Closed-Form Computational Possibility Spaces for Bioinformatic Discovery, Genome Design, and Constant-Time Computing

Alexis Eleanor Fagan

Plan A Biotech

Abstract

This white paper defines four mathematically rigorous, closed-form possibility spaces for scalable, exact computation in biology and computing. The first section develops exact Bayesian inference over multiplexed biomarker vectors using radial integration. The second introduces a genome design and search framework by inverting phenotype-to-genotype mappings over embedded manifolds. The third formalizes the space of constant-time (O(1)) classical acceleration based on logic and linear operators. The fourth fully defines the conditions and limitations under which O(1) genome search can be achieved through idealized or structurally constrained settings.

Contents

1 Closed-Form Inference in Multiplexed Biomarker Assays

1.1 Problem Definition

Let $y \in \mathbb{R}^N$ be a vector of observed noisy biomarker levels. Suppose each coordinate is a channel in a multiplexed assay (e.g. expression level, methylation, or protein concentration). The generative model is:

$$y = s + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$$

Here, $s \in \mathbb{R}^N$ is the latent biological signal, modeled as:

$$s \sim \mathcal{N}(0, \tau^2 I)$$

This reflects prior isotropic uncertainty across biomarkers.

1.2 Exact Posterior and Estimation

By Gaussian–Gaussian conjugacy, the posterior distribution is:

$$p(s \mid y) = \mathcal{N}\left(\mu_{\text{post}}, \Sigma_{\text{post}}\right)$$

with:

$$\mu_{\text{post}} = \frac{\tau^2}{\sigma^2 + \tau^2} y, \quad \Sigma_{\text{post}} = \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2} I$$

This yields a closed-form point estimate and confidence radius per channel.

1.3 Detection and Hypothesis Testing

We test $H_0 : s = 0$ vs. $H_1 : \|s\| > 0$ using the radial statistic $\|y\|$. Under H_0, we have:

$$\frac{\|y\|}{\sigma} \sim \chi_N$$

Setting a threshold T, the false-alarm rate is:

$$P_{FA} = \mathbb{P}(\|y\| > T) = 1 - F_{\chi_N}(T/\sigma)$$

If a true signal s_0 is present, $\|y\| \sim \chi'_N(\lambda)$ with noncentrality $\lambda = \|s_0\|/\sigma$, giving closed-form detection probabilities via the generalized Marcum Q-function.

2 Closed-Form Genome Retrieval from Trait Space

2.1 Problem Definition

Let $g \in \{0,1\}^L$ be a genome of fixed length L, and let $\phi : \{0,1\}^L \to \mathbb{R}^T$ map a genome to its expressed traits $t = \phi(g)$.

Given a desired trait vector $t_0 \in \mathbb{R}^T$, the goal is to find a genome g^* such that:

$$\phi(g^*) \approx t_0$$

2.2 Radial Integral over Trait Manifold

We treat trait space as an embedded manifold. Let $\rho(t)$ be the trait density induced by sampling genomes. The probability mass near t_0 is:

$$P(\|t - t_0\| < \varepsilon) = \int_{r=0}^{\varepsilon} \int_{S^{T-1}} \rho(t_0 + ru)\, r^{T-1} dS(u) dr$$

This formulation enables exact marginal likelihood estimation for genome matching.

2.3 Closed-Form Surrogates

If $\phi(g) \approx Ag$ is approximately linear (e.g. via regression or learned encoder), then:

$$g^* = \arg \min_{g \in \{0,1\}^L} \|Ag - t_0\|^2$$

Relaxing $g \in \{0,1\}^L$ to $g \in [0,1]^L$ yields a convex program with closed-form minimizer:

$$g_{relaxed} = A^+ t_0$$

Round or decode to the nearest binary genome.

3 Constant-Time (O(1)) Classical Acceleration

3.1 Definition of O(1) Possibility Space

Let $f : \{0,1\}^n \to \{0,1\}^m$ be a function. f is computable in O(1) if it belongs to one of the following classes:

- **$\mathbf{NC_0}$**: Constant-depth, bounded-fanin Boolean circuits.
- **$\mathbf{AC_0}$**: Constant-depth, unbounded-fanin circuits with AND/OR gates.
- **Linear transforms**: $f(x) = Ax \mod 2$, computable with bitwise operations.

3.2 Conditions for Realization

1. Logic gates or arithmetic units must operate in constant time.
2. Vector or SIMD instructions must allow all bits to be processed simultaneously.
3. Operand size must remain fixed (not scaling with input).

3.3 Feasible Classes

- Constant-degree polynomials over reals (real-RAM model)
- Constant-width neural networks
- Fixed-size Fourier transforms

4 O(1) Genome Search: Full Formulation

4.1 Problem Formulation

Let $\mathcal{G} = \{0,1\}^L$ be the set of all possible genomes. A search query is defined by a predicate $P : \mathcal{G} \to \{0,1\}$.

Goal: Find any genome $g \in \mathcal{G}$ such that $P(g) = 1$, using O(1) time per query.

4.2 Impossibility Without Structure

Brute-force search is $O(2^L)$. To reduce this to $O(1)$, the following must hold:

- A perfect hash function $h(g)$ computable in constant time
- A lookup table $T[h(g)] = P(g)$
- Memory availability for $T \in \{0,1\}^{2^L}$

This is infeasible for large L (e.g., $L > 30$).

4.3 Structural Reduction

To achieve feasible O(1) genome search:

1. Restrict predicates to a fixed, known set (e.g., presence of motifs)
2. Precompute lookup tables for only these k predicates
3. Use compressed indexes (suffix tries, bitvectors)

4.4 Compressed Search with Word-Parallelism

- Build a trie of known sequences or motifs
- Assign each node a unique prefix code
- Use word-aligned search via SIMD or hash folding

4.5 Feasible Possibility Space Summary

O(1) genome search is realizable when:

- Genome length L is small (e.g., $L \leq 20$)
- Queries are fixed (e.g., 100 motifs)
- Word-parallel structures or hashes are used

5 Conclusion

We have defined four closed-form possibility spaces:

1. Exact biomarker inference using radial vector calculus
2. Genome design via inverse trait projection
3. Constant-time classical function classes (e.g., NC_0)
4. O(1) genome search via structural or indexed constraints

Each is immediately applicable to scalable diagnostics, trait-based genome engineering, and efficient computing architectures central to Plan A Biotech's roadmap.