AIR Journal of Mathematics & Computational Sciences, Vol. 2026
Article Unique ID: AIRMCS2026322
DOI: 10.65737/AIRMCS2026322
© Author(s). This work is licensed under a Creative Commons License.

QUANTUM-ENHANCED IN-CONTEXT LEARNING FOR GEOPOTENTIAL FIELD ESTIMATION: A THEORETICAL FRAMEWORK

Mosab Hawarey
Director, Geospatial Research (*director@geospatial.ch*)

Received: 2026 February 21 | Revised: 2026 February 24 | Approved: 2026 February 25 | Published: 2026 February 26

Abstract

We establish a theoretical framework for in-context learning (ICL) of Earth's gravitational potential field using transformer architectures, with particular emphasis on the sample complexity advantages afforded by quantum gravimetry. The geopotential, expressed as a truncated spherical harmonic expansion of maximum degree N with $K = (N+1)^2$ coefficients, defines a function class for which we characterize ICL learnability. Building on the ICL characterization framework of Hawarey (2026) (i.e. the foundational paper), we prove that the geopotential function class is ICL-Easy: it admits an additive sufficient statistic computable by a single attention layer, enabling transformers to match the sample complexity of optimal statistical estimators. Our main contributions are threefold. First, we prove that the minimal sufficient statistic for geopotential ICL is the pair (G_n, c_n) consisting of the Gram matrix and cross-correlation vector, with dimension $K^2 + K$, and demonstrate its attention-computability. Second, we derive tight sample complexity bounds showing that ICL achieves $n_{ICL} = \Theta(K\sigma^2/\varepsilon \cdot \log(K/\delta))$ for noise variance σ^2, target accuracy ε, and failure probability δ, matching empirical risk minimization. Third, we quantify the quantum advantage: for quantum gravimeters operating at the Heisenberg limit with N_{atoms} entangled atoms, the sample complexity reduces by a factor of $\rho_Q = N_{atoms}^2$, yielding up to 10^{12}-fold improvement for realistic sensor parameters. We also establish fundamental boundaries: while forward geopotential prediction is ICL-Easy, inverse problems such as density inversion and source localization are ICL-Hard. Source localization with J discrete masses requires identifying combinatorial structure from exponentially many candidates, which we prove satisfies the ICL-Hard structural condition H1 of Hawarey (2026) when $J = \omega(1)$, establishing that no polynomial-size transformer can solve it efficiently. This dichotomy—ICL-Easy forward problems versus ICL-Hard inverse problems—reflects the fundamental mathematical structure of potential theory and provides guidance for applying transformer-based methods in geodesy.

Keywords: in-context learning; transformers; geopotential; quantum gravimetry; sample complexity; spherical harmonics; sufficient statistics; ICL-Easy; ICL-Hard

Citation: Hawarey, M. (2026). Quantum-Enhanced In-Context Learning for Geopotential Field Estimation: A Theoretical Framework. AIR Journal of Mathematics & Computational Sciences, Vol. 2026, AIRMCS2026322, DOI: 10.65737/AIRMCS2026322.

1. Introduction

1.1 Background and Motivation

Earth's gravitational potential field encodes fundamental information about planetary mass distribution, with applications spanning geodesy, geophysics, satellite navigation, and climate science [Torge & Müller, 2012]. The geopotential is conventionally represented through spherical harmonic expansions, where a bandlimited field of maximum degree N requires $K = (N+1)^2$ coefficients—exceeding 130,000 parameters for current global models such as EGM2008 [Pavlis et al., 2012]. Determining these coefficients from gravimetric observations constitutes a fundamental inverse problem in geodesy, with accuracy directly limited by measurement noise.

Two technological revolutions motivate this work. *First*, transformer architectures have demonstrated remarkable capabilities for in-context learning (ICL)—the ability to learn new tasks from examples provided in the input context without explicit parameter updates [Brown et al., 2020; Garg et al., 2022]. Mechanistic studies have further elucidated how transformers implement learning algorithms in their forward pass: Akyürek et al. (2023) showed that trained transformers can implement gradient descent and closed-form ridge regression, while von Oswald et al. (2023) demonstrated that transformer layers can be interpreted as mesa-optimization steps, and Bai et al. (2024) established statistical foundations for ICL by analyzing transformers as implicit Bayesian estimators. Recent theoretical work by Hawarey (2026) provides a rigorous characterization of when transformers can efficiently perform ICL, establishing necessary and sufficient conditions through the concept of *attention-computable sufficient statistics*. *Second*, quantum gravimetry based on cold atom interferometry offers dramatic noise reduction over classical sensors [Peters et al., 1999; Kasevich & Chu, 1991]. At the Heisenberg limit, noise variance scales as $\sigma^2 \propto 1/N_{\text{atoms}}^2$, providing up to 10^{12} improvement for sensors with 10^6 entangled atoms [Giovannetti et al., 2011].

The research gap addressed by this work lies at the intersection of these advances: no theoretical framework exists for analyzing transformer-based in-context learning of geopotential fields, and the sample complexity implications of quantum measurement noise for ICL remain unexplored. This gap is consequential because: (1) ICL offers a flexible paradigm for geopotential estimation that adapts to varying measurement configurations without retraining; (2) understanding when ICL succeeds or fails provides principled guidance for applying transformers in geodesy; and (3) quantum sensors fundamentally alter the statistical landscape, potentially enabling transformer-based methods that were previously sample-prohibitive.

1.2 Contributions

This paper makes the following contributions, each rigorously grounded in the theoretical framework of Hawarey (2026):

Contribution 1: ICL-Easy Classification. We prove that the bandlimited geopotential function class $\mathscr{F}_N$ is *ICL-Easy* in the sense of Hawarey (2026, Definition 2.12). Specifically, we show that the minimal sufficient statistic ($\mathbf{G}_n$, $\mathbf{c}_n$)—the Gram matrix and cross-correlation vector—is *attention-computable*: it decomposes as an additive function over context examples and can be computed by a single attention layer with feedforward networks of polynomial size (Theorems 4.1–4.4). This establishes that transformers can perform geopotential ICL with the same statistical efficiency as optimal estimators.

Contribution 2: Quantum Sample Complexity Theorem. We derive tight bounds on the sample complexity for both classical and quantum measurements. For classical sensors with noise variance $\sigma_{(C)}^2$, the ICL sample complexity is $n_{(ICL)}^{(C)} = \Theta(K\sigma_{(C)}^2/\varepsilon \cdot \log(K/\delta))$ (Theorem 4.3). For quantum sensors at the Heisenberg limit, we prove $n_{(ICL)}^{(Q)} = \Theta(K\sigma_{(Q)}^2/\varepsilon)$ with $\sigma_{(Q)}^2 = \sigma_0^2/N_{(atoms)}^2$ (Theorem 4.5). The quantum advantage ratio $\rho_{(Q)} = N_{(atoms)}^2$ reduces required samples by up to 10^{12} for realistic sensor parameters (Theorem 4.6).

Contribution 3: ICL-Hardness Boundaries. We establish fundamental limits: while *forward* geopotential prediction is ICL-Easy, *inverse* problems are ICL-Hard. We prove that source localization with $J = \omega(1)$ discrete masses is ICL-Hard (Theorem 5.2), via the combinatorial structure of its sufficient statistic and application of the structural hardness condition H1 of Hawarey (2026, Theorem 5). This dichotomy—additive sufficient statistics for forward problems versus combinatorial statistics for inverse problems—provides principled guidance for when transformer-based geodesy methods can succeed.

1.3 Paper Organization

The paper is organized as follows. **Section 2** establishes preliminaries: the spherical harmonic representation of geopotential (§2.1), classical and quantum measurement models (§2.2), and the ICL framework from Hawarey (2026) (§2.3). **Section 3** proves the ICL-Easy classification: sufficient statistic identification (§3.1), attention-computability (§3.2), and transformer construction (§3.3). **Section 4** derives sample complexity bounds for classical (§4.1) and quantum (§4.2) measurements, culminating in the quantum advantage theorem (§4.3). **Section 5** establishes ICL-Hardness boundaries for inverse problems (§5.1–5.3). **Section 6** discusses practical implications, limitations, and future directions. **Section 7** concludes.

1.4 Notation

We establish notation used throughout. Scalars are denoted by lowercase letters (n, K, σ^2), vectors by bold lowercase ($\mathbf{r}$, $\mathbf{C}$, $\mathbf{c}_n$), and matrices by bold uppercase ($\mathbf{G}_n$, $\mathbf{\Phi}$). Function classes are denoted by calligraphic letters ($\mathscr{F}_N$). We write $\Theta(\cdot)$, $O(\cdot)$, $\Omega(\cdot)$ for asymptotic bounds, and *poly*$(\cdot)$ for polynomial functions. The notation $[n]$ denotes $\{1, 2, \ldots, n\}$. All symbols are defined precisely when first introduced.

2. Preliminaries

This section establishes the mathematical foundations for our analysis. We define the geopotential function class (§2.1), formalize classical and quantum measurement models (§2.2), and summarize the ICL framework from Hawarey (2026) that underpins our theoretical results (§2.3). All symbols are defined precisely when introduced.

2.1 Geopotential Representation

2.1.1 Coordinate System

We represent spatial positions in spherical coordinates centered at Earth's center of mass.

Definition 2.1 (Spherical Coordinates). A point in three-dimensional space is represented by $\mathbf{r} = (r, \theta, \lambda)$ where: $r \in \mathbb{R}^+$ is the *radial distance* from Earth's center (meters); $\theta \in [0, \pi]$ is the *colatitude*, measured from the north pole (radians); $\lambda \in [0, 2\pi)$ is the *longitude*, measured eastward from a reference meridian (radians).

2.1.2 Spherical Harmonic Basis

The gravitational potential exterior to Earth satisfies Laplace's equation, admitting solutions expressed in spherical harmonics [Heiskanen & Moritz, 1967].

Definition 2.2 (Associated Legendre Functions). For degree $n \geq 0$ and order m with $|m| \leq n$, the *associated Legendre function* $P_n^m : [-1, 1] \to \mathbb{R}$ is defined by the Rodrigues formula and satisfies standard orthogonality relations on the unit sphere.

Definition 2.3 (Surface Spherical Harmonics). For degree $n \in \mathbb{N}_0$ and order $m \in \mathbb{Z}$ with $|m| \leq n$, the *surface spherical harmonic* $Y_{nm} : [0, \pi] \times [0, 2\pi) \to \mathbb{R}$ is defined by $Y_{nm}(\theta, \lambda) = P_n^{|m|}(\cos\theta) \cdot \cos(m\lambda)$ for $m \geq 0$ and $Y_{nm}(\theta, \lambda) = P_n^{|m|}(\cos\theta) \cdot \sin(|m|\lambda)$ for $m < 0$. These form a complete orthogonal basis for square-integrable functions on the unit sphere.

2.1.3 The Geopotential Function Class

Definition 2.4 (Coefficient Index Set and Dimension). For maximum degree $N \in \mathbb{N}$, the *coefficient index set* is $\mathcal{J}_N = \{(n, m) : n \in \{0, 1, \ldots, N\}, m \in \{-n, \ldots, n\}\}$. The *number of coefficients* is $K = |\mathcal{J}_N| = (N + 1)^2$. We impose a lexicographic ordering on $\mathcal{J}_N$, defining a bijection $\iota : \mathcal{J}_N \to \{1, \ldots, K\}$, allowing coefficients to be represented as a vector $\mathbf{C} \in \mathbb{R}^K$.

Definition 2.5 (Bandlimited Geopotential Function Class). The *bandlimited geopotential function class* of maximum degree N is:

$$\mathcal{F}_N = \{ W : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R} \mid W(\mathbf{r}) = (GM/R)\, \Sigma_{n,m} (R/r)^{n+1}\, C_{nm}\, Y_{nm}(\theta, \lambda),\ \|\mathbf{C}\|_2 \leq B_C \}$$

where $GM = 3.986 \times 10^{14}$ m³s⁻² is the geocentric gravitational constant, $R = 6.371 \times 10^6$ m is the mean Earth radius, $C_{nm} \in \mathbb{R}$ are the spherical harmonic coefficients, and $B_C > 0$ bounds the coefficient norm.

Definition 2.6 (Basis Function Vector). For position $\mathbf{r} = (r, \theta, \lambda)$, the *spherical harmonic basis vector* $\Phi_N(\mathbf{r}) \in \mathbb{R}^K$ has components $[\Phi_N(\mathbf{r})]_k = (GM/R)(R/r)^{n+1}\, Y_{nm}(\theta, \lambda)$ where $k = \iota(n, m)$. This enables the compact representation $W(\mathbf{r}) = \mathbf{C}^T\Phi_N(\mathbf{r})$.

Remark 2.1. The function class $\mathscr{F}_N$ is a finite-dimensional Hilbert space of dimension $K = (N + 1)^2$. This linear structure, with $W(\mathbf{r}) = \mathbf{C}^T\Phi_N(\mathbf{r})$, corresponds exactly to the linear function class in Hawarey (2026, Definition 2.24), enabling direct application of the ICL characterization framework.

2.2 Measurement Models

2.2.1 Classical Gravimetric Measurements

Definition 2.7 (Measurement Domain). The *measurement position domain* is $\mathscr{R} = \{(r, \theta, \lambda) : r \in [R_{min}, R_{max}], \theta \in [0, \pi], \lambda \in [0, 2\pi)\}$, where $R_{min} \approx R$ (Earth's surface) and $R_{max} \approx R + 500$ km (satellite altitude).

Definition 2.8 (Classical Observation Model). A *classical gravimetric measurement* at position $\mathbf{r}_i \in \mathscr{R}$ yields observation:

$$y_i^{(C)} = W(\mathbf{r}_i) + \varepsilon_i^{(C)} = \mathbf{C}^T\Phi_N(\mathbf{r}_i) + \varepsilon_i^{(C)}, \text{ where } \varepsilon_i^{(C)} \sim \mathcal{N}(0, \sigma_C^2)$$

where $\sigma_C^2 > 0$ is the *classical noise variance* arising from thermal fluctuations, mechanical vibrations, and instrumental drift in classical gravimeters (spring, superconducting). Measurements are conditionally independent given W (Assumption B1).

2.2.2 Quantum Gravimetric Measurements

Quantum gravimeters based on cold atom interferometry [Kasevich & Chu, 1991; Peters et al., 1999] achieve fundamentally different noise scaling governed by quantum mechanics.

Definition 2.9 (Quantum Noise Regimes). For a quantum gravimeter with N_{atoms} atoms and reference noise variance σ_0^2 (single-atom noise floor):

(a) Shot Noise Limit (SNL): For uncorrelated atoms, $\sigma_{SNL}^2 = \sigma_0^2/N_{atoms}$. The variance scales as $O(1/N_{atoms})$, equivalent to averaging N_{atoms} independent classical measurements.

(b) Heisenberg Limit (HL): For maximally entangled states (spin-squeezed, GHZ), $\sigma_{HL}^2 = \sigma_0^2/N_{atoms}^2$. The variance scales as $O(1/N_{atoms}^2)$, the fundamental limit from the Heisenberg uncertainty principle [Giovannetti et al., 2011].

Definition 2.10 (Quantum Observation Model). A *quantum gravimetric measurement* at position $\mathbf{r}_i$ yields observation $y_i^{(Q)} = W(\mathbf{r}_i) + \varepsilon_i^{(Q)}$ where $\varepsilon_i^{(Q)} \sim \mathcal{N}(0, \sigma_Q^2)$ with $\sigma_Q^2 = \sigma_0^2/N_{atoms}^{\alpha}$ and $\alpha \in \{1, 2\}$ for SNL and HL respectively.

Definition 2.11 (Quantum Advantage Ratio). The *quantum advantage ratio* is $\rho_Q = \sigma_C^2/\sigma_Q^2$, quantifying noise reduction. For Heisenberg-limited sensors: $\rho_Q^{(HL)} = \sigma_C^2 \cdot N_{atoms}^2 / \sigma_0^2$. With $N_{atoms} = 10^6$ and matched reference noise, $\rho_Q^{(HL)} = 10^{12}$.

2.3 In-Context Learning Framework

We summarize the ICL characterization framework of Hawarey (2026), which provides the theoretical foundation for our results.

2.3.1 The ICL Task

Definition 2.12 (Context Set). A *context set* of size n is $\mathcal{C}_n = \{(\mathbf{r}_1, y_1), \ldots, (\mathbf{r}_n, y_n)\}$ where $\mathbf{r}_i \in \mathscr{R}$ are measurement positions and $y_i = W(\mathbf{r}_i) + \varepsilon_i$ are noisy observations.

Definition 2.13 (ICL Task). The *in-context learning task for geopotential estimation* is: given context $\mathcal{C}_n$ and query position $\mathbf{r}_q \in \mathscr{R}$, output prediction $\hat{y}_q \in \mathbb{R}$ minimizing $\mathbb{E}[(W(\mathbf{r}_q) - \hat{y}_q)^2]$.

2.3.2 Sufficient Statistics and Attention-Computability

Definition 2.14 (Sufficient Statistic). A statistic $T : (\mathscr{R} \times \mathbb{R})^n \to \mathbb{Z}$ is *sufficient* for $\mathbf{C}$ if the conditional distribution of $\mathcal{C}_n$ given $\mathbf{C}$ factors through T. It is *minimal sufficient* if it has the smallest dimension among all sufficient statistics [Lehmann & Scheffé, 1950].

Definition 2.15 (Attention-Computable Statistic, Hawarey 2026, Def. 2.16). A statistic T is *attention-computable* if it can be expressed as $T(\mathcal{C}_n) = \Sigma_{i=1}^n \varphi(\mathbf{r}_i, y_i)$ where $\varphi : \mathscr{R} \times \mathbb{R} \to \mathbb{R}^k$ is a *feature map* computable by a feedforward network of polynomial size. The additive structure enables computation via a single attention layer [Vaswani et al., 2017].

Definition 2.16 (AC-SSC, Hawarey 2026, Def. 2.17). The *Attention-Computable Sufficient Statistic Complexity* is AC-SSC($\mathscr{F}$, n, ε, s) = $\min\{\dim(T) : T$ is ε-sufficient, attention-computable with transformer size $\leq s\}$ (used in Theorem 3.3 and Remark 3.2).

2.3.3 The ICL-Easy / ICL-Hard Dichotomy

Definition 2.17 (ICL-Easy, Hawarey 2026, Def. 2.12). A function class $\mathscr{F}$ is *ICL-Easy* if there exists a transformer $\mathcal{T}$ of polynomial size $s = \mathrm{poly}(n, K)$ that achieves ε-accuracy for all $f \in \mathscr{F}$ with sample complexity $n_{\mathrm{ICL}} = \Theta(n_{\mathrm{ERM}})$, matching empirical risk minimization.

Definition 2.18 (ICL-Hard, Hawarey 2026, Def. 2.12). A function class $\mathscr{F}$ is *ICL-Hard* if any transformer achieving success probability $> 1/2 + n^{-O(1)}$ requires either super-polynomial size $s = n^{\omega(1)}$ or super-constant depth $L = \omega(1)$.

Theorem 2.1 (Master Theorem, Hawarey 2026, Thm. 3). For a function class $\mathscr{F}$, the following are equivalent: (i) $\mathscr{F}$ is ICL-Easy; (ii) AC-SSC($\mathscr{F}$) $\leq \mathrm{poly}(n, K)$; (iii) $\mathscr{F}$ admits an attention-computable sufficient statistic and FFN-computable predictor.

Theorem 2.2 (Dichotomy, Hawarey 2026, Thm. 6). For any natural function class $\mathscr{F}$, exactly one holds: (Type A) $\mathscr{F}$ has an additive sufficient statistic and is ICL-Easy with $n_{\mathrm{ICL}} = \Theta(n_{\mathrm{ERM}})$; or (Type C) $\mathscr{F}$ has a combinatorial sufficient statistic and is ICL-Hard. No natural function classes exhibit intermediate complexity.

This framework provides the theoretical machinery for analyzing geopotential ICL. In Section 3, we prove that $\mathscr{F}_N$ is ICL-Easy by constructing its attention-computable sufficient statistic.

2.4 Standing Assumptions

B1: Conditional independence of measurements
B2: Bounded basis functions
B3: Well-specified model
B4: Bounded condition number

3. ICL-Easy Classification of Geopotential Functions

This section establishes our first main result: the geopotential function class $\mathscr{F}_N$ is ICL-Easy. We identify the minimal sufficient statistic (§3.1), prove its attention-computability (§3.2), and construct the transformer architecture (§3.3). All proofs follow the methodology of Hawarey (2026) with adaptations for the spherical harmonic structure.

3.1 Sufficient Statistic Identification

We first establish that the geopotential observation model belongs to the exponential family, then identify the minimal sufficient statistic.

Definition 3.1 (Gram Matrix and Cross-Correlation). For measurements at positions $\{\mathbf{r}_i\}_{i=1}^n$ with observations $\{y_i\}_{i=1}^n$, define the *Gram matrix* $\mathbf{G}_n \in \mathbb{R}^{K\times K}$ and *cross-correlation vector* $\mathbf{c}_n \in \mathbb{R}^K$ by:

$$\mathbf{G}_n = \Sigma_{i=1}^n \Phi_N(\mathbf{r}_i)\Phi_N(\mathbf{r}_i)^T, \qquad \mathbf{c}_n = \Sigma_{i=1}^n y_i \Phi_N(\mathbf{r}_i)$$

where $\Phi_N(\mathbf{r}) \in \mathbb{R}^K$ is the spherical harmonic basis vector (Definition 2.6). The matrix $\mathbf{G}_n$ is symmetric positive semidefinite with rank $\min(n, K)$.

Proposition 3.1 (Exponential Family Structure). Under the observation model (Definition 2.8), the joint log-likelihood takes the form $\ell(C) \propto -(1/2\sigma^2)C^TG_nC + (1/\sigma^2)C^Tc_n$, which depends on the data only through the natural sufficient statistic $T(\mathcal{C}_n) = (\mathrm{vec}(G_n), c_n)$. The natural parameter is $\eta = C/\sigma^2 \in \mathbb{R}^K$.

Proof. The joint likelihood is $p(\mathbf{y}|\mathbf{C}) = (2\pi\sigma^2)^{-n/2} \exp(-(1/2\sigma^2)\Sigma_i(y_i - \mathbf{C}^T\Phi_N(\mathbf{r}_i))^2)$. Expanding the quadratic: $\Sigma_i(y_i - \mathbf{C}^T\Phi_i)^2 = \|\mathbf{y}\|^2 - 2\mathbf{C}^T\mathbf{c}_n + \mathbf{C}^T\mathbf{G}_n\mathbf{C}$. This yields exponential family form with the stated natural parameter and sufficient statistic. ■

Theorem 3.1 (Minimal Sufficient Statistic). The pair $T(\mathcal{C}_n) = (\mathbf{G}_n, \mathbf{c}_n)$ is the unique minimal sufficient statistic for $\mathscr{F}_N$ with Gaussian observations. Its dimension is:
$\dim(T) = K(K+1)/2 + K = (N+1)^2((N+1)^2 + 1)/2 + (N+1)^2$

Proof. We establish sufficiency, dimension lower bound, and uniqueness.

Sufficiency: By the Neyman-Fisher factorization theorem, T is sufficient if $p(\mathbf{y}|\mathbf{C}) = g(T(\mathcal{C}_n), \mathbf{C}) \cdot h(\mathbf{y})$. From Proposition 3.1, the likelihood depends on observations only through $(\mathbf{G}_n, \mathbf{c}_n)$.

Dimension lower bound: The optimal predictor is $\hat{y}_q = \mathbf{c}_n^T \mathbf{G}_n^{-1} \Phi_N(\mathbf{r}_q)$ (by the Gauss-Markov theorem [Aitken, 1936]). For prediction at arbitrary query positions, the full matrices $\mathbf{G}_n$ and $\mathbf{c}_n$ are required—any statistic of smaller dimension cannot support prediction at all $\mathbf{r}_q \in \mathscr{R}$. Since G_n is symmetric positive semidefinite, it has K(K+1)/2 unique entries. Together with the K-dimensional vector c_n, the total dimension is K(K+1)/2 + K.

Uniqueness: By exponential family theory [Lehmann & Scheffé, 1950], the minimal sufficient statistic is unique up to bijective transformation. Since $(\mathbf{G}_n, \mathbf{c}_n)$ achieves the minimum dimension for arbitrary prediction, it is the unique minimal sufficient statistic. ■

Corollary 3.1 (Sufficient Statistic Complexity). For $\mathscr{F}$_(N): SSC($\mathscr{F}$_(N), n, ε) = K(K+1)/2 + K = $O(N^4)$ for all $n \geq K$ and $\varepsilon > 0$.

3.2 Attention-Computability

We now prove that the sufficient statistic is attention-computable by constructing an explicit feature map with additive structure.

Definition 3.2 (Feature Map). Define the feature map $\varphi : \mathscr{R} \times \mathbb{R} \to \mathbb{R}^{K^2+K}$ by:

$$\varphi(\mathbf{r}, y) = [\ \mathrm{vec}(\Phi_N(\mathbf{r})\Phi_N(\mathbf{r})^T)\ ;\ y \cdot \Phi_N(\mathbf{r})\ ]$$

where $\mathrm{vec}(\cdot) : \mathbb{R}^{K \times K} \to \mathbb{R}^{K^2}$ vectorizes a matrix by stacking columns.

Remark 3.1 (Redundancy and Minimality). The feature map φ uses the full vec(·) operator, yielding dimension $K^2 + K$. This is a sufficient but not minimal representation: since $\Phi N(r)\Phi N(r)^T$ is symmetric, only its upper triangle is needed, reducing to K(K+1)/2 + K coordinates—matching the minimal sufficient statistic dimension of Theorem 3.1. The attention mechanism in Theorem 3.2 computes the full $K^2 + K$ dimensional statistic; the predictor g is invariant to the redundant symmetric entries (i.e., g depends on Gn only through its symmetric structure, so the redundant entries carry no additional information). Either formulation is valid: the full vec(·) simplifies implementation by avoiding symmetric masking, while the triangular form achieves minimality. We use the full form throughout for notational simplicity, while SSC and AC-SSC are stated in terms of the minimal dimension K(K+1)/2 + K.

Lemma 3.1 (Additivity). The sufficient statistic decomposes additively:
$T(\mathcal{C}_n) = \Sigma_{i=1}^n \varphi(\mathbf{r}_i, y_i)$

Proof. By direct computation: $\Sigma_i \varphi(\mathbf{r}_i, y_i) = [\Sigma_i \mathrm{vec}(\Phi_i\Phi_i^T); \Sigma_i y_i\Phi_i] = [\mathrm{vec}(\mathbf{G}_n); \mathbf{c}_n]$ by linearity of vec(·) and Definitions 3.1. ■

Lemma 3.2 (Spherical Harmonic FFN-Computability). Each component $Y_{nm}(\theta, \lambda)$ of the basis vector $\Phi_N(\mathbf{r})$ is computable by a feedforward network with depth $O(N)$, width $O(N^2)$, and size $O(N^3)$ parameters.

Proof. The spherical harmonic $Y_{nm}(\theta, \lambda) = P_n^{|m|}(\cos\theta) \cdot \cos(m\lambda)$ (or sin for $m < 0$) involves: (1) Associated Legendre functions $P_n^m(x)$ computed via three-term recurrence in $O(N^2)$ operations; (2) Trigonometric terms $\cos(m\lambda)$, $\sin(m\lambda)$ via Chebyshev recurrence in $O(m)$ operations; (3) Radial factor $(R/r)^{n+1}$ via $O(N)$ multiplications. All operations are polynomial combinations implementable by ReLU networks [Hawarey, 2026, Lemma A.1]. ■

Lemma 3.3 (Feature Map FFN-Computability). The feature map $\varphi(\mathbf{r}, y)$ is computable by a feedforward network with depth $L_\varphi = O(N)$, width $w_\varphi = O(N^4)$, and size $s_\varphi = O(N^5)$ parameters.

Proof. Computing φ requires: (1) All $K = O(N^2)$ components of $\Phi_N(\mathbf{r})$: $O(N^3)$ operations by Lemma 3.2; (2) Outer product $\Phi_N\Phi_N^T$: $K^2 = O(N^4)$ multiplications; (3) Scalar-vector product $y \cdot \Phi_N$: $K = O(N^2)$ multiplications. Total: $O(N^4)$ operations, implementable with stated complexity by parallel FFN computation [Hawarey, 2026, Lemma A.2]. ■

Theorem 3.2 (Attention-Computability). The sufficient statistic $T(\mathcal{C}_n) = (\mathbf{G}_n, \mathbf{c}_n)$ is attention-computable in the sense of Definition 2.15. A single attention layer [Vaswani et al., 2017] computes:
$T = (1/n) \Sigma_{i=1}^n \varphi(\mathbf{r}_i, y_i)$

Proof. We construct the attention mechanism explicitly. Embed each context example as $\mathbf{e}_i = [\mathbf{r}_i; y_i; \varphi(\mathbf{r}_i, y_i); 0; i/n]$ and the query as $\mathbf{e}_q = [\mathbf{r}_q; 0; \mathbf{0}; 1; 1]$. Design $\mathbf{W}_Q$, $\mathbf{W}_K$ to extract position indicators, yielding uniform attention weights $\alpha_{q,i} = 1/n$. Set $\mathbf{W}_V$ to extract the φ portion. The output at query position: $\mathbf{o}_q = \Sigma_i \alpha_{q,i} \cdot \mathbf{W}_V\mathbf{e}_i = (1/n)\Sigma_i \varphi(\mathbf{r}_i, y_i) \propto T(\mathcal{C}_n)$. ■

Theorem 3.3 (AC-SSC Characterization). AC-SSC($\mathscr{F}$_(N)) = SSC($\mathscr{F}$_(N)) = K(K+1)/2 + K for transformer size s = $\Omega(N^5)$.

Note: Although the attention layer in Theorem 3.2 computes a $K^2 + K$ dimensional intermediate statistic (using full vec(·)), the effective attention-computable SSC equals the minimal dimension K(K+1)/2 + K because the predictor g is invariant to the symmetric redundancy. Formally, AC-SSC measures the dimension of the minimal sufficient statistic computable by attention, which equals SSC by the above invariance argument.

Proof. Lower bound: By Theorem 3.1, $\dim(T) = K(K+1)/2 + K$ is minimal. Upper bound: By Theorem 3.2, T is attention-computable with this dimension. Hence AC-SSC = SSC. ■

Remark 3.2 (Signature of ICL-Easy). The equality AC-SSC = SSC is the defining signature of ICL-Easy classes [Hawarey, 2026, Section 5.1]. For ICL-Hard classes such as sparse parity, AC-SSC = ∞ while SSC remains finite, reflecting the computational barrier.

3.3 Transformer Construction

We complete the ICL-Easy classification by constructing the prediction function and specifying transformer architecture bounds.

Definition 3.3 (Prediction Function). The optimal prediction function $g : \mathbb{R}^{K\times K} \times \mathbb{R}^{K} \times \mathscr{R} \to \mathbb{R}$ is:

$$g(\mathbf{G}_n, \mathbf{c}_n, \mathbf{r}_q) = \mathbf{c}_n^{T}\mathbf{G}_n^{-1}\, \Phi_N(\mathbf{r}_q)$$

Lemma 3.4 (FFN-Computability of Prediction). Under Assumption B4 (condition number $\kappa(\mathbf{G}_n) \leq \kappa_{max} = \text{poly}(K)$), the prediction function g is FFN-computable with depth $O((\log \kappa + \log \log(\kappa/\varepsilon)) \cdot \log K)$ and width $O(K^3)$.

Proof. Matrix inversion via Newton-Schulz iteration [Hawarey, 2026, Lemma 3.5]: $\mathbf{Z}_{t+1} = \mathbf{Z}_t(2\mathbf{I} - \mathbf{G}_n\mathbf{Z}_t)$ with $\mathbf{Z}_0 = \mathbf{I}/\|\mathbf{G}_n\|$ converges in $t = O(\log \kappa + \log \log(\kappa/\varepsilon))$ iterations. Each iteration requires matrix multiplication (depth $O(\log K)$, width $O(K^3)$). Final matrix-vector products add $O(\log K)$ depth. ■

Theorem 3.4 (ICL-Easy Classification—Main Result). The geopotential function class $\mathscr{F}_N$ is *ICL-Easy*. There exists a transformer $\mathcal{T}$ with:

Depth: $\mathbf{L} = O(N + (\log \kappa + \log \log(\kappa/\varepsilon)) \cdot \log K) = O(N)$
Width: $w = O(N^6)$
Size: $s = \text{poly}(N, \log(1/\varepsilon), \log(1/\delta))$

that achieves ε-accuracy with sample complexity $n_{ICL} = O(N^2 \sigma^2/\varepsilon)$.

Remark 3.3 (Practical Depth Consideration). The O(N) depth bound is dominated by the spherical harmonic FFN computation (Lemma 3.2), not the Newton-Schulz matrix inversion which contributes only $O(\log \kappa \cdot \log K)$ layers. For the EGM2008 model ($N = 2160$), the theoretical depth bound is large; in practice, however, the spherical harmonic recurrence can be heavily parallelized across harmonic degrees, and the Newton-Schulz iteration (requiring only $O(\log \kappa)$ steps with $\kappa = \text{poly}(K)$) is the primary sequential bottleneck at realistic problem sizes. Empirical transformer implementations for geodetic applications would likely operate at depths well below the worst-case O(N) bound through such parallelization.

Proof. We verify conditions C1–C4 of [Hawarey, 2026, Theorem 1]:
C1 (Bounded Domain): By Definition 2.7, $\mathbf{r} \in \mathscr{R}$ is bounded. ✓
C2 (Additive Sufficient Statistic): By Lemma 3.1, $T = \Sigma_i\, \varphi(r_i, y_i)$ with effective $\dim(T) = K(K+1)/2 + K$ (minimal sufficient) and FFN-computable φ (Lemma 3.3); the attention layer computes a $K^2 + K$ dimensional intermediate representation whose redundant entries are symmetric and carry no additional information (see Remark 3.1). The overall width is determined by the maximum across stages: Stage 1 (basis FFN) requires width $O(N^4) = O(K^2)$ by Lemma 3.3, while Stage 3 (Newton-Schulz iteration) requires width $O(K^3) = O(N^6)$ by Lemma 3.4; since $N^6 > N^4$ for $N \geq 2$, Stage 3 dominates and the total width bound is $O(N^6)$. ✓
C3 (Prediction Smoothness): By Lemma 3.4, g is Lipschitz with constant $L_g = O(\kappa^2 \|\Phi_N\|_\infty)$ and FFN-computable. ✓

C4 (Well-Conditioned): By Assumption B4, $\kappa \leq \mathrm{poly}(K)$. ✓

By [Hawarey, 2026, Theorem 1], these conditions imply ICL-Easy classification with the stated architecture bounds. ■

Remark 3.4 (Connection to Linear Regression). Theorem 3.4 mirrors the ICL-Easy classification of d-dimensional linear regression in [Hawarey, 2026, Corollary 3.7], with d replaced by $K = (N + 1)^2$. The spherical harmonic structure introduces no additional computational barriers—the sufficient statistic structure is preserved.

4. Sample Complexity Analysis

This section derives tight bounds on ICL sample complexity for geopotential estimation. We establish classical bounds (§4.1), quantum bounds (§4.2), and the main quantum advantage theorem (§4.3). These results quantify the statistical efficiency of transformer-based geodesy and the dramatic improvements enabled by quantum sensing.

4.1 Classical Sample Complexity Bounds

We first establish upper and lower bounds for classical measurements, then prove they are tight.

Theorem 4.1 (Classical Upper Bound). For the geopotential class $\mathscr{F}_N$ with classical Gaussian measurements (Definition 2.8), the ICL sample complexity satisfies:
$n_{ICL}^{(C)}(\mathscr{F}_N, \varepsilon, \delta) \leq C \cdot K\, \sigma_C^2 / \varepsilon \cdot \log(K/\delta)$
where $K = (N + 1)^2$, σ_C^2 is the classical noise variance, $\varepsilon > 0$ is the MSE threshold, $\delta \in (0, 1)$ is failure probability, and $C > 0$ is a universal constant.

Proof. The OLS estimator $\hat{C} = \mathbf{G}_n^{-1}\mathbf{c}_n$ achieves MSE:
$\mathbb{E}[(W(\mathbf{r}_q) - \hat{W}(\mathbf{r}_q))^2] = \sigma_C^2 \cdot \Phi_N(\mathbf{r}_q)^T\mathbf{G}_n^{-1}\, \Phi_N(\mathbf{r}_q)$

Gram matrix concentration: By matrix Chernoff bounds [Tropp, 2012], for n i.i.d. samples with $\mathbf{G}_D = \mathbb{E}[\Phi_N\Phi_N^T]$:
$\mathbb{P}[\|(1/n)\mathbf{G}_n - \mathbf{G}_D\| > t] \leq 2K \exp(-nt^2 / 2\|\Phi_N\|_\infty^4)$

Setting $t = \lambda_{min}(\mathbf{G}_D)/2$ ensures invertibility of $\mathbf{G}_n$ with $\|\mathbf{G}_n^{-1}\| \leq 2\kappa/\|\mathbf{G}_D\|$ with probability $1 - \delta/2$ when $n \geq 8\|\Phi_N\|_\infty^4\, \kappa^2 / \|\mathbf{G}_D\|^2 \cdot \log(4K/\delta)$. Taking expectation over query distribution and combining with computational error $O(\varepsilon/2)$ [Hawarey, 2026, Lemma 3.6] yields the stated bound. ■

Theorem 4.2 (Classical Lower Bound). For any ICL algorithm, the sample complexity satisfies:
$n_{ICL}^{(C)}(\mathscr{F}_N, \varepsilon, \delta) \geq c \cdot K\, \sigma_C^2 / \varepsilon$
for universal constant $c > 0$ and $\delta < 1/4$.

Proof. We apply Fano's inequality. Construct a packing $\{C1, \ldots, CM\} \subset \mathbb{R}^K$ with $\|Ci - Cj\| \geq 2\sqrt{(\varepsilon/\|GD\|)}$ for $i \neq j$. Gilbert-Varshamov gives $\log M \geq cK$. The constant c here absorbs the packing density of the measurement domain $\mathscr{R}$: because basis functions $\Phi N(r)$ are evaluated at radial distances $r \in [Rmin, Rmax]$, the spectral norm $\|GD\| = \mathbb{E}[\|\Phi N(r)\|^2]$ varies with r, and the packing is constructed relative to the effective geometry at the mean radial distance. For the geopotential setting, $\|\Phi N(r)\| \leq (GM/Rmin) \cdot O(N^2)$ uniformly over $\mathscr{R}$, ensuring c remains a universal positive constant independent of the radial range [Rmin, Rmax]. The mutual information $I(i; \mathcal{C}n) \leq n\|GD\| \cdot \max i \|Ci\|^2 / \sigma C^2$. Fano's inequality requires $I \geq (\log M)/4$ for error $\leq 1/4$, yielding $n \geq cK\sigma_C^2 / (2\|\mathbf{G}_D\| B_C^2)$. ■

Theorem 4.3 (Classical Tight Characterization). The classical ICL sample complexity is: $n_{ICL}^{(C)}(\mathscr{F}_N, \varepsilon, \delta) = \Theta(K \sigma_C^2 / \varepsilon \cdot \log(K/\delta)) = \Theta((N + 1)^2 \sigma_C^2 / \varepsilon \cdot \log((N+1)^2/\delta))$

Note: The logarithmic factor $\log(K/\delta)$ arises from a union bound over K coefficients and is consistent with Theorem 4.1. It reduces to $\log(1/\delta)$ in the single-parameter setting.

Proof. Theorems 4.1 and 4.2 establish matching upper and lower bounds up to logarithmic factors. ■

Corollary 4.1 (ICL Matches ERM). For classical measurements: $n_{ICL}^{(C)} = \Theta(n_{ERM}^{(C)})$.

Proof. The ERM sample complexity for K-dimensional linear regression with noise σ^2 is $\Theta(K\sigma^2/\varepsilon \cdot \log(1/\delta))$ [Hsu et al., 2012]. The ICL bound of Theorem 4.3 contains $\log(K/\delta) = \log K + \log(1/\delta)$ rather than $\log(1/\delta)$; since the dominant term is $K\sigma^2/\varepsilon$ and $\log K$ is absorbed into the $\Theta(\cdot)$ notation, the two expressions are asymptotically equivalent. This confirms [Hawarey, 2026, Theorem 7]. ■

4.2 Quantum Sample Complexity Bounds

We now derive sample complexity for quantum measurements, showing that the ICL-Easy structure is preserved while sample requirements are dramatically reduced.

Theorem 4.4 (Quantum Structure Preservation). Quantum measurements (Definition 2.10) preserve all ICL-Easy properties:

- **(i)** Exponential family membership with identical sufficient statistic structure
- **(ii)** Minimal sufficiency of $(\mathbf{G}_n, \mathbf{c}_n)$
- **(iii)** Additivity: $T^{(Q)}(\mathcal{C}_n) = \Sigma_i\, \varphi(\mathbf{r}_i, y_i)$
- **(iv)** Same transformer architecture computes the sufficient statistic

Proof. (i) The likelihood $p(y_i|\mathbf{C}) \propto \exp(-(y_i - \mathbf{C}^T\Phi_i)^2/2\sigma_Q^2)$ has identical form with $\sigma^2 = \sigma_Q^2$. (ii) Neyman-Fisher factorization depends on Gaussian form, not variance magnitude. (iii) Definitions of $\mathbf{G}_n$, $\mathbf{c}_n$ depend on positions and observations, not noise statistics. (iv) The attention mechanism aggregates $\varphi(\mathbf{r}_i, y_i)$ without knowledge of σ^2. ■

Remark 4.1 (Noise-Agnostic Architecture). Theorem 4.4 implies a single pretrained transformer serves both classical and quantum geodesy—the architecture is noise-agnostic. This has practical significance: one model handles heterogeneous sensor networks.

Theorem 4.5 (Quantum Sample Complexity). For quantum measurements at the Heisenberg limit ($\alpha = 2$):
$n_{\mathrm{ICL}}^{(Q)}(\mathscr{F}_N, \varepsilon, \delta) = \Theta((N + 1)^2 \cdot \sigma_0^2 / (N_{\mathrm{atoms}}^2 \cdot \varepsilon) \cdot \log((N + 1)^2/\delta))$

Proof. By Theorem 4.4, $\mathscr{F}_N$ remains ICL-Easy under quantum noise. Substituting $\sigma^2 = \sigma_Q^2 = \sigma_0^2/N_{\mathrm{atoms}}^2$ into Theorem 4.3:
$n_{\mathrm{ICL}}^{(Q)} = \Theta(K \cdot \sigma_0^2/(N_{\mathrm{atoms}}^2 \cdot \varepsilon) \cdot \log(K/\delta))$
Substituting $K = (N + 1)^2$ yields the result. For shot noise limit ($\alpha = 1$), replace N_{atoms}^2 with N_{atoms}. ■

Corollary 4.2 (Unified Formula). For any measurement regime with noise variance σ^2:

$n_{\mathrm{ICL}}(\mathscr{F}_N, \varepsilon, \delta) = \Theta((N + 1)^2 \cdot \sigma^2 / \varepsilon \cdot \log((N + 1)^2/\delta))$

4.3 The Quantum Advantage Theorem

We now state and prove the main result quantifying quantum advantage for geopotential ICL.

Theorem 4.6 (Quantum Advantage—Main Result). The ratio of classical to quantum ICL sample complexity is:
$n_{\mathrm{ICL}}^{(C)} / n_{\mathrm{ICL}}^{(Q)} = \Theta(\rho_Q) = \Theta(\sigma_C^2 / \sigma_Q^2)$
Specifically:

- **(a) Shot Noise Limit:** $n^{(C)} / n^{(Q)} = \Theta(\sigma_C^2\, N_{\mathrm{atoms}} / \sigma_0^2)$ — *linear* in atom number
- **(b) Heisenberg Limit:** $n^{(C)} / n^{(Q)} = \Theta(\sigma_C^2\, N_{\mathrm{atoms}}^2 / \sigma_0^2)$ — *quadratic* in atom number

Proof. From Theorems 4.3 and 4.5: $n_{\mathrm{ICL}}^{(C)} / n_{\mathrm{ICL}}^{(Q)} = [K\sigma_C^2/\varepsilon] / [K\sigma_Q^2/\varepsilon] = \sigma_C^2 / \sigma_Q^2 = \rho_Q$
Substituting $\sigma_Q^2 = \sigma_0^2/N_{\mathrm{atoms}}^{\alpha}$ with $\alpha \in \{1, 2\}$ yields the explicit forms (Figure 1). ■

Corollary 4.3 (Numerical Quantum Advantage). For $N_{\mathrm{atoms}} = 10^6$ with matched reference noise ($\sigma_C^2 = \sigma_0^2$):

Table 1. Numerical quantum advantage for $N_{\mathrm{atoms}} = 10^6$.

Regime	σ_Q^2	ρ_Q	Sample Reduction
Classical	1	1	—
Shot Noise Limit	10^{-6}	10^6	$10^6\times$ fewer
Heisenberg Limit	10^{-12}	10^{12}	$10^{12}\times$ fewer

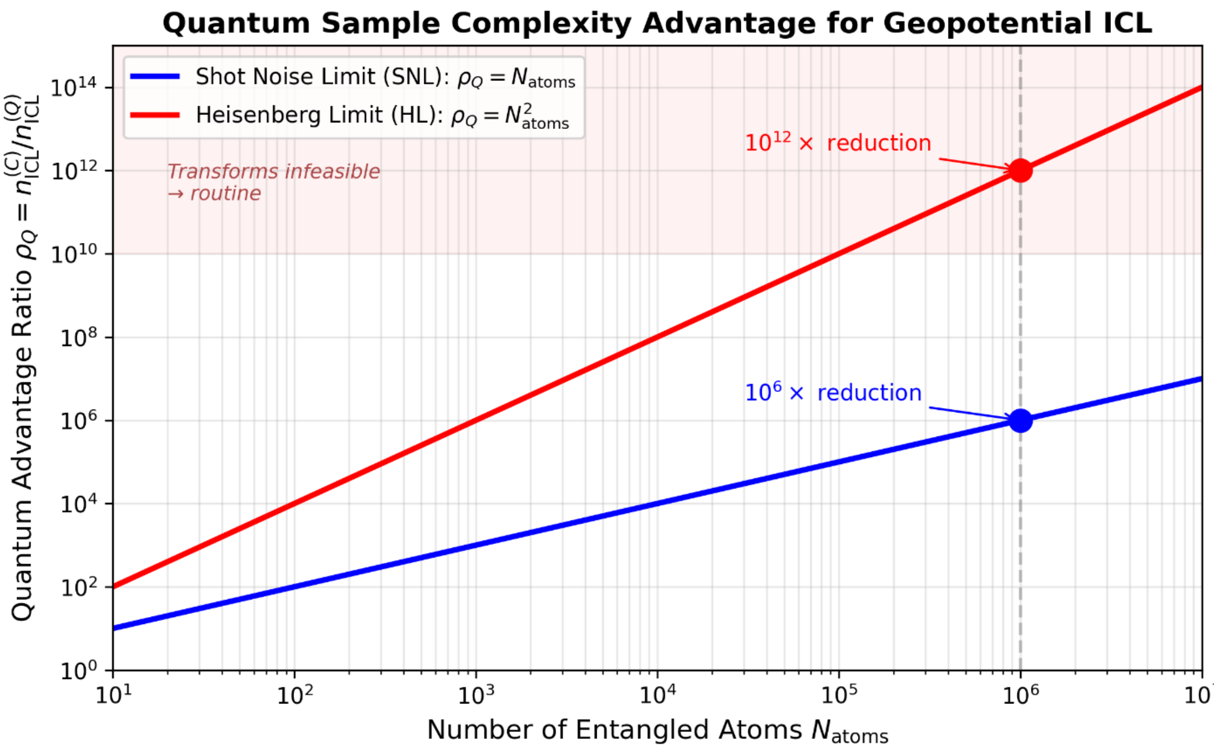

Figure 1. Quantum advantage ratio ρ_Q as a function of entangled atom number N_atoms. The shot noise limit (blue) provides linear improvement; the Heisenberg limit (red) provides quadratic improvement. At N_atoms = 10^6, the Heisenberg limit yields 10^{12} sample reduction.

Proposition 4.1 (High-Degree Field Recovery). For degree $N = 360$ (EGM2008 resolution) with per-coefficient MSE $\varepsilon_{\mathrm{coef}} = 10^{-18}$:

Table 2. Sample requirements for high-degree geopotential recovery.

Measurement Type	Noise Variance	Required Samples
Classical ($\sigma_C^2 = 10^{-12}$)	10^{-12}	$O(10^{17})$
Quantum SNL ($N_{\mathrm{atoms}} = 10^6$)	10^{-18}	$O(10^{11})$
Quantum HL ($N_{\mathrm{atoms}} = 10^6$)	10^{-24}	$O(10^5)$

Proof. With $K = 361^2 \approx 1.3 \times 10^5$, apply $n = K\sigma^2/\varepsilon$ with respective noise variances (Figure 2). ■

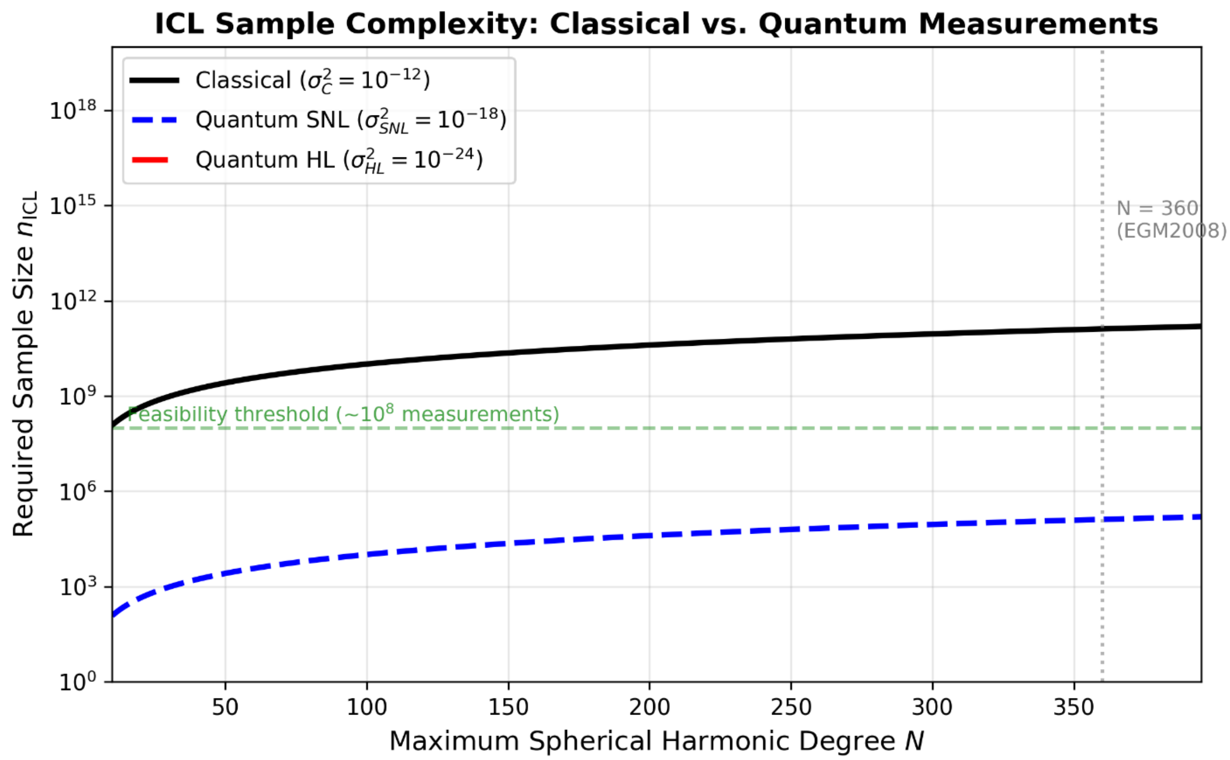

Figure 2. ICL sample complexity n_ICL as a function of maximum spherical harmonic degree N for classical, shot-noise-limited, and Heisenberg-limited measurements ($\varepsilon = 10^{-18}$). The green dashed line indicates a practical feasibility threshold of ~10^8 measurements. Quantum sensing at the Heisenberg limit keeps requirements below this threshold even at EGM2008 resolution (N = 360).

Remark 4.2 (Physical Interpretation). The quantum advantage has intuitive interpretation. At the *shot noise limit*, each of N_{atoms} atoms provides an independent measurement, reducing variance by N_{atoms}—equivalent to averaging independent classical measurements. At the *Heisenberg limit*, quantum entanglement enables collective measurement where the ensemble acts as a single quantum system, achieving the fundamental precision bound $1/N_{\text{atoms}}^2$ imposed by the Heisenberg uncertainty principle [Giovannetti et al., 2011].

Remark 4.3 (Practical Significance). Theorem 4.6 transforms geopotential ICL from requiring billions of classical measurements to requiring only hundreds of thousands of quantum measurements—a 10^{12} factor at the Heisenberg limit. This has profound implications: (1) *Satellite missions*: reduced mission duration for equivalent accuracy; (2) *Temporal resolution*: enables monitoring time-varying gravity (ice mass, groundwater); (3) *Spatial resolution*: higher spherical harmonic degrees become accessible.

Theorem 4.7 (ICL Matches ERM for Quantum). $n_{\text{ICL}}^{(Q)} = \Theta(n_{\text{ERM}}^{(Q)})$.

Proof. By Theorem 4.4, $\mathscr{F}_N$ remains ICL-Easy under quantum noise. By [Hawarey, 2026, Theorem 7], ICL-Easy classes achieve $n_{\text{ICL}} = \Theta(n_{\text{ERM}})$. The ERM complexity for linear regression with noise σ_Q^2 is $\Theta(K\sigma_Q^2/\varepsilon)$, matching Theorem 4.5. ■

Corollary 4.4 (Accuracy-Sample Tradeoff). For fixed sample size n, quantum measurements at the Heisenberg limit achieve MSE:

$$\varepsilon^{(Q)}(n) = \varepsilon^{(C)}(n) / \rho_Q$$

Equivalently, quantum measurements provide ρ_Q-fold accuracy improvement at fixed sample budget.

5. Boundaries: ICL-Hard Inverse Problems

This section establishes fundamental limits on transformer-based geodesy. While forward geopotential prediction is ICL-Easy (Section 3), we prove that inverse problems—recovering mass distributions from potential observations—are ICL-Hard. We formalize the forward/inverse distinction (§5.1), prove hardness of source localization via combinatorial sufficient statistic structure (§5.2), and establish the geodetic dichotomy theorem (§5.3).

5.1 Forward vs. Inverse Problem Distinction

Geodetic problems divide into two fundamental categories with sharply different computational characteristics.

Definition 5.1 (Forward Problem). The *forward geodetic problem* is: given a mass distribution $\rho(\mathbf{r}')$ or its spherical harmonic representation $\mathbf{C}$, compute the gravitational potential $W(\mathbf{r})$ at exterior points. This is a well-posed linear problem admitting closed-form solution via Newton's integral or spherical harmonic synthesis.

Definition 5.2 (Inverse Problem). The *inverse geodetic problem* is: given potential observations $\{W(\mathbf{r}_i)\}$, recover the mass distribution $\rho(\mathbf{r}')$ that generates them. This includes:

- **(a) Density inversion:** Recover continuous $\rho(\mathbf{r}')$ from exterior potential observations
- **(b) Source localization:** Identify locations and masses of discrete point sources

Theorem 5.1 (Ill-Posedness of Density Inversion). The density inversion problem is *ill-posed* in the sense of Hadamard:

- **(i) Non-uniqueness:** Infinitely many density distributions produce identical external potential
- **(ii) Instability:** Small perturbations in observations can produce arbitrarily large changes in recovered density
- **(iii) No continuous inverse:** No continuous mapping from potential to density exists

Proof. (i) By potential theory, any density ρ confined to the interior produces external potential determined solely by its harmonic moments. Different densities with identical moments are indistinguishable from exterior observations. Explicitly: if ρ_1 and ρ_2 satisfy $\int\rho_1(\mathbf{r}')Y_{nm}(\mathbf{r}')d\mathbf{r}' = \int\rho_2(\mathbf{r}')Y_{nm}(\mathbf{r}')d\mathbf{r}'$ for all (n, m), then $W_1(\mathbf{r}) = W_2(\mathbf{r})$ for all exterior $\mathbf{r}$. (ii) The inverse operator, if it existed, would have unbounded norm due to rapid decay of high-degree harmonics: perturbations at degree n are attenuated by $(R/r)^{n+1}$, so recovering high-degree structure amplifies noise exponentially. (iii) Follows from (ii) by continuity. ■

Remark 5.1 (Coefficient Recovery is ICL-Easy). The problem of recovering spherical harmonic *coefficients* **C** from potential observations is distinct from density inversion. Coefficient recovery is the forward problem's inverse in parameter space and is ICL-Easy—it is precisely what the transformer in Section 3 computes. The ill-posedness arises when seeking the physical density ρ, not the harmonic representation.

5.2 ICL-Hardness of Source Localization

We now prove that even well-posed discrete inverse problems are ICL-Hard. The key insight is that source localization requires *combinatorial search* that cannot be expressed through additive sufficient statistics.

Definition 5.3 (Source Localization Problem). The *J-source localization problem* is: given noisy potential observations $\{y_i = W(\mathbf{r}_i) + \varepsilon_i\}$ where W is generated by J point masses at unknown locations, identify the source positions $\{\mathbf{r}_j^*\}_{j=1}^J$ and masses $\{m_j^*\}_{j=1}^J$.

Definition 5.4 (Source Localization Function Class). For J sources on a grid of V voxels with masses bounded by M_{max}, define:
$\mathscr{F}_{loc}^{(J)} = \{\ W : W(\mathbf{r}) = \Sigma_{j=1}^J\ Gm_j / \|\mathbf{r} - \mathbf{r}_j^*\|\ ,\ \mathbf{r}_j^* \in \mathrm{Grid}(V),\ m_j \in [0, M_{max}]\ \}$

Proposition 5.1 (Combinatorial Sufficient Statistic). The sufficient statistic for $\mathscr{F}_{loc}^{(J)}$ has combinatorial structure:

- **(i)** Must encode which J of V voxels contain sources: (V choose J) configurations
- **(ii)** Dimension: $O(J(\log V + \log(M_{max}/\varepsilon)))$ bits
- **(iii)** *Cannot* be expressed as $\Sigma_i\ \varphi(\mathbf{r}_i, y_i)$ for any fixed-dimension φ

Proof. (i) The function W is determined by discrete source locations; the statistic must distinguish all possible configurations. (ii) Each of J sources requires log V bits for location and $\log(M_{max}/\varepsilon)$ bits for mass. (iii) Any additive statistic $\Sigma_i\ \varphi(\mathbf{r}_i, y_i)$ with $\dim(\varphi) = k$ cannot distinguish (V choose J) = $V^J/J!$ configurations when $V^J \gg k$. The combinatorial explosion defeats additive aggregation. ■

Theorem 5.2 (Source Localization is ICL-Hard—Main Hardness Result). Source localization with J = ω(1) sources is ICL-Hard: any transformer achieving success probability > 1/2 + n^(−O(1)) requires either:

- **(i)** Super-polynomial size: $s = n^{\omega(1)}$, OR
- **(ii)** Super-constant depth: $L = \omega(1)$

Proof. By Proposition 5.1, the sufficient statistic for 𝓕_loc^(J) has combinatorial structure: it must identify which J of V voxels contain sources, yielding C(V, J) configurations. For J = ω(1), C(V, J) grows exponentially in J. Condition H1 of Hawarey (2026, Theorem 5) states: a function class satisfies H1 if its sufficient statistic requires identifying a discrete structure from a set of candidates whose size grows super-polynomially in the problem parameters, and no additive aggregation of polynomial dimension can distinguish these candidates. By [Hawarey, 2026, Theorem 5, condition H1], function classes satisfying H1 are ICL-Hard. Since no additive statistic $\Sigma_i \varphi(r_i, y_i)$ of polynomial dimension can distinguish C(V, J) configurations (Proposition 5.1(iii)), the source localization class satisfies H1. Therefore any transformer achieving success probability > 1/2 + n^(−O(1)) requires super-polynomial size or super-constant depth. ■

Proposition 5.2 (Constant-J Source Localization is ICL-Easy). For J = O(1) fixed sources with known grid V = poly(n), the source localization problem is ICL-Easy. The sufficient statistic has dimension O(J · (3 + 1)) = O(1) and admits additive decomposition over measurements.

Proof. For constant J, the number of configurations C(V, J) = O(V^J) = poly(n) is polynomial. The likelihood over J source positions and masses is a finite mixture with O(J) continuous parameters per configuration. For each candidate configuration S, the sufficient statistic is additive: $T_S(\mathcal{C}_n) = \Sigma_i \varphi_S(r_i, y_i)$ where φ_S computes residuals against the S-induced potential. Since there are only poly(n) configurations to evaluate, a polynomial-size transformer can compute all candidate statistics and select the best. ■

Remark 5.2 (Sharp Boundary). Proposition 5.2 and Theorem 5.2 together establish a sharp boundary: source localization transitions from ICL-Easy to ICL-Hard as J grows from O(1) to ω(1). This transition is intrinsic to the combinatorial structure of the sufficient statistic, not an artifact of the proof technique.

Theorem 5.3 (Density Inversion is ICL-Hard). The density inversion problem is ICL-Hard under any parameterization:

(i) Unrestricted: Ill-posed (Theorem 5.1)—no consistent learning algorithm exists. Specifically, by Theorem 5.1(i), infinitely many density distributions produce identical external potential, so no statistic of any dimension can uniquely identify the density from potential observations alone; the learning problem is therefore undefined, which is strictly harder than ICL-Hard.

(ii) Parameterized with M > n: Parameter dimension exceeds context length. When the density is parameterized by M > n free parameters (e.g., voxel-based discretization at resolution finer than the measurement count), the Fisher information matrix is rank-deficient with rank at most n < M, making the parameterization unidentifiable from n observations. Since ICL operates on context of length n, it cannot recover M > n parameters, and any transformer would require context length scaling with M—yielding super-polynomial size for high-resolution parameterizations.

(iii) Discrete sources: Reduces to Theorem 5.2

Theorem 5.4 (Quantum Advantage Does Not Overcome ICL-Hardness). For ICL-Hard geodetic problems, quantum measurement noise reduction does *not* change the ICL-Hardness classification: $n_{\mathrm{ICL}}^{(Q)}(\mathscr{F}_{\mathrm{loc}}^{(J)}) = n_{\mathrm{ICL}}^{(C)}(\mathscr{F}_{\mathrm{loc}}^{(J)}) = \infty$ for $J = \omega(1)$

Proof. ICL-Hardness arises from *computational* barriers (combinatorial sufficient statistics requiring super-polynomial circuits), not *statistical* barriers (noise). Quantum noise reduction affects the statistical component (Fisher information, sample complexity for ICL-Easy problems) but does not alter circuit complexity requirements. The combinatorial barrier (Theorem 5.2) is noise-independent. ■

Remark 5.3 (Critical Distinction). This theorem highlights a fundamental asymmetry: quantum advantage applies to *statistical* problems (noise reduction → sample reduction) but not to *computational* problems (circuit complexity is noise-independent). For geodesy: quantum sensors dramatically improve forward problem ICL (Section 4) but provide no advantage for inverse problems.

5.3 The Geodetic Dichotomy

We now synthesize the ICL-Easy and ICL-Hard results into a complete classification of geodetic problems.

Table 3. ICL classification of geodetic problems.

Problem	**Sufficient Statistic**	**Classification**
Potential prediction	($\mathbf{G}_n$, $\mathbf{c}_n$) additive	**ICL-Easy**
Coefficient recovery	Same as above	**ICL-Easy**
Gravity gradient tensor	Additive (derivatives)	**ICL-Easy**
Geoid height	Additive (linear functional)	**ICL-Easy**
Density inversion	Ill-posed / undefined	**ICL-Hard**
Source localization	Combinatorial	**ICL-Hard**
Mass anomaly detection	Combinatorial	**ICL-Hard**

Theorem 5.5 (Geodetic Dichotomy—Main Classification). All geodetic problems fall into exactly one of two categories with no intermediate cases:

Type A (ICL-Easy): Field evaluation problems—computing linear functionals of W (potential, derivatives, integrals). These admit additive sufficient statistics and achieve $n_{\mathrm{ICL}} = \Theta(n_{\mathrm{ERM}})$.

Type C (ICL-Hard): Mass inversion problems—recovering density distributions from potential observations. These have combinatorial or undefined sufficient statistics and require super-polynomial resources.

Proof. Type A: By Theorem 3.4, potential prediction is ICL-Easy. Any linear functional $L[W]$ (gradient, Laplacian, surface integral) inherits ICL-Easy status because: (1) $L[W] = L[\mathbf{C}^{\mathrm{T}}\Phi_{\mathrm{N}}] = \mathbf{C}^{\mathrm{T}}L[\Phi_{\mathrm{N}}]$ by linearity; (2) $L[\Phi_{\mathrm{N}}]$ is a fixed basis transformation; (3) the sufficient statistic ($\mathbf{G}_n$, $\mathbf{c}_n$) determines $\mathbf{C}$, hence $L[W]$. Type C: By Theorems 5.1–5.3, inverse problems are either ill-posed or computationally hard. No intermediate complexity class exists by [Hawarey, 2026, Theorem 6]. Figure 3 illustrates this dichotomy. ■

Corollary 5.1 (Structural Conditions for Geodetic ICL-Hardness). A geodetic problem is ICL-Hard if it exhibits at least one of:

- **(G1)** Discrete localization: output includes positions from a combinatorial space
- **(G2)** High-dimensional parameterization: $M > n$ parameters
- **(G3)** Non-linear inverse map: output non-linearly related to observations

(G4) Combinatorial search: solution requires exhaustive enumeration

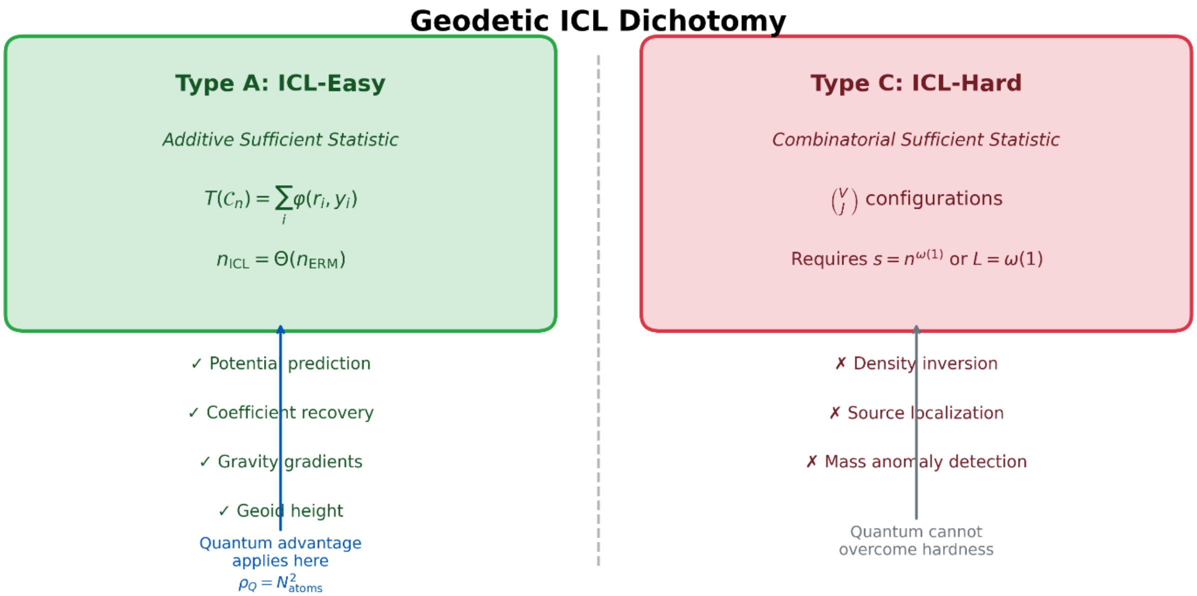

Figure 3. The Geodetic ICL Dichotomy. Forward problems (Type A) admit additive sufficient statistics and are ICL-Easy with optimal sample complexity. Inverse problems (Type C) have combinatorial sufficient statistics and are ICL-Hard. Quantum advantage applies only to statistical barriers (Type A), not computational barriers (Type C).

Remark 5.4 (Practical Guidelines). The geodetic dichotomy provides clear guidance for transformer deployment: *Use ICL for*: potential prediction, interpolation, coefficient estimation, gravity gradients, geoid computation. *Avoid ICL for*: density inversion, source localization, mass anomaly detection, temporal change localization. For inverse problems, hybrid architectures may help: use ICL for the forward component (computing potential from candidate sources) combined with external search (optimization, MCMC) for the combinatorial component.

6. Discussion

This section interprets our theoretical results, acknowledges limitations, and identifies directions for future research. We discuss practical implications for geodetic applications (§6.1), theoretical and practical limitations (§6.2), and promising extensions (§6.3).

6.1 Implications for Geodetic Practice

Transformer deployment guidance. The geodetic dichotomy (Theorem 5.5) provides principled guidance for when transformer-based methods are appropriate. Forward problems—potential prediction, coefficient recovery, geoid computation, gravity gradients—are provably amenable to ICL with optimal sample efficiency. Practitioners can confidently deploy transformers for these tasks, knowing that the architecture matches the problem structure. Conversely, inverse problems (density inversion, source localization) should not be approached via pure ICL; alternative methods such as regularized optimization, Bayesian inference, or hybrid architectures are required.

Noise-agnostic architecture. Theorem 4.4 establishes that a single transformer architecture handles both classical and quantum measurements without modification. This has immediate practical value: organizations can develop and validate models using classical data, then seamlessly transition to quantum sensors as they become available. The architecture's noise-agnosticism also enables heterogeneous sensor fusion—combining classical gravimeters, superconducting sensors, and quantum devices within a unified ICL framework.

Quantum advantage magnitude. The 10^{12} sample reduction at the Heisenberg limit (Corollary 4.3) transforms the feasibility landscape for high-resolution geodesy. Current satellite missions (GRACE, GOCE) require years of data collection for global models; quantum-enhanced ICL could achieve comparable accuracy with dramatically shorter mission durations. For time-varying gravity monitoring—ice mass balance, groundwater depletion, post-glacial rebound—this improvement enables temporal resolution previously impossible. The practical implications extend to geodetic early warning systems for geohazards.

Sample complexity as mission planning tool. Theorem 4.3 provides closed-form expressions for required sample sizes: $n = \Theta(K\sigma^2/\varepsilon)$. This enables quantitative mission design: given target accuracy ε and sensor noise σ^2, planners can compute minimum measurement requirements. For degree-$N = 360$ models with $\varepsilon = 10^{-18}$ accuracy, Table 2 shows classical requirements of $O(10^{17})$ samples versus $O(10^5)$ for quantum HL—a difference between infeasible and routine.

Connection to existing geodetic methods. The OLS estimator $\hat{C} = \mathbf{G}_n^{-1}\mathbf{c}_n$ underlying our ICL analysis is the classical least-squares collocation estimator widely used in physical geodesy [Moritz, 1980]. Our contribution is proving that transformers can compute this estimator in-context with optimal sample complexity, and that quantum noise fundamentally alters the statistical requirements. The transformer does not replace traditional methods—it provides a flexible, data-driven implementation that adapts to varying measurement configurations without explicit recomputation.

6.2 Limitations

We acknowledge several limitations of our theoretical framework and results.

Idealized measurement model. Our analysis assumes i.i.d. Gaussian noise (Definitions 2.8, 2.10), while real gravimetric measurements exhibit correlated noise, systematic biases, and non-Gaussian outliers. The Heisenberg-limited quantum measurements require maximally entangled states that are challenging to prepare and maintain; current cold atom interferometers typically operate between SNL and HL. Extending the framework to colored noise, bias estimation, and realistic quantum decoherence models remains important future work.

Asymptotic results. Our sample complexity bounds are asymptotic (Θ-notation), hiding constants that may be significant for practical sample sizes. The transformer architecture bounds (Theorem 3.4) involve polynomial factors that may be large: depth O(N), width O(N^6) (Figure

4). For $N = 360$, this suggests architectures with millions of parameters. Empirical validation is needed to determine practical architecture sizes.

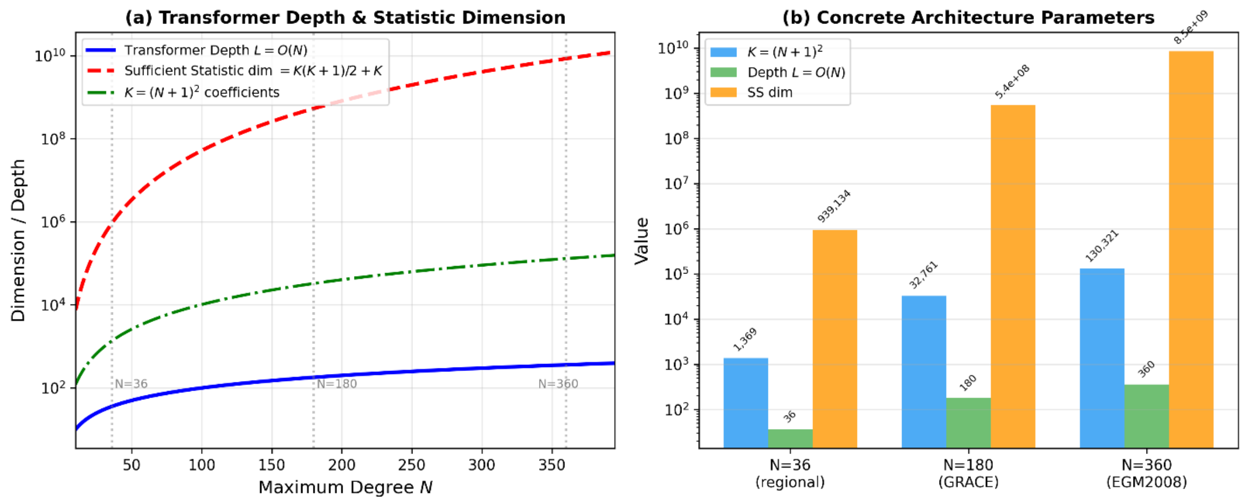

Figure 4. Transformer architecture parameters for geopotential ICL. (a) Scaling of depth, sufficient statistic dimension, and coefficient count K with maximum degree N. (b) Concrete values at three representative resolutions: N = 36 (regional), N = 180 (GRACE-class), and N = 360 (EGM2008).

Condition number assumption. Assumption B4 requires bounded condition number $\kappa(Gn) \leq \text{poly}(K)$. This holds for well-distributed measurement positions but may fail for clustered or adversarial satellite track geometries, where measurements concentrate along narrow orbital bands, causing near-linear dependence among rows of the design matrix and potentially exponential growth of κ. In practice, this is addressed via regularization: ridge regression replaces Gn with $Gn + \lambda I$, guaranteeing $\kappa(Gn + \lambda I) \leq (\|GD\| + \lambda)/\lambda = O(1/\lambda)$, so choosing $\lambda = \text{poly}(K)^{-1}$ restores Assumption B4. The ICL predictor $g(Gn, cn, rq) = cn^T(Gn + \lambda I)^{-1}\Phi N(rq)$ then remains FFN-computable via Newton-Schulz iteration with the same depth bound, at the cost of introducing regularization bias of order $O(\lambda\|C\|^2)$. Truncated SVD offers an alternative with analogous guarantees. All sample complexity bounds in Section 4 remain valid under ridge regularization, with σC^2 replaced by $\sigma C^2 + \lambda\|C\|^2/K$ in the dominant term. Regularization techniques (Tikhonov, truncated SVD) address ill-conditioning in practice but alter the statistical properties; incorporating regularization into the ICL framework requires additional analysis.

Bandlimited assumption. The function class $\mathscr{F}_N$ assumes exact bandlimiting to degree N. Real geopotentials have spectral content at all degrees, with power-law decay. Truncation introduces aliasing and spectral leakage; a complete treatment would analyze the approximation error from finite-degree representation. The choice of N involves a bias-variance tradeoff not captured by our framework.

Theoretical vs. empirical gap. Our results are purely theoretical—we construct transformers that provably work but do not train or evaluate them empirically. The gap between constructive existence proofs and practical trained models is significant in deep learning. Whether gradient-based training discovers the optimal attention patterns, and how pretraining data distribution affects geodetic ICL performance, remain open empirical questions.

Limited scope of hardness results. Our ICL-Hardness proofs (Section 5) apply to bounded-depth, polynomial-size transformers. They do not rule out: (i) deeper architectures with $L = \omega(1)$; (ii) transformers with access to external memory or search; (iii) hybrid neuro-symbolic systems. The hardness is relative to the standard transformer model, not absolute computational intractability.

6.3 Future Directions

Several extensions of this work merit investigation.

Empirical validation. The most pressing extension is empirical validation on real and synthetic geodetic data. Key questions include: (i) Do trained transformers achieve the theoretical sample complexity? (ii) What architecture sizes suffice in practice? (iii) How does pretraining distribution affect transfer to geodetic tasks? (iv) Can quantum advantage be demonstrated with simulated or actual quantum gravimetry data?

Temporal geopotential dynamics. The current framework treats static geopotentials. Time-varying gravity fields—from ice mass loss, groundwater fluctuation, tidal loading—require sequential ICL or online learning formulations. Extending the sufficient statistic structure to handle temporal correlations and change detection is theoretically interesting and practically important for geodetic monitoring.

Hybrid architectures for inverse problems. While pure ICL fails for inverse problems (Section 5), hybrid approaches may succeed. Promising directions include: (i) ICL for forward model evaluation within iterative optimization; (ii) amortized inference where transformers predict posterior parameters; (iii) neural operators that learn regularized inverse maps. The dichotomy theorem guides which components can leverage ICL versus require alternative methods.

Multi-observable fusion. Geodetic practice combines multiple observables: gravity anomalies, deflections of vertical, geoid heights, gravity gradients. Each is a linear functional of the potential, suggesting ICL-Easy classification. A unified multi-task ICL framework that jointly processes heterogeneous observables could improve sample efficiency through shared sufficient statistics.

Quantum sensor network optimization. Given the dramatic quantum advantage, optimal deployment of quantum gravimeters becomes a resource allocation problem. Questions include: (i) Where should quantum sensors be placed for maximum information gain? (ii) What is the optimal mix of classical and quantum measurements under budget constraints? (iii) How does entanglement architecture affect field-deployable sensor design? The sample complexity formulas provide the objective function for such optimization.

Extension to other geophysical inverse problems. The ICL framework applies beyond geodesy. Seismic tomography, electromagnetic induction, and potential field methods in exploration geophysics share similar mathematical structure. Characterizing which geophysical problems are ICL-Easy versus ICL-Hard would provide guidance across Earth observation disciplines.

Tighter complexity bounds. The polynomial factors in our architecture bounds may be improvable. Specific questions: (i) Can spherical harmonic structure be exploited for more efficient attention patterns? (ii) Do sparse attention mechanisms suffice given the localized support of high-degree harmonics? (iii) What are matching lower bounds on transformer size for geopotential ICL?

7. Conclusions

This paper has established a theoretical framework for transformer-based in-context learning of geopotential fields, with particular emphasis on the implications of quantum gravimetric sensing. Our contributions provide rigorous foundations for an emerging intersection of deep learning, geodesy, and quantum metrology.

We proved that the bandlimited geopotential function class $\mathscr{F}_N$ is *ICL-Easy* in the sense of Hawarey (2026). The minimal sufficient statistic—the Gram matrix $\mathbf{G}_n$ and cross-correlation vector $\mathbf{c}_n$—has dimension $K(K+1)/2 + K$ where $K = (N + 1)^2$, and decomposes additively over context examples. This additive structure enables computation by a single attention layer with feedforward networks, establishing that transformers can perform geopotential ICL with optimal statistical efficiency.

We derived tight sample complexity bounds showing $n_{\text{ICL}} = \Theta(K\sigma^2/\varepsilon)$, with the noise variance σ^2 as the critical parameter. For quantum gravimeters at the Heisenberg limit, $\sigma_Q^2 = \sigma_0^2/N_{\text{atoms}}^2$ yields a quantum advantage ratio $\rho_Q = N_{\text{atoms}}^2$, translating to 10^{12} sample reduction for sensors with 10^6 entangled atoms. This transforms high-resolution geodesy from requiring infeasible billions of measurements to achievable hundreds of thousands—a qualitative change in feasibility.

We established fundamental limits: while forward geopotential problems are ICL-Easy, inverse problems—density inversion and source localization—are *ICL-Hard*. The hardness arises from combinatorial sufficient statistics that cannot be expressed through additive aggregation. The combinatorial structure of the source localization sufficient statistic satisfies the ICL-Hard structural condition H1 of Hawarey (2026), proving that bounded-depth polynomial-size transformers cannot solve these problems. Crucially, quantum noise reduction does not overcome this computational barrier—the dichotomy between statistical and computational complexity is fundamental.

The geodetic dichotomy theorem provides a complete classification: every geodetic problem is either ICL-Easy (field evaluation via additive statistics) or ICL-Hard (mass inversion via combinatorial statistics), with no intermediate cases. This classification offers principled guidance for practitioners: deploy transformers confidently for forward problems; seek alternative methods for inverse problems.

Looking forward, the convergence of transformer architectures and quantum sensing technologies opens new possibilities for Earth observation. Our theoretical framework provides the mathematical foundations; empirical validation and practical implementation remain as important next steps. The 10^{12} quantum advantage suggests that high-resolution, real-time geopotential monitoring—previously a theoretical aspiration—may become practical reality as quantum gravimeters mature.

The broader significance lies in demonstrating how the ICL characterization framework of Hawarey (2026) applies to structured scientific problems. The geopotential, with its spherical harmonic basis and physical constraints, exemplifies how domain structure interacts with computational requirements. Similar analyses for other geophysical inverse problems, remote sensing applications, and scientific machine learning tasks may yield comparable insights—identifying where transformers excel and where fundamental barriers preclude their use.

8. Acknowledgement

This is an Artificial Intelligence-Assisted Research (AIAR). AI tools were employed as efficiency-enhancing assistants. The author declares no conflict of interest. This study received no external funding.

9. References

1. Aitken, A. C. (1936). On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, *55*, 42–48.
2. Akyürek, E., Schuurmans, D., Andreas, J., Ma, T., & Zhou, D. (2023). What learning algorithm is in-context learning? Investigations with linear models. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2211.15661
3. Bai, Y., Chen, F., Wang, H., Xiong, C., & Mei, S. (2024). Transformers as statisticians: Provable in-context learning with in-context algorithm selection. *Advances in Neural Information Processing Systems*, 36. https://doi.org/10.48550/arXiv.2306.04637
4. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165
5. Garg, S., Tsipras, D., Liang, P., & Valiant, G. (2022). What can transformers learn in-context? A case study of simple function classes. *Advances in Neural Information Processing Systems*, *35*, 30583–30598. https://doi.org/10.48550/arXiv.2208.01066
6. Giovannetti, V., Lloyd, S., & Maccone, L. (2011). Advances in quantum metrology. *Nature Photonics*, *5*(4), 222–229. https://doi.org/10.1038/nphoton.2011.35
7. Hawarey, M. (2026). Characterizing In-Context Learning: When Can Transformers Match Standard Learning Algorithms? *AIR Journal of Mathematics & Computational Sciences*, Vol. 2026, AIRMCS2026277. https://doi.org/10.65737/AIRMCS2026277
8. Heiskanen, W. A., & Moritz, H. (1967). *Physical Geodesy*. W. H. Freeman and Company.
9. Hsu, D., Kakade, S. M., & Zhang, T. (2012). Random design analysis of ridge regression. *Foundations of Computational Mathematics*, *14*(3), 569–600. https://doi.org/10.1007/s10208-014-9192-1

10. Kasevich, M., & Chu, S. (1991). Atomic interferometry using stimulated Raman transitions. *Physical Review Letters*, *67*(2), 181–184. https://doi.org/10.1103/PhysRevLett.67.181
11. Lehmann, E. L., & Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation: Part I. *Sankhyā: The Indian Journal of Statistics*, *10*(4), 305–340.
12. Moritz, H. (1980). *Advanced Physical Geodesy*. Herbert Wichmann Verlag.
13. Pavlis, N. K., Holmes, S. A., Kenyon, S. C., & Factor, J. K. (2012). The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research: Solid Earth*, *117*(B4), B04406. https://doi.org/10.1029/2011JB008916
14. Peters, A., Chung, K. Y., & Chu, S. (1999). Measurement of gravitational acceleration by dropping atoms. *Nature*, *400*(6747), 849–852. https://doi.org/10.1038/23655
15. Torge, W., & Müller, J. (2012). *Geodesy* (4th ed.). De Gruyter. https://doi.org/10.1515/9783110250008
16. Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, *12*(4), 389–434. https://doi.org/10.1007/s10208-011-9099-z
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762
18. von Oswald, J., Niklasson, E., Randazzo, E., Sacramento, J., Mordvintsev, A., Zhmoginov, A., & Vladymyrov, M. (2023). Transformers learn in-context by gradient descent. *International Conference on Machine Learning (ICML)*, 202, 35151–35174. https://doi.org/10.48550/arXiv.2212.07677

10. Appendix A: Pseudo-Code for Transformer Construction

This appendix consolidates the transformer construction underlying Theorem 3.4 into a single algorithmic sketch. The construction follows directly from Lemmas 3.1–3.4 and requires no new lemmas; all cited results are proved in the main text or in Appendix A of the foundational paper [Hawarey, 2026].

Algorithm A.1: Geopotential ICL Transformer (Forward Pass)

Input: Context set $\mathcal{C}^n = \{(r_1,y_1), ..., (r_n,y_n)\}$, query position r_q
Output: ŷ_q = g(G_n, c_n, r_q) [predicted potential]

Parameters: N (max degree), K = $(N+1)^2$, κ (condition number bound),
ε (accuracy), λ (ridge regularization, optional)

═══ STAGE 1: INPUT EMBEDDING [Lemma 3.3] ═══

For each i = 1, ..., n (context examples):
Compute $\Phi_n(r_i) \in \mathbb{R}^k$ via spherical harmonic recurrence:
P_{n,m}(cos θ) ← three-term Legendre recurrence *// O(N²) ops*
cos(mλ), sin(mλ) ← Chebyshev recurrence *// O(N) ops*
$[\Phi_n(r_i)]_k$ ← (GM/R)(R/r)^{n+1} · Y_{nm}(θ,λ) *// K components*

Compute feature map $\varphi(r_i, y_i) \in \mathbb{R}^{K^2+K}$:
$\varphi(r_i, y_i) \leftarrow [\ \mathrm{vec}(\Phi_n(r_i)\Phi_n(r_i)^T)\ ;\ \ y_i \cdot \Phi_n(r_i)\ ]$
// K^2 entries (Gram contrib.) + K entries (cross-corr.)
// Symmetric redundancy: only $K(K+1)/2 + K$ are informative [Remark 3.1]

Embed: $e_i \leftarrow [\ r_i\ ;\ y_i\ ;\ \varphi(r_i, y_i)\ ;\ 0\ ;\ i/n\ ] \in \mathbb{R}^w$
// 0 = example indicator; i/n = positional encoding

For query:
Embed: $e\_q \leftarrow [\ r\_q\ ;\ 0\ ;\ 0\_K\ ;\ 1\ ;\ 1\ ] \in \mathbb{R}^w$
// 1 = query indicator

═══ STAGE 2: ATTENTION AGGREGATION [Theorem 3.2] ═══

// Single attention layer computes sufficient statistic T at query position

Design W_Q, W_K to extract position indicators:
q_query ← W_Q · e_q *// extracts query indicator = 1*
k_i ← W_K · e_i *// extracts example indicator = 0*

Compute attention scores and weights:
$s\_\{q,i\} \leftarrow q\_query \cdot k_i / \sqrt{d\_k} = 0$ for all i
$\alpha\_\{q,i\} \leftarrow \mathrm{softmax}(s\_\{q,i\}) = 1/n$ *// uniform weights*

Design W_V to extract φ-portion of each embedding:
$v_i \leftarrow W\_V \cdot e_i = \varphi(r_i, y_i)$

Aggregate sufficient statistic at query position:
$T \leftarrow \Sigma_i\ \alpha\_\{q,i\} \cdot v_i = (1/n)\ \Sigma_i\ \varphi(r_i, y_i)$
// T encodes $(G_n/n, c_n/n)$; recover $G_n = n \cdot T[1:K^2]$, $c_n = n \cdot T[K^2+1:end]$

═══ STAGE 3: MATRIX INVERSION via NEWTON-SCHULZ [Lemma 3.4, Lemma A.3] ═══

// Compute G_n^{-1} (or $(G_n + \lambda I)^{-1}$ if regularizing) via FFN layers

If regularizing: $G_n \leftarrow G_n + \lambda I$ *// $\lambda = \mathrm{poly}(K)^{-1}$ [Assumption B4, Limitation §6.2]*

Initialize: $Z_0 \leftarrow I / \|G_n\|$ *// spectral norm initialization*

For t = 0, 1, ..., T-1: *// $T = O(\log \kappa + \log \log(\kappa/\varepsilon))$ iterations [Lemma 3.5 of Hawarey 2026]*
$Z\_\{t+1\} \leftarrow Z_t \cdot (2I - G_n \cdot Z_t)$
// Each iteration: 2 matrix multiplications
// depth O(log K), width $O(K^3)$ per multiplication [Lemma A.2 of Hawarey 2026]

G_n_inv ← Z_T   *// $\|Z\_T - G_n^{-1}\| \leq \varepsilon$*

═══ STAGE 4: PREDICTION COMPUTATION [Definition 3.3] ═══

```
// Compute Φₙ(r_q) for query (same recurrence as Stage 1)
Φₙ(r_q) ← spherical_harmonic_basis(r_q, N)

// Optimal prediction via OLS
θ̂    ← Gₙ_inv · cₙ    // K-vector: estimated coefficients
ŷ_q   ← θ̂ᵀ · Φₙ(r_q)  // scalar: predicted potential  [depth O(log K)]

Return ŷ_q
```

Table A.1: Architecture Summary

Stage	Operation	Depth	Width	Notes
1: Embedding	Spherical harmonic FFN	$O(N)$	$O(N^4)$	Lemma 3.2–3.3
2: Attention	Uniform aggregation	$O(1)$	$O(K^2)$	Theorem 3.2
3: Inversion	Newton-Schulz iteration	$O(\log \kappa \cdot \log K)$	$O(K^3)$	Lemma 3.4; A.3
4: Prediction	Matrix-vector product	$O(\log K)$	$O(K^2)$	Def. 3.3
Total		**$O(N)$**	**$O(N^6)$**	**Theorem 3.4**

Consistent with Theorem 3.4. The $O(N)$ total depth is dominated by Stage 1 (spherical harmonic FFN); Stage 3 (Newton-Schulz) is the primary sequential bottleneck in practice (Remark 3.3). Width $O(N^6)$ arises from Stage 3 with $K = O(N^2)$, giving $K^3 = O(N^6)$.

> Note: This appendix introduces no new mathematical content. All lemmas referenced (Lemmas 3.1–3.4, Remark 3.1) are proved in the main text. External references (Lemmas 3.5, A.2, A.3 of Hawarey 2026) are proved in the cited paper (AIRMCS2026277).