Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cognitive Value-Aware Semantic Communication and Disagreement-Driven Formation Reconfiguration for Large-Model- Enabled UAV Swarms

Yangming Wu
Nanjing University of Aeronautics and Astronautics
Zhisheng Wang
wangzhisheng@nuaa.edu.cn
Nanjing University of Aeronautics and Astronautics
Quan Ouyang
Nanjing University of Aeronautics and Astronautics
Nannan Li
Beijing Institute of Control and Electronics Technology
Rufei Zhang
Beijing Institute of Control and Electronics Technology
Shuijuan Gao
Nanjing Vocational Institute of Transport Technology

Research Article

Keywords: UAV swarm, semantic communication, information value, cognitive freshness, cognitive disagreement, dynamic formation reconfiguration, large-model inference, edge intelligence

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11757576/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Cognitive Value-Aware Semantic Communication and Disagreement-Driven Formation Reconfiguration for Large-Model-Enabled UAV Swarms

Yangming Wu[1,2],Zhisheng Wang[1*],Quan Ouyang[1],Nannan Li[3],Rufei Zhang[3],Shuijuan Gao[1,2]

Abstract. Large-model-enabled UAV swarms now care less about moving more bits than about agreeing on what those bits mean. The communication and formation literature, though, still measures throughput, delay, packet loss and geometric shape, and none of these quantities captures how much a given message is worth to the decision the swarm is about to take. We couple cognitive value-aware semantic communication (CVSC) with disagreement-driven formation reconfiguration (DFR) in one framework.Transmission is no longer triggered by a change in the data but by a change in the decision value, which is formalised through an information value function $V(z)$ that adds the gains in task success, cognitive consistency and risk detection and subtracts the cost of transmission. All are computable: the effect of a task is the KL divergence between policy distributions at consecutive decision steps, and the risk impact is the mutual information carried by the posterior over risk targets. Cognitive freshness $\Phi(z,t)$ is also employed to ascertain whether an existing conclusion can still support the current decision, rather than considering the age of the underlying data. Selection of key visual tokens in the encoding stage is a monotone submodular maximisation problem with a cardinality constraint, and the greedy rule has a $1-1/e$ guarantee; multi-precision semantic coding has been shown to be a trade-off among rate, semantic distortion and task performance, and its Lagrangian dual yields a water-filling allocation plus an integer solution. Reconfiguration in the formation is caused by cognitive disunity rather than geometric separation; thus, a disagreement field is formed, its origin is localised, and both observation and relay nodes are moved simultaneously under connectivity and dynamic constraints, and stability is guaranteed by a Lyapunov function. The four decisions of what to send, at what precision, from where, and along which inference path form a mixed-integer non-convex program that is solved by alternating optimisation, successive convex approximation and ADMM, along with convergence and complexity results. On 32 heterogeneous UAVs and 30 Monte Carlo runs, CVSC-DFR reduced the communication volume per task by 71.3%, increased the task success rate by 19.6 percentage points, achieved a cognitive consistency of 0.938, shortened the link-recovery time by 58.4%, and predicted the value of semantic information with an accuracy of 0.917 compared to raw image transmission; this was done by fixing the communication rate, triggering based on state-change events, age-of-information scheduling, centralised scheduling, and disagreement-blind formation control.

Index Terms— UAV swarm; semantic communication; information value; cognitive freshness; cognitive disagreement; dynamic formation reconfiguration; large-model inference; edge intelligence

✉ Yangming Wu
wuyangming@nuaa.edu.cn
Zhisheng Wang
wangzhisheng@nuaa.edu.cn
Quan Ouyang
ouyangquan@nuaa.edu.cn
Nannan Li
beyondlnn@163.com
Rufei Zhang
zrf_lucky@sina.com
Shuijuan Gao
2685191913@qq.com

1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China
2. College of Electronic and Information Engineering, Nanjing Vocational Institute of Transport Technology, Nanjing, 211188, China
3. Beijing Institute of Control and Electronics Technology, Beijing, 100038, China

1 Introduction

UAV swarms are becoming more intelligent and are equipped with many sensors and onboard computers. Large models are deployed at the edge of the swarm in the embodied-agent paradigm for on-device inference, resource scheduling and offloading decisions [1]. Edge routing for ubiquitous LLM inference extends the concept of embodied reasoning to account for a swarm as a carrier and improves KV-cache reuse and goodput by tracking user mobility along with the state of neighbouring nodes [2]. Semantics have reached the physical layers of communication and control as well. Reference [3] defines and prioritizes the semantic areas of path planning to ensure that air-ground cooperation can maintain the planning accuracy of the links under degradation; [4] encodes a global obstacle image for the leader and broadcasts it to the followers to achieve decentralized formation global awareness via local

sensing; [5] learns a translation model that maps percept symbols to executable trajectories and preserves semantic intent and physical feasibility in a single representation. The capacity of the sharing model in a swarm and training without centralisation of data is also being studied [6-7].

The Objectives of Communication and Education have not changed either. Swarm communication is still measured by rate, delay, packet loss, energy and spectral efficiency [8, 9]; the scheduler at the link layer needs high throughput and low latency, and a high compression ratio for the codec. Age of information has introduced task- and semantics-oriented variants to give new life to the objective [10-12], and now it is no longer about whether a packet has reached its destination but how current the destination's knowledge is at the link level. However, the above indicators only show the mode of transmission. A change in the risk level from low to high can be observed, so a large number of frames with no visual change of the ground do not need to be kept and consume 1000 times more bandwidth.

Formation Control Module is not available. According to consensus theory and graph theory, multi-agent formation has been developed [13], and now it addresses event triggering, topology switching and time-varying shapes [14,15]; heterogeneous composition and topology maintenance in low-altitude operations introduce capability constraints and link quality directly into online decisions [16, 17]. All the above ways are for solving the problem of formation goals geometric quantities: desired spacing, shape, or coverage. When the viewports of the swarm do not meet the semantic judgment requirements of the task—for example, if all the platforms are looking down on the same facade from the same side and cannot distinguish between shadows and structures — a formation that is geometrically perfect will systematically increase the error. The organisation of the formation should be based on what the swarm needs to know, not on how far apart the aircraft are.

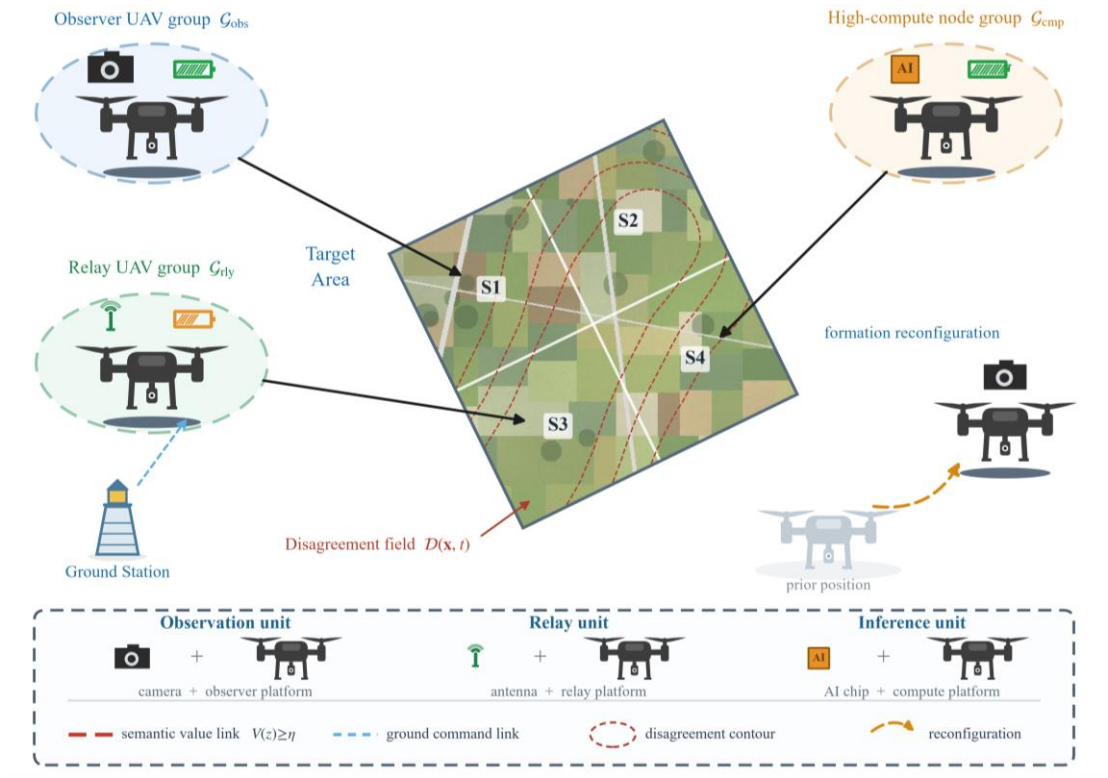

Fig. 1. The CVSC-DFR task scenario. Observation UAVs $\mathcal{G}_{\mathrm{obs}}$ occupy the upper left, relay UAVs $\mathcal{G}_{\mathrm{rly}}$ the lower left and high-compute nodes $\mathcal{G}_{\mathrm{cmp}}$ the right; together they cover the target area in the centre. The area is taken from low-altitude satellite imagery and divided by sub-task into four sub-regions S1–S4, and the red dashed contours trace the cognitive disagreement field $D(x,t)$ along the shadow corridor. Black arrows from the three groups towards S1, S3 and S4 mark semantic value links that satisfy $\mathrm{V}(z) \geq \eta$; the blue dotted line is the ground-station command link; the orange dashed arrows show a platform manoeuvring towards a complementary viewpoint, i.e. reconfiguration, with the pale grey platform and dashed footprint at lower right marking its position before the move. The bottom panel lists the three roles of a cognitive node (camera with observation platform, antenna with relay platform, compute chip with high-compute platform) and the legend.

Communication and formation are considered to be the same idea; therefore, a joint framework of cognitive value-aware semantic communication (CVSC) and disagreement-driven formation reconfiguration (DFR) has been proposed. CVSC has changed and therefore triggered. A UAV does not need to wait for an update signal of observation changes, a timer expiry, or a full queue; instead, it predicts how beneficial the message will be to the decision that the large model is about to make, transmits only when this value exceeds a certain threshold, and then selects the content based on this value—such as key visual tokens, object relations, event summaries, confidences, counterfactual judgments, etc. Only the DrivingAmount of DFR is altered. When the swarm disagrees on the risk in a certain area, it does not need to wait for consensus or be led by a central node; instead, it can reposition observation platforms, relay structures, high-compute nodes, task partitions, and target-confirmation geometry to provide the physical arrangement of information required to resolve the disagreement. Together, they will decide what to talk about and where to obtain information. Figure 1 is the loop of a typical low-altitude sensing task.

Four Follow-up Actions:

(1) A transmission trigger built for large-model decisions. The criterion moves from a change in the message to a change in decision value. We define an information value function $V(z) = \lambda_1 \Delta P_{\mathrm{task}} + \lambda_2 \Delta C_{\mathrm{cog}} + \lambda_3 \Delta R_{\mathrm{risk}} - \lambda_4 \tau(z)$ and make all four terms computable: the task increment follows from the KL divergence between successive policy distributions, the consistency increment from the drop in the entropy of the cognitive state, the risk increment from the mutual information of the risk posterior, and the cost from a weighted sum of delay, energy and bandwidth.

(2) The Concept of Cognitive Freshness and Its Evolution Equation. AoI is the age of the data; cognitive freshness is how far the existing conclusion has deviated from the current decision. The speed of decay is affected

by the circumstances of the event, drift in the decision context, and strength of supporting evidence; therefore, an old but stable prior, such as terrain obstacles, and a new but fleeting observation are given different weights.

(3) Joint optimisation of what is transmitted and how the model reasons. Key-token selection becomes monotone submodular maximisation with a cardinality constraint and admits a $1 - 1/e$ guarantee for the greedy rule. Multi-precision semantic coding is posed as a three-way trade-off among rate, semantic distortion and task performance, and the Lagrangian dual of the rate–distortion function yields a water-filling allocation in closed form; the transmitted content also carries counterfactual judgements, which lets the inference path be pruned.

(4) Disagreement-driven Reconfiguration and its Solver. Build a disagreement field and assign it to perceptual, viewpoint or timeliness differences; then, jointly move observation nodes, relay nodes and task partitions under connectivity constraints to achieve bounded stability via a Lyapunov function. The three stages of the mixed-integer non-convex problem for communication decisions, compression levels, positions and inference paths are alternations of optimisations, successive convex approximations and ADMM, along with convergence and complexity results.

Section 2: Review of Related Work and Research Gap. Section 3 introduces the network and task and cognitive-state models, defines information value, cognitive freshness and disagreement, and presents the joint problem. Section 4 develops the value estimator and its generalisation bound; Section 5 introduces the adaptive semantic communication mechanism; Section 6 presents the disagreement-aware reconfiguration method and its stability analysis; Section 7 develops the joint solver. Section 8 presents the experiments and Section 9 concludes.

2 Related Work

2.1 Swarm Communication and Networking

Swarm Communication Research has long studied Link Reliability and Spectral Efficiency. For air-ground channels, line-of-sight probability and height-dependent path loss are often used as the basis for multi-objective optimisation of trajectory, power and delay in hierarchical aerial access networks [9]; research has also been conducted to add emergency scenarios, multi-hop connectivity and fleet size to a single cost function to achieve extended seamless coverage with a reduced number of platforms [18]. Based on a survey of autonomous swarms in networking and resource management, the design dimensions of autonomy, resource management, multiple access and routing have been proposed [19], and a cognitive dynamic system has been introduced to organise swarm management via a sense-cognition-reconfiguration loop [20]. Key-node identification [21] and state monitoring and tamper defence [22] provide observability and security for network design, and task-driven interaction over a digital twin improves the timeliness of information by selecting between physical and virtual links [23]. Model sharing and cooperative inference have been gaining increasing attention in the field of computing: Coded distributed training reduces stragglers through redundant coding [24], and early exit with task offloading distributes inference across devices, the edge, and the cloud [25]. Topology preservation at a low elevation reduces link quality and causes a fault. Reference [17] has a topology-inference accuracy of over 90% with only beacon information in a GNSS-denied environment; [26] maximises routing utility for a cognitive swarm via Q-learning; and [27] can restore formation after leader failure using Raft-style voting.

The optimisation variables in this literature are links and resources, and the objectives are rate, delay, and coverage; surveys of swarm systems have also treated what to transmit as a given input [28-29]. Content-level redundancy is therefore disregarded. When the twenty UAVs arrive at the same judgment for a patch of airspace, the twenty-first frame contains little new information and still uses the same bandwidth as the first. Optimise the selection of content, and thus only need to make one decision for sending and selecting content.

2.2 Semantic and Task-Oriented Communication

Semantic Communication intends to transmit meaning rather than the exact values of bits. A deep semantic autoencoder for multi-UAV cooperation jointly trains the transceiver and task execution, and uses a semantic-similarity gate to decide whether to establish a link based on semantic consistency under poor imagery and tight bandwidth [30]; perception-enhanced multimodal

semantic communication strengthens the decisive features with attention and directs resource allocation according to semantic-importance feedback [31]. Therefore, the aim is to promote the application of reconstruction quality in actual operations, and corresponding triggers and compression policies will be introduced. According to Reference [3], there are required semantic regions for path planning and transmission. As shown in reference [12], to achieve semantic granularity for mobile users in the channel, the decision variable of compression depth is set by combining offloading ratio, trajectory and power. Reference [32] is a drift-plus-penalty problem with a power cost that combines semantic transmission and cooperative tracking control in decentralized MIMO fading channels, and closed-form solutions along with Lyapunov stability conditions have been obtained. Reference [4] substitutes high-dimensional image sharing with semantic broadcast. Semantic orchestration can optimise the extraction, transmission and computation of data over a long period to reduce the end-to-end delay of a task with high accuracy [11].

Therefore, the direction of the work has not been altered. It is generally periodic or buffer-based, and the compression level is fixed after offline training. When the task changes from object recognition to threat assessment of that object, the required semantic precision for the same frame is increased by several orders of magnitude, and a fixed-rate codec cannot be employed to take advantage of this change. CVSC sets the compression level and transmission of the value estimator to determine the coding granularity.

2.3 Information Value and Age of Information

The Age of Information provides a method to assess and plan for the decay of old information. Mean-field multi-agent forms of this work have been applied to timeliness in swarm data collection to determine trajectories and sampling schedules in mixed continuous-discrete action spaces [10], and a semantic freshness index has been proposed to combine semantic value and delay into a single objective for low-end-to-end delay under bandwidth constraints [11]. Radio-resource optimisation in cooperative sensing for the utility side combines AoI, power consumption and the delivery rate of cooperative-awareness messages into a single multi-task reward, and it has been shown that timeliness and task return can be traded within a single learning framework [8]; truth discovery in crowdsensing uses UAV reference data to adjust sensing credibility and expands the range of applicable information from the perspective of data quality [33].

Research has shown a decision value, but its value function is generally restricted to the reward of the immediate task or combined with age. They do not model how a message changes the cognitive state of the swarm, and they are not related to formation, which is a physical degree of freedom. The two advantages of our formulation are as follows: first, the value is a weighted sum of the variational distance between the decision distributions and the mutual information of the risk posterior, and it is computable, learnable, and suitable for a generalisation bound; second, the value signal drives both reconfiguration and transmission.

2.4 Dynamic Formation Control

The first purpose of the formation control is to maintain a certain shape and reach a particular goal. Consensus, potential fields and graph theory are the traditional frameworks; distributed assignment and optimisation can reach a stable state in a short time and adapt smoothly to changes in the number of vehicles [13]. Engineering-driven work includes switching control based on event triggering [14], formation switching in the presence of DoS attacks and vehicle loss [15], and environment-aware dynamic constraints under the influence of an external attacker [34]. At the level of control synthesis, distributed optimal tracking of the position-attitude loop can use an LQR index in conjunction with a disturbance observer for a relatively low computational complexity [35], or partition-based distributed predictive control can reduce the computation load of a large swarm by dividing it into subgroups [36].

Shape generation and reconfiguration are relatively simple and intuitive in physical space. A molecular-dynamics model of attraction, repulsion and alignment gathers holding, obstacle avoidance and splitting into a single set of virtual forces to rearrange the topology of the swarm without central coordination [37]; a magnetic reconfiguration protocol uses Hungarian assignment with virtual repulsion to change the configuration of hundreds of aircraft with almost no collisions [38]. Task-oriented formation formation work sets the goal, and then cooperative containment of multiple maneuvering

targets is achieved by fixed-time observers with auction mechanisms for dynamic regrouping [39]; heterogeneous air-ground formations maintain their shape through narrow passages with rigid graphs and adaptive safety potentials [40]; multi-region coverage keeps the formation stable via orthogonal potentials and heuristically orders region visits [41]; and a joint assignment and trajectory framework ensures both optimal assignment and feasible trajectories in the presence of dense obstacles [42].

Multi-agent reinforcement learning can learn a joint policy through local observations at the algorithmic level [43]. Hierarchical attention reduces the state and action spaces of large-scale confrontation [44]; evolutionary multi-objective policy optimisation produces a set of policies with different preferences in a single pass [45]; and evolutionary operators combined with attention simplify the policy cycles of self-play [46]. Cooperative decision-making for multi-target tracking adds an evolutionary mechanism to speed up learning [47] and a separate safety network to filter dangerous behaviour, thus balancing tracking success with the prevention of inter-agent collisions [48]; target search models trajectories, recharging, offloading and compute allocation together [49]. Adversarial Scenarios have constructed an attack-defense Stackelberg game for distributed predictive control [50], and conflict-free solutions for task allocation have been achieved by means of self-attention and command-priority resolution [51].

The advantages of these methods are that the desired structure can be specified geometrically or achieved according to the needs of life, such as coverage and connection. If the bottleneck is in cognition rather than geometry, for example, how to resolve a disagreement about the risk in a certain area, then a geometrically optimal configuration does not need to be informative. The purpose of the formation is to be a function of the disagreement field and thus reduce cognitive uncertainty.

2.5 Distributed Large-Model Inference and Edge Intelligence

The general trend now is to use a large model as the inference core of a swarm, but both sides of deployment are limited by on-board computing and link bandwidth. One line is model compression and cooperative inference: early exit with task offloading distributes delay-constrained inference across local, edge and aerial compute centres to maximise accuracy under limited compute [25]; coded distributed computing cancels stragglers with redundant coding, enabling a swarm to train models larger than any single platform can hold [24]. The other is a network-level optimisation of inference. An embodied agent addresses deployment, association, transmission scheduling and compute allocation at two time scales [1]; a quantum-enhanced multi-agent actor-critic introduces KV-cache reuse and inference routing in swarm decisions to increase goodput for effective tokens [2]; edge caching designs trajectory and cache jointly with multi-agent federated reinforcement learning to reduce the long-run cost of content updates [52]. Adaptive swarm intelligence is used for siting and computing to balance coverage with reduced redundancy and energy, and a lightweight solver is provided for aerial edge nodes [53]; a swarm framework is employed in dynamic environments to maximise the computation rate as user positions and demands change [54].

The Interface of Inference and Communication is also needed. The Semantic Broadcasting Organises the world's observations in an organised way that is suitable for distribution to followers [4], and a learned translation model directly maps percept symbols to executable actions [5]; both aim to achieve inference rather than raw data. Graph attention with proximal policy optimisation is used in the network to determine topology and offloading ratio simultaneously; convex approximations are then performed to allocate bandwidth and power [55], and it has been shown that topology, offloading and resources can be coordinated in a single learning framework, and grouping and role assignment for heterogeneous swarms explicitly separate communication nodes from compute nodes to avoid the delay caused by multi-hop forwarding [56]. Based on a survey of distributed machine learning, four frameworks have been identified: federated learning, multi-agent reinforcement learning, distributed inference and split learning; at present, the fusion of semantic communication with distributed inference is being explored to improve bandwidth efficiency [7].

Based on the above results, the swarm has good semantic reasoning ability, but reasoning and communication are not integrated; furthermore, the depth, branching and sampling of the inference path do

not consider communication cost, and the transmitted content is not organised for pruning. Swarm intelligence surveys have also identified the same open problems, and it has been pointed out that there is no unified framework for cross-layer optimisation with real-time guarantees [29,57]. When a swarm needs to perform a cross-linking operation that requires inference based on evidence from multiple hops, this separation will be computationally expensive and bandwidth-intensive. The Inference Path can be taken as the optimisation variable and solved jointly with communication.

2.6 Research Gap and Position

Four gaps have emerged: (i) The transmission trigger is still based on information change or elapsed time, and a computable value function for large-model decisions has not been established, nor has a generalisation bound been obtained for it. (ii) Semantic coding precision is fixed offline, and the three-way rate-semantic-distortion-task trade-off has no closed-form solution. (iii) Reconfiguration is driven by geometric quantities rather than cognitive quantities, and neither the mechanism mapping disagreement to configuration nor a stability proof has been provided. (iv) Communication decisions, compression levels, formation geometry and inference paths belong to different optimisation traditions, and a unified mixed-integer non-convex model and a distributed solver have not been developed. CVSC-DFR can meet the above four needs. Methodologically, it is based on evidential fusion, distributed alternating optimisation and submodular maximisation; all of these have been validated in cooperative sensing [8] and distributed predictive control [36], and the hierarchical account of swarm intelligence provides the rationale for combining them at different levels [58].

3 Network, Task and Cognitive Models

3.1 Dynamic Network and Channel Model

Consider a swarm $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$ of N heterogeneous UAVs carrying out cooperative sensing and risk assessment over a low-altitude area $\Omega \subset \mathbb{R}^3$. The position of platform u_i at time t is written $p_i(t) = [x_i(t), y_i(t), h_i(t)]^\top$ and its velocity $v_i(t)$. Communication in slot t is described by an undirected graph $G(t) = (\mathcal{U}, \mathcal{E}(t))$, in which the edge $(i,j) \in \mathcal{E}(t)$ exists if and only if the received signal-to-noise ratio between the two platforms is at least the demodulation threshold. The air–ground link uses a line-of-sight probability model to represent blockage, and the elevation angle between u_i and u_j is

$$\theta_{ij}(t) = \frac{180}{\pi}\arcsin\left(\frac{h_i(t)-h_j(t)}{d_{ij}(t)}\right), d_{ij}(t) = \|p_i(t) - p_j(t)\|_2 \quad (1)$$

where $d_{ij}(t)$ is the separation between platforms and $h_i(t)$ the height; (1) turns the height difference into the elevation angle that enters (2) as its argument. In suburban and semi-structured environments the line-of-sight probability is well approximated by a sigmoid

$$P_{\mathrm{LoS}}(\theta_{ij}) = \frac{1}{1+a_e \exp[-b_e(\theta_{ij}-c_e)]} \quad (2)$$

where a_e, b_e, c_e is a fitting parameter tied to the environment type (building density, mean height, vegetation cover). Weighting the two components by their probabilities gives the equivalent channel gain

$$g_{ij}(t) = [P_{\mathrm{LoS}}\eta_L + (1 - P_{\mathrm{LoS}})\eta_N] g_0 d_{ij}^{-\alpha}(t) |\tilde{h}_{ij}(t)|^2 \quad (3)$$

where η_L and η_N are the excess attenuation of the line-of-sight and non-line-of-sight components, g_0 the channel gain at the reference distance, α the path-loss exponent and $\tilde{h}_{ij}(t)$ the small-scale fading, a complex Gaussian variable of unit variance. Eq. (3) shows how a small change in line-of-sight probability is amplified into a large change in equivalent gain, through the wide gap between η_L and η_N; this is the physical reason links fail so often along shadow corridors. The achievable rate from u_i to u_j follows as

$$R_{ij}(t) = B_{ij}(t)\log_2\left(1 + \frac{P_t g_{ij}(t)|\tilde{h}_{ij}(t)|^2}{\sigma^2 \Gamma}\right) \quad (4)$$

where $B_{ij}(t)$ is the bandwidth given to edge (i,j), P_t the transmit power, σ^2 the noise power and Γ the signal-to-noise gap of the modulation and coding scheme. The bandwidth variable $B_{ij}(t)$ in (4) belongs to the communication decisions of the joint problem, and its value is limited by the total bandwidth and by semantic priority.

3.2 Energy and Delay

The three components of the energy consumption for a UAV per slot are flight, computation and communication.

$$E_i(t) = (\kappa_1 v_i^3(t) + \frac{\kappa_2}{v_i(t)})T_s + \epsilon_c \phi_i(q_i(t))T_s + \frac{P_t}{R_{ij}(t)} b_{ij}(t) \quad (5)$$

where $\kappa_1 v_i^3$ is the induced-drag power term, κ_2 / v_i the hover and lift term that varies inversely with speed, T_s the slot length, ϵ_c the computation-energy coefficient, $\phi_i(q_i(t))$ the floating-point operations per unit time required by compression level $q_i(t)$, and $b_{ij}(t)$ the number of transmitted bits. The three are flight, computation and communication; as the latter is directly related to rate allocation, the choice of coding precision in Section 5 will affect both communication energy and inference delay. End-to-end delay has five parts, sensing, encoding, transmission, propagation and inference:

$$\tau_{ij}(t) = \tau_{\text{sen}} + \tau_{\text{enc}}(q_i(t)) + \frac{b_{ij}(t)}{R_{ij}(t)} + \tau_{\text{prop}} + \tau_{\text{inf}}(\mathcal{P}_i(t)) \quad (6)$$

where τ_{sen} is the sensor exposure and readout delay, $\tau_{\text{enc}}(q_i)$ the semantic encoding delay (which grows with the compression level, because a lower rate demands heavier quantisation and entropy coding), τ_{prop} the propagation delay and $\tau_{\text{inf}}(\mathcal{P}_i)$ the inference delay, whose argument is the inference path $\mathcal{P}_i(t)$—whether a counterfactual branch is invoked, for instance. The model keeps both feasible compromises in view, transmitting less and reasoning longer, or transmitting more and reasoning less; that is the space in which the joint optimisation of Section 7 operates.

3.3 Task State and Swarm Cognitive State

The task is modelled as a set of interdependent critical sub-tasks. Let $\mathcal{K}_{\text{crit}} = \{1,2,\dots,K_c\}$ Be the set of necessary sub-tasks, k a sub-task, t the time, and $p_k(t)$ the probability that the sub-task has been completed correctly by then. The success probability of the task at time t is

$$P_{\text{task}}(t) = 1 - \prod_{k \in \mathcal{K}_{\text{crit}}} [1 - p_k(t)]^{\varpi_k}, \varpi_k \in (0,1] \quad (7)$$

where ϖ_k is the weight of sub-task k and indicates its necessity in the task chain. A product is used instead of a sum because there are failures in the sub-tasks; a single failure will reduce the total probability significantly, and only a product takes into account this weakest-link effect. Eqs. (7) and (8) show the degree of task completion and the level of swarm understanding, respectively, and the first two terms of (10) are a product of these two.

The cognitive state of the swarm is modelled as a belief over semantic classes held on each platform. Let $\mathcal{K} = \{1,2,\dots,K\}$ be the set of classes (low risk, suspicious, high risk, confirmed threat) and u_i the local belief of a platform, a vector on the simplex

$$b_i(t) = [b_i^1(t),\dots,b_i^K(t)]^\top \in \Delta^{K-1}, b_i^k(t) \geq 0, \sum_{k=1}^K b_i^k(t) = 1 \quad (8)$$

where Δ^{K-1} is the $K-1$ -dimensional probability simplex and $b_i^k(t)$ the posterior probability that u_i assigns to class k. The local belief comes from the on-board model given its own observation and context, and later evidence is fused into it by Bayes' rule; the simplex constraint in (8) keeps it a valid distribution and excludes the negative probabilities that numerical error would otherwise produce. The swarm-level state is the output of an admissible fusion operator $\mathcal{T}$:

$$b^c(t) = \mathcal{T}(\{b_i(t)\}_{i=1}^N) \quad (9)$$

where $\mathcal{T}$ is an evidence-weighted fusion operator whose weights account for platform reliability, semantic freshness and observation geometry, specified in Section 5.6. Eq. (9) feeds directly into (13). The swarm state is not an arithmetic mean; for instance, if two aircraft observe the same facade from the same side, the count will be doubled by one source and manufactured confidence. Therefore, we take $\mathcal{T}$ to be a log-pooling operator that is sensitive to correlation. Prediction in the swarm also considers context: locally, an embodied agent can observe changes in the environment and the state of neighboring agents to make future plans [1], and a learned translation model can directly convert perception symbols into executable actions [5]; together, they enable answering the question of how a decision would change given a new message without new observations. This way, the value function of Section 4 can be calculated.

3.4 Information Value and Cognitive Freshness

The central notion is information value. Let z be a candidate message (an encoded frame, a set of key tokens, an event summary, a counterfactual judgement), s the current decision context and $\mathcal{H}$ the history.

A unit of information content.

$$V(z) = \lambda_1 \Delta P_{\text{task}}(z) + \lambda_2 \Delta C_{\text{cog}}(z) + \lambda_3 \Delta R_{\text{risk}}(z) - \lambda_4 \tau(z) \quad (10)$$

$\Delta P_{\text{task}}(z)$ is the gain in task success probability after receiving message z, $\Delta C_{\text{cog}}(z)$ is the gain in cognitive

consistency, $\Delta R_{\text{risk}}(z)$ is the gain in risk discrimination, $\tau(z)$ is the cost of transmitting and processing message z, and $\lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$ are trade-off coefficients. Section 4 presents the calculable form of all terms. It is not the same as the previous method; it does not depend on the magnitude of z or how much it has been altered, but rather on how far z shifts the decision and belief distributions. Therefore, a message that can reduce the risk assessment by a small amount is more practical; a large-scale fulfilment of expectations is not very effective.

Values should also be appropriate for the time. We define cognitive freshness $\Phi(z,t)$, the degree to which historical information z remains credible in the current decision context, as a function of evidential support and cognitive drift:

$$\Phi(z,t) = C_{\text{cog}}^{\text{supp}}(z,t) \cdot \exp[-\int_{t_z}^{t} \rho_{\text{cog}}(\varsigma) d\varsigma] \quad (11)$$

where $C_{\text{cog}}^{\text{supp}}(z,t) \in [0,1]$ is the cognitive support that message z receives (from the number of platforms backing the conclusion, the complementarity of their viewpoints and the independence of their evidence), t_z the moment z was generated and $\rho_{\text{cog}}(\varsigma) \geq 0$ the cognitive drift rate. Age of information sees only $t - t_z$. The support term in (11) lets information that is old but corroborated keep its freshness, and the exponential term makes information that is new but unconfirmed decay quickly.

The drift rate is also a function of the decision context and can be modelled as a baseline plus a contextual term:

$$\frac{d\Phi(z,t)}{dt} = -[\rho_0 + \rho_1 D_{\text{KL}}(\pi_t \| \pi_{t_z})]\Phi(z,t) + \chi m(z,t) \quad (12)$$

where ρ_0 is the decision-independent baseline drift caused by target motion and environmental change, $D_{\text{KL}}(\pi_t \| \pi_{t_z})$ the KL divergence between the current policy distribution and the one at the moment the message was generated, which measures how far the decision context has moved, ρ_1 its sensitivity coefficient, $\chi m(z,t)$ a refresh term and $m(z,t)$ the number of independent supporting items added per unit time. Two regimes follow from (12). While the context is stable ($\pi_t \approx \pi_{t_z}$) old information holds its value; when the context shifts abruptly, as when a cruise task turns into a search, the freshness of the entire history decays at once and a burst of semantic refresh is triggered. Communication then follows the task phase instead of a fixed period.

3.5 Cognitive Dissonance

Cognitive disunity is the separation of platform nodes in a particular area of the swarm. Conflicts that have already been stored in memory and weak inconsistencies that have not reached the threshold are included. Kernel-weighted disagreement field at a specific position is

$$\mathcal{D}(x,t) = \frac{\sum_{i,j} w_{ij}(x,t) \| b_i(t) - b_j(t) \|_1}{2 \sum_{i,j} w_{ij}(x,t)},$$

$$w_{ij}(x,t) = \exp(-\frac{\|p_i(t)-x\|^2}{2\sigma_d^2}) \exp(-\frac{\|p_j(t)-x\|^2}{2\sigma_d^2}) \quad (13)$$

where $x \in \Omega$ is the spatial position, σ_d the influence radius of the field, $w_{ij}(x,t)$ the joint spatial weight of platforms i and j with respect to x, and the factor 2 in the denominator normalises the L_1 distance to $[0,1]$. Three properties follow. Disagreement is continuous in space, so a gradient exists and can be used to decide where to look. It is based on the difference in beliefs rather than their absolute values; thus, a swarm that is confidently and unanimously incorrect is not considered low-disagreement. The kernel is far away from the region of decision for distant disagreements.

Uniformity can also be shown in numbers.

$$\mathcal{H}_{\text{cog}}(t) = -\sum_{k=1}^{K} \overline{b}^{k}(t) \log \overline{b}^{k}(t), C_{\text{cog}}(t) = 1 - \frac{\mathcal{H}_{\text{cog}}(t)}{\log K} \quad (14)$$

where $\overline{b}^{k}(t)$ is the kth component of the fused swarm belief, $\mathcal{H}_{\text{cog}}(t)$ the semantic entropy of that belief and $C_{\text{cog}}(t) \in [0,1]$ a normalised consistency index. When all platforms agree $C_{\text{cog}} = 1$; when the belief is uniform over K classes $C_{\text{cog}} = 0$. The two quantities are complementary: $\mathcal{D}(x,t)$ describes local pairwise inconsistency, $C_{\text{cog}}(t)$ the global concentration of the distribution.

3.6 Joint Problems and Constraints

Let $\alpha_{ij}(t) \in \{0,1\}$ be the communication decision in slot t (whether a message is sent over edge (i,j)), $q_i(t) \in \mathcal{Q}$ the semantic compression level, $p_i(t)$ the platform positions and $\mathcal{P}_i(t)$ the inference path. The joint problem is

$$min_{\{\alpha_{ij}, q_i, p_i, \mathcal{P}_i\}} \mathbb{E}[\sum_{t=1}^{T} (\gamma_1 \Omega(t) + \gamma_2 \overline{\tau}(t) + \gamma_3 E(t) + \gamma_4 \overline{\mathcal{D}}(t))] \quad (15)$$

where $\Omega(t) = \sum_{i,j}\alpha_{ij}(t)|z_{ij}(t)|$ is total traffic, $\overline{\tau}(t)$ mean end-to-end delay, $E(t) = \sum_i E_i(t)$ total energy, $\overline{\mathcal{D}}(t)$ the area-weighted mean disagreement and $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ the trade-off coefficients after normalisation. The constraints cover bandwidth and energy budgets, admissible compression levels, the flight envelope, connectivity and the inference budget:

$$\sum_{j\in\mathcal{N}_i(t)}\alpha_{ij}(t)b_{ij}(t) \le C_i(t),$$
$$q_i(t) \in \mathcal{Q} = \{q^{(1)}, q^{(2)}, \dots, q^{(Q)}\}, p_i(t) \in \Omega_{\text{safe}} \quad (16)$$

where $\mathcal{N}_i(t)$ is the neighbour set of u_i, $C_i(t)$ the bit budget of the slot, $\mathcal{Q}$ the discrete set of compression levels ($Q = 4$ here) and Ω_{safe} the flyable region set by minimum safe altitude, no-fly zones and collision-avoidance distance. Connectivity and the on-board inference budget add two more:

$$\lambda_2(L(t)) \ge \lambda_{min}, \sum_i |\mathcal{P}_i(t)| \le \Pi_{max}, E_i(t) \le E_i^{max}, \forall i \in \mathcal{U} \quad (17)$$

where $L(t)$ is the Laplacian of the communication graph, $\lambda_2(\cdot)$ its second-smallest eigenvalue (the algebraic connectivity), λ_{min} the connectivity threshold, $|\mathcal{P}_i(t)|$ the inference-path length and Π_{max} the total compute budget. Problem (15) is stochastic, mixed-integer, non-convex and coupled: α_{ij} is binary, q_i discrete, p_i enters the constraints and the objective nonlinearly through channel gain and coverage, and $\mathcal{P}_i$ couples into the objective through delay and value. No analytical global optimum exists. Fig. 2 shows the route taken here.

Section 4 sets the stage for building a computable objective function; Section 5 lists the elements and levels of precision; Section 6 performs the restructuring; and Section 7 connects the three-tier solver.

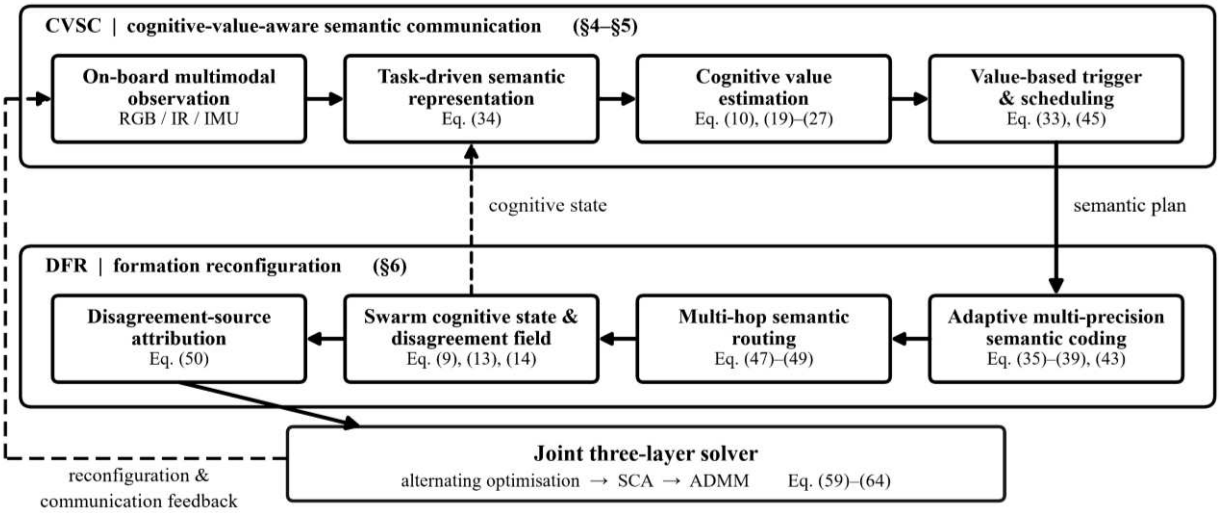

Figure 2. Overall Framework of CVSC-DFR. The upper layer determines the cognitive value and begins transmission (Sections 4 and 5), and the lower layer modifies the formation based on disagreement (Section 6). The two are connected by a loop of swarm cognitive state and disagreement field. The three-layer decomposition of the joint problem and the method at each layer (Section 7) are shown in the lower figure.

4 Cognitive Value Estimation

The four terms in equation (10) are converted into computable quantities in this section, and then the structure of the value-prediction network, its training, and a generalisation bound for the estimation error are introduced. All four terms have the same requirement: they must be approximated in a single forward pass of the board, otherwise, the value estimation will not be real-time and thus impractical.

4.1 Effect of a message on the task decision

The output of a large model or policy network is a distribution over actions. With context s and policy parameters θ, that distribution is

$$\pi_\theta(a|s) = \frac{\exp([f_\theta(s)]_a/T)}{\sum_{a'\in\mathcal{A}}\exp([f_\theta(s)]_{a'}/T)} \quad (18)$$

where $f_\theta(s) \in \mathbb{R}^{|\mathcal{A}|}$ is the action-score vector produced by the policy head, $\mathcal{A}$ the discrete action set (keep cruising, change heading, report a target, request re-observation, shrink the task area) and T a temperature. When message z arrives, the context moves from s to $s \oplus z$, where $\oplus$ denotes concatenating or injecting the

semantic embedding of z into the context representation, and the distribution shifts with it. We define the effect of the message on the decision as the KL divergence of that shift passed through a monotone map, with π_θ taken from (18):

$$\Delta P_{\text{task}}(z) = \varrho[D_{\text{KL}}(\pi_\theta(\cdot\,|s \oplus z)\|\pi_\theta(\cdot\,|s))] \quad (19)$$

where $\varrho: [0,\infty) \to [0,1]$ is an increasing link function that maps the size of the distribution shift onto a gain in success probability. The concrete form of ϱ is obtained by fitting the calibration data; we use the saturating exponential $\varrho(u) = 1 - \exp(-u/u_0)$, where u_0 is the calibration scale. Reading of (19) is as follows:A message that visibly changes what the swarm will do next carries high decision value; a message the context absorbs with hardly any change in the action distribution is redundant at the decision level, however different it looks at the pixel level.

Due to the limitations of the small on-board hardware, Evaluation (19) cannot be implemented; an extra forward pass of the policy network would be required. When the perturbation due to z is small, the second-order approximation of the divergence is given by the Fisher Information Matrix:

$$D_{\text{KL}}(\pi_\theta(\cdot\,|s \oplus z)\|\pi_\theta(\cdot\,|s)) \approx \frac{1}{2} g_z^{\top} F_\theta(s) g_z \quad (20)$$

g_z is the gradient of lo$\partial \log \pi_\theta(a|s)/\partial z$ with respect to the message embedding, averaged over actions, and $F_\theta(s)$ is the Fisher information of the policy distribution with respect to embedding parameters.A single backward pass can be employed to replace the two forward passes; although the measured value-estimation delay is approximately 0.7ms, it exhibits an underestimation problem in the presence of large perturbations. A two-threshold rule guards against that bias: when the approximation in (20) lands near the trigger threshold, the exact two-pass computation takes over.

4.2 Effect of a message on the safety judgement

Safety value cannot be read off an action distribution, because safety events are sparse. Not taking an avoidance manoeuvre is the right decision almost all of the time, so the action distribution hardly reacts when risk changes. We measure the effect on the safety judgement through the posterior over risk instead. Let $r \in \{0,1\}^K$ be the risk-class indicator vector, $\mathcal{H}$ the history and z the new message; the risk value of the message is its mutual information with respect to r:

$$\Delta R_{\text{risk}}(z) = \mathbb{E}_z[\log \frac{p(r|z,\mathcal{H})}{p(r|\mathcal{H})}] = I(r; z|\mathcal{H}) \quad (21)$$

where $p(r|z,\mathcal{H})$ is the risk posterior after z is admitted, $p(r|\mathcal{H})$ the prior, and the expectation runs over the generating distribution of z. The form of mutual information is also sensitive to the prior. When the prior is already confident that there is no risk, an absolute change of 0.01 in the posterior is still a relatively large value; it is even larger near a prior of 0.5. Risk management is implemented in practice as follows: The marginal value of information from a high-risk area exceeds that of the same information in a low-risk area. By the chain rule, (21) is the expected reduction in posterior entropy.

$$\Delta R_{\text{risk}}(z) = H(r|\mathcal{H}) - H(r|z,\mathcal{H}) \quad (22)$$

where $H(\cdot\,|\,\cdot)$ is a conditional entropy. The value of (22) is that a variational bound suffices: if the risk posterior is approximated as Gaussian and observation z is treated as a noisy linear measurement of the risk state, the entropy drop is available analytically as

$$\Delta R_{\text{risk}}(z) \approx \frac{1}{2} \log|I + \Lambda_z \Sigma_{\text{prior}}| \quad (23)$$

where I is the identity matrix, Σ_{prior} the prior covariance of the risk state and Λ_z the information that message z contributes about it, namely $\Lambda_z = J_z^{\top} \Sigma_{\text{obs}}^{-1} J_z$, with $J_z = \partial r/\partial z$ the observation Jacobian and Σ_{obs} the measurement-noise covariance. The expression follows from the information form of the Gaussian posterior: once z arrives the covariance becomes $\Sigma_{\text{post}} = (\Sigma_{\text{prior}}^{-1} + \Lambda_z)^{-1}$, and substituting into the entropy difference $\frac{1}{2}\log|\Sigma_{\text{prior}}| - \frac{1}{2}\log|\Sigma_{\text{post}}|$ with the determinant product rule gives the result. This is the log-determinant criterion of D-optimal experiment design: the message that most reduces uncertainty about the risk state is the one worth sending.

4.3 Effect of a Message on Cognitive Consistency

The effect of the message on consistency can be shown as a decrease in the semantic entropy of the swarm. From (14) and (13), the increased consistency is

$$\Delta C_{\text{cog}}(z) = \frac{1}{\log K}[\mathcal{H}_{\text{cog}}(t) - \mathcal{H}_{\text{cog}}(t|z)] \quad (24)$$

where $\mathcal{H}_{\text{cog}}(t|z)$ is the semantic entropy after the swarm has absorbed message z. To make the term

computable the fused belief is modelled as a Dirichlet posterior $\mathrm{Dir}(\alpha)$, with α_k the equivalent evidence count for class k. Accepting message z then adds $\eta(z)$ equivalent observations to one class, and the entropy becomes

$$\mathcal{H}_{\mathrm{cog}}(t|z) = \psi(\alpha_0^+ + 1) - \frac{1}{\alpha_0^+ + 1}\sum_{k=1}^{K}\alpha_k^+ + \psi(\alpha_k^+ + 1),$$

$$\alpha_k^+ = \alpha_k + \eta(z)\delta_{k,k(z)} \quad (25)$$

where $\psi(\cdot)$ is the digamma function and $\alpha_0^+ = \sum_k \alpha_k^+$; $k(z)$ is the dominant class supported by message z, $\delta_{k,k(z)}$ the Kronecker delta, and $\eta(z)$ the number of equivalent observations that z supplies, fixed by the independence of the evidence and the reliability of its source. Eq. (25) accounts for a behaviour we met repeatedly in the experiments. When the swarm is close to agreement (α concentrated), more evidence of the same kind results in almost no entropy decrease; when it is split (α uniform), any independent evidence for one class leads to a large increase. Thus, the consistency term is employed to send the information with the largest disagreement without an additional heuristic.

4.4 Communication Costs

The three resources of the cost term in (10) are delay, energy and spectrum.

$$\tau(z) = \underbrace{\frac{|z|}{R_{\mathrm{eff}}(t)}}_{\mathrm{delay}} + \varsigma_p \underbrace{\frac{P_t|z|}{R_{\mathrm{eff}}(t)E_i^{max}}}_{\mathrm{energy}} + \varsigma_s \underbrace{\frac{|z|}{B_{\mathrm{tot}}T_s}}_{\mathrm{spectrum}} \quad (26)$$

where $|z|$ is the bit length of message z, $R_{\mathrm{eff}}(t)$ the equivalent achievable rate of the current link (the bottleneck rate when the path is multi-hop), B_{tot} the total bandwidth available to the swarm, and ς_p and ς_s normalisation coefficients for energy and spectrum. The scale is comparable with the first three terms: compressing a message to a lower rate lowers all three costs together, so cheap messages are easier to justify. Note that R_{eff} in (26) depends on the formation and on relay selection, which couples the cost term to the reconfiguration of Section 6. Relocating the relay to a more advantageous position lowers the R_{eff} at all nodes in the multi-hop path and extends the range of high-value messages in the hazardous area.

4.5 Value-prediction Network and Training

Of the four terms in (10), ΔP_{task} needs a policy forward pass and ΔR_{risk} a posterior update; neither can be run per message on board in real time. We therefore train a lightweight value-prediction network V_ω that regresses value from the semantic embedding of message z and the current cognitive state. It has three stages: per-token encoding, set aggregation and a regression head.

$$V_\omega(z) = \phi_{\omega_2}\left(\underset{m\in z}{\mathrm{Agg}}\{h_{\omega_1}(x_m)\}\right), \underset{m}{\mathrm{Agg}}\{h_m\} = \sum_m \varsigma_m h_m \quad (27)$$

where x_m is the mth semantic unit of message z (a token, a bounding box, a relation triple or a summary sentence), h_{ω_1} a shared lightweight encoder of two feed-forward layers with width 128, Agg an attention-weighted permutation-invariant aggregation, ς_m the normalised attention weights and ϕ_{ω_2} a scalar regression head. The output does not depend on input order, and cost grows linearly in $|z|$; value prediction takes 0.21 ms per message against 2.4 ms for the policy forward pass.

Labels are obtained from the precise evaluation of (10) and continuously generated in a bootstrapped way during the operation of the swarm.

$$V_{\mathrm{tgt}}(z) = r(z) + \xi\mathbb{E}_{s'}[max_{z'\in\mathcal{Z}(s')}V_{\omega^-}(z')] \quad (28)$$

where $r(z)$ is the decision gain observed within one step of admitting message z (extra targets confirmed, extra risk alarms hit), $\xi \in (0,1)$ a discount factor, $\mathcal{Z}(s')$ the candidate messages available at the next state s' and ω^- the parameters of the target network. Thus, the estimate is no longer a single-step heuristic but a multi-step return; it can now recognize messages that do not bring immediate gains but change what can be done later, such as a summary report that a corridor is blocked. Labels are generated online, and the training loss is

$$\mathcal{L}(\omega) = \mathbb{E}_z[(V_{\mathrm{tgt}}(z) - V_\omega(z))^2] + \mu\|\omega\|_2^2 + \nu\mathbb{E}_z[(V_\omega(z) - V_\omega(z_{\mathrm{aug}}))^2] \quad (29)$$

where μ is the weight-decay coefficient, ν the consistency-regularisation coefficient and z_{aug} a copy of z obtained by value-preserving semantic augmentation (a paraphrase within the same class, a reordering of tokens that does not alter object relations). The third term asks the prediction to be invariant to meaning-preserving transformations, which is what improves generalisation to unseen tasks. Since the task distribution drifts, the

network is fine-tuned on board under (29); and because the observation distributions of the platforms are not identically distributed, parameter updates follow the clustering and aggregation strategy of federated learning for heterogeneous swarms, which limits the degradation drift would otherwise cause [6]. The trigger threshold η is not tuned by hand but back-solved on a separate calibration set from a target false-trigger rate, so no prior is imposed on the task distribution. Fig. 3 shows the information flow and the training loop.

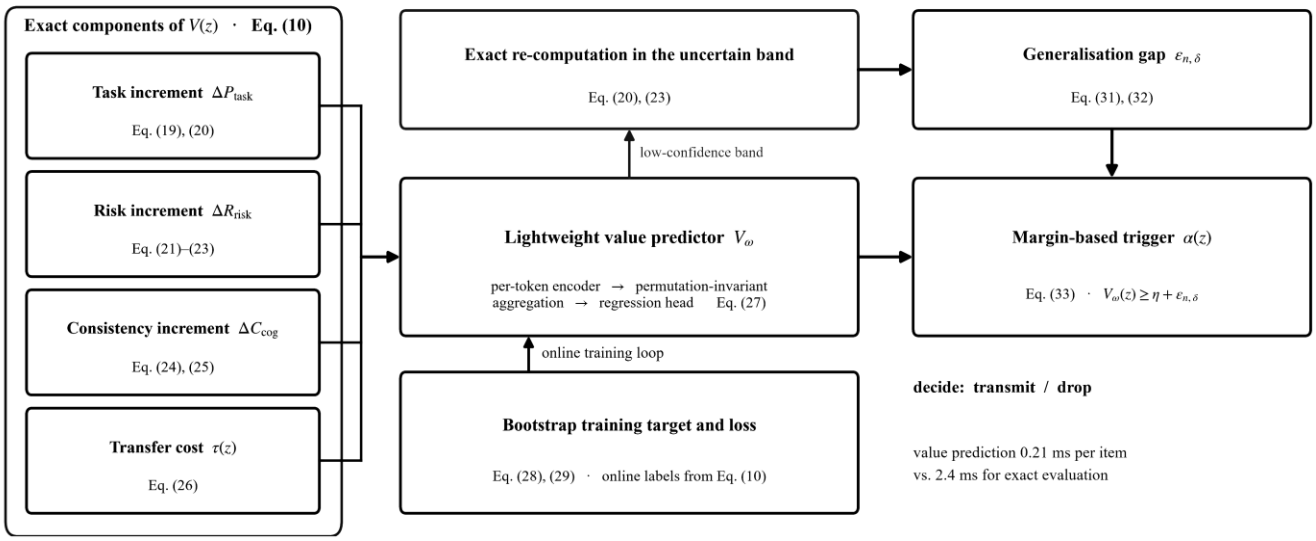

Figure 3. The Cognitive Value Estimation Network and Its Information Flow. The specific computation paths of the four value terms are shown on the left; in the middle is the lightweight prediction network (per-token encoding, permutation-invariant aggregation, regression head) and its bootstrapped training loop; and on the right is the margin-based trigger decision built on the generalisation gap. All components are labelled in the equations.

4.6 Estimation Error and Generalisation Bound

The reliability of the trigger rests on the estimation error, so we give a uniform convergence bound for it. Let $\mathcal{F}_\omega = \{z \mapsto V_\omega(z)\}$ be the class of value predictors, $\{z_i\}_{i=1}^n$ an i.i.d. sample, $V(z)$ the true value, $\mathcal{R}_n(f) = \frac{1}{n}\sum_i \ell(f(z_i), V(z_i))$ the empirical risk and ℓ an L-Lipschitz bounded loss. The empirical Rademacher complexity of the class is

$$\mathrm{Rad}_n(\mathcal{F}_\omega) = \mathbb{E}_\varsigma[sup_{f\in\mathcal{F}_\omega} \frac{1}{n}\sum_{i=1}^n \varsigma_i f(z_i)], \varsigma_i \sim \mathrm{Unif}\{-1,+1\} \quad (30)$$

where $\varsigma = (\varsigma_1, \dots, \varsigma_n)$ are independent Rademacher variables and sup runs over the class. The quantity bounds how well the class fits random label noise and does not depend on the sample, so it can stand in for generalisation error; (30) is its empirical estimate. For the architecture used here—linear encoding, permutation-invariant aggregation and a linear head—the complexity obeys a spectral-norm bound:

$$\mathrm{Rad}_n(\mathcal{F}_\omega) \leq \frac{2\sqrt{\log 2}}{C_\mathcal{F}\sqrt{n}} \left(\prod_{l=1}^L \|W_l\|_2\right)\left(\sum_{l=1}^L \frac{\|W_l\|_F^2}{\|W_l\|_2^2}\right)^{\frac{1}{2}} \sqrt{\sum_{i=1}^n \|z_i\|_2^2} \quad (31)$$

where L is the number of layers, $\|W_l\|_2$ and $\|W_l\|_F$ the spectral and Frobenius norms of the weight matrix at layer l, and $C_\mathcal{F}$ a constant tied to the Lipschitz constant of the activation. The bound links generalisation to the product of spectral norms and to the total energy of the input embeddings, which suggests two ways of regularising: constrain the spectral norms through weight decay, or constrain embedding energy by normalisation, which we apply to semantic units before aggregation (L_2 normalisation). From (31) and standard uniform convergence, for any $\delta \in (0,1)$ with probability at least $1-\delta$

$$sup_{f\in\mathcal{F}_\omega} |\mathcal{R}(f) - \mathcal{R}_n(f)| \leq 2L\mathrm{Rad}_n(\mathcal{F}_\omega) + 3B_\ell\sqrt{\frac{\log(2/\delta)}{2n}} \quad (32)$$

where $\mathcal{R}(f)$ is the true risk, B_ℓ an upper bound on the loss, and the right-hand side is written $\varepsilon_{n,\delta}$, the population generalisation gap. The first factor $2L$ comes from the contraction lemma for an L-Lipschitz loss, the second is the usual tail term. Eq. (32) supplies the reliability guarantee for the trigger. With true threshold η and trigger indicator $\alpha(z)$, the decision rule carries a margin:

$$\alpha(z) = \mathbf{1}\{V_\omega(z) \geq \eta + \varepsilon_{n,\delta}\} \quad (33)$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and $\varepsilon_{n,\delta}$ the gap from (32). The rule ensures that the transmitted message

has a true value of at least η with a probability of at least $1-\delta$, and thus the false-trigger rate is bounded. With $n = 2 \times 10^5$ training samples and the spectral norm bounded by 3.2, (31) gives $\varepsilon_{n,\delta} \approx 0.041$,

As shown in Figure 8, the absolute prediction error of the measurement is 0.043, and at this level, the bound is relatively small.

5 Adaptive Semantic Communication

5.1 Task-driven Semantic Representation

Based on the value signal from (10), this module determines various transmission parameters, including the content, accuracy, order and path of transmission. First, create a representation. Raw images are represented by pixels, which are highly redundant and unrelated to the task; fixed-rate semantic communication uses a uniform feature map, which removes some redundancy but still does not distinguish the units that matter for a decision. We use a task-driven structured representation, in which a message z is decomposed into four kinds of semantic unit,

$$z = \{(e_m, \varrho_m, r_m, a_m)\}_{m=1}^{|z|}, r_m = \mathbf{1}\{\text{relation}\}, \\ a_m = \mathbf{1}\{\text{counterfactual}\} \quad (34)$$

where e_m is the embedding of the mth unit (a patch token from the visual encoder, a box feature from the detector, a relation triple from the language side), $\varrho_m \in [0,1]$ its confidence, r_m a flag for whether the unit takes part in a relation between objects (vehicle A inside the mouth of corridor B) and a_m a flag for whether it carries a counterfactual judgement (if the object is not a heat source, the shadow explanation fails). The last two types are where this design is different from the old plan. Relation units preserve the topological semantics of targets and therefore do not need to perform relational inference for the receiver; counterfactual units directly support pruning by enabling the receiver to skip unprofitable branches. Visual units originate from the on-board semantic encoder, are arranged in multiple UAVs, a query is broadcast by the task aircraft, and then, based on semantic similarity, supporting aircraft decide whether to establish a link; feature consistency is maintained even in conditions of poor imagery [30], and for multimodal fusion, perception-enhanced attention weights the decisive features and a coarse classification guides the fusion [31]. The Backbones only provide modules. A unit will be sent if and only if it meets the value criterion.

5.2 Multi-precision Semantic Coding and the Rate-Task Trade-off

Once the set of units is fixed, each of them needs a rate. Too low destroys the semantics, too high wastes bandwidth. Distortion is measured semantically, as D_{sem}, rather than in pixels, through an exponential rate–distortion relation:

$$D_{\text{sem}}(R) = D_{min} + (D_0 - D_{min})2^{-\beta R} \quad (35)$$

where R is the rate given to the unit (bits per unit), D_0 the semantic distortion at zero rate, D_{min} the saturated lower bound on distortion, fixed by the intrinsic ambiguity of the unit (shadow and dark ground are hard to separate in a single frame), and β the decay coefficient. The form mirrors a classical rate–distortion function, but its parameters are fitted to calibration data for the semantic task, so different units end up with different D_{min} and β. Multi-precision coding corresponds to a discrete set of levels $\mathcal{Q}$, obtained by quantising the rates that (35) returns. A model for the map of semantic distortion and task performance is needed; therefore, a linear saturating function is used.

$$P_{\text{task}}(R) = P_{max}[1 - \kappa_s D_{\text{sem}}(R)] \quad (36)$$

P_{max} is the upper bound of the success probability at zero distortion, and κ_s is the coefficient for converting semantic distortion into task performance. It is valid only within a small range. Distortion is manifested as a few missed targets or broken relationships; however, the success rate gradually decreases, and beyond the saturation point ($\kappa_s D_{\text{sem}} \geq 1$), the task has failed completely, so P_{task} is set to 0. Together, Equations (35) and (36) form the three-way relationship of rate, semantic distortion and task performance.

Therefore, the goal is to maximise utility subject to a total-rate limit.

$$max_{\{R_i\}_{i=1}^N} \sum_i w_i P_{\text{task}}(R_i) - \lambda_t \sum_{i=1}^N R_i \\ \text{s.t.} \sum_{i=1}^N R_i \leq R_{\text{tot}}, R_{min} \leq R_i \leq R_{max} \quad (37)$$

where $w_i \in [0,1]$ is the value weight of the i th message (obtained by normalising (10), which is exactly where value estimation meets coding), λ_t the price of rate and R_{tot} the total rate available. Problem (37) is concave in $\{R_i\}$: $P_{\text{task}}(R)$ is linear in R and $2^{-\beta R}$ convex in R, so P_{task} is convex, and maximising a concave

objective under linear constraints is a convex programme. The Lagrangian Dual is the water-filling solution:

$$R_i^* = [\frac{1}{\beta}\log_2(\frac{w_i P_{max}\kappa_s(D_0-D_{min})\beta\ln 2}{\lambda_t^*})]_{R_{min}}^{R_{max}} \quad (38)$$

where $[\cdot]_{R_{min}}^{R_{max}}$ denotes projection onto the bounds and λ λ_t^* is the multiplier for the total-rate constraint, updated as

$$\lambda_t^{k+1} = [\lambda_t^k + \kappa_d(\sum_{i=1}^N R_i^*(\lambda_t^k) - R_{\text{tot}})]^+ \quad (39)$$

with $\kappa_d > 0$ the dual step size. As shown in (38), the rate increases logarithmically with the value weight w_i; thus, high-value messages are assigned a higher rate without exhausting the budget, and a message with a weight lower than that indicated by λ_t^* will receive a minimal rate or no rate. At the same total rate, allocation outperforms fixed-rate coding by a considerable margin; as shown in the ablation study in Section 8.6, adding water-filling to replace the fixed rate increases task success by 4.8 percentage points.

5.3 Key-token Selection and its Approximation Guarantee

Under rate and transmission-time budgets not every unit can be sent, so a subset has to be chosen. It does not need to be noticeable; it must help to alter the decision. The purpose of the mth unit is as follows:

$$I_m = \mathbb{E}[|V_\omega(\mathcal{H} \cup \{e_m\}) - V_\omega(\mathcal{H})|] \quad (40)$$

where $\mathcal{H}$ is the set already selected and V_ω the value network of (27). Different Situations Require Different Levels of Importance for the Same Unit. When the swarm is divided and has had little luck finding the target by now, it has lost its way. Using the chosen set S, the expected value of transmission is

$$F(S) = \mathbb{E}[\Delta P_{\text{task}}(S)] + \alpha_r \mathbb{E}[\Delta R_{\text{risk}}(S)] \quad (41)$$

where α_r is the weight of the risk term, and $\Delta P_{\text{task}}(S)$ and $\Delta R_{\text{risk}}(S)$ are the gains in success probability and risk discrimination provided by S. First of all, there is Submodularity.

Theorem 1 (submodularity). If the marginal contribution of a unit to the value function V_ω respects the monotonicity of information processing, then the set function F defined by (41) is monotone submodular: for any $S \subseteq T$ and $e \notin T$,

$$F(S \cup \{e\}) - F(S) \geq F(T \cup \{e\}) - F(T) \quad (42)$$

which is the diminishing-returns property. Proof sketch: by (40) the marginal value of unit e is the conditional mutual information between e and the decision variable given the set. The information-processing inequality makes conditional mutual information non-increasing as the conditioning set grows, so $I(r; e|S) \geq I(r; e|T)$ holds for $S \subseteq T$. The larger T is, the less new information unit e can add; diminishing returns follow and F is monotone submodular. Repeated observation must lose marginal value—which is what the experiments show.

Theorem 2 (greedy ratio). Under the cardinality constraint $|S| \leq M$, maximising (41) is NP-hard, yet the greedy rule that repeatedly takes the unit of largest marginal gain (step 8 of Algorithm 1) returns

$$\frac{F(S_{\text{greedy}})}{F(S^*)} \geq 1 - (1 - \frac{1}{M})^M \geq 1 - \frac{1}{e} \approx 0.632 \quad (43)$$

where S^* is the optimal set and M the cardinality bound. The result is classical for submodular maximisation and follows from (42) together with the limit of $1 - (1 - 1/\mathrm{M})^{\mathrm{M}}$. Our constraint couples cardinality with rate, and the greedy bound for cardinality alone is known to be unimprovable on general monotone submodular functions, so the ratio is tight. At the parameters used here ($\mathrm{M} = 64$) the bound is about 0.632, while the greedy solution in Section 8 reaches 94.7% of the optimum; the gap is wide because the value function is close to additive in most situations, and submodularity shows mainly in highly redundant regions. Figure 4 shows the flow of multi-precision coding and token selection.

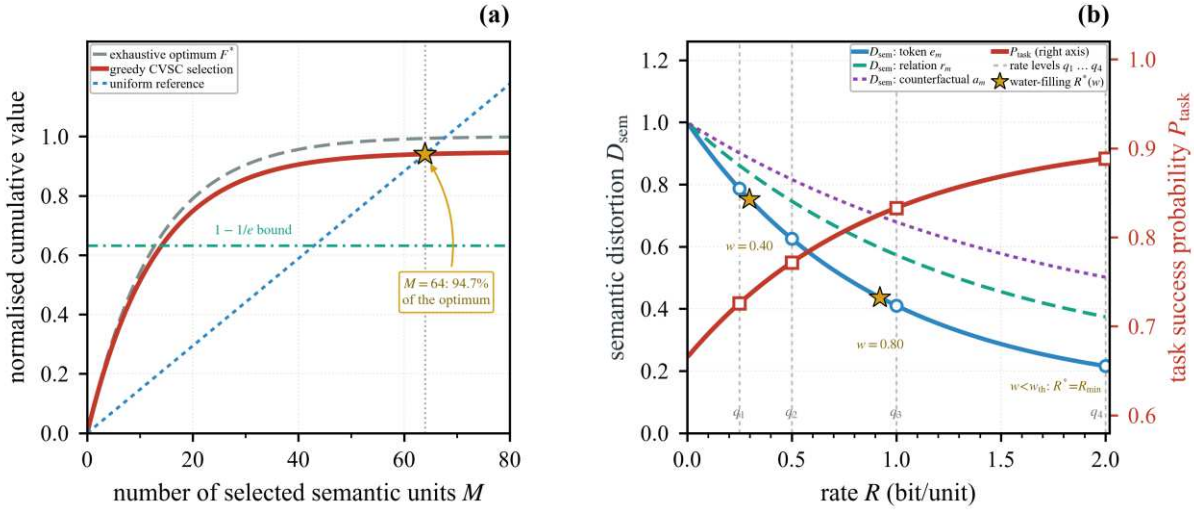

Fig. 4. Task-driven semantic representation and multi-precision coding. (a) Greedy selection of key semantic units: normalised cumulative value against the cardinality bound M, where the greedy solution reaches 94.7% of the optimum at $M = 64$ and stays above the theoretical lower bound $1 - 1/e$ throughout. (b) The rate–semantic distortion–success relation of (35) and (36): the three unit types (visual tokens e_m, relations r_m, counterfactuals a_m) carry individually calibrated D_{min} and β; the solid red line is the success rate for visual tokens (right axis), the vertical grey dashed lines mark the discrete rate levels $q_1 \dots q_4$, and the open circles and squares are the operating points reached by quantising.

the continuous optimal rate to q ∈ $\mathcal{Q}$; the star is the water-filling optimum $R^*(w)$ of (38), with w the information value weight;

when $w < w_{th}$ the optimum is projected onto the lower bound R_{min}.

5.4 Transmission Priority and Scheduling

Triggering and Selection specify what is transmitted and when. The connections vary; therefore, organise according to the importance of information and queue length. We use the Lyapunov drift-plus-penalty framework. Let $Q_i(t)$ be the length of the queue holding u_i, with dynamics $Q_i(t+1) = max\{Q_i(t) - C_i(t), 0\} + A_i(t)$, where $C_i(t)$ is the number of bits the slot can carry and $A_i(t)$ the bits newly queued. With Lyapunov function $L(t) = \frac{1}{2}\sum_i Q_i^2(t)$, the drift-plus-penalty bound is

$$\Delta(t) - V\mathbb{E}[U(t)|Q(t)] \le B + \sum_{i=1}^{N} Q_i(t)\mathbb{E}[A_i(t) - C_i(t)|Q(t)] - V\mathbb{E}[U(t)|Q(t)] \quad (44)$$

where $\Delta(t) = \mathbb{E}[L(t+1) - L(t)|Q(t)]$ is the conditional drift, $U(t) = \sum_z \alpha(z) V_\omega(z)$ the semantic value accrued in the slot, $V > 0$ the value–delay trade-off parameter, B a constant tied to the second moment of the arrival process, and $Q(t) = [Q_1(t), \dots, Q_N(t)]^\top$. The first term on the right of (44) does not depend on the decision, so minimising the bound reduces to a decision taken message by message:

$$\alpha(z) = 1 \Leftrightarrow V_\omega(z) \ge \frac{1}{V}\sum_{i=1}^{N} Q_i(t)\frac{|z|}{R_{ij}(t)} \quad (45)$$

where $|z|$ is the bit length of message z and $R_{ij}(t)$ the achievable rate of its intended link. Equation (45) is relatively simple to implement; the value threshold is increased by the backlog itself. Under congestion (Q_i is large), the threshold is raised, and only important messages are transmitted over the channel; when the network is idle, the threshold is reduced to send less valuable messages to improve consistency. The behaviour does not require any tuning of the trigger rate, and thus, it is not subject to event triggering at a fixed threshold. The policy provides the conditions for stability and optimisation.

$$lim\,sup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[Q_i(t)] \le \frac{B + V(U_{max} - U_{\text{opt}})}{\epsilon},$$
$$\overline{U}_{\text{avg}} \ge U_{\text{opt}} - \frac{B}{V} \quad (46)$$

where $\epsilon > 0$ is the expected service margin of the queue, U_{opt} is the return of a policy that knows the future, and U_{max} is a per-slot bound on the return. According to equation (46), increasing V can get close to the maximum semantic gain, but it will be delayed by an overhead that increases with O(V). A practical application of this trade-off is that if the task can tolerate an average delay of no more than τ_{max}, then V can be

bounded accordingly. Algorithm 1 is as follows: Trigger collection, Selection, Coding and Scheduling.

Algorithm 1: Cognitive value-aware semantic communication trigger and content selection (CVSC-TC)

Input: candidate messages $\mathcal{Z}$, value threshold η, generalisation gap $\varepsilon_{n,\delta}$, total rate R_{tot}, cardinality bound M, queue state $Q(t)$

Output: transmission plan Π

1: for each candidate message $z \in \mathcal{Z}$ do

2: compute $V_\omega(z)$ from (27); if $V_\omega(z)$ falls inside the threshold neighbourhood $[\eta - \varepsilon_{n,\delta}, \eta + \varepsilon_{n,\delta}]$ then

3: recompute the value terms exactly from (20) and (23) and correct $V_\omega(z)$

4: end if

5: test the trigger with (33); skip the message if it does not fire

6: end for

7: for every triggered message, compute the marginal value of each semantic unit from (40) and sort in decreasing order

8: with the set function of (41), take greedily the unit of largest marginal gain until the cardinality bound M is reached, which gives S_{greedy} (ratio guaranteed by (43))

9: with the normalised value of (10) as weight w_i, allocate rate from (38) and adjust λ_t^* under (39) until the total-rate constraint holds

10: quantise the rate onto the discrete level set $\mathcal{Q}$ and check distortion and task performance with (35) and (36)

11: decide transmission for each message from (45) and occupy the channel in decreasing order of value density

12: for each message to be sent, set up multi-hop semantic routing with the path cost of (47)

13: estimate end-to-end semantic distortion from (49); if it exceeds the threshold, fall back to a higher precision level

14: return transmission plan Π

5.5 Multi-hop Semantic Routing

When the destination lies beyond radio range the semantics must be relayed. Conventional routing weighs hop count against link quality; semantic routing must also account for the semantic loss each hop introduces. The cost of a path $\mathcal{R}$ is

$$C(\mathcal{R}) = \sum_{(i,j)\in\mathcal{R}} \left(\frac{|z|}{R_{ij}(t)} + \varsigma_d D_{\mathrm{sem}}^{ij}\right) + \varsigma_h |\mathcal{R}| \quad (47)$$

where D_{sem}^{ij} is the extra semantic distortion that bandwidth limits introduce on edge (i,j), ς_d and ς_h are normalised weights on semantic loss and hop count, and $|\mathcal{R}|$ is the hop count. The cost is additive along the path, so Dijkstra returns the optimum in $O(|\mathcal{E}|\log|\mathcal{U}|)$ without changing the complexity class of the routing protocol. End-to-end semantic distortion, by contrast, accumulates multiplicatively:

$$D_{\mathrm{sem}}^{\mathrm{e2e}} = 1 - \prod_{(i,j)\in\mathcal{R}} (1 - D_{\mathrm{sem}}^{ij}) \quad (48)$$

the product coming from a model in which every hop independently discards part of the semantic units. Eq. (48) shows how strongly hops amplify distortion: 5% per hop grows to 22.6% over five hops. Hence the relay positions in the reconfiguration problem.Therefore, the relay positions in the reconfiguration problem. Moving the relay reduces the distance and improves the connection quality; thus, $D_{\mathrm{sem}}^{\mathrm{e2e}}$ and the hop count are lowered.

5.6 Semantic Recovery and Distortion Analysis

At the receiver end, the units need to be converted back into usable information, and then the quality of reconstruction is evaluated. As it is an organised object rather than a picture, semantic consistency needs to be achieved instead of pixel-level accuracy. The End-to-End Recovery Error is as follows:

$$D_{\mathrm{rec}} = \frac{1}{|\mathcal{E}|}\sum_{m\in\mathcal{E}}[1 - \mathrm{sim}(h_m, \hat{h}_m)] + \frac{1}{|\mathcal{R}_{\mathrm{rel}}|}\sum_{r\in\mathcal{R}_{\mathrm{rel}}} \mathbf{1}\{\hat{r} \neq r\} \quad (49)$$

where $\mathcal{E}$ is the transmitted unit set, h_m and $\hat{h}_m$ the recovered and original semantic embeddings, $\mathrm{sim}(\cdot,\cdot)$ a cosine similarity, $\mathcal{R}_{\mathrm{rel}}$ the set of relation units and $r, \hat{r}$ the recovered and true relation labels. The second term is a discrete indicator rather than a continuous similarity, because a shift in relation semantics—reading inside as outside—sits close in embedding space yet is as wrong for the decision as a total failure, so relations are compared exactly, one by one. The metric serves two purposes here: it calibrates D_{min} and β in (35), and it tells the receiver whether to request a higher-precision retransmission.

6. 2024 Disagreement-Driven Formation Reconfiguration

6.1 Constructing the Disagreement Field

The input to reconfiguration is the disagreement field. The field $\mathcal{D}(x,t)$ of (13) is continuously differentiable over the task area, so its gradient gives the direction in

which disagreement concentrates. To reduce the on-board computation, we do not perform point-by-point calculations; instead, the area is divided into macro-cells $X = \{X_1, \dots, X_G\}$, and the field is calculated at the cell centers and then reconstructed via bilinear interpolation, thus reducing the cost to be proportional to the number of cells \$G\$ rather than the number of pixels. $G = 64$ is used in this implementation; therefore, the construction time of 0.35ms is small compared to $T_s = 2$ s slot.

6.2 Locating the Sources of Disagreement

Deviation from the mean is called a deviation. Therefore, the reasons for these problems are different; a difference in perception requires a high-resolution platform, a change in viewpoint needs an auxiliary angle, and a problem of old data only needs new data. Disagreement is due to first-order causes. Let c be the cause and ϑ_c be the observable feature associated with it; the relative contribution is

$$\rho_c = \frac{|\partial \mathcal{D}(x,t)/\partial \vartheta_c|}{\sum_{c'=1}^{3} |\partial \mathcal{D}(x,t)/\partial \vartheta_{c'}|}, \sum_{c=1}^{3} \rho_c = 1 \tag{50}$$

The three causes are shown as the ratio of ground sample distances $\vartheta_1 = |\log(\mathrm{GSD}_i/\mathrm{GSD}_j)|$, the angle $\vartheta_2 = \beta_{ij}$ between two lines of sight, and the timestamp difference $\vartheta_3 = |t_i - t_j|$. The partial derivatives are approximated by finite differences: perturb each feature by one standard deviation, recompute the fused belief and read the change in the field, from which a three-point stencil gives the whole gradient. Since ϑ_c has three dimensions only, the step costs about two belief fusions, 0.9 ms in our implementation. The attribution sets the node-selection rule of Section 6.3.

6.3 Selecting Observation and Relay Nodes

Observation nodes are chosen so that the largest disagreement is resolved at the lowest flight cost. Let $X_d = \{x : \mathcal{D}(x,t) \geq D_{\mathrm{th}}\}$ be the set of macro-cells whose disagreement exceeds the threshold; the observation value of platform u_k for cell x is the reduction it can deliver:

$$\mathrm{OV}(u_k, x) = \underbrace{(1-\rho_2)\mathcal{D}(x,t)}_{\text{availability}} \cdot \underbrace{\frac{|\cos\beta_k(x)|}{\|p_k - x\| + \delta_0}}_{\text{viewgeometry}} \cdot \underbrace{\mathrm{GSD}_k^{-1}(x)}_{\text{resolution}} \tag{51}$$

$\beta_k(x)$ is the angle between the line-of-sight of uk and the cell normal, δ_0 is a distance-smoothing constant, and $\mathrm{GSD}_k(x)$ is the ground sample distance of the platform at x. The weight of the first term is $(1-\rho_2)$; when the disagreement is due to viewpoin(ρ_2 large), a high-resolution platform is not very helpful, and the geometry needs to be altered first. The design does not have the problem of repeatedly showing the same incorrect view with a different camera. Assigning observation tasks is a set-partition problem with a cardinality constraint, and to reduce the marginal returns, the submodular greedy framework of Theorems 1 and 2 can be employed.

Relay nodes are chosen to minimise multi-hop semantic cost and relocation cost while keeping the graph connected. Let $\mathcal{P}_r$ be the candidate relay positions and $y_n \in \{0,1\}$ the relay-selection variables; the problem is

$$\begin{aligned} &min_{\{y_n, p_n\}_{n \in \mathcal{P}_r}} \sum_n y_n [\varsigma_m \|p_n - p_n^0\| + C(\mathcal{R}_n)] \\ &\text{s.t.}\, \lambda_2(L) \geq \lambda_{min}, y_n \in \{0,1\} \end{aligned} \tag{52}$$

where p_n^0 is the current position of a candidate, ς_m a relocation-cost coefficient, $C(\mathcal{R}_n)$ the mean path cost when traffic is forwarded through node n (from (47)) and $\lambda_2(L)$ the algebraic connectivity of the communication graph. The connectivity constraint is convex and monotone; adding an edge can only increase λ_2, so a two-stage rule is employed: greedily add edges according to the cost-benefit ratio, and then check connectivity to obtain a feasible solution in $O(|\mathcal{P}_r|\log|\mathcal{P}_r|)$. The Constraint is needed in a shadowed area. Minimise the path cost to drive all relays to the lower-cost end and reduce the graph to a tree that fails at a single point.

6.4 Formation Reconfiguration

Set up the observation and relay nodes, and then place all the platforms according to the requirements. The four components of fixed-cost are:

$$\begin{aligned} J_f = \varsigma_1 \sum_i \|p_i - p_i^d\|^2 + \varsigma_2 \sum_{x \in X_d} \mathcal{D}(x,t) \exp\left(-\frac{r_{\mathrm{cov}}^2(x)}{2\sigma_c^2}\right) + \\ \varsigma_3 \tau_{\mathrm{com}} + \varsigma_4 \sum_{i,j} \psi(\|p_i - p_j\|) \end{aligned} \tag{53}$$

Here, p_i^d is the desired position of u_i (the result of observation and relay assignment), $r_{\mathrm{cov}}(x)$ is the distance from x to the nearest observation platform, σ c is a coverage decay scale, τ_{com} is the mean communication delay, and $\psi(\cdot)$ is a collision penalty function that increases sharply when the separation falls below the safety distance. Firstly, there is a platform near its destination; secondly, it is away from the conflict area; thirdly, the degeneration of the shape does

not lead to a loss of communication; and finally, it protects the swarm. Clustering is a cluster of points that links nearby points and increases the probability of collision. Therefore, the optimum is neither clustered nor dispersed; it is a compromise determined by the four coefficients.

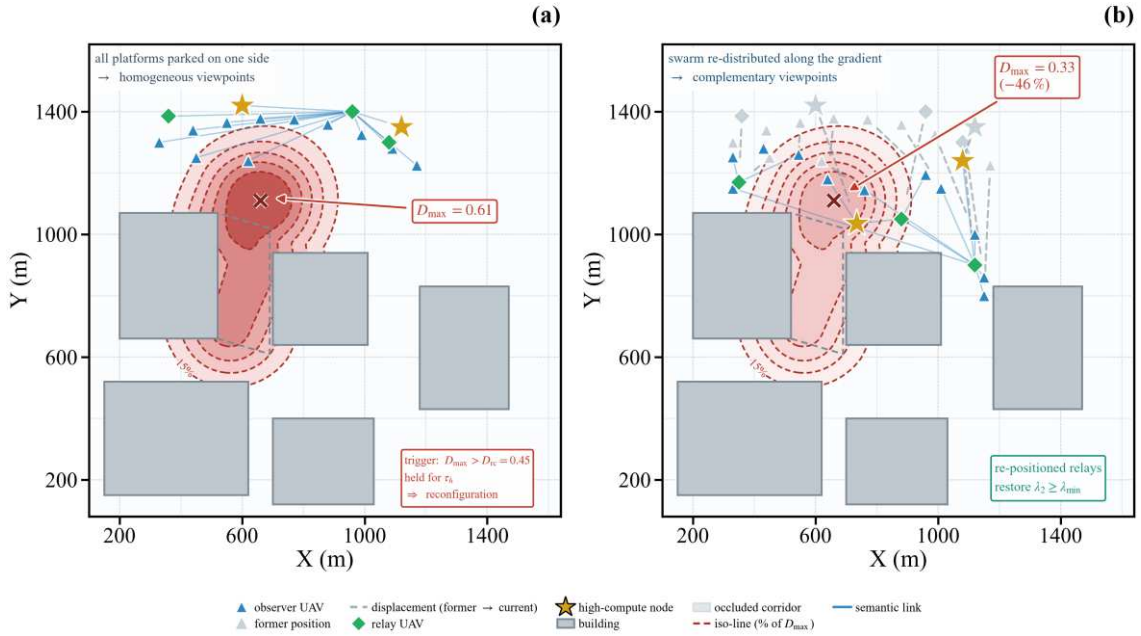

Fig. 5. Disagreement-driven reconfiguration, shown for the same 16 platforms over the same task area in both panels. The field factorises as $D = D_{max}\varphi$; its normalised distribution φ is fixed by the geometry of building occlusion and is independent of the formation, so the red dashed contours in the two panels are one and the same set of scaled level lines of φ, labelled at 15%, 30%, 45%, 60% and 75% of the peak, with identical shape, spacing, position and labels. Absolute amplitude is taken from the peak D_{max} (0.61 in (a), 0.33 in (b)), and shading is used, which shares the same 0-0.61 scale; thus, the whole field appears paler after the manoeuvre. (a) Before the manoeuvre, all platforms are located north of the area with nearly identical viewing geometry and cannot observe the shadow corridor cast by the buildings; the peak D_{max}x=0.61 (red cross) exceeds the trigger thresholdD_{rc}=0.45 and remains above τ h, thus initiating the manoeuvre. (b) After the manoeuvre, the swarm has redistributed along the disagreement gradient—the high-compute node moves forward to the center of the disagreement to carry out inference, four observation platforms shift east for supplementary views, and three relays relocate so that the communication graph maintains $\lambda_2 \geq \lambda_{min}$; pale grey marks the pre-manoeuvre location of each platform, and grey dashed segments connect the two positions, whose length is the displacement, all within the per-step limit of (54); the peak has dropped to D_{max}=0.33, approximately 45.9% lower than the previous 0.61 (about 46% according to the text), fallen below the threshold, and the system exits the reconfiguration state.

Algorithm 2: Disagreement-driven dynamic formation reconfiguration (DAFR)

Input: beliefs $\{b_i(t)\}$, positions $\{p_i(t)\}$, disagreement threshold D_{th}, hysteresis time τ_h, connectivity threshold λ_{min}

Output: reconfigured positions $\{p_i^*\}$ and routing structure $\mathcal{R}$

1: build the disagreement field $\mathcal{D}(x,t)$ on macro-cells from (13) and reconstruct it by interpolation

2: if $max_x\mathcal{D}(x,t) < D_{\mathrm{th}}$ stays above τ_h then keep the current shape and return

3: attribute the source of disagreement for every macro-cell above threshold with (50), giving ρ_1, ρ_2, ρ_3

4: compute the observation value of candidate platforms for each macro-cell from (51) and assign observation nodes by submodular greedy selection (ratio guaranteed by (43))

5: for each candidate relay position $n \in \mathcal{P}_r$ do

6: compute the mean path cost of forwarding through n from (47) and form the cost-benefit ratio

7: end for

8: add edges greedily in ascending cost-benefit order until the algebraic connectivity constraint $\lambda_2 \geq \lambda_{min}$ of (52) holds

9: initialise $p^{(0)} \leftarrow p(t)$; repeat

10: build the convex upper bound of the formation cost (53) from (60)

11: solve the convex subproblem (54) for a temporary position p

12: apply control-barrier safety filtering under (55) and project onto Ω_{safe} while keeping the minimum separation

13: if the Lyapunov function of (57) does not decrease then halve the trust-region radius and re-solve

14: until the decrease of the objective falls below ϵ or the iteration bound is reached

15: generate velocity and heading commands under (55) and issue them smoothly to the platforms
16: return $\{p_i^*\}$ and $\mathcal{R}$

The Reconfiguration Problem follows:

$$min_{\{p_i\}} J_f \,\mathrm{s.t.}\, p_i \in \Omega_{\mathrm{safe}}, \|p_i - p_j\| \geq d_{\mathrm{safe}},$$
$$\|p_i(t) - p_i(t - T_s)\| \leq v_{max} T_s \quad (54)$$

where d_{safe} is the minimum safe separation and v_{max} the maximum speed. The second term of (53) is non-convex in p_i, being an exponential kernel composed with a nearest-neighbour distance, and the collision constraint is non-convex too, so the problem is solved by successive convex approximation as set out in Section 7.2. Reconfiguration does not run every slot. It is triggered only when the peak of the disagreement field exceeds the threshold D_{rc} (0.45, Table 1) and stays above it for longer than the hysteresis time τ_h; after the manoeuvre the peak falls back and the system returns to quiescence, which spares the airframes the wear and the network the churn of constant reshaping. Fig. 5 illustrates the process and Algorithm 2 states it.

6.5 Dynamic and Communication Constraints

A Rearranged Position needs to be in the range of the airframe. The dynamics of the platform are modeled as a second-order system with input constraints, and a control barrier function maintains the forward invariance of the safe set.

$$\dot{p}_i = v_i, \dot{v}_i = u_i, \|u_i\| \leq u_{max}, \dot{h}(p_i) \geq -\alpha_b h(p_i) \quad (55)$$

where v_i is the velocity, u_i the acceleration input, u_{max} the acceleration bound, $h(p_i) \geq 0$ the defining function of the safe set (no-fly zones, minimum altitude, inter-platform separation) and $\alpha_b > 0$ the barrier decay rate.

platform separation) and $\alpha_b > 0$ the barrier decay rate. The differential inequality in (55) is the barrier condition, and it keeps the trajectory within the safe set starting from inside it. We introduce it to the convex subproblem in Section 6.4 as a quadratic constraint, thus position planning and safety filtering are performed in a single step, and the feasibility conflict of a two-layer design is never encountered. Constraints of this type are the foundation for safe reinforcement learning with control barrier functions [59], multi-target tracking with a separate safety network to filter dangerous actions [48], and constrained reinforcement learning for formation control with safety penalties [60]. The difference is that the constraint is created online by the disagreement field and not set in advance.

In addition to the algebraic connectivity in (17), there is a communication constraint per hop and a maximum number of hops:

$$R_{ij}(t) \geq R_{min}, |\mathcal{R}| \leq H_{max}, \sum_i |z_{ij}(t)|\, \alpha_{ij}(t) \leq C_{\mathrm{tot}}(t) \quad (56)$$

R_{min} is the minimum rate for reliable demodulation, H_{max} is the maximum hop count, and $C_{\mathrm{tot}}(t)$ is the total bit budget; thus, end-to-end semantic distortion is bounded by (48). Eqs. (56) and (52) show the chain from formation geometry to link quality and then to semantic distortion; therefore, formation and communication are addressed together in Section 7.

6.6 Stability

Reconfiguration needs to move the swarm from its current position to the destination quickly and stably, keep the connection and safety of the swarm, and handle the resulting instability. The candidate Lyapunov function is an integral function of shape error, residual mismatch and queue backlog.

$$W(t) = \frac{1}{2}\sum_{i=1}^{N} \|p_i(t) - p_i^*\|^2 + \frac{\varsigma_q}{2}\sum_{i=1}^{N} Q_i^2(t) + \varsigma_d \sum_x \mathcal{D}^2(x,t) \quad (57)$$

where p_i^* is the target position and ς_q, ς_d are weights. The three terms describe shape convergence, communication saturation and cognitive convergence. Differentiation along the closed-loop trajectory of (55) and substitution of the assignment in (51) and the schedule in (45) yields

$$\dot{W}(t) \leq -\alpha_1 \sum_i \|p_i - p_i^*\|^2 - \alpha_2 \sum_x \mathcal{D}^2(x,t) - \alpha_3 \sum_i Q_i^2(t) + c_0 \quad (58)$$

where $\alpha_1, \alpha_2, \alpha_3 > 0$ is fixed by the control gains, a lower bound on observation value and the service rate of the schedule, and c_0 is a constant bound caused by external disturbance (target motion, wind, arriving messages). Eq. (58) shows that the system is practically stable; the state converges to the set $\{W \leq c_0/\alpha_{min}\}$ with $\alpha_{min} = min\{\alpha_1, \alpha_2, \alpha_3\}$. The Three Effects in Practice. Perturbation does not prevent convergence; it only increases the size of the neighbourhood around the limit point, and thus a swarm in a dynamic environment does not need to be perfectly stable. Raise the lower

bound of the observation value and require that each manoeuvre actually yields information; thus, increase α_2 and reduce the steady-state residual disagreement, as shown in (51). And when ɑ 3 is too small, if the service rate drops below the arrival rate, the queue term can no longer be reduced, and the value threshold must be raised above (45) to restore stability; this is how the traffic fall of itself under congestion.

7 Joint Optimization Algorithm

7.1 Decomposition

Problem (15) is a combination of binary, discrete and continuous variables with mutual constraints that cannot be solved in its current form. The three blocks are divided according to the strength of coupling:

$$(\alpha, q) \to \mathcal{B}_1, p \to \mathcal{B}_2, \mathcal{P} \to \mathcal{B}_3 \tag{59}$$

where $\mathcal{B}_1$ is the communication and coding block (it fixes α_{ij} and q_i), $\mathcal{B}_2$ the formation block (it fixes p_i) and $\mathcal{B}_3$ the inference block (it fixes $\mathcal{P}_i$). The blocks exchange only three intermediate quantities: the channel gain g_{ij}, set by p; the effective rate R_{eff}, set jointly by p and α; and the cognitive disagreement $\mathcal{D}$, set by the messages injected through $\mathcal{B}_1$ together with the viewing geometry of $\mathcal{B}_2$. The coupling is sparse and one-directional, which suits alternating optimisation: hold two blocks, solve the third, rotate. Each subproblem is convex or nearly so with respect to the others, so the objective function does not increase from one iteration to the next.

7.2 Alternating Optimisation and Successive Convex Approximation

The formation block $\mathcal{B}_2$ shows how the convex approximation is built. The second term of (53) composes an exponential kernel with a nearest-neighbour distance, and the third carries a reciprocal rate; neither is convex. Expanding to first order at the point $p^{(k)}$ of iteration k and adding a quadratic trust-region term gives a convex upper bound:

$$\begin{aligned}\tilde{J}_f(p; p^{(k)}) = J_f(p^{(k)}) + \nabla J_f^{\top}(p^{(k)})(p - p^{(k)}) \\ + \frac{L_J}{2}\|p - p^{(k)}\|^2,\end{aligned}$$

$$\tilde{J}_f \geq J_f \tag{60}$$

where ∇J_f is the gradient of the formation cost with respect to position, L_J the Lipschitz constant of that gradient and $\frac{L_J}{2}\|p - p^{(k)}\|^2$ the trust-region penalty. The quadratic term makes $\tilde{J}_f$ a global upper bound on J_f, so each convex minimisation of $\tilde{J}_f$ is guaranteed to lower the original objective. The nonlinear constraints (separation, rate threshold, connectivity) are linearised at the same point or expressed as second-order cones; thus, the subproblem can be transformed into a second-order cone programme and solved by an interior-point method in polynomial time. Convergence of the m-th SCA iteration is as follows: Under Slater's condition, the iterates converge to a KKT point of the original problem.

7.3 Distributed Solution via ADMM

Alternating Optimisation is carried out in the swarm, and each platform only obtains local information. The alternating direction method of multipliers is employed to divide the global problem into per-platform subproblems and a low-dimensional consistency update. Auxiliary variables can be added to the problem in consensus form for the augmented Lagrangian.

$$\mathcal{L}_\zeta(x, y, \eta) = \sum_i f_i(x_i) + \varsigma g(y) + \frac{\zeta}{2}\|\sum_i A_i x_i - y + \eta\|_2^2 - \frac{\zeta}{2}\|\eta\|_2^2 \tag{61}$$

where x_i is the local variable of u_i (its communication decisions, compression level and position), f_i the local cost (its own energy, delay and observation value), y the global consistency variable (total traffic, total disagreement, total rate), $g(y)$ the indicator function of the global constraints, A_i the coupling matrix, η the scaled dual variable and $\zeta > 0$ the penalty. The three standard updates follow:

$$\begin{aligned}x_i^{k+1} &= \backslash\text{arg}min_{x_i}\mathcal{L}_\zeta(x_i, y^k, \eta^k), \\ y^{k+1} &= \backslash\text{arg}min_y\mathcal{L}_\zeta(x^{k+1}, y, \eta^k), \\ \eta^{k+1} &= \eta^k + \sum_i A_i x_i^{k+1} - y^{k+1}\end{aligned} \tag{62}$$

The first step depends only on the state of u_i and on two broadcast scalars, the traffic residual and the disagreement residual, so it is solved on board without help. The second involves the low-dimensional y and is finished in constant time by the current relay or a rotating node. The third updates the duals and needs nothing but the residuals. A round therefore exchanges $O(N)$ scalars regardless of problem dimension. Algorithm 3 states the full three-layer procedure.

Algorithm 3: Three-layer solver for joint communication, semantic compression and reconfiguration (JCSF-ADMM)

Input: initial positions $p^{(0)}$, initial inference paths $\mathcal{P}^{(0)}$, convergence threshold ϵ, penalty ζ, maximum rounds K_{max}

Output: $\{\alpha^*, q^*, p^*, \mathcal{P}^*\}$

1: initialise x^0, y^0, η^0 and cache g_{ij}, R_{eff} and $\mathcal{D}$

2: repeat outer alternating optimisation

3: communication and coding block $\mathcal{B}_1$: with p and $\mathcal{P}$ fixed, estimate value from (27), decide the trigger from (33), select units greedily from (43) and allocate rate from (38)

4: schedule from (45), set up multi-hop semantic routing from (47) and update R_{eff} and $\mathcal{D}$

5: formation block $\mathcal{B}_2$: with α, q and $\mathcal{P}$ fixed, build the convex upper bound from (60) and solve (54) for a temporary position

6: inner repeat: barrier filtering under (55); if (57) does not decrease, halve the trust region and re-solve

7: until the inner loop converges; output p

8: inference block $\mathcal{B}_3$: with the other variables fixed, prune and extend the inference path by the ratio of value gain to inference delay, giving $\mathcal{P}$

9: ADMM consistency update: update x, y and η from (62) and broadcast the residual scalars; adapt ζ to the residual ratio

10: if neither the unit selection nor the formation positions change in two consecutive rounds, lock the block and skip the remaining inner iterations

11: until the relative decrease of objective (15) falls below ϵ or the round count exceeds K_{max}

12: return $\{\alpha^*, q^*, p^*, \mathcal{P}^*\}$

7.4 Convergence and Complexity

The Three conditions for Convergence.. Alternating optimisation, with every block subproblem convex and feasible, lowers the objective monotonically and is bounded below, hence reaches a stationary point of block coordinate descent. The trust-region step preserves feasibility and decreases the objective at each iteration, converging to a KKT point under Slater's condition. And ADMM with a single coupling matrix has an ergodic rate: the primal and dual residuals at iteration k satisfy

$$\|r^k\|_2^2 + \|s^k\|_2^2 \le \frac{C_A}{k}, |F^k - F^*| \le \frac{C_B}{k} \qquad (63)$$

where r^k and s^k are the primal and dual residuals at round k, F^k the objective at round k, F^* the optimum, and C_A, C_B constants tied to the initial deviation and the penalty. The $O(1/k)$ rate of (63) means that roughly a hundred iterations bring the residual to 1% of its initial value; with our parameters one iteration costs about 1.8 ms per platform, so the three layers converge in about 0.2 s, well inside the $T_s = 2$ s slot. A single iteration has four cost components: value estimation $O(N|z|d_e)$, with d_e the embedding dimension of a semantic unit; greedy unit selection $O(M^2|z|)$; the formation subproblem $O(N^{3.5})$, at the interior-point cost of a second-order cone programme; and the consensus update $O(N)$. The total is

$$C_{\text{total}} = O(N|z|d_e + M^2|z| + N^{3.5})\text{perround} \qquad (64)$$

where M is the cardinality bound on unit selection. The $N^{3.5}$ term stays below the order of 10^6 while $N \le 50$, and both value estimation and unit selection scale linearly in N, so the algorithm runs in real time at the swarm sizes considered here. At larger sizes the formation subproblem can be split by spatial partition and solved in parallel, which reduces the complexity to $O(N/N_p \cdot N_p^{3.5})$, with N_p the number of platforms per partition, effectively linear.

8. Experiments

8.1 Simulation Setup

Experiments use a low-altitude cooperative sensing and risk-assessment task with 32 heterogeneous UAVs. The Area is 1600m x 1600m of semi-structured suburban terrain with six building blocks, two communication shadow corridors and a patch of vegetation. It holds 24 real targets (moving people, heat sources, temporary obstacles) and 24 decoys (reflective roofs, air-conditioning units, tree-crown hot spots), of which 9 targets are marked as critical for the risk decision. The swarm must complete search, recognition and risk confirmation within 1500 s, and the task counts as successful only if every critical target is confirmed correctly and the miss rate stays under its threshold. On-board compute is modelled on a Jetson Orin NX: a

quantised 7B model performs inference with the τ_{inf} term of (6) at about 78 ms per forward pass, and the value network, two feed-forward layers of width 128, predicts in 0.21 ms per message. Table 1 lists the main parameters.

Table 1. Principal Simulation Parameters

Parameter	Value	Parameter	Value
Task area	1600 m × 1600 m	Swarm size N	8–48 platforms
Real / decoy / critical targets	24 / 24 / 9	Mission time limit	1500 s
Camera resolution	1920×1080 / 3840×2160	Pixel size	2.2–3.45 μm
Flight altitude	60–140 m	Flight speed	8–16 m/s
Bandwidth	1–4 MHz	Path-loss exponent α	2.4
Transmit / noise power	20 dBm / −104 dBm	LOS probability parameters (a_e, b_e, c_e)	(9.61, 0.16, 10.0)
Excess attenuation η_L/η_N	1.00 / 0.18	Slot length T_s	2 s
Bit budget per platform	2–12 kB	Compression levels Q	4 levels (0.25 / 0.5 / 1.0 / 2.0)
Cardinality bound M	64	Value weights $(\lambda_1, \dots, \lambda_4)$	(0.34, 0.28, 0.26, 0.12)
Trigger threshold η	0.35	Confidence $1-\delta$ / generalisation gap $\varepsilon_{n,\delta}$	0.95 / 0.041
Disagreement threshold D_{th} / reconfiguration trigger D_{rc}	0.18 / 0.45	Hysteresis time τ_h	20 s
Connectivity threshold λ_{min}	0.35	Minimum safe separation	25 m
Maximum acceleration	4 m/s ^{2}	Monte Carlo repetitions	30 (seed 2026 + run index)

Average the results of 30 independent trials, and then show the error bars and 95% confidence intervals according to the t-distribution with 29 degrees of freedom. Repetition k uses seed $2026+k$ to prevent randomness in the comparison, and all methods have the same target layout, channel realisation, and decoy arrangement. Training samples for the value network are generated online by (10), and each label is obtained from the exact computation; with $n = 2 \times 10^5$ samples and a spectral norm bound of 3.2, the gap between (31) and (32) is 0.041.

8.2 Baselines and Indicators

The six baseline families and their Design Choices are presented in Table 2. They all have the same perception backend and upper-level task allocator, as well as total bandwidth; only the trigger conditions, semantic coding, formation basis and scheduling are different, so the comparison is fair. Table 3 is the list of ten indicators and their improvements.

Table 2: Baselines and Their Design Choices

ID	Method	Trigger	Semantic coding	Formation basis
B1	Raw image transmission	Periodic, fixed length	None (JPEG)	Geometric coverage
B2	Fixed-rate semantic communication	Periodic, fixed length	Fixed 0.5 bit·token $^{-1}$	Geometric coverage
B3	State-change event triggering	Observed change	Fixed 0.5 bit·token $^{-1}$	Geometric coverage
B4	Age-of-information optimisation	AoI threshold	Fixed 0.5 bit·token $^{-1}$	Geometric coverage
B5	Centralised communication scheduling	Central decision	Greedy truncation	Centre-assigned
B6	Formation control without disagreement	Decision value (as here)	Water-filling (as here)	Geometric coverage
B7	CVSC-DFR (this paper)	Decision value + generalisation margin	Water-filling + submodular selection	Disagreement field

Fig. 6 shows a representative snapshot ($t = 780$ s, $N = 32$). Grey polygons are buildings; solid red circles are confirmed targets and open orange circles are targets still under dispute; the red dashed contours are scaled level lines (15%-75%) of the disagreement field of (13) drawn on the normalised distribution D/D_{max}, shading shows the absolute amplitude on the 0-0.61 scale shared with Fig. 5, blue lines are semantic links that satisfy $V(z) \geq \eta$, and the

red cross marks the peak. The two characteristics are: The peak is in the shadow corridor of the two buildings, and the platforms look at the scene from the same side and have not agreed for a long time. Three observation nodes are arranged unevenly around the peak, and the high-compute node is located on the far side of the corridor; this is what Algorithm 2 does after (50) identifies the viewpoint as the main reason. It is about 46 per cent lower than the previous uniform formation.

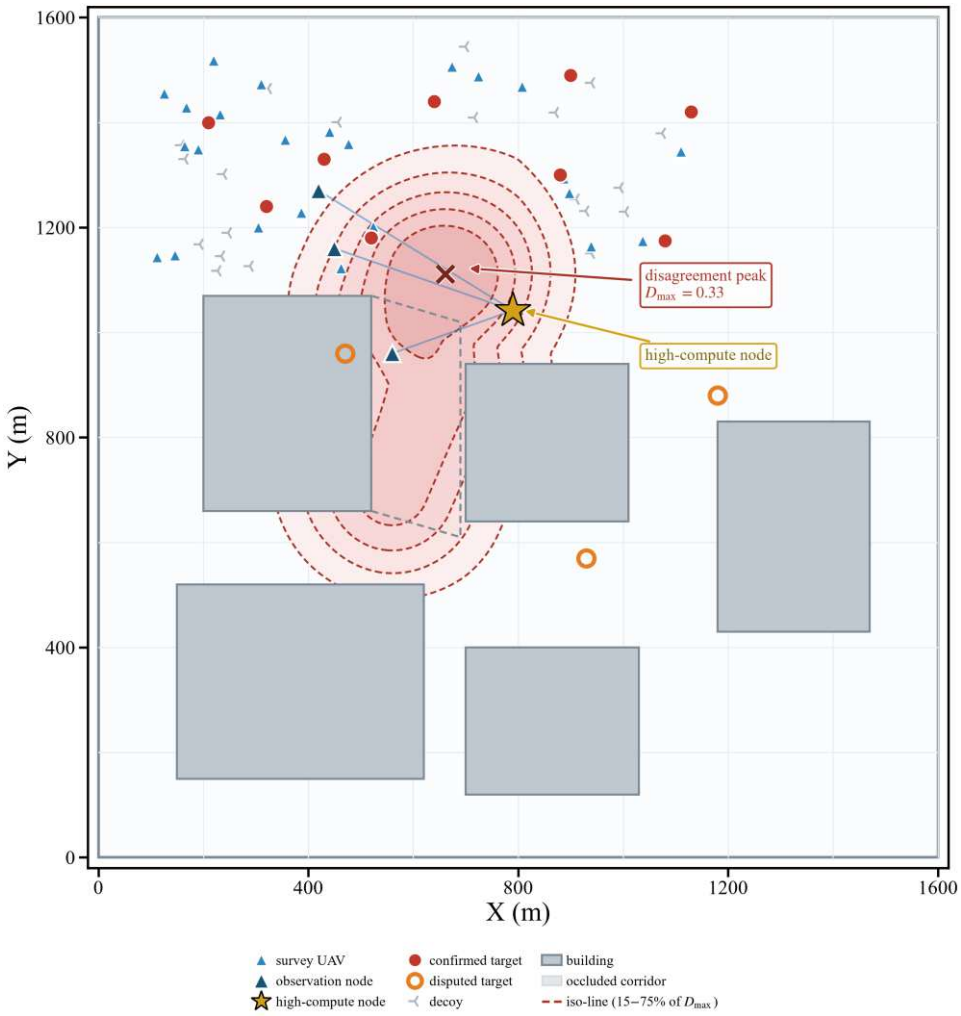

Fig. 6. A representative snapshot ($t = 780$ s, $N = 32$). The overlay carries the contours and shading of the cognitive disagreement field, the semantic links, confirmed and disputed targets, and the UAV positions. Contours use the same normalised distribution as Fig. 5 at 15%–75%, shading gives amplitude on the absolute 0–0.61 scale; the high-compute node has moved to the far side of the shadow corridor to carry the inference and three observation nodes form a complementary geometry; the red cross marks the peak.

Table 3: Definitions of the Metrics

Metric	Definition	Direction
Traffic per task	total bits transmitted to complete the task / targets confirmed	lower
Task success rate	fraction of missions in which every critical target is confirmed correctly and the miss rate meets its bound	higher
Cognitive consistency	time average of the consistency index C_{cog} over the final stage of the mission	higher
Recognition accuracy	ratio of correctly recognised targets and decoys	higher
Mean delay	mean end-to-end delay, the sum of sensing, encoding, transmission, propagation and inference	lower
Total energy	cumulative flight, computation and communication energy over the mission	lower
Reconfigurations	number of planned reconfiguration events (more than $\geq$ = 2 platforms displaced by over 15 m)	lower
Link-recovery time	mean time from a link failure until algebraic connectivity returns above threshold	lower
Risk miss rate	critical targets not judged high-risk / total critical targets	lower
Value-prediction	fraction of messages whose predicted value lies within 0.05 of the exact value	higher

Metric	Definition	Direction
accuracy		

8.3 Performance of Bandwidth and Packet Loss

Table 4 shows the seven methods at N=32, 2 MHz and 1% loss. CVSC-DFR (B7) is the best of the six. Reducing the volume of raw image transmission (B1) to 118.5 MB has dropped the traffic per task by 71.3% and increased the success rate from 0.702 to 0.898, an improvement of 19.6 percentage points. It is not a problem of awareness but of action. B1 uniformly compresses all observations and, about 62 per cent of the traffic, are repeated views of areas that the swarm has already agreed upon. The message has a task and consistency increment close to zero in (10) and is thus not triggered.

Table 4. Communication and task performance at $N = 32$ (mean ± 95% CI over 30 runs)

Method	Traffic / MB	Task success	Consistency	Recognition	Delay / ms	Energy / kJ
B1 Raw image	412.6±14.2	0.702±0.021	0.641±0.018	0.812±0.014	168.4±7.6	96.3±3.8
B2 Fixed-rate semantic	268.4±11.6	0.743±0.019	0.688±0.017	0.848±0.013	132.7±6.4	78.5±3.2
B3 State-change trigger	186.2±9.4	0.771±0.018	0.703±0.016	0.861±0.012	108.3±5.5	64.2±2.7
B4 AoI optimisation	214.7±10.1	0.786±0.017	0.724±0.015	0.869±0.012	96.5±5.1	69.8±2.9
B5 Centralised scheduling	152.9±8.2	0.804±0.016	0.746±0.015	0.883±0.011	84.7±4.6	58.1±2.4
B6 No-disagreement formation	121.5±6.8	0.842±0.014	0.812±0.013	0.902±0.010	76.2±4.2	52.4±2.1
B7 CVSC-DFR (ours)	118.5±6.1	0.898±0.011	0.938±0.009	0.951±0.007	62.4±3.5	43.6±1.8

The three will be introduced now. Age-of-information scheduling (B4) is used to trigger multiple state changes (B3), and it has been more successful more often: 0.786 times compared with 0.771 times. Therefore, the AoI criterion does not meet the two goals; at the threshold, it will update the old information, such as the background of the confirmed area, and ignores data that have not changed much but are semantically important. Centralised Scheduling (B5) has reduced traffic to 37.1% of B1, but the average delay is still 84.7ms because the centre needs to collect state information before making a decision, and the round trip is 41%; at 48 platforms, the decision queue backs up and throughput stops increasing.

Only by changing the trigger to a decision value (B6 instead of B5) was the success rate increased from 0.804 to 0.842; thus, value-driven triggering is effective. Adding a disagreement-driven reconfiguration on top of it (B7 against B6) increased the success to 0.898 and improved consistency from 0.812 to 0.938. The second gain is due to the risk-related sub-tasks; as shown in Table 5, the miss rate of B7 is 0.043 and that of B6 is 0.079, a reduction of 45.6%. Rearrangement can improve the appearance of the area in the shade, and even if communication is optimised, others will not know about it.

Table 5. Formation and risk metrics at $N = 32$ (mean ± 95% CI over 30 runs)

Method	Reconfigurations	Link recovery / s	Risk miss rate	Value-prediction accuracy
B1 Raw image	68±3.4	18.5±1.2	0.164±0.013	—
B2 Fixed-rate semantic	61±3.1	16.2±1.1	0.142±0.012	—
B3 State-change trigger	55±2.8	14.1±1.0	0.128±0.011	—
B4 AoI optimisation	58±2.9	12.8±0.9	0.118±0.010	—
B5 Centralised scheduling	49±2.5	11.4±0.8	0.104±0.009	—
B6 No-disagreement formation	44±2.2	9.6±0.7	0.079±0.008	0.884±0.012
B7 CVSC-DFR (ours)	31±1.7	7.7±0.5	0.043±0.006	0.917±0.009

Figure 7 is the sensitivity versus the trigger threshold η. As η increases from 0.15 to 0.65, traffic decreases continuously, the peak of success occurs around $\eta \approx 0.25$ and at $\eta = 0.35$, the traffic is 34.8% lower than that at $\eta = 0.15$ for a 0.6 percentage point increase in success; thus, $\eta = 0.35$ is chosen as the operating point. Equation (33) is for the shape. A low threshold will use up bandwidth for low-value messages and fail to achieve its goal due to congestion; a high threshold will not obtain the necessary marginal information and fail due to starvation. The gap of 0.041 from (32) accounts for the flat region around η^*, as the triggered set is barely altered within ± 0.041 of the threshold. As shown in Figure 7(b), the success-traffic Pareto front at three different bandwidths has been shifted

upwards and to the right when increasing the bandwidth from 1 MHz to 4 MHz; thus, both semantic selection and a larger raw bandwidth have improved the gain.

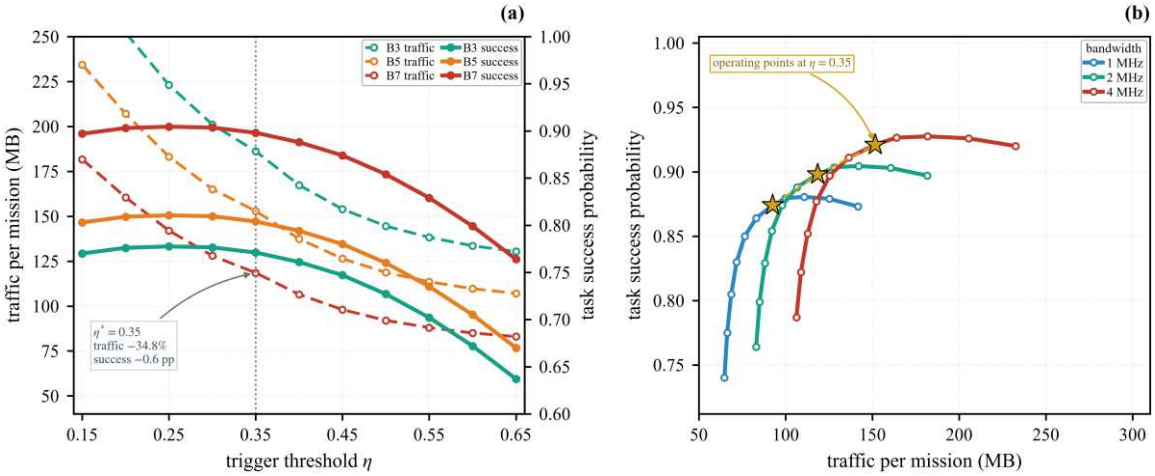

Fig. 7. Sensitivity to the trigger threshold η. (a) Traffic per task (dashed) and success (solid) against η for B3, B5 and B7. (b) Success–traffic Pareto fronts at 1, 2 and 4 MHz, the gold dashed line joining the compromise operating points at $\eta = 0.35$.

Bandwidth is presented in the Pareto front of Figure 7(b). After dropping to 1 MHz from 4 MHz, the success rate of B7 decreased by 5.1% (from 0.921 to 0.874), and that of B1 and B5 were 16.0% and 14.3% respectively (0.712-0.598 and 0.819-0.702). Robustness is achieved through water-filling and submodular selection; given the limited bandwidth (38), high-rate messages are preferentially selected, and (43) can still choose the most important units with a small cardinality; thus, the loss mainly affects the precision of low-value messages. Fig. 8 turns to packet loss. B7 decreases from 0.5% to 8% B7 loses 0.043 of the success (0.902 to 0.859) and maintains consistency above 0.91 (0.938 to 0.913); B1, B2 and B5 drop to 0.587, 0.628 and 0.695 respectively, while B3 and B4 are at 0.643 and 0.665. After the loss B7 estimation, end-to-end semantic distortion is reduced (as in equation 49), and only essential units are selectively retransmitted with high precision; thus, the event-triggered scheme avoids retransmission of value-based data and saves bandwidth. As shown in Figure 8(c), the value-prediction accuracy in the same range is as follows: Only B6 and B7 have value networks, and B7 decreases from 0.917 to 0.888 and B6 decreases from 0.884 to 0.850; therefore, the prediction is also relatively stable with respect to the loss. The remaining units can still be used for judgment; therefore, B7 is still above 0.86 after an 8% reduction.

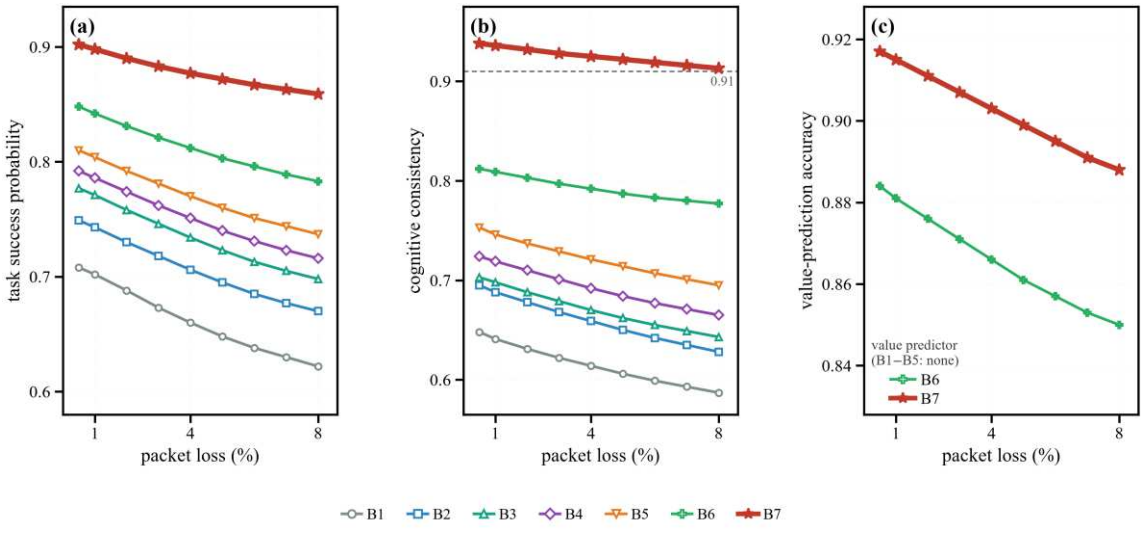

Figure 8. Effect of Packet Loss on (a) Task Success, (b) Cognitive Consistency and (c) Value-Prediction Accuracy. Panels (a) and (b) have the same legend at the bottom of the figure, and the grey dashed line in (b) is the 0.91 floor. Only B6 and B7 are in the value network; the other rows of Table 5 are dashes, so (c) has two curves. The average of the three runs at that time is taken, and the 95% interval is not calculated separately.

8.4 Disagreement Experiments

Fig. 9(a) shows consistency and the resolution of discrepancies. In the first 300 seconds, the methods are relatively close (0.62-0.69) due to a lack of overlapping observations; after 300 seconds, B7 increases sharply to a plateau of over 0.93 around 620 seconds, reaches 0.943 at 1500 seconds, and B1 and B2 remain at 0.64-0.68 and fluctuate periodically. A dip occurs due to the diffusion of a decoy-induced error; when multiple platforms are misled by the same decoy, a large reflection roof is formed, and their beliefs move in tandem, resulting in a temporary loss of consistency. B7 recognises the low-yield messages via the consistency term in (24) and suppresses them; thus, synchronous spread is interrupted.

Fig. 9(b) shows the reasons for the disagreement and how to solve it. The total number of triggered events for perceptual differences, viewpoint differences and timeliness is about 31%, 44% and 25% respectively; due to building occlusion, many platforms have been forced to take the same path. Resolution was achieved in 0.83, 0.79, and 0.94 of the three cases. Timeliness is simple; only need new data. Viewpoint is relatively difficult because of the dynamic constraints in (54) and (55); at some times, the platform cannot obtain an auxiliary angle, and thus Algorithm 2 delays this operation to a later slot. Ordering only by size (without attribution), as shown in Table 6, reduces the resolution to 0.68, which is the result of the decomposition in (50).

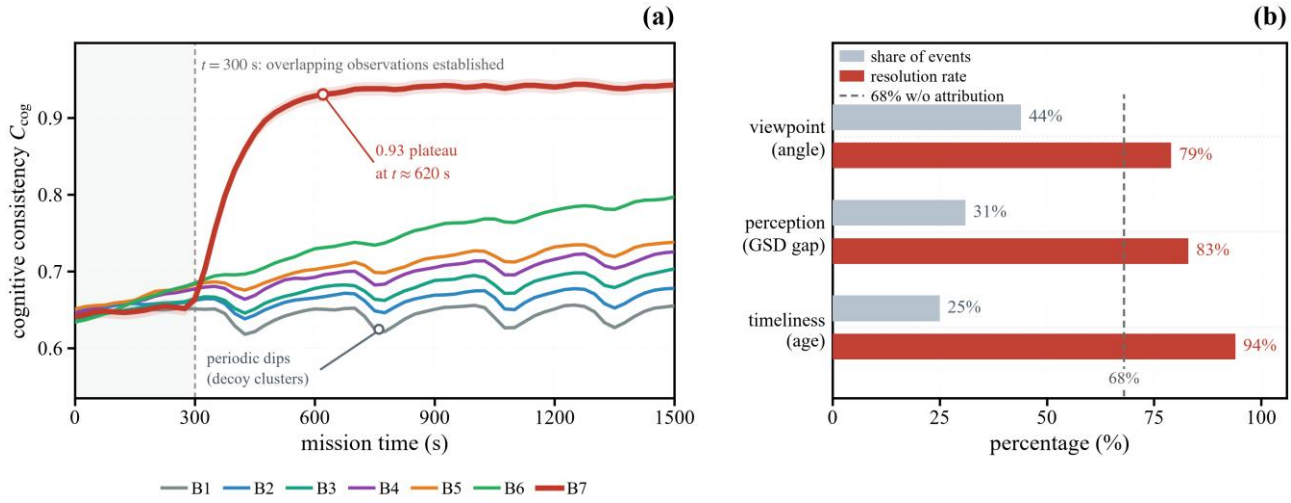

Figure 9. Disagreement Experiments. (a) Cognitive consistency with the mission time for the seven methods; the pale red band is the 95% interval of B7, the grey dashed line at t = 300s shows when overlapping observations occur and the methods start to diverge, and the periodic dips correspond to the spread of decoy-induced error. (b) Attribution of disagreement (grey-blue bars, share of each cause over all events) with the resolution success of each cause (red bars); the grey dashed line at 68% is the case without attribution. Reasons are given in proportion.

8.5 Reconfiguration Experiments

Fig. 10(a) summarises reconfiguration. As the swarm increases from 8 to 48 platforms, most methods have to reconfigure frequently and the slopes are different. Geometric Coverage Methods (B1-B4) increase almost linearly, from about 14 events at 8 platforms to about 96 at 48 platforms, because more platforms have opened up additional coverage gaps that need to be filled. B7 increases from 9 to 42 sublinearly due to the spatial aggregation of the disagreement field in (13); scale adds extent to the disagreement rather than more peaks, and hysteresis ($\tau_h = 20$ s) suppresses repeats.

Link recovery at the 48 platforms is 9.2s for B7 and 22.1s for B1; it has decreased by 58.4% from N=32 and the same reduction has occurred with scaling. As shown in equation (52), it does not have to be a shortest-path shape; rather, it is an explicit trade-off between path cost and connectivity, and thus redundant routes are not lost after the removal of a node. Release the constraint, minimize only the path cost, and the recovery time at Platform 48 is 20.4s, close to the 21.7s of B5. The risk miss rate of B7 is less than 0.05 at all sizes, but for B1, it remains at 0.171 even after adding 48 platforms; thus, the systematic blind spot of homogeneous viewing geometry has not been resolved.

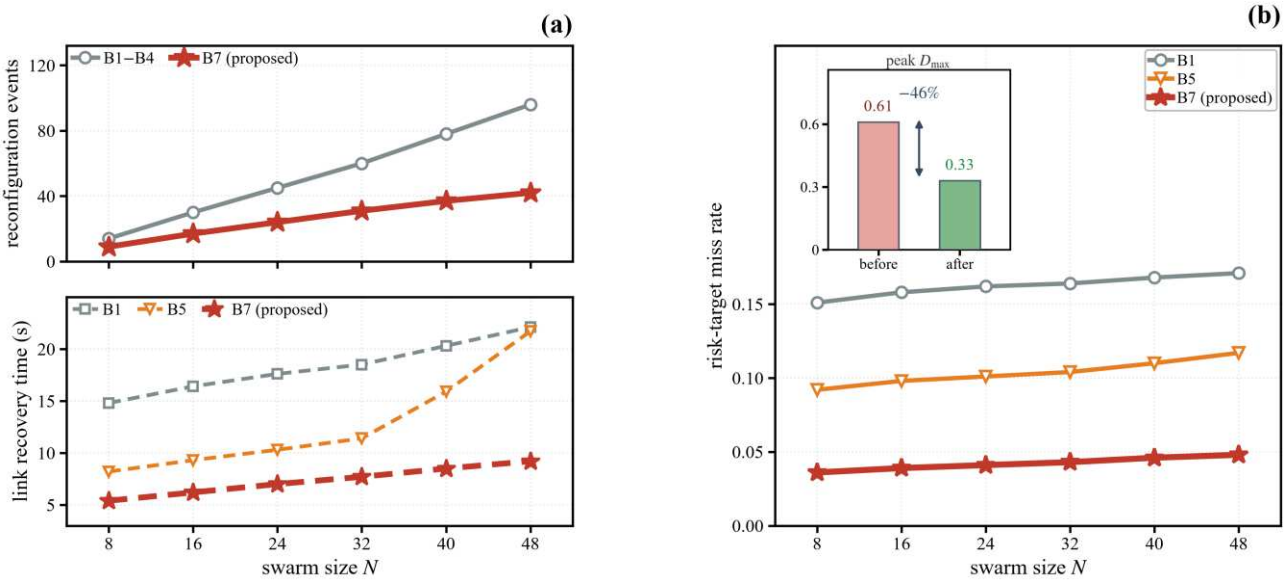

Fig. 10. Reconfiguration experiments. (a) Number of reconfigurations (top) and link-recovery time (bottom) against swarm size, sharing the horizontal axis and each with its own vertical axis. (b) Risk miss rate against swarm size, with an inset comparing the peak of the disagreement field D_{max} before and after reconfiguration (down 45.9%).

8.6 Ablation

In addition to the full method, we have constructed six variants of the module: V\T (value trigger replaced by the state-change criterion of B3), V\S (submodular selection replaced by truncation on confidence), V\W (water-filling replaced by a fixed rate), V\F (cognitive freshness replaced by classical AoI), V\A (attribution removed, ordering by magnitude only), and V\R (reconfiguration removed, geometric formation kept). Table 6 is as follows:

Table 6. Ablation results ($N = 32, M = 64$)

Variant	Task success	Consistency	Traffic / MB	Reconfigurations	Risk miss rate
CVSC-DFR-Full	0.898	0.938	118.5	31	0.043
V\T (state-change trigger)	0.842	0.874	186.2	34	0.079
V\S (non-submodular selection)	0.861	0.902	131.7	33	0.068
V\W (fixed rate)	0.850	0.913	142.3	32	0.071
V\F (AoI only)	0.869	0.887	128.4	35	0.062
V\A (no attribution)	0.883	0.921	122.1	38	0.055
V\R (no reconfiguration)	0.842	0.812	121.5	44	0.079

The reduction in task success due to the removal of value triggering is 5.6; that due to a fixed rate is 4.8; submodular selection is 3.7; and AoI substitution is 2.9. The Value Triggering function determines when and to whom to send a message, and Value-Filling accounts for 4.8 points on its own; therefore, in a narrow bandwidth, who receives a particular rate is as important as the size of that rate. V&A is Learning. Although the traffic is close to that of the full-method (122.1 vs. 118.5), it has been reconfigured 38 times instead of 31, and the miss rate has increased from 0.043 to 0.055; therefore, Attribution buys not traffic reduction but the efficiency of each move. Otherwise, the algorithm will continue to send the same type of platform to the same disputed area and repeatedly observe it for no reason. Most of the value of V\R is lost due to inconsistency (0.938 to 0.812, 13.4%), and no communication policy can overcome this physical limitation caused by the diversity of viewing geometry. Figure 11 shows the relative degradation and the rank of all variants; it is relatively sensitive.

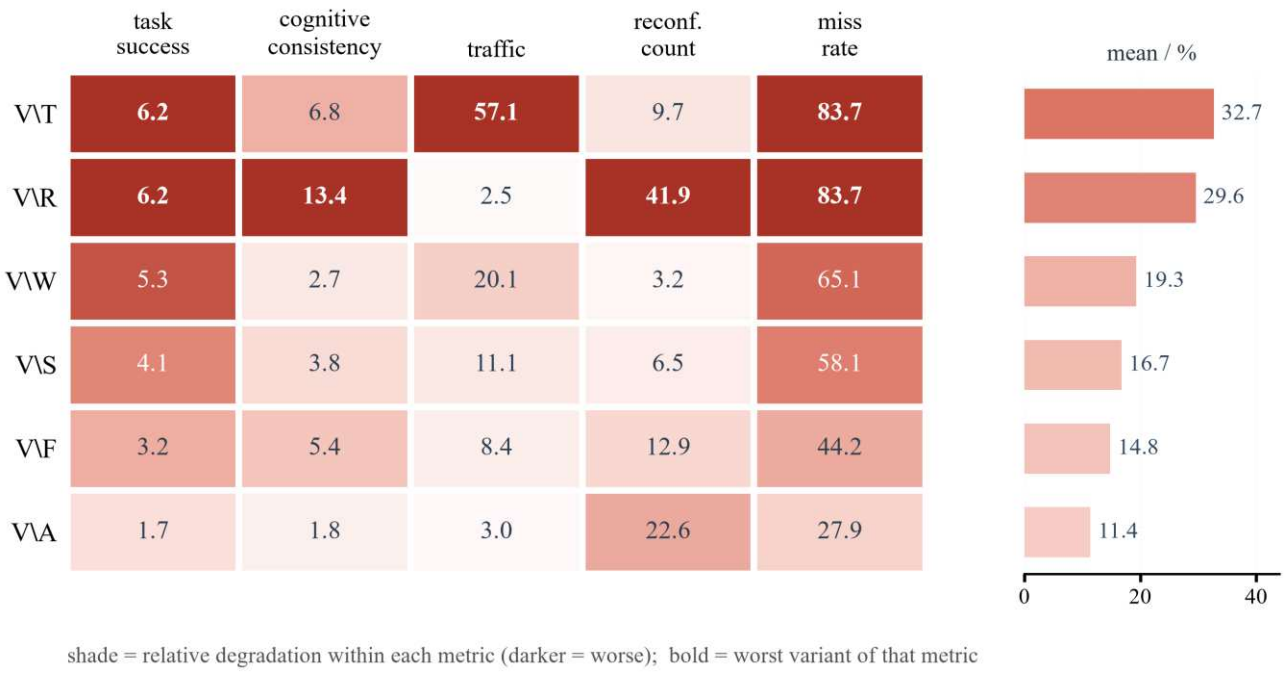

Fig. 11. Heat Map of Relative Degradation in the Ablation Study. Each cell is the degradation (%) of a variant compared to the full CVSC-DFR, and a higher value is worse; colours are normalised by column, darker colours indicate a larger loss in that metric, and bold entries are the worst variants in the column. The bars on the right are the five-metric means, and the rows are ordered according to these means. The degradation values in percentage points in the text are the same as the relative percentages here (e.g., 5.6 points is 6.2%).

8.7 Computational and Communication Costs

Table 7 compares the cost at $N = 32$. One round on one platform takes 11.4 ms with CVSC-DFR: 1.9 ms for value prediction, 2.6 ms for submodular unit selection, 0.8 ms for rate allocation, 4.4 ms for the formation subproblem, 0.9 ms for the ADMM consensus update and 0.8 ms for building and attributing the field. Against $T_s = 2$ s slots that is under 0.6% occupancy, comfortably inside on-board real-time limits. Centralised scheduling (B5) spends 38.6 ms per round and grows superlinearly, reaching 148 ms at 48 platforms: it still runs at a 2 s slot, but can no longer support slots below 0.5 s.

Table 7. Computational and communication cost at $N = 32$

Method	Compute per round / ms	Traffic per platform per round / kB	Total traffic / MB	On-board storage / MB
B1 Raw image	2.1	12.8	412.6	0.4
B2 Fixed-rate semantic	6.8	8.4	268.4	9.6
B3 State-change trigger	3.4	5.8	186.2	8.8
B4 AoI optimisation	4.2	6.6	214.7	9.1
B5 Centralised scheduling	38.6	9.2	152.9	10.4
B6 No-disagreement formation	9.7	3.9	121.5	12.2
B7 CVSC-DFR (ours)	11.4	3.7	118.5	12.6

Fig. 12 gives the scaling.Per-round compute for B5 rises from 5.2ms at 8 platforms to 148ms at 48 platforms, an increase of approximately O(($N^{1.85}$), and for B7, it increases from 4.1ms to 26.3ms by about O$N^{1.05}$). B7 scales according to the structure in (64); the value prediction and unit selection are linear in N, only the formation subproblem has an $N^{3.5}$ term, and its empirical exponent is significantly lower than the theoretical limit. Per-round traffic has risen from 2.9kB to 4.6kB (58.6%), and for B5, it has increased from 6.4kB to 18.2kB (184%). B7 is almost decoupled from the swarm size because it transmits semantic units and value residuals instead of the full state, and the queue threshold in (45) suppresses the redundant reports that would be generated by scaling.

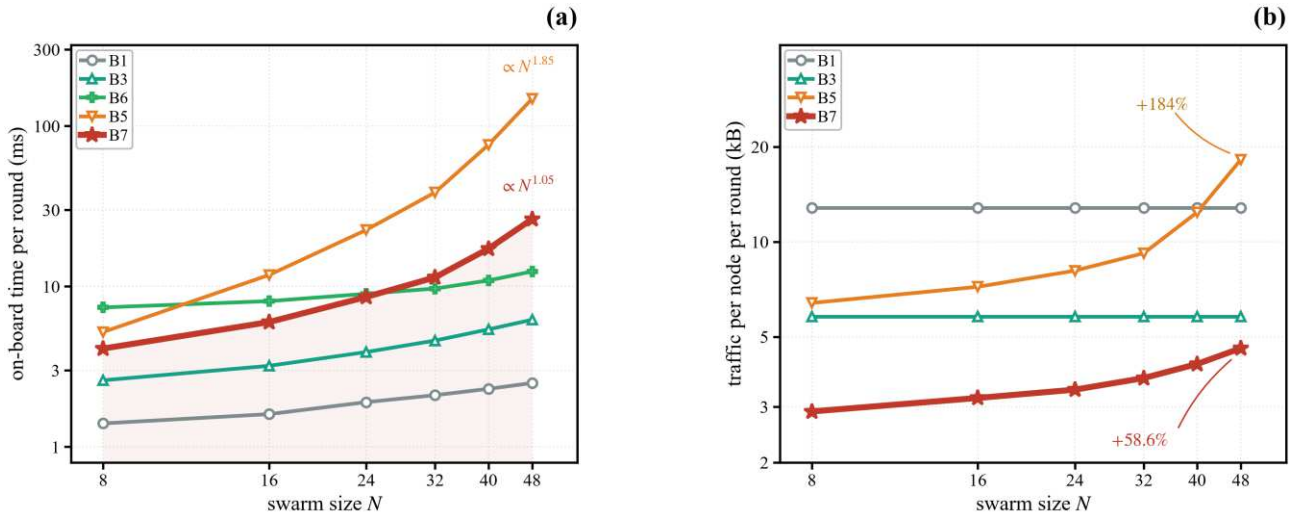

Figure 12. Cost Versus. Swarm Size: (a) Per platform per round, (b) Traffic per platform per round. Both panels are log-log plots, a power law is represented by a straight line, the exponents of the B5 and B7 curves are shown next to them, and panel (b) shows the relative increase from 8 to 48 platforms.

8.8 General Performance

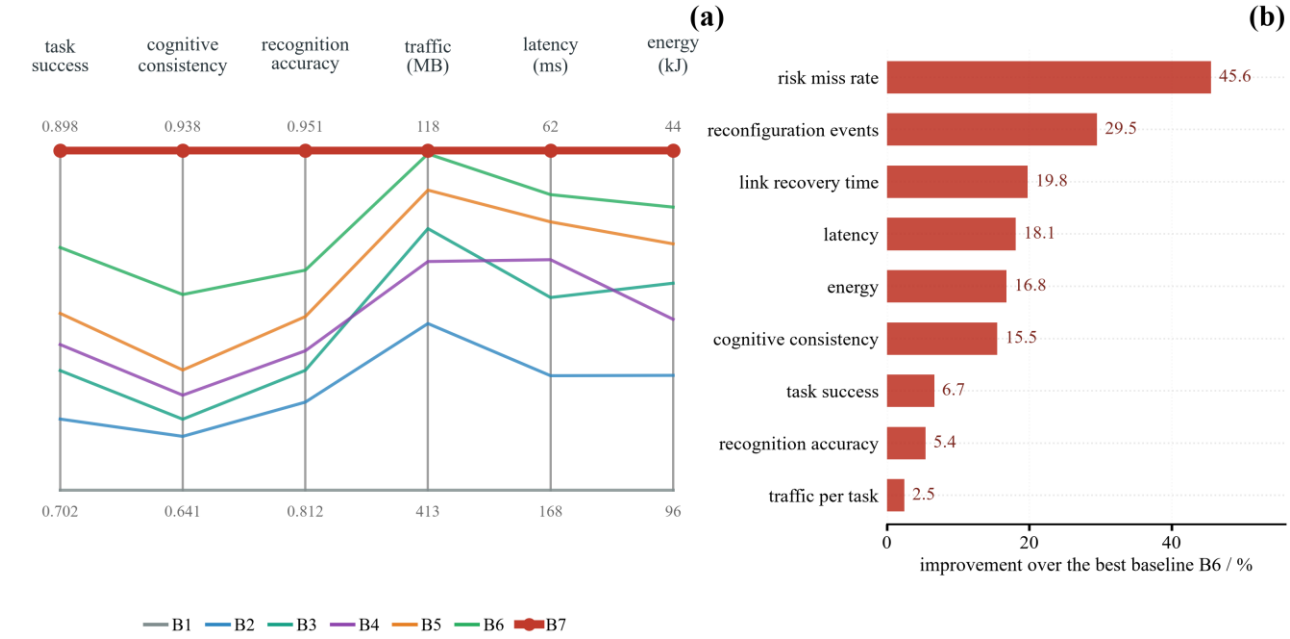

Figure 13. Overall Performance. (a) Normalised parallel-coordinate profiles of the six representative metrics, with each axis normalised to the seven methods, and a better value at the top and a worse value at the bottom; raw values are shown at both ends. (b) Improvement of B7 over the strongest baseline (B6) in nine metrics, and the positive values favour B7.

Figure 13 consists of the two parts mentioned above. Panel (a) shows the six normalised profiles of them: task success, cognitive consistency, recognition accuracy, traffic per task, mean delay and total energy, and each axis is one of the seven methods; the best value is at the top. The Red Polyline of B7 is the largest of the six axes. B1 is at the end of the road and going slowly. B6 has a good shape and is close to B7 in traffic, but it lacks the consistency, recognition accuracy and risk omission of B7; therefore, only communication-layer optimisation is insufficient to solve the problems caused by viewing geometry.

Panel (b) shows the margins of B7 over B6, and the highest baseline values are presented here: risk misses decreased by 45.6%, reconfigurations by 29.5%, link recovery by 19.8%, mean delay by 18.1%, energy by 16.8%, consistency increased by 15.5%, and task success, recognition accuracy and traffic per task improved by 6.7%, 5.4% and 2.5%, respectively. Traffic is a problem for B7 and is still relatively high; optimisation of the communication module has brought the per-platform traffic close to the information-theoretic limit, so the next step in improvement will be to change the focus of attention, and disagreement-driven reconfiguration can achieve this. All nine indicators have been rising together; therefore, the increase is not due to a drop in any single one.

9 Conclusion

Swarm systems based on large models have lost the original optimisation goal for their decisions and, therefore, CVSC-DFR has been developed to address this problem. On the communication side, the trigger is moved from a change in message to a change in decision value; an information value function consisting of task, consistency, risk and cost terms is assembled, computed, estimated online by a lightweight network trained through bootstrapping, has a generalisation bound from Rademacher complexity, and a trigger rule that includes a margin is set. Monotone submodular maximisation is used for key-token selection, with a $1 - 1/e$ guarantee; Multi-precision coding is a rate-semantic-distortion-task trade-off problem that can be solved by a closed-form water-filling allocation, and the scheduling method is a Lyapunov drift-plus-penalty rule that maintains queue stability and semantic return. A disagreement field at the source attribution level drives the joint relocation of observation and relay nodes under connectivity and dynamic constraints in the formation, and a Lyapunov function guarantees practical stability. The four classes of variables on the solution side are handled by alternating optimisation, successive convex approximation and ADMM, and have an $O(1/k)$ convergence rate and polynomial per-round complexity.

In simulations of 32 heterogeneous UAVs and 30 Monte Carlo repetitions, CVSC-DFR outperformed raw image transmission, fixed-rate semantic communication, state-change event triggering, age-of-information scheduling, centralised scheduling and formation control in all ten indicators: traffic per task, task success, cognitive consistency, recognition accuracy, mean delay, total energy, number of reconfigurations, link-recovery time, risk miss rate and value-prediction accuracy; it reduced traffic by 71.3%, increased success by 19.6 percentage points, shortened link recovery by 58.4%, and predicted value with an accuracy of 0.917. The Margin is not the sum of the modules. Value triggering modifies the decision to send (5.6 points), changes the amount of the send (4.8), alters what is sent (3.7), and adjusts where to look for consistency (13.4%), thus switching the swarm from a sending mode to a learning mode. As shown in Figure 13, Reading (B7) is outside the six normalised profiles and performs better than B6 (the best baseline) in all nine indicators by 45.6% for risk misses and 29.5% for reconfigurations.

Several Add-ons are available.. Linear weighting of the value function assumes that the four terms can be calibrated independently; however, in practice, risk and consistency may be coupled non-additively, and the marginal value of consistency increases sharply under high risk, so making the weights state-dependent is a natural extension. Cognitive freshness is a first-order drift model that may be too coarse for targets with periodic or bursty behaviour; therefore, the event dynamics of each semantic class can be learned. On-device fine-tuning of the value network will be subject to drift in the task distribution over time, and how to share updates across multiple platforms without creating a communication bottleneck is still an open problem in federated value learning. Finally, the experiments are in software-loop and ignore the combined effect of wind, visual degradation and link outages. Flight tests on a multi-UAV platform will be conducted next, along with a study of how value-aware behaviour accumulates over multiple missions.

References

[1] Yang Z, Yang J, Li Z, et al. Joint network configuration and resource management in autonomous UAV-assisted MEC systems via embodied agentic AI. IEEE Transactions on Cognitive Communications and Networking, 2026, 12: 8372-8386.
[2] Zheng X, Luo L, Yu H, et al. Quantum-enhanced multi-agent reinforcement learning for ubiquitous LLM inference via embodied UAV swarms. IEEE Transactions on Cognitive Communications and Networking, 2026, 12: 11006-11019.
[3] Zhao F, Sun Y, Lan J, et al. Adaptive semantic communication for UAV/UGV cooperative path planning. IEEE Transactions on Vehicular Technology, 2026, 75(7): 14356-14367.

[4] Xu X, Zhang B, Li R. A leader-assisted decentralized adaptive formation method for UAV swarms integrating a pre-trained semantic broadcast communication model. Drones, 2025, 9(10): 681. doi:10.3390/drones9100681.
[5] Starbuck B, Jang W, Sholapurkar S, et al. Implicit semantic control manifolds for learning-enabled multi-UAV coordination. Autonomous Robots, 2026, 50: 37. doi:10.1007/s10514-026-10265-4.
[6] He W, Yao H, Mai T, et al. Three-stage Stackelberg game enabled clustered federated learning in heterogeneous UAV swarms. IEEE Transactions on Vehicular Technology, 2023, 72(7): 9366-9380.
[7] Ding Y, Yang Z, Pham Q V, et al. Distributed machine learning for UAV swarms: Computing, sensing, and semantics. IEEE Internet of Things Journal, 2024, 11(5): 7447-7473.
[8] Li M, Dong A, Wang G, et al. Wireless resource optimization for UAV swarm cooperative sensing via multiagent multitask deep reinforcement learning. IEEE Internet of Things Journal, 2026, 13(9): 18429-18447.
[9] Jia Z, He J, He L, et al. Dynamic trajectory optimization and power control for hierarchical UAV swarms in 6G aerial access network. IEEE Transactions on Wireless Communications, 2026, 25: 3349-3362.
[10] Emami Y, Gao H, Li K, et al. Age of information minimization using multi-agent UAVs based on AI-enhanced mean field resource allocation. IEEE Transactions on Vehicular Technology, 2024, 73(9): 13368-13380.
[11] Li H, Li X, Wang H, et al. Semantic-aware UAV swarms for low-delay and energy-efficient data collection. IEEE Transactions on Wireless Communications, 2026, 25: 21701-21716.
[12] Han W, Du Y, Guo Y, et al. Computation offloading and resource allocation strategy considering user mobility in multi-UAV assisted semantic communication networks. Electronics, 2025, 14(20): 4067. doi:10.3390/electronics14204067.
[13] Tian J, Wei R, Jiang L. Formation construction and reconfiguration control of UAV swarms: A perspective from distributed assignment and optimization. Nonlinear Dynamics, 2024, 112(23): 21171-21191. doi:10.1007/s11071-024-10024-z.
[14] Wei L, Chen M, Shi S. Dynamic event-triggered consensus cost-based switching control for UAV formation with disturbances. IEEE Transactions on Intelligent Vehicles, 2024, 9(2): 3531-3543.
[15] Lu C, Wang T, Li T, et al. Event-triggered formation switching control for fixed-wing UAV swarm under DoS attacks and outside disturbances. International Journal of Systems Science, 2026. doi:10.1080/00207721.2026.2672075.
[16] Zhang Z, Hu L, Zhang D, et al. Constrained optimization and dynamic trade-off method for formation assignment of heterogeneous UAV swarms. Drones, 2026, 10(6): 428. doi:10.3390/drones10060428.
[17] Gaydamaka A, Samuylov A, Moltchanov D, et al. Dynamic topology organization and maintenance algorithms for autonomous UAV swarms. IEEE Transactions on Mobile Computing, 2024, 23(5): 4423-4439.
[18] Liu C, Xin X, Dai Y, et al. Cost optimization of UAV swarm network for persistent emergency communication. IEEE Transactions on Green Communications and Networking, 2026, 10: 1734-1748.
[19] Sarkar N I, Gul S. Artificial intelligence-based autonomous UAV networks: A survey. Drones, 2023, 7(5): 322. doi:10.3390/drones7050322.
[20] Jia Z, You J, Dong C, et al. Cooperative cognitive dynamic system in UAV swarms: Reconfigurable mechanism and framework. IEEE Vehicular Technology Magazine, 2024, 19(3): 90-101.
[21] Qi G, Kong R, Li Y, et al. A critical node identification method for group targets based on structural similarity and effective communication distance. Journal of the Franklin Institute, 2026, 363: 108340. doi:10.1016/j.jfranklin.2025.108340.
[22] Xu H, Ding Y, Dong Y, et al. Mathematics-driven state monitoring and tampering defense for a hierarchical heterogeneous UAV swarm communication system. Chaos, Solitons & Fractals, 2026, 197: 119306. doi:10.1016/j.chaos.2026.119306.
[23] Wang B, Sun Y, Dobre O A, et al. Digital twin-enabled task-driven UAV communications under uncertainty. IEEE Transactions on Vehicular Technology, 2025, 74(5): 8454-8459.
[24] Hu M, Zhang C, Deng W. Constructed encoded data based coded distributed DNN training for edge computing scenario. Physical Communication, 2024, 66: 102499. doi:10.1016/j.phycom.2024.102499.
[25] Zhang P, Tian H, Luo H, et al. A hybrid fast inference approach with distributed neural networks for edge computing enabled UAV swarm. Physical Communication, 2023, 60: 102129. doi:10.1016/j.phycom.2023.102129.
[26] Zhang L, Ma X, Zhuang Z, et al. Q-learning aided intelligent routing with maximum utility in cognitive UAV swarm for emergency communications. IEEE Transactions on Vehicular Technology, 2023, 72(3): 3707-3723.
[27] Zuo Y, Yao W, Chang Q, et al. Voting-based scheme for leader election in lead-follow UAV swarm with constrained communication. Electronics, 2022, 11(14): 2143. doi:10.3390/electronics11142143.
[28] Du Z, Luo C, Min G, et al. A survey on autonomous and intelligent swarms of uncrewed aerial vehicles (UAVs). IEEE Transactions on Intelligent Transportation Systems, 2025, 26(10): 14477-14500.
[29] Javed S, Hassan A, Ahmad R, et al. State-of-the-art and future research challenges in UAV swarms. IEEE Internet of Things Journal, 2024, 11(11): 19023-19045.
[30] Han X, Lin B, Wu N, et al. Design of multi-UAV cooperative deep semantic autoencoders for communication networks. IEEE Transactions on Vehicular Technology, 2026, 75(8): 17063-17077.
[31] Guo Z, Liu D, Zhang Z. Perception enhanced multimodal multitask semantic communication and resource management for UAV-assisted ISAC systems. IEEE Transactions on Vehicular Technology, 2026, 75(6): 11051-11065.
[32] Tang M, Feng C, Quek T Q S. Decentralized semantic communication and cooperative tracking control for a UAV swarm over wireless MIMO fading channels. IEEE Transactions on Vehicular Technology, 2026, 75(2): 3354-3359.
[33] Lu J, Qu Z, Liu A, et al. MLM-WR: A swarm-intelligence-based cloud-edge-terminal collaboration data collection scheme in the era of AIoT. IEEE Internet of Things Journal, 2024, 11(1): 243-255.
[34] Hu Z, Jin X. Formation control for a UAV team with environment-aware dynamic constraints. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 1465-1480.
[35] Luo Y, Tang F, Zhang H, et al. Synchronous position-attitude loop regulation-based distributed optimal trajectory tracking control for multi-UAVs formation with external disturbances.

IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(12): 7445-7456.
[36] Delansnay G, Ocampo-Martinez C, Dewasme L, et al. Partitioning and distributed predictive control of UAV swarms. Robotics and Autonomous Systems, 2026, 202: 105473. doi:10.1016/j.robot.2026.105473.
[37] Huang Y P, Lin C L, Chen Y Y. A molecular-dynamics-inspired decentralized framework for cooperative control and path planning of UAV swarms. International Journal of Aerospace Engineering, 2026, 2026: 2518752. doi:10.1155/ijae/2518752.
[38] Sarabia Sauquillo N, Gashaw H, Wubben J, et al. A safe in-flight reconfiguration solution for UAV swarms based on attraction/repulsion principles. Electronics, 2025, 14(19): 3799. doi:10.3390/electronics14193799.
[39] Yu H, Yang X X, Zhang Y. Adaptive cooperative fencing and dynamic reorganization control of UAV swarm for multiple maneuvering targets. Complex & Intelligent Systems, 2026, 12: 102. doi:10.1007/s40747-026-02229-z.
[40] Chen Q, Liu X, Nie W, et al. Safe formation scaling and motion planning for heterogeneous UAV-UGV teams in cluttered environments. Scientific Reports, 2026, 16: 37211. doi:10.1038/s41598-026-37211-9.
[41] Pu Q, Yang L, Li T. Dynamic coverage path planning method for UAV formations in multi-region aerial tasks. Aerospace Science and Technology, 2025, 167: 110683. doi:10.1016/j.ast.2025.110683.
[42] Alqudsi Y, Makaraci M. Towards optimal guidance of autonomous swarm drones in dynamic constrained environments. Expert Systems, 2025, 42: e70067. doi:10.1111/exsy.70067.
[43] Ekechi C C, Elfouly T, Alouani A, et al. A survey on UAV control with multi-agent reinforcement learning. Drones, 2025, 9(7): 484. doi:10.3390/drones9070484.
[44] Nian X, Li M, Wang H, et al. Large-scale UAV swarm confrontation based on hierarchical attention actor-critic algorithm. Applied Intelligence, 2024, 54(4): 3279-3294. doi:10.1007/s10489-024-05293-5.
[45] Feng L, Zheng H, Zhao Y, et al. Autonomous flight control for UAV swarm using evolutionary multi-agent multi-objective reinforcement learning. IEEE Transactions on Cognitive Communications and Networking, 2026, 12: 4673-4686.
[46] Wang B, Gao X, Xie T. An evolutionary multi-agent reinforcement learning algorithm for multi-UAV air combat. Knowledge-Based Systems, 2024, 299: 112000. doi:10.1016/j.knosys.2024.112000.
[47] Jiao K, Chen J, Xin B, et al. Multiagent reinforcement learning with evolution for multitarget tracking by unmanned aerial vehicle swarm. Applied Soft Computing, 2025, 170: 113463. doi:10.1016/j.asoc.2025.113463.
[48] Wang Y, Ling Y, Wang Y, et al. Multiple targets tracking with unmanned aerial vehicle swarm based on multi-agent deep reinforcement learning framework with safety network. Engineering Applications of Artificial Intelligence, 2026, 158: 115370. doi:10.1016/j.engappai.2026.115370.
[49] Zhu H, Hui J, Guo Z. Two-level-attention-based continuous trajectory design and computation offloading for multi-UAV cooperative target search. IEEE Transactions on Mobile Computing, 2026, 25(3): 3196-3214.
[50] Sun Z, He G. A simulation-driven hierarchical Stackelberg-DMPC framework for UAV swarm interception. Electronics, 2026, 15(16): 3740. doi:10.3390/electronics15163740.
[51] Zhao Z, Chen J, Xin B, et al. Learning scalable task assignment with imperative-priori conflict resolution in multi-UAV adversarial swarm defense problem. Journal of Systems Science and Complexity, 2024, 37(1): 369-388. doi:10.1007/s11424-024-4029-8.
[52] Li X, Liu J, Chen X, et al. Caching on the sky: A multiagent federated reinforcement learning approach for UAV-assisted edge caching. IEEE Internet of Things Journal, 2024, 11(17): 28213-28226.
[53] Teng F, Naouri A, Nouri N A, et al. Adaptive swarm intelligence optimization for unmanned aerial vehicle-assisted edge computing. Swarm and Evolutionary Computation, 2026, 91: 102355. doi:10.1016/j.swevo.2026.102355.
[54] Chen Y, Pi D, Yang S, et al. A dynamic optimization framework for computation rate maximization in UAV-assisted mobile edge computing. IEEE Transactions on Vehicular Technology, 2025, 74(7): 11395-11409.
[55] Lei G, Liang T, Ding H, et al. Joint communication topology formation and task offloading for heterogeneous UAV swarms. IEEE Wireless Communications Letters, 2026, 15: 4503-4507.
[56] Huang W, Guo H, Liu J. Task offloading in UAV swarm-based edge computing: Grouping and role division. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM), 2021. doi:10.1109/GLOBECOM46510.2021.9685920.
[57] Chen N, Zhu Q, Xi M, et al. Application of artificial intelligence in unmanned aerial vehicles: A survey. International Journal of Machine Learning and Cybernetics, 2026, 17: 383. doi:10.1007/s13042-026-03218-x.
[58] Zhou Y, Rao B, Wang W. UAV swarm intelligence: Recent advances and future trends. IEEE Access, 2020, 8: 183856-183878.
[59] Wu Q, Chen L, Liu K, et al. Reinforcement learning-based pathfinding for multiple UAVs facing abrupt hazardous areas. IEEE Transactions on Automation Science and Engineering, 2026, 23: 4848-4860.
[60] Gu J, Wang Y. A constrained reinforcement learning based approach for cooperative control of multi-UAV in dense obstacle environments. Science China Technological Sciences, 2026, 69(1): 1120601. doi:10.1007/s11431-025-3076-2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile.docx