Research

Bayesian and Probabilistic Inference Frameworks for Intelligent Parameter Tuning in Next-Generation Computational Systems

Youssef El Mansouri[1],Omar Benkacem[2]

Tangier Institute of Technology, Department of Electronics and Communication Engineering, 12 Avenue Moulay Ismail, Tangier, Morocco[1]
Marrakech University of Applied Sciences, Department of Electrical and Information Engineering, 45 Rue Ibn Sina, Marrakech, Morocco[2]

Abstract: Emerging computational systems increasingly rely on algorithmic components whose performance depends on high-dimensional and context-sensitive parameters. As workloads diversify and hardware architectures evolve, manual tuning rarely scales and static heuristics degrade when operational conditions change. Bayesian and probabilistic inference offer a principled language for representing uncertainty about latent performance functions, for combining heterogeneous sources of evidence, and for making sequential decisions that trade off exploration and exploitation. This paper develops a neutral and technically detailed account of how Bayesian modeling and related probabilistic frameworks enable intelligent parameter tuning in next-generation computational environments, including accelerators, distributed runtimes, and adaptive data processing pipelines. The exposition emphasizes modular generative models for performance signals, posterior inference mechanisms suitable for low-latency control loops, and decision-theoretic criteria aligned with service-level objectives and safety constraints. Sequential design techniques, variational approximations, and probabilistic programming tools are described in the context of real-time feedback and multi-fidelity measurements, while robustness is treated through risk-sensitive objectives and distribution shift diagnostics. The presentation avoids domain-specific claims and restricts itself to model constructions, algorithmic templates, and analysis strategies that can be composed with system-level scheduling and monitoring. The discussion also highlights implementation considerations such as amortized inference, streaming updates, and compute–communication trade-offs on heterogeneous platforms. The overall aim is to delineate precise probabilistic formulations for tuning problems, to articulate their computational realizations, and to summarize evaluation protocols that quantify uncertainty-aware adaptation without presupposing particular benchmarks or vendor-specific stacks.

1. Introduction

Intelligent parameter tuning represents a unifying paradigm for optimizing complex computing and learning systems whose performance depends on numerous interacting configuration variables [1]. These variables may control compiler flags, kernel launch parameters, cache hierarchies, memory allocation policies, learning rates, degrees of parallelism, and mappings across heterogeneous hardware devices. The central challenge arises from the fact that the relationship between configuration parameters and observed performance is typically nonconvex, nonlinear, and highly context-dependent. Furthermore, measurements of performance metrics—such as latency, throughput, energy consumption, or predictive accuracy—are inherently noisy and often only partially observable due to monitoring limitations or external system interference. Consequently, the search for optimal parameter settings cannot rely solely on deterministic optimization but instead benefits from a probabilistic formulation that explicitly represents uncertainty and partial observability.

Table 1. Core Stages of Intelligent Parameter Tuning

Stage	Function	Typical Techniques
Modeling	Define surrogate or prior linking parameters, context, and performance	Gaussian processes, Bayesian linear models, neural ensembles
Inference	Update posterior over latent performance and uncertainty	Variational inference, MCMC, Bayesian filtering, Laplace approximation
Acquisition	Select next configuration balancing exploration–exploitation	Expected improvement, UCB, Thompson sampling, constrained optimization
Update	Integrate new observations into posterior model	Online Bayesian optimization, amortized updates, meta-learning

By introducing random variables for latent performance, measurement processes, and hidden confounders, the tuning problem becomes a well-structured inference and decision-making task. This probabilistic framing enables the decomposition of uncertainty into two components: epistemic uncertainty, which reflects limited knowledge arising from finite data or unobserved regions of the parameter space, and aleatoric uncertainty, which captures intrinsic randomness in performance outcomes due to stochastic execution, workload variability, or environmental fluctuations. Treating these forms of uncertainty separately allows for more nuanced decision strategies that adapt exploration behavior to the confidence level of the current model. The ultimate objective becomes inference over latent performance functions and the selection of parameter configurations that optimize expected utility subject to reliability, safety, and service-level constraints.

In its general form, intelligent parameter tuning can be viewed as a sequential decision process. At each iteration, the system observes a context vector—comprising workload descriptors, input data statistics, and hardware state variables—selects a configuration vector according to its current beliefs, executes the system under these settings, and records stochastic feedback in the form of performance measurements [2]. The process repeats, continually refining the model of the underlying performance surface. This dynamic interaction defines a feedback loop consisting of four core stages: modeling, inference, acquisition, and update. Each stage plays a distinct yet interdependent role in the overall adaptation mechanism.

The modeling stage involves specifying a prior or surrogate model that captures assumptions about the functional relationship between configuration parameters, contexts, and outcomes. Structured priors such as Gaussian processes, neural network ensembles, or Bayesian linear models encode smoothness, correlation, and scaling properties that guide generalization from limited samples [3]. Additionally, explicit noise models define the variability and reliability of observed metrics, allowing for heteroscedastic or correlated measurement errors. Hierarchical structures can further capture dependencies across workloads or devices, enabling transfer of information between related tuning tasks while maintaining robustness to domain-specific variations.

Inference translates observed data into posterior beliefs about the performance function and its uncertainty. Depending on the complexity of the model, inference may be performed exactly—using analytic posterior updates—or approximately through techniques such as variational inference, expectation propagation, or Monte Carlo sampling. In online settings, where observations arrive sequentially, recursive updates analogous to Bayesian filtering allow the model to maintain real-time estimates of posterior means and variances without reprocessing historical data. The quality of inference directly influences the effectiveness of subsequent decisions, as inaccurate or overconfident posteriors can lead to premature exploitation or unsafe exploration. [4]

The acquisition stage defines the strategy for selecting the next configuration to evaluate, balancing the competing objectives of exploration and exploitation. Exploration seeks to reduce epistemic uncertainty by sampling configurations in regions where the model is uncertain, thereby improving overall knowledge of the performance landscape. Exploitation, by contrast, prioritizes configurations predicted to yield high expected performance based on current beliefs. Acquisition strategies are often formalized through utility functions such as expected improvement, upper confidence bound, or Thompson sampling, each reflecting a different trade-off between risk and reward. In many practical systems, the acquisition step must also consider operational constraints, ensuring that chosen configurations do not violate safety limits, throughput requirements, or hardware compatibility. [5]

The update stage closes the loop by incorporating new observations into the posterior model. Upon receiving fresh performance data, the system refines its estimates of latent quantities and recalibrates its uncertainty measures. This continual update allows the tuner to adapt to changing workloads, evolving hardware conditions, or nonstationary performance surfaces. Online Bayesian optimization frameworks implement these ideas efficiently, combining sequential inference with bounded computation to maintain responsiveness under strict latency constraints. When computational resources permit, ensemble or meta-learning approaches can further accelerate adaptation by leveraging prior experience across related tasks or hardware generations.

In real deployments, intelligent tuning operates under multiple layers of constraints [6]. Wall-clock time limits bound the duration of tuning cycles, particularly in interactive or latency-sensitive applications. Throughput and utilization targets restrict the number of permissible measurement runs, since each evaluation consumes resources that could otherwise serve production workloads. Moreover, the inference and acquisition computations themselves must respect hardware budgets; complex posterior updates or acquisition optimizations can become bottlenecks if not carefully managed. Consequently, practical implementations often employ surrogate approximations, low-rank updates, or amortized inference schemes that balance statistical fidelity with computational efficiency.

This probabilistic and iterative view of parameter tuning unifies ideas from Bayesian optimization, reinforcement learning, and adaptive control [7]. It transforms performance tuning from an ad hoc process of empirical search into a principled exercise in sequential decision-making under uncertainty. By explicitly maintaining beliefs about unknown performance landscapes and quantifying uncertainty in those beliefs, intelligent tuning systems can allocate exploration effort where it is most informative, exploit stable performance regions confidently, and respect operational constraints with statistical guarantees. As computing architectures and workloads continue to grow in complexity and heterogeneity, such probabilistic tuning frameworks provide the mathematical and conceptual foundation for self-optimizing systems capable of sustaining efficiency, reliability, and adaptability in uncertain and dynamic environments.

The following sections develop hierarchical Bayesian models for parameter–performance relationships, probabilistic graphical encodings for multi-component systems, sequential design via Gaussian-process and nonparametric priors, variational and amortized inference mechanisms for rapid adaptation, risk- and constraint-aware formulations for robustness, and implementation techniques for accelerators and distributed runtimes. Evaluation protocols are articulated in terms of calibration, regret, reliability, and reproducibility under controlled workload drifts.

2. Bayesian Foundations for Intelligent Parameter Tuning

Table 2. Uncertainty in Probabilistic Tuning

Uncertainty Type	Source / Meaning	Impact on Decision Strategy
Epistemic	Limited data or unobserved regions of parameter space	Drives exploration to improve model confidence
Aleatoric	Inherent stochasticity in execution or measurements	Encourages risk-aware or robust optimization
Hierarchical / Transfer	Shared structure across contexts or workloads	Enables cross-task generalization and knowledge sharing

Table 3. Bayesian Components for Performance Modeling

Component	Role / Example
Prior $p(f, \eta)$	Encodes smoothness, correlation, or noise structure assumptions
Likelihood $p(y \mid f(x,c), \eta)$	Connects observations to latent performance; can be Gaussian, Student-t, or mixture
Posterior $p(f, \eta \mid \mathcal{D}_t)$	Updated belief over performance surfaces given data
Hierarchical Structure	$\phi \sim p(\phi)$, $f \vert \phi \sim \mathcal{F}(\phi)$; transfers knowledge between contexts
Risk-Aware Objective	$x_{t+1} = \arg\min_x \mathbb{E}[\ell(y)]$ for robust or quantile-based tuning

Table 4. Probabilistic Graphical Models for Systemic Adaptation

Model Element	Purpose	Examples / Techniques
Factor Graphs	Encode dependencies among subsystems and metrics	Potentials ψ_α, message passing, EP / BP approximations
Latent Variables h	Capture hidden conditions (e.g., congestion, temperature)	Context-dependent priors $p(h \mid c)$
Multi-task Priors	Share information across components	Coregionalized GP: $\mathrm{cov}(f^{(k)}, f^{(k')}) = B_{kk'}k(\cdot)$
Sparse Priors	Promote selective or low-dimensional effects	Spike-and-slab, horseshoe, group sparsity

A baseline generative view treats the performance signal as a random function of a configuration vector and a context vector. Let $x \in \mathbb{R}^d$ denote the parameter setting to be tuned and $c \in \mathbb{R}^m$ the observed context summarizing workload and hardware state. A latent function f maps (x, c) to a performance scalar or vector. The measurement model records y, a noisy realization of $f(x, c)$ according to a likelihood $p(y \mid f(x,c), \eta)$ with nuisance parameters η. Priors over f and η complete the model [8]. Posterior inference computes the distribution of latent quantities given a dataset $\mathcal{D}_t = \{(x_i, c_i, y_i)\}_{i=1}^t$,

$$p(f, \eta \mid \mathcal{D}_t) \propto p(\eta)\, p(f) \prod_{i=1}^{t} p(y_i \mid f(x_i, c_i), \eta).$$

Decision-making uses the posterior predictive for a candidate (x, c), $p(y^\star \mid x, c, \mathcal{D}_t) = \int p(y^\star \mid f(x,c), \eta)\, p(f, \eta \mid \mathcal{D}_t)\, df\, d\eta$.

When parameters directly influence resource allocation or physical limits, it is useful to endow f with hierarchical structure that pools information across related contexts. Let c be associated with a latent embedding $z(c)$, and parameterize a family of local response surfaces with global hyperparameters ϕ. A hierarchical prior yields

$$\phi \sim p(\phi), \qquad f(\cdot,\cdot) \mid \phi \sim \mathcal{F}(\phi), \qquad y_i \mid f, \eta \sim p(y_i \mid f(x_i, c_i), \eta),$$

where $\mathcal{F}(\phi)$ denotes a stochastic process prior. Posterior concentration properties depend on design coverage in (x, c) and on the regularity encoded by $\mathcal{F}(\phi)$. The information geometry of the posterior can be summarized through the Fisher metric of the likelihood and the curvature of the log-prior, with the Laplace approximation delivering a Gaussian local surrogate [9]

$$p(\theta \mid \mathcal{D}_t) \approx \mathcal{N}\Big(\hat{\theta},\ \big[-\nabla_\theta^2 \log p(\theta, \mathcal{D}_t)\big|_{\hat{\theta}}\big]^{-1}\Big),$$

where θ collects finite-dimensional parameters after discretization or basis expansion. Such approximations support rapid posterior predictive updates when evaluating many candidate settings.

Risk-aware tuning expresses preferences via a utility $u(y)$ or a loss $\ell(y)$ and selects x by minimizing posterior expected loss or maximizing expected utility:

$$x_{t+1} \in \arg\min_{x} \mathbb{E}_{y \sim p(\cdot | x, c_{t+1}, \mathcal{D}_t)}[\ell(y)].$$

When ℓ encodes latency thresholds or tail-sensitive penalties, the objective targets quantiles rather than means. Heavy-tailed or heteroskedastic noise motivates mixture likelihoods and latent scale variables, e.g.,

$$y \mid f, \lambda \sim \mathcal{N}(f, \lambda^{-1}), \qquad \lambda \sim \text{Gamma}(a, b),$$

which imply Student-t marginal errors and automatically down-weight outliers in acquisition computations. [10]

3. Probabilistic Graphical Models for Systemic Adaptation

Many systems expose dozens of tunables across interacting components such as schedulers, memory allocators, communication layers, and device-specific kernels. Probabilistic graphical models represent conditional dependencies among intermediate performance indicators and provide a controlled factorization of the joint likelihood. Let $x = (x^{(1)}, \dots, x^{(K)})$ denote grouped tunables across K subsystems, and let $y = (y^{(1)}, \dots, y^{(L)})$ be observable metrics such as latency, throughput, and power. A factor graph with potentials ψ_α over cliques $\mathcal{C}_\alpha$ encodes

$$p(y \mid x, \theta) \propto \prod_{\alpha} \psi_\alpha(y_{\mathcal{C}_\alpha}, x_{\mathcal{C}_\alpha}; \theta_\alpha).$$

Latent variables h capture unobserved states, including queueing congestion and thermal conditions, with priors $p(h \mid c)$ that depend on contexts. Message passing produces local posteriors for h and intermediate performance variables,

$$m_{u \to v}(s_v) \propto \int \psi_{uv}(s_u, s_v) \prod_{w \in \text{ne}(u) \setminus v} m_{w \to u}(s_u)\, ds_u,$$

which deliver calibrated marginal beliefs for acquisition calculations [**?**]. When exact inference is intractable, loopy belief propagation or expectation propagation yields approximations, and their fixed points define moment-matching criteria that translate into differentiable surrogate objectives for tuning under partial observability.

Structured priors over component responses shrink estimates toward shared trends while allowing deviations. For example, with $f^{(k)}$ the effect of $x^{(k)}$ on a latent efficiency score under context c, a multi-task Gaussian process prior with coregionalization matrix B and kernel k gives

$$\text{cov}\big(f^{(k)}(x, c), f^{(k')}(x', c')\big) = B_{kk'}\, k\big((x, c), (x', c')\big),$$

automatically transferring information between related subsystems. Sparsity-promoting priors such as spike–and–slab on regression weights or horseshoe shrinkage on basis coefficients further regularize high-dimensional effects, enabling focused exploration in a large configuration space.

4. Sequential Design: Bayesian Optimization and Bandit Formulations

When evaluations are costly, Bayesian optimization places a prior on f and applies an acquisition rule that trades off exploration and exploitation. Under a Gaussian process prior with mean $\mu_t(\cdot)$ and covariance $k_t(\cdot, \cdot)$ conditioned on $\mathcal{D}_t$, the posterior predictive at a candidate $z = (x, c)$ has mean–variance pair $(\mu_t(z), \sigma_t^2(z))$. Common acquisition rules include upper confidence bound and expected improvement [11]. With confidence parameter $\beta_t > 0$,

$$x_{t+1} \in \arg\max_x \ \mu_t(x, c_{t+1}) + \sqrt{\beta_t}\,\sigma_t(x, c_{t+1}),$$

while expected improvement with incumbent $f^\star$ selects

$$x_{t+1} \in \arg\max_x \ \alpha_{\mathrm{EI}}(x, c_{t+1}), \qquad \alpha_{\mathrm{EI}}(z) = \mathbb{E}\big[(f(z) - f^\star)_+ \mid \mathcal{D}_t\big].$$

Batch and asynchronous settings require joint or penalized selections, for which determinantal point process regularizers and posterior covariance penalties reduce redundancy. Multi-fidelity variants introduce cheap surrogates f_ℓ at fidelities $\ell = 1, \ldots, L$ with cost w_ℓ and correlation priors across fidelities so that the acquisition optimizes value per unit cost,

$$z_{t+1} \in \arg\max_{(x,\ell)} \frac{\alpha_\ell(x)}{w_\ell}.$$

Bandit formulations view tuning as sequential allocation with partial feedback [12]. In contextual bandits with posterior $p(\theta \mid \mathcal{D}_t)$ over parametric models f_θ, Thompson sampling draws $\tilde{\theta} \sim p(\theta \mid \mathcal{D}_t)$ and chooses $x_{t+1} \in \arg\max_x f_{\tilde{\theta}}(x, c_{t+1})$, achieving exploration through posterior randomness. Under sub-Gaussian noise and appropriately regularized linear features $\phi(x, c)$, confidence sets

$$\mathcal{C}_t = \{\theta : \ \|\theta - \hat{\theta}_t\|_{V_t} \le \beta_t\}, \qquad V_t = \lambda I + \sum_{i=1}^{t} \phi_i \phi_i^\top$$

yield regret bounds of order $\tilde{O}(d\sqrt{t})$ with β_t scaling with $\log(1/\delta)$ and $\log\det(V_t)$. In nonparametric cases with kernelized predictors, information-theoretic complexity via the maximum information gain

$$\gamma_t = \max_{A \subset \mathcal{X} \times \mathcal{C}, |A| = t} I\big(y_A; f\big)$$

characterizes achievable regret rates. These results guide choices of kernels, features, and regularization in practice by translating smoothness assumptions into sampling budgets.

Constraints arising from safety and service-level objectives are expressed as chance constraints under auxiliary latent functions g_j that model violations. With tolerance levels ϵ_j, [13]

$$\max_x \ \mathbb{E}[f(x, c)] \quad \text{s.t.} \quad \mathbb{P}\big(g_j(x, c) \le 0 \mid \mathcal{D}_t\big) \ge 1 - \epsilon_j \ \ \forall j,$$

and posterior moment bounds or Gaussian approximations convert these constraints into tractable surrogates using quantiles $q_{1-\epsilon_j}$:

$$\mu_{g_j}(x, c) + q_{1-\epsilon_j}\,\sigma_{g_j}(x, c) \le 0.$$

5. Variational Inference, Amortization, and Meta-Priors

Low-latency adaptation is critical when measurements must be incorporated in near real time. Variational inference selects a tractable family $q_\phi(\theta)$ and maximizes an evidence lower bound,

$$\mathcal{L}(\phi) = \mathbb{E}_{q_\phi(\theta)}\big[\log p(\mathcal{D}_t, \theta)\big] - \mathbb{E}_{q_\phi(\theta)}\big[\log q_\phi(\theta)\big],$$

with gradients estimated by reparameterization or score-function estimators. For reparameterizable $q_\phi(\theta)$,

$$\nabla_\phi \mathcal{L}(\phi) = \mathbb{E}_\epsilon\big[\nabla_\theta \log p(\mathcal{D}_t, \theta)\, \nabla_\phi \theta(\epsilon, \phi) - \nabla_\phi \log q_\phi(\theta(\epsilon, \phi))\big].$$

Amortization replaces per-instance variational parameters with outputs of an inference network that maps sufficient statistics or raw data to the parameters of q_ϕ [14]. Given summary $s(\mathcal{D}_t)$, an encoder produces $q_\phi(\theta \mid s(\mathcal{D}_t))$, enabling constant-time posterior updates as t grows, with accuracy governed by the capacity and training distribution of the encoder.

Meta-priors capture invariances and cross-task regularities. Suppose tuning tasks are indexed by $r = 1, \ldots, R$ and share parameters through a hyperprior $p(\theta_r \mid \psi)$ with $\psi \sim p(\psi)$. The integrated posterior for a new task $r^\star$ leverages pooled evidence,

$$p(\theta_{r^\star} \mid \mathcal{D}_{1:R}, \mathcal{D}_{r^\star}) \propto \int p(\theta_{r^\star} \mid \psi) \prod_{r=1}^{R} p(\mathcal{D}_r \mid \psi)\, p(\psi)\, d\psi.$$

Variational families that factorize across tasks with a shared $q(\psi)$ and task-specific $q(\theta_r)$ admit coordinate updates,

$$\log q^\star(\psi) \leftarrow \mathbb{E}_{\{q(\theta_r)\}}\big[\log p(\psi) + \sum_r \log p(\theta_r \mid \psi)\big] + \text{const},$$

$$\log q^\star(\theta_r) \leftarrow \mathbb{E}_{q(\psi)}\big[\log p(\theta_r \mid \psi)\big] + \log p(\mathcal{D}_r \mid \theta_r) + \text{const},$$

producing adaptive warm starts for new contexts and thereby reducing exploration costs.

Risk-sensitive variational objectives introduce divergences beyond Kullback–Leibler to calibrate tail risk [15]. With a χ^2-divergence penalty scaled by $\tau > 0$,

$$\max_{q_\phi}\ \mathbb{E}_{q_\phi}[u(\theta)] - \tau\, D_{\chi^2}\big(q_\phi \,\|\, p(\theta \mid \mathcal{D})\big),$$

the optimum q_ϕ overweights high-utility regions while controlling deviation from the posterior, and the dual problem yields tempering schemes that interpolate between risk-neutral and robust selections.

6. Robustness under Distribution Shift and Safety-Constrained Inference

Operational workloads drift as data distributions, concurrency levels, and hardware conditions change. A shift variable s indexes environments with prior $p(s)$ and transition dynamics $p(s_{t+1} \mid s_t)$. Robust objectives consider worst-case or ambiguity-averaged risk over a set $\mathcal{P}$ of plausible environment distributions. For loss $\ell(\theta, s)$,

$$\min_\theta \sup_{q \in \mathcal{P}} \mathbb{E}_{s\sim q}[\ell(\theta, s)] \quad \text{subject to} \quad D(q \| p) \le \rho,$$

which admits a dual representation with risk-sensitive exponential tilting,

$$\min_\theta \frac{1}{\eta} \log \mathbb{E}_{s\sim p}\big[\exp\big(\eta\, \ell(\theta, s)\big)\big] + \frac{\rho}{\eta}.$$

When ℓ is induced by the predictive distribution over y, this criterion emphasizes tail events arising from congestion, interference, or thermal throttling [16]. Posterior predictive checks test calibration by comparing realized residuals to simulated draws, and miscalibration triggers tempering or likelihood reweighting.

Safety constraints may be enforced probabilistically through credible sets or deterministically through robust surrogates. Let $g(x, s)$ denote a violation margin; a chance constraint with level ϵ reads

$$\mathbb{P}\big(g(x, s) \le 0 \mid \mathcal{D}_t\big) \ge 1 - \epsilon.$$

Under Gaussian approximations, this converts to

$$\mu_g(x) + \Phi^{-1}(1-\epsilon)\,\sigma_g(x) \le 0,$$

and in non-Gaussian settings, monotone concentration inequalities bound violation probability using sub-exponential parameters (ν, b), [17]

$$\mathbb{P}\big(g(x) - \mathbb{E}[g(x)] \ge t\big) \le \exp\!\left(-\frac{t^2}{2\nu^2}\right) \text{ for } 0 \le t \le \frac{\nu^2}{b}.$$

Adaptive constraint learning couples performance and safety through a joint posterior over (f, g). A Lagrangian view with multiplier $\lambda \ge 0$ yields the acquisition surrogate

$$x_{t+1} \in \arg\max_x \; \mathbb{E}[f(x)] - \lambda\, \mathbb{E}\big[g(x)_+\big],$$

with λ updated from posterior estimates of constraint satisfaction rates to achieve target violation budgets over long horizons. In time-varying systems, a discount factor $\gamma \in (0,1)$ is applied to constraint penalties so that recent violations weigh more in the update.

7. System Realization: Distributed Runtimes, Accelerators, and Overheads

Deploying probabilistic tuning in practice requires accounting for computational cost, memory footprint, and communication in distributed settings. Consider a cluster of R workers observing local data $\mathcal{D}_r$ and evaluating configuration candidates in parallel. A consensus posterior over θ can be formed with quadratic penalties, [18]

$$\min_{\{\theta_r\}, z} \sum_{r=1}^{R} \Big(-\log p(\mathcal{D}_r \mid \theta_r) - \log p(\theta_r) \Big) + \frac{\rho}{2} \sum_{r=1}^{R} \|\theta_r - z\|_2^2,$$

with alternating updates

$$\theta_r^{k+1} \leftarrow \arg\min_{\theta_r} \; -\log p(\mathcal{D}_r \mid \theta_r) - \log p(\theta_r) + \frac{\rho}{2}\|\theta_r - z^k + u_r^k\|_2^2,$$

$$z^{k+1} \leftarrow \frac{1}{R}\sum_{r=1}^{R}\Big(\theta_r^{k+1} + u_r^k\Big), \qquad u_r^{k+1} \leftarrow u_r^k + \theta_r^{k+1} - z^{k+1},$$

implementing a proximal consensus scheme whose per-iteration complexity depends on local likelihood structure. Communication-efficient natural gradients precondition updates with approximate Fisher information computed from mini-batches, and curvature subsampling keeps memory constant per worker.

On accelerators, Gaussian process models with N observations face cubic scaling. Structured kernels with Kronecker or Toeplitz factorizations, inducing-point approximations, and random Fourier features reduce training and prediction costs [19]. Let $\Phi \in \mathbb{R}^{N \times M}$ denote a feature matrix with $M \ll N$; then

$$\mu_t(z) \approx \phi(z)^\top A^{-1} \Phi^\top y, \quad \sigma_t^2(z) \approx k(z,z) - \phi(z)^\top A^{-1} \phi(z), \quad A = \Phi^\top \Phi + \sigma^2 I_M,$$

with matrix solves accelerated by mixed precision on tensor cores. Streaming updates maintain Cholesky factors or Sherman–Morrison corrections for rank-one updates when new observations arrive.

Acquisition optimization over high-dimensional x integrates gradient-based local search and randomized restarts informed by posterior uncertainty. When x parameterizes discrete or mixed categorical choices such as kernel variants and tiling schemes, Gumbel–Softmax relaxations produce differentiable surrogates. Let $x \in \Delta^{K-1}$ represent a relaxed categorical vector; a sample is

$$x_k = \frac{\exp\left((\log \pi_k + g_k)/\tau\right)}{\sum_{j=1}^{K} \exp\left((\log \pi_j + g_j)/\tau\right)},$$

where g_k are i.i.d. Gumbel and τ controls smoothness [20]. Annealing τ during acquisition search yields near-discrete proposals that can be projected to feasible configurations.

8. Bayesian Equalization and Joint Optimization of Feed-Forward and Decision-Feedback Filters

High-speed serial links, optical interconnects, and wideband wireless front-ends commonly deploy feed-forward and decision-feedback equalization to mitigate intersymbol interference and colored noise introduced by channels with limited bandwidth, reflections, and dispersion. The configuration of the equalizers affects error rates, latency through buffering and retiming, and power draw due to tap activity and comparator thresholds. In operational environments with temperature drift, process variation, varying crosstalk, and changes in cable or backplane impedance, the optimal settings shift over time and across contexts. A probabilistic formulation provides a compact language for modeling uncertainty over the effective channel, the noise characteristics, the decision process within the slicer, and the interaction between feed-forward equalization and post-cursor cancellation [21]. The aim is to describe a coherent modeling and inference stack that selects tap patterns, step sizes, and loop gains in a way that integrates measurement budgets, safety constraints on error bursts, and hardware-induced discretizations such as coefficient word lengths and clipping.

A discrete-time baseband model expresses received symbols as a convolution between the transmitted sequence and an effective channel, with additive noise that may be colored and amplitude dependent. Let $x_t \in \mathcal{X}$ denote symbols, $h \in \mathbb{R}^{L_h}$ channel taps, and n_t noise. The observation r_t is

$$r_t = \sum_{k=0}^{L_h-1} h_k \, x_{t-k} + n_t,$$

with n_t modeled by a distribution capturing thermal noise, shot noise, and possible impulsive components. The feed-forward equalizer with taps $w \in \mathbb{R}^{L_f}$ forms a pre-decision output

$$u_t = \sum_{k=0}^{L_f-1} w_k \, r_{t-k},$$

and the decision-feedback term subtracts a weighted sum of past detected symbols using feedback taps $b \in \mathbb{R}^{L_b}$,

$$v_t = u_t - \sum_{k=1}^{L_b} b_k \, \hat{x}_{t-k},$$

where $\hat{x}_{t-k}$ are hard or soft decisions. The slicer produces $\hat{x}_t = \mathrm{decide}(v_t)$ according to thresholds or a probabilistic rule. In deterministic linear analysis, w and b are chosen to approximate the inverse of the channel polynomial up to noise amplification. Under uncertainty, the coefficients and even the decision process itself become random variables, and the coupling between w and b is modeled explicitly in the likelihood. [22]

A Bayesian specification places priors over the latent effective channel, equalizer coefficients, and noise hyperparameters, and couples them with regimes representing environmental or workload contexts. Let c_t denote a context vector summarizing temperature, supply voltage, lane coupling indicators, and characteristic impedances. A hierarchical prior allows the equalizer coefficients to depend smoothly on c_t through an embedding $\phi(c_t)$ and a latent mapping g,

$$w = g_w(\phi(c_t)) + \epsilon_w, \qquad b = g_b(\phi(c_t)) + \epsilon_b,$$

with ϵ_w and ϵ_b capturing idiosyncratic deviations [23]. Nonparametric priors over g_w and g_b support flexible dependence while pooling strength across similar contexts. The observation model acknowledges the discrete decision process by introducing a latent continuous variable for the slicer input and a probabilistic decision link. For binary PAM, one convenient choice uses a logistic or probit link so that

$$\mathbb{P}\big(\hat{x}_t = +1 \mid v_t\big) = \sigma(\alpha\, v_t + \delta),$$

with gain α and offset δ that themselves vary with comparator calibration and thermal drift. This link captures the soft-decision regime available in some receivers and allows likelihood-based learning from soft counts without saturating at extreme signal-to-noise ratios.

The likelihood under known transmitted symbols during a training phase accounts for the colored nature of noise and for nonlinearities in front-end analog components [24]. Let N training symbols be sent. With $r_{1:N}$ collected and decisions disabled, the probabilistic data model becomes

$$p(r_{1:N} \mid x_{1:N}, h, \theta_n) = \mathcal{N}\big(A x_{1:N}, \Sigma(\theta_n)\big),$$

where A is the Toeplitz convolution operator built from h and $\Sigma(\theta_n)$ encodes noise covariance with parameters θ_n capturing spectral coloring. When training with decision feedback engaged, the model involves the joint distribution over decisions and residuals. A data augmentation scheme introduces latent Gaussian variables that linearize the logistic link via Pólya–Gamma variables, turning the joint posterior into a conditionally Gaussian form amenable to Gibbs or variational inference. The augmented representation yields closed-form updates for the equalizer coefficients when conditioned on augmented variables, making it suitable for low-latency updates during short adaptation windows.

Posterior inference targets $p(w, b, h, \theta_n, \alpha, \delta \mid \mathcal{D})$ given observed training sequences, soft counts, and context traces. Exact inference is generally intractable, so factorized or structured variational distributions and Laplace approximations around modes are adopted [25]. A structured Gaussian variational family with block covariances aligned to (w, b) versus (h, θ_n) offers a balance between speed and calibration. The evidence lower bound takes the form

$$\mathcal{L} = \mathbb{E}_q\big[\log p(r_{1:N}, \hat{x}_{1:N}, w, b, h, \theta_n, \alpha, \delta \mid x_{1:N}, c_{1:N})\big] - \mathbb{E}_q\big[\log q(w, b, h, \theta_n, \alpha, \delta)\big],$$

with gradients computed by reparameterization and natural gradient updates that exploit the local curvature of the likelihood. The natural gradient step uses an approximate Fisher information assembled from mini-batches of symbol windows, reducing variance in updates when operating under strict wall-clock constraints.

Tap selection and coefficient shrinkage are central to reliable operation under limited sample sizes and nonstationary interference [26]. A spike-and-slab prior encourages sparse taps without committing to a fixed number of active coefficients. Let $\gamma_k \in \{0, 1\}$ be inclusion indicators with Bernoulli priors, and condition $w_k \mid \gamma_k$ on a Gaussian with variance modulated by γ_k. The joint prior reads

$$p(w, \gamma) = \prod_{k=0}^{L_f - 1} \big[(1 - \pi_k)\, \delta_0(\gamma_k) + \pi_k\, \delta_1(\gamma_k)\big] \cdot \mathcal{N}\big(w_k; 0, \gamma_k \sigma_w^2 + (1 - \gamma_k)\epsilon\big),$$

with ϵ a small variance that approximates a point mass at zero. A parallel construction applies to b. Variational updates on γ exploit local evidence from the marginal posterior over each coefficient and produce posterior inclusion probabilities that serve as uncertainty-aware tap masks under changing conditions. The decision-feedback path particularly benefits from shrinkage when error propagation risks increase, because the posterior adapts by temporarily deactivating taps that raise susceptibility to burst errors. [27]

Quantization and saturation of coefficients are intrinsic to hardware realizations. A common representation uses q-bit fixed-point for taps, and saturation occurs when updates exceed representable ranges. Treating quantization within the posterior requires discrete variables that are expensive to sample directly. A relaxation computes a continuous posterior over unconstrained coefficients and projects them through a differentiable approximation of quantization. The Gumbel–Softmax trick provides a surrogate for selecting discrete codebook values, mapping a latent continuous parameter to a near-one-hot vector over codebook entries [28]. If $C = \{c_1, \ldots, c_M\}$ is a tap codebook, a relaxed selection variable $z \in \Delta^{M-1}$ determines $w_k \approx \sum_m z_m c_m$ with

$$z_m = \frac{\exp\left((\log \pi_m + g_m)/\tau\right)}{\sum_{j=1}^{M} \exp\left((\log \pi_j + g_j)/\tau\right)},$$

where g_m are i.i.d. Gumbel variables and τ is a temperature that is annealed during training to sharpen selections. The resulting stochastic computational graph allows backpropagation through quantization surrogates within the variational objective, aligning the statistical optimization with implementable coefficients.

In nonstationary channels, coefficients drift in response to environmental state evolution and slow variations in the analog front-end. A state-space model for (w_t, b_t) driven by a latent state s_t captures gradual changes and abrupt shifts. A linear-Gaussian prior [29]

$$\begin{aligned} s_{t+1} &= F s_t + \xi_t, \qquad \xi_t \sim \mathcal{N}(0, Q), \\ \begin{bmatrix} w_{t+1} \\ b_{t+1} \end{bmatrix} &= G s_{t+1} + \zeta_t, \qquad \zeta_t \sim \mathcal{N}(0, R), \end{aligned}$$

combined with the augmented logistic observation model for decisions, yields a hybrid dynamical system. Extended Kalman filters, unscented filters, or particle filters propagate beliefs over (s_t, w_t, b_t) and deliver online updates with complexity controlled by the small latent state dimension. The predictive distribution over error counts on short blocks informs whether the loop should allocate symbols to training or remain in data mode, thereby enacting a budgeted exploration–exploitation trade-off.

Selecting training sequences is itself a design variable that modifies the Fisher information with respect to equalizer coefficients and link parameters. Given a candidate pilot pattern with power spectrum $S_{xx}(\omega)$ and an effective channel transfer function $H(\omega)$, the expected information about w under a linearized observation model is proportional to integrals of $|S_{xx}(\omega)H(\omega)|^2$ modulated by the noise spectrum. A constrained design problem allocates a fraction ρ of symbols to pilots while preserving throughput requirements, and chooses pilot spectral content within transmitter constraints [30]. A Lagrangian of the form

$$\max_{S_{xx}, \rho} \operatorname{Tr}\left(W \mathcal{I}(S_{xx}, \rho)\right) - \lambda \rho$$

trades information gain against training overhead, with W emphasizing directions in parameter space that most affect error bursts or energy consumption. The optimization is performed within regulator-imposed spectral masks and peak-to-average-power ratio constraints, and its solution guides firmware-level pilot schedulers to adjust training patterns under uncertainty.

Error propagation in decision-feedback equalization introduces dependence of present errors on past decisions, which complicates likelihood formulations if hard decisions are used. A soft-decision model treats $\hat{x}_{t-k}$ as latent variables with distributions conditioned on v_{t-k} and introduces an approximate posterior $q(\hat{x}_{1:N})$ that factorizes across time conditioned on a small set of summary statistics. Expectation-consistent inference aligns moments between the true posterior marginals and the approximating family by minimizing a sum of Kullback–Leibler divergences over factors. The resulting fixed-point equations compute soft estimates of past decisions and stabilize updates to b in regimes where the hard slicer would otherwise create feedback loops that degrade convergence. [31]

In receivers supporting multiple modulation orders and coding configurations, the equalizer must coordinate with the demapper and decoder to avoid redundant adaptations and misaligned objectives. A joint model includes a mapping from v_t to log-likelihood ratios for decoder input, with a parameterization that depends on equalizer settings and noise statistics. The posterior predictive distribution for frame error rate under a given code at a specified target rate approximates via a binomial–beta conjugate model on short test windows, with parameters updated from observed error counts. Specifically, if E errors are observed in N coded blocks during a probe interval, a $\mathrm{Beta}(a, b)$ prior on error probability yields a posterior

$$p(\theta \mid E, N) = \mathrm{Beta}(a + E,\, b + N - E),$$

and predictive tail probabilities for exceeding a service-level threshold θ_0 inform whether the adaptation should further reduce uncertainty or commit to the current configuration. This treatment allows frame-level criteria to drive coefficient updates rather than relying solely on symbol-level proxies. [32]

Robust noise modeling is critical in the presence of jitter, crosstalk bursts, and power-supply transients. A scale-mixture of Gaussians produces heavy-tailed marginals that retain computational tractability. For each sample n_t, introduce a latent precision λ_t and write

$$n_t \mid \lambda_t \sim \mathcal{N}(0, \lambda_t^{-1}), \qquad \lambda_t \sim \mathrm{Gamma}(a, b),$$

which integrates to a Student-t distribution for n_t. The augmented variables $\{\lambda_t\}$ yield closed-form updates for w and b conditional on λ_t in the linearized segments of the model and reduce the undue influence of rare, large deviations. Posterior predictive checks compare the empirical distribution of residuals with draws from the fitted model to detect underestimation of tail mass that would otherwise cause optimistic acquisition decisions.

When multiple lanes or parallel channels share substrates or packages, crosstalk creates dependencies between equalizers that can be exploited for data efficiency [33]. A multi-task prior correlates equalizer coefficients across lanes via a coregionalization matrix that encodes structural similarity. If $w^{(\ell)}$ denotes the FFE coefficients for lane ℓ and z indexes tap positions, a Gaussian process prior over z and ℓ with covariance

$$\mathrm{cov}\big(w_z^{(\ell)}, w_{z'}^{(\ell')}\big) = B_{\ell\ell'}\, k(z, z')$$

shares information across both tap index and lane identity. The posterior coupling reduces the number of pilot symbols required per lane to achieve a target calibration accuracy, and the learned B indicates which lanes co-adapt strongly due to proximity or shared clock trees. Similar constructions apply to DFE taps, allowing the system to localize feedback effort to lanes whose post-cursor structures resemble those for which reliable evidence exists.

Computational efficiency drives the feasibility of deploying Bayesian adaptation in firmware or microcontroller environments with strict time budgets. Low-rank features constructed from frequency-domain templates and sparse time-domain bases reduce the dimension of the optimization and enable constant-time updates per symbol block [34]. Let $\Phi \in \mathbb{R}^{L_f \times M}$ be a dictionary with $M \ll L_f$ atoms representing plausible tap profiles derived from channel measurements or electromagnetic simulations. Modeling $w = \Phi\theta$ places priors on the low-dimensional coefficients θ and constrains updates to a physically interpretable subspace. The posterior over θ is cheaper to maintain, and matrix operations such as $A^\top \Sigma^{-1} A$ in the variational bound can be preconditioned and cached at the level of $M \times M$ matrices. A related approach in the decision-feedback path represents b in a basis favoring exponentially decaying structures, reflecting typical post-cursor shapes in limited-bandwidth channels.

Acquisition strategies under uncertainty determine how aggressively to modify coefficients when observations are limited or changing. A posterior sampling rule draws a sample of (w, b) from the current posterior and applies it for a short interval, recording error counts and analog metrics such as eye height and jitter. The randomness produces exploration proportional to uncertainty without explicit optimism bonuses, while a simple risk-aware filter prevents excursions that would cause unacceptable peaks in error rates [35]. With a constraint on burst error probability per frame of at most ϵ, a chance-constrained optimizer chooses coefficient adjustments $\Delta w, \Delta b$ that maximize expected improvement while satisfying

$$\mathbb{P}(\mathrm{BER}(\Delta w, \Delta b) > \mathrm{BER}_{\max} \mid \mathcal{D}) \leq \epsilon,$$

approximated by Gaussian or bootstrap quantiles drawn from the predictive distribution over block error rates. The filter temporarily reduces step sizes or increases soft-decision weighting when predicted risk exceeds the budget, then resumes exploration after sufficient evidence accumulates.

Hardware–software co-design considerations surface in the distribution of computation between fast analog loops and slower digital adaptation. Analog equalizer stages such as continuous-time linear equalizers expose a small number of continuous knobs with local effects on signal spectra, while digital FFE and DFE provide higher-dimensional, more precise adjustments. A hierarchical control strategy models the analog stage as an outer loop that coarsely shapes the channel, with a prior over its effect on the digital equalizer's feasible region [36]. The posterior over the analog parameters influences the prior over (w, b), encouraging the digital equalizer to operate in regions where the analog front-end yields favorable conditioning. Calibration sequences alternate between outer and inner loops, updating beliefs about analog impact and refining the digital coefficients accordingly.

Evaluation of adaptation policies in situ relies on brief measurement windows that yield small-sample statistics. Rather than relying on asymptotic approximations, short-window inference uses exact or tight finite-sample bounds. For a window with N opportunities and E observed symbol errors, a Clopper–Pearson interval for the error probability derives from beta quantiles and informs whether current settings satisfy a bound with high probability [37]. If the upper bound at level $1 - \alpha$ remains above a threshold, the system maintains exploration or increases training density; if it falls below, the settings are retained. This method integrates seamlessly into the probabilistic framework by representing interval endpoints as functionals of the posterior and avoids underestimating uncertainty in short windows that occupy a small fraction of total runtime yet carry most of the information.

Comparisons with classical stochastic gradient adaptation such as least mean squares are informative when framed in terms of point estimators versus full posteriors. Stochastic gradient updates approximate a maximum-likelihood or regularized solution under specific cost functions, and their step-size schedules determine the bias–variance trade-off implicitly. In contrast, the Bayesian approach expresses uncertainty explicitly, allowing acquisition rules to tailor exploration to regions where posterior predictive variance is high and to freeze components where credible intervals are tight. Under nonstationary interference and intermittent burst noise, maintaining calibrated uncertainty on b can prevent aggressive cancellation that would amplify error propagation [38]. Empirical studies across diverse channels report improved generalization of tuned coefficients across contexts and faster convergence under structured priors that encode physically plausible responses. As one illustration among many, Dikhaminjia et al. (2021) report that a Bayesian machine learning strategy for jointly selecting hyperparameters and covariance functions across feed-forward and decision-feedback components achieves faster convergence than least mean squares and maintains performance in previously unseen operating conditions, aligning with the advantages predicted by the uncertainty-aware framework without requiring the details of any single benchmark or hardware stack [39].

9. Experimental Protocols and Evaluation Metrics

To evaluate intelligent tuning under uncertainty, protocols should separate design quality from raw search budget and should measure calibration and decision quality. Given a horizon T with measurement budget B, a run is defined by a random seed, an initial design set, and a fixed workload stream [40]. Posterior predictive calibration is summarized by coverage of credible intervals: for a nominal level α, define empirical coverage

$$\hat{\kappa}_\alpha = \frac{1}{M}\sum_{j=1}^{M} \mathbb{I}\{y_j \in I_\alpha(x_j)\},$$

and examine deviations $|\hat{\kappa}_\alpha - \alpha|$ across contexts. Decision quality is quantified by cumulative regret relative to the unknown optimum $f^\star(c_t)$,

$$\mathrm{Regret}(T) = \sum_{t=1}^{T} \big(f^\star(c_t) - f(x_t, c_t)\big),$$

estimated by high-fidelity emulation or by post-hoc validation measurements at the final incumbent configurations. Safety performance records empirical violation rates against budgets, with target levels stated explicitly as ϵ and realized frequencies constrained below $\epsilon + \delta$ with δ margins interpreted as statistical variability.

Reproducibility is supported by logging posterior states, acquisition choices, and random seeds. Posterior drift across repeated runs indicates instability; drift metrics can be expressed as Wasserstein distances between posterior samples at checkpoints, [41]

$$W_2\big(p_t(\theta), p'_t(\theta)\big) = \left(\inf_{\gamma \in \Gamma(p_t, p'_t)} \mathbb{E}_{(\theta,\theta')\sim\gamma} \|\theta - \theta'\|_2^2\right)^{1/2}.$$

When posterior differences correlate with hardware contention or temperature traces, diagnostics attribute sensitivity to specific subsystems. Multi-fidelity protocols record cost-normalized performance so that an algorithm achieving the same improvement with 40% less wall-clock is recognized appropriately. Finally, uncertainty-aware early stopping rules halt unpromising runs when credible intervals on achievable improvements shrink below a threshold Δ, converting tuning budgets to deterministic schedules.

10. Advanced Decision-Theoretic Criteria and Information Budgets

A decision-theoretic framing tracks the value of information under resource constraints. Let a denote an action that chooses both a parameter x and a measurement fidelity [42]. The expected value of information at time t for a candidate a is

$$\mathrm{EVI}_t(a) = \mathbb{E}_{y\sim p(\cdot|a,\mathcal{D}_t)}\left[\max_{x'} \mathbb{E}[u(y'_{x'}) \mid \mathcal{D}_t \cup \{(a,y)\}] - \max_{x'} \mathbb{E}[u(y'_{x'}) \mid \mathcal{D}_t]\right],$$

with $y'_{x'}$ a hypothetical outcome under x'. When utility is concave in predictive mean and penalizes variance, a first-order approximation yields acquisition rules proportional to mutual information between y and the maximizer $x^\star$, with tractable surrogates based on differential entropy reductions,

$$\mathrm{MI}(y; f(z^\star)) = H\big(f(z^\star) \mid \mathcal{D}_t\big) - \mathbb{E}_y\big[H\big(f(z^\star) \mid \mathcal{D}_t \cup \{(a,y)\}\big)\big].$$

In Gaussian process models, entropy reductions admit closed forms involving log determinants of conditional covariances, facilitating batch selection with diminishing-returns guarantees characterized by submodularity approximations.

Budgeted optimization formalizes a constraint on cumulative cost $C_T = \sum_{t=1}^{T} w(a_t)$ with w the cost of action a_t. A Lagrangian relaxation optimizes

$$\max_{\pi} \mathbb{E}\left[\sum_{t=1}^{T} u(y_t) - \lambda\, w(a_t)\right],$$

where policy π maps posteriors to actions and λ tunes the cost–benefit trade-off. Dual gradient ascent updates λ based on observed costs to meet long-run budget targets [43]. In nonstationary environments, discounting γ yields a Bellman recursion for the value of a belief state $b_t = p(\theta \mid \mathcal{D}_t)$,

$$V(b_t) = \max_{a} \mathbb{E}_{y \sim p(\cdot \mid a, b_t)}\big[u(y) + \gamma\, V\big(\mathcal{U}(b_t, a, y)\big)\big],$$

where $\mathcal{U}$ denotes the Bayesian update. Particle approximations represent b_t by samples with weights updated via likelihoods, and value function regression provides a parametric approximation for V using invariants of the posterior such as predictive moments.

11. Uncertainty Calibration, Diagnostics, and Failure Modes

Calibrated uncertainty is essential for safe and efficient exploration. Posterior predictive checks simulate replicated data $\tilde{y}$ under the fitted model and compare discrepancy statistics $T(y)$ with $T(\tilde{y})$. A tail-area measure

$$p_{\mathrm{ppc}} = \mathbb{P}\big(T(\tilde{y}) \geq T(y) \mid \mathcal{D}\big)$$

indicates misfit when concentrated near 0 or 1. For tuning, discrepancy functions target tail metrics such as the 0.95-quantile of latency or the fraction of requests exceeding a threshold. Heteroskedasticity is handled by modeling log-variance as another latent function $h(x, c)$ with a link $\sigma^2(x, c) = \exp(h(x, c))$ [44]. The joint posterior for (f, h) produces predictive distributions with asymmetric tails, changing acquisition decisions that penalize uncertainty in variance explicitly.

Failure modes include posterior collapse in overconfident variational approximations, acquisition myopia that overexploits transient improvements, and confounding from unobserved covariates such as background network traffic. Diagnostic interventions add control variables, introduce randomized exploration to break correlations, and employ conservative initialization priors with larger length-scales or heavier tails. A tempered posterior

$$p_\tau(\theta \mid \mathcal{D}) \propto p(\theta)^\tau\, p(\mathcal{D} \mid \theta)^\tau$$

with $\tau \in (0, 1)$ broadens credible sets and has the effect of delaying aggressive exploitation until sufficient evidence accumulates [45]. In streaming contexts, forgetting factors apply exponential decay to older data,

$$\log p(\mathcal{D}_t \mid \theta) \leftarrow \sum_{i=1}^{t} \gamma^{t-i} \log p(y_i \mid x_i, c_i, \theta), \quad \gamma \in (0, 1),$$

which stabilizes adaptation under gradual drift.

12. Data Efficiency, Multi-Task Transfer, and Multi-Objective Trade-offs

Data efficiency improves through transfer across related tasks and through structured priors that capture invariances. Suppose tasks share a low-dimensional latent representation $r \in \mathbb{R}^q$ with $q \ll d$ so that $f(x, c) \approx \tilde{f}(U^\top x, V^\top c)$ for projection matrices U and V. Bayesian matrix factorization on sensitivity matrices identifies these subspaces with uncertainty,

$$U \sim \mathcal{N}(0, \sigma_U^2 I), \;\; V \sim \mathcal{N}(0, \sigma_V^2 I), \;\; \tilde{f} \sim \mathrm{GP}(0, k),$$

reducing the effective dimension seen by the acquisition optimizer. In multi-objective tuning with objectives $f^{(1)}, \ldots, f^{(J)}$, posterior samples induce random Pareto fronts. Scalarization with random weights w and temperature τ selects [46]

$$x_{t+1} \in \arg\max_{x} \sum_{j=1}^{J} w_j \, s_\tau\big(f^{(j)}(x)\big), \quad s_\tau(u) = \frac{1}{\tau} \log\big(1 + \exp(\tau u)\big),$$

while maintaining coverage of the Pareto set by varying w across iterations. Posterior entropy over the Pareto set can be reduced directly by acquisitions that target uncertainty at the front rather than over the full space.

Transfer learning via empirical Bayes fits hyperparameters $\hat{\phi}$ on historical logs and initializes new campaigns at $\hat{\phi}$ while maintaining uncertainty through a hyperprior around $\hat{\phi}$. The hyperposterior variance controls how aggressively structure is transferred; large variance defaults to weak sharing, while small variance enforces strong pooling. When conflicts arise between historical and current environments, Bayesian model averaging over alternative kernels or likelihoods avoids premature commitment.

13. Conclusion

The preceding development has articulated a Bayesian and probabilistic perspective on intelligent parameter tuning, focusing on the integration of uncertainty modeling, inference, and decision-making in dynamic computing environments [47]. By framing the tuning process as inference over latent performance functions conditioned on context and configuration, the formulation captures both epistemic uncertainty due to limited observations and aleatoric variability inherent in stochastic system behavior. This generative modeling approach treats the performance signal itself as a random process governed by hierarchical dependencies among tunable parameters, hardware conditions, and workload descriptors. The result is a coherent framework in which performance optimization is no longer a purely empirical search, but a structured inference problem that accommodates nonstationarity, measurement noise, and system-level constraints.

The modeling constructs rely on hierarchical and multi-task architectures to share statistical strength across related tuning scenarios. Parameters that govern similar kernels, workloads, or devices can be jointly represented through shared priors or latent factors, allowing information gathered in one setting to inform posterior estimates in another. This sharing not only accelerates adaptation but also provides regularization that mitigates overfitting to transient noise or idiosyncratic workloads [48]. Risk-sensitive objectives further refine the decision process by explicitly penalizing uncertainty or constraint violations, leading to conservative exploration in safety-critical or resource-limited environments. Through this probabilistic formalism, performance tuning gains a degree of interpretability and robustness that deterministic heuristics typically lack.

Approximate inference plays a central role in enabling real-time operation under bounded computational budgets. Exact Bayesian updates are intractable for high-dimensional parameter spaces or complex likelihood models; thus, practical systems employ variational approximations, expectation propagation, or sampling-based techniques to maintain posterior estimates. Variational inference provides fast, differentiable updates through optimization of an evidence lower bound, while Monte Carlo and sequential sampling schemes offer flexibility in representing multimodal or heavy-tailed posteriors [49]. These methods transform the challenge of exact inference into one of approximation design—choosing parameterizations that balance fidelity, stability, and computational efficiency. The online nature of tuning further demands streaming-friendly algorithms capable of updating beliefs incrementally as new measurements arrive, without the need for full retraining.

Acquisition strategies derived from information-theoretic principles connect the exploration–exploitation trade-off to measures of posterior complexity and expected information gain. Criteria such as expected improvement, entropy search, and knowledge gradient quantify how much each potential configuration is expected to reduce predictive uncertainty or improve expected performance. By integrating these measures with safety and cost constraints, acquisition functions can explicitly account for resource budgets, latency ceilings, and energy consumption. In distributed or accelerator-based environments, these constraints often dominate algorithmic design, as the cost of evaluating candidate configurations can exceed that of inference itself [50]. Consequently, acquisition mechanisms are implemented with cost-aware adjustments, prioritizing configurations that offer favorable information-to-cost ratios.

Implementation on accelerators and distributed runtimes introduces additional layers of complexity related to data movement, synchronization, and hardware heterogeneity. Probabilistic tuning frameworks must map inference and acquisition computations onto parallel hardware efficiently, minimizing memory traffic and communication overhead while maintaining numerical stability. Techniques such as model partitioning, asynchronous updates, and mini-batch sampling allow scalable deployment across clusters and multi-device configurations. Streaming operation further constrains design: inference updates must proceed in lockstep with ongoing computation or data processing, requiring low-latency communication between the tuner and the executing system [51]. The integration of tuning logic into runtime schedulers or compiler feedback loops exemplifies how probabilistic inference can be embedded directly into system control flows.

Evaluation protocols for uncertainty-aware tuning methods emphasize calibration, regret, and reliability as key performance metrics. Calibration measures the alignment between predicted uncertainties and observed errors, ensuring that posterior confidence intervals reflect actual variability in performance. Regret quantifies the cumulative loss relative to an oracle that always selects the best configuration, serving as a proxy for adaptation efficiency. Reliability assesses the system's ability to maintain consistent performance across varying workloads, hardware conditions, and temporal shifts. Together, these metrics provide a comprehensive picture of how well the tuning process manages uncertainty and adapts to nonstationary conditions [52]. Benchmarking under changing workloads and heterogeneous platforms further validates whether probabilistic tuning methods generalize beyond controlled environments.

Throughout this development, the presentation has maintained a neutral and integrative stance, emphasizing composability rather than prescriptive methodology. The Bayesian and probabilistic principles outlined here do not dictate a single algorithmic pathway but rather define a design space in which different modeling choices, inference techniques, and acquisition rules can be combined according to application constraints. Whether implemented as a lightweight online optimizer for embedded devices or as a distributed learning component for large-scale systems, the same foundational constructs—latent variable modeling, posterior updating, and uncertainty-aware decision-making—remain central. This modularity makes the framework adaptable to emerging challenges in next-generation computational systems, where dynamic workloads, heterogeneous resources, and safety-critical requirements converge. Ultimately, the probabilistic formulation offers not just a toolkit for parameter tuning, but a conceptual foundation for adaptive, introspective computing systems capable of optimizing themselves under uncertainty. [53]

References

1. Kumar, G.K.; Narayanam, B. High Speed FIR Filter Using Radix-2r Multiplier and Its Application for Denoising EOG Signal. *Journal of Circuits, Systems and Computers* **2021**, *30*. https://doi.org/10.1142/s0218126621502376.
2. Casseres, D.A.M.; Valencia, G.E.; Forero, J.D. In-Cylinder Pressure Signal Processing Applied to the Diagnosis of Combustion in Low Displacement Diesel Engine. *International Journal on Communications Antenna and Propagation (IRECAP)* **2020**, *10*, 137–144. https://doi.org/10.15866/irecap.v10i2.18640.
3. Wang, M.; Zhang, S.; Liu, Z.; Zhang, X.; He, Y.; Ma, Y.; Zhang, Y.; Zhang, Z.; Liu, Y. High-frequency characterization of high-speed modulators and photodetectors in a link with low-speed photonic sampling. *Journal of Semiconductors* **2021**, *42*, 042303–. https://doi.org/10.1088/1674-4926/42/4/042303.

4. Yee, C.F.; Isa, M.M.; Al-Hadi, A.A.; Arshad, M.K.M. Techniques of impedance matching for minimal PCB channel loss at 40 GBPS signal transmission. *Circuit World* **2019**, *45*, 132–140. https://doi.org/10.1108/cw-01-2019-0004.
5. Huhn, F.; Wentzel, A.; Heinrich, W. A new modulator for digital RF power amplifiers utilizing a wavetable approach. *International Journal of Microwave and Wireless Technologies* **2017**, *9*, 1251–1260. https://doi.org/10.1017/s1759078717000770.
6. Guan, K.; Keusgen, W.; Fan, W.; Briso, C.; Sun, B. Guest Editorial: Antennas and propagation at millimetre, sub-millimetre wave and terahertz bands. *IET Microwaves, Antennas & Propagation* **2023**, *17*, 415–418. https://doi.org/10.1049/mia2.12374.
7. Vardhana, M.; Bhat, A.K. VLSI Implementation and Software Co-Simulation of Digital Image Watermarking with Increased Security. *International Journal of Sensors, Wireless Communications and Control* **2020**, *10*, 634–638. https://doi.org/10.2174/2210327909666190319150314.
8. fast terahertz detection. *Electronics Letters* **2017**, *53*, 55–55. https://doi.org/10.1049/el.2016.4592.
9. Jiang, D.; Wang, H.; Wang, R.; Liu, Z.; Liu, Y.; Lin, Z.; Wan, W. Design and experiments of isolated gate driver using discrete devices for silicon carbide MOSFET. *IET Power Electronics* **2022**, *16*, 118–127. https://doi.org/10.1049/pel2.12368.
10. Isabona, J.; Srivastava, V.M. Radio Channel Propagation Characterization and Link Reliability Estimation in Shadowed Suburban Macrocells. *International Journal on Communications Antenna and Propagation (IRECAP)* **2017**, *7*, 57–63. https://doi.org/10.15866/irecap.v7i1.10343.
11. Fujita, K.; Hasegawa, M. Position sensorless starting method for IPMSMs by all-pass filters. *Electrical Engineering in Japan* **2021**, *214*. https://doi.org/10.1002/eej.23348.
12. He, F.; Zhang, J. Research on time–frequency characteristics of contact force based on wavelet transform. *IET Electrical Systems in Transportation* **2022**, *12*, 287–297. https://doi.org/10.1049/els2.12053.
13. Kota, P.N.; Gaikwad, A.N. Analysis of Detection Methods for Space Frequency Block Code MIMO-OFDM System in Fast Fading Channels. *International Journal on Communications Antenna and Propagation (IRECAP)* **2017**, *7*, 635–640. https://doi.org/10.15866/irecap.v7i7.13613.
14. Abuelhaija, A.; Saleh, G. A Pi-Shaped Compact Dual Tuned 1H/23Na Microstripline-Based Switch for 7-Tesla MRI. *International Journal on Communications Antenna and Propagation (IRECAP)* **2021**, *11*, 57–64. https://doi.org/10.15866/irecap.v11i1.20302.
15. Liu, J.; Zhou, F.; Zhao, C.; Wang, Z. Mechanism analysis and suppression strategy research on permanent magnet synchronous generator wind turbine torsional vibration. *ISA transactions* **2019**, *92*, 118–133. https://doi.org/10.1016/j.isatra.2019.02.006.
16. Ahifar, A.; Noee, A.R.; Rahmani, Z. Terminal synergetic design of a nonlinear robot manipulator in the presence of disturbances. *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering* **2018**, *37*, 208–223. https://doi.org/10.1108/compel-11-2016-0490.
17. Samanth, R.; Nayak, S.G. An Efficient Two-phase Clocked Sequential Multiply -Accumulator Unit for Image Blurring. *International Journal of Electronics and Telecommunications* **2021**, pp. 307–313. https://doi.org/10.24425/ijet.2022.139883.
18. Bouaouina, M.S.; Lebbal, M.R.; Bouchemat, M.; Bouchemat, T. Enhanced prompt trigger optical switching using nonlinear photonic crystal ring resonator for application on all-optical AND/NAND and OR/NOR logic function. *Frequenz* **2021**, *76*, 145–155. https://doi.org/10.1515/freq-2021-0110.
19. Zamiri, F.; Nabavi, A. A modified Fresnel-based algorithm for 3D microwave imaging of metal objects. *International Journal of Microwave and Wireless Technologies* **2018**, *11*, 313–325. https://doi.org/10.1017/s175907871800123x.
20. Singh, H.; Sheetal, A. Suitability of FBG for Gain Flatness of 64×10Gbps DWDM System Using Hybrid (EDFA+YDFA) Optical Amplifier in C+L Band up to 50GHz (0.4nm) Channel Spacing. *Journal of Optical Communications* **2019**, *44*, 517–525. https://doi.org/10.1515/joc-2019-0134.
21. Poluektov, A.; Pinomaa, A.; Ahola, J.; Kosonen, A. Motor cable communication concept based on DSSS for inverter-fed electric drives. *IET Electric Power Applications* **2019**, *13*, 786–794. https://doi.org/10.1049/iet-epa.2019.0093.
22. Sun, W.; Vertatschitsch, L.; Sahr, J.D. Development design and demonstration of very high-speed multi-antenna digital receiver. *IET Radar, Sonar & Navigation* **2018**, *12*, 523–529. https://doi.org/10.1049/iet-rsn.2017.0383.
23. Hussain, M.; Ahmed, R.; Cheema, H.M. Segmented Radon Fourier Transform for Long-Time Coherent Radars. *IEEE Sensors Journal* **2023**, *23*, 9582–9594. https://doi.org/10.1109/jsen.2023.3260024.

24. Xiao, Y.; Dai, L.; Yu, X.; Zhou, Y.; Zhao, Z. Generation of a synthetic aperture radar deception jamming signal based on a deep echo inversion network. *IET Radar, Sonar & Navigation* **2023**, *17*, 801–812. https://doi.org/10.1049/rsn2.12379.
25. Lv, Z.; Lloret, J.; Song, H. 5G for mobile augmented reality. *International Journal of Communication Systems* **2022**, *35*. https://doi.org/10.1002/dac.5107.
26. Han, L.; Li, S.; Liu, X.; Jia'an, G. Online burr video denoising by learning sparsifying transform. *IET Image Processing* **2019**, *13*, 1138–1145. https://doi.org/10.1049/iet-ipr.2018.6243.
27. Filatov, A.; Hawkins, L. Novel Combination 3DOF Radial/Axial Eddy-Current Position Sensor. *IEEE Sensors Journal* **2022**, *22*, 23743–23755. https://doi.org/10.1109/jsen.2022.3220556.
28. Lee, J.Y.; null LeeDongYeol.; Jeong, M.S.; Dongwoo, K. Research on Hall Sensor Fault Diagnosis and Compensation Method to Improve Motor Control Reliability. *Journal of Magnetics* **2018**, *23*, 648–653. https://doi.org/10.4283/jmag.2018.23.4.648.
29. Azhdari, H.; Javahernia, S. An all optical photonic crystal half adder suitable for optical processing applications. *Journal of Optical Communications* **2021**, *44*, s321–s326. https://doi.org/10.1515/joc-2020-0261.
30. Amiri, I.S.; Rashed, A.N.Z.; Mohammed, A.E.N.A.; El-din, E.S.; Yupapin, P.P. Spatial Continuous Wave Laser and Spatiotemporal VCSEL for High-Speed Long Haul Optical Wireless Communication Channels. *Journal of Optical Communications* **2019**, *44*, 43–51. https://doi.org/10.1515/joc-2019-0061.
31. Ioannopoulos, G.A.; Anagnostou, D.E.; Chryssomallis, M.T. Evaluating the Effect of Small Number of Snapshots and Signal-to-Noise-Ratio on the Efficiency of MUSIC Estimations. *IET Microwaves, Antennas & Propagation* **2017**, *11*, 755–761. https://doi.org/10.1049/iet-map.2016.0308.
32. John, T.M.; Chacko, S. Efficient VLSI architecture for FIR filter design using modified differential evolution ant colony optimization algorithm. *Circuit World* **2020**, *47*, 243–251. https://doi.org/10.1108/cw-05-2020-0097.
33. Habbab, M.; Hazzab, A.; Sicard, P. Real Time Implementation of Fuzzy Adaptive PI-sliding Mode Controller for Induction Machine Control. *International Journal of Electrical and Computer Engineering (IJECE)* **2018**, *8*, 2883–2893. https://doi.org/10.11591/ijece.v8i5.pp2883-2893.
34. Yin, J.; Tian, M.; Duan, X.; Zhang, Y. Effect of External Perturbation and System Parameters on Optical Secure Communication Models. *Journal of Optical Communications* **2019**, *42*, 141–147. https://doi.org/10.1515/joc-2018-0048.
35. F, F.V.; Zuev, A.; Zhirabok, A.; Protsenko, A.A. Development of Accommodation System for Faults in Thrusters of Underwater Robots. *Mekhatronika, Avtomatizatsiya, Upravlenie* **2021**, *22*, 262–271. https://doi.org/10.17587/mau.22.262-271.
36. Hwang, B.C.; Min, B.G.; Kim, D.J.; Yoon, C.H. Implementation and Performance Evaluation of TSMB Bus System with Time Synchronized Multiple Bi-Directional TMDS Data Transmissions for Intelligent Electronic Devices. *The transactions of The Korean Institute of Electrical Engineers* **2023**, *72*, 37–45. https://doi.org/10.5370/kiee.2023.72.1.37.
37. Xie, W.; Li, Y.; Zhou, W. Hyperspectral image filtering with adaptive manifold for classification. *Journal of Electronic Imaging* **2017**, *26*, 033025–033025. https://doi.org/10.1117/1.jei.26.3.033025.
38. Oluwajobi, F.I.; Dong-Nhat, N.; Malekmohammadi, A. Novel Manchester-Based Multilevel Signaling for High-Speed Optical Communication Systems. *Journal of Optical Communications* **2019**, *42*, 655–667. https://doi.org/10.1515/joc-2018-0139.
39. Dikhaminjia, N.; Tsintsadze, G.; Kiguradze, Z.; He, J.; Tsiklauri, M.; Drewniak, J.; Chada, A.; Mutnury, B. Optimization of joint equalization of high-speed signals using Bayesian machine learning. In Proceedings of the 2021 IEEE International Joint EMC/SI/PI and EMC Europe Symposium. IEEE, 2021, pp. 48–52.
40. Simon, J.; Prabakaran, N. Optimal wavelets with Binary Phase Shift Keying modulator utilized simulation analysis of 4G/5G OFDM systems. *International Journal of Communication Systems* **2021**, *34*. https://doi.org/10.1002/dac.4775.
41. Zare, S.; Ayati, M. Simultaneous fault diagnosis of wind turbine using multichannel convolutional neural networks. *ISA transactions* **2020**, *108*, 230–239. https://doi.org/10.1016/j.isatra.2020.08.021.
42. Sathyanarayana, S.G.; Wang, Z.; Sun, N.; Ning, B.; Hu, S.; Hossack, J.A. Recovery of blood flow from undersampled photoacoustic microscopy data using sparse modeling. *IEEE transactions on medical imaging* **2021**, *41*, 1–1. https://doi.org/10.1109/tmi.2021.3104521.

43. Ibrahim, A.; Ismail, T.; Elsayed, K.M.F.; Darweesh, M.S. Odd clipping optical orthogonal frequency division multiplexing for VLC system. *International Journal of Communication Systems* **2019**, *32*. https://doi.org/10.1002/dac.4150.
44. Kaur, H.; Bhamrah, M.S.; Kaur, B. A comprehensive study on radio over fiber systems: present evaluations and future challenges. *Journal of Optical Communications* **2022**, *45*, 829–843. https://doi.org/10.1515/joc-2022-0040.
45. Wang, Y.; Liu, G.; Han, F.; Qu, H.; Chen, Q. SS-OFDM: an enhanced multicarrier transmission scheme based on variable granularity spectrum allocation for 5G. *IET Communications* **2018**, *12*, 2418–2428. https://doi.org/10.1049/iet-com.2018.5518.
46. Kumar, A.; Anjaneyulu, L. Design and Development of Redeployable Underwater Data Communication Link for Defence Application. *Defence Science Journal* **2017**, *68*, 111–116. https://doi.org/10.14429/dsj.68.11837.
47. Freag, H.; Hassan, E.S.; El-Dolil, S.A.; Dessouky, M.I. New Hybrid PAPR Reduction Techniques for OFDM-Based Visible Light Communication Systems. *Journal of Optical Communications* **2018**, *39*, 427–435. https://doi.org/10.1515/joc-2017-0002.
48. Liu, J.; Zhang, Y.; Dong, X. Linear-FM random radar waveform compressed by dechirping method. *IET Radar, Sonar & Navigation* **2019**, *13*, 1107–1115. https://doi.org/10.1049/iet-rsn.2018.5432.
49. Wang, X.; Zhang, B.; Zhao, H.; Su, Y.; Muhammad, A.; Guo, D.; Jin, Z. A 75 GHz regenerative dynamic frequency divider with active transformer using InGaAs/InP HBT technology. *Journal of Semiconductors* **2017**, *38*, 085001–. https://doi.org/10.1088/1674-4926/38/8/085001.
50. Saad, M.; Maali, A.; Azzaz, M.S.; Benssalah, M. An efficient FPGA-based implementation of UWB radar system for through-wall imaging. *International Journal of Communication Systems* **2023**, *36*. https://doi.org/10.1002/dac.5510.
51. Bhardwaj, S.; Raghuraman, S.; Adapa, B.; Acharyya, A. Vector Cross Product and Coordinate Rotation Based nD Hybrid FastICA. *Journal of Low Power Electronics* **2018**, *14*, 351–364. https://doi.org/10.1166/jolpe.2018.1551.
52. Eid, M.M.A.; Rashed, A.N.Z.; El-Meadawy, S.; Ahmed, K. RETRACTED ARTICLE: Simulation study of signal gain optimization based on hybrid composition techniques for high-speed optically dense multiplexed systems. *Journal of Optical Communications* **2020**, *45*, s149–s155. https://doi.org/10.1515/joc-2020-0150.
53. Zhou, Y.; Sharma, A.; Shabaz, M. Feature recognition of state signal of electromechanical integration railway turnout over health parameters using CMOS area array technology. *The Journal of Engineering* **2022**, *2022*, 715–724. https://doi.org/10.1049/tje2.12154.