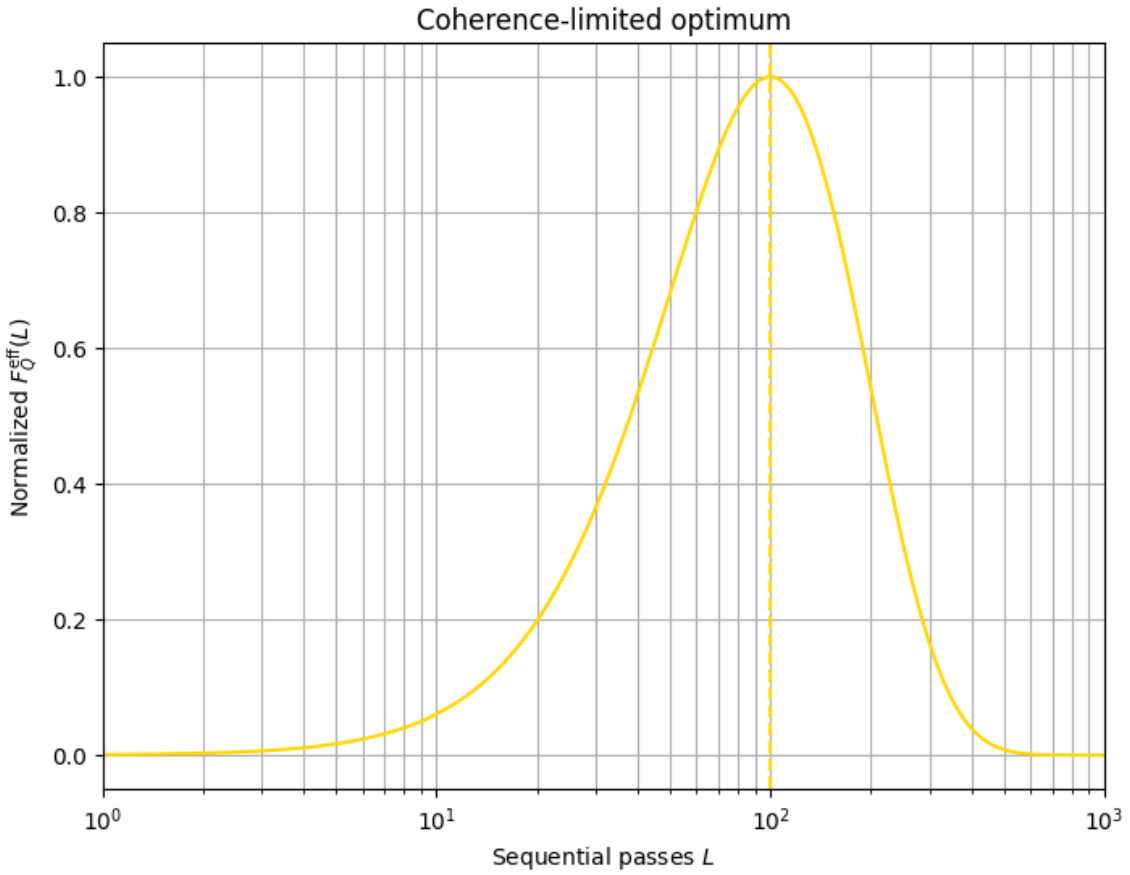

Figure 2: **Coherence-limited optimum.** Normalized effective QFI $F_Q^{\text{eff}}(L) \propto (NL)^2 e^{-\gamma L}$ as a function of sequential depth L (log-scale), showing the universal coherence-limited optimum at $L^\star \approx 2/\gamma$. Here $\gamma = 0.02$ is used consistently across all figures.

4 Why All Platforms Saturate

The saturation behavior arises from the competition between information gain ($\propto L^2$ in the ideal sequential degree) and coherence loss ($\propto e^{-\gamma L}$). For sequential interrogation at fixed M,

$$F_Q^{\text{eff}}(L, M) = (LM)^2 e^{-\gamma L}. \tag{15}$$

Equivalently, up to a constant (M^2) factor, the L-dependence is

$$F_Q^{\text{eff}}(L, M) \propto L^2 e^{-\gamma L}, \tag{16}$$

which peaks at

$$\frac{d}{dL}\left(L^2 e^{-\gamma L}\right) = 0 \quad \Rightarrow \quad L^* = \frac{2}{\gamma}. \tag{17}$$

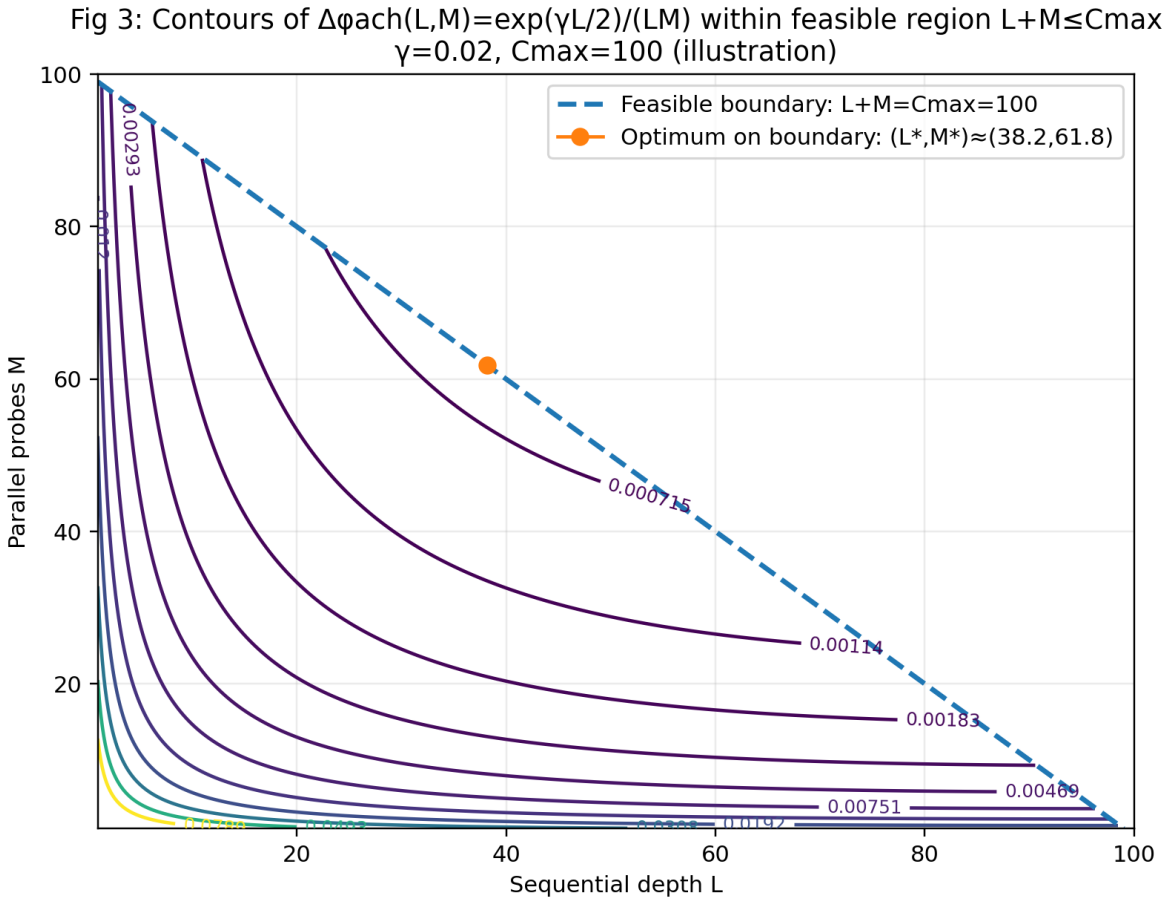

Figure 3: **Contours of achievable precision within a feasible design region.** Contours of $\Delta\phi_{\text{ach}}(L, M) = e^{\gamma L/2}/(LM)$ over the (L, M) design space, masked to the feasible region $L + M \leq C_{\max}$ (illustration: $C_{\max} = 100$). Lower contours are better. The optimum for this illustrative cost model occurs on the feasibility boundary $L + M = C_{\max}$, providing an architecture-level interpretation of the knee as a transition between coherence-beneficial and coherence-dominated regimes.

4.1 Interplay between L and M under a fixed budget $N = ML$: Contour maps

The preceding discussion often fixes M and optimizes over coherent depth L, which cleanly reveals the noise-limited optimum L^*. However, practical design typically faces a constrained budget: $N = ML$ is fixed, so increasing L forces a reduction in M (and vice versa). The correct question is therefore: *for fixed total budget N, how should resources be split between parallel depth M and coherent depth L under noise?*

To visualize this trade-space, we plot contours of the effective QFI $F_Q^{\text{eff}}(M, L) = (LM)^2 e^{-\gamma L}$ and the corresponding achievable precision $\Delta\phi_{\text{ach}}(M, L) = e^{\gamma L/2}/(ML)$ over the (L, M) plane, under the physical constraint $N = ML = 100$ and $L \geq 1$.

Design insight from the contour geometry. The contour maps make the mechanism of "saturation" concrete: for fixed N, there is no monotone path to improved precision by increasing L. Instead, noise creates a bounded "useful-L" region; pushing L past this region reduces F_Q^{eff} and worsens $\Delta\phi_{\text{ach}}$ even though N is unchanged. In other words, the knee is the one-dimensional projection of a two-

dimensional constrained optimum over (M, L).

4.2 Architectural optimization in the (M, L) plane

While Figures 1–2 illustrate Heisenberg-saturation as a function of the total resource N, deeper insight is obtained by examining the joint dependence of performance on parallelization M and coherent depth L under the constraint $f(L, M) \leq C_{\max}$.[1]

Figure 3 shows contour lines of achievable phase precision $\Delta\phi_{\text{ach}}(L, M) = \exp(\gamma L/2)/(LM)$ over the (M, L) design space within an illustrative engineering feasibility region $L + M \leq C_{\max}$ (shown as a triangular domain), with boundary $L + M = C_{\max}$ overlaid. Lower contours are better. For this model, the global optimum over the feasible set occurs on the boundary; the marked point indicates the minimizing (L^*, M^*) along $L + M = C_{\max}$. This 2-D geometry provides an architecture-level explanation of the 1-D "knee" in Fig. 2: beyond the coherence-limited neighborhood, increasing L exponentially penalizes precision even though more sequential depth is being invested.

4.3 Implications for Sensor Design: A Resource-Partitioning Paradigm

The decomposition $N = ML$ converts Heisenberg-saturation from an abstract scaling limitation into a concrete architectural design principle. The central lesson is that quantum enhancement is governed not by the total resource alone, but by how that resource is partitioned between parallelization and coherent depth under a finite coherence budget.

As shown analytically and confirmed by the (M, L) contour plots, increasing coherent interrogation depth beyond L^* does not merely yield diminishing returns—it actively degrades performance by converting phase accumulation into decoherence. In this regime, additional resources fail to increase quantum Fisher information, marking a noise-induced termination of Heisenberg scaling.

This observation motivates a paradigm shift: optimal quantum sensors must dynamically allocate resources between M and L rather than pursue maximal entanglement depth or maximal interrogation time in isolation. The Heisenberg-saturation knee thus defines a boundary between coherence-limited and resource-limited operation.

We refer to this principle as *dynamic quantum resource partitioning*. It provides a unifying design framework applicable across sensing platforms, independent of specific noise models or measurement strategies. Within this framework, sensor optimization becomes an architectural problem—selecting the resource allocation that maximizes retained quantum advantage—rather than an exercise in saturating abstract bounds.

5 Summary and Conclusions

The breakdown of Heisenberg scaling in realistic quantum sensors is often treated as an unavoidable consequence of noise or as a channel-specific limitation. Here we have shown that this breakdown admits a constructive, architecture-level interpretation. By decomposing the total sensing resource as $N = ML$ and explicitly accounting for coherence loss with increasing sequential depth, the appearance

[1] In Fig. 3 we use an illustrative engineering feasibility constraint $L + M \leq C_{\max}$ to visualize the architecture trade-space. The Heisenberg scaling statements in Sections 3–4 are defined with respect to conserved physical resources (e.g., total interrogation quanta).

of a "knee" in precision scaling emerges as the outcome of a constrained optimization problem rather than a plotting artifact, estimator inefficiency, or channel-specific anomaly.

The Heisenberg-saturation knee marks a noise-induced termination of coherent phase accumulation: beyond an optimal interrogation depth L^*, additional resources no longer increase quantum Fisher information and instead accelerate decoherence. This behavior is directly visible in both analytic expressions and in the (M, L) contour plots, which reveal how quantum advantage is redistributed across sensing architectures under a fixed total resource budget.

Crucially, the termination of Heisenberg scaling is traced to a bound on *coherent depth* rather than on particle number itself. In the absence of noise ($\gamma = 0$), sequential accumulation can extend indefinitely and $\Delta\phi \propto 1/N$ is recovered. With any finite decoherence rate, however, coherent depth becomes the limiting factor, and the optimal operating point is determined by the platform's finite coherence budget.

These results motivate dynamic quantum resource partitioning as a foundational design principle for scalable quantum sensing. Rather than attempting to maximize entanglement or interrogation time in isolation, optimal sensors must allocate resources between parallelization and coherent depth to maximize *retained* quantum advantage. This perspective applies uniformly across optical, atomic, solid-state, and superconducting platforms, providing a common language for understanding, comparing, and designing quantum sensors beyond idealized limits. In the companion work, this principle is elevated into a concrete architectural framework for block-partitioned sensing systems.

Interpretation of the saturation knee. The saturation knee observed in Fig. 1 is not merely a crossover between Heisenberg and SQL scaling, nor a channel-specific artifact. Rather, it reflects a constrained optimum in the (M, L) resource plane under the fixed-budget condition $N = ML$. Noise imposes an exponential penalty on coherent depth L, terminating Heisenberg scaling even as total resources increase. Beyond this boundary, additional resources cannot be converted into improved phase precision through coherent accumulation. The knee therefore represents a noise-induced architectural optimization boundary, rather than a failure of quantum metrology itself.

References

[1] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Physical Review Letters*, 96(1):010401, 2006.

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

[3] Rafał Demkowicz-Dobrzański, Jan Kołodyński, and Madalin Guta. The elusive Heisenberg limit in quantum-enhanced metrology. *Nature Communications*, 3:1063, 2012.

[4] Rafał Demkowicz-Dobrzański, Marcin Jarzyna, and Jan Kołodyński. Quantum limits in optical interferometry. In Emil Wolf, editor, *Progress in Optics*, volume 60, pages 345–435. Elsevier, 2015.