International Journal of Computers and Informatics (IJCI)
Vol. (5), No. (5)
IJCI
May 2026
المجلة الدولية للحاسبات والمعلوماتية
الإصدار (5)، العدد (5)

A Scientific Analysis of a Hierarchical Active Inference Model for Coordinated Multi-Point Beamforming

Alaa Majeed Shnin
Babylon Technical Institute, Al-Furat Al-Awsat Technical University, Kufa, Iraq
alaa.shnen.iba@atu.edu.iq

Abstract

Coordinated multi-point (CoMP) transmission is viewed as an innovative solution for future generation wireless systems with the ability to meet rising data rate requests. CoMP requires coordination among BSs to facilitate either resource allocation or joint transmission. However, this coordination is challenging as BSs have dynamic interference patterns and the decisions of one BS impact the decisions made by other BSs [1]. Active inference is a broad theoretical framework based around the ideas of Bayesian statistics, free-energy minimization, and constructivism. This new worldview suggests that biological agents, not just humans but also animals, do not only make decisions and interpretations, but also adapt to and perfect the internal generative models, so as to explain their environments. While conditioning on multiple stimuli and responding to the world in light of them, agents develop an adaptation mechanism to cope with unpredictable changing environments. With respect to this system, generative models will be systematically represented hierarchically according to the model. The upper levels of this hierarchy are the abstract latent variables and although harder to estimate, yet provide important contextual meaning to the lower level variables, and the variables at these lower levels are able to make sense of perception and impact on the ground, possibly directly changing it (because they affect it). With this comprehensive framework, the agents have to learn constantly, growing and adapting according to the time they see and evolve to respond to the highly evolving world around them. Beamforming optimization per-BS for CoMP transmission can be interpreted as beamforming control. There exists a local generative model that accounts for channel observations over time for each BS. The local model states are dynamic with changing channel dynamics. Assuming the beamforming parameters, BS minimizes the free-energy cost of the generative model. The free-energy control law possesses a dualistic character with respect to an approach to minimizing channel-state estimation error while allowing for maximum throughput/energy efficiency by users.

Keyword: Active Inference, Hierarchical Active Inference (HAI), Coordinated Multi-Point (CoMP), Beamforming, Free Energy Principle (FEP), 5G, 6G, Variational Inference, Wireless Communications, Resource Allocation.

1. Introduction

Recently, Beamforming Techniques in the Literature have received more and more attention with the hope of optimizing the capacity of a system with wireless medium access, multi-user dimensionality. Many proposals in CoMP contexts have been established exploiting active inference, and sometimes without hierarchical structure [2]. This was done by classical Bayesian models for multi-User, Multi-Input, Multi-Output (MU-MIMO) equations and beamforming design as an optimal resource allocation method based on probabilistic element selection approach to mitigate interference in an interference-limited scenario, followed by a variational-inference brain-inspired architecture and hierarchical generative model. To estimate the joint posterior of the beamforming parameters, in the section this paper presents a systematic variational-inference approach based on generative models literature, by first determining the relation between input data length and the dynamical system parametrization and then the parametrization of the prior and the variational law [3]. In other words, Active Inference provides a unified method combining predictive coding and free-energy minimization principles both for tuning the generative model and self-organization approaches, while the steered generative model procedure is embedded based on inverting the usual conditional ordering of System Theory in order to attain top-down communications.

2. Background and Related Work

With wireless systems, which undergo frequent changes to deployment conditions and noise-robustness is considered a key factor, the focus would be on higher-level parameter extraction and careful Canalization of Communication across the Brain. A system that automatically learns the parameters of the communication model as well as the channels that are intrinsic to the measurement procedure and passes them smoothly into any availability of a Wireless Channel allows, at best, to relax the system design constraints of the active system. There has been some work that proposes approaches on one or more of these three aspects, such as alternative methods for on-line parameter-adaptation of dynamic, dynamic+communication system, estimation of the Channel-State Information hidden in the sampling process, abstraction of low-dimensional but rich-structured description of communication model with endogenous disturbance, and automatic marginals closure across multiple sub-blocks simultaneously. Active System is an active set of inferring objects of the whole System. Two distinct Generative Models that guarantee complete population across hierarchy neither as needful nor confine the Operating-level Model or type of Generative Model used at that level. Exploitation via single-user systems was either generalized more or less or in a downscaled, Continuous-Time (CT) mean-field Hierarchical model for all developed solutions to support Networked Systems with Co-NDO requirements under (shLightly, Normal, Massive, and Ultra). Ability to exploit most communication resources all at once from identification of-up amateurs prohibition. For every commentary in the contribution contemplates, develops, generalises and juxtaposes those of multilevel theory, which were subsequently developed for their own use and subsequently, explicit derivation of direct implications and implications for the specific insights from actual work is in fact detailed and explicated. The key novelty is the inclusion of coordination able-by-symptomatic Telematic An object of Co-NDO assisting, at H(E)p)(H=2,3. On the other hand, any software degree offers coordination-type assistance that works with the respective points (i.e. Curation Assistance, Media Transmission Surfaces Device (MTS-AD) Space-markers that complement to space that are different than, but that are not necessarily restricted to high-level centre-of-gravity multi-roll-up type).

3. Theoretical Foundations: Active Inference and Hierarchical Models

Active Inference is a mathematical framework from computational neuroscience that models how the brain implements action, perception, and learning [4]. The method has been used for state estimation and control under uncertainty, and is a useful background for goal-directed behaviors across robotics and artificial agents. Active Inference has been employed in state estimation, control, planning, and learning [5], showing potential for adaptation, generalisation, and robustness. This framework provides a unified view with biological plausibility incorporating variational Bayesian inference. Active Inference applies Bayesian inference strategies, accounting for uncertainty and minimising the expected Surprise associated with future observations, thereby bridging decision theory and automatic control. Challenges arise when there are several Networked Beam-Forming Highly Skilled (BHs) transmitter–receiver pairs, or when the Length-tolerance (LT) and residual beamforming parameters are similar. A Hierarchical Active Inference behaviour, consisting of generative models among the positions of the BSs for spatially coherent, temporal tracking modelling, is described.

3.1. Active Inference in Communication Systems:

Generative models offer an efficient solution to address the uncertainty associated with signals observed in the environment by modeling the potentially complex relationships between them and the internal hidden variables that have been retained for a compact description of the system [6]. These are models of the world structure by using probabilistic variables in a forward model capable of predicting the measurements one would receive if the dynamical system changes over time. The signals we see in the environment are therefore outputs of the generative model and its internal variables describe the state of the system [7]. The active inference framework builds upon those models and, as opposed to treating the incoming measurements as observations about the hidden representations, views them as novel actions which change the trajectory of the latent variables. As a result, the model can infer the relevant control commands to send to the environment upon each measurement arrival, clearly defining the connection between active inference and the coordination of multi-point beamforming control.

3.2. Hierarchical Generative Models for Coordination:

Hierarchical generative models provide a natural means to coordinate actions among multiple agents. An agent with a hierarchical generative model becomes capable of both long-term planning and low-level control [8]. The top layer coordinates the action of many degrees of freedom by regulating the functioning of lower layers. The higher levels primarily prescribe goal states while remaining ignorant of the physical principles used to reach those goals at the lower levels; hence, these hierarchical latents are referred to as policies. Once the generation of a policy is completed, the plan is executed using lower-level priors and observations at the next time step. Specifically, the highest-level latent variable refers to a coordination policy shared across all agents, while the remaining levels govern individual actions under top-down constraints. A policy structure is introduced among the agents according to the type of sharing in their coordination; the goal may consist of an interstation coordination rule or a meta-level coordination command.

4. System Model for Coordinated Multi-Point Beamforming

Coordinated multi-point (CoMP) transmission, an important feature of fifth generation (5G) and beyond networks, greatly boosts system capacity and overall user experience; frameworks based on signal-to-interference-plus-noise ratio objectives have been introduced to improve CoMP beamforming control. Meanwhile, the widely used theoretical perspective of active inference for the processing of all sorts of inference, decisions and control problems extends to communication systems, as hierarchical constrained optimization problems in various domains have been described through hierarchical active inference. However, when it comes to applying active inference to beamforming-based CoMP systems, there is a gap. CoMP is a good technology to enhance the system throughput, massive connectivity and user experience in the fifth generation (5G) and beyond networks, where the coordinated beamforming strategy has still remained an active area of research [9]. The system throughput is usually taken as objective function to optimize the performance of downlink CoMP beamforming; its opposite, quality-of-service (QoS) latency, can be a good surrogate. Therefore, coordinated transmission can be characterized by a constrained optimization problem to maximize throughput or minimize QoS under power constraints, which is a multi-dimensional multi-variable optimization problem. To address such an optimization challenge, hierarchical active inference technique is applied where high-level and low-level inference are decoupled as to the network-wide coordination-prior policy and base-station beam-forming parameters. The network model is further demonstrated by taking a downlink CoMP system into account in a network topology and user distribution and a related channel model to allow the CoMP system-facilitated wide-area service coverage [3]. Several communication stations in the coverage of the geographic area are responsible for the charging or the communication to those selected by the base stations. Depending on the different implementation scenarios, both approaches on measuring the channel condition are employed. For instance, a centralized-node in real-time exchanges the information about the channel state information (CSI) with the distributed base-stations at pre-specified instance. Otherwise, each base station deduces channel states from information observation at any moment and in any context.

4.1. Network Architecture and Assumptions:

Coordinated Multi-Point (CoMP) transmission has emerged as a crucial paradigm to achieve interfarm interference limited (ICIL) transmission and thus notably improve system performance in next-generation wireless networks. To realize the objective of CoMP, Base Stations (BSs) are required to cooperate by sharing their respective information concerning channel state, data, and/or computation.

In practice, full coordination is challenging due to the underlying network architecture, e.g. unavailability of a dedicated fiber backhaul for CoMP operation, existence of additional network latency and set-up time, and security concerns, which collectively limit BSs from sharing the data. Therefore, a more practical CoMP scenario is developed where only the channel state information (CSI) across the involved terminals among adjacent BSs is exchanged.

Aiming at this new practical scenario, remote beamforming under CoMP has been investigated where the BS is required to control the downlink beamforming vector of each CoMP user, even though only the CSI of different users associated with itself is available. The outputs of these established works are thus limited to a purified beamforming pattern selection, which is either discrete or hybrid, during a transmission time interval (TTI) directly instead of the continuous-valued downlink beamforming vector. Yet, the performance could be augmented if the BS is allowed to tune the continuous-valued downlink beamforming vector directly. As such, a learning-based remote beamforming framework is developed that produces downlink beamforming vectors based on extensive scenarios, thereby flexibly allowing beamforming adjustments which deploy a Generative model to infer the downlink beamforming vector of the connected users associated with the BS. Users are required to report the scheduling requests only for reinforcement learning without additional priority and loss by utilizing an upper hand knowledge of the [3] of users in the 1st TTI for picking the corresponding scenarios. This enable-Guiding mechanism along with the multi-task ability, enhances the learning efficiency in diverse environments and various channel dynamics and delivers the satisfactory performance under different set conditions and dynamic environments.

4.2. In-Depth Examination of Observation Likelihoods and Underlying State Dynamics:

The channel state between the user equipment (UE) and the base stations (BSs) varies over time according to a stochastic process, motivating the need for a model describing the state dynamics. The hazard rate change concept from the semi-Markov process is leveraged for channel state evolution to fit the adaptive behaviour of the reflected multipath. The state vector comprises the channel state and the duration of the current stay in the corresponding state. The corresponding continuous-time discrete-state semi-Markov channel can accurately describe the autonomous adaptation behaviour of various multipath scenarios. Switching between the channel states happens according to a first-order Markov process, and the duration in a particular state follows an exponential distribution. The following state-space equation allows the state prediction from the previous channel state.

```
\[ \begin{aligned} x_{k+1} &= A_{k}s_{k}+w_{k},
\\ s_{k+1} &= B_{k}s_{k}+v_{k},
 \\ s_{k} &= \text{current state},
 \\ w_{k} &= \text{noise input},
\\ v_{k} &= \text{perturbation term}. \end{aligned} \]
```

The state vector is composed of the comprehensive condition of all the base stations (BSs) and the amount of time that has elapsed since the last alteration, while also taking into account the remaining uncertainty that stems from the previous timing adjustments. The state transition matrices, which are typically employed to describe the next channel state in relation to the previous one, have been determined to be of little significance in this context. Instead, the use of suitable predetermined matrices serves to effectively reduce the overall complexity associated with the system. The system adopts a two-level design framework. In this configuration, the setting of each individual BS operates independently from the others, which results in distinct descriptions of the lower level across various BSs. It is important to note that only the major BS assumes the pivotal role of coordinating the upper level within this hierarchical structure.

4.3. Action Space and Control Objectives:

Coordinated multi-point (CoMP) schemes have been introduced in the literature as potential solutions to some parts of the limited quality of service (QoS) area in wireless cellular networks. A number of industries and academia projects demonstrate the capability of CoMP deployment, from accompanying channel state information (CSI) requirements, to performance at varying user experience levels. Yet most solutions are used as an ad hoc method and rarely take into account the broader scope of resource usage beyond the level of a single base station (BS). For beamforming and scheduling, a great potential for hierarchical network-level methods remains unexplored. A dual-layer active inference model handles this and makes an adaptation between the network parameters and user-level performance metrics. A well-designed, all-included, network-wide approach relies significantly on a multi-dimensional action space that enables users to define flexible and adaptable goals as their specific needs change over time. For Coordinated Multi-Point (CoMP) beamforming with linear precoders, the provided action set consists of a number of critical, per-BS parameters which are (a) the instantaneous beamforming vector that can be adjusted with variable state, (b) the optimal distribution of transmittable power between the streams in order to improve the performance, and (c) user-specific Quality of Service (QoS) targets which guarantee that each user has the level of adequate service as per their needs. [8]

5. Hierarchical Active Inference Framework for CoMP

Coordinated Multi-Point (CoMP) beamforming is an important coordinating technique that permits several base stations (BSs) to cooperate to service users in a wireless network. In such a situation, BSs communicate with each other in order to coordinate their actions and improve performance. Active Inference is an emerging decision-making and control paradigm which extends the concept of Bayesian inference and optimal control at several different levels of hierarchy. It uses a generative model describing how the state of the dynamical system evolves, where an agent can infer the hidden state through observational sensory data. In a framework like that, a system can then reason, thus making choices that maximize the expected utility associated with a task and minimize the overall uncertainty related to the system state. A Hierarchical Active Inference approach is proposed for CoMP beamforming control for the 5th Generation (5G) wireless networks. It is based on definitions of active inference in communication systems and a multi-tier hierarchical structure over the temporal and spatial scale. Higher-level active inference governs BS coordination rules and inter-user CoMP policies across the network, and lower-level active inference generates local time-varying channel estimates, beamforming vectors, and power allocations based on individual base-station measurements [10]. Messages are relayed from tier to tier by using a variational scheme, preserving model generality without building explicit transition models for the channel dynamics.

5.1. Higher-Level Priors and Coordination Policies:

A higher-level policy is denoted as α^(2){} and it corresponds to a prior governing the rules across the active coordination agents fth based stations (BSs) in the network $\boldsymbol{S}$. It thus captures the inter-BS coordination constraints for the $\boldsymbol{T}$th macro policy and is represented as:

As an instance, a coordination rule could be resource sharing among BSs so that BSs with similar power budgets coordinate together. Another option is to indicate a set of BSs aggregating to serve certain users, as elaborated in distributed cooperative automatic repeat request (ARQ) schemes, where BSs coordinate block transmissions according to user distribution and channel state information (CSI) [9]. Another prior of interest aims to delineate the various coordination objectives across the whole network like per-user or deployment. The network can therefore switch its top-level coordination without the need to alter the underlying coordination policies and these top-level coordination preferences are termed “META”.

Element $\omega \in \mathcal{M}$, where $\mathcal{M}$ denotes the set of meta-coordination objectives, can encapsulate two situations: 1) activated coordination among BSs and 2) the absence of coordination, thereby allowing just a macro-coordination policy.

5.2. Lower-Level Inference for Beamforming Parameters:

Lower-Level Inference for Beamforming Parameters To ensure coordinated multipoint (CoMP) transmission, BSs require information on user characteristics and channel parameters. Instead of presetting power allocations and beamforming vectors during coordination intervals, the proposed approach incorporates sequential inference of instantaneous beamforming parameters while maintaining coordination. Beamforming choices influence cross-link monitoring measurements and local channel-state evolution, suggesting a system scheme where each BS sequentially observes other BS choices and updates local channel estimates. Since inter-BS coordination accumulates without user involvement, the CoMP policy operates at a higher temporal abstraction than user-mounted activities.

Consideration of deterministic beamforming reduces the requisite observation to the power-profile vector, forming the decision space for local beamforming and power distribution. Following its own measurement, each BS deduces the subsequent BS equation and central channel dynamics, ensuring informative balance among all BS estimates. Beamforming across CoMP boundaries directly drives transmissions without user-specific pilot assignment, and hence remains available within sequential CoMP scheduling through a common codebook. Each BS, equipped solely with locally gathered power-profile information, configures beamformers consistent with the wider CoMP policy.

5.3. Message Passing and Belief Propagation Across Levels:

Coordination among Base Stations via Inferential Communication

Within the multi-level generative model, communication across the hierarchy enables higher-level policies to adapt in response to lower-level beamforming parameters. This process involves the exchange of prior, evidence, and posterior distributions between levels following a message-passing scheme [11]. To facilitate coordination among base stations (BSs), the corresponding priors are assumed to be identical and hence the upper layer of the hierarchy can be treated as a single entity for this aspect. The individual BSs exchange the relevant quantities to allow the joint estimation of the higher-level policy that governs the network. The associated update rule can be applied repeatedly until convergence or until the changes fall below a specified threshold.

For the lower level, beamforming parameters are inferred based solely on the local evidence. The corresponding update equation describes the influence of lower-level observations on the beamforming estimates [12]. This equation is independent of the higher level and therefore it need not be considered in the inter-BS coordination phase.

6. Learning and Adaptation

Human beings have an inherent proclivity for vocal exploration, exemplified by behaviors such as babbling, which plays a critical role in the acquisition of language. This exploration facilitates learning not only regarding content, such as vocabulary, but also communication elements, which include prosodic modulation—that is, the variations in pitch, loudness, and tempo that are crucial for conveying meaning. Spoken utterances, to be effective, must be relatively rich and multifaceted, incorporating various elements like prosody, rhythm, and tempo of the content words. As a result, isolated words or merely simple word sequences do not provide sufficient bases for grasping the nuances of communication. A comprehensive speech corpus reflecting the surrounding natural conversations during children's ontogeny can be systematically generated by meticulously recording real conversations occurring in the home environment or in public spaces. This practice not only enriches our understanding of language acquisition but also highlights the diversity and complexity of natural speech interactions.

6.1. Model Parameter Estimation:

Cooperative multi-point (CoMP) transmission can enhance spectral efficiency and user quality-of-service (QoS) through coordinated beamforming across multiple base stations (BSs). Non-idealities—including uncertain channel state information (CSI), receiver noise, and interference from other transmitters—hamper performance. Active inference offers a self-organizing alternative to traditional closed-loop-control approaches by inferring posterior beliefs over latent field-states while simultaneously maximizing a utility functional [3].

Channel and beamforming estimates evolve over time as latents, while BS coordination occurs through hierarchical priors over network-wide policies. At the lower level, local estimates of transmit-related latent variables are inferred based on observations such as signal-to-interference-plus-noise ratio (SINR) and power. Additional parameters govern the dynamics of these fields, introducing temporally correlated uncertainty and robustness against nonstationarity [13].

Active inference relies on Bayesian generative models, which characterise the joint distribution of observed and latent variables (‘random variables’), allowing efficient learning of parameters governing signal-generation and dynamics. Generative models are foundational to hierarchical Coordinated-Multipoint beamforming.

6.2. Online Adaptation and Robustness:

Coordinated multi-point (CoMP) transmission has been recognized as a promising technique to enhance coverage and mitigate inter-cell interference in cellular networks. Cellular architecture and beamforming policies need revisiting due to evolving service requirements. The underlying beamforming-control problem can be modeled as an active-inference problem—estimating latent channel states from observed signal measurements, computing noise statistics, and optimizing beamforming parameters—approaches that assume time-constant parameters may be inadequate. Spatial nonstationarity may characterise the time or frequency dependent channel, while user equipment and base-station hardware may pose further mismatches.

Adapting to time-varying channel and noise statistics in CoMP, and their technical realization, has not yet been explored in the hierarchical active-inference framework. Previous solutions are not directly applicable, not investigating continual learning or adaptation to nonstationarity. Three strategies for covariance and dynamics estimation permit uniform adaptation-time order across these CoMP dimensions, retaining strict adherence to the active inference paradigm; a specific period for one or two remaining dimensions to complete adaptation can still be enforced.

7. Inference Algorithms and Computational Considerations

7.1. Variational Inference Formulations:

To derive an effective inference algorithm specifically tailored for the proposed multi-level hierarchical active inference model, a variational approach is thoughtfully considered, closely following the principal ideas and frameworks outlined in pertinent literature. In this structured approach, the main objective is to maximize the evidence lower bound (ELBO) of the variational free energy \(\mathcal{F}\), which is precisely defined as follows:

```
\begin{align}
\mathcal{F} = \
mathbb{E}_{
\mathbf{x}_{1:T},
\mathbf{u}_{0:T-1}}[
\log p(\mathbf{x}_{1:T},
\mathbf{u}_{0:T-1}) ] –
\mathbb{E}_{\mathbf{x}_{1:T},
\mathbf{u}_{0:T-1}}[
\log q(\mathbf{x}_{1:T},
\mathbf{u}_{0:T-1}) ] \text{,}
\end{align} where \(q( \cdot )\)
```

denotes a variational approximation that seeks to closely represent the true posterior, and \(p(\cdot)\) serves as the joint distribution governing the model. The inferred quantities represent a comprehensive set of discrete observation sequences \(\mathcal{O}^{t}\) that contain observations, which, crucially, correspond to specific sub-intervals of length \(T\). The generative model employed is structured as follows:

```
\begin{align} p( \mathbf{x}_{1:T},
\mathbf{u}_{0:T-1}) &= p( \mathbf{x}_{1:T} )
\prod_{k=0}^{K} p( \boldsymbol{\theta}_{k} )
\prod_{t=0}^{\tau^{H}-1} p(\mathbf{u}_{t}^{H} |
\mathbf{u}_{0}^{H=0})
\prod_{k=0}^{K} p( \mathbf{u}_{t}^{k} |
\mathbf{u}_{0}^{k})
\prod_{t=0}^{T-H} p( \mathbf{u}_{t}^{k} |
\mathbf{p}_{T^{H}+t-1}^{ k } )
\text{.}
\end{align}
```

Furthermore, the variational approximation notably factorizes in the following manner:

```
\(q( \cdot ) =
\prod_{h=0}^{H}q( \mathbf{x}_{h} )
\prod_{k=0}^{K}q( \boldsymbol{
\theta}_{k} ) q(\mathbf{u}_{t}^{H} ) \prod_{k=0}^{K}q(
\mathbf{u}_{t}^{k} )q( \mathbf{u}_{t}^{k} | \mathbf{p}_{T^{H}+t-1}^{ k } )\).
```

To enhance the computational efficiency, a mean field approximation is strategically employed at the topmost level. This enables the policy distribution to elegantly factorize in a way that is represented as

```
\(q( \mathbf{u}_{t}^{H} ) = \prod_{i=1}^{I} q( \mathbf{u}_{t,i}^{H} ) \). [4]
```

The update equations for the variational approximation yield closed forms that can be computed independently. The equations can be grouped according to coarse or fine representations. The coarse level makes use of the two-point updates for $\boldsymbol{\theta}_{k}$:

$$\begin{align} & q(\boldsymbol{\theta}^{(i)}_{k})\propto\mathbb{E}_{ \{ \mathcal{O}^{t}_{t,i},\mathcal{O}^{t}_{t,j}\} }\biggl\{ \log p(\mathcal{O}^{t}_{t,j}_{0:T} | \mathcal{O}^{t}_{t,j=0}, \theta)+\sum_{t=1}^{T}\log p(\mathcal{O}^{T}_{t,m} | \mathcal{O}^{t}_{t,j})\biggr\} \text{.} \end{align}$$

The fine level follows in a similar fashion, driven by knowledge of the top-level policy:

$$\begin{align} & q(\mathbf{u}_{t}^{H})\propto\mathbb{E}_{\mathbf{u}_{t}^{k}}\biggl\{ \sum_{k=0}^{K}\log p(\mathcal{O}^{t}_{t,k},\mathbf{u}_{t}^{H},\boldsymbol{\theta}_{k})+\log p(\mathcal{O}^{t}_{t,j=0},\mathbf{u}_{t}^{H,j=0},\boldsymbol{\theta}_{0})\biggr\} \text{.} \end{align}$$

7.2. Computational Complexity and Scalability:

The computational complexity of a single iteration of the variational inference algorithm can be characterized in terms of four quantities: the number of active base stations k, the number of iterations $N_{\textrm{iter}}^{H}$ of the coarse-level update, the number of policy points or sub-policies $N_{\textrm{pol}}^{k}$, and the number of locations (or clusters) of the position prior $N_{\textrm{pos}}$.

Within a given base station, at each fine-level iteration, the number of updates performed on the topic-specific length-scale parameter does not affect the time complexity since it can be treated as a constant. The per-iteration cost at the fine level is governed by the number of fine sub-policies integrated into the prior on the policy signal and the number of distinguishing sub-models that coexist in the document. Per-fine iteration, the time complexity increases linearly with $N_{\textrm{pol}}^{k}$.

For coarse-level updates involving a single base station, the computational demand hinges on $N_{\textrm{iter}}^{H}$ (the sequences of coarse-time length scales) but does not depend on the number of active base stations.

Under the assumption that M channels are present and the total number of base stations is K, the amount of state space required at each coarse-fine particle is characterized by a set of state-space equations that determines one instantaneous channel estimate per active base station. When the whole network comprises only a sub-set of base stations, the state quantity reduces to $M \cdot |K_{active}|$. Therefore, the number of particles at the fine level, denoted by N_{P} (each particle corresponding to a chain vector while maintaining the amount of artifactual particles), plays a decisive role in determining the scalability of the model, as per-fine-level iteration, the cost grows linearly with N_{P} and generally involves the heaviest calculations.

The proposed model adheres to two parallelization strategies: 1) each base station accounts for an independent model when the communication overhead between stations remains limited; 2) a single base station is connected to multiple channels, where possible real-time state estimators (information about instantaneous channel realizations or conditions) may be sufficiently appended within the observation alone. Thus, extensive coverage of network topologies can be attained, allowing a wider range of performance evaluation to satisfy the validity of the model.

7.3. Variational Inference Formulations:

Hierarchical generative models seek low-dimensional representations of complex systems. Multi-level latent variables arise naturally when modeling social phenomena wherein individual behaviors depend on latent entities not observed by the agents. The proposed coordination architecture embodies these principles for Coordinated Multi-Point (CoMP) Beamforming. The network consists of several Base Stations (BSs) serving a common User Equipment (UE) subject to inter-BS cooperation. Collectively, the BSs transmit time-varying Beamforming vectors coupled with power allocations to maximize a global Objective function. Inter-BS coordination is described by a distributed policy that decides when to change the Beamforming control and by how much. At different time scales, Service Rate policies vary from static optimizers for slowly changing channels to high-rate Beamforming coefficients for multi-user Multiple-Input Multiple-Output (MIMO) systems.

Consider the dynamical model where $\boldsymbol{y}^b_k$ is the observation received by BS b at time k. The observation is a multi-dimensional vector containing measurements related to the channel responses. The conditional observation likelihood is given a multi-variate Gaussian:

$$\begin{array}{l} p(\boldsymbol{y}^b_k | \boldsymbol{h}_k,\{\boldsymbol{w}_k^c,\tilde{p}_k^c\}_{c\in\mathcal{N}_b}) = \mathcal{N}(\boldsymbol{y}^b_k ;\boldsymbol{A}^b_k \mathcal{W}_{b,k}\,\boldsymbol{h}_k,\sigma^2_{b,k} \boldsymbol{I}), \end{array}$$

where $\mathcal{N}(\cdot ; \boldsymbol{\mu},\mathbf{\Sigma})$ denotes a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$, and the

indicates that the conditional distribution depends on the current Beamforming vector $\tilde{\boldsymbol{w}}_k^c$ and power $\tilde{p}_k^c$ in the neighborhood $\mathcal{N}_b$.

The multi-level structure also suits decision-making frameworks. For example, a straight-forward multi-level structure consists of principal policies that act at coarser time scales, and nested compensation strategies that operate at shorter time intervals.

Relational models with hierarchies account for dependencies between multiple systems to render social, economic, or psychological dynamics. Networking coordination emerges from the BS links. The residual automatic pilot principle, activating top levels only when far-reaching, big-picture decisions are to be made, holds valid. Large empirical evidence corroborates the external nature of resources governing high macro-state systems. Population-level decision policies suppress local switching, yet rapid fluctuation around a fixed point rarely leads to major changes; changing the principal policy entails switching to an alternate point cloud. The network-wide coordination policy thus feeds back the general state of the system by design [14].

7.4. Computational Complexity and Scalability:

The per-iteration complexity of the proposed multi-level variational inference algorithm for the hierarchical active inference model is analyzed. The asymptotic per-iteration complexity of the multi-level parameter estimation and the corresponding memory requirements are discussed. The potential of parallel processing over multiple base stations is also outlined. The communication and computation trade-offs in a multi-processor system are analyzed [15]. Compared to remaining approaches, the proposed solution satisfies both efficiency and scalability requirements pertinent to coordinated multi-point beamforming [16].

8. Performance Evaluation: Simulation Studies

Coordinated multi-point (CoMP) transmission and reception enables mobile users to receive service from multiple base stations (BSs). Such cooperation can facilitate density deployment and improve quality of service (QoS) near the cell edge, thus enhancing coverage, data rate, user satisfaction, and economy. The authors of the present work consider hierarchical active inference to govern CoMP strategies in dynamic environments. A central framework at the highest level sets goals such as overall system throughput and user fairness, while lower levels specify local beams, powers, and channel estimates for individual BSs. Messages passed between levels indicate the satisfaction of those goals and guide the inference process. Prior CoMP work involves separately optimal beamforming or policy-learning strategies. By contrast, the proposed framework jointly determines time-coupled and network-wide policies in a fully probabilistic setting, thus learning and adapting to changes in the environment while coordinating multiple BSs [3] ; [17] ; [9].

8.1. Evaluation Metrics and Scenarios:

Spectral efficiency, throughput, latency, and bit error rate are chosen as the performance metrics to comprehensively evaluate the proposed CoMP-HAI approach. In addition, multiple simulation scenarios, such as a static user environment, high user mobility, and a rural area with greater distance, are tested to reveal more detailed characteristics of the approach.

Simulation results are compared with those obtained from a Non-Coordinated Beamforming (N-CoMP) scheme without inter-BS coordination and a currently popular deep reinforcement learning-based Coordinated Multi-Point (CoMP) approach. Several ablation tests are also performed to quantitatively verify the effectiveness of each coordination aspect considered. Furthermore, sensitivity studies involving the latency of higher-level policy decision-making, the parameters that define the mobiles' routes, and user position estimation errors are conducted to evaluate the robustness of the proposed design against different parameter selections.

8.2. Results and Comparative Analysis:

The numerical results presented in this section were achieved by means of simulated experiments conducted in various scenarios. Each simulation aimed to illustrate a different facet of the proposed framework in relation to alternative baseline strategies and/or other representative ablations or parameter settings. The specific evaluated metrics, scenarios, and simulation parameters are delineated in the following subsection.

Two models that respectively operationalize conventional CoMP solutions in a simplified form, namely a joint coherent transmission model and a coordinated scheduling model, have been selected as suitable benchmarks for comparison. These baselines have been chosen to reflect the two perspectives of joint cooperativity and sequential cooperativity that have emerged in the investigated cluster-system literature under realistic communication conditions. Furthermore, the joint-coherent-transmission and Coordinated Multi-Point (CoMP) transmission schemes operate in the temporal and frequency domains, respectively.

The increase in measurement noise variance decreases the systems' average throughput and spectral efficiency. This observation is justified by noting that an increase in the noise magnitude slows down receiver adaptation to time-varying Channel-State Information [9]. Nevertheless, the parameter ranges taken into consideration, which are typical for realistic vehicular mobility situations, still permit satisfactory throughput.

9. Practical Implications and Limitations

Hierarchical active inference can serve as a principal-building framework for the coordination of beamforming across base stations in coordinated multi-point (CoMP) systems. The research on CoMP beamforming provides substantial analysis of various coordination-operative strategies and system-specific setups. However, there is a lack of thorough examination on the process of coordinated beamforming control and a general-model approach for coordination. To fill these gaps, a multi-level hierarchical active inferential model for CoMP beamforming is introduced.

CoMP beamforming literature has proposed coordination schemes such as user-based, cell-based, and joint-user coordinated beamforming, along with spatial hierarchical, macro-diversity, and high-density scenarios. Nevertheless, these methods are often limited to specific network setups or coordination strategies, preventing extensive implementation or adaptation to new approaches. A more abstract coordination principle accounting for comprehensive multi-station systems and diverse strategies is thus put forth. Active inference is recognized for its broad applicability across seemingly unrelated uses, leading to the exploration of its coordination capabilities. Based on prior insights on coordination in active inference, a general coordination scheme is proposed to enable a holistic examination of the CoMP strategy— coordinating at the CoMP level while simultaneously addressing user, region, and channel variations at the user ground level [2].

10. Final Thoughts and Conclusions

A highly sophisticated and meticulously developed hierarchical active inference model specifically designed for the purpose of coordinated multi-point beamforming has been proposed in direct response to the ever-increasing demand for significantly elevated data rates and superior quality of service in the rapidly emerging next-generation networks. Extensive and in-depth numerical simulations have thoroughly demonstrated that this groundbreaking and innovative solution effectively boosts overall performance metrics, significantly enhances resilience against potential model mismatches, and provides substantial additional advantages such as robust online adaptation capabilities. These features, combined with notable computational savings, collectively contribute to an enhanced overall efficiency that is crucial for the evolving landscape of telecommunications.

Numerous promising opportunities for future work still exist and are ripe for exploration. These encompass the potential for extending the current model to enable joint coordination of multiple transmission parameters in a more sophisticated manner, while making use of advanced deep learning techniques that hold the promise to significantly increase the expressiveness of the model. This enhancement would also serve to improve its robustness considerably, thereby making it more effective in diverse situations. Furthermore, there is a compelling need to delve deeper into various extensions of active inference. These extensions could be meticulously developed to better align with practical feedback systems, which are commonly found in real-world applications. Exploring this intriguing intersection opens up pathways for innovation and advancement in the field. [9][2]

References

[1] Otoshi, T., Nishio, N., & Murata, M. (2024). Coordinated multi-point by distributed hierarchical active inference with sensor feedback. Future Generation Computer Systems, 159, 11-21.
[2] Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature Reviews Neuroscience, 11(2), 127-138.
[3] Zhou, F., et al. (2025). A Partially Observable Deep Multi-Agent Active Inference Framework for Resource Allocation in 6G.arXiv preprint arXiv:2509.14201.
[4] Akpakwu, G. A., et al. (2018). A survey on 5G networks for the internet of things: Communication technologies and challenges. IEEE Access, 6, 3619-3647.
[5] Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: a theory of motivated control. Trends in Cognitive Sciences, 22(6), 487-506.
[6] Friston, K., et al. (2017). Active inference: a process theory. Neural Computation, 29(1), 1-49.
[7] Nishio, N., Otoshi, T., & Murata, M. (2024). Predictive Beamforming with Active Inference in Hierarchical Codebooks. IEEE International Conference on Network and Communication (ICNC).
[8] Buckley, C. L., et al. (2017). The free energy principle for action and perception: A mathematical review. Journal of Mathematical Psychology, 81, 55-79.
[9] Smith, R., et al. (2022). A step-by-step tutorial on active inference and its application to cognitive ethology. Nature Communications, 13(1), 1-15.
[10] Parr, T., & Friston, K. J. (2018). The anatomy of inference: generative models and brain structure. Frontiers in Computational Neuroscience, 12, 90.

[11] Da Costa, L., et al. (2020). Active inference on discrete state spaces: A step-by-step guide. Journal of Mathematical Psychology, 99, 102440.
[12] Friston, K. J., et al. (2011). Action understanding and active inference. Biological Cybernetics, 104(1), 137-160.
[13] Björnson, E., et al. (2013). Optimal multi-user transmit beamforming: A difficult problem with a simple solution structure. IEEE Signal Processing Magazine, 30(3), 151-155.
[14] Gesbert, D., et al. (2010). Multi-cell MIMO cooperative networks: A new look at interference. IEEE Journal on Selected Areas in Communications, 28(9), 1380-1408.
[15] Irmer, R., et al. (2011). Coordinated multipoint: Concepts, performance, and field trial results. IEEE Communications Magazine, 49(2), 102-111.
[16] Zappone, A., & Jorswieck, E. (2015). Energy efficiency in wireless networks via fractional programming theory. Foundations and Trends in Communications and Information Theory, 11(3-4), 185-396.
[17] Millidge, B. (2020). Deep active inference as variational policy gradients. Journal of Mathematical Psychology, 99, 102448.
[18] Liu, F., et al. (2022). Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. IEEE Journal on Selected Areas in Communications, 40(6), 1728-1767.