Beamforming with Free Energy Principle under Hierarchical Codebook

Tatsuya Otoshi
Graduate School of Economics
Osaka University
Osaka, Japan
t-otoshi@econ.osaka-u.ac.jp

Masayuki Murata
Graduate School of Information Science and Technology
Osaka University
Osaka, Japan
murata@ist.osaka-u.ac.jp

***Abstract*—Beamforming plays a crucial role in enhancing the performance of wireless communication systems. However, achieving optimal beamforming entails a trade-off between exploration and exploitation, where the system needs to balance the exploration of different beam directions with the exploitation of the best available beam. Motivated by the exploration-exploitation trade-off, we propose the FEP method, which leverages hierarchical modeling and adaptive beam switching to optimize this trade-off. Through simulations in a dynamic environment, we evaluate the performance of the FEP method in terms of expected free energy minimization and signal-to-interference-plus-noise ratio (SINR) maximization. The results demonstrate that the FEP method effectively maintains high SINR levels through adaptive beam direction selection, reflecting efficient exploitation. Comparative analysis with other beamforming methods further highlights the superior performance of the FEP method in terms of average signal quality and stability.**

***Index Terms*—Beamforming, Beyond 5G, Active Inference, Free Energy Principle, Hierarchical Codebook**

I. Introduction

Massive MIMO technology has shown great potential for enhancing coverage and improving spatial utilization in wireless communication systems [1]. However, beamforming, a key technique in Massive MIMO, faces challenges in accurate channel state estimation. This challenge arises due to the need for feedback from the terminals, resulting in round-trip delays and reduced throughput due to observation-based radio transmissions. Particularly in Massive MIMO systems with a large number of antennas, the dimensionality of channel state estimation increases, leading to increased overhead for estimation.

In the context of 5G, beamforming techniques utilizing codebooks have been standardized to alleviate the aforementioned challenges. By predefining a fixed number of candidate beams, the overhead associated with beam search can be reduced. However, as the number of antennas increases, the number of candidate beams also escalates, giving rise to concerns regarding search overhead. To tackle this issue, hierarchical codebooks have been proposed [2], [3], which enable efficient beam search by employing broader-angle beams in the upper layers and narrower-angle beams in the lower layers. This approach gradually refines the beam selection from coarse granularity to fine granularity.

Nonetheless, existing methods for hierarchical beamforming assume a static environment and depend on sequential search of the hierarchical codebooks from top to bottom in the spatial domain. In practice, it is necessary to simultaneously explore both higher and lower hierarchical levels, adapting to changing conditions. These challenges arise due to a system model that separates the stages of exploratory observation-based estimation and utilization-based data transfer.

Recently, the Free Energy Principle (FEP) has garnered significant attention in the realm of neuroscience due to its potential applications in various domains [4], [5]. FEP introduces a unifying framework that tackles estimation and decision-making by framing them as processes aimed at minimizing free energy. This notion of free energy encompasses both epistemic and pragmatic values, fostering a harmonious equilibrium between exploration and exploitation during the estimation and decision-making phases. When integrated into the context of beamforming, FEP stands out for its capability to dynamically and intelligently adjust the equilibrium between exploration and exploitation in response to evolving environmental conditions. Also, FEP has the capability to iteratively update the internal model itself. This aspect stands in contrast to conventional methods (e.g. neural network [6]) that require time-consuming retraining. FEP, on the other hand, enables swift adaptation to changing circumstances while simultaneously enhancing the accuracy of the internal model. As a result, FEP not only aligns with the principles of adaptive beamforming but also offers a novel approach to optimizing the performance of beamforming techniques amidst changing contexts.

In this paper, we present a novel approach that utilizes Free Energy Principle (FEP) agents for beam selection at each level of the hierarchical codebook. These FEP agents make use of feedback information from the terminals to select beams at their respective levels. The FEP agents in the upper and lower levels simultaneously make independent decisions, but they exchange state information to ensure coherent beam selection across levels. To account for the disparities in spatial and temporal scales between the upper and lower levels, we set longer intervals for updating the state in the higher levels. This allows for the reflection of the differences in spatial and temporal scales between the upper and lower levels.

Through comprehensive evaluations, we demonstrate the effectiveness of our proposed method. We show that our approach consistently maintains a higher SINR(Signal-to-Inference-pluse-Noise Ratio) compared to conventional beam training techniques and policy-based approaches in multi-armed bandit problems.

In Section II, we review the existing beamforming studies and FEP as related studies. Section III describes the assumed beamforming system model. In Section IV, we propose a beamforming method based on FEP. In Section V, we present the evaluation results of the proposed method by simulations. Finally, in Section VI, we summarize and discuss future issues.

II. Related Work

A. Beamforming with Hierarchical Codebook

Previous literature has focused on the design of hierarchical codebooks. In many cases, hierarchical codebooks are constructed by dividing the angular domain from coarse to fine granularity, primarily targeting the direction of beams [7], [8].

The mainstream approach in hierarchical codebooks is to divide the beam directions hierarchically and store beam vectors corresponding to each direction in the codebook, as shown in [7], [8]. The upper levels contain candidate angles with coarse granularity, while the lower levels have candidate angles with fine granularity. In many cases, at each level, two alternative angles are selected based on the angle determined in the upper level as the center. However, when performing the search from the topmost to the bottommost level, the minimum number of beam vectors to be explored is achieved through a three-partition search, which has led to the proposal of codebooks for ternary tree search [9].

When the receiving terminal is far enough, the direction of beam emission from each antenna aligns with the beam direction. However, when the receiving terminal is nearby, adjustments in the emission direction for each antenna are necessary based on the distance. Therefore, methods that consider both the beam direction and the distance and construct codebooks by hierarchically partitioning the two variables have also been proposed [2], [3]. The upper levels have candidate angles and distances with coarse granularity, while the lower levels have candidate angles and distances with fine granularity. In this case, at each level, a four-choice selection is performed by dividing the range of angles and distances into two parts for each variable.

In beam determination, as feedback is obtained for each beam, increasing the number of beams results in a larger overhead for feedback. When using hierarchical codebooks, the overhead is reduced by limiting the number of beams to be explored.

Beamforming using hierarchical codebooks often involves beam training, where beams are determined. Beam training is performed by exploring and determining the beam with the highest SINR before communication. The search is conducted from coarse beams at the upper levels to fine beams at the lower levels, following the hierarchical structure of the codebook.

Thus, conventional beamforming using hierarchical codebooks assumes offline beam exploration. On the other hand, active inference allows for online beamforming, which performs communication and exploration simultaneously.

Although not using hierarchical codebooks, a method that performs online beam training known as the UCB (Upper Confidence Bound) algorithm is proposed in the literature [10]. This method narrows down the candidate beams using terminal location information and performs online beam exploration as a multi-armed bandit problem.

The beam exploration selects the top B_{tr} beams to explore per round, where the beams with the highest UCB values are explored, as defined by:

$$\mathrm{UCB}_i = \frac{X_i}{T_i} + \sqrt{\frac{2\log(n)}{T_i}} \tag{1}$$

Here, n represents the round number, X_i is the cumulative reward for the i-th beam, and T_i is the number of times the i-th beam has been explored. The reward is 1 if the beam has the highest received signal strength and 0 otherwise.

The first term on the right-hand side represents the selection criterion for beams with higher average rewards, while the second term ensures the selection of beams with fewer explorations relative to the number of rounds. This allows for beam exploration considering the trade-off between exploration and exploitation. However, this method does not consider the hierarchical structure of the codebook, and the location information of the terminal is obtained from the satellite, which is limited to the case where the satellite is in the line of sight.

Our FEP-based beamforming also allows for the use of these conventional codebooks, and in this paper, we employ a standard binary partition codebook based on angles. However, conventional hierarchical codebooks assume a unidirectional search from the topmost to the bottommost level and may not be optimal when performing simultaneous searches between upper and lower levels as required by the attention mechanism of the FEP.

B. Free Energy Principle

FEP is used to explain a wide range of phenomena related to neural activity and behavior in the brain and provide a common underlying principle for biological systems [4], [5]. According to FEP, biological systems are programmed to minimize free energy. In order for biological systems to achieve goals such as cognition, perception, and behavior, they must process sensory information appropriately and minimize free energy.

In FEP, the brain is assumed to compute an approximate distribution $q(x)$ of the posterior distribution $p(x|s)$ of the state x of the environment under the observed information s. The free energy is defined by

$$F(q,s) = \sum_x q(x)E(x,s) - (-\sum_x q(x)\log q(x)) \tag{2}$$

$$= D_{KL}(q(x)||p(x|s)) - \log p(s) \tag{3}$$

where $E(x,s) = -\log p(x,s)$ is the logarithm of the simultaneous distribution of x and s, and D_{KL} is the KL divergence. $-\log p(s)$ is called the Shannon surprise. Equation (2) corresponds to the Helmholtz free energy $F = U - TS$ in statistical mechanics if the temperature T is neglected ($T = 1$).

III. System Model

This section describes the model of beamforming assumed in this paper. We also describe the time scales of the observation and control loops assumed.

A. Beamforming

The base station uses M antennas to transmit radio waves, which are transformed by spatial characteristics to reach N antennas at the receiving terminal (UE: User Equipment). Note that $N = 1$ in the cases of [6], [11]. The spatial characteristics are represented by the matrix $H(t)$, where the element H_{ij} represents the channel coefficient for the transmission from antenna i to antenna j.

The transmitter estimates $H(t)$ and performs beamforming by transforming the phase and amplitude of the transmitted radio wave for each antenna according to $H(t)$, thereby improving the efficiency of radio propagation. Beamforming is specified by the beam vector $\boldsymbol{w}(t)$ and the transmit power $P(t)$, and the following relationship is established between the transmit signal $x(t)$ and the receive signal $y(t)$:

$$\boldsymbol{y}(t) = \sqrt{P(t)} H(t) \boldsymbol{w}(t) \circ \boldsymbol{x}(t), \tag{4}$$

where $\circ$ denotes the element-wise (Hadamard) product.

In this case, the signal-to-interference-plus-noise ratio (SINR) is expressed as follows:

$$\gamma(t) = \frac{P(t)|H(t)\boldsymbol{w}(t)|^2}{I(t) + \sigma^2(t)}, \tag{5}$$

where $I(t)$ is the strength of the interfering radio wave and $\sigma^2(t)$ is the strength of the noise.

The transmission rate is also estimated using SINR as follows:

$$C(t) = \log(1 + \gamma(t)), \tag{6}$$

where $C(t)$ is the transmission rate.

B. Hierarchical Codebook

In beamforming, it is necessary to estimate the channel state $H(t)$ using feedback information and optimize the beamforming vector $w(t)$. However, estimating the channel state incurs significant overhead and is not practical. Therefore, an approach is used where beamforming vectors are selected from a pre-defined set of candidates called a codebook. However, in scenarios with a large number of antennas, such as Massive MIMO, the codebook itself becomes extensive, requiring the introduction of hierarchy to improve the efficiency of the search process.

In a hierarchical codebook, there are K_l beamforming vectors $w_k^{(l)}$ for each layer l. The value of K_l increases for lower layers, and in the case of constructing a codebook with binary splits, $K_l = 2^l$.

When $w_{a^l}^{(l)}$ is selected at layer l, at the next layer $(l+1)$, beamforming vectors similar to $w_{a^l}^{(l)}$ are further divided and selected. In the case of binary splits, the selection is made between $w_{2a^l-1}^{(l+1)}$ and $w_{2a^l}^{(l+1)}$.

The hierarchical codebook used in the paper is based on the one described in the literature [7]. The n-th beamforming vector at layer l is as follows:

$$\boldsymbol{w}_n^{(l)} = (\boldsymbol{a}(2^l, -1 + \tfrac{2n-1}{2^l}), \mathbf{0}_{N-2^l}) \tag{7}$$

$$\boldsymbol{a}(N, \Omega) = \tfrac{1}{\sqrt{N}}(\exp(i\pi 0\Omega), \cdots, \exp(i\pi(N-1)\Omega)) \tag{8}$$

At layer l, the beam is formed using 2^l antennas, and as l increases, it allows for finer beam configurations.

C. Observable Information on Base Station

Estimating the channel coefficients is essential in beamforming, but since the coefficients themselves cannot be directly observed, they must be estimated from the signals at the receiving UE, which are observable.

There are two main methods for measuring signal information at the receiving UE: using the synchronization signal (SS) and using Channel State Information (CSI).

In the case of SS, the signal strength and SINR are measured using the synchronization signal received in the SS/PBCH block, which is a set of SS and PBCH (Physical Broadcast Channel), as a reference signal. SS/PBCH is transmitted periodically (e.g., every two frames) for initial access, so there is no need to transmit a new reference signal.

When using CSI, a reference signal is sent for observation, and the received signal is used to measure signal strength and SINR.

In either case, information from the receiving side is fed back to the transmitting side, enabling the transmitter to estimate the downstream channel coefficients.

In limited situations, such as when the same frequency is used in time division, it is also possible to use the uplink reference signal (SRS: Sounding Reference Signal) for measurement due to the symmetry of the uplink and downlink channel coefficients. In this case, the base station observes the signal directly, so there is no need for round-trip feedback.

We assume that the signal strength and SINR are observed using these methods in the proposed method.

D. Timescales

Beamforming itself can be switched at a high speed. In fact, for initial access, beam sweeping switches between two different beams within one slot. Specifically, during beam sweeping, the SS/PBCH block consists of four symbols transmitted by the base station, with a two-symbol gap between them [12].

Observation information, on the other hand, is acquired on a slightly slower timescale than beamforming switching. Feedback of the receiver's SINR to the base station requires at least a round-trip time between the UE and the base station to reply to the transmitted signal from the base station. Therefore, observations are collected at intervals of multiple symbols or multiple slots. Specifically, in the case of SS, channel

information is measured using SS/PBCH blocks [13], which are transmitted periodically every 20 ms, so observations are obtained every 20 ms. When using CSI, channel information can be obtained in a shorter time, and according to the literature [14], the feedback delay for CSI is a minimum of 1 ms. Since beamforming can be switched within a symbol, it is assumed that control switching can be done fast enough for the timescale of observation.

IV. Beamforming with Active Inference

We propose beamforming using the FEP to dynamically perform beamforming while mitigating search overhead. Our method leverages limited feedback information to estimate the environmental state and determine the beamforming vector accordingly based on FEP. It automatically balances the beam selection for exploration and optimization for data transmission, assuming feedback is obtained only for the selected beam. Additionally, to facilitate hierarchical decision-making with a hierarchical codebook, we deploy agents at each layer to make decisions and enable state exchange among them.

A. Hierarchical Decision with FEP

Hierarchical decision-making based on the Free Energy Principle (FEP) is employed in our proposed method. In the context of FEP, hierarchical active inference through predictive coding has been advocated [15]. Predictive coding suggests that the brain predicts input and acts to minimize the prediction error, which is the discrepancy between the predicted input and the actual input. In this framework, each layer's internal state serves as input for the higher layer, and the lower layer predicts the lower-level states, forming a multi-level hierarchy of active inference. It is assumed that higher layers operate on longer timescales than lower layers.

Considering layer $l = 1$ as the topmost layer and $l = L$ as the bottommost layer, let $o_t^{(l)}$ represent the input at layer l, and $s_t^{(l)}$ denote the internal state. For information flow from lower to higher layers, the lower layer's estimated internal state distribution is provided as input:

$$o_t^{(l)} = P(s_t^{(l+1)}|o_{0:t}^{l+1}) \tag{9}$$

Conversely, for information flow from higher to lower layers, the lower layer's internal state prediction serves as the prior distribution:

$$P(s_t^{(l+1)}) \quad = \quad \hat{o}_t^{(l)} \tag{10}$$

where $\hat{o}_t^{(l)} = E_{P(o_t^{(l)})|s_t^{(l)})}[P(o_t^{(l)})]$ represents the expected value of observations under the posterior distribution of the higher layer.

At each layer, the policy $\pi^{(l)} = (a_1^{(l)}, \cdots, a_T^{(l)})$, representing a sequence of selected actions, is determined to minimize the expected free energy, guided by the Free Energy Principle:

$$\begin{aligned} G(o_{1:T}^{(l)}, s_{1:T}^{(l)}, \pi^{(l)}) \quad = \quad & E_Q[\log Q(s_{1:T}^{(l)}, \pi^{(l)}) \\ & - \log \tilde{P}(o_{1:T}^{(l)}, s_{1:T}^{(l)}, \pi^{(l)})] \end{aligned} \tag{11}$$

Here, Q represents the approximate posterior distribution, and $\tilde{P}$ represents the target distribution.

In the work [15], the timescales of the lower and higher layers are predetermined. After a fixed number of iterations (20 times) of estimation and action at the lower layer, one iteration of estimation and action at the higher layer occurs. Then, the lower layer repeats the estimation and action based on the new prior obtained from the higher layer. Similarly, in our paper, we assume that the higher layer operates on a slower time cycle T_h compared to the lower layer.

B. Definition of FEP Agents

We design agents corresponding to each hierarchical codebook layer in the proposed method. Specifically, we define the agent's observations, states, actions, and preference distribution.

a) Observations: The base station can observe the transmission rate $\boldsymbol{C}(t)$ to each UE based on feedback from the terminals. Additionally, except for the lowest layer, higher layers can also observe the states estimated by the lower layer.

Therefore, the observations at layer l correspond to $\boldsymbol{o}^{(l)}(t) = (P(s_t^{(l+1)}|o_{0:t}^{(l+1)}), \boldsymbol{C}(t))$. The lowest layer only observes feedback from the terminals, $\boldsymbol{o}^{(L)}(t) = \boldsymbol{C}(t)$.

As feedback, we assume that the receiving side's SINR is measured using CSI or SS, and the transmission rate is estimated from the SINR.

However, the higher the layer, the longer the time interval for obtaining observations. The specific time intervals and the impact of temporal smoothing of observations will be investigated through simulations.

b) States: Based on the observed transmission rates $\boldsymbol{C}(t)$ and their own beamforming information, the base station estimates the channel coefficients $H^u(t)$. Therefore, the system state corresponds to $s^{(l)}(t) = H(t)$.

State estimation is performed to minimize the error between the predicted distribution $P(o^{(l)}(t)|s^{(l)}(t))$ of the observations and the actual observations. At higher layers, the state does not necessarily have the same dimensions as the state at the lower layer and can be an abstract representation. The same applies to the lower layers, where the system state $s^{(l)}(t)$ corresponds indirectly to $H(t)$ through the observed values, and there is no strict requirement for them to match.

The top-down information from higher layers to lower layers is achieved by overwriting the predicted distribution $P(o^{(l)}(t)|s^{(l)}(t))$ with the distribution of the lower layer's state $P(s^{(l+1)}(t))$.

c) Actions: The base station specifies the beam shape using the beamforming vector $\boldsymbol{w}^{(l)}(t)$ and determines the transmission power $P^{(l)}(t)$. Therefore, $a^{(l)}(t) = (\boldsymbol{w}^{(l)}(t), P^{(l)}(t))$ corresponds to the actions. However, at each layer, there may be undetermined choices $a^{(l)}(t) = \emptyset$, in which case the choice of the higher layer is adopted.

d) Preference Distribution: The preference distribution sets the desired states as a prior distribution for control.

In existing beamforming methods, the objective function is often the power efficiency or maximization of the trans-

mission rate. However, when considering traffic demands, it is desirable to set the effective throughput as the objective function. By choosing a Boltzmann distribution as the preference distribution with the objective function $R_b(o_\tau)$, which measures utility in terms of logarithm, we can reflect $R(o_\tau)$ in the preference distribution:

$$\tilde{P}_R(o_\tau) \propto \exp(\beta R(o_\tau)) \tag{12}$$

Here, β is a parameter that determines how much the preference is biased based on the magnitude of the objective function.

C. Agents Behavior

Each agent in the proposed method performs state estimation upon receiving feedback and determines its action policy based on the estimated state. Furthermore, when a certain amount of observations has accumulated, the internal model is updated through model learning [16]. The following describes each operation in detail.

1) Inference: When the observed value o_τ regarding the radio wave strength is obtained, the state S_τ of the estimation regarding the channel is updated. Here, the state is estimated by minimizing the variational free energy according to the FEP. The state is estimated as a distribution, and the distribution $Q(s_\tau|\theta_\tau)$ follows a categorical distribution.

$$\theta_\tau = \arg\min_{\theta}\{D_{KL}[Q(s_\tau|\theta)||P(s_\tau|o_\tau)] - \log P(o_\tau)\} \tag{13}$$

That is, θ_τ is calculated as the minimization of the sum of the true posterior distribution and $Q(s_\tau|\theta_\tau)$ with the Cullback-Leibler information content and the Shannon surprise $-\log P(o_\tau)$.

2) Action: Based on the estimated state, beamforming a_τ is determined as the appropriate action for the situation at that time. Again, according to the FEP, the action with the smallest Expected Free Energy is selected. The policy that determines the series of actions $a_1, \cdots, a_T$ is π, and the probability of the policy is determined as follows.

$$P(\pi) = \mathrm{softmax}(-\sum_{\tau} G_\tau(\pi)) \tag{14}$$

where $\mathrm{softmax}(x) = \frac{e^x}{\sum_{x'} e^{x'}}$ and G_τ at each time is defined by

$$\begin{aligned} G_\tau(\pi) &= E_{Q(o_\tau, s_\tau|\pi)}[\log Q(s_\tau|\pi) \\ &\quad - \log \tilde{P}_C(o_\tau) P(s_\tau|o_\tau, \pi)] \end{aligned} \tag{15}$$

where $\tilde{P}_C(o_\tau)$ denotes the preference for the observed value.

3) Learning: For the internal models $P(o_\tau|s_\tau), P(s_{\tau+1}|s_\tau)$, it is possible to set designed values or learn from data. Therefore, it is also possible to operate by giving a base model and then updating the internal model itself according to the situation.

Update $P(o_\tau|s_\tau)$ as Dirichlet distribution $Q(A)$ with parameter A and $P(s_{\tau+1}|s_\tau)$ as Dirichlet distribution $Q(B)$ with parameter B. Again, A, B is obtained by minimizing the variational free energy as in the state estimation.

$$\begin{aligned} A_T &= \arg\min_{A} E_{Q(s_{1:T}, A, \pi)} [\log Q(s_{1:T}, A, \pi) \\ &\quad - \log P(o_{1:T}, s_{1:T}, A, \pi)] \end{aligned} \tag{16}$$

$$\begin{aligned} B_T &= \arg\min_{B} E_{Q(s_{1:T}, B, \pi)} [\log Q(s_{1:T}, B, \pi) \\ &\quad - \log P(o_{1:T}, s_{1:T}, B, \pi)] \end{aligned} \tag{17}$$

V. Evaluation

To evaluate the effectiveness of the proposed method, we conducted simulations. The evaluation aimed to assess the performance of the method in dynamic scenarios where UE is moving, requiring continuous updates of beamforming. We compared the proposed method against existing approaches such as online beam training and periodic beam training.

A. Setting

In the evaluation, we considered a scenario where a single UE with one antenna receives signals from a base station equipped with $N = 4$ antennas. To begin with, we verified the correct functioning of the expected free energy minimization in a static environment by keeping the channel coefficients fixed without temporal variations.

The base station observes the UE's signal-to-interference-plus-noise ratio (SINR) as feedback and, upon receiving the feedback, updates the internal models and states of each layer and determines the power and beamforming vector for the next transmission. The beamforming vector is selected from a predefined hierarchical codebook defined by Eq. (7).

The UE performs circular motion around the base station, moving once per unit time. The distance between the base station and the UE is set to 200m, and the noise signal strength at the UE is -114dBm.

In the simulation setting, we consider the time-varying channel coefficients modeled by L-path Rayleigh fading, following the approach in [6]. The channel coefficient $h_{ij}(t)$ between base station i and user equipment j at time t is given by:

$$h_{ij}(t) = \sqrt{\frac{\beta_{i,j}}{L}} \sum_{l=1}^{L} g_{i,j}(t,l) a^{\dagger}_{i,j}(\theta) \tag{18}$$

$$g_{i,j}(t,l) = \rho g_{i,j}(t-1,l) + \sqrt{1-\rho^2} e_{i,j}(t) \tag{19}$$

where $\beta_{i,j}$ represents the average channel power gain between the base station and user equipment, $a^{\dagger}_{i,j}(\theta_l)$ is the steering vector representing the direction of the l-th path, $e_{i,j}(t)$ represents a complex Gaussian-distributed noise term, and ρ controls the temporal continuity of the Rayleigh fading process. θ_l, are assumed to follow a uniform distribution within the range $(\bar{\theta}_l - \vartheta/2, \bar{\theta}_l + \vartheta/2)$, where $\bar{\theta}_l$ represents the average angle and ϑ denotes the angular spread.

B. Methods

In the evaluation, we compared the following beamforming methods in the simulations:

1) FEP-based BF(FEP-based Beamforming): This is the beamforming method based on the FEP described in Section IV. The internal models $P(o_\tau|s_\tau,\pi)$ and $P(s_\tau|s_{\tau-1},\pi)$ are learned as expected free energy minimization. The base station has a two-layer codebook, with the upper layer operating at a timescale $T_h = 10$ times slower than the lower layer.

2) BT(Beam Training): For comparison, we implemented beamforming with periodic beam training. Specifically, we set the training interval to $T_{train} = 100$, and after performing T_{train} communications, beam training is conducted. During beam training, communication is not interrupted, but the beams used for communication are selected sequentially through hierarchical exploration. Once beam training is completed, the final determined beam is used for communication until the next beam training.

The beams during beam training are selected as follows:

1) Start from layer $l = 0$.
2) Select a beam $a \in A^{(l)}(a^{(l-1)})$ at layer l.
3) Measure the SINR under the selected beam.
4) If beams at layer l have not been fully explored, go back to step 2.
5) Select the $a^{(l)}$ with the highest SINR.
6) If l is not equal to L, update l to $l+1$ and go back to step 2.
7) $a^{(L)}$ is the final beam.

Here, $A^{(l)}(a^{(l-1)})$ represents the set of refined beams available at layer l when a beam $a^{(l-1)}$ is selected at the upper layer. $A^{(0)}$ at the top layer represents the set of all coarsest beams.

3) fixBT: This method is the same as BT, but with $T_{train} = \inf$. After the initial beam training, the selected beam is continuously used for communication.

4) UCB(Upper Confidence Bound): This method uses the UCB algorithm to select a set of beams that maximizes the UCB defined in Eq. (1). The set size is set to 2, and the reward is defined as SINR.

5) RAUCB(Risk-Aware UCB): In the UCB method, this approach stochastically rejects the selection based on UCB and instead chooses beams with higher average rewards. The rejection probability $\tilde{Z}_n$ follows a Beta distribution as given below:

$$\tilde{Z}_n \sim Beta(1 + Z_{tot}[i], 1 + T_i - Z_{tpt}[i]) \tag{20}$$

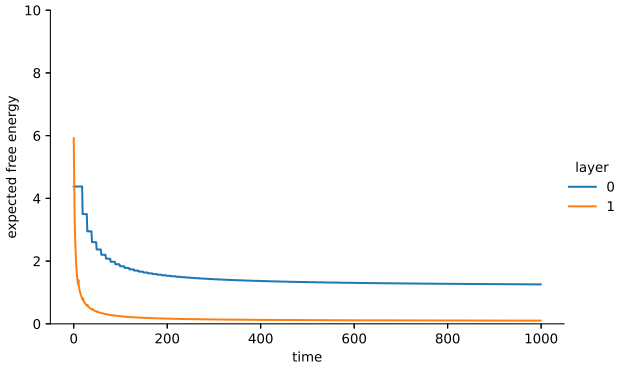
$$Z_{tot}[i] = \sum_{t=1}^{n} z_t[i] \tag{21}$$

$$z_t[i] = \begin{cases} 1 & \text{if } \frac{max_{k\in S}\gamma_{k,t}}{\gamma_{i,t}} > \Gamma \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

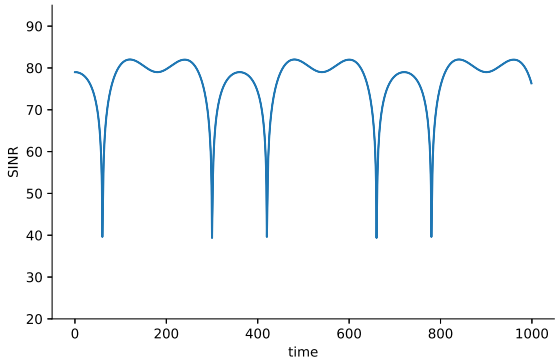
Here, $\gamma_{k,t}$ represents the received signal strength of beam k at time t, S denotes the set of selected beams, and Γ is the parameter controlling the rejection for UCB, set to $\Gamma = 5$dB.

C. Results

1) Behavior of FEP in No Fading Environment: Figure 1 shows the results of the FEP method in terms of the temporal changes in expected free energy and SINR for both the upper and lower layers in no fading environment. The time unit in the figure corresponds to one loop of observation and control, which is the same as the number of received SINR feedbacks.



(a) Expected Free Energy



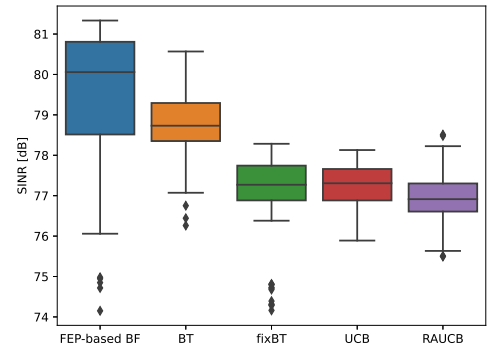
(b) SINR

Fig. 1. Timeseries of Expected Free Energy and SINR in no fading environment

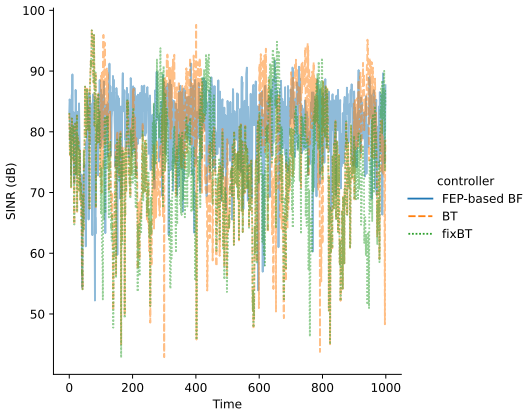
From the figure, it can be observed that both the expected free energy of the upper and lower layers monotonically decrease, indicating that the minimization of free energy is effectively functioning. Additionally, the SINR varies with changes in the angle between the terminal and the base station. However, the SINR is maintained at a high level, indicating that the FEP method is effectively exploiting the best beam direction.

2) Comparison: Figure 2(a) presents box plots depicting the average time-varying SINR achieved by each method during beamforming. To account for the influence of channel coefficients and stochastic selection, we performed beamforming 100 times for each method using different seeds. Moreover, Figure 2(b) shows examples of SINR time series. However, UCB and RAUCB exhibit strong exploratory behavior and large SINR variations; hence, Figure 2(b) displays only the results of other methods excluding UCB and RAUCB.

From the figures, it is evident that the proposed method achieves the highest median SINR among all techniques. Subsequently, the SINR for BT, which performs periodic



(a) Average SINR



(b) Timeseries of SINR

Fig. 2. Average and Timeseries of SINR for Each Method

beam training, is comparatively high, while fixBT, UCB, and RAUCB result in lower SINR values. The fixed beam selection in fixBT leads to reduced SINR as it is unable to adapt to changing environmental conditions. In the case of UCB and RAUCB, frequent exploratory selections and susceptibility to noise-induced errors contribute to lower SINR values.

Figure 2(b) further illustrates that the proposed method maintains consistently high SINR levels throughout the entire duration. Conversely, BT and fixBT experience temporary SINR decreases due to environmental fluctuations, which likely lead to the observed reduction in average SINR. The continuous high performance of the proposed method indicates its robustness and effectiveness in handling dynamic environments.

VI. Conclusion

In this paper, we proposed a beamforming method based on the FEP for multi-antenna communication systems. The FEP approach leverages hierarchical modeling and adaptive beam switching to optimize the trade-off between exploration and exploitation. Through simulation evaluations, we compared the FEP method with other beamforming methods, including regular beam training, fixed beam selection, UCB, and RAUCB. The results showed that the FEP method achieved the highest SINR and the minimum variation of SINR, indicating superior performance in terms of both average signal quality and stability.

Further research can focus on extending the FEP approach to more complex scenarios and investigating its performance under different system parameters and constraints.

References

[1] R. Chataut and R. Akl, "Massive mimo systems for 5g and beyond networks—overview, recent trends, challenges, and future research direction," *Sensors*, vol. 20, no. 10, p. 2753, 2020.

[2] X. Shi, J. Wang, Z. Sun, and J. Song, "Hierarchical codebook-based beam training for extremely large-scale massive mimo," *arXiv preprint arXiv:2210.03345*, 2022.

[3] J. Chen, F. Gao, M. Jian, and W. Yuan, "Hierarchical codebook design for near-field mmwave mimo communications systems," *arXiv preprint arXiv:2212.07847*, 2022.

[4] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of physiology-Paris*, vol. 100, no. 1-3, pp. 70–87, 2006.

[5] K. Friston, "The free-energy principle: a unified brain theory?" *Nature reviews neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[6] K. Yu, G. Wu, S. Li, and G. Y. Li, "Local observations-based energy-efficient multi-cell beamforming via multi-agent reinforcement learning," *Journal of Communications and Information Networks*, vol. 7, no. 2, pp. 170–180, 2022.

[7] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3380–3392, 2016.

[8] H. Yu, P. Guan, W. Qu, and Y. Zhao, "An improved beam training scheme under hierarchical codebook," *IEEE Access*, vol. 8, pp. 53 627–53 635, 2020.

[9] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive mimo with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1376–1393, 2021.

[10] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Online learning for position-aided millimeter wave beam training," *IEEE Access*, vol. 7, pp. 30 507–30 526, 2019.

[11] S. Sharifi, S. Shahbazpanahi, and M. Dong, "A POMDP-based antenna selection for massive mimo communication," *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 2025–2041, 2021.

[12] A. Chakrapani, "On the design details of ss/pbch, signal generation and prach in 5g-nr," *IEEE Access*, vol. 8, pp. 136 617–136 637, 2020.

[13] K. Shamaei and Z. M. Kassas, "Receiver design and time of arrival estimation for opportunistic localization with 5g signals," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4716–4731, 2021.

[14] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive mimo in 5g," *IEEE communications magazine*, vol. 52, no. 5, pp. 44–51, 2014.

[15] R. C. Heins, M. B. Mirza, T. Parr, K. Friston, I. Kagan, and A. Pooresmaeili, "Deep active inference and scene construction," *Frontiers in Artificial Intelligence*, vol. 3, p. 509354, 2020.

[16] C. Heins, B. Millidge, D. Demekas, B. Klein, K. Friston, I. Couzin, and A. Tschantz, "pymdp: A python library for active inference in discrete state spaces," *arXiv preprint arXiv:2201.03904*, 2022.