

POMDPs for Autonomous Science Exploration

Daniel Guirguis, Nathan Wallace, Hanna Kurniawati, and Salah Sukkarieh

Abstract—Autonomous exploration missions require decision-making under sensor uncertainty and computational constraints, yet integrating scientific representations into POMDP planning has remained intractable due to high-dimensional observation spaces. Information-theoretic planners overcome this by assuming deterministic observations, sacrificing the principled uncertainty quantification that POMDPs provide. We introduce the Science Hypothesis Map POMDP (SHM-POMDP), which makes science-driven belief-space planning more tractable by branching on inferred physical properties rather than raw sensor data. This preserves full sensor information through learned observation models while enabling the planner to reason jointly about navigation and scientific properties under uncertainty. On an extended RockSample domain with 50-dimensional observations, SHM-POMDP achieves 18.6% higher rewards and 32.9% reduced computation time per step than continuous-observation baselines. On realistic geologic exploration using Cuprite hyperspectral data, SHM-POMDP achieves $2.5\times$ higher information gain than the best information-theoretic baseline by maintaining beliefs and replanning adaptively—reaching 80% of oracle performance using only uniform priors. These results demonstrate that integrating hierarchical probabilistic models into belief-space planning enables tractable, principled autonomous science that outperforms both traditional POMDP methods and science-aware information-theoretic approaches.

I. INTRODUCTION

Modern planetary robotic exploration relies heavily on scientists specifying waypoints based on expert knowledge and expectations about where to gather mission-critical information [1]. As measurements are collected, scientists reinterpret data with growing contextual knowledge and frequently re-plan [2]. However, in many exploration scenarios with low bandwidth and high latency communications, pre-planned waypoint navigation fails to capitalise on in-situ observations and unexpected discoveries, necessitating greater onboard autonomy [3].

This work presents a framework to overcome communication bottlenecks in robotic exploration by shifting from route-based command requiring human-in-the-loop replanning to guidance driven by an evolving model of scientific belief, enabling the robot to replan autonomously without ground-station intervention. The proposed approach accounts for uncertainty in observations, information gathering, and navigation under partial observability. The Partially Observable Markov Decision Process (POMDP) framework provides a principled approach for decision-making under uncertainty by maintaining belief distributions over states and optimising expected cumulative rewards [4]. However, solving POMDPs for scientific exploration becomes computationally challenging as multimodal sensors such as spectrometers and cameras produce high-dimensional observation vectors, and

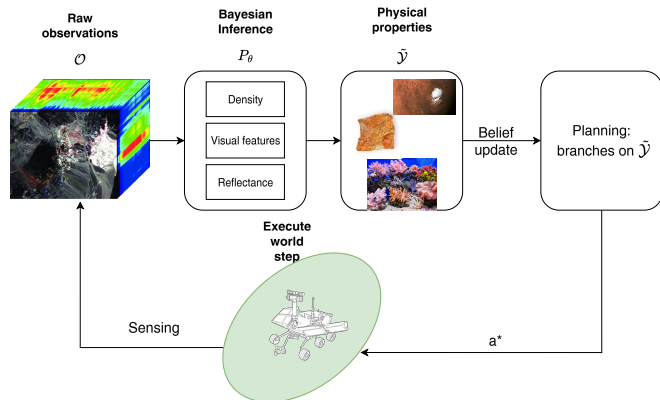


Fig. 1: Overview of the SHM-POMDP framework: SHM-POMDP makes POMDP planning tractable by planning on inferred physical properties $\tilde{y}$ rather than raw observations, abstracted through the SHM model P_θ to update the robot’s belief for autonomous exploration.

these sensors often exhibit complex correlations that further complicate direct planning. Prior work on Science Hypothesis Maps (SHM) addressed the scientific reasoning gap but assumed deterministic observations, while POMDP solvers handle uncertainty but struggle with high-dimensional observation spaces. This paper introduces the SHM-POMDP, a framework that integrates the SHM’s hierarchical probabilistic structure into POMDP planning. This integration provides a principled approach for scientific exploration under observation uncertainty while improving tractability through observation abstraction, enabling autonomous exploration within the computational budget of onboard flight processors (Fig. 1).

Scientists conceptualise exploration goals in terms of abstract hypotheses while robots must operate on raw sensor measurements such as spectral data, imagery, and other high-dimensional observations [5]. This gap between scientific reasoning and robotic sensing makes it difficult for autonomous systems to make decisions that are both scientifically meaningful and computationally feasible. However, it is possible to construct simpler hierarchical probabilistic models relating scientific reasoning to measurable data. For example, NASA missions formalise this relationship through the Science Traceability Matrix (STM), which employs a tripartite division into investigation objectives, physical properties, and instrument measurements [1]. For instance, an investigation objective such as “assess past habitability” traces to latent physical properties like organic molecule concentration, which are not directly observed but inferred from instrument measurements such as mass spectrometry

readings [6].

We exploit this hierarchical structure in SHM-POMDP by learning the mapping between robotic raw sensor measurements and latent scientific properties of interest. Standard POMDP solvers branch on raw observations, which may be high-dimensional or continuous. This leads to computational intractability as the solver creates separate branches for observations that differ only by sensor noise yet indicate the same underlying mineral, which SHM-POMDP addresses through observation abstraction, branching instead on inferred latent scientific properties. Each inferred property is a branch weighted by its posterior probability and updates use the full observation likelihood, so the belief over properties remains a distribution rather than a point estimate. This preserves the uncertainty that gives information-gathering actions their value while significantly reducing the branching complexity handled by the planner. The robot can thereby jointly reason about navigation decisions and their impact on scientific discovery while maintaining explicit uncertainty quantification throughout. We demonstrate this approach on two robotic domains: an extended rock sample problem for testing scalability, and a realistic geologic exploration task where a rover maps mineral compositions using spectral data from the Cuprite mining district.

The contributions of this paper include: (1) extending the SHM framework to explicitly model observation uncertainty, addressing a critical limitation of prior work that assumed deterministic sensor measurements; (2) a principled integration of the SHM with POMDP planning that improves tractability through observation abstraction, branching on physical properties rather than raw sensor data; and (3) experiments validate SHM-POMDP achieving up to 18.6% higher reward and 32.9% reduction in computation time per step on synthetic benchmarks, and $2.5\times$ higher information gain than the best information-theoretic baseline on realistic geologic exploration.

II. RELATED WORK

POMDPs provide a principled framework for decision-making under uncertainty, and Monte Carlo tree search (MCTS) methods have emerged as effective online solvers for continuous POMDPs. POMCPOW [7] handles continuous observation spaces through progressive widening, while PFT-DPW [7] uses particle filter trees for belief representation. DESPOT- α [8] and LABECOP [9] extend this to very large observation spaces through sparse sampling-based approaches. However, these are general-purpose solvers that operate on raw observation spaces, and when applied to robotic exploration with multimodal sensors, the observation space grows rapidly, making planning intractable.

Hierarchical approaches address this complexity by decomposing problems into manageable subproblems, improving computational tractability [10]–[15]. However, most prior work has focused on action hierarchies, where complex behaviours are decomposed into primitives to improve tractability [14]. These methods attempt to manage observation intractability by restricting which observations each

sub-POMDP processes [16]. This is insufficient because each distinct observation produces a different belief update through the POMDP likelihood model, and large observation spaces therefore increase the branching factor of the search tree—the primary driver of computational cost in online planning [17], [18]. Partitioning observations across sub-POMDPs does not resolve this, since the full observation must be maintained somewhere in the hierarchy [16]. Action-focused hierarchies therefore do not address the fundamental challenge of compressing the observation space itself.

This hierarchical structure between scientific properties and sensor measurements offers a principled basis for observation compression. While Bayesian networks have been explored to represent latent scientific properties in robotic exploration [1], [4], they have not addressed how navigation and sensing jointly affect belief evolution. Learning-based approaches such as Differentiable Particle Filters [19] and Visual Tree Search [20] take an alternative approach, training generative observation models offline and incorporating them into online POMDP solvers to address high-dimensional observations. While effective for visual domains, these methods require large-scale training and operate in uninterpretable latent spaces, limiting their applicability to scientific missions where interpretability is essential.

The most directly relevant work is [1], which operationalised the STM’s tripartite structure as Science Hypothesis Maps, implementing the hierarchy between investigation objectives, physical properties, and raw measurements as a Bayesian network. However, their implementation separates navigation from sensing and assumes deterministic observations, failing to capture how movement affects future observations under uncertainty. This limitation is shared by information-theoretic path planning, which provides performance guarantees but treats sensing and navigation independently [21]–[23]. Similar limitations are seen in deployed Mars systems such as AEGIS [24] and PIXL [25] which demonstrate autonomous capabilities but operate reactively, responding only to current observations rather than planning future actions under uncertainty. SHM-POMDP extends [1] hierarchical structure to explicitly model observation uncertainty, integrating physical properties into the POMDP observation model. This integration would be intractable without observation abstraction; by branching on physical properties rather than raw observations, we improve tractability while maintaining principled uncertainty quantification.

III. PRELIMINARIES

A. POMDP

A POMDP is typically represented by the tuple $(S, \mathcal{A}, \mathcal{O}, T, Z, R, \gamma)$, where S is the state space, $\mathcal{A}$ is the action space, and $\mathcal{O}$ is the observation space. The transition model $T(s, s', a) = P(s' | s, a)$ specifies the conditional probability of moving from state s to s' after performing action a , while the observation model $Z(s', a, o) = P(o | s', a)$ specifies the conditional probability of observing o after taking action a in state s' . The reward function $R(s, a)$ quantifies the utility of executing action a in state s . The

discount factor $\gamma \in [0, 1)$ balances immediate and future rewards. The agent maintains a belief distribution $b_t(s)$ over states, updated after each action-observation pair through Bayes' rule [26]:

$$b_{t+1}(s') \propto Z(s', a_t, o_t) \sum_{s \in \mathcal{S}} T(s, s', a_t) b_t(s). \quad (1)$$

B. Science Hypothesis Map

Scientific exploration requires reasoning about physical properties that cannot be measured directly but must be inferred from sensor data. We formalise this relationship through a SHM, adapting the hierarchical probabilistic structure introduced by [1] for integration with online POMDP planning.

The SHM defines a three-level hierarchy $\mathcal{H} \rightarrow \mathcal{Y} \rightarrow \mathcal{O}$ relating scientific concepts to observable measurements. At the highest level, investigation objectives $\mathcal{H} = \{h_1, \dots, h_l\}$ represent the scientific hypotheses a mission seeks to evaluate—for instance, determining whether a region formed through hydrothermal or sedimentary processes. At the intermediate level, latent physical scientific properties $\mathcal{Y} = \{y_1, \dots, y_m\}$ are attributes that serve as diagnostic indicators of these objectives; in geologic exploration, these are mineral types such as chlorite, serpentine or calcite. These properties are physical (grounded in measurable reality), scientific (meaningful to domain experts), and latent (partially observed through noisy sensors). For brevity, we refer to these as physical properties throughout. At the lowest level, the observation space $\mathcal{O}$ captures raw sensor observations such as hyperspectral reflectance data, where $\mathcal{O} \subseteq \mathbb{R}^d$ and d is the number of spectral bands.

These three levels are linked by conditional distributions. The distribution $P(y | h)$ specifies how physical properties relate to investigation objectives, such as which minerals are characteristic of which geologic formations. The observation likelihood $P_\theta(o | y)$, parameterised by learned parameters θ , specifies the distribution over sensor measurements given the underlying physical property. Together, these define a joint distribution that factorises as:

$$P(h, y, o) = P(h) \cdot P(y | h) \cdot P_\theta(o | y). \quad (2)$$

While the full hierarchy supports inference up to investigation objectives, the planning algorithm we present operates at the $\mathcal{Y} \rightarrow \mathcal{O}$ interface, using physical properties as the abstraction level for observation branching. This level provides sufficient abstraction for computational tractability while remaining grounded in scientifically meaningful categories.

The physical property space $\mathcal{Y}$ plays a central role in our formulation, serving as the bridge between raw sensor data and scientific interpretation. A physical property $y \in \mathcal{Y}$ is an attribute of the environment that satisfies three criteria. First, it must be observationally grounded, meaning distinct properties induce distinct measurement distributions: $P_\theta(o | y_i) \neq P_\theta(o | y_j)$ for any $y_i \neq y_j$. This guarantees that the property can be inferred from sensor data. Second, it must be interpretable, corresponding to a discrete set of domain-meaningful concepts $\mathcal{C}$ (e.g., mineral types, chemical

Algorithm 1: SHM-POMDP Planning and Execution

Input: Belief b , depth d , SHM model $P_\theta(o | y, s', a)$, threshold ϵ , simulations N

Output: Action a^* and updated belief b'

Procedure *TreeSearch*(b, d):

```

if  $d = 0$  or IsTerminal( $b$ ) then
  return 0
if  $b \notin \text{Tree}$  then
  InitialiseNode( $b$ )
 $a \leftarrow \text{SelectAction}(b)$ 
 $s = (y, s_r) \sim b$ 
 $b' \leftarrow \text{PredictBelief}(b, a)$ 
 $s' = (y', s'_r) \sim b'$ 
 $o_{exp} \leftarrow \text{ComputeExpectedObs}(s')$ 
foreach  $\tilde{y}_i \in \tilde{\mathcal{Y}}$  do
   $p_i \leftarrow P_\theta(o_{exp} | \tilde{y}_i, s', a) \cdot P(\tilde{y}_i | s', a)$ 
  Normalise( $p_1, \dots, p_k$ ) to get  $P(\tilde{y} | o_{exp}, s', a)$ 
   $\tilde{\mathcal{Y}}_{branch} \leftarrow \{\tilde{y}_i : P(\tilde{y}_i | o_{exp}, s', a) > \epsilon\}$ 
   $Q \leftarrow R(s, a)$ 
  foreach  $\tilde{y}_i \in \tilde{\mathcal{Y}}_{branch}$  do
    if  $(a, \tilde{y}_i) \notin \text{Children}(b)$  then
       $b_{child} \leftarrow \text{UpdateBelief}(b', a, \tilde{y}_i)$ 
      AddChild( $b, a, \tilde{y}_i, b_{child}$ )
    else
       $b_{child} \leftarrow \text{GetChild}(b, a, \tilde{y}_i)$ 
       $Q \leftarrow Q + \gamma \cdot P(\tilde{y}_i | o_{exp}, s', a) \cdot \text{TreeSearch}(b_{child}, d - 1)$ 
   $Q(b, a) \leftarrow Q$ 
return  $Q(b, a)$ 

```

Procedure *Execute*(b, a^*):

```

Execute  $a^*$  in environment
 $o \leftarrow \text{GetObservation}()$ 
foreach particle  $s_i = (y_i, s_{r,i})$  in  $b$  do
   $s'_i = (y'_i, s'_{r,i}) \sim T(s_i | s_i, a^*)$ 
   $w'_i \leftarrow w_i \cdot P_\theta(o | y'_i, s'_i, a^*)$ 
return WeightedResample( $\{s'_i, w'_i\}$ )

```

repeat

```

  // plan
  for  $n \leftarrow 1$  to  $N$  do
     $\text{TreeSearch}(b, d)$ 
  // execute
   $b \leftarrow \text{Execute}(b, \arg \max_a Q(b, a))$ 

```

until *mission complete*;

compositions, or terrain classes) such that $\mathcal{Y} \subseteq \mathcal{C}$, rather than an arbitrary continuous latent variable. Third, in the context of exploration missions, a physical property is static: the robot's actions do not alter the underlying property, which is formally defined by the transition deterministic mapping $T(y, a, y') = 1$ if $y = y'$ and 0 otherwise.

IV. SHM-POMDP

We integrate the Science Hypothesis Map with the POMDP to enable science-driven planning under uncertainty; we term the resulting formulation an SHM-POMDP. The state space factors as $\mathcal{S} = \mathcal{Y} \times \mathcal{S}_r$, where $\mathcal{Y}$ represents physical properties and $\mathcal{S}_r$ is the remaining state containing all other task-relevant variables. A complete state $s = (y, s_r)$ specifies both the physical property of interest and the robot's operational context. The variables $y \in \mathcal{Y}$ are partially

observed, while $s_r \in \mathcal{S}_r$ may be fully observed, partially observed, or a mixture of both, depending on the application.

The transition function is then defined as $T((y, s_r), a, (y', s'_r)) = T(y, a, y') \times T(s_r, a, s'_r)$, where $T(y, a, y') = 1$ for $y = y'$ and 0 otherwise, as physical properties remain static during the mission.

The SHM defines a generative model $P_\theta(o | y)$ specifying how physical properties produce observations. Within the SHM-POMDP, this model is extended to $P_\theta(o | y, s', a)$ to account for measurement conditions such as distance to target and sensor noise. During planning, the robot predicts the belief forward through the transition model (PredictBelief), then inverts the observation model to reweight the predicted belief into a posterior over property values (UpdateBelief), applying no further transition. Here PredictBelief gives $b'(s') = \sum_s T(s' | s, a)b(s)$ and UpdateBelief reweights, $b_{\text{child},i}(s') \propto P(\tilde{y}_i | s', a)b'(s')$ then resamples. We denote these inferred values as $\tilde{y} \in \tilde{\mathcal{Y}}$ to distinguish the robot's inferred estimate from the true physical property y in the state. The inferred set $\tilde{\mathcal{Y}}$ need not equal $\mathcal{Y}$ —it may refine $\mathcal{Y}$ or add an uncertain outcome—but each element is scientifically meaningful and $|\tilde{\mathcal{Y}}| \ll |\mathcal{O}|$.

To formalise this inversion, we augment the standard POMDP observation function to include the inferred property as an intermediate variable. The observation function becomes:

$$Z(s', a, o) = P(o, \tilde{y} | s', a) = P_\theta(o | \tilde{y}, s', a) \cdot P(\tilde{y} | s', a), \quad (3)$$

where $P_\theta(o | \tilde{y}, s', a)$ is the SHM's likelihood of observations given the inferred property and measurement conditions, and $P(\tilde{y} | s', a)$ is the prior distribution over inferred properties given the sampled state. This factorisation allows us to compute the posterior distribution over inferred properties via Bayes' rule. For a state $s' = (y', s'_r)$ drawn from the predicted belief b' , an expected observation o_{exp} (ComputeExpectedObs) is generated. The posterior over inferred properties given this expected observation is:

$$P(\tilde{y}_i | o_{\text{exp}}, s', a) = \frac{P_\theta(o_{\text{exp}} | \tilde{y}_i, s', a) \cdot P(\tilde{y}_i | s', a)}{\sum_j P_\theta(o_{\text{exp}} | \tilde{y}_j, s', a) \cdot P(\tilde{y}_j | s', a)}, \quad (4)$$

where $\tilde{y}_i$ is a candidate inferred property, o_{exp} is the characteristic observation from the sampled state s' , P_θ is the learned SHM generative observation model, $P(\tilde{y}_i | s', a)$ is the prior probability of observing $\tilde{y}_i$ given the sampled state and action, and j indexes over all possible properties in $\tilde{\mathcal{Y}}$.

We use the computed distribution $P(\tilde{y}_i | o_{\text{exp}}, s', a)$ to determine which inferred properties to branch on. The algorithm identifies all $\tilde{y}_i$ values with non-negligible probability, specifically $\{\tilde{y}_i : P(\tilde{y}_i | o_{\text{exp}}, s', a) > \epsilon\}$ where ϵ is a threshold parameter. The solver then creates branches for each of these inferred properties, treating them as discrete outcomes rather than raw sensor data [7]. In the resulting tree structure, nodes represent belief states b over (y, s_r) , and edges represent action-observation pairs $(a, \tilde{y}_i)$ where $\tilde{y}_i$ is an inferred property. Each branch $(a, \tilde{y}_i)$ leads to a child belief representing the posterior after inferring property $\tilde{y}_i$.

The tree thus encodes uncertainty over physical properties that might be inferred, rather than specific sensor readings. This abstraction reduces the branching factor from the size of the continuous observation space $|\mathcal{O}|$ to the number of likely inferred properties $|\{\tilde{y}_i : P(\tilde{y}_i | o_{\text{exp}}, s', a) > \epsilon\}| \leq |\tilde{\mathcal{Y}}|$, where $|\tilde{\mathcal{Y}}| \ll |\mathcal{O}|$. Algorithm 1 provides the complete procedure for SHM-POMDP tree search with observation abstraction.

The probability $P(\tilde{y}_i | o_{\text{exp}}, s', a)$ serves as a weight that determines which branches are created and how much each branch contributes to the expected value calculation during planning. While the use of a single representative observation o_{exp} to determine branches is an approximation during planning, the full observation likelihood $P_\theta(o | y, s', a)$ is used without approximation during execution for belief updates, preserving sensing fidelity where it most affects downstream planning. During tree search, the value function aggregates over branches as $Q(b, a) = R(s, a) + \gamma \sum_i P(\tilde{y}_i | o_{\text{exp}}, s', a) \cdot V(b_{\text{child},i})$, where $s \sim b$ is the state sampled at this node and each child belief represents the posterior after discovering $\tilde{y}_i$. This weighted aggregation ensures the planning process accounts for the full distribution of possible inferences while achieving computational efficiency through discrete branching.

During execution, the robot takes the best action a^* determined by the tree search and receives a high-dimensional observation $o \in \mathbb{R}^d$ from its sensors. Rather than using the inferred properties from planning, the belief update employs the full observation likelihood from the SHM. Particle weights are updated as $w'_i = w_i \cdot P_\theta(o | y'_i, s'_i, a^*)$ using the SHM likelihood, then resampled to concentrate on likely classifications. This combination of discrete branching for planning and full observations for execution achieves tractability without sacrificing sensing fidelity.

The SHM provides a learnable probabilistic generative observation model that can be learned offline from training data consisting of paired observations and ground-truth labels. The model captures the conditional distribution $P_\theta(o | y, s', a)$, which represents the likelihood of observing data o given a physical property y , measurement conditions s' and action a . In the experimental domains presented, all sensing actions use identical observation mechanisms, $P_\theta(o | y, s', a) = P_\theta(o | y, s')$ for all a . The specific implementation can vary by domain—Gaussian Mixture Models for spectral data, neural networks for images, or other probabilistic models appropriate for the sensor modality.

V. EXPERIMENTS

We evaluate SHM-POMDP across two domains that address the challenge of autonomous exploration in high-dimensional observation spaces where scientific reasoning is required. The robot must infer physical properties from noisy, continuous high-dimensional sensor data while jointly optimising navigation and information gathering under budget constraints. Since movement incurs a step penalty, cumulative reward serves as a joint proxy for scientific and navigational efficiency, where higher reward reflects effective

and efficient paths to achieve them. Section V-A uses an extended RockSample benchmark with continuous multi-modal observations to demonstrate scalability: as observation dimensions grow from 5D to 50D, SHM-POMDP maintains constant branching factor $O(|\tilde{\mathcal{Y}}|^d)$ while standard solvers face $O(|\mathcal{O}|^d)$. Section VI uses realistic geologic mapping with authentic spectral confusion to compare SHM-POMDP against state of the art information-theoretic planners; when spectral overlap creates ambiguity, belief-based planning detours to disambiguate while static planners cannot adapt.

We assess two aspects of performance. First, computational efficiency captures planning time per decision step, reflecting feasibility for onboard processors with limited computational budgets. Second, scientific return measures the value of exploration: cumulative reward for RockSample and realised information gain for Cuprite. Statistical significance is assessed via paired t-tests ($\alpha = 0.05$) across 100 trials with identical random seeds. All experiments were conducted on a laptop with an Intel Ultra 7 Processor 155H 4.8 GHz and 32GB RAM.

A. Extended RockSample

1) *Setup*: We modify the Information Search RockSample (ISRS) domain [6] to model a rover exploring an $n \times n$ grid with k rocks of unknown scientific value (good or bad). The state $s = (y, s_r)$ consists of latent rock types $y \in \{good, bad, uncertain\}^k$ (partially observable) and $s_r = (x_{rover}, y_{rover}, \{(x_i, y_i)\}_{i=1}^k)$, which contains the rover position and rock locations (fully observable). The rover can move in the four cardinal directions, sample rocks for reward, and sense rocks with continuous-valued observations whose fidelity decays with distance. Movement incurs a penalty, requiring efficient path planning that balances information gathering against traversal costs.

We simulate realistic sensing through five modalities with distinct noise characteristics and structured intra-modality correlations. The rover’s multi-sensor observation is modelled as a multivariate normal whose mean encodes the expected rock signature and whose covariance captures distance dependent noise [26]. While rock states are binary, the SHM-POMDP maps these high-dimensional observations to three abstracted values $\tilde{y} \in \{good, bad, uncertain\}$ for tree branching. High confidence observations map to $\tilde{y} = good$ or bad , while ambiguous observations yield $\tilde{y} = uncertain$. This 3-way branching reduces computational complexity while preserving full observational information $P_\theta(o | y, s')$ for belief updates.

We compare SHM-POMDP against a continuous-observation POMDP baseline, both using the POMCPOW solver. The key difference is that SHM-POMDP branches on abstracted observations $\tilde{y}$ while the baseline branches on continuous observations o . Hyperparameters including ϵ , were tuned on 5D observations via preliminary sweeps and fixed for higher dimensions to isolate scaling effects, yielding $\epsilon = 10^{-4}$. Both methods used identical action selection, depth limits, and particle counts to ensure comparable computational requirements.

TABLE I: Scalability comparison: SHM-POMDP vs Continuous baseline across increasingly difficult RockSample problems (mean $\pm$ 95% CI, 100 trials). SHM-POMDP achieves up to 18.6% higher reward and 32.9% reduction in computation time per step as observation dimensionality increases.

Env	Rocks	Dim	Reward		Time/step (sec/step)	
			CONT	SHM	CONT	SHM
5x5	4	5D	8.58 $\pm$ 0.60	9.04$\pm$0.73	0.22	0.23
		10D	8.08 $\pm$ 0.75	8.99$\pm$0.74	0.38	0.31
		30D	8.59 $\pm$ 0.72	9.81$\pm$0.82	1.00	0.82
		50D	8.27 $\pm$ 0.85	8.86$\pm$0.87	1.26	0.90
7x7	8	5D	6.66 $\pm$ 0.92	7.34$\pm$0.82	0.45	0.45
		10D	8.51 $\pm$ 1.05	8.85$\pm$1.01	0.77	0.63
		30D	8.47 $\pm$ 0.89	8.99$\pm$0.96	2.18	1.67
		50D	7.19 $\pm$ 0.88	7.35$\pm$0.92	2.57	1.69
15x15	15	5D	3.13$\pm$1.61	3.09 $\pm$ 1.62	0.60	0.59
		10D	3.51 $\pm$ 1.68	3.92$\pm$1.58	1.01	0.73
		30D	3.56 $\pm$ 1.65	4.07$\pm$1.56	3.52	2.72
		50D	2.79 $\pm$ 1.57	3.31$\pm$1.62	6.81	4.57

2) *Results*: Table I summarises performance across systematically scaled problem complexity ($5 \times 5, 7 \times 7, 15 \times 15$ maps) and observation dimensionality (5D - 50D). Both methods executed identical episode counts per planning step using 100 identical random seeds.

SHM-POMDP’s primary advantage emerges in computational efficiency, with gains that scale dramatically with problem complexity and observation dimensionality. The computational efficiency showed minimal advantages on small problems (5×5 with 5D: -4.5%) but substantial improvements on complex problems (15×15 with 50D: 32.9%). This scaling reflects the different architecture of the algorithms as standard POMCPOW’s progressive widening produces highly varied samples in high-dimensional spaces, forcing the tree wide and shallow by fragmenting computation across $O(|\mathcal{O}|^d)$ branches where d is the search depth. SHM-POMDP circumvents this explosion through hierarchical decomposition. During tree expansion, the planner branches only over the abstracted observations $\tilde{\mathcal{Y}}$, maintaining constant $O(|\tilde{\mathcal{Y}}|^d) = O(3^d)$ branching regardless of observation dimensionality. When an observation is received during rollout, SHM-POMDP evaluates the likelihood $P_\theta(o | y, s')$ through the pre-defined sensor model without creating explicit tree branches for each possible observation. This architectural difference grows on complex 15×15 maps with many rocks and long horizons, where deeper search matters most.

The observed reward improvements (up to 18.6% in $15 \times 15, 50D$) reflect more efficient navigation to valuable rocks, as reduced branching enables deeper tree search under fixed computational budget, revealing path value differences shallow trees cannot distinguish. Additionally, SHM-POMDP exploits sensor correlations encoded in the observation model. The observation model $P(o | y, s')$ incorporates both intra-modal (specified correlation coefficients within each sensor type) and inter-modal (all sensors respond proportionally to the same underlying rock property) correlations. This

correlation structure means numerically different but physically equivalent observations naturally cluster into three abstracted values $\tilde{y}$, redirecting computation from redundant distinctions towards deeper search.

The reward improvements are largest in the 10–30D range rather than at 50D. At intermediate dimensions, progressive widening creates enough branches to fragment computational effort across observation distinctions, yet not enough to cover the space meaningfully—the worst case for unstructured observation handling, which SHM-POMDP’s fixed branching circumvents. At 50D, the gap narrows not because SHM-POMDP degrades, but because the solver’s behaviour changes. With extreme sparsity, every sampled observation is unique, so progressive widening creates a very wide tree that restricts search depth. SHM-POMDP’s smaller branching factor still enables a deeper search, as reflected in the computational efficiency gain.

VI. CUPRITE GEOLOGIC EXPLORATION PATH PLANNING

1) *Setup*: We evaluate our approach using the Cuprite Hills mining district, Nevada; a well characterised site with extensive ground-truth mineralogical maps and airborne hyperspectral imagery. The site contains diverse hydrothermal alteration zones containing distinctive mineral assemblages, providing a challenging yet realistic testbed for autonomous exploration [27]. We discretise a 1×1 km region into a 10×10 grid where each cell represents 100×100 m of terrain.

The POMDP state comprises mineral types y for each of 10 rocks scattered across the map (partially observable) and s_r containing the rover and rock locations (fully observable). The rover must identify mineral classes of interest (e.g., alunite, chlorite) from hyperspectral reflectance vectors o spanning the visible to short-wave infrared range. Many minerals exhibit characteristic absorption features due to their molecular composition, enabling supervised learning to associate spectra with mineral labels.

To construct the SHM observation model, we aligned USGS mineralogical maps [1] with co-registered Next Generation Airborne Visible Infrared Imaging Spectrometer (AVIRIS-NG) hyperspectral data [28], yielding pixel-level (y, o) pairs across five mineral classes. We applied Principal Component Analysis (PCA) to reduce dimensionality while preserving diagnostic variance, then fit a Gaussian Mixture Model (GMM) with one component per class to obtain $P_\theta(o | y)$. This captures both the central tendency of each mineral’s spectral signature and the covariance structure induced by sensor noise and natural variability [29]. During online planning, we augment this model with distance-dependent observation noise to create $P_\theta(o | y, s')$, where measurement quality degrades with rover-to-rock distance. The branching threshold $\epsilon = 10^{-6}$ was selected via preliminary tuning to account for the finer spectral distinctions between mineral classes. This formulation provides likelihoods that can be used for Bayesian belief updates in the POMDP. It also accounts for inter-class overlap by representing minerals as overlapping Gaussian components rather than hard classification boundaries.

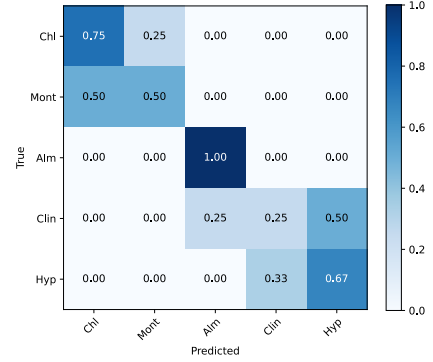


Fig. 2: Confusion matrix for five-mineral LDA classifier trained on AVIRIS-NG spectra. Off-diagonal values indicate spectral overlap between minerals (e.g., Chlorite–Montmorillonite), demonstrating the observation uncertainty that SHM-POMDP explicitly models.

To validate the spectral classification, we trained a linear discriminant analysis (LDA) classifier on five representative minerals using AVIRIS-NG. The confusion matrix in Fig. 2 illustrates classification errors due to overlapping absorption features. These results underscore the need for a probabilistic observation model, as a hard classifier would obscure uncertainty in spectral interpretation [30]. Similar to the VTS framework [20], we combine offline-learned probabilistic models with online POMDP planning, but here our SHM enables the planner to branch on abstracted observations while still maintaining beliefs over true mineral states.

In each trial, 10 rocks are randomly placed on the discretised Cuprite Hills grid. The rover navigates from the northwest to the southeast corner under a budget of twice the start to goal Manhattan distance, with the objective of maximising information gain about chlorite presence. The rover receives a terminal reward for reaching the goal and a per-step penalty so that information gathering is traded against reaching the goal within budget. The true mineral state $y \in \{\text{chlorite}, \text{not-chlorite}\}$ is the target variable, while abstracted observations $\tilde{y} \in \{\text{chlorite (chl)}, \text{montmorillonite (mont)}, \text{almandine (alm)}, \text{clinoclone (clin)}, \text{hypersthene (hyp)}\}$ represent classification outcomes used for tree branching. Information gain is computed as:

$$IG = H(y_{\text{chlorite}}) - H(y_{\text{chlorite}} | o). \quad (5)$$

Where H denotes binary entropy in nats and posteriors are updated using the learned GMM observation model $P_\theta(o | y, s')$. Per-trial information gain is computed as the cumulative sum of per-step values across all sensing actions in the episode, where entropy is evaluated using the particle belief approximation maintained by the planner.

We do not repeat the POMDP comparison from Section V-A as the computational advantages remains unchanged. Instead, we compare against science-aware information-theoretic path planners to test SHM-POMDP’s ability to handle real spectral confusion through adaptive belief-based planning versus static information-gain heuristics. We evaluate two scenarios: (1) uniform prior, where all planners begin with $p(\text{chlorite}) = 0.2$, testing performance under maximum

TABLE II: Performance comparison on Cuprite geologic exploration (mean $\pm$ 95% CI, 100 trials). Realised information gain is measured in nats; path length is the number of grid cells traversed. SHM-POMDP with uniform prior achieves $2.5\times$ higher information gain than the best baseline and 80% of oracle performance.

Planner	Realised IG (nats)	Path length (cells)
<i>Uniform prior</i>		
Greedy	0.242 ± 0.053	19.1 ± 0.1
Random	0.156 ± 0.061	30.8 ± 1.1
pSPIEL	0.305 ± 0.080	30.3 ± 0.6
SHM-POMDP	0.760 ± 0.156	31.8 ± 4.9
<i>Ground truth prior</i>		
Greedy (oracle)	0.935 ± 0.097	31.2 ± 0.5
pSPIEL (oracle)	0.944 ± 0.101	30.8 ± 0.5

uncertainty; and (2) oracle initialisation, where baseline planners (Greedy, pSPIEL) receive ground-truth mineral maps for planning, representing best-case performance in a perfect information ‘oracle planning’ scenario. The oracle baselines receive ground-truth maps as a deliberate best-case upper bound. They compute optimal paths assuming perfect knowledge but still receive noisy observations during execution. Notably, SHM-POMDP uses only uniform priors in both cases. The fundamental difference is that baseline planners follow pre-computed paths regardless of observations, while SHM-POMDP updates beliefs and replans based on actual spectral measurements. We compare three science-aware planners that incorporate mineralogical information into their decision making, and one science-blind baseline. *Random* selects actions uniformly at random. *pSPIEL* uses submodular orienteering approach based on [21], with budget-aware lookahead. *Greedy* selects neighbouring cells maximising expected entropy reduction. *SHM-POMDP* maintains full belief distribution and replans under uncertainty.

2) *Results*: Table II presents the performance comparison across all trials. SHM-POMDP achieves significantly higher information gain (0.760 ± 0.156 nats) compared to the best baseline planner pSPIEL (0.305 ± 0.080 nats), representing a $2.5\times$ improvement while maintaining comparable path lengths.

SHM-POMDP’s advantage stems from its ability to maintain and update beliefs during execution. Heuristic planners collapse each measurement to a single label via maximum likelihood, discarding measurement confidence levels and treating ambiguous spectra as certain. When noise or spectral overlap causes deviations from expectation, these planners cannot adapt, causing overconfident early choices that cause errors to propagate into the rest of the trajectory.

Notably, SHM-POMDP with only uniform priors achieves 80% of the information gain obtained by oracle planners (0.760 vs 0.944 nats). This performance gap reveals the fundamental limitation of static planning: even with perfect prior knowledge, static planners cannot adapt when observations deviate from expectations. SHM-POMDP maintains beliefs through particle filtering weighted by $P_\theta(o | y, s')$, where each observation reweighs particles and drives exploration toward observations that disambiguate between chlorite and

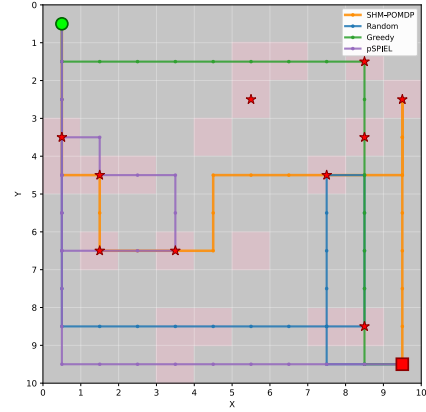


Fig. 3: Cuprite exploration paths (start: north-west; goal: south-east). Belief-based planning (SHM-POMDP) enables deliberate detours to disambiguate uncertain observations, unlike static information-theoretic planners.

non-chlorite. Particle diversity encourages exploration in uncertain regions, and noisy observations are naturally integrated across the belief distribution, producing robustness absent in heuristic methods. Because SHM-POMDP computes a policy rather than a fixed path, its behaviour adapts to realised observations, explaining the higher variance in path length compared to deterministic baselines.

Fig. 3 illustrates how algorithmic differences manifest in path structure. With only a uniform prior, heuristic planners cannot discriminate between rocks before observing them, so their strategies reflect static heuristics rather than adaptive reasoning. Random achieves 3/10 detections purely through chance. Greedy’s short trajectory (19 steps, 3/10 detections) reflects myopic selection by $IG(\text{rock}) - \lambda \times \text{detour cost}$, leaving many rocks unexplored. pSPIEL achieves only 4/10 detections despite its submodular optimisation. It groups rocks into clusters and plans systematic coverage within each cluster, but the uniform prior provides no guidance about which clusters actually contain valuable minerals. This produces consistent short paths with low variance (around 30 steps on average), but much of this efficiency is spent covering a singular uninformative region while missing high-value areas identifiable through adaptive belief updates, since its precomputed plan cannot respond when observations deviate from expectation.

In contrast, SHM-POMDP produces longer and more variable paths (around 32 steps with higher variance) that emerge from belief updates during execution. Starting from the same uninformative prior, the planner gathers initial observations to identify promising regions, then systematically explores areas where updated beliefs indicate chlorite probability exceeds the baseline. The difference is not path shape but the reason for revisiting. Random revisits by chance; pSPIEL revisits due to precomputed cluster coverage; SHM-POMDP revisits in response to observations. When an ambiguous measurement increases uncertainty, the planner deliberately seeks additional observations to disambiguate, building an increasingly accurate belief map. This adaptive replanning explains both the additional path length and the significant

information gain advantage.

VII. CONCLUSION

SHM-POMDP enables autonomous scientific exploration by integrating domain knowledge into belief-space planning through hierarchical probabilistic models. First, the SHM's physical property space provides a principled basis for observation abstraction, enabling branching on physical properties rather than raw sensor data to avoid the computational explosion that cripples traditional POMDP solvers while preserving full sensor information through learned observation models that capture both inter-modal and intra-modal correlations. Second, explicit modelling of observation uncertainty enables adaptive exploration that myopic and heuristic methods cannot achieve. The ability to maintain full belief distributions rather than point estimates allows the planner to recognise the value of partial information and adapt strategies based on realised observations.

Future work should extend SHM-POMDP beyond simple categorical physical properties to richer representations with probabilistic dependencies, thereby better capturing complex relationships between multi-modal observation and underlying phenomena. Online adaptation of observation models would increase robustness when sensor characteristics drift or new environments differ from training data. As missions venture to increasingly remote environments where communication latency precludes human intervention, frameworks enabling principled reasoning about scientific uncertainty will become essential for maximising discovery potential.

REFERENCES

- [1] A. Candela, D. Thompson, E. N. Dobre, and D. Wettergreen, "Planetary robotic exploration driven by science hypotheses for geologic mapping," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada, Sep. 2017, pp. 3811–3818.
- [2] A. N. Hock, N. A. Cabrol, J. M. Dohm, J. Piatek, K. Warren-Rhodes, S. Weinstein, D. S. Wettergreen, E. A. Grin, J. Moersch, C. S. Cockell *et al.*, "Life in the atacama: a scoring system for habitability and the robotic exploration for life," *Journal of Geophysical Research: Biogeosciences*, vol. 112, no. G4, 2007.
- [3] D. R. Thompson, D. S. Wettergreen, and F. J. C. Peralta, "Autonomous science during large-scale robotic survey," *Journal of Field Robotics*, vol. 28, no. 4, pp. 542–564, 2011.
- [4] M. J. Kochenderfer, *Decision Making Under Uncertainty: Theory and Application*. Cambridge, MA, USA: MIT Press, 2015.
- [5] A. Arora *et al.*, "An approach to autonomous science by modeling geological knowledge in a bayesian framework," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada, 2017, pp. 3803–3810.
- [6] S. Choudhury, N. Gruver, and M. J. Kochenderfer, "Adaptive informative path planning with multimodal sensing," in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 30, 2020, pp. 57–65.
- [7] Z. Sunberg and M. Kochenderfer, "Online algorithms for POMDPs with continuous state, action, and observation spaces," in *Proc. Int. Conf. Automated Planning and Scheduling (ICAPS)*, vol. 28, Delft, Netherlands, 2018, pp. 259–263.
- [8] N. P. Garg, D. Hsu, and W. S. Lee, "DESPOT- α : Online POMDP planning with large state and observation spaces," in *Proc. Robotics: Science and Systems (RSS)*, vol. 3, no. 3, 2019, pp. 3–11.
- [9] M. Hoerger and H. Kurniawati, "An on-line pomdp solver for continuous observation spaces," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 7643–7649.
- [10] C. Chuck, S. Chockchowwat, and S. Niekum, "Hypothesis-driven skill discovery for hierarchical deep reinforcement learning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2020, p. 5572–5579.
- [11] H. Kurniawati, Y. Du, D. Hsu, and W. S. Lee, "Motion planning under uncertainty for robotic tasks with long time horizons," *The International Journal of Robotics Research*, vol. 30, no. 3, pp. 308–323, 2011.
- [12] Y. Lee, P. Cai, and D. Hsu, "Magic: Learning macro-actions for online pomdp planning," in *Robotics: Science and Systems (RSS)*, 2021.
- [13] J. Pineau and S. Thrun, "High-level robot behavior control using pomdps," in *AAAI-02 Workshop on Cognitive Robotics*, vol. 107, 2002, p. 4.
- [14] M. Sridharan, J. Wyatt, and R. Dearden, "Planning to see: A hierarchical approach to planning visual actions on a robot using pomdps," *Artificial Intelligence*, vol. 174, no. 11, pp. 704–725, 2010.
- [15] S. A. Serrano, E. Santiago, J. Martinez-Carranza, E. F. Morales, and L. E. Sucar, "Knowledge-based hierarchical pomdps for task planning," *Journal of Intelligent & Robotic Systems*, vol. 101, no. 4, p. 82, 2021.
- [16] J. Pineau and S. Thrun, *An integrated approach to hierarchy and abstraction for POMDPs*. Carnegie Mellon University, the Robotics Institute, 2002.
- [17] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.
- [18] N. Roy, G. Gordon, and S. Thrun, "Finding approximate pomdp solutions through belief compression," *Journal of artificial intelligence research*, vol. 23, pp. 1–40, 2005.
- [19] R. Jonschkowski, D. Rastogi, and O. Brock, "Differentiable particle filters: End-to-end learning with algorithmic priors," in *Proceedings of Robotics: Science and Systems (RSS)*, Pittsburgh, PA, USA, 2018.
- [20] S. Deglurkar, M. H. Lim, J. Tucker, Z. N. Sunberg, A. Faust, and C. J. Tomlin, "Compositional learning-based planning for vision POMDPs," in *Proc. 5th Annual Conference on Learning for Dynamics and Control (L4DC)*, vol. 211. PMLR, 2023, pp. 1–14.
- [21] A. Singh, "Nonmyopic adaptive informative path planning for multiple robots," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 34, 2009, pp. 1843–1850.
- [22] G. A. Hollinger and G. S. Sukhatme, "Sampling-based robotic information gathering algorithms," *The International Journal of Robotics Research*, vol. 33, no. 9, pp. 1271–1287, 2014.
- [23] C. Chekuri and M. Pal, "A recursive greedy algorithm for walks in directed graphs," in *46th annual IEEE symposium on foundations of computer science (FOCS'05)*. IEEE, 2005, pp. 245–253.
- [24] T. A. Estlin, B. J. Bornstein, D. M. Gaines, R. C. Anderson, D. R. Thompson, M. Burl, R. Castano, and M. Judd, "Aegis automated science targeting for the mer opportunity rover," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 3, no. 3, pp. 1–19, 2012.
- [25] A. C. Allwood, L. A. Wade, M. C. Foote, W. T. Elam, J. A. Hurowitz, S. Battel, D. E. Dawson, R. W. Denise, E. M. Ek, M. S. Gilbert *et al.*, "Pixl: Planetary instrument for x-ray lithochemistry," *Space Science Reviews*, vol. 216, no. 8, p. 134, 2020.
- [26] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA: MIT Press, 2005.
- [27] G. A. Swayze, R. N. Clark, A. F. Goetz, K. E. Livo, G. N. Breit, F. A. Kruse, S. J. Sutley, L. W. Snee, H. A. Lowers, J. L. Post *et al.*, "Mapping advanced argillic alteration at cuprite, nevada, using imaging spectroscopy," *Economic geology*, vol. 109, no. 5, pp. 1179–1221, 2014.
- [28] L. Hamlin, R. Green, P. Mouroulis, M. Eastwood, D. Wilson, M. Dudik, and C. Paine, "Imaging spectrometer science measurements for terrestrial ecology: Aviris and new developments," in *2011 Aerospace conference*. IEEE, 2011, pp. 1–7.
- [29] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 763–767, 2002.
- [30] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote sensing of environment*, vol. 113, pp. S110–S122, 2009.