Principled Learning-to-Communicate with Quasi-Classical Information Structures

Xiangyu Liu[†] Haoyi You[†] Kaiqing Zhang[†]

***Abstract*— Learning-to-communicate (LTC) in partially observable environments has emerged and received increasing attention in deep multi-agent reinforcement learning, where the control and communication strategies are *jointly* learned. On the other hand, the impact of communication has been extensively studied in control theory. In this paper, we seek to formalize and better understand LTC by bridging these two lines of work, through the lens of *information structures* (ISs). To this end, we formalize LTC in decentralized partially observable Markov decision processes (Dec-POMDPs) under the common-information-based (CIB) framework, and classify LTCs based on the ISs before additional information sharing. We first show that non-classical LTCs are computationally intractable in general, and thus focus on quasi-classical (QC) LTCs. We then propose a series of necessary conditions for QC LTCs, violating which can cause computational hardness in general. Further, we develop provable planning and learning algorithms for QC LTCs, and show that examples of QC LTCs satisfying the above conditions can be solved without computationally intractable oracles. Along the way, we also establish some relationship between (strictly) QC IS and the strategy-independence condition in the CIB framework (SI-CIB), as well as solving general Dec-POMDPs without computationally intractable oracles beyond those with SI-CIB, which may be of independent interest.**

I. Introduction

Learning-to-communicate (LTC) has emerged and gained traction in the area of (deep) multi-agent reinforcement learning (MARL) [1], [2], [3]. Unlike classical MARL, which aims to learn a *control* strategy that minimizes the expected accumulated costs, LTC seeks to *jointly* minimize over both the *control* and the *communication* strategies of all the agents, as a way to mitigate the challenges due to the agents' *partial observability* of the environment. Despite the promising empirical successes, theoretical understandings of LTC remain largely underexplored.

On the other hand, in control theory, a rich literature has investigated the role of *communication* in decentralized/networked control [4], [5], [6], [7], [8], inspiring us to examine LTCs from such a principled and rigorous perspective. Most of these studies, however, focused on linear systems, and did not explore the computational or sample complexity guarantees when the system knowledge is not (fully) known. A few recent studies [9], [10] started to explore the settings with general discrete spaces, with special communication protocols and state transition dynamics.

More broadly, (the design of) communication strategy dictates the *information structure* (IS) of the control system, which characterizes *who knows what and when* [11]. IS and its impact on the *optimization tractability*, especially for linear systems, have been extensively studied in decentralized control, see [12], [13] for comprehensive overviews. In this work, we seek a more principled understanding of LTCs through the lens of information structures, with a focus on the computational and sample complexities of the problem.

Specifically, we formalize LTCs in the general framework of decentralized partially observable Markov decision processes (Dec-POMDPs) [14], as in the empirical works [1], [2], [3]. To achieve finite-time and sample guarantees, we resort to the recent development in [15] on partially observable MARL, based on the common-information-based (CIB) framework [16], [17] from decentralized control to model the communication and information sharing among agents. We detail our contributions as follows.

Contributions. (i) We formalize learning-to-communicate in Dec-POMDPs under the common-information-based framework [16], [17], [15], allowing the sharing of *historical* information, and the modeling of communication costs; (ii) We classify LTCs through the lens of *information structure*, according to the ISs before additional information sharing. We then show that LTCs with *non-classical* [12] baseline IS is computationally intractable. (iii) Given the hardness, we thus focus on *quasi-classical* (QC) LTCs, and propose a series of conditions under which LTCs preserve the QC IS after sharing, while violating which can cause computational hardness in general. (iv) We propose both planning and learning algorithms for QC LTCs, by reformulating them as Dec-POMDPs with *strategy-independent (SI) common-information-based beliefs* (SI-CIB) [17], [15], a condition shown to be critical for efficient computation and learning [15]. (v) Quasi-polynomial time and sample complexities of the algorithms are established for QC LTC examples that satisfy the conditions in (iii). Along the way, we also establish some relationship between *(strictly) quasi-classical* ((s)QC) ISs and the SI-CIB condition in the framework of [17] under certain assumptions, as well as solving general Dec-POMDPs without computationally intractable oracles beyond those with SI-CIBs, and thus advancing the results in [15]. These results may be of independent interest besides studying LTCs. We conclude with some experimental results.

A. Related Work

Learning-to-communicate in deep MARL. Learning-to-communicate has received increasing attention in the deep

[†]The authors are ordered alphabetically, and are affiliated with the University of Maryland, College Park, MD, USA, 20742. Emails: {xyliu999, yuriiyou, kaiqing}@umd.edu. This work was supported by the Army Research Office (ARO) grant W911NF-24-1-0085 and the NSF CAREER Award 2443704.

MARL literature [1], [2], [3]. However, these algorithms optimize the communication strategies *end-to-end*, without any theoretical analyses or guarantees.

Communication-control joint optimization. The joint design of control and communication strategies has been studied in the control literature [7], [6], [8], [9], [10]. However, even with model knowledge, the computational complexity (and associated necessary conditions) of solving these models remains elusive, let alone the sample complexity when it comes to learning. Moreover, these models mostly have more special structures, e.g., with linear systems [6], [7], [8], or allowing to share only instantaneous observations [9], [10].

Information sharing and information structures. Information structure has been extensively studied to characterize *who knows what and when* in decentralized control [12], [13]. Our paper aims to formally understand LTC through the lens of information structures. The common-information-based approaches to formalize *information sharing* in [16], [17] serve as the basis of our work. In comparison, these results focused on the *structural results*, without concrete computational (and sample) complexity analysis.

Partially observable MARL theory. Planning and learning in partially observable MARL are known to be hard [18], [19], [20], [14]. Recently, [21], [22] developed polynomial-sample complexity algorithms for partially observable stochastic games, but with computationally intractable oracles; [15] developed quasi-polynomial-time and sample algorithms for such models, leveraging information sharing. In contrast, our paper focuses on *optimizing/learning to share*, together with control strategy optimization/learning.

II. Preliminaries

Notation. We use $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ to denote the sets of all the natural, rational, and real numbers, respectively. For an integer $m > 0$, we denote $[m] := \{1, 2, \cdots, m\}$. For a finite set $\mathcal{X}$, we use $|\mathcal{X}|$ to denote the cardinality of $\mathcal{X}$, and use $\Delta(\mathcal{X})$ to denote the probability simplex over $\mathcal{X}$. For a random variable x, we use $\sigma(x)$ to denote the sigma-algebra generated by x. For σ-algebras $\mathscr{F}_1$ on the space $\mathcal{X}_1$ and $\mathscr{F}_2$ on the space $\mathcal{X}_2$, we denote by $\mathscr{F}_1 \otimes \mathscr{F}_2$ the product σ-algebra on the space $\mathcal{X}_1 \times \mathcal{X}_2$. We use $\mathbb{1}[]$ to denote the indicator function. Unless otherwise noted, the set $\{\}$ considered is ordered, such that elements in the set are indexed.

A. Learning-to-Communicate (with Communication Cost)

For $n > 1$ agents, a *Learning-to-Communicate* problem can be depicted by a tuple $\mathcal{L} = \langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}$, $\{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{M}_{i,h}\}_{i\in[n],h\in[H]}, \mathbb{T}, \mathbb{O}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}$, $\{\mathcal{K}_h\}_{h\in[H]}\rangle$, where H denotes the length of each episode, and other components are introduced as follows.

a) Decision-making components: We use $\mathcal{S}$ to denote the state space, and $\mathcal{A}_{i,h}$ to denote the *control action* space of agent i at timestep $h \in [H]$. We denote by $s_h \in \mathcal{S}$ the state and by $a_{i,h}$ the control action of agent i at timestep h. We use $a_h := (a_{1,h}, \cdots, a_{n,h}) \in \mathcal{A}_h := \prod_{i\in[n]} \mathcal{A}_{i,h}$ to denote the joint control action for all the n agents at timestep h. We denote by $\mathbb{T} = \{\mathbb{T}_h\}_{h\in[H]}$ the collection of state transition kernels, where $s_{h+1} \sim \mathbb{T}_h(\cdot \mid s_h, a_h) \in \Delta(\mathcal{S})$ at timestep h. We use $\mu_1 \in \Delta(\mathcal{S})$ to denote the initial state distribution. We denote by $\mathcal{O}_{i,h}$ the observation space and by $o_{i,h} \in \mathcal{O}_{i,h}$ the observation of agent i at timestep h. We use $o_h := (o_{1,h}, o_{2,h}, \cdots, o_{n,h}) \in \mathcal{O}_h := \mathcal{O}_{1,h} \times \mathcal{O}_{2,h} \times \cdots \mathcal{O}_{n,h}$ to denote the joint observation of all the n agents at timestep h. We use $\mathbb{O} = \{\mathbb{O}_h\}_{h\in[H]}$ to denote the collection of emission functions, where $o_h \sim \mathbb{O}_h(\cdot \mid s_h) \in \Delta(\mathcal{O}_h)$ at timestep h and state $s_h \in \mathcal{S}$. Also, we denote by $\mathbb{O}_{i,h}(\cdot \mid s_h)$ the emission for agent i, the marginal distribution of $o_{i,h}$ given $\mathbb{O}_h(\cdot \mid s_h)$ for all $s_h \in \mathcal{S}$. At each timestep h, agents will receive a common reward $r_h = \mathcal{R}_h(s_h, a_h)$, where $\mathcal{R}_h : \mathcal{S} \times \mathcal{A}_h \to [0, 1]$ denotes the reward function shared by the agents.

b) Communication components: In addition to reward-driven decision-making, agents also need to decide and learn (what) to communicate with others. At timestep h, agents share part of their information $z_h \in \mathcal{Z}_h$ with other agents, where $\mathcal{Z}_h$ denotes the collection of all possible shared information at timestep h. Here we consider a general setting where the shared information z_h may contain two parts, the *baseline-sharing* part z_h^b that comes from some existing sharing protocol among agents, and the *additional-sharing* part $z_{i,h}^a$ for each agent i that comes from explicit communication *to be decided/learned*, with the joint additional-sharing information $z_h^a := \cup_{i=1}^n z_{i,h}^a$. This general setting covers those considered in most empirical works on LTC [1], [2], [3], with a void baseline sharing part. We kept the baseline sharing since our focus is on the *finite-time* and *sample* tractability of LTC, for which a certain amount of information sharing is known to be necessary [15]. Note that $z_h = z_h^b \cup z_h^a$ and $z_h^b \cap z_h^a = \emptyset$. The shared information is part of the historical observations and (both *control* and *communication*) actions. We denote by $\mathcal{Z}_h^b$, $\mathcal{Z}_h^a$, and $\mathcal{Z}_{i,h}^a$ the collections of all possible z_h^b, z_h^a, and $z_{i,h}^a$ at timestep h.

At timestep h, the *common information* among all the agents is thus defined as the union of all the *shared information* so far: $c_{h^-} = \cup_{t=1}^{h-1} z_t \cup z_h^b$, and $c_{h^+} = \cup_{t=1}^{h} z_t$, where c_{h^-} and c_{h^+} denote the (accumulated) common information *before* and *after* additional sharing, respectively. Hence, the *private information* of agent i at time h *before* and *after* additional sharing is defined accordingly as $p_{i,h^-} = \{o_{i,1}, a_{i,1}, \cdots, a_{i,h-1}, o_{i,h}\} \backslash c_{h^-}, p_{i,h^+} = \{o_{i,1}, a_{i,1} \cdots, a_{i,h-1}, o_{i,h}\} \backslash c_{h^+}$, respectively. We denote by $p_{h^-} := (p_{1,h^-}, \cdots, p_{n,h^-})$ the joint private information *before* additional sharing, by $p_{h^+} := (p_{1,h^+}, \cdots, p_{n,h^+})$ the joint private information *after* additional sharing, at timestep h. We then denote by $\tau_{i,h^-} = p_{i,h^-} \cup c_{h^-}, \tau_{i,h^+} = p_{i,h^+} \cup c_{h^+}$ the *information available* to agent i at timestep h, before and after additional sharing, respectively, with $\tau_{h^-} = p_{h^-} \cup c_{h^-}, \tau_{h^+} = p_{h^+} \cup c_{h^+}$ denoting the associated joint information. We use $\mathcal{C}_{h^-}, \mathcal{C}_{h^+}, \mathcal{P}_{i,h^-}, \mathcal{P}_{i,h^+}, \mathcal{P}_{h^-}, \mathcal{P}_{h^+}, \mathcal{T}_{i,h^-}, \mathcal{T}_{i,h^+}, \mathcal{T}_{h^-}, \mathcal{T}_{h^+}$ to denote, respectively, the corresponding collections of all possible $c_{h^-}, c_{h^+}, p_{i,h^-}, p_{i,h^+}, p_{h^-}, p_{h^+}, \tau_{i,h^-}, \tau_{i,h^+}, \tau_{h^-}, \tau_{h^+}$.

We use $m_{i,h}$ to denote the *communication action* of agent

i at timestep h, and it will determine what information $z^a_{i,h}$ she will share, through the way specified later. We denote by $\mathcal{M}_{i,h}$ the space of $m_{i,h}$, and by $m_h := (m_{1,h}, \cdots, m_{n,h}) \in \mathcal{M}_h := \mathcal{M}_{1,h} \times \cdots \mathcal{M}_{n,h}$ the joint communication action of all the agents. $\mathcal{K}_h : \mathcal{Z}^a_h \to [0, 1]$ denotes the *communication cost* function, and $\kappa_h = \mathcal{K}_h(z^a_h)$ denotes the incurred communication cost at timestep h, due to additional sharing.

c) System evolution: The system evolves by alternating between the communication and the control steps as follows.

Communication step: At each timestep h, each agent i observes $o_{i,h}$ and may share part of her private information via baseline sharing, receives the baseline sharing of information from others, and forms p_{i,h^-} and c_{h^-}. Then, each agent i chooses her communication action, which determines the additional sharing of information, receives the additional-sharing of information from others, forms p_{i,h^+} and c_{h^+}, and incurs some communication cost κ_h. Formally, the evolution of the information is formalized as follows, which, unless otherwise noted, will be assumed throughout the paper.

Assumption II.1 (*Information evolution*)**.** For each $h \in [H]$,

(a) (Baseline sharing). $z^b_{h+1} = \chi_{h+1}(p_{h^+}, a_h, o_{h+1})$ for some fixed transformation χ_{h+1};

(b) (Additional sharing). For each agent $i \in [n]$, $z^a_{i,h} = \phi_{i,h}(p_{i,h^-}, m_{i,h})$ for some function $\phi_{i,h}$, given communication action $m_{i,h}$, and $m_{i,h} \in z^a_{i,h}$; and the joint sharing $z^a_h := \cup_{i\in[n]} z^a_{i,h}$ is thus generated by $z^a_h = \phi_h(p_{h^-}, m_h)$, for some function ϕ_h;

(c) (Private information before sharing). For each agent $i \in [n]$, $p_{i,(h+1)^-} = \xi_{i,h+1}(p_{i,h^+}, a_{i,h}, o_{i,h+1})$ for some fixed transformation $\xi_{i,h+1}$, and the joint private information thus evolves as $p_{(h+1)^-} = \xi_{h+1}(p_{h^+}, a_h, o_{h+1})$ for some fixed transformation ξ_{h+1};

(d) (Private information after sharing). For each agent $i \in [n]$, $p_{i,h^+} = p_{i,h^-} \backslash z^a_{i,h}$;

(e) (Full memory). For each agent $i \in [n]$, $\tau_{i,h^-} \subseteq \tau_{i,h^+} \subseteq \tau_{i,(h+1)^-}$, and $o_{i,h} \in \tau_{i,h^-}$.

Note that as *fixed transformations* (e.g., χ_h and $\xi_{i,h}$ above), they are not affected by the *realized values* of the random variables, but dictate some *pre-defined* transformation of the input random variables. See [16], [17] and §B in [15] for common examples of baseline sharing that admit such fixed transformations when there is no additional sharing, and examples in §A how they are extended in the LTC setting. It should not be confused with some general *function* (e.g., $\phi_{i,h}$ above), which may depend on the *realized values* of the input random variables. (a) and (c) on baseline sharing follow from those in [17], [15]; (b) and (d) on additional sharing dictate how the communication action affects the sharing based on private information. For example, a common choice of $(\mathcal{M}_{i,h}, \phi_{i,h})$ is that $\mathcal{M}_{i,h} = \{0, 1\}^{|p_{i,h^-}|}$, where 1 or 0 denotes if the associated element in p_{i,h^-} is shared or not; for any $p_{i,h^-} \in \mathcal{P}_{i,h^-}$ and $m_{i,h} \in \mathcal{M}_{i,h}$, $\phi_{i,h}(p_{i,h^-}, m_{i,h})$ consists of the k-th element ($k \in [|p_{i,h^-}|]$) of p_{i,h^-} if and only if the k-th element of $m_{i,h}$ is 1. As $m_{i,h}$ (depicting what to share) will be known given $z^a_{i,h}$ (what has been shared), $m_{i,h}$ is thus also modeled as being shared, i.e., $m_{i,h} \in z^a_{i,h}$. This is also consistent with the models in [9], [10] on control/communication joint optimization. (e) means that the agent has full memory of the information she has in the past and at present. We emphasize that this is closely related, but different from the common notion of *perfect recall* [23] (see also Definition .20), where the agent has to recall all her own *past actions*. Condition (e), in contrast, relaxes the memorization of the actions, but includes the instantaneous observation $o_{i,h}$. This condition is satisfied by the models and examples in [12], [16], [17], [15]. See also §A for more examples that satisfy this assumption.

Decision-making step: After the communication, each agent i chooses her control action $a_{i,h}$, receives a reward r_h, and the joint action a_h drives the state to $s_{h+1} \sim \mathbb{T}_h(\cdot \mid s_h, a_h)$.

d) Strategies and solution concept: At timestep h, each agent i has two strategies, a *control* strategy and a *communication* strategy. We define a control strategy as $g^a_{i,h} : \mathcal{T}_{i,h^+} \to \mathcal{A}_{i,h}$ and a communication strategy as $g^m_{i,h} : \mathcal{T}_{i,h^-} \to \mathcal{M}_{i,h}$. We denote by $g^a_h = (g^a_{1,h}, \cdots, g^a_{n,h})$ the joint control strategy and by $g^m_h = (g^m_{1,h}, \cdots, g^m_{n,h})$ the joint communication strategy. Note that the communication strategies are based on all the information available right before additional sharing, i.e., τ_{i,h^-}, to determine the communication action $m_{i,h}$; the control strategies are based on all the information available after additional sharing, i.e., τ_{i,h^+}, to determine the control action $a_{i,h}$. We denote by $\mathcal{G}^a_{i,h}, \mathcal{G}^m_{i,h}, \mathcal{G}^a_h, \mathcal{G}^m_h$ the corresponding spaces of $g^a_{i,h}, g^m_{i,h}, g^a_h, g^m_h$, respectively.

The objective of the agents in the LTC problem is to maximize the expected accumulated sum of the reward and the negative communication cost from timestep $h = 1$ to H:

$$J_{\mathcal{L}}(g^a_{1:H}, g^m_{1:H}) := \mathbb{E}_{\mathcal{L}}\left[\sum_{h=1}^{H}(r_h - \kappa_h) \,\middle|\, g^a_{1:H}, g^m_{1:H}\right],$$

where the expectation $\mathbb{E}_{\mathcal{L}}$ is taken over all the randomness in the system evolution, given the strategies $(g^a_{1:H}, g^m_{1:H})$. With this objective, for any $\epsilon \geq 0$, we can define the solution concept of *ϵ-team optimum* for $\mathcal{L}$ as follows.

Definition II.2 (ϵ-team optimum)**.** We call a joint strategy $(g^a_{1:H}, g^m_{1:H})$ an ϵ-team optimal strategy of the LTC $\mathcal{L}$ if

$$\max_{\widetilde{g}^a_{1:H} \in \mathcal{G}^a_{1:H}, \widetilde{g}^m_{1:H} \in \mathcal{G}^m_{1:H}} J_{\mathcal{L}}(\widetilde{g}^a_{1:H}, \widetilde{g}^m_{1:H}) - J_{\mathcal{L}}(g^a_{1:H}, g^m_{1:H}) \leq \epsilon.$$

B. Information Structures of LTC

In decentralized stochastic control, the notion of information structure [24], [12] captures *who knows what and when* as the system evolves. In LTC, as the additional sharing via communication will also affect the IS and is *not* determined *beforehand*, when we discuss the *IS of an LTC problem*, we will refer to that of the problem *with only baseline sharing*. In particular, an LTC $\mathcal{L}$ without additional sharing is essentially a Dec-POMDP (with potential baseline information sharing), as defined in §D for completeness. We formally define such a Dec-POMDP *induced* by $\mathcal{L}$ as follows.

Definition II.3 (Dec-POMDP (with information sharing) induced by LTC)**.** For an LTC $\mathcal{L} =$

$\langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{M}_i\}_{i\in[n]}, \mathbb{T}, \mathbb{O}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}, \{\mathcal{K}_h\}_{h\in[H]}\rangle$, we call a Dec-POMDP (with information sharing) $\overline{\mathcal{D}}_{\mathcal{L}}$ *the Dec-POMDP (with information sharing) induced by* $\mathcal{L}$ if the agents share information only following the baseline sharing protocol of $\mathcal{L}$, i.e., without additional sharing, which can be characterized by the tuple $\overline{\mathcal{D}}_{\mathcal{L}} := \langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \mathbb{T}, \mathbb{O}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}\rangle$. We may refer to it as the *Dec-POMDP induced by LTC* or the *induced Dec-POMDP* for short.

In §II-A, we introduced LTC in the *state-space model*. Information structure is oftentimes more conveniently discussed under the equivalent framework of *intrinsic models* [24] (see the instantiation for Dec-POMDPs in §D.1 for completeness). In an intrinsic model, each agent only *acts once* throughout the problem evolution, and the same agent in the state-space model at different timesteps is now treated as *different agents*. There are thus $n \times H$ agents in total. Formally, for completeness, we extend the intrinsic-model-based reformulation to LTCs in §D.3.

Quasi-classical and strictly quasi-classical ISs are important subclasses of ISs, which were first introduced for decentralized stochastic control [24], [25], [13] (see the instantiation for Dec-POMDPs in §D.2). An IS that is not QC is *non-classical* [12], [13]. We extend such a categorization to LTC problems with different ISs as follows.

Definition II.4 ((Strictly) quasi-classical LTC)**.** *We call an LTC* $\mathcal{L}$ *(strictly) quasi-classical if the Dec-POMDP induced by* $\mathcal{L}$ *(c.f. Definition II.3) is (strictly) quasi-classical. Namely, each agent in the intrinsic model of* $\overline{\mathcal{D}}_{\mathcal{L}}$ *knows the information (and the actions) of the agents who influence her, either directly or indirectly.*

Similarly, an LTC $\mathcal{L}$ that is not QC is called *non-classical*. Note that the categorization above is defined based on the ISs *before* additional sharing, as an inherent property of the LTC problem per se, since additional sharing is the solution *to be* decided/learned, based on the given LTC model. We focus on finding such a solution in the next sections.

III. Hardness and Necessary Assumptions

It is known that computing an (approximate) team-optimum in Dec-POMDPs, which are LTCs *without* information-sharing, is `NEXP-hard` [14]. The hardness cannot be fully circumvented even when agents are allowed to share information: even if agents share all the information, the LTC problem becomes a Partially Observable Markov Decision Process (POMDP), which is known to be `PSPACE-hard` [18], [19]. Hence, additional assumptions are necessary to make LTCs computationally tractable. We introduce several such assumptions and their justifications below, whose proofs can be found in §B.

Recently, [26] showed that *observable* POMDPs, a class of POMDPs with relatively *informative* observations, allow *quasi-polynomial time* algorithms to solve. Such a condition was then generalized to the *joint* emission function of Dec-POMDPs in [15], enabling *quasi-polynomial time* algorithms. As solving LTCs is at least as hard as solving the Dec-POMDPs considered in [15] (with void additional sharing), we first also make such an observability assumption, to avoid computationally intractable oracles.

Assumption III.1 (γ-observability [27], [26], [15])**.** There exists a $\gamma > 0$ such that $\forall h \in [H]$, the emission $\mathbb{O}_h$ satisfies that $\forall b_1, b_2 \in \Delta(\mathcal{S})$ $\left\|\mathbb{O}_h^\top b_1 - \mathbb{O}_h^\top b_2\right\|_1 \geq \gamma \left\|b_1 - b_2\right\|_1$.

However, we show next that, Assumption III.1 is not enough when it comes to LTC, if the baseline sharing IS is not favorable, in particular, *non-classical* [12]. The hardness persists even under a few additional assumptions to be introduced later that will make LTCs tractable.

Lemma III.2 (Non-classical LTCs are hard)**.** There exists an $\epsilon > 0$, such that even under Assumption III.1, together with Assumptions III.4, III.5, III.7, and IV.7, computing an ϵ/H-team optimum for non-classical LTCs is `PSPACE-hard`.

Note that Assumptions III.1 and IV.7 were sufficient for finding an ϵ-approximate team-optimum when there is no additional sharing [15], and Assumption IV.7, in particular, rules out the source of hardness due to the intractability of *one-step* team-decision problems [28]. Rather, the hardness comes from the *non-classicality* of the baseline sharing IS.

By Lemma III.2, we will hence focus on the *quasi-classical* LTCs hereafter. Indeed, QC is also known to be critical for efficiently solving *continuous-space* and *linear* decentralized control [29], [30]. However, in our discrete setting, even QC LTCs may not be computationally tractable: the additional sharing may *break* the QC IS, and introduce computational hardness. We formalize this intuition with the following discussions on when *QC may break*, and computational hardness results to justify the associated assumptions.

Firstly, QC may break by additional sharing, if an agent influences others (only) through such sharing, while others cannot fully access the information used for determining the *communication action*. Indeed, the general communication-strategy space in §II-A.0.d allows the dependence on agents' *private information*, making this case possible. We show next that this may cause computational hardness in general.

Lemma III.3 (QC LTCs with full-history-dependent communication strategies are hard)**.** For QC LTCs, even under Assumption III.1, together with Assumptions III.5, III.7, and IV.7, computing a team-optimum in the general space of $(\mathcal{G}^a_{1:H}, \mathcal{G}^m_{1:H})$ with $\mathcal{G}^m_{i,h} := \{g^m_{i,h} : \mathcal{T}_{i,h^-} \to \mathcal{M}_{i,h}\}$ is `NP-hard`.

To avoid this hardness, we thus focus on communication strategies that only condition on the *common information*. Intuitively, this assumption is not unreasonable, as it means that *which historical information to share* is determined by *what has been shared* (in the common information). Note that, this does not lose the generality in the sense that the private information p_{i,h^-} *can still* be shared. It only means that the communication action is not determined based on p_{i,h^-}, and the additional sharing is still dictated by $z^a_{i,h} = \phi_{i,h}(p_{i,h^-}, m_{i,h})$ (c.f. Assumption II.1), depending on p_{i,h^-}.

Assumption III.4 (Common-information-based communication strategy)**.** The communication strategies take *common information* as input, with the following form:

$$\forall i \in [n], h \in [H], \quad g^m_{i,h} : \mathcal{C}_{h^-} \to \mathcal{M}_{i,h}. \tag{III.1}$$

Secondly, QC may break by additional sharing if it makes an agent *influence* others(' available information) by *sharing* her *control* actions, while these other agents were *not influenced* by the agent in the baseline sharing, and thus did not have to access the available information that the agent decided her control actions upon. We make the following two assumptions to avoid the related pessimistic cases, followed by the hardness results when they are missing. The common idea behind the hardness results in both Lemmas III.6 and III.8 exactly follows from this insight.

Specifically, in some special cases, the action of some agents may not influence the state transition. Such actions are thus *useless* in terms of decision-making, when there is *no* information sharing. However, if they were deemed *non-influential* (and thus the later agents do not have to access the information this action was based on, still satisfying QC), but shared via additional sharing, then QC breaks in LTC. We thus make the following assumption that such useless actions will never be shared, followed by a justification result.

Assumption III.5 (Control-useless action is not used)**.** For each $i \in [n], h \in [H]$, if agent i's action $a_{i,h}$ does not influence the state s_{h+1}, namely, $\forall s_h \in \mathcal{S}, a_h \in \mathcal{A}_h, a'_{i,h} \in \mathcal{A}_{i,h}, a'_{i,h} \neq a_{i,h}, \mathbb{T}_h(\cdot \,|\, s_h, a_h) = \mathbb{T}_h(\cdot \,|\, s_h, (a'_{i,h}, a_{-i,h}))$. Then, $\forall h' > h, a_{i,h} \notin \tau_{h'^-}$ and $a_{i,h} \notin \tau_{h'^+}$.

Lemma III.6 (QC LTCs without Assumption III.5 are hard)**.** For QC LTCs, even under Assumptions III.1, III.4, III.7 and IV.7, finding a team-optimum without Assumption III.5 is `NP-hard`.

Note that other than the justification above based on computational hardness, Assumption III.5 has been implicitly made in the IS examples in the literature when there are *uncontrolled* state dynamics, see e.g., [17], [15][Kaiqing: is there more ref?]. Moreover, we emphasize that for common cases where actions *do* affect the state transition, this assumption becomes irrelevant and thus not needed.

Other than *not influencing* state transition, an action may also be non-influential if the emission functions of other agents are *degenerate*: they cannot sense the influence from previous agents' actions. We thus make the following assumption on the emissions, followed by a justification result.

Assumption III.7 (Other agents' emissions are non-degenerate)**.** For $\forall h \in [H], i \in [n]$, $\mathbb{O}_{-i,h}$ satisfies $\forall b_1, b_2 \in \Delta(\mathcal{S}), b_1 \neq b_2,\ \mathbb{O}^\top_{-i,h} b_1 \neq \mathbb{O}^\top_{-i,h} b_2$.

Lemma III.8 (QC LTCs without Assumption III.7 are hard)**.** For QC LTCs, even under Assumption III.1, III.4, III.5, and IV.7, finding an ϵ-team optimum without Assumption III.7 is `PSPACE-hard`.

Finally, for both the baseline and additional sharing protocols, we follow the convention in the series of works on partial history/information sharing [16], [17], [15], [9], [10] that, if an agent decides to share part of her information, she will share the information with *all other* agents. We make it more formally as follows.

Assumption III.9. $\forall i_1, i_2 \in [n], h_1, h_2 \in [H], i_1 \neq i_2, h_1 < h_2$, if $\sigma(o_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^-})$, then $\sigma(o_{i_1,h_1}) \subseteq \sigma(c_{h_2^-})$, and if $\sigma(a_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^-})$, then $\sigma(a_{i_1,h_1}) \subseteq \sigma(c_{h_2^-})$; if $\sigma(o_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^+})$, then $\sigma(o_{i_1,h_1}) \subseteq \sigma(c_{h_2^+})$, and if $\sigma(a_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^+})$, then $\sigma(a_{i_1,h_1}) \subseteq \sigma(c_{h_2^+})$.

As will be shown later (c.f. Theorem IV.2), LTCs under Assumptions III.4, III.5, III.7, and III.9 can indeed *preserve* the QC/sQC information structure after additional sharing, making it possible for the overall LTC problem to be computationally tractable, as we will show next. Some examples that satisfy these assumptions can also be found in §A[Kaiqing: is this correct?].

IV. Solving LTC Problems Provably

We now study how to solve LTC provably, via either *planning* (with model knowledge) or *learning* (without model knowledge). Proofs of the results can be found in §B.

A. An Equivalent Dec-POMDP

Given any LTC $\mathcal{L}$, we can define a Dec-POMDP $\mathcal{D}_\mathcal{L}$ given by $\langle \widetilde{H}, \widetilde{\mathcal{S}}, \{\widetilde{\mathcal{A}}_{i,h}\}_{i\in[n],h\in[\widetilde{H}]}, \{\widetilde{\mathcal{O}}_{i,h}\}_{i\in[n],h\in[\widetilde{H}]}, \widetilde{\mathbb{T}}, \widetilde{\mathbb{O}}, \widetilde{\mu}_1, \{\widetilde{\mathcal{R}}_h\}_{h\in[\widetilde{H}]}\rangle$, such that these two problems are equivalent. The elements in $\mathcal{D}_\mathcal{L}$ can be specified as follows:

$$\begin{gathered}\widetilde{H} = 2H, \quad \widetilde{\mathcal{S}} = \mathcal{S}, \quad \widetilde{s}_{2h-1} = \widetilde{s}_{2h} = s_h, \quad \widetilde{\mathcal{A}}_{i,2h-1} = \mathcal{M}_{i,h}, \\ \widetilde{\mathcal{A}}_{i,2h} = \mathcal{A}_{i,h}, \quad \widetilde{a}_{i,2h-1} = m_{i,h}, \quad \widetilde{a}_{i,2h} = a_{i,h}, \\ \widetilde{\mathcal{O}}_{i,2h-1} = \mathcal{O}_{i,h}, \quad \widetilde{\mathcal{O}}_{i,2h} = \{\emptyset\}, \quad \widetilde{o}_{i,2h-1} = o_{i,h}, \quad \widetilde{o}_{i,2h} = \emptyset, \\ \widetilde{\mathbb{T}}_{2h-1}(\widetilde{s}_{2h} \,|\, \widetilde{s}_{2h-1}, \widetilde{a}_{2h-1}) = \mathbb{1}[\widetilde{s}_{2h} = \widetilde{s}_{2h-1}], \\ \widetilde{\mathbb{T}}_{2h}(\widetilde{s}_{2h+1} \,|\, \widetilde{s}_{2h}, \widetilde{a}_{2h}) = \mathbb{T}_h(\widetilde{s}_{2h+1} \,|\, \widetilde{s}_{2h}, \widetilde{a}_{2h}), \\ \widetilde{\mu}_1 = \mu_1, \quad \widetilde{\mathbb{O}}_{2h-1} = \mathbb{O}_h, \quad \widetilde{\mathcal{R}}_{2h-1} = -\mathcal{K}_h, \quad \widetilde{\mathcal{R}}_{2h} = \mathcal{R}_h, \\ \widetilde{p}_{i,2h-1} = p_{i,h^-}, \quad \widetilde{p}_{i,2h} = p_{i,h^+}, \quad \widetilde{c}_{2h-1} = c_{h^-}, \quad \widetilde{c}_{2h} = c_{h^+}, \\ \widetilde{z}_{2h-1} = z^b_h, \quad \widetilde{z}_{2h} = z^a_h, \quad \widetilde{\tau}_{i,2h-1} = c_{h^-}, \quad \widetilde{\tau}_{i,2h} = \tau_{i,h^+},\end{gathered} \tag{IV.1}$$

for all $(i,h) \in [n] \times [H]$, $s_h \in \mathcal{S}, a_{i,h} \in \mathcal{A}_{i,h}, o_{i,h} \in \mathcal{O}_{i,h}, m_{i,h} \in \mathcal{M}_{i,h}, p_{i,h^-} \in \mathcal{P}_{i,h^-}, p_{i,h^+} \in \mathcal{P}_{i,h^+}, c_{h^-} \in \mathcal{C}_{h^-}, c_{h^+} \in \mathcal{C}_{h^+}, \tau_{i,h^-} \in \mathcal{T}_{i,h^-}, \tau_{i,h^+} \in \mathcal{T}_{i,h^+}$. It is important to note that, at the odd timestep $2h-1$ with $h \in [H]$, for each agent $i \in [n]$, we set $\widetilde{\tau}_{i,2h-1} = c_{h^-}$ under Assumption III.4, i.e., in $\mathcal{D}_\mathcal{L}$, it is *as if* the available information for decision-making at timestep $2h-1$ is only the common information so far. Correspondingly, for any $h \in [\widetilde{H}], i \in [n]$, we denote by $\widetilde{g}_{i,h}, \widetilde{g}_h$ the (joint) strategy and by $\widetilde{\mathcal{G}}_{i,h}, \widetilde{\mathcal{G}}_h$ the (joint) strategy spaces. Similarly, the objective of $\mathcal{D}_\mathcal{L}$ is defined as $J_{\mathcal{D}_\mathcal{L}}(\widetilde{g}_{1:\widetilde{H}}) = \mathbb{E}_{\mathcal{D}_\mathcal{L}}[\sum_{h=1}^{\widetilde{H}} \widetilde{r}_h \,|\, \widetilde{g}_{1:\widetilde{H}}]$.

Essentially, this reformulation splits the H-step decision-making and communication procedure into a $2H$-step one. A similar splitting of the timesteps was also used in [9], [10]. In comparison, we consider a more general setting, where the state is not decoupled and agents are allowed to share the observations and actions at the *previous* timesteps, due to the generality of our LTC formulation. The equivalence between $\mathcal{L}$ and $\mathcal{D}_{\mathcal{L}}$ is more formally stated as follows.

Proposition IV.1 (Equivalence between $\mathcal{L}$ and $\mathcal{D}_{\mathcal{L}}$)**.** Let $\mathcal{D}_{\mathcal{L}}$ be the reformulated Dec-POMDP from $\mathcal{L}$, then the solutions of the two problems are equivalent, in the sense that $\forall g^m_{1:H} \in \mathcal{G}^m_{1:H}, g^a_{1:H} \in \mathcal{G}^a_{1:H}, i \in [n]$[Kaiqing: what are these $\mathcal{G}^m_{1:H}, \mathcal{G}^a_{1:H}$ notations.. pls check all..][Haoyi:defined in strategy section.], let $\widetilde{g}_{1:\widetilde{H}} = (g^m_1, g^a_1, \cdots, g^m_H, g^a_H)$, then $J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) = J_{\mathcal{L}}(g^m_{1:H}, g^a_{1:H})$. Also, $\forall \widetilde{g}_{1:\widetilde{H}} \in \widetilde{\mathcal{G}}_{1:\widetilde{H}}, i \in [n]$, let $g^m_{1:H} = (\widetilde{g}_1, \widetilde{g}_3, \cdots, \widetilde{g}_{\widetilde{H}-1}), g^a_{1:H} = (\widetilde{g}_2, \widetilde{g}_4, \cdots, \widetilde{g}_{\widetilde{H}})$, then $J_{\mathcal{L}}(g^m_{1:H}, g^a_{1:H}) = J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}})$.

Also, the Dec-POMDP $\mathcal{D}_{\mathcal{L}}$ inherits the QC IS from $\mathcal{L}$.

Theorem IV.2 (Preserving (s)QC)**.** If $\mathcal{L}$ is (s)QC and satisfies Assumptions III.4, III.5, III.7, and III.9, then the reformulated Dec-POMDP $\mathcal{D}_{\mathcal{L}}$ is also (s)QC.

By Proposition IV.1, it suffices to solve the reformulated $\mathcal{D}_{\mathcal{L}}$ that are QC/sQC, which will be our focus next.

B. Strict Expansion of $\mathcal{D}_{\mathcal{L}}$

Despite being QC/sQC, it is not clear if one can solve the $\mathcal{D}_{\mathcal{L}}$ in finite-time without computationally intractable oracles. Note that this is different from the continuous-space, linear quadratic case, where QC problems can be reformulated and solved efficiently [29], [31]. With discrete spaces, to the best of our knowledge, the only known finite-time computational complexity results for planning in such decentralized control models were in [15], which were established under the *strategy independence* assumption [17] on the common-information-based beliefs [16], [17]. In particular, this SI assumption was critical for computational purposes – it eliminates the need for *enumerating* the past strategies for the backward induction in dynamic programming, which would otherwise be prohibitively large. Hence, we need to connect QC/sQC to SI for more efficient computation.

Interestingly, under certain conditions, one can connect QC with SI-CIB for the reformulated Dec-POMDP $\mathcal{D}_{\mathcal{L}}$. As the first step, we will *expand* the QC $\mathcal{D}_{\mathcal{L}}$ by adding the *actions* of the agents who influence the later agents in the intrinsic model of $\mathcal{D}_{\mathcal{L}}$ to the shared information. We denote the strictly expanded Dec-POMDP as $\mathcal{D}^{\dagger}_{\mathcal{L}}$. We replace the $\widetilde{\ }$ notation in $\mathcal{D}_{\mathcal{L}}$ by the $\breve{\ }$ notation in $\mathcal{D}^{\dagger}_{\mathcal{L}}$. The horizon ($\breve{H} = \widetilde{H} = 2H$), states, actions, observations, transitions, and reward functions remain the same, but the sets of information $\breve{p}_h, \breve{c}_h, \breve{\tau}_h, \breve{p}_{i,h}, \breve{\tau}_{i,h}$ are different: for any $h \in [\widetilde{H}], i \in [n]$

$$\begin{aligned} \breve{c}_h =& \widetilde{c}_h \cup \{\widetilde{a}_{j,t} \,|\, j \in [n], t < h, \sigma(\widetilde{\tau}_{j,t}) \subseteq \sigma(\widetilde{c}_h)\} \\ \breve{p}_{i,h} =& \widetilde{p}_{i,h} \backslash \{\widetilde{a}_{i,t} \,|\, t < h, \sigma(\widetilde{\tau}_{i,t}) \subseteq \sigma(\widetilde{c}_h)\}. \end{aligned} \tag{IV.2}$$

It is not hard to verify the following.

Lemma IV.3. If $\mathcal{D}_{\mathcal{L}}$ is QC, then $\mathcal{D}^{\dagger}_{\mathcal{L}}$ is sQC.

In contrast to the reformulation in §IV-A, the expansion here cannot guarantee the equivalence between $\mathcal{D}_{\mathcal{L}}$ and $\mathcal{D}^{\dagger}_{\mathcal{L}}$: the strategy spaces of $\mathcal{D}^{\dagger}_{\mathcal{L}}$ are larger than those of $\mathcal{D}_{\mathcal{L}}$, as each agent can now access more information, i.e., $\widetilde{\tau}_{i,h} \subseteq \breve{\tau}_{i,h}$. Fortunately, the team-optimal value and strategy of both Dec-POMDPs are related, as shown in the following theorem.

Theorem IV.4. Let $\mathcal{D}_{\mathcal{L}}$ be the QC Dec-POMDP reformulated from a QC LTC $\mathcal{L}$, and $\mathcal{D}^{\dagger}_{\mathcal{L}}$ be the sQC expansion of $\mathcal{D}_{\mathcal{L}}$. Then, for any ϵ-team-optimal strategy $\breve{g}^*_{1:\breve{H}}$ of $\mathcal{D}^{\dagger}_{\mathcal{L}}$, there exists a function φ such that $\widetilde{g}^*_{1:\widetilde{H}} = \varphi(\breve{g}^*_{1:\breve{H}}, \mathcal{D}_{\mathcal{L}})$ is an ϵ-team-optimal strategy of $\mathcal{D}_{\mathcal{L}}$, with $J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}^*_{1:\widetilde{H}}) = J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\breve{g}^*_{1:\breve{H}})$.

Theorem IV.4 shows that one can solve the QC $\mathcal{D}_{\mathcal{L}}$ by first solving the sQC expansion $\mathcal{D}^{\dagger}_{\mathcal{L}}$, and then using an oracle φ to translate it back as a solution in the strategy spaces of $\mathcal{D}_{\mathcal{L}}$, without loss of optimality. Importantly, we show in Algorithm 4 that such φ can be implemented efficiently.

As shown below, a benefit of obtaining an *sQC* $\mathcal{D}^{\dagger}_{\mathcal{L}}$ is that, it is also *SI-CIB*, making it possible to be solved without computationally intractable oracles as in [15].

Theorem IV.5. Let $\mathcal{D}^{\dagger}_{\mathcal{L}}$ be an sQC Dec-POMDP generated from $\mathcal{L}$ after reformulation and strict expansion, then $\mathcal{D}^{\dagger}_{\mathcal{L}}$ has *strategy-independent common-information-based beliefs* [17], [15]. More formally, for any $h \in [\breve{H}]$, any two different joint strategies $\breve{g}_{1:h-1}$ and $\breve{g}'_{1:h-1}$, and any common information $\breve{c}_h$ can be reached under strategy $\breve{g}_{1:h-1}$, for any joint private information $\breve{p}_h \in \breve{\mathcal{P}}_h$ and state $\breve{s}_h \in \breve{\mathcal{S}}$,

$$\mathbb{P}^{\mathcal{D}^{\dagger}_{\mathcal{L}}}_h(\breve{s}_h, \breve{p}_h \,|\, \breve{c}_h, \breve{g}_{1:h-1}) = \mathbb{P}^{\mathcal{D}^{\dagger}_{\mathcal{L}}}_h(\breve{s}_h, \breve{p}_h \,|\, \breve{c}_h, \breve{g}'_{1:h-1}). \tag{IV.3}$$

C. Refinement of $\mathcal{D}^{\dagger}_{\mathcal{L}}$

Despite of being SI, $\mathcal{D}^{\dagger}_{\mathcal{L}}$ is still not eligible for applying the results in [15]: the information evolution rules of $\mathcal{D}^{\dagger}_{\mathcal{L}}$ break those in [17], [15] (c.f. Assumption 1 therein). Specifically, due to Assumption III.4, we set $\widetilde{\tau}_{2t-1} = \widetilde{c}_{2t-1}, \widetilde{p}_{2t-1} = \emptyset, \forall t \in [H]$ in $\mathcal{D}_{\mathcal{L}}$, which violates Assumption 1 in [17], [15]. Since the strict expansion only adds actions, the expanded $\mathcal{D}^{\dagger}_{\mathcal{L}}$ thus also violates the assumption. To address this issue, we propose to further *refine* the $\mathcal{D}^{\dagger}_{\mathcal{L}}$ to obtain a Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$, which satisfies the information evolution rules. We replace the $\breve{\ }$ notation in $\mathcal{D}^{\dagger}_{\mathcal{L}}$ by the $\bar{\ }$ notation in $\mathcal{D}'_{\mathcal{L}}$. The elements in $\mathcal{D}'_{\mathcal{L}}$ remain the same as those in $\mathcal{D}^{\dagger}_{\mathcal{L}}$, except that the private information at odd steps is refined as

$$\overline{p}_{i,2t-1} = p_{i,t-} \backslash \breve{c}_{2t-1}. \tag{IV.4}$$

The new Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ is not equivalent to $\mathcal{D}^{\dagger}_{\mathcal{L}}$ in general, since it enlarges the strategy space at the odd timesteps. However, if we define new strategy spaces in $\mathcal{D}'_{\mathcal{L}}$ as $\overline{\mathcal{G}}_{i,2t-1} : \overline{\mathcal{C}}_{2t-1} \to \overline{\mathcal{A}}_{i,2t-1}, \overline{\mathcal{G}}_{i,2t} : \overline{\mathcal{T}}_{i,2t} \to \overline{\mathcal{A}}_{i,2t}$ for each $t \in [H], i \in [n]$, and thus define $\overline{\mathcal{G}}_h$ to be the corresponding joint space, then solving $\mathcal{D}^{\dagger}_{\mathcal{L}}$ is equivalent to finding a *best-in-class* team-optimal strategy of $\mathcal{D}'_{\mathcal{L}}$ within the space $\overline{\mathcal{G}}_{1:\overline{H}}$. Formally, we have the following theorem.

Theorem IV.6. Let $\mathcal{D}^{\dagger}_{\mathcal{L}}$ be an sQC Dec-POMDP generated from $\mathcal{L}$ after reformulation and strict expansion, and $\mathcal{D}'_{\mathcal{L}}$ be the refinement of $\mathcal{D}^{\dagger}_{\mathcal{L}}$ as above. Then, $\mathcal{D}'_{\mathcal{L}}$ satisfies the information evolution rules that for each $h \in [\overline{H}]$:

$$\begin{aligned} \overline{c}_{h+1} = \overline{c}_h \cup \overline{z}_{h+1}, \overline{z}_{h+1} = \overline{\chi}_{h+1}(\overline{p}_h, \overline{a}_h, \overline{o}_{h+1}) \\ \text{for each } i \in [n], \quad \overline{p}_{i,h+1} = \overline{\xi}_{i,h+1}(\overline{p}_{i,h}, \overline{a}_{i,h}, \overline{o}_{i,h+1}), \end{aligned}$$

with some functions $\{\overline{\chi}_{h+1}\}_{h\in[\overline{H}}, \{\overline{\xi}_{i,h+1}\}_{i\in[n],h\in[\overline{H}]}$. Furthermore, $\mathcal{D}'_{\mathcal{L}}$ is SI-CIB with respect to the strategy spaces $\overline{\mathcal{G}}_{1:\overline{H}}$, i.e., for any $h \in [\overline{H}], \overline{s}_h \in \overline{\mathcal{S}}, \overline{p}_h \in \overline{\mathcal{P}}_h, \overline{c}_h \in \overline{\mathcal{C}}_h, \overline{g}_{1:h-1}, \overline{g}'_{1:h-1} \in \overline{\mathcal{G}}_{1:h-1}$, it holds that

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}). \quad \text{(IV.5)}$$

D. Planning in QC LTC with Quasi-polynomial Time

Now we focus on how to solve the SI Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ *computationally efficiently*, which has been studied in [15]. Given any such a Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$, [15] proposed to construct an (ϵ_r, ϵ_z)-*expected*-approximate common information model $\mathcal{M}$ through *finite memory* (as defined in §.9), when $\mathcal{D}'_{\mathcal{L}}$ is γ-observable. ϵ_r and ϵ_z here denote the approximation errors for rewards and transitions respectively, for which we defer a detailed introduction to XXX[Kaiqing: reminder...]. However, the Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ obtained from LTC has two key differences from those considered in [15]. First, $\mathcal{D}'_{\mathcal{L}}$ does not satisfy the γ-observability assumption throughout the whole $2H$ timesteps. Fortunately, the emissions of $\mathcal{D}'_L$ at all the odd steps are still γ-observable, which can lead to a similar result of belief contraction and near-optimality of finite-memory truncation as in [32], [15]. Second, the rewards at the odd steps can depend on the private information $\overline{p}_h$, instead of the underlying state $\overline{s}_h$. Thanks to the approximate common-information-based beliefs for $\mathcal{M}$ defined as $\{\mathbb{P}_h^{\mathcal{M}}(\overline{s}_h, \overline{p}_h \,|\, \widehat{c}_h)\}_{h\in[H]}$, which provide the *joint* (conditional) probability of both the state $\overline{s}_h$ and the private information $\overline{p}_h$, we can still properly evaluate the rewards at the odd steps in the algorithms of [15].

Hence, we can leverage a similar common-information-based approach as in [15] to find the optimal strategy $\overline{g}^*_{1:\overline{H}}$ by finding an optimal prescriptions $\gamma^*_{1:\overline{H}}$ under each possible $\overline{c}_{1:\overline{H}}$ in a fashion of backwards induction over the time step $h = \overline{H}, \cdots, 1$. Meanwhile, it is worth mentioning that at each step $h \in [\overline{H}]$, it requires maximizing the Q-value functions (as defined in §B.10) as follows

$$\left(\widehat{g}^*_{1,h}(\cdot \,|\, \widehat{c}_h, \cdot), \cdots, \widehat{g}^*_{n,h}(\cdot \,|\, \widehat{c}_h, \cdot)\right) \leftarrow \operatorname*{argmax}_{\gamma_h \in \Gamma_h} Q_h^{*,\mathcal{M}}(\widehat{c}_h, \gamma_h). \quad \text{(IV.6)}$$

Note that solving Eq. (IV.6) is `NP-hard` in general. Hence, the guarantee for the algorithms in [15] also relies on the following assumption, on the tractability of the *one-step* team-decision problem [28]. Note that this assumption is minimal for the computational tractability of finding a team-optimum in Dec-POMDPs/LTCs, as otherwise, even the $H = 1$ case is intractable [28]. That said, the structural results so far still hold without this assumption, and the hardness results in §III still hold even with this assumption.

Assumption IV.7 (One-step tractability)**.** Suppose the *one-step* team-decision problem of Eq. (IV.6) can be solved in polynomial time.

Assumption IV.7 can be satisfied for several classes of Dec-POMDPs with information sharing [15], which could result from structures of either the decision-making components (transition kernels, emissions, rewards), or the information structures. We also include several such structural conditions in §E for completeness.

Finally, we can develop an algorithm obtaining an $(2\overline{H}\epsilon_r + \overline{H}^2\epsilon_z)$-team optimal policy for $\mathcal{D}'_L$ (and thus also for $\mathcal{L}$) by planning in $\mathcal{M}$ with polynomial time complexity on the size of the space of the approximate common information $|\widehat{\mathcal{C}}_h|$ and other problem parameters, e.g., $|\mathcal{S}|$, $\overline{H}$, etc. Furthermore, if $\mathcal{L}$ has the baseline sharing in Appendix A, there exists an algorithm that can compute an ϵ-team optimal policy for $\mathcal{L}$ in quasi-polynomial time w.r.t problem parameters, e.g., $|\mathcal{S}|$, $\overline{H}$, etc, for any fixed $\epsilon > 0$. We defer the formal guarantees to XXX [Kaiqing: reminder..]

E. LTC with Quasi-polynomial Time and Samples

Based on the previous result on planning in QC LTC, we are ready to solve the *learning* problem without model knowledge using both quasi-polynomial time and sample complexities. In the learning setting, one can only sample from $\mathcal{L}$, making it difficult to obtain an SI $\mathcal{D}'_{\mathcal{L}}$ from $\mathcal{L}$ as before. Fortunately, the *reformulation* step (§IV-A) does not change the system dynamics, but only maps the information to different random variables; the *expansion* step (§IV-B) only requires agents to share more actions with each other, without changing the input and output of the environment; the *refinement* step (§IV-C) only recovers the private information the agents had in original $\mathcal{L}$ problem. [Kaiqing: but what about the third step, refinement?]. Therefore, we can treat the samples from $\mathcal{L}$ as the samples from the associated Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ obtained in §IV-D. This way, we can utilize similar algorithmic ideas in [15] to develop both time and sample (quasi-)efficient algorithms.

Specifically, we construct an expected approximate common information model that *depends on some given a strategy $\overline{g}_{1:\overline{H}}$ that generates the data for such a construction, which is denoted by $\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})$*. For such a model, one could *simulate* and *sample* by running strategy $\overline{g}_{1:\overline{H}}$ in the true model $\mathcal{G}$. The choice of $\overline{g}_{1:\overline{H}}$ will be carefully specified to ensure $\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})$ to be a good approximation of $\mathcal{D}'_{\mathcal{L}}$. Then one can learn an empirical estimator $\widehat{\mathcal{M}}(\overline{g}_{1:\overline{H}})$ of $\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})$ by sampling under $\overline{g}_{1:\overline{H}}$ and solving the planning problem in $\widehat{\mathcal{M}}(\overline{g}_{1:\overline{H}})$. Meanwhile, the sample complexity analysis of such an algorithm will depend on the notion of *length* for the approximate common information, denoted as $\widehat{L}$. We defer the formal introduction for $\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})$, $\widehat{L}$, and corresponding algorithm to XXX[Kaiqing: reminder..]. Finally, we present our main results for solving the LTC problem.

Theorem IV.8. Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, III.7, and IV.7, we can construct an SI Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ such that the following holds. Given a strategy $\overline{g}_{1:\overline{H}}$, $\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})$, and the length of approximate common information $\widehat{L}$, where each $\overline{g}^h$ is a complete strategy with $\overline{g}_{h,h-\widehat{L}:h} = \text{Unif}(\mathcal{A})$ for $h \in [\overline{H}]$, we define the statistical error for estimating $\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})$ using samples under $\overline{g}_{1:\overline{H}}$ as $\epsilon_{apx}(\overline{g}_{1:\overline{H}}, \widehat{L}, \zeta_1, \zeta_2, \theta_1, \theta_2, \phi)$ for some parameters $\delta_1, \zeta_1, \zeta_2, \theta_1, \theta_2, \phi > 0$. Then there exists an algorithm that can learn an ϵ-team optimal strategy for $\mathcal{L}$ with probability at least $1 - \delta_1$, using a sample complexity $N_0 = \texttt{poly}(\max_{h \in [\overline{H}]} |\mathcal{P}_h|, \max_{h \in [\overline{H}]} |\widehat{\mathcal{C}}_h|, H, \max_{h \in [\overline{H}]} |\mathcal{A}_h|, \max_{h \in [\overline{H}]} |\mathcal{O}_h|, 1/\zeta_1, 1/\zeta_2, 1/\theta_1, 1/\theta_2) \cdot \log(1/\delta_1)$, where $\epsilon := \overline{H}\epsilon_r(\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})) + \overline{H}^2\epsilon_z(\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})) + (\overline{H}^2 + \overline{H})\epsilon_{apx}(\overline{g}_{1:\overline{H}}, \widehat{L}, \zeta_1, \zeta_2, \theta_1, \theta_2, \phi) + \overline{H}\epsilon_e$. Specifically, if $\mathcal{L}$ has the baseline sharing as in Appendix A, there exists an algorithm that can construct an $\widetilde{\mathcal{M}}(\overline{g}_{1:\overline{H}})$, learn an $\widehat{\mathcal{M}}(\overline{g}_{1:\overline{H}})$, and finally compute an ϵ-team optimal strategy for $\mathcal{L}$ with both quasi-polynomial time and sample complexities.

V. Solving General QC Dec-POMDPs

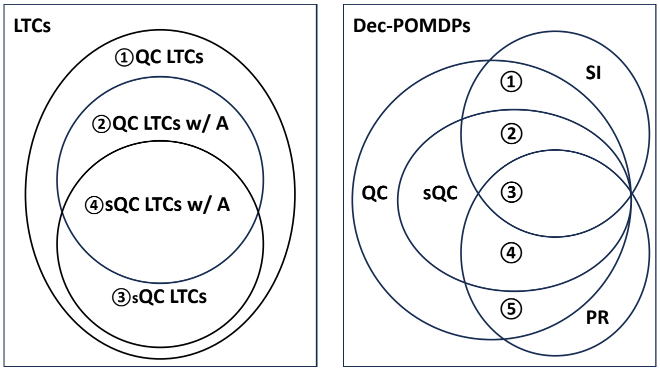
In §IV, we developed a pipeline for solving a special class of QC Dec-POMDPs generated by LTCs, without computationally intractable oracles. In fact, the pipeline can also be extended to solving general QC Dec-POMDPs, which thus advances the results in [15] that can only address *SI-CIB* Dec-POMDPs, a result of independent interest. Without much confusion given the context, we will adapt the notation of LTC to study general Dec-POMDPs: we set $h^+ = h^- = h$ and void the additional sharing protocol. We extend the results in §IV to general QC Dec-POMDPs as follows.

Theorem V.1. Consider a Dec-POMDP $\mathcal{D}$ satisfying Assumptions II.1 (e). If $\mathcal{D}$ is sQC and satisfies Assumptions III.5, and III.7, then $\mathcal{D}$ is SI. Meanwhile, if $\mathcal{D}$ is SI and has perfect recall, then $\mathcal{D}$ is sQC.

Note that perfect recall [23] means agents will never forget their information and actions in the past, and is formally defined in §V. Assumption II.1 (e) is similar but different from perfect recall. It will be implied by perfect recall with $o_{i,h} \in \tau_{i,h}$. Also, Assumptions III.5 and III.7 were originally made for LTCs, and here we meant to impose them for Dec-POMDPs with $h^+ = h^- = h$. Given Theorem V.1 and the results in §IV, we can illustrate the relationship between LTCs and Dec-POMDPs under different assumptions and ISs in Fig. 1, which may be of independent interest.

VI. Experimental Results

For the experiments, we validate both the implementability and performance of our learning-to-communicate approaches, and conduct an ablation study for LTCs with different communication costs. We conduct the experiments in problems Dectiger and Grid3x3, and the setup details are deferred to §G. We train the agents in each LTC problem in two environments with 20 different random seeds and different communication cost functions, and execute them in problems with horizons $[4, 6, 8, 10]$. The attained average-values are presented in Fig. 2, and the learning curves are shown in Fig. 3. The results shows additional sharing is beneficial for agents to obtain higher reward. And, cheaper communication cost will encourage agents to share more and achieve better optimal strategy. [Kaiqing: say more? esp. related to our theory?][Haoyi:it is hard to relate to our theory..]



(a) (b)

Fig. 1: (a) Venn diagram of LTCs with different ISs: ① QC LTCs. ② QC LTCs satisfying Assumptions III.4, III.5, and III.7. ③ sQC LTCs. ④ sQC LTCs satisfying Assumptions III.4, III.5, and III.7, whose reformulated Dec-POMDPs have SI-CIB; (b) Venn diagram of general Dec-POMDPs with different ISs. PR denotes perfect recall. ③ are the Dec-POMDPs we mainly consider, e.g., the examples in [17], [15]. Other examples in various areas can be found in §F.

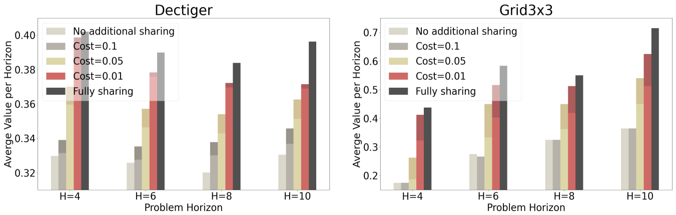


Fig. 2: The average-values achieved under different communication costs and horizons. Each full bar, the dark part, and the light part denote the values associated with the reward, the communication cost, and the overall objective (reward minus cost) of the agents, respectively. Note that, as baselines, there is no communication cost in the *no additional sharing* and *fully sharing* cases.

References

[1] J. Foerster, I. A. Assael, N. De Freitas, and S. Whiteson, "Learning to communicate with deep multi-agent rl," in *NeurIPS*, 2016.

[2] S. Sukhbaatar, R. Fergus, *et al.*, "Learning multiagent communication with backpropagation," in *NeurIPS*, 2016.

[3] J. Jiang and Z. Lu, "Learning attentional communication for multi-agent cooperation," in *NeurIPS*, 2018.

[4] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, pp. 1056–1068, 2004.

[5] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceed. of the IEEE*, vol. 95, pp. 108–137, 2007.

[6] A. Chamaken and L. Litz, "Joint design of control and communication in wireless networked control systems: A case study," in *American Control Conf.* IEEE, 2010, pp. 1835–1840.

[7] L. Xiao, M. Johansson, H. Hindi, S. Boyd, and A. Goldsmith, "Joint optimization of wireless communication and networked control systems," *Switching and Learning Feedback Sys.*, pp. 248–272, 2005.

[8] S. Yüksel, "Jointly optimal LQG quantization and control policies for multi-dimensional systems," *IEEE Trans. Autom. Control*, vol. 59, pp. 1612–1617, 2013.

[9] S. Sudhakara, D. Kartik, R. Jain, and A. Nayyar, "Optimal communication and control strategies in a multi-agent mdp problem," *arXiv preprint arXiv:2104.10923*, 2021.

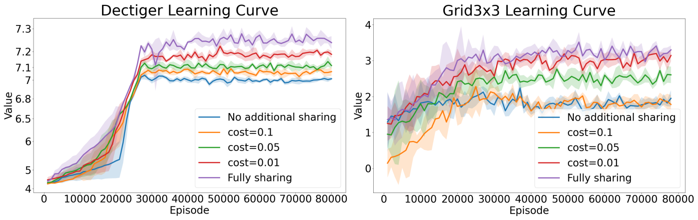


Fig. 3: Each line represents the learning curve of agents in the problem of different communication cost settings.

[10] D. Kartik, S. Sudhakara, R. Jain, and A. Nayyar, "Optimal communication and control strategies for a multi-agent system in the presence of an adversary," in *IEEE Conf. on Dec. and Control (CDC)*, 2022.

[11] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proceed. of the IEEE*, vol. 59, pp. 1557–1566, 1971.

[12] A. Mahajan, N. C. Martins, M. C. Rotkowitz, and S. Yüksel, "Information structures in optimal decentralized control," in *IEEE Conf. on Dec. and Control (CDC)*, 2012.

[13] S. Yüksel and T. Başar, *Stochastic Teams, Games, and Control under Information Constraints*. Springer Nature, 2023.

[14] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of markov decision processes," *Math. Oper. Res.*, vol. 27, pp. 819–840, 2002.

[15] X. Liu and K. Zhang, "Partially observable multi-agent reinforcement learning with information sharing," *arXiv preprint arXiv:2308.08705 (short version accepted at ICML 2023)*, 2023.

[16] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Trans. Autom. Control*, vol. 58, no. 7, pp. 1644–1658, 2013.

[17] A. Nayyar, A. Gupta, C. Langbort, and T. Başar, "Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games," *IEEE Trans. Autom. Control*, vol. 59, pp. 555–570, 2013.

[18] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Oper. Res.*, vol. 12, pp. 441–450, 1987.

[19] C. Lusena, J. Goldsmith, and M. Mundhenk, "Nonapproximability results for partially observable Markov decision processes," *J. Artif. Intell. Res.*, pp. 83–103, 2001.

[20] C. Jin, S. Kakade, A. Krishnamurthy, and Q. Liu, "Sample-efficient reinforcement learning of undercomplete pomdps," in *NeurIPS*, 2020.

[21] Q. Liu, C. Szepesvári, and C. Jin, "Sample-efficient reinforcement learning of partially observable Markov games," in *NeurIPS*, 2022.

[22] A. Altabaa and Z. Yang, "On the role of information structure in reinforcement learning for partially-observable sequential teams and games," in *NeurIPS*, 2024.

[23] H. W. Kuhn, "Extensive games and the problem of information," in *Contrib. Theory Games, Vol. II*. Princeton Univ. Press, 1953.

[24] H. S. Witsenhausen, "The intrinsic model for discrete stochastic control: Some open problems," in *Control Theory, Numer. Methods Comput. Syst. Model., Int. Symp., Rocquencourt*, 1975, pp. 322–335.

[25] A. Mahajan and S. Yüksel, "Measure and cost dependent properties of information structures," in *Amer. Control Conf. (ACC)*. IEEE, 2010, pp. 6397–6402.

[26] N. Golowich, A. Moitra, and D. Rohatgi, "Planning and learning in partially observable systems via filter stability," in *Proc. 55th Annu. ACM Symp. Theory Comput. (STOC)*, 2023.

[27] E. Even-Dar, S. M. Kakade, and Y. Mansour, "The value of observation for monitoring dynamic systems," in *IJCAI*, 2007.

[28] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Autom. Control*, vol. 30, pp. 440–446, 1985.

[29] Y.-C. Ho *et al.*, "Team decision theory and information structures in optimal control problems – part i," *IEEE Trans. Autom. Control*, vol. 17, pp. 15–22, 1972.

[30] A. Lamperski and L. Lessard, "Optimal decentralized state-feedback control with sparsity and delays," *Automatica*, pp. 143–151, 2015.

[31] M. Rotkowitz, "On information structures, convexity, and linear optimality," in *IEEE Conf. on Dec. and Control (CDC)*, 2008.

[32] N. Golowich, A. Moitra, and D. Rohatgi, "Planning in observable pomdps in quasipolynomial time," 2022. [Online]. Available: https://arxiv.org/abs/2201.04735

[33] ——, "Learning in observable POMDPs, without computationally intractable oracles," in *NeurIPS*, 2022.

[34] J. Filar and K. Vrieze, *Competitive Markov decision processes*. Springer, 2012.

[35] Y. Bai and C. Jin, "Provable self-play algorithms for competitive reinforcement learning," in *ICML*, 2020.

[36] J. Peralez, A. Delage, O. Buffet, and J. S. Dibangoye, "Solving hierarchical information-sharing Dec-POMDPs: an extensive-form game approach," *arXiv preprint arXiv:2402.02954*, 2024.

[37] C. Boutilier, "Multiagent systems: Challenges and opportunities for decision-theoretic planning," *AI magazine*, vol. 20, pp. 35–35, 1999.

[38] R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella, "Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings," in *IJCAI*, 2003.

[39] C. Amato, J. Dibangoye, and S. Zilberstein, "Incremental policy generation for finite-horizon Dec-POMDPs," in *Proc. Int. Conf. Autom. Plan. Sched. (ICAPS)*, vol. 19, 2009, pp. 2–9.

Appendix

A. Examples of QC LTC

[Haoyi:to check] In this section, we introduce 8 examples of QC LTC problems, and 4 of them are extended from the information structures of the baseline sharing protocol considered in the literature [17], [15]. It can be shown that LTC with any of these 8 examples as baseline sharing is QC.

- **Example 1: One-step delayed information sharing:** At timestep $h \in [H]$, agents will share all the action-observation history in the private information until timestep $h-1$. Namely, $\forall h \in [H], i \in [n], c_{h^-} = \{o_{1:h-1}, a_{1:h-1}\}$ and $p_{i,h^-} = \{o_{i,h}\}$.
- **Example 2: State controlled by one controller with asymmetric delayed information sharing:** [Kaiqing: be mindful about this one.] The state dynamics and reward are controlled by only one agent (without loss of generality, agent 1), i.e., $\mathbb{T}_h(\cdot \,|\, s_h, a_{1,h}, a_{-1,h}) = \mathbb{T}_h(\cdot \,|\, s_h, a_{1,h}, a'_{-1,h}), \mathcal{R}_h(\cdot \,|\, s_h, a_{1,h}, a_{-1,h}) = \mathcal{R}_h(\cdot \,|\, s_h, a_{1,h}, a'_{-1,h})$ for all $s_h, a_{1,h}, a_{-1,h}, a'_{-1,h}$. Agent 1 will share all of her information immediately, while others will share their information with a delay of $d \geq 1$ timesteps in the baseline sharing. Namely, for any $h \in [H], i \neq 1$, $c_{h^-} = c_{(h-1)^+} \cup \{a_{1,h-1}, o_{1,h}, o_{-1,h-d}\}, p_{1,h^-} = \emptyset, p_{i,h^-} = p_{i,(h-1)^+} \cup \{o_{i,h}\} \backslash \{o_{i,h-d}\}$.
- **Example 3: Information sharing with one-directional-one-step-delay:** For convenience, we assume there are 2 agents, and this case can be generalized to multi-agent cases. In this case, agent 1 will share the information immediately, while agent 2 will share information with one-step delay, i.e. $c_{h^-} = \{o_{1:h-1}, a_{1:h-1}, o_{1,h}\}, p_{1,h^-} = \emptyset, p_{2,h^-} = \{o_{2,h}\}$.
- **Example 4: Uncontrolled state process:** The state transition does not depend on the action of agents, i.e., $\mathbb{T}_h(\cdot \,|\, s_h, a_h) = \mathbb{T}_h(\cdot \,|\, s_h, a'_h)$ for any s_h, a_h, a'_h. All agents will share their information with a delay of $d \geq 1$. For any $h \in [H], i \in [n]$, $c_{h^-} = c_{(h-1)^+} \cup \{o_{h-d}\}, p_{i,h^-} = p_{i,(h-1)^+} \cup \{o_{i,h}\} \backslash \{o_{i,h-d}\}$.
- **Example 5: One-step delayed observation sharing:** At timestep $h, h \in [H]$, each agent has access to observations of all agents until timestep $h-1$ and her present observation. Namely, $\forall h \in [H], i \in [n], c_{h^-} = \{o_{1:h-1}\}$ and $p_{i,h^-} = \{o_{i,h}\}$.
- **Example 6: One-step delayed observation and two-step delayed control sharing:** At time $h, h \in [H]$, each agent will share the observations history until timestep $h-1$ and actions history until timestep $h-2$ in the private information. Namely, $\forall h \in [H], i \in [n], c_{h^-} = \{o_{1:h-1}, a_{1:h-2}\}, p_{i,h^-} = \{o_{i,h}, a_{i,h-1}\}$.
- **Example 7: State controlled by one controller with asymmetric delayed observation sharing:** The state dynamics and reward are controlled by only one agent (, system dynamics are the same as example 2). Agent 1 will share all of her observations immediately, while others will share their observations with a delay of $d \geq 1$ timesteps in baseline sharing. Namely, for any $h \in [H], i \neq 1, c_{h^-} = c_{(h-1)^+} \cup \{o_{1,h}, o_{-1,h-d}\}, p_{1,h^-} = \emptyset, p_{i,h^-} = p_{i,(h-1)^+} \cup \{o_{i,h}\} \backslash \{o_{i,h-d}\}$.
- **Example 8: State controlled by one controller with asymmetric delayed observation and two-step delayed action sharing:** The state dynamics and reward are controlled by only one agent (, system dynamics are the same as example 2). At timestep $h, h \in [H]$, agent 1 will share all of her observations immediately and her actions history until timestep $h-2$, while others will share their observations with a delay of $d \geq 1$. Namely, for any $h \in [H], i \neq 1, c_{h^-} = c_{(h-1)^+} \cup \{o_{1,h}, a_{1,h-2}, o_{-1,h-d}\}, p_{1,h^-} = \{a_{1,h-1}\}, p_{i,h^-} = p_{i,(h-1)^+} \cup \{o_{i,h}\} \backslash \{o_{i,h-d}\}$.

In fact, the first 4 examples are all sQC LTC problems, while the rest 4 examples are QC but not sQC problems, as shown in the following lemma.

Lemma .1. Given an LTC problem $\mathcal{L}$. If the baseline sharing of $\mathcal{L}$ is one of the first 4 examples above, then $\mathcal{L}$ is sQC. If the baseline sharing of $\mathcal{L}$ is one of the last 4 examples above, then $\mathcal{L}$ is QC but not sQC.

Proof. Let $\overline{\mathcal{D}}_{\mathcal{L}}$ be the Dec-POMDP induced by $\mathcal{L}$. We prove this lemma by cases. For convenience, we use $\dot{x}$ to denote the elements in $\overline{\mathcal{D}}_{\mathcal{L}}$.

- **Example 1:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \in [n], \dot{c}_h = \{\dot{o}_{1:h-1}, \dot{a}_{1:h-1}\}$ and $\dot{p}_{i,h} = \{\dot{o}_{i,h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{a}_{1:h_1-1}, \dot{o}_{i_1,h_1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$, and $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$, and thus $\mathcal{L}$ is sQC.
- **Example 2:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \neq 1, \dot{c}_h = \{\dot{a}_{1,1:h-1}, \dot{o}_{1,1:h-1}, \dot{o}_{-1,1:h-d}\}, \dot{p}_{1,h} = \emptyset, \dot{p}_{i,h} = \{o_{i,h-d+1:h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$. If $i_1 \neq 1$, then agent (i_1, h_1) will not influence agent (i_2, h_2). If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1,1:h_1}, \dot{a}_{1,1:h_1-1}, \dot{o}_{-1,1:h_1-d}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$, and $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,\underline{h_1}}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$ if agent (i_1, h_1) influences agent (i_2, h_2), and thus $\mathcal{L}$ is sQC.
- **Example 3:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], \dot{c}_h = \{\dot{o}_{1:h-1}, \dot{a}_{1:h-1}, \dot{o}_{1,h}\}$ and $\dot{p}_{1,h} = \emptyset, \dot{p}_{2,h} = \{\dot{o}_{i,h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{a}_{1:h_1-1}, \dot{o}_{1,h_1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. If $i_1 = 2$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1}, \dot{a}_{1:h_1-1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$, and thus $\mathcal{L}$ is sQC.
- **Example 4:** Since in $\overline{\mathcal{D}}_{\mathcal{L}}$, for any $i_1, i_2 \in [n], h_1, h_2 \in [H]$, agent (i_1, h_1) does not influence agent (i_2, h_2), then $\mathcal{L}$ is sQC.

- **Example 5:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \in [n], \dot{c}_h = \{\dot{o}_{1:h-1}\}$ and $\dot{p}_{i,h} = \{\dot{o}_{i,h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{o}_{i_1,h_1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. However, agent $(1,1)$ may influence agent $(1,2)$ but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{1,2})$. Hence, $\mathcal{L}$ is QC but not sQC.
- **Example 6:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \in [n], \dot{c}_h = \{\dot{o}_{1:h-1}, \dot{a}_{1:h-2}\}$ and $\dot{p}_{i,h} = \{\dot{o}_{i,h}, \dot{a}_{i,h-1}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{a}_{1:h_1-2}, \dot{o}_{i_1,h_1}, \dot{a}_{i_1,h_1-1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$, and $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. However, agent $(1,1)$ may influence agent $(2,2)$ but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{2,2})$. Hence, $\mathcal{L}$ is QC but not sQC.
- **Example 7:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \neq 1, \dot{c}_h = \{\dot{o}_{1,1:h-1}, \dot{o}_{-1,1:h-d}\}, \dot{p}_{1,h} = \emptyset, \dot{p}_{i,h} = \{o_{i,h-d+1:h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$. If $i_1 \neq 1$, then agent (i_1, h_1) will not influence agent (i_2, h_2). If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1,1:h_1}, \dot{o}_{-1,1:h_1-d}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$ if agent (i_1, h_1) influences agent (i_2, h_2). However, agent $(1,1)$ may influence agent $(1,2)$ but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{1,2})$. Hence, $\mathcal{L}$ is QC but not sQC.
- **Example 8:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \neq 1, \dot{c}_h = \{\dot{o}_{1,1:h-1}, \dot{a}_{1,1:h-2}, \dot{o}_{-1,1:h-d}\}, \dot{p}_{1,h} = \{\dot{a}_{1,h-1}\}, \dot{p}_{i,h} = \{o_{i,h-d+1:h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$. If $i_1 \neq 1$, then agent (i_1, h_1) will not influence agent (i_2, h_2). If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1,1:h_1}, \dot{a}_{1,h_1-1}, \dot{o}_{-1,1:h_1-d}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$ if agent (i_1, h_1) influences agent (i_2, h_2). However, agent $(1,1)$ may influence agent $(2,2)$ but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{2,2})$. Hence, $\mathcal{L}$ is QC but not sQC.

This completes the proof. □

B. *Deferred Proof Details*

Lemma .2. Given any QC LTC $\mathcal{L}$, its induced Dec-POMDP $\overline{\mathcal{G}}_{\mathcal{L}}$ and any $i_1, i_2 \in [n], h_1, h_2 \in [H]$. If agent (i_1, h_1) influences agent (i_2, h_2) in $\overline{\mathcal{G}}_{\mathcal{L}}$, then for the random variables $\tau_{i_1,h_1^-}, \tau_{i_2,h_2^-}$ in $\mathcal{L}$, we have $\sigma(\tau_{i_1,h_1^-}) \subseteq \sigma(\tau_{i_2,h_2^-})$.

1) *Proof of Lemma .2:*

Proof. Consider the random variable $\overline{\tau}_{i_1,h_1}, \overline{\tau}_{i_2,h_2}$ be the information of agent $(i_1, h_1), (i_2, h_2)$ in the problem $\overline{\mathcal{G}}_{\mathcal{L}}$. From the definition, if agent (i_1, h_1) influences agent (i_2, h_2), then $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h_2})$. Since for any $h \in [H], i \in [n], \overline{\tau}_{i,h}$ is the information of agent (i, h) without additional sharing. Then, $\tau_{i,h^-} \backslash \overline{\tau}_{i,h} \subseteq \cup_{t=1}^{h-1} z_t^a, \tau_{i,h^+} \backslash \overline{\tau}_{i,h} \subseteq \cup_{t=1}^{h} z_t^a$. Therefore, we know that $\sigma(\tau_{i_1,h_1^-} \backslash \overline{\tau}_{i_1,h_1}) \subseteq \sigma(\cup_{t=1}^{h-1} z_t^a) \subseteq \sigma(c_{h_1^-}) \subseteq \sigma(c_{h_2^-}) \subseteq \sigma(\tau_{i_2,h_2^-})$. Also, we know $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h_2}) \subseteq \sigma(\tau_{i_2,h_2^-})$. We can conclude $\sigma(\tau_{i_1,h_1^-}) \subseteq \sigma(\tau_{i_2,h_2^-})$. □

2) *Proof of Lemma III.2:*

Proof. We first have the following proposition on the hardness of solving POMDPs.

Proposition .3. There exists an $\epsilon > 0$, such that computing an ϵ-additive optimal strategy in POMDPs is `PSPACE-hard`.

One can adapt the proof of [19, Theorem 4.11], which proved the `PSPACE-hardness` of computing an ϵ-*relative* optimal strategy in POMDPs, to obtain such a result for an ϵ-*additive* one. In particular, any ϵ-additive optimal strategy in the POMDP constructed in the proof of Theorem 4.11 therein is also an ϵ-relative optimal strategy.

Now we proceed with the proof of Lemma III.2. Given any POMDP $\mathcal{P} = (\mathcal{S}^{\mathcal{P}}, \mathcal{A}^{\mathcal{P}}, \mathcal{O}^{\mathcal{P}}, \{\mathbb{O}_h^{\mathcal{P}}\}_{h\in[H^{\mathcal{P}}]}, \{\mathbb{T}_h^{\mathcal{P}}\}_{h\in[H^{\mathcal{P}}]}, \{\mathcal{R}_h^{\mathcal{P}}\}_{h\in[H^{\mathcal{P}}]}, \mu_1^{\mathcal{P}})$, we can construct an LTC $\mathcal{L}$ as follows:

- Number of agents: $n = 3$; length of episode: $H = 2H^{\mathcal{P}}$.
- Underlying state space: $\mathcal{S} = \mathcal{S}^{\mathcal{P}} \times [2]$. For any $s \in \mathcal{S}$, we can split $s = (s^1, s^2)$, where $s^1 \in \mathcal{S}^{\mathcal{P}}, s^2 \in [2]$. Intial state distribution: $\forall s \in \mathcal{S}, \mu_1(s) = \mu_1^{\mathcal{P}}(s^1)/2$.
- Control action space: For any $h \in [H], \mathcal{A}_{1,h} = \mathcal{A}^{\mathcal{P}}, \mathcal{A}_{2,h} = [2], \mathcal{A}_{3,h} = \{\emptyset\}$.
- Transition functions: For any $h \in [H-1], s_h, s_{h+1} \in \mathcal{S}, a_h \in \mathcal{A}_h$, if $h = 2t-1$ with $t \in [H^{\mathcal{P}}], \mathbb{T}_h(s_{h+1} \,|\, s_h, a_h) = \mathbb{T}_t^{\mathcal{P}}(s_{h+1}^1 \,|\, s_h^1, a_{1,h})\mathbb{1}[s_{h+1}^2 = s_h^2]$; if $h = 2t$ with $t \in [H-1], \mathbb{T}_h(s_{h+1} \,|\, s_h, a_h) = \mathbb{1}[s_{h+1}^1 = s_h^1, s_{h+1}^2 = a_{2,h}]$.
- Observation space: For any $h \in [H]$, if $h = 2t-1$ with $t \in [H^{\mathcal{P}}], \mathcal{O}_{1,h} = \mathcal{O}_t^{\mathcal{P}}, \mathcal{O}_{2,h} = \mathcal{O}_{3,h} = \mathcal{S}$; if $h = 2t$ with $t \in [H^{\mathcal{P}}]$, $\mathcal{O}_{1,h} = [2], \mathcal{O}_{2,h} = \mathcal{O}_{3,h} = \mathcal{S}$.
- Emission matrix: For any $h \in [H]$, if $h = 2t-1$ with $t \in [H^{\mathcal{P}}], \forall o_h \in \mathcal{O}_h, s_h \in \mathcal{S}, \mathbb{O}_h(o_h \,|\, s_h) = \mathbb{O}_h^{\mathcal{P}}(o_{1,h} \,|\, s_h^1)\mathbb{1}[o_{2,h} = o_{3,h} = s_h]$; if $h = 2t$ with $t \in [H^{\mathcal{P}}], \forall o_h \in \mathcal{O}_h, s_h \in \mathcal{S}, \mathbb{O}_h(o_h \,|\, s_h) = \mathbb{O}_h^{\mathcal{P}}(o_{1,h} \,|\, s_h^2)\mathbb{1}[o_{2,h} = o_{3,h} = s_h]$
- The baseline sharing: null.
- The communication action space: For any $h \in [H], i \in [3], \mathcal{M}_{i,h} = \{0,1\}^h$. For any $i \in [3], p_{i,h^-} \in \mathcal{P}_{i,h^-}, \phi_{i,h}(p_{i,h^-}, m_{i,h}) = \{o_{i,k} \,|\, k\text{-th digit of } p_{i,h^-} \text{ is 1 and } o_{i,k} \in p_{i,h^-}\} \cup \{m_{i,h}\}$.
- Reward function: For any $h \in [H], i \in [3], s_h \in \mathcal{S}, a_h \in \mathcal{A}_h$, if $h = 2t-1$ with $t \in [H^{\mathcal{P}}], \mathcal{R}_h(s_h, a_h) = \mathcal{R}_t^{\mathcal{P}}(s_h^1, a_{1,h})/H$; if $h = 2t$ with $t \in [H^{\mathcal{P}}], \mathcal{R}_h(s_h, a_h) = \mathbb{1}[a_{2,h} = 1]$.
- Communication cost function: For any $h \in [H], z_h^a \in \mathcal{Z}_h^a, \mathcal{K}_h(z_h^a) = \mathbb{1}[z_h^a \neq \{m_h\}]$. It means the communication cost is 1 until there is no additional sharing.

- We restrict communication strategy can only use c_h as input, and remove $a_{3,t}$ in τ_h for any $h > t$.

We first verify that such a construction satisfies Assumptions III.1, III.4, III.5, III.7, and IV.7.

- $\mathcal{L}$ satisfies Assumption III.1, III.7 because agent 2 and agent 3 has individual γ-observability. That is, for any $b_1, b_2 \in \Delta(\mathcal{S}), i = 2, 3$, we have

$$\begin{aligned} ||\mathbb{O}_{i,h}^\top(b_1 - b_2)||_1 &= \sum_{o_{i,h} \in \mathcal{O}_h} |\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))\mathbb{O}_{i,h}(o_{i,h} \,|\, s_h)| \\ &= \sum_{o_{i,h} \in \mathcal{O}_h} |\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))\mathbb{1}[o_{i,h} = s_h]| \\ &= \sum_{o_{i,h} \in \mathcal{O}_h} (b_1(o_{i,h}) - b_2(o_{i,h}))| = ||b_1 - b_2||_1. \end{aligned}$$

- $\mathcal{L}$ satisfies Assumption III.4 because we restrict communication strategy can only use $\widehat{c}_h$ as input.
- $\mathcal{L}$ satisfies Assumption III.5 since only $a_{3,h}, h \in [H-1]$ do not influence underlying state, and $\mathcal{A}_{3,h} = \{\emptyset\}$.
- $\mathcal{L}$ satisfies Assumption IV.7 since it satisfies the **Turn-based structures** condition in §E, with $ct(2t-1) = 1, ct(2t) = 2$ for any $t \in [H^{\mathcal{P}}]$.

In LTC problem $\mathcal{L}$, agent 2 will always choose $a_{i,2t} = 1$ at even steps to obtain $r_{2h} = 1$. And there will be no additional sharing since any additional sharing at timestep h will suffer the communication cost $\kappa_h = 1 > \max \sum_{t=1}^{H^{\mathcal{P}}} \mathcal{R}_{2t-1}(s_{2t-1}, a_{2t-1})$, and it cannot achieve optimum. Therefore, state $s_h^2, h \in [H]$ are dummy state, and agent $2, 3$ are dummy agents. Then, any $(g_{1:H}^{a,*}, g_{1:H}^{m,*})$ be an ϵ/H-team optimal strategy of $\mathcal{L}$ will directly gives the ϵ-optimal of $\mathcal{P}$ as $\{g_{1,2t-1}^{a,*}\}_{h \in [H^{\mathcal{P}}]}$. From the Proposition .3, we can complete the proof. □

3) Proof of Lemma III.3:

Proof. We prove this result by showing a reduction from the Team Decision problem [28].

Definition .4 (Team decision problem (TDP))**.** Given finite sets Y_1, Y_2, U_1, U_2, a rational probability mass function $p : Y_1 \times Y_2 \to \mathbb{Q}$, and an integer cost function $c : Y_1 \times Y_2 \times U_1 \times U_2 \to \mathbb{N}$, find decision rules $\gamma_i : Y_i \to U_i, i = 1, 2$ that minimize the expected cost

$$J(\gamma_1, \gamma_2) = \sum_{y_1 \in Y_1, y_2 \in Y_2} c(y_1, y_2, \gamma_1(y_1), \gamma_2(y_2))p(y_1, y_2). \tag{.1}$$

We show the `NP-hardness` of solving LTC from problem TDP. Given any TDP $\mathcal{TD} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{U}_1, \widetilde{U}_2, \widetilde{c}, \widetilde{p}, \overline{J})$ with $|\widetilde{U}_1| = |\widetilde{U}_2| = 2$, let $\widetilde{U}_1 = \{1, 2\}, \widetilde{U}_2 = \{1, 2\}$, then we can construct an $H = 4$ and 2 agents LTC $\mathcal{L}$ with two parameters $n_1 \in \mathbb{N}, \alpha_1 \in \mathbb{R}, \alpha_2 \in (0, 1)$ (to be specified later) such that:

- Number of agents: $n = 2$; length of episode: $H = 4$.
- Underlying state: $\mathcal{S} = [2]^4$. For each $s_1 \in \mathcal{S}$, we can split s_1 into 4 parts as $s_1 = (s_1^1, s_1^2, s_1^3, s_1^4)$, where $s_1^1, s_1^2, s_1^3, s_1^4 \in [2]$. Similarly, $s_2, s_3, s_4 \in \mathcal{S}$ can be splitted in the same way.
- Initial state distribution: $\forall s_1 \in \mathcal{S}, \mu_1(s_1) = \frac{1}{16}$.
- Control action space: For the first 2 timesteps, $\forall i = 1, 2, \mathcal{A}_{i,1} = \mathcal{A}_{i,2} = \{\emptyset\}$; for $h = 3, \mathcal{A}_{1,3} = [2], \mathcal{A}_{2,3} = \{\emptyset\}$; for $h = 4, \mathcal{A}_{2,4} = [2], \mathcal{A}_{1,4} = \{\emptyset\}$.
- Transition: $\forall s \in \mathcal{S}, a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2, a_3 \in \mathcal{A}_3, \mathbb{T}_1(s \,|\, s, a_1) = \mathbb{T}_2(s \,|\, s, a_2) = \mathbb{T}_3(s \,|\, s, a_3) = 1$. Note that under the transition dynamics above, $s_1 = s_2 = s_3 = s_4$ always holds, for any $s_1 \in \mathcal{S}$.
- Observation space: $\mathcal{O}_{1,1} = \mathcal{O}_{2,1} = \mathcal{O}_{1,2} = \mathcal{O}_{2,2} = [2] \times \mathcal{S}$, $\mathcal{O}_{1,3} = \widetilde{Y_1} \times \mathcal{S}, \mathcal{O}_{2,3} = \widetilde{Y_2} \times \mathcal{S}, \mathcal{O}_{1,4} = \mathcal{O}_{2,4} = \mathcal{S}$; For each $i \in [2], h \in [2], o_{i,h} \in \mathcal{O}_{i,h}$, we can split $o_{i,h}$ into 2 parts as $o_{i,h} = (o_{i,h}^1, o_{i,h}^2)$, where $o_{i,h}^1 \in [2], o_{i,h}^2 \in \mathcal{S}$. For each $i \in [n], o_{i,3} \in \mathcal{O}_{i,3}$, similarly, we can split $o_{i,3}$ into 2 parts as $o_{i,3} = (o_{i,3}^1, o_{i,3}^2)$, where $o_{i,3}^1 \in \widetilde{Y}_i, o_{i,3}^2 \in \mathcal{S}$.
- The baseline sharing is null.
- Communication action space: For $i \in [2], h \in \{1, 2, 4\}, \mathcal{M}_{i,h} = \{0, 1\}^h, \mathcal{M}_{i,3} = \{1, 2\}$; For each $i \in [2], \phi_{i,h}$ is defined as $\forall h \in \{1, 2, 4\}, \phi_{i,h}(p_{i,h^-}, m_{i,h}) = \{o_{i,k} \in p_{i,h^-} \,|\, k \leq h, k^{\text{th}}$ digit of $m_{i,h}$ is $1\}$; For $h = 3, \phi_{i,h}(p_{i,h^-}, 1) = \{o_{i,1}, o_{i,3}, m_{i,h}\}, \phi_{i,h}(p_{i,h^-}, 2) = \{o_{i,2}, o_{i,3}, m_{i,h}\}$.
- Emission matrix: For any $i \in [2], h \in [2], s_h \in \mathcal{S}, o_{i,h} \in \mathcal{O}_{i,h}, \mathbb{O}_h(o_h \,|\, s_h) = \Pi_{i=1}^2 \mathbb{O}_{i,h}(o_{i,h} \,|\, s_h)$ and $\mathbb{O}_{i,h}(o_{i,h} \,|\, s_h)$ is defined as:

$$\mathbb{O}_{i,h}(o_{i,h} \,|\, s_h) = \begin{cases} \frac{1-\alpha_2}{16} & o_{i,h}^1 = s_h^{i+2h-2}, o_{i,h}^2 \neq s_h \\ \frac{1-\alpha_2}{16} + \alpha_2 & o_{i,h}^1 = s_h^{i+2h-2}, o_{i,h}^2 = s_h \\ 0 & \text{o.w.} \end{cases}.$$

For $i \in [2], s_3 \in \mathcal{S}, o_3 \in \mathcal{O}_3, \mathbb{O}_3(o_3 \,|\, s_3) = \mathbb{O}_3^1(o_3^1 \,|\, s_3)\mathbb{O}_3^2(o_3^2 \,|\, s_3), \mathbb{O}_3^2 = \Pi_{i=1}^2 \mathbb{O}_{i,3}^2(o_{i,3}^2 \,|\, s_3)$ is defined as:

$$\mathbb{O}_3^1(o_3^1 \,|\, s_3) = \widetilde{p}(o_{1,3}^1, o_{2,3}^1)$$

$$\mathbb{O}_{i,3}^2(o_3^2 \,|\, s_3) = \begin{cases} \frac{1-\alpha_2}{16} & o_{i,3}^2 \neq s_3 \\ \frac{1-\alpha_2}{16} + \alpha_2 & o_{i,3}^2 = s_3 \end{cases}.$$

And for $i \in [2], s_4 \in \mathcal{S}, o_{i,4} \in \mathcal{O}_{i,4}, \mathbb{O}_4(o_4 \,|\, s_h) = \Pi_{i=1}^2 \mathbb{O}_{i,4}(o_{i,4} \,|\, s_4)$ and $\mathbb{O}_{i,4}(o_{i,4} \,|\, s_4)$ is defined as:

$$\mathbb{O}_{i,4}(o_{i,4} \,|\, s_4) = \begin{cases} \frac{1-\alpha_2}{16} & o_{i,4} \neq s_4 \\ \frac{1-\alpha_2}{16} + \alpha_2 & o_{i,4} = s_4 \end{cases}.$$

Such an emission matrix means that for each $h \in [2]$ and $i \in [2]$, agent i will accurately observe part of the underlying state s_h^{i+2h-2} and vaguely observe the whole underlying state s_h. And for $h = 4, i \in [2]$, agent i can only vaguely observe the whole underlying state s_h. The reward functions are defined as:

$$\mathcal{R}_1(s_1, a_1) = \mathcal{R}_2(s_2, a_2) = 0, \quad \forall s_1, s_2 \in \mathcal{S}, a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2;$$

$$\mathcal{R}_3(s_3, a_3) = \begin{cases} 1 & \text{if } a_{1,3} = s_3^2 \text{ or } a_{1,3} = s_3^4 \\ 0 & \text{o.w.} \end{cases};$$

$$\mathcal{R}_4(s_4, a_4) = \begin{cases} 1 & \text{if } a_{2,4} = s_4^1 \text{ or } a_{2,4} = s_4^3 \\ 0 & \text{o.w.} \end{cases}.$$

The communication cost functions are defined as:

$$\forall h \in \{1, 2, 4\}, z_h^a \in \mathcal{Z}_h^a, \mathcal{K}_h(z_h^a) = 1 \text{ if } z_h^a \neq \{m_{1,h}, m_{2,h}\} \text{ else } 0;$$

$$\mathcal{K}_3(z_3^a) = \begin{cases} \widetilde{c}(o_{1,3}^1, o_{2,3}^1, 1, 1)/\alpha_1 & \text{if } \{o_{1,1}, o_{2,1}\} \subseteq z_3^a \text{ and } \{o_{1,2}, o_{2,2}\} \cap z_3^a = \emptyset \\ \widetilde{c}(o_{1,3}^1, o_{2,3}^1, 2, 1)/\alpha_1 & \text{if } \{o_{1,2}, o_{2,1}\} \subseteq z_3^a \text{ and } \{o_{1,1}, o_{2,2}\} \cap z_3^a = \emptyset \\ \widetilde{c}(o_{1,3}^1, o_{2,3}^1, 1, 2)/\alpha_1 & \text{if } \{o_{1,1}, o_{2,2}\} \subseteq z_3^a \text{ and } \{o_{1,2}, o_{2,1}\} \cap z_3^a = \emptyset \\ \widetilde{c}(o_{1,3}^1, o_{2,3}^1, 2, 2)/\alpha_1 & \text{if } \{o_{1,2}, o_{2,2}\} \subseteq z_3^a \text{ and } \{o_{1,1}, o_{2,1}\} \cap z_3^a = \emptyset \end{cases}.$$

Let $\alpha_0 = \max_{y_1, y_2, u_1, u_2} \widetilde{c}(y_1, y_2, u_1, u_2)$, and set $\alpha_1 = 2\alpha_0$. Under such a construction, $\mathcal{L}$ satisfies the following conditions:

- Problem $\mathcal{L}$ is QC: For $\forall i_1, i_2 \in [2], h_1, h_2 \in [4]$, agent (i_1, h_1) does not influence (i_2, h_2) because agent (i_1, h_1) cannot influence the observation of agent (i_2, h_2), and baseline sharing is null.
- Problem $\mathcal{L}$ satisfies Assumptions III.1 and III.7: We prove this by showing that each agent $i \in [2]$ satisfies γ-observability. For $\forall i \in [2], h \in [2], b_1, b_2 \in \Delta(\mathcal{S})$, let

$$\begin{aligned}
||\mathbb{O}_{i,h}^\top(b_1 - b_2)||_1 &= \sum_{o_{i,h}^1 \in [2]} \sum_{o_{i,h}^2 \in \mathcal{S}} |\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))\mathbb{O}_{i,h}((o_{i,h}^1, o_{i,h}^2) \,|\, s_h)| \\
&\geq \sum_{o_{i,h}^2 \in \mathcal{S}} |\sum_{o_{i,h}^1 \in [2]} \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))\mathbb{O}_{i,h}((o_{i,h}^1, o_{i,h}^2) \,|\, s_h)| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} |\sum_{s_h \in \mathcal{S}} \sum_{o_{i,h}^1 \in [2]} (b_1(s_h) - b_2(s_h))\mathbb{1}[o_{i,h}^1 = s_h^{i+2h-2}](\frac{1-\alpha_2}{16} + \alpha_2 \mathbb{1}[o_{i,h}^2 = s_h])| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} |\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))(\frac{1-\alpha_2}{16} + \alpha_2 \mathbb{1}[o_{i,h}^2 = s_h])| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} |\frac{1-\alpha_2}{16}(\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))) + \alpha_2(b_1(o_{i,h}^2) - b_2(o_{i,h}^2))| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \alpha_2 |b_1(o_{i,h}^2) - b_2(o_{i,h}^2)| = \alpha_2 ||b_1 - b_2||_1.
\end{aligned}$$

  For $\forall i \in [2], h = 3, 4$, the proof is similar, by replacing $o_{i,h}^1 \in [2]$ with $o_{i,h}^1 \in \widetilde{Y}_i$ for $h = 3$ and replacing the space $o_{i,h}^1 \in [2]$ with $\emptyset$ for $h = 4$.
- Problem $\mathcal{L}$ satisfies Assumption III.5, because control actions $a_{1:4}$ does not influence underlying states and we restrict the communication and control strategies do not use them as input.
- Problem $\mathcal{L}$ satisfies Assumption IV.7 since it satisfies the **Turn-based structures** condition in §E, with $ct(1) = ct(2) = ct(3) = 1, ct(4) = 2$.

We will show as follows that computing a team-optimal strategy can give us a team-optimal strategy in $\mathcal{TD}$. Given $(g_{1:4}^{a,*}, g_{1:4}^{m,*})$ as team optimal strategy of $\mathcal{L}$, firstly it will have no additional sharing at timesteps $h = 1, 2, 4$, namely, for $h = 1, 2, 4, \mathbb{P}(z_h^a \neq \{m_{1,h}, m_{2,h}\} \,|\, g_{1:4}^{a,*}, g_{1:4}^{m,*}) = 1$, since any additional sharing at timesteps $h = 1, 2, 4$ will incur the cost as high as 1 and cannot achieve optimum. Also, for the additional sharing at timestep $h = 3$, agent i will definitely share $o_{i,3}$ and choose to share $o_{i,1}$ or $o_{i,2}$. Then $\forall \tau_{1,3^+} \in \mathcal{T}_{1,3^+}, g_{1,3}^{a,*}(\tau_{1,3^+}) = \begin{cases} o_{2,1} & \text{if } o_{2,1} \in \tau_{1,3^+} \\ o_{2,2} & \text{if } o_{2,2} \in \tau_{1,3^+} \end{cases}$ and $\forall \tau_{2,4^+} \in \mathcal{T}_{2,4^+}, g_{2,4}^{a,*}(\tau_{2,4^+}) = \begin{cases} o_{1,1} & \text{if } o_{1,1} \in \tau_{2,4^+} \\ o_{1,2} & \text{if } o_{1,2} \in \tau_{2,4^+} \end{cases}$, since such action can achieve the optimal reward $r_3 = r_4 = 1$. Therefore, $J_{\mathcal{L}}(g_{1:H}^{a,*}, g_{1:H}^{m,*}) = \mathbb{E}[\sum_{h=1}^4 r_h - \kappa_h \,|\, g_{1:H}^{a,*}, g_{1:H}^{m,*}] = 2 - \mathbb{E}[\kappa_3 \,|\, g_{1:H}^{a,*}, g_{1:H}^{m,*}] = 2 - \mathbb{E}[\widetilde{c}(o_{1,3}^1, o_{2,3}^1, m_{1,3}, m_{2,3})]$, where $m_{1,3} = g_{1,3}^{m,*}(\{o_{1,1}, o_{1,2}, o_{1,3}\})$. Since κ_3 is independent of $o_{1,1}, o_{1,2}, o_{1,3}^1$, $o_{1,1}, o_{1,2}, o_{1,3}^1$ are useless information for agent 1 to choose $m_{1,3}$ and minimize the κ. Therefore, not using them in $g_{1,3}^{m,*}$ does not lose any optimality. Hence, we can consider the $g_{1,3}^{m,*}$ that only has $o_{1,3}^1$ as input. In the same way, we consider the $g_{2,3}^{m,*}$ that has $o_{2,3}^1$ as input. Therefore, $J_{\mathcal{L}}(g_{1:H}^{a,*}, g_{1:H}^{m,*}) = 2 - \sum_{o_{1,3}^1, o_{1,3}^2, m_{1,3}, m_{2,3}} \frac{\widetilde{c}(o_{1,3}^1, o_{2,3}^1, m_{1,3}, m_{2,3}}{\alpha_1} g_{1,3}^{m,*}(m_{1,3} \,|\, o_{1,3}^1) g_{2,3}^{m,*}(m_{2,3} \,|\, o_{2,3}^1) \widetilde{p}(o_{1,3}^1, o_{2,3}^1)$. Then we can construct $\gamma_1 = g_{1,3}^{m,*}, \gamma_2 = g_{2,3}^{m,*}, \gamma_1 : \widetilde{Y}_1 \to \Delta(\widetilde{U}_1), \gamma_2 : \widetilde{Y}_2 \to \Delta(\widetilde{U}_2)$.
We extend $\widetilde{J}$ for stochastic decision rules, that is

$$\forall \widehat{\gamma}_1 : \widetilde{Y_1} \to \Delta(\widetilde{U}_1), \widehat{\gamma}_2 : \widetilde{Y}_2 \to \Delta(\widetilde{U}_2), \widetilde{J}(\widehat{\gamma}_1, \widehat{\gamma}_2) = \sum_{y_1, y_2, u_1, u_2} \widetilde{c}(y_1, y_2, u_1, u_2) \widehat{\gamma}_1(u_1 \,|\, y_1) \widehat{\gamma}_2(u_2 \,|\, y_2) p(y_1, y_2).$$

From the optimality of $g_{1,3}^{m,*}, g_{2,3}^{m,*}$, we know that γ_1, γ_2 minimizes the $\widetilde{J}$, namely, $\forall \widehat{\gamma}_1, \widehat{\gamma}_2, \widetilde{J}(\widehat{\gamma}_1, \widehat{\gamma}_2) \geq \widetilde{J}(\gamma_1, \gamma_2)$. Then, defining $\overline{\gamma}_1$ as

$$\forall y_1 \in Y_1, \overline{\gamma}_1(y_1) \in \operatorname*{argmin}_{u \in U_1} \sum_{y_2 \in Y_2, u_2 \in U_2} \widetilde{c}(y_1, y_2, u, u_2) \gamma_2(u_2 \,|\, y_2) \widetilde{p}(y_1, y_2),$$

we know that

$$\begin{aligned} \widetilde{J}(\gamma_1, \gamma_2) = & \sum_{y_1, y_2, u_1, u_2} \widetilde{c}(y_1, y_2, u_1, u_2) \gamma_1(u_1 \,|\, y_1) \gamma_2(u_2 \,|\, y_2) \widetilde{p}(y_1, y_2) \\ = & \sum_{y_1, u_1} \gamma_1(u_1 \,|\, y_1) \sum_{y_2, u_2} \widetilde{c}(y_1, y_2, u_1, u_2) \gamma_2(u_2 \,|\, y_2) \widetilde{p}(y_1, y_2) \\ \geq & \sum_{y_1} \min_{u_1} \sum_{y_2, u_2} \widetilde{c}(y_1, y_2, u_1, u_2) \gamma_2(u_2 \,|\, y_2) \widetilde{p}(y_1, y_2) \\ = & \sum_{y_1, y_2, u_2} \widetilde{c}(y_1, y_2, \overline{\gamma}_1(y_1), u_2) \gamma_2(u_2 \,|\, y_2) \widetilde{p}(y_1, y_2) \\ = & \widetilde{J}(\overline{\gamma}_1, \gamma_2). \end{aligned}$$

Similarly, define $\overline{\gamma}_2$ as

$$\forall y_2 \in Y_2, \overline{\gamma}_2(y_2) = \operatorname*{argmin}_{u \in U_2} \sum_{y_1} \widetilde{c}(y_1, y_2, \overline{\gamma}_1(y_1), u_2) \widetilde{p}(y_1, y_2),$$

and we can obtain $\widetilde{J}(\overline{\gamma}_1, \overline{\gamma}_2) \leq \widetilde{J}(\overline{\gamma}_1, \gamma_2)$. Then, we obtain the decision rules $\overline{\gamma}_1, \overline{\gamma}_2$ which minimize the $\overline{J}$ for the original TDP problem $\mathcal{TD}$. □

4) Proof of Lemma III.8:

Proof. We prove this by showing a reduction from the hardness of finding ϵ-optimal strategy in POMDP. Given any POMDP $\mathcal{P} = (\mathcal{S}^{\mathcal{P}}, \mathcal{A}^{\mathcal{P}}, \mathcal{O}^{\mathcal{P}}, \{\mathbb{O}_h^{\mathcal{P}}\}_{h \in [H^{\mathcal{P}}]}, \{\mathbb{T}_h^{\mathcal{P}}\}_{h \in [H^{\mathcal{P}}]}, \{\mathcal{R}_h^{\mathcal{P}}\}_{h \in [H^{\mathcal{P}}]}, \mu_1^{\mathcal{P}})$, we can construct a LTC $\mathcal{L}$ with 2 agents as follows:

- Number of agents: $n = 2$; length of episode: $H = H^{\mathcal{P}}$.
- $\mathcal{S} = \mathcal{S}^{\mathcal{P}}, \forall s \in \mathcal{S}$.
- Initial state distribution: $\forall s_1 \in \mathcal{S}, \mu_1(s_1) = \mu_1^{\mathcal{P}}(s_1)$.
- Control action space: $\forall h \in [H]$, $\mathcal{A}_{1,h} = \mathcal{A}_h^{\mathcal{P}}, \mathcal{A}_{2,h} = \{\emptyset\}$.
- Transition: $\forall s_h, s_{h+1} \in \mathcal{S}, a_h \in \mathcal{A}_h, \mathbb{T}_h(s_{h+1} \,|\, s_h, a_h) = \mathbb{T}_h^{\mathcal{P}}(s_{h+1} \,|\, s_h, a_{1,h})$.
- Observation space: $\forall h \in [H], \mathcal{O}_{1,h} = \mathcal{O}^{\mathcal{P}}, \mathcal{O}_{2,h} = \mathcal{S}$.
- Emission matrix: For any $h \in [H], \forall o_h \in \mathcal{O}_h, s_h \in \mathcal{S}$, $\mathbb{O}_h(o_h \,|\, s_h) = \mathbb{O}_h^{\mathcal{P}}(o_{1,h} \,|\, s_h) \mathbb{1}[o_{2,h} = s_h]$.
- Reward functions: For any $h \in [H], i \in [2], s_h \in \mathcal{S}, a_h \in \mathcal{A}_h, \mathcal{R}_h(s_h, a_h) = \mathcal{R}^{\mathcal{P}}(s_h, a_{1,h})/H$.
- The baseline sharing: For any $h \in [H]$, $z_h^b = \{o_{1,h}, a_{1,h-1}\}$.

- Communication action space: For any $h \in [H], \mathcal{M}_{1,h} = \{\emptyset\}, \mathcal{M}_{2,h} = \{0,1\}^h$. For any $p_{1,h^-} \in \mathcal{P}_{1,h^-}, p_{2,h^-} \in \mathcal{P}_{2,h^-}, m_h \in \mathcal{M}_h, \phi_{1,h}(p_{1,h^-}, m_{1,h}) = \{m_{1,h}\}, \phi_{2,h}(p_{2,h^-}, m_{2,h}) = \{o_{2,k} \,|\, k\text{-th digit of } p_{2,h^-} \text{ is 1 and } o_{2,k} \in p_{i,h^-}\} \cup \{m_{2,h}\}$.
- Communication cost functions: For any $h \in [H], z_h^a \in \mathcal{Z}_h^a, \mathcal{K}_h(z_h^a) = \mathbb{1}[z_h^a \neq \{m_h\}]$. It means the communication cost is 1 until there is no additional sharing.
- We restrict communication strategy can only use c_h as input, and remove $a_{2,t}$ in τ_h for any $h > t$.

We first verify that $\mathcal{L}$ is QC and satisfies Assumptions III.1, III.4, III.5, and IV.7.

- $\mathcal{L}$ is QC: For any $\forall h_1 < h_2 \leq H$, agent $(2, h_1)$ does not influence agent $(1, h_2)$ under baseline sharing since agent $(2, h_1)$ does not influence $s_h^1, \forall h \in [H]$, then does not influence $o_{1,h}, \forall h \in [H]$, and thus not influencing agent $(1, h_1)$. For any $\forall h_1 < h_2 \leq H$, under baseline sharing, $p_{1,h^-} = \emptyset$. Then $\sigma(\tau_{1,h_1^-}) \subseteq \sigma(c_{h_1^-}) \subseteq \sigma(c_{h_2^-}) \subseteq \sigma(\tau_{2,h_2^-})$.
- $\mathcal{L}$ satisfies Assumption III.1: For any $h \in [H], b_1, b_2 \in \Delta(\mathcal{S})$, $\mathbb{O}_h$ satisfies

$$\begin{aligned}
||\mathbb{O}_h^\top (b_1 - b_2)||_1 &= \sum_{o_{1,h} \in \mathcal{O}^{\mathcal{P}}} \sum_{o_{2,h} \in \mathcal{S}} \Big| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_h((o_{1,h}, o_{2,h}) \,|\, s_h) \Big| \\
&\geq \sum_{o_{2,h} \in \mathcal{S}} \Big| \sum_{o_{1,h} \in \mathcal{O}^{\mathcal{P}}} \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{1,h}(o_{1,h} \,|\, s_h) \mathbb{O}_{2,h}(o_{2,h} \,|\, s_h) \Big| \\
&= \sum_{o_{2,h} \in \mathcal{S}} \Big| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{2,h}(o_{2,h} \,|\, s_h) \sum_{o_{1,h} \in \mathcal{O}^{\mathcal{P}}} \mathbb{O}_{1,h}(o_{1,h} \,|\, s_h) \Big| \\
&= \sum_{o_{2,h} \in \mathcal{S}} \Big| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{1}[o_{2,h} = s_h] \\
&= \sum_{o_{2,h} \in \mathcal{S}} |b_1(o_{2,h}) - b_2(o_{2,h})| = ||b_1 - b_2||_1.
\end{aligned}$$

- $\mathcal{L}$ satisfies Assumption III.4: For any $h \in [H]$, we restrict that each agent decides $m_{i,h}$ based on c_h.
- $\mathcal{L}$ satisfies Assumption III.5: For any $h \in [H], a_{2,h}$ does not influence s_{h+1}, and it is removed from τ.
- $\mathcal{L}$ satisfies Assumption IV.7 since it satisfies the **Turn-based structures** condition in §E, with $ct(h) = 1$ for any $h \in [H]$.

Agent 2 will share nothing through additional sharing, otherwise it will suffer the communication cost $\kappa_h = 1 > \max \sum_{h=1}^{H} \mathcal{R}_h(s_h, a_h)$ and cannot achieve optimum. Hence, Agent 2 is the dummy player. Therefore, any $(g_{1:H}^{a,*}, g_{1:H}^{m,*})$ be an ϵ/H-team optimal strategy of $\mathcal{L}$ will directly gives the ϵ-optimal of $\mathcal{P}$ as $\{g_{1,1:H}^{a,*}\}_{h \in [H]}$. From Proposition .3, we can complete our proof. □

5) *Proof of Theorem IV.2:*

Proof. We prove the following lemma first.

Lemma .5. Let the $\mathcal{G}$ be the QC LTC problem satisfying Assumptions III.5, III.7, and III.9, and $\mathcal{D}_{\mathcal{L}}$ be the reformulated Dec-POMDP. Then for $i_1, i_2 \in [n], t_1, t_2 \in [H]$, if agent $(i_1, 2t_1)$ influences agent $(i_2, 2t_2)$ in $\mathcal{D}_{\mathcal{L}}$, then $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$ in $\mathcal{L}$. Moreover, if $\mathcal{L}$ is sQC, then $\sigma(a_{i_1,t_1}) \subseteq \sigma(\tau_{i_2,t_2^-})$.

Proof. If $i_1 = i_2$, then $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_1,t_2^-})$ under perfect recall in the LTC $\mathcal{L}$. So in the following part, we assume $i_1 \neq i_2$.

- If a_{i_1,t_1} influences the underlying state s_{t_1+1}, then from Assumption III.7, agent (i_1, t_1) influences o_{-i_1,t_1+1}, so there must exist $i_3 \neq i_1$, such that agent (i_1, t_1) influences o_{i_3,t_1+1}. From part (e) of Assumption II.1 and $t_1 < t_2$, we know $o_{i_3,t_1+1} \in \tau_{i_3,t_2^-}$ even under no additional sharing, and then we get agent (i_1, t_1) influences agent (i_3, t_2) under no additional sharing, hence $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_3,t_2^-})$. From Assumption III.9 and $i_3 \neq i_1$, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(c_{t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$. (Similarly, if $\mathcal{L}$ is sQC, we have $\sigma(a_{i_1,t_1}) \subseteq \sigma(\tau_{i_3,t_2^-})$ from Assumption III.9, and $\sigma(a_{i_1,t_1}) \subseteq \sigma(c_{t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$ from Assumption III.9).
- If a_{i_1,t_1} does not influence s_{t_1+1}, from Assumption III.5, $\forall t > t_1, a_{i_1,t_1} \notin \tau_{t^-}$ and $a_{i_1,t_1} \notin \tau_{t^+}$. If exists $t_3 \in [H], t_1 < t_3 \leq t_2$ such that a_{i_1,t_1} influences baseline sharing at t_3, then a_{i_1,t_1} influences $z_{t_3}^b$. From $z_{t_3}^b \subseteq \tau_{i_2,t_2^-}$, we know agent (i_1, t_1) influences agent (i_2, t_2) under $\mathcal{L}$ without additional sharing, then from definition of QC of $\mathcal{L}$, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(c_{t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$ (and $\sigma(a_{i_1,t_1}) \subseteq \sigma(c_{t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$ if $\mathcal{L}$ is sQC).
- If a_{i_1,t_1} does not influence baseline sharing between t_1 and t_2. Then in $\mathcal{D}_{\mathcal{L}}$, agent $(i_1, 2t_1)$ does not influence $\widetilde{\tau}_{i,2t_1+1}, \forall i \in [n]$, hence it does not influence $\widetilde{a}_{i,2t_1+1}, \forall i \in [n]$. Then it does not influence $\widetilde{z}_{2t_1+1}$, and further does not influence $\widetilde{a}_{i,2t_1+2}, \forall i \in [n]$. From induction, we know agent $(i_1, 2t_1)$ does not influence agent $(i_2, 2t_2)$, which leads to a contradiction.

This completes the proof of this lemma. □

We now go back to prove the theorem. Firstly, we prove the QC cases. To show $\mathcal{D}_{\mathcal{L}}$ is QC, we need to prove $\forall i_1, i_2 \in [n], h_1, h_2 \in [\widetilde{H}]$, if agent (i_1, h_1) influences agent (i_2, h_2) with $h_1 < h_2$, then $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$, where we use $\widetilde{\tau}_{i,h}$ to denote the available information of agent (i, h) in $\mathcal{D}_{\mathcal{L}}$. We prove this by considering the following cases:

1) If $h_1 = 2t_1 - 1$ with $t_1 \in [H]$, by the construction of $\mathcal{D}_{\mathcal{L}}$ and Assumption III.4, we have $\widetilde{\tau}_{i_1,h_1} = \widetilde{c}_{h_1} = c_{t_1^-} \subseteq \widetilde{\tau}_{i_2,h_2}$, since common information accumulates over time by definition, and will always be included in the available information $\widetilde{\tau}_{i,h}$ in later steps. Thus, $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.
2) If $h_1 = 2t_1, h_2 = 2t_2$ with $t_1, t_2 \in [H]$, then $\widetilde{\tau}_{i_1,h_1} = \tau_{i_1,t_1^+} = \tau_{i_1,t_1^-} \cup z^a_{t_1}$ by definition. Consider agent (i_1, t_1) and (i_2, t_2) in $\mathcal{L}$. From Lemma .5, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_2,t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^+})$. Also, $z^a_{t_1} \subseteq c_{t_1^+} \subseteq c_{t_2^+} \subseteq \tau_{i_2,t_2^+} = \widetilde{\tau}_{i_2,h_2}$ by the accumulation of c_{h^+} over time. Thus, we have $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.
3) If $h_1 = 2t_1, h_2 = 2t_2 - 1, t_1, t_2 \in [H]$, then $\widetilde{\tau}_{i_2,h_2} = \widetilde{c}_{h_2}$, then $\exists i_3 \in [n], i_3 \neq i_1, \widetilde{\tau}_{i_2,h_2} \subseteq \widetilde{c}_{h_2+1} \subseteq \widetilde{\tau}_{i_3,h_2+1}$. From agent (i_1, h_1) influences (i_2, h_2), we know agent (i_1, h_1) also influences agent $(i_3, h_2 + 1)$ in $\mathcal{D}_{\mathcal{L}}$, hence agent (i_1, t_1) influences agent (i_2, t_2) in $\mathcal{L}$. Since $\mathcal{L}$ is QC, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_3,t_2^-})$. From Assumption III.9 and $i_1 \neq i_3$, we know $\sigma(\widetilde{\tau}_{i_1,h_1}) = \sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(c_{t_2^-}) = \sigma(\widetilde{\tau}_{i_2,h_2})$.

Second, we prove the sQC case. In $\mathcal{D}_{\mathcal{L}}$, for $\forall i_1, i_2 \in [n], h_1, h_2 \in [\widetilde{H}]$, agent (i_1, h_1) influences (i_2, h_2). From the prove above, we know $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$. We only need to prove $\sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.

1) If $h_1 = 2t_1 - 1$ with $t_1 \in [H]$, then we know $\widetilde{a}_{i_1,h_1} = m_{i_1,t}$. From Assumption II.1, we know that $m_{i_1,t} \subseteq z^a_{i_1,t}$. Then we get $\sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\widetilde{z}_{i_1,h_1+1}) \subseteq \sigma(\widetilde{c}_{h_2}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.
2) If $h_1 = 2t_1, h_2 = 2t_2$ with $t_1, t_2 \in [H]$, then from Lemma .5, we know that $\sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.
3) If $h_1 = 2t_1, h_2 = 2t_2 - 1, t_1, t_2 \in [H]$, then $\widetilde{\tau}_{i_2,h_2} = \widetilde{c}_{h_2}$, then $\exists i_3 \in [n], i_3 \neq i_1, \widetilde{\tau}_{i_2,h_2} \subseteq \widetilde{c}_{h_2+1} \subseteq \widetilde{\tau}_{i_3,h_2+1}$. From agent (i_1, h_1) influences (i_2, h_2), we know agent (i_1, h_1) also influences agent $(i_3, h_2 + 1)$ in $\mathcal{D}_{\mathcal{L}}$, hence agent (i_1, t_1) influences agent (i_2, t_2) in $\mathcal{L}$. Since $\mathcal{L}$ is sQC, we know $\sigma(a_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_3,t_2^-})$. From Assumption III.9 and $i_1 \neq i_3$, we know $\sigma(\widetilde{a}_{i_1,h_1}) = \sigma(a_{i_1,t_1}) \subseteq \sigma(c_{t_2^-}) = \sigma(\widetilde{\tau}_{i_2,h_2})$.

This completes the proof. □

6) *Proof of Lemma IV.3:*

Proof. From the construction of $\mathcal{D}^\dagger_{\mathcal{L}}$, since $\mathcal{D}^\dagger_{\mathcal{L}}$ requires agent to share more than $\mathcal{D}_{\mathcal{L}}$, it is easy to observe the fact that $\forall h \in [\widetilde{H}], i \in [n], \widetilde{c}_h \subseteq \breve{c}_h, \widetilde{\tau}_{i,h} \subseteq \breve{\tau}_{i,h}$.

Let $i_1, i_2 \in [n], h_1, h_2 \in [\widetilde{H}], h_1 < h_2$, and agent (i_1, h_1) influences agent (i_2, h_2) in $\mathcal{D}^\dagger_{\mathcal{L}}$.

- If $h_1 = 2t_1 - 1$ with $t_1 \in [H]$, then h_1 is communication step. So $\breve{\tau}_{i_1,h_1} = \breve{c}_{h_1} \subseteq \breve{c}_{h_2}$, and $\widetilde{a}_{i_1,h_1} = m_{i_1,t_1} \subseteq \breve{c}_{h_1+1} \subseteq \breve{c}_{h_2}$ from Assumption II.1. Therefore, we have $\sigma(\breve{\tau}_{i_1,h_1}) \cup \sigma(\breve{a}_{i_1,h_1}) \subseteq \sigma(\breve{c}_{h_1}) \subseteq \sigma(\breve{\tau}_{i_2,h_2})$.
- If $h_1 = 2t_1, h_2 = 2t_2 - 1$ with $t_1, t_2 \in [H]$, then $\breve{\tau}_{i_2,h_2} = \breve{c}_{h_2}$. If agent (i_1, h_1) does not influence (i_2, h_2) in $\mathcal{D}_{\mathcal{L}}$, but agent (i_1, h_1) influences (i_2, h_2) in $\mathcal{D}^\dagger_{\mathcal{L}}$, then it means $\breve{a}_{i_1,h_1} \in \breve{\tau}_{i_2,h_2}$ but $\widetilde{a}_{i_1,h_1} \notin \widetilde{\tau}_{i_2,h_2}$. This can only happen when $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_{h_2}) \subseteq \sigma(\breve{c}_{h_2})$, and $\widetilde{a}_{i_1,h_1} \subseteq \breve{c}_{h_2}$. Also, from the construction of $\mathcal{D}^\dagger_{\mathcal{L}}$, we know that $\breve{\tau}_{i_1,h_1} \backslash \widetilde{\tau}_{i_1,h_1} \subseteq \breve{c}_{h_1}$. Therefore, we have $\sigma(\breve{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\breve{c}_{h_2}) \subseteq \sigma(\breve{\tau}_{i_2,h_2})$.
  If agent (i_1, h_1) influences (i_2, h_2) in $\mathcal{D}_{\mathcal{L}}$, then from QC of $\mathcal{D}_{\mathcal{L}}$, we know that $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_{h_2})$, then $\widetilde{a}_{i_1,h_1} \in \breve{c}_{h_2}$. Still, we have $\breve{\tau}_{i_1,h_1} \backslash \widetilde{\tau}_{i_1,h_1} \subseteq \breve{c}_{h_1}$. Therefore, $\sigma(\breve{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\breve{\tau}_{i_2,h_2})$.
- If $h_1 = 2t_1, h_2 = 2t_2$ with $t_1, t_2 \in [H]$. If agent (i_1, h_1) does not influence (i_2, h_2) in $\mathcal{D}_{\mathcal{L}}$, then it means sharing $\widetilde{a}_{i_1,h_1}$ leads to the influence. Then, $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_{h_2}) \subseteq \sigma(\breve{c}_{h_2})$, and $\widetilde{a}_{i_1,h_1} \subseteq \breve{c}_{h_2}$. We can conclude $\sigma(\breve{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\breve{c}_{h_2}) \subseteq \sigma(\breve{\tau}_{i_2,h_2})$.
  Now we consider the case that agent (i_1, h_1) influences (i_2, h_2) in $\mathcal{D}_{\mathcal{L}}$. If $i_1 \neq i_2$, then we have $\widetilde{\tau}_{i_1,h_1} \subseteq \widetilde{\tau}_{i_2,h_2}$. From Assumption III.9, and $i_1 \neq i_2$, we know $\widetilde{\tau}_{i_1,h_1} \subseteq \widetilde{c}_{h_2}$. Then, from the construction of $\mathcal{D}^\dagger_{\mathcal{L}}$, we have $\widetilde{a}_{i_1,h_1} \subseteq \breve{c}_{h_2}$. Finally, we have $\sigma(\breve{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\breve{\tau}_{i_2,h_2})$.
  If $i_1 = i_2$, then from the perfect recall of $\mathcal{L}$, we know that $\widetilde{\tau}_{i_1,h_1} \cup \widetilde{a}_{i_1,h_1} \subseteq \widetilde{\tau}_{i_2,h_2}$. From $\breve{\tau}_{i_1,h_1} \backslash \widetilde{\tau}_{i_1,h_1} \subseteq \breve{c}_{h_1}$, we conclude $\sigma(\breve{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\breve{\tau}_{i_2,h_2})$.

This completes the proof. □

7) *Proof of Theorem IV.4:*

Proof. We firstly prove that given any strategy $\breve{g}_{1:H}$ and $\widetilde{g}_{1:H} = \varphi(\breve{g}_{1:H}, \mathcal{D}_{\mathcal{L}})$, $J_{\mathcal{D}^\dagger_{\mathcal{L}}}(\breve{g}_{1:H}) = J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:H})$. Since $\mathcal{D}^\dagger_{\mathcal{L}}$ only changes what to share, $\widetilde{\tau}_h = \breve{\tau}_h$ always hold. Then, for any $i \in [n], h \in [\widetilde{H}], \widetilde{\tau}_h \in \widetilde{\mathcal{T}}_h$, let $\widetilde{\tau}_{i,h}, \breve{\tau}_{i,h}$ be the corresponding information of agent i in $\mathcal{D}_{\mathcal{L}}, \mathcal{D}^\dagger_{\mathcal{L}}$, respectively. From Algorithm 3, we know that $\widetilde{g}_{i,h}(\widetilde{\tau}_{i,h}) = \breve{g}_{i,h}(\breve{\tau}_{i,h})$. This is because, for any $\widetilde{a}_{j,t} \in \breve{\tau}_{i,h} \backslash \widetilde{\tau}_{i,h}, j \in [n], t < h$, there must holds that $\sigma(\widetilde{\tau}_{j,t}) \subseteq \sigma(\widetilde{c}_{i,h})$. Therefore, we can always recover $\widetilde{a}_{j,t}$ from $\breve{\tau}_{i,h}$ and $\widetilde{g}_{i,h}$. As a result, we can have $J_{\mathcal{D}^\dagger_{\mathcal{L}}}(\breve{g}_{1:H}) = J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:H})$.

Since $\mathcal{D}^\dagger_{\mathcal{L}}$ has larger strategy spaces, $\max_{\widetilde{g}_{1:\widetilde{H}} \in \widetilde{G}_{1:\widetilde{H}}} J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) \leq \max_{\breve{g}_{1:\breve{H}} \in \breve{G}_{1:\breve{H}}} J_{\mathcal{D}^\dagger_{\mathcal{L}}}(\breve{g}_{1:\breve{H}})$. Let $\breve{g}^*_{1:\breve{H}}$ be the strategy satisfying $J_{\mathcal{D}^\dagger_{\mathcal{L}}}(\breve{g}^*_{1:\breve{H}}) \geq \max_{\breve{g}_{1:\breve{H}} \in \breve{G}_{1:\breve{H}}} J_{\mathcal{D}^\dagger_{\mathcal{L}}}(\breve{g}_{1:\breve{H}}) - \epsilon$. Then, we have $J_{\mathcal{D}_{\mathcal{L}}}(\varphi(\breve{g}^*_{1:\breve{H}}, \mathcal{D}_{\mathcal{L}})) = J_{\mathcal{D}^\dagger_{\mathcal{L}}}(\breve{g}^*_{1:\breve{H}}) \geq \max_{\breve{g}_{1:\breve{H}} \in \breve{G}_{1:\breve{H}}} J_{\mathcal{D}^\dagger_{\mathcal{L}}}(\breve{g}_{1:\breve{H}}) - \epsilon \geq \max_{\widetilde{g}_{1:\widetilde{H}} \in \widetilde{G}_{1:\widetilde{H}}} J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) - \epsilon$. Then $\varphi(\breve{g}^*_{1:\breve{H}}, \mathcal{D}_{\mathcal{L}})$ is the ϵ-team optimal strategy of $\mathcal{D}_{\mathcal{L}}$. □

8) *Proof of Theorem IV.5:*

Proof. For any fixed $h \in [\overline{H}]$, we want to prove the belief $\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h)$ is independent to the strategy of any agent before h, namely, $\forall h_1 \in [h-1], i_1 \in [n], \overline{g}_{1:h-1} \in \overline{\mathcal{G}}_{1:h-1}, \overline{g}'_{i_1,h_1} \in \overline{\mathcal{G}}_{i_1,h_1}$, let $\overline{g}'_{1:h-1} = (\overline{g}_{1,1}, \cdots, \overline{g}_{i_1-1,h_1}, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h-1})$, the following holds

$$\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}). \tag{.2}$$

If $\exists i_2 \in [n], i_2 \neq i_1$, $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h}), \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h})$. Then since $\mathcal{L}$ satisfies Assumption III.9, we know $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h), \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h)$. Then, there exists a unique $\overline{\tau}'_{i_1,h_1}$ and a unique $\overline{a}'_{i_1,h_1}$ that $\mathbb{P}(\overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1} \,|\, \overline{c}_h) = 1$, then

$$\begin{aligned}
\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) &= \sum_{\substack{\overline{\tau}_{i_1,h_1} \in \overline{\mathcal{T}}_{i_1,h_1} \\ \overline{a}_{i_1,h_1} \in \overline{\mathcal{A}}_{i_1,h_1}}} \mathbb{P}(\overline{s}_h, \overline{p}_h, \overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1} \,|\, \overline{c}_h, \overline{g}_{1:h-1}) \\
&= \mathbb{P}(\overline{s}_h, \overline{p}_h, \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1} \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1}, \overline{c}_h, g_{1:h-1}) \mathbb{P}(\overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1} \,|\, \overline{c}_h, \overline{g}_{1:h-1}) \\
&= \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1} \backslash \overline{g}_{i_1,h_1}).
\end{aligned}$$

The last equality is because the input and output of $\overline{g}_{i_1:h_1}$ are $\overline{\tau}'_{i_1,h_1}$ and $\overline{a}'_{i_1,h_1}$. Then we know $\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1})$.
If $\forall i_2 \neq i_1, \sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h})$ or $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h})$, then agent (i_1, h_1) does not influence agent $(i_2, h), \forall i_2 \neq i_1$ because $\mathcal{D}'_{\mathcal{L}}$ is sQC. And $h_1 = 2k_1$ with $k_1 \in [n]$. (If h_1 is odd, then $\overline{\tau}_{i_1,h_1} = \overline{c}_{h_1} \subseteq \overline{c}_h \subseteq \overline{\tau}_{i_2,h}$, and $\sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{z}_{h_1+1}) \subseteq \overline{c}_h$, contradiction!) Then:

- If $h = 2k-1, k \in [n]$, then $\overline{p}_h = \emptyset$. If agent (i_1, h_1) influences $\overline{s}_h$, then consider the timestep $h+1$, $\overline{o}_h \subseteq \overline{\tau}_{h+1}$. Since $\mathcal{L}$ satisfies the Assumption III.7, we know she also influences $\overline{o}_{-i,h}$, then $\exists i_3 \neq i_1$, agent (i_1, h_1) influences $\overline{\tau}_{i_3,h+1}$. Then agent (i_1, h_1) must influences agent $(i_3, h+1)$ in the problem $\mathcal{D}_{\mathcal{L}}$. From Lemma .5, we know $\sigma(\tau_{i_1,k_1^-}) \subseteq \sigma(\tau_{i_2,k^-})$ in $\mathcal{L}$. Also, $\widetilde{\tau}_{i_1,h} = \widetilde{\tau}_{i_1,h-1} \cup \widetilde{z}_h$ because h is even. Then, we have $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h})$. We only requires agent to share more in $\mathcal{D}'_{\mathcal{L}}$, then we can further get $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h})$. Also, from the construction from QC problem to sQC, we know $\sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h)$, contradiction! Hence, we know agent (i_1, h_1) does not influence state $\overline{s}_h$.
  Also, $\overline{\tau}_{i_1,h} = \overline{c}_h = \overline{\tau}_{i_2,h}, \forall i_2 \neq i_1$, then agent (i_1, h_1) does not influence agent (i_1, h).
- If $h = 2k, k \in [n]$. If agent (i_1, h_1) influences $\overline{s}_{h_1+1}$, then from Assumption III.7, she influences $\overline{o}_{-i,h_1+1}$, and $\overline{o}_{-i,h_1+1} \subseteq \overline{\tau}_{-i,h_1+2}$, then $\exists i_3 \neq i_1$, agent (i_1, h_1) influence $\overline{\tau}_{i_3,h_1+2}$, so $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \overline{c}_{h_1+1} \subseteq \overline{c}_{h_2}$ and $\sigma(\overline{a}_{i_1,h_1}) \subseteq \overline{c}_{h_1+2} \subseteq \overline{c}_{h_2}$. Contradiction! So agent (i_1, h_1) does not influences $\overline{s}_{h_1+1}$.
  Also, since (i_1, h_1) does not influences $\overline{s}_{h_1+1}$, from the Assumption III.5, $\overline{a}_{i_1,h_1} \notin \overline{\tau}_{i_1,h'}, \forall h' > h_1$. So agent (i_1, h_1) does not influence $\overline{\tau}_{i_1,h}$.

Therefore, we know agent (i_1, h_1) does not influence $\overline{s}_h$, and does not influence $\overline{\tau}_{i,h}, \forall i \in [n]$.

$$\begin{aligned}
\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) &= \mathbb{P}(\overline{s}_h, \overline{p}_h, \overline{c}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \{\overline{\tau}_{i,h}\}_{i \in [n]} \,|\, \overline{c}_h, \overline{g}_{1:h-1}) \\
&= \mathbb{P}(\overline{s}_h, \{\overline{\tau}_{i,h}\}_{i \in [n]} \,|\, \overline{c}_h, \overline{g}'_{1:H}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}).
\end{aligned}$$

This completes the proof. □

9) *Proof of Theorem IV.6:*

Proof. We first prove that the Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ satisfies information evolution rules. For each $t \in [H]$, we define the random variable $\widehat{p}_{i,2t-1} = p_{i,t^-}, \widehat{p}_{2t-1} = p_{t^-}$. Recall the reformulation, $\widetilde{p}_{i,2t-1} = \emptyset$ rather than p_{i,t^-}. Then from the 2H-reformulation and Assumption II.1, it holds that, for any $i \in [n], h \in [\overline{H}]$, if $h = 2t-1$ with $t \in [2:H]$

$$\widetilde{z}_h = \chi_t(\widetilde{p}_{h-1}, \widetilde{a}_{h-1}, \widetilde{o}_h), \qquad \widehat{p}_{i,h} = \xi_{i,t}(\widetilde{p}_{i,h-1}, \widetilde{a}_{i,h-1}, \widetilde{o}_{i,h});$$

if $h = 2t$ with $t \in [H]$

$$\widetilde{z}_h = \phi_t(\widehat{p}_{h-1}, \widetilde{a}_{h-1}), \qquad \widetilde{p}_{i,h} = \widehat{p}_{i,h-1} \backslash \phi_{i,t}(\widehat{p}_{i,h-1}, \widetilde{a}_{i,h-1}),$$

where $\chi_t, \xi_{i,t}$ are fixed transformations and $\phi_h, \phi_{i,h}$ are additional-sharing functions. Then, we can construct the $\{\overline{\chi}_{h+1}\}_{h \in [\overline{H}]}, \{\overline{\xi}_{i,h+1}\}_{i \in [n], h \in [\overline{H}]}$ accordingly as follows:

- If $h = 2t-1$ with $t \in [H]$, for any $\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h$, since $\overline{p}_{h-1} = \breve{p}_{h-1}$ from construction of $\mathcal{D}'_{\mathcal{L}}$, we can select an $\widetilde{p}_{h-1}$ that $\breve{p}_{h-1}$ can be generated from $\widetilde{p}_{h-1}$ through expansion (such $\widetilde{p}_{h-1}$ might not be unique). Then, define $\overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h) = \chi_t(\widetilde{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h) \cup \{\overline{a}_{j,h_1} \,|\, j \in [n], h_1 < h, \overline{a}_{j,h_1} \in \overline{p}_{h-1}, \sigma(\widetilde{\tau}_{j,h_1}) \subseteq \sigma(\widetilde{c}_h)\} \backslash (\widetilde{p}_{h-1} \backslash \overline{p}_{h-1})$. Since χ_t is a fixed transformation and we remove the $\widetilde{p}_{h-1} \backslash \overline{p}_{h-1}$ part from $\overline{z}_h$, the value $\overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$ is the same no matter what $\widetilde{p}_{h-1}$ we select, and thus such $\overline{\chi}_h$ is well-defined. Similarly, we can define $\overline{\xi}_{i,h}(\overline{p}_{i,h-1}, \overline{a}_{i,h-1}, \overline{o}_{i,h-1}) = \xi_{i,t}(\widetilde{p}_{i,h-1}, \overline{a}_{i,h-1}, \overline{o}_{i,h}) \backslash \{\overline{a}_{i,h_1} \,|\, h_1 < h, \overline{a}_{i,h_1} \in \overline{p}_{i,h-1}, \sigma(\widetilde{\tau}_{i,h_1}) \subseteq \sigma(\widetilde{c}_h)\} \backslash (\widetilde{p}_{i,h-1} \backslash \overline{p}_{i,h-1})$.

- If $h = 2t$ with $t \in [H]$, for any $\overline{p}_{h-1}, \overline{a}_{h-1}$, from the construction of $\mathcal{D}'_{\mathcal{L}}$, we can select an $\widehat{p}_{h-1}$ that $\overline{p}_{h-1}$ can be generated from $\widehat{p}_{h-1} = p_{t^-}$ through expansion (such $\widehat{p}_{h-1}$ might not be unique). Also, it holds that $\overline{o}_h = \emptyset$, then define $\overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h) = \phi_t(\widehat{p}_{h-1}, \overline{a}_{h-1}) \cup \{\overline{a}_{j,h_1} \,|\, j \in [n], h_1 < h, \overline{a}_{j,h_1} \in \overline{p}_{h-1}, \sigma(\widetilde{\tau}_{j,h_1}) \subseteq \sigma(\widetilde{c}_h)\} \backslash (\widehat{p}_{h-1} \backslash \overline{p}_{h-1})$. Still, since ϕ_t is the addition-sharing function, which part of $\widehat{p}_{h-1}$ to share only depends on $\overline{a}_{h-1}$, and not depends on the value of $\widehat{p}_{h-1}$, and we remove the $\widehat{p}_{h-1} \backslash \overline{p}_{h-1}$ part from $\overline{z}_h$, the value of $\overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$ is the same no matter what $\widehat{p}_{h-1}$ we select, and thus such $\overline{\chi}_h$ is well-defined. Similarly, we can define $\overline{\xi}_{i,h}(\overline{p}_{i,h-1}, \overline{a}_{i,h-1}, \overline{o}_{i,h-1}) = \overline{p}_{i,h-1} \backslash \{\overline{a}_{i,h_1} \,|\, h_1 < h, \overline{a}_{i,h_1} \in \overline{p}_{i,h-1}, \sigma(\widetilde{\tau}_{i,h_1}) \subseteq \sigma(\widetilde{c}_h)\} \backslash \phi_{i,t}(\widehat{p}_{i,h-1}, \overline{a}_{i,h-1})$.

Therefore, common and private information of $\mathcal{D}'_{\mathcal{L}}$ satisfies

$$\overline{c}_{h+1} = \overline{c}_h \cup \overline{z}_{h+1}, \overline{z}_{h+1} = \overline{\chi}_{h+1}(\overline{p}_h, \overline{a}_h, \overline{o}_{h+1})$$
$$\text{for each } i \in [n], \overline{p}_{i,h+1} = \overline{\xi}_{i,h+1}(\overline{p}_{i,h}, \overline{a}_{i,h}, \overline{o}_{i,h+1}),$$

with some functions $\{\overline{\chi}_{h+1}\}_{h \in [\overline{H}]}, \{\overline{\xi}_{i,h+1}\}_{i \in [n], h \in [\overline{H}]}$.
Then, we prove that such a Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ is SI with respect to the strategy space $\overline{\mathcal{G}}_{1:\overline{H}}$. This is equivalent to that for any $h \in [\overline{H}], \overline{s}_h \in \overline{\mathcal{S}}, \overline{p}_h \in \overline{\mathcal{P}}_h, \overline{c}_h \in \overline{\mathcal{C}}_h, i_1 \in [n], h_1 < h, \overline{g}_{1:h-1}, \overline{g}'_{i_1,h_1} \in \overline{\mathcal{G}}_{i_1:h_1}$, let $\overline{g}'_{1:h-1} = (\overline{g}_{1,1}, \cdots, \overline{g}_{i_1-1,h_1}, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h-1})$, it holds

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}). \tag{.3}$$

We prove this case-by-case. If $h = 2t$ with $t \in [H]$, then from the result of Theorem IV.5, it holds

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}^\dagger_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}^\dagger_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}).$$

If $h = 2t - 1$ with $t \in [H]$, and $h_1 = 2t_1 - 1$ with $t_1 \in [H]$, which means h_1 is communication timestep. Then it holds that $\overline{c}_{h_1} \subseteq \overline{c}_h, \overline{a}_{i_1,h_1} = m_{i_1, \frac{h_1+1}{2}} \in \overline{c}_h$, then

$$\begin{aligned} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) &= \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_{h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1}) \\ &= \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_{h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1} \backslash \overline{g}_{i_1,h_1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}), \end{aligned}$$

where the second equal sign is because the input and output of $\overline{g}_{i_1,h_1}$ are $\overline{c}_{h_1}$ and $\overline{a}_{i_1,h_1}$.
If $h = 2t - 1$ with $t \in [H]$, and $h_1 = 2t_1$ with $t_1 \in [H]$, which means h_1 in control timestep. then if agent (i_1, h_1) influences the underlying state $\overline{s}_{h_1+1}$, then from Assumption $III.7$, we know that there exists $i_2 \neq i_1$ that, agent (i_1, t_1) influences $o_{i_2,t}$, and thus influences agent (i_2, t) in problem $\mathcal{L}$ even there is no additional sharing. From QC of $\mathcal{L}$ and Assumption III.9, we know that $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_2,t^-}) \subseteq \sigma(c_t)$. Also, from $\tau_{i_1,t^-} \backslash \tau_{i_1,t_1^+} \subseteq c_{t^+}$, we get $\sigma(\tau_{i_1,t_1^+}) \subseteq \sigma(c_t)$. After reformulation, we have $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_h)$. From the definition of strict expansion in Eq. (IV.2), we have $\overline{a}_{i_1,h_1} \in \overline{c}_h$, and $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h)$. Then, we conclude

$$\begin{aligned} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) &= \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1}) \\ &= \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1} \backslash \overline{g}_{i_1,h_1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}), \end{aligned}$$

where the second equal sign is because the input and output of $\overline{g}_{i_1,h_1}$ are $\overline{\tau}_{i_1,h_1}$ and $\overline{a}_{i_1,h_1}$.
If agent (i_1, h_1) does not influence the underlying state $\overline{s}_{h_1+1}$, then from Assumption III.5, $\overline{a}_{i_1,h_1} \notin \overline{\tau}_{h_2}$ for any $h_2 > h_1$. Then, agent (i_1, h_1) will not influence $\overline{s}_h$ and $\overline{p}_h$. Then, it directly holds that

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}).$$

□

10) Important Definitions of SI Dec-POMDP: Given a Dec-POMDP SI $\mathcal{D}'_{\mathcal{L}}$ from $\mathcal{L}$ after reformulation and refinement. In this part, we only need to discuss how to solve this $\mathcal{D}'_{\mathcal{L}}$. To avoid abuse of notation, we use $\overline{x}$ to represent the notation in $\mathcal{D}'_{\mathcal{L}}$.

First, we define the following quantities.

Definition .6 (Value function). For each $i \in [n]$ and $h \in [\overline{H}]$, given common information $\overline{c}_h$ and strategy $\overline{g}_{1:H}$, the value function conditioned on the common information is defined as:

$$V_h^{\overline{g}, \mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) := \mathbb{E}_{\overline{g}}^{\mathcal{D}'_{\mathcal{L}}}[\sum_{h'=h}^{H} \overline{\mathcal{R}}_{h'}(\overline{s}'_h, \overline{a}'_h, \overline{p}'_h) \,|\, \overline{c}_h], \tag{.4}$$

where $\overline{\mathcal{R}}_{h'}$ takes $\overline{s}'_h, \overline{a}'_h, \overline{p}'_h$ as input, since after reformulation, the reward may come from communication cost, which is a function of $\overline{p}'_h$ and $\overline{a}'_h$.

Definition .7 (Prescriptions and Q-Value function)**.** Prescriptions is an important concept in the common-information-based framework [16], [17]. The prescription of agent i at timestep h is defined as $\gamma_{i,h}:\overline{\mathcal{P}}_{i,h}\to\overline{\mathcal{A}}_{i,h}$. We use γ_h to denote joint prescription and $\Gamma_{i,h},\Gamma_h$ to denote the prescription space. The prescriptions are marginalization of strategy $\overline{g}_h$, i.e. $\gamma_{i,h}(\cdot\,|\,\overline{p}_{i,h})=\overline{g}_{i,h}(\cdot\,|\,\overline{c}_h,\overline{p}_{i,h})$. Then we can define the Q-value function as

$$Q_h^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h,\gamma_h):=\mathbb{E}_{\overline{g}}^{\mathcal{D}'_{\mathcal{L}}}\left[\sum_{h'=h}^{H}\overline{\mathcal{R}}_{h'}(\overline{s}'_h,\overline{a}'_h,\overline{p}'_h)\,|\,\overline{c}_h,\gamma_h\right]. \tag{.5}$$

Remark .8. In this paper, for any Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ generated by an $\mathcal{L}$ after reformulation, strict expansion and refinement, we only consider the strategy spaces at odd timesteps as $\overline{\mathcal{G}}_{i,2t-1}:\overline{\mathcal{C}}_{2t-1}\to\overline{\mathcal{A}}_{i,2t-1}$ and aim to find the optimal strategy in these classes. Therefore, we define the prescriptions at odd timesteps as $\forall h\in[H],i\in[n],\Gamma_{i,2h-1}=\overline{\mathcal{A}}_{i,2h-1},\Gamma_{2h-1}=\overline{\mathcal{A}}_{2h-1}$

Definition .9 (Expected approximate common information model)**.** We define an expected approximate common information model of $\mathcal{D}'_{\mathcal{L}}$ as

$$\mathcal{M}:=\left(\{\widehat{\mathcal{C}}_h\}_{h\in[\overline{H}]},\{\widehat{\phi}_h\}_{h\in[\overline{H}]},\{\mathbb{P}_h^{\mathcal{M},z}\}_{h\in[\overline{H}]},\Gamma,\{\widehat{\mathcal{R}}_h^{\mathcal{M}}\}_{h\in[\overline{H}]}\right), \tag{.6}$$

where Γ is the joint prescription space, $\widehat{\mathcal{C}}_h$ is the space of approximate common information at step h. $\mathbb{P}_h^{\mathcal{M},z}:\widehat{\mathcal{C}}_h\times\Gamma_h\to\Delta(\overline{Z}_{h+1})$ gives the probability of $\overline{z}_{h+1}$ under $\widehat{c}_h$ and γ_h. $\widehat{\mathcal{R}}_h^{\mathcal{M}}:\widehat{\mathcal{C}}_h\times\Gamma_h\to[0,1]$ gives the reward at timestep h given $\widehat{c}_h$ and γ_h. Then, we call $\mathcal{M}$ os an $(\epsilon_r(\mathcal{M}),\epsilon_z(\mathcal{M}))$-expected-approximate common information model of $\mathcal{D}'_{\mathcal{L}}$ with some compression function Compress$_h$ that $\widehat{c}_h$ =Compress$_h(\overline{c}_h)$ satisfying following:

- There exists a transformation function $\widehat{\phi}_h$ such that

  $$\widehat{c}_h=\widehat{\phi}_h(\widehat{c}_{h-1},\overline{z}_h), \tag{.7}$$

  where $\overline{z}_h=\overline{c}_h\backslash\overline{c}_{h-1}$ in $\mathcal{D}'_{\mathcal{L}}$.
- For any $\overline{g}_{1:h-1}$ and any prescription $\gamma_h\in\Gamma_h$, it holds that

  $$\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}|\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\overline{\mathcal{R}}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)\,|\,\overline{c}_h,\gamma_h]-\widehat{\mathcal{R}}_h^{\mathcal{M}}(\widehat{c}_h,\gamma_h)|\leq\epsilon_r(\mathcal{M}). \tag{.8}$$
- For any $\overline{g}_{1:h-1}$ and any prescription $\gamma_h\in\Gamma_h$, it holds that

  $$\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot\,|\,\overline{c}_h,\gamma_h)-\mathbb{P}_h^{\mathcal{M},z}(\cdot\,|\,\widehat{c}_h,\gamma_h)||_1\leq\epsilon_z(\mathcal{M}). \tag{.9}$$

Definition .10 (Value function under $\mathcal{M}$)**.** Given an Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$and its expected approximate common information model $\mathcal{M}$. For any strategy $\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}},h\in[H]$, we define the value function as

$$V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)=\widehat{\mathcal{R}}_h^{\mathcal{M}}(\text{Compress}_h(\overline{c}_h),\{\overline{g}_{j,h}(\cdot\,|\,\overline{c}_h,\cdot)\}_{j\in[n]})+\mathbb{E}^{\mathcal{M}}[V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1})\,|\,\text{Compress}_h(\overline{c}_h),\{\overline{g}_{j,h}(\cdot\,|\,\overline{c}_h,\cdot)\}_{j\in[n]}]. \tag{.10}$$

Definition .11 (Model-belief consistency)**.** We say the expected approximate common information model $\mathcal{M}$ is *consistent with* some belief $\{\mathbb{P}_h^{\mathcal{M},c}(s_h,p_h\,|\,\widehat{c}_h)\}_{h\in[H]}$ if it satisfies the following for all $i\in[n]$, $h\in[H]$:

$$\mathbb{P}_h^{\mathcal{M},z}(z_{h+1}\,|\,\widehat{c}_h,\gamma_h)=\sum_{\substack{s_h,p_h,a_h,o_{h+1}:\\ \chi_{h+1}(p_h,a_h,o_{h+1})=z_{h+1}}}\left(\mathbb{P}_h^{\mathcal{M},c}(s_h,p_h\,|\,\widehat{c}_h)\prod_{j=1}^{n}\gamma_{j,h}(a_{j,h}\,|\,p_{j,h})\times\sum_{s_{h+1}}\mathbb{T}_h(s_{h+1}\,|\,s_h,a_h)\mathbb{O}_{h+1}(o_{h+1}\,|\,s_{h+1})\right), \tag{.11}$$

$$\widehat{r}_{i,h}^{\mathcal{M}}(\widehat{c}_h,\gamma_h)=\sum_{s_h,p_h,a_h}\mathbb{P}_h^{\mathcal{M},c}(s_h,p_h\,|\,\widehat{c}_h)\prod_{j=1}^{n}\gamma_{j,h}(a_{j,h}\,|\,p_{j,h})r_{i,h}(s_h,a_h). \tag{.12}$$

Definition .12 (Policy-dependent approximate common information model)**.** Given a model $\widetilde{\mathcal{M}}$ (as in Definition **??**) and H joint policies $\pi^{1:H}$, where each $\pi^h\in\Delta(\Pi^{\mathrm{det}})$ for $h\in[H]$, we say $\widetilde{\mathcal{M}}$ is a *policy-dependent expected approximate common information model*, denoted as $\widetilde{\mathcal{M}}(\pi^{1:H})$, if it is consistent with the *policy-dependent* belief $\{\mathbb{P}_h^{\pi^h,\mathcal{G}}(s_h,p_h\,|\,\widehat{c}_h)\}_{h\in[H]}$ (as per Definition .11).

Definition .13 (Length of approximate common information)**.** Given the compression functions $\{\text{Compress}_h\}_{h\in[H+1]}$, we define the integer $\widehat{L}>0$ as the minimum length such that there exists a mapping $\widehat{f}_h:\mathcal{A}^{\min\{\widehat{L},h\}}\times\mathcal{O}^{\min\{\widehat{L},h\}}\to\widehat{\mathcal{C}}_h$ such that for each $h\in[H+1]$ and joint history $\{o_{1:h},a_{1:h-1}\}$, we have $\widehat{f}_h(x_h)=\widehat{c}_h$, where $x_h=\{a_{\max\{h-\widehat{L},1\}},o_{\max\{h-\widehat{L},1\}+1},\cdots,a_{h-1},o_h\}$.

Finally, we provide the formal guarantees for planning in QC LTC.

Theorem .14. Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, III.7, and IV.7, we can construct an SI Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ such that the following holds. Fix $\epsilon_r, \epsilon_z, \epsilon_e > 0$. Given any (ϵ_r, ϵ_z)-expected-approximate common information model $\mathcal{M}$ for $\mathcal{D}'_{\mathcal{L}}$ that is consistent with some given approximate belief $\{\mathbb{P}_h^{\mathcal{M},c}(s_h, p_h \,|\, \widehat{c}_h)\}_{h\in[\overline{H}]}$, Algorithm 1 can compute an $(2\overline{H}\epsilon_r + \overline{H}^2\epsilon_z + \overline{H}\epsilon_e)$-team optimal policy for the original LTC problem $\mathcal{L}$ with time complexity $\max_{h\in[\overline{H}]} |\widehat{\mathcal{C}}_h| \cdot \mathtt{poly}(|\mathcal{S}|, |\mathcal{A}_h|, |\mathcal{P}_h|, \overline{H}, 1/\epsilon_e)$. In particular, for fixed $\epsilon > 0$, if $\mathcal{L}$ has the baseline sharing as in Appendix A, one can construct an $\mathcal{M}$ and apply Algorithm 1 to compute an ϵ-team optimal policy for $\mathcal{L}$ in quasi-polynomial time.

11) Proof of Theorem .14:

Proof. We divide the proof into the following three **Parts**.

Part I: Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, and III.7, we can construct an SI Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ such that finding an ϵ-team optimal strategy can give us an ϵ-team optimal strategy of $\mathcal{L}$, as shown in Algorithm 1.
We can construct a Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ from $\mathcal{L}$ through Algorithm 1. From Proposition IV.1 and Theorems IV.4, IV.5. We know that $\mathcal{D}'_{\mathcal{L}}$ is SI and ϵ-team optimal strategy of $\mathcal{D}'_{\mathcal{L}}$ can give us ϵ-team optimal strategy of $\mathcal{L}$.

Part II: Given any ϵ-expected-approximate common information model $\mathcal{M}$ of Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$, then there exists an algorithm, Algorithm 6, that can output an ϵ-team optimal strategy of $\mathcal{D}'_{\mathcal{L}}$.
Solving SI Dec-PODMP. We need to prove that solving $\mathcal{M}$ can get the ϵ-team optimal strategy of $\mathcal{D}'_{\mathcal{L}}$. We prove following 2 lemmas first.

Lemma .15. For any strategy $\overline{g}_{1:\overline{H}}$, and $h \in [\overline{H}]$, we have

$$\mathbb{E}_{\overline{g}_{1:\overline{H}}}^{\mathcal{D}'_{\mathcal{L}}}[|V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)|] \leq (\overline{H} - h + 1)\epsilon_r + \frac{(\overline{H} - h + 1)(\overline{H} - h)}{2}\epsilon_z. \tag{.13}$$

Proof. We prove it by induction. For $h = \overline{H} + 1$, we have $V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) = V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = 0$.
For the step $h \leq \overline{H}$, we have

$$\begin{aligned}
&\mathbb{E}_{\overline{g}_{1:\overline{H}}}^{\mathcal{D}'_{\mathcal{L}}}[|V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)|] \\
\leq&\mathbb{E}_{\overline{g}_{1:\overline{H}}}^{\mathcal{D}'_{\mathcal{L}}}\left[|\mathbb{E}^{\mathcal{D}_{\mathcal{L}}}[\overline{\mathcal{R}}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h) \,|\, \overline{c}_h, \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h, \cdot)\}_{j\in[n]}] - \widehat{\mathcal{R}}_h^{\mathcal{M}}(\widehat{c}_h, \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h, \cdot)\}_{j\in[n]})|\right] \\
&+ \mathbb{E}_{\overline{g}_{1:\overline{H}}}^{\mathcal{D}'_{\mathcal{L}}}\left[|\mathbb{E}_{\overline{z}_{h+1}\sim\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot \,|\, \overline{c}_h, \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h, \cdot)\}_{j\in[n]})}[V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h \cup \overline{z}_{h+1})] - \mathbb{E}_{\overline{z}_{h+1}\sim\mathbb{P}_h^{\mathcal{M},z}(\cdot \,|\, \widehat{c}_h, \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h, \cdot)\}_{j\in[n]})}[V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h \cup \overline{z}_{h+1})]|\right] \\
\leq&\epsilon_r + (\overline{H} - h)\mathbb{E}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}^{\mathcal{D}'_{\mathcal{L}}}||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot \,|\, \overline{c}_h, \gamma_h) - \mathbb{P}_h^{\mathcal{M},z}(\cdot \,|\, \widehat{c}_h, \gamma_h)||_1 + \mathbb{E}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}^{\mathcal{D}'_{\mathcal{L}}}\left[|V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_{h+1}) - V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1})|\right] \\
\leq&\epsilon_r + (\overline{H} - h)\epsilon_z + (\overline{H} - h)\epsilon_r + \frac{(\overline{H} - h)(\overline{H} - h - 1)}{2}\epsilon_z \\
\leq&(\overline{H} - h + 1)\epsilon_r + \frac{(\overline{H} - h)(\overline{H} - h + 1)}{2}\epsilon_z.
\end{aligned}$$

The proof mainly follows from the proof of Lemma 2 in [15]. But the difference is that $\mathcal{D}'_{\mathcal{L}}$ may not satisfy Assumption II.1. In the third line of this proof, the part $\overline{z}_{h+1} \sim \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot \,|\, \overline{c}_h, \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h, \cdot)\}_{j\in[n]})$, $\overline{z}_{h+1}$ generates as

$$\begin{aligned}
&\gamma_h = \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h, \cdot)\}_{j\in[n]}, \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{z}_{h+1} \,|\, \overline{c}_h, \gamma_h) \\
=&\sum_{\overline{s}_h\in\overline{\mathcal{S}}, \overline{p}_h\in\overline{\mathcal{P}}_h} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \quad \overline{p}_h \,|\, \overline{c}_h) \sum_{\overline{s}_{h+1}\in\overline{\mathcal{S}}, \overline{o}_{h+1}\in\overline{\mathcal{O}}_{h+1}} \overline{\mathbb{T}}_{h+1}(\overline{s}_{h+1} \,|\, \overline{s}_h, \gamma_h(\overline{p}_h))\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1} \,|\, \overline{s}_{h+1})\mathbb{1}[\overline{\chi}_{h+1}(\overline{p}_h, \gamma_h(\overline{p}_h), \overline{o}_{h+1}, \overline{c}_h)].
\end{aligned}$$

More specifically, $\overline{z}_{h+1} = \overline{\chi}_{h+1}(\overline{p}_h, \overline{a}_h, \overline{o}_{h+1}, \overline{c}_h)$, where $\overline{z}_{h+1}$ may be implicit influenced by $\overline{c}_h$, rather than $\overline{z}_{h+1} = \overline{\chi}_{h+1}(\overline{p}_h, \overline{a}_h, \overline{o}_{h+1})$. □

Lemma .16. Let $\widehat{g}_{1:\overline{H}}^*$ be the strategy output by Algorithm 6, then for any $h \in [\overline{H}], \overline{c}_h \in \overline{\mathcal{C}}_h, \overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}$, it holds that

$$V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) \leq V_h^{\widehat{g}_{1:\overline{H}}^*,\mathcal{M}}(\overline{c}_h). \tag{.14}$$

Proof. We prove it by induction. For $h = \overline{H} + 1$, we have $V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = 0$.
For the step $h \leq H$, we have

$$\begin{aligned}
V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) &= \mathbb{E}^{\mathcal{M}}[\widehat{r}_h^{\mathcal{M}}(\widehat{c}_h) + V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1}) \,|\, \widehat{c}_h, \overline{g}_{1:\overline{H}}] \\
&\leq \mathbb{E}^{\mathcal{M}}[\widehat{r}_h^{\mathcal{M}}(\widehat{c}_h) + V_{h+1}^{\widehat{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1}) \,|\, \widehat{c}_h, \overline{g}_{1:\overline{H}}] \\
&= Q_h^{\widehat{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h, \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h)\}_{j\in[n]}) \\
&\leq Q_h^{\widehat{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h, \{\overline{g}_{j,h}(\cdot \,|\, \overline{c}_h)\}_{j\in[n]}) \\
&= V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)
\end{aligned}$$

For the first inequality sign, we use the induction hypothesis. For the second inequality sign, we use the property of $argmax$ in algorithm and $V_h^{\widehat{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\widehat{c}_h)$. According to induction, we complete the proof. □

We go back to the proof of theorem. Let $\widehat{g}^*_{1:\overline{H}}$ be the solution output by Algorithm 6, then for any $\overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}, h \in [\overline{H}], \overline{c}_h \in \overline{\mathcal{C}}_h$, we have

$$\begin{aligned}
&\mathbb{E}_{\overline{g}_{1:\overline{H}}}^{\mathcal{D}'_{\mathcal{L}}}\left[V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h)\right] \\
&\quad= \mathbb{E}_{\overline{g}_{1:\overline{H}}}^{\mathcal{D}'_{\mathcal{L}}}\left[\left(V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)\right) + \left(V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h)\right)\right] \\
&\quad\leq \mathbb{E}_{\overline{g}_{1:\overline{H}}}^{\mathcal{D}'_{\mathcal{L}}}\left[\left(V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)\right) + \left(V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h)\right)\right] \qquad (.15) \\
&\quad\leq (\overline{H} - h + 1)\epsilon_r + \frac{(\overline{H} - h)(\overline{H} - h + 1)}{2}\epsilon_z + (\overline{H} - h + 1)\epsilon_r + \frac{(\overline{H} - h)(\overline{H} - h + 1)}{2}\epsilon_z \\
&\quad= 2(\overline{H} - h + 1)\epsilon_r + (\overline{H} - h)(\overline{H} - h + 1)\epsilon_z.
\end{aligned}$$

For the first inequality sign, we use Lemma .16. For the second inequality sign, we use Lemma .15. Then apply $h = 1$, we have $J_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) \leq J_{\mathcal{D}'_{\mathcal{L}}}(\widehat{g}^*_{1:\overline{H}}) + 2\overline{H}\epsilon_r + \overline{H}^2\epsilon_z$. This completes the proof of **Part II**.

Part III: If the baseline sharing of $\mathcal{L}$ is one of the 4 cases in §A, we can construct an expected-approximate common information model of $\mathcal{D}'_{\mathcal{L}}$.
We first define the following notion.

Definition .17. We say an expected approximate common information model $\mathcal{M}$ is consistent with some belief $\{\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h, \,|\, \widehat{c}_h)\}_{h\in\overline{H}}$ if it satisfies the following, for all $i \in [n], h \in [\overline{H}]$:

$$\begin{aligned}
&\mathbb{P}_h^{\mathcal{M},z}(\overline{z}_{h+1} \,|\, \widehat{c}_h, \gamma_h) \\
&= \sum_{\overline{s}_h,\overline{p}_h,\overline{a}_h,\overline{o}_{h+1}} \mathbb{1}[\overline{z}_{h+1} = \overline{\chi}_{h+1}(\overline{p}_h, \overline{a}_h, \overline{o}_{h+1}), \overline{a}_h = \gamma_h(\overline{p}_h)] \left(\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h, \,|\, \widehat{c}_h) \sum_{\overline{s}_{h+1}} \overline{\mathbb{T}}_h(\overline{s}_{h+1} \,|\, \overline{s}_h, \overline{a}_h)\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1} \,|\, \overline{s}_{h+1})\right) \\
&\widehat{\mathcal{R}}_h^{\mathcal{M}}(\widehat{c}_h, \gamma_h) = \sum_{\overline{s}_h,\overline{p}_h,\overline{a}_h} \mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \,|\, \widehat{c}_h)\mathbb{1}[\overline{a}_h = \gamma_h(\overline{p}_h)]\mathcal{R}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h).
\end{aligned}$$

We aim to bound the ϵ_r, ϵ_z using the following lemma.

Lemma .18. Given any belief $\{\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h\}_{h\in[\overline{H}]}$ of expected-approximate-common-information model $\mathcal{M}$, it holds that for any $h \in [\overline{H}], \overline{\mathcal{C}}_h, \gamma_h \in \Gamma_h$:

$$\begin{aligned}
||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot \,|\, \overline{c}_h, \gamma_h) - \mathbb{P}_h^{\mathcal{M},z}(\cdot \,|\, \widehat{c}_h, \gamma_h)||_1 &\leq ||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot, \cdot \,|\, \overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot, \cdot \,|\, \widehat{c}_h)||_1, \\
|\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h) \,|\, \overline{c}_h, \gamma_h] - \widehat{\mathcal{R}}_h^{\mathcal{M}}(\widehat{c}_h, \gamma_h)| &\leq ||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot, \cdot \,|\, \overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot, \cdot \,|\, \widehat{c}_h)||_1,
\end{aligned}$$

where $\widehat{c}_h$ =Compress$_h(\overline{c}_h)$.

Proof. Adapted from Lemma 3 in [15]. □

Then we define the belief states following the notation in [32], [15] as $\overline{\boldsymbol{b}}_1(\emptyset) = \mu_1$, $\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h}) = \mathbb{P}(\overline{s}_h = \cdot \,|\, \overline{o}_{1:h}, \overline{a}_{1:h-1}), \overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h-1}) = \mathbb{P}(\overline{s}_h = \cdot \,|\, \overline{o}_{1:h-1}, \overline{a}_{1:h-1})$, where $\overline{\boldsymbol{b}} \in \Delta(\mathcal{S})$. Also, we define the approximate belief state using the most recent L-step history, that

$$\begin{aligned}
\overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h-1}) &= \mathbb{P}(\overline{s}_h = \cdot \,|\, \overline{s}_{h-L} \sim \text{Unif}(\mathcal{S}), \overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h}) \\
\overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h-1}) &= \mathbb{P}(\overline{s}_h = \cdot \,|\, \overline{s}_{h-L} \sim \text{Unif}(\mathcal{S}), \overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h}).
\end{aligned}$$

Also, for any set $N \subseteq [n]$, we define $\overline{a}_{N,h} = \{\overline{a}_{i,h}\}_{i \in N}$, and the same for $\overline{o}_{N,h}$. And we can also define the belief states with part of the observations that for any $N \subseteq [n]$,

$$\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h-1}, \overline{o}_{N,h}) = \mathbb{P}(\overline{s}_h = \cdot \,|\, \overline{a}_{1:h-1}, \overline{o}_{1:h-1}, \overline{o}_{N,h})$$
$$\overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h-1}, \overline{o}_{N,h}) = \mathbb{P}_h(\overline{s}_h = \cdot \,|\, \overline{s}_{h-L} \sim \text{Unif}(\mathcal{S}), \overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h-1}, \overline{o}_{N,h}).$$

Then, we have the following lemma.

Lemma .19. There is a constant $C \geq 1$ such that the following holds. Given any LTC problem $\mathcal{L}$ satisfying Assumption III.1, and let $\mathcal{D}'_{\mathcal{L}}$ be the Dec-POMDP after reformulation and expansion. Let $\epsilon \geq 0$, fix a strategy $\overline{g}_{1:\overline{H}}$ and indices $1 \leq h - L < h - 1 \leq \overline{H}$. If $L \geq C\gamma^{-4}\log(\frac{S}{\epsilon})$, then the following set of propositions hold

$$\mathbb{E}_{\overline{a}_{1:h-1}, \overline{o}_{1:h} \sim \overline{g}_{1:\overline{H}}} ||\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h}) - \overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h})||_1 \leq \epsilon \tag{.16}$$
$$\mathbb{E}_{\overline{a}_{1:h-1}, \overline{o}_{1:h} \sim \overline{g}_{1:\overline{H}}} ||\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h-1}) - \overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h-1})||_1 \leq \epsilon \tag{.17}$$
$$\mathbb{E}_{\overline{a}_{1:h-1}, \overline{o}_{1:h} \sim \overline{g}_{1:\overline{H}}} ||\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h-1}, \overline{o}_{N,h}) - \overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h-1}, \overline{o}_{N,h})||_1 \leq \epsilon. \tag{.18}$$

Proof. Given any LTC problem $\mathcal{L}$, we can construct a Dec-POMDP $\check{\mathcal{D}}$ that the transition and observation functions of $\check{\mathcal{D}}$ are the same as $\mathcal{L}$. And the information of $\check{\mathcal{D}}$ is fully sharing. Since $\mathcal{D}'_{\mathcal{L}}$ is reformulated from $\mathcal{L}$, we have

$$\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h}) = \boldsymbol{b}_{\lfloor \frac{h+1}{2} \rfloor}(a_{1:\lfloor \frac{h-1}{2} \rfloor}, o_{1:\lfloor \frac{h+1}{2} \rfloor}) = \check{\boldsymbol{b}}_{\lfloor \frac{h+1}{2} \rfloor}(\check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{1:\lfloor \frac{h+1}{2} \rfloor})$$
$$\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h-1}) = \boldsymbol{b}_{\lfloor \frac{h+1}{2} \rfloor}(a_{1:\lfloor \frac{h-1}{2} \rfloor}, o_{1:\lfloor \frac{h}{2} \rfloor}) = \check{\boldsymbol{b}}_{\lfloor \frac{h+1}{2} \rfloor}(\check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{1:\lfloor \frac{h}{2} \rfloor}).$$

And for the approximate belief state, we have

$$\overline{b}'_{h+1}(\overline{a}_{h-L:h}, \overline{o}_{h-L+1:h}) = \boldsymbol{b}'_{\lfloor \frac{h+2}{2} \rfloor}(a_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h}{2} \rfloor}, o_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h+1}{2} \rfloor}) = \check{\boldsymbol{b}}'_{\lfloor \frac{h+2}{2} \rfloor}(\check{a}_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h}{2} \rfloor}, \check{o}_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h+1}{2} \rfloor})$$
$$\overline{b}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h}) = \boldsymbol{b}'_{\lfloor \frac{h+1}{2} \rfloor}(a_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h-1}{2} \rfloor}, o_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h+1}{2} \rfloor}) = \check{\boldsymbol{b}}'_{\lfloor \frac{h+1}{2} \rfloor}(\check{a}_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h}{2} \rfloor}).$$

Also, since for any $t \in [H], \overline{a}_{2t-1}$ are communication actions, $\overline{o}_{2t} = \emptyset$ is null, and $\overline{s}_{2t-1} = \overline{s}_{2t}$ always holds. Then we can write Eq. (.16) and Eq. (.17) as

$$\mathbb{E}_{\{\overline{a}_{2t}\}_{t=1}^{\lfloor \frac{h-1}{2} \rfloor}, \{\overline{o}_{2t-1}\}_{t=1}^{\lfloor \frac{h+1}{2} \rfloor} \sim \overline{g}_{1:\overline{H}}} ||\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h}) - \overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h})||_1 \leq \epsilon \tag{.19}$$
$$\mathbb{E}_{\{\overline{a}_{2t}\}_{t=1}^{\lfloor \frac{h-1}{2} \rfloor}, \{\overline{o}_{2t-1}\}_{t=1}^{\lfloor \frac{h+1}{2} \rfloor} \sim \overline{g}_{1:\overline{H}}} ||\overline{\boldsymbol{b}}_h(\overline{a}_{1:h-1}, \overline{o}_{1:h-1}) - \overline{\boldsymbol{b}}'_h(\overline{a}_{h-L:h-1}, \overline{o}_{h-L+1:h-1})||_1 \leq \epsilon. \tag{.20}$$

Since $\check{\mathcal{D}}$ has fully sharing, for any $i \in [n], h \in [\overline{H}]$ and information $\overline{\tau}_{i,h}, \overline{\tau}_{i,2h}$, we have $\sigma(\overline{\tau}_{i,h}) \subseteq \sigma(\check{\tau}_{i,\lfloor \frac{h+1}{2} \rfloor})$. Therefore, given any strategy $\overline{g}_{1:\overline{H}}$, we can construct a strategy $\check{g}_{1:H}$ such that, for any $\overline{a}_{1:h-1}, \overline{o}_{1:h}$

$$\mathbb{P}(\{\overline{a}_{2t}\}_{t=1}^{\lfloor \frac{h-1}{2} \rfloor}, \{\overline{o}_{2t-1}\}_{t=1}^{\lfloor \frac{h+1}{2} \rfloor} \,|\, \overline{g}_{1:\overline{H}}) = \mathbb{P}(\check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{1:\lfloor \frac{h+1}{2} \rfloor} \,|\, \check{g}_{1:H}).$$

Since $\check{\mathcal{D}}$ satisfies Assumption III.1, we can apply the Theorem 10 in [15] with $\check{g}_{1:H}$ to get the result that there is a constant $C_0 \geq 1$ that if $L' \geq C_0\gamma^{-4}\log(\frac{S}{\epsilon})$, the following holds

$$\mathbb{E}_{\check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{1:\lfloor \frac{h+1}{2} \rfloor} \sim \check{g}_{1:H}} ||\check{\boldsymbol{b}}_{\lfloor \frac{h+1}{2} \rfloor}(\check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{1:\lfloor \frac{h+1}{2} \rfloor}) - \check{\boldsymbol{b}}'_{\lfloor \frac{h+1}{2} \rfloor}(\check{a}_{\lfloor \frac{h}{2} \rfloor - L':\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{\lfloor \frac{h+1}{2} \rfloor - L'+1:\lfloor \frac{h+1}{2} \rfloor})||_1 \leq \epsilon \tag{.21}$$
$$\mathbb{E}_{\check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{1:\lfloor \frac{h+1}{2} \rfloor} \sim \check{g}_{1:H}} ||\check{\boldsymbol{b}}_{\lfloor \frac{h+1}{2} \rfloor}(\check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{1:\lfloor \frac{h}{2} \rfloor}) - \check{\boldsymbol{b}}'_{\lfloor \frac{h+1}{2} \rfloor}(\check{a}_{\lfloor \frac{h}{2} \rfloor - L':\lfloor \frac{h-1}{2} \rfloor}, \check{o}_{\lfloor \frac{h+1}{2} \rfloor - L'+1:\lfloor \frac{h}{2} \rfloor})||_1 \leq \epsilon. \tag{.22}$$

We choose $C = 3C_0, L = 2L' + 1$. If $L \geq C\gamma^{-4}\log(\frac{S}{\epsilon})$, there must have $L' \geq C_0\gamma^{-4}\log(\frac{S}{\epsilon})$. Therefore, we directly get Eq. (.19) and Eq. (.20).
For Eq. (.18), we cannot directly apply the Theorem 10 in [15], but we can slightly change the Equation(E.11) of Theorem 10 in [15] as

$$\mathbb{E}^{\mathcal{G}}_{a_{1:h-1}, o_{1:h} \sim \pi'} ||\boldsymbol{b}_h(a_{1:h-1}, o_{1:h-1}, o_{N,h}) - \boldsymbol{b}'_h(a_{h-L:h-1}, o_{h-L+1:h-1}, o_{N,h})||_1 \leq \epsilon. \tag{.23}$$

It still holds if posterior update $F^q(P : o_{1,h})$ is changed to $F^q(P : o_{N,h})$ when applying Lemma 9 in the proof of Theorem 10 of [15]. Therefore, we can use the same process to prove Eq. (.18) from Eq. (.23) as above, and this completes the proof. □

Then we can compress the common information using a finite-memory truncation. Here, we discuss case-by-case how to compress it for the 8 examples of QC LTC given in §A. After reformulation and strict expansion, Examples 5 and 6 will be the same as Example 1, and Examples 7 and 8 will be the same as Example 2. Therefore, here we just need to show how to compress the common information for the first 4 examples.

Example 1: Baseline sharing of $\mathcal{L}$ is one-step delayed sharing[Kaiqing: why don't we just refer to the]. Then, common information should be that for any $t \in [H], \overline{c}_{2t-1} = \{\overline{o}_{1:2t-2}, \overline{a}_{1:2t-2}\}, \overline{c}_{2t} = \{\overline{o}_{1:2t-2}, \overline{a}_{1:2t-1}, \overline{a}_{N,2t-1}\}, N \subseteq [n]$, where N is the set of agents choose to share their observations through additional sharing, and N can be infer from $\overline{c}_{2t}$. Then we have that $\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t-1}(\overline{s}_{2t-1}, \overline{p}_{2t-1} \,|\, \overline{c}_{2t-1}) = \overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-2}, \overline{o}_{1:2t-2})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1} \,|\, \overline{s}_{2t-1})$. Fix compress length $L > 0$, we define the approximate common information as $\widehat{c}_{2t-1} = \{\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}\}$, and the common information conditioned belief as $\mathbb{P}^{\mathcal{M},c}_{2t-1}(\overline{s}_{2t-1}, \overline{p}_{2t-1} \,|\, \widehat{c}_{2t-1}) = \overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1} \,|\, \overline{s}_{2t-1})$. Also, we have $\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t}(\overline{s}_{2t}, \overline{p}_{2t} \,|\, \overline{c}_{2t}) = \overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-1}, \overline{o}_{1:2t-2}, \overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{N,2t-1})$. Fix compress length $L > 0$, we define the approximate common information a $\widehat{c}_{2t} = \{\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{N,2t-1}\}$, and the common information conditioned belief as $\mathbb{P}^{\mathcal{M},c}_{2t}(\overline{s}_{2t}, \overline{p}_{2t} \,|\, \widehat{c}_{2t}) = \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{N,2t-1})$, where $\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{N,2t-1}) = \frac{\overline{\mathbb{O}}_{2t-1}(\overline{o}_{N,2t-1}, \overline{o}_{-N,2t-1} \,|\, \overline{s}_{2t-1})}{\sum_{\overline{o}'_{-N,2t-1}} \overline{\mathbb{O}}_{2t-1}(\overline{o}_{N,2t-1}, \overline{o}'_{-N,2t-1} \,|\, \overline{s}_{2t-1})}$. Now, we need to verify that the Definition .9 is satisfied.

- The $\{\widehat{c}_h\}_{h \in [\overline{H}]}$ satisfied the Equation (.7) since for any $h \in [H]$, $\widehat{c}_{h+1} \subseteq \widehat{c}_h \cup \overline{z}_h$.
- Note that for any $\overline{c}_{2t-1}$ and the corresponding $\widehat{c}_{2t-1}$ constructed above:

$$
\begin{aligned}
&||\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t-1}(\cdot, \cdot \,|\, \overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_{2t-1}(\cdot, \cdot \,|\, \widehat{c}_h)||_1 \\
&\quad = \sum_{\overline{s}_{2t-1}, \overline{o}_{2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-2}, \overline{o}_{1:2t-2})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1} \,|\, \overline{s}_{2t-1}) \\
&\qquad\qquad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-1})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1} \,|\, \overline{s}_{2t-1})| \\
&\quad = ||\overline{\boldsymbol{b}}_{2t-1}(\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-2}, \overline{o}_{1:2t-2}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-1})||_1.
\end{aligned}
$$

For any $\overline{c}_{2t}$ and the corresponding $\widehat{c}_{2t}$ constructed above:

$$
\begin{aligned}
&||\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t}(\cdot, \cdot \,|\, \overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_{2t}(\cdot, \cdot \,|\, \widehat{c}_h)|| \\
&\quad = \sum_{\overline{s}_{2t-1}, \overline{o}_{-N,2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-1}, \overline{o}_{1:2t-2}, \overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{N,2t-1}) \\
&\qquad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{N,2t-1})| \\
&\quad = ||\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-1}, \overline{o}_{1:2t-2}, \overline{o}_{N,2t-1}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{N,2t-1})||_1.
\end{aligned}
$$

If we choose $L \geq C\gamma^{-4}\log(\frac{S}{\epsilon})$, for any $h \in [\overline{H}]$

$$
\mathbb{E}_{\overline{a}_{1:h-1}, o_{1:h} \sim \overline{g}_{1:\overline{H}}} ||\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_h(\cdot, \cdot \,|\, \overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_h(\cdot, \cdot \,|\, \widehat{c}_h)||_1 \leq \epsilon.
$$

Therefore, such a model is an ϵ-expected-approximate common information model.

Example 2: Baseline sharing of $\mathcal{L}$ is one-directional-one-delayed sharing. Then, common information common information should be that for any $t \in [H], \overline{c}_{2t-1} = \{\overline{o}_{1:2t-2}, \overline{a}_{1:2t-2}, \overline{o}_{1:2t-1}\}, \overline{c}_{2t} = \{\overline{o}_{1:2t-2}, \overline{a}_{1:2t-1}, \overline{a}_{N,2t-1}\}, N \subseteq [n], 1 \in N$. Here N is the same as defined in case 1, but it must satisfies that $1 \in N$. Then we similarly as case 1, we construct $\widehat{c}_{2t-1} = \{\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-L-1:2t-2}, \overline{o}_{1:2t-1}\}, \widehat{c}_{2t} = \{\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{N,2t-1}\}$, and approximate common information conditioned belief as $\mathbb{P}^{\mathcal{M},c}_{2t-1}(\overline{s}_{2t-1}, \overline{p}_{2t-1} \,|\, \widehat{c}_{2t-1}) = \overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{1,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-1,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{1,2t-1}), \mathbb{P}^{\mathcal{M},c}_{2t}(\overline{s}_{2t}, \overline{p}_{2t} \,|\, \widehat{c}_{2t}) = \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{N,2t-1})$. Now, we need to verify Definition .9 is satisfied.

- The $\{\widehat{c}_h\}_{h \in [\overline{H}]}$ satisfies the Equation (.7) since for any $h \in [H]$, $\widehat{c}_{h+1} \subseteq \widehat{c}_h \cup \overline{z}_h$.
- Note that for any $\overline{c}_{2t-1}$ and the corresponding $\widehat{c}_{2t-1}$ constructed above:

$$
\begin{aligned}
&||\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t-1}(\cdot, \cdot \,|\, \overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_{2t-1}(\cdot, \cdot \,|\, \widehat{c}_h)||_1 \\
&\quad = \sum_{\overline{s}_{2t-1}, \overline{o}_{-1,2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-1}, \overline{o}_{1:2t-2}, \overline{o}_{1,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-1,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{1,2t-1}) \\
&\quad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{1,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-1,2t-1} \,|\, \overline{s}_{2t-1}, \overline{o}_{1,2t-1})| \\
&\quad = ||\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-1}, \overline{o}_{1:2t-2}, \overline{o}_{1,2t-1}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2}, \overline{o}_{2t-L:2t-2}, \overline{o}_{1,2t-1})||_1.
\end{aligned}
$$

For any $\overline{c}_{2t}$ and the corresponding $\widehat{c}_{2t}$ constructed above:

$$\begin{aligned}
&||\mathbb{P}_{2t}^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot\,|\,\overline{c}_h) - \mathbb{P}_{2t}^{\mathcal{M},c}(\cdot,\cdot\,|\,\widehat{c}_h)||_1 \\
&\quad = \sum_{\overline{s}_{2t-1},\overline{o}_{-N,2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-1},\overline{o}_{1:2t-2},\overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}\,|\,\overline{s}_{2t-1},\overline{o}_{N,2t-1}) \\
&\qquad\qquad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2},\overline{o}_{2t-L:2t-2},\overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}\,|\,\overline{s}_{2t-1},\overline{o}_{N,2t-1})| \\
&\quad = ||\overline{\boldsymbol{b}}_{2t-1}(\overline{a}_{1:2t-1},\overline{o}_{1:2t-2},\overline{o}_{N,2t-1}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{a}_{2t-1-L:2t-2},\overline{o}_{2t-L:2t-2},\overline{o}_{N,2t-1})||_1.
\end{aligned}$$

If we choose $L \geq C\gamma^{-4}\log(\frac{S}{\epsilon})$, then from Lemma .19 we have, for any $h \in [\overline{H}]$

$$\mathbb{E}_{\overline{a}_{1:h-1},o_{1:h}\sim\overline{g}_{1:\overline{H}}}||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot\,|\,\overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot\,|\,\widehat{c}_h)||_1 \leq \epsilon.$$

Therefore, such a model is an ϵ-expected-approximate common information model.

Example 3: Baseline sharing of $\mathcal{L}$ is state controlled by one controller with asymmetric delay sharing. Then the common information should be that, for any $h \in [\overline{H}], \overline{c}_h = \{\overline{o}_{1:h-2d}, \overline{a}_{1,1:h-1}, \{\overline{a}_{-1,2t-1}\}_{t=\lfloor\frac{h-2d+1}{2}\rfloor}^{\lfloor\frac{h}{2}\rfloor}, \overline{o}_{1,h-2d+1:h}, \overline{o}_M\}$, where $M \subset \{(i,t)\,|\,1 < i \leq n, h-2d+1 \leq t \leq h\}$ and $\overline{o}_M = \{o_{i,t}\,|\,(i,t) \in M\}$, and corresponding $\overline{p}_h = \{\overline{o}_{i,t}\,|\,1 < i \leq n, h-2d < t \leq h, (i,t) \notin M\}$. Actually, $\overline{o}_M$ are the observations shared by the additional sharing in $\mathcal{L}$. Denote $f_{\tau,h-2d} = \{\overline{a}_{1:h-2d-1}, \overline{o}_{h-2d}, \{\overline{a}_{-1,2t-1}\}_{t=\lfloor\frac{h-2d+1}{2}\rfloor}^{\lfloor\frac{h}{2}\rfloor}\}, f_a = \{\overline{a}_{1,h-2d:h-1}\}, f_o = \{\overline{o}_{1,h-2d+1:h}, \overline{o}_M\}$. We can compute the common-information-based belief as

$$\begin{aligned}
\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{c}_h) &= \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{s}_{h-2d},f_a,f_o)\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}\,|\,f_{\tau,h-2d},f_a,f_o) \\
&= \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{s}_{h-2d},f_a,f_o)\frac{\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d},f_a,f_o\,|\,f_{\tau,h-2d})}{\sum_{\overline{s}'_{h-2d}}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}'_{h-2d},f_a,f_o\,|\,f_{\tau,h-2d})}.
\end{aligned}$$

Denote the probability $P_h(f_o\,|\,\overline{s}_{h-2d},f_a) := \Pi_{t=1}^{2d}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{o}_{1,h-2d+t},\overline{o}_{M_{h-2d+t}}\,|\,\overline{s}_{h-2d},\overline{a}_{1,h-2d:h-2d+t})$, where $M_{h-2d+t} = \{(i,h-2d+t)\,|\,(i,h-2d+t) \in M\}$ denotes the set of observations at timestep $h-2d+t$ and shared through additional sharing. With such notation, we have

$$\begin{aligned}
\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}\,|\,f_{\tau,h-2d},f_a,f_o) =& \frac{\overline{\boldsymbol{b}}_{h-2d}(\overline{a}_{1:h-2d-1},\overline{o}_{1:h-2d})(\overline{s}_{h-2d})P_h(f_o\,|\,\overline{s}_{h-2d},f_a)}{\sum_{\overline{s}'_{h-2d}}\overline{\boldsymbol{b}}_{h-2d}(\overline{a}_{1:h-2d-1},\overline{o}_{1:h-2d})(\overline{s}'_{h-2d})P_h(f_o\,|\,\overline{s}'_{h-2d},f_a)} \\
=& F^{P_h(\cdot\,|\,\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{a}_{1:h-2d-1},\overline{o}_{1:h-2d});f_o)(\overline{s}_{h-2d}),
\end{aligned}$$

where $F^{P_h(\cdot\,|\,\cdot,f_a)}(\cdot;f_o) : \Delta(\mathcal{S}) \rightarrow \Delta(\mathcal{S})$ is a posterior belief update function. The formal definition is shown in Lemma 9 in [15].
Then, we define the approximate common information as $\widehat{c}_h := \{\overline{o}_{1,h-2d-L+1:h}, \overline{a}_{1,h-2d-L:h-1}, \overline{o}_M\}$ and corresponding approximate common information conditioned belief as

$$\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h,\overline{p}_h\,|\,\widehat{c}_h) = \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{s}_{h-2d},f_a,f_o)F^{P_h(\cdot\,|\,\cdot,f_a)}(\overline{\boldsymbol{b}}'_{h-2d}(\overline{a}_{h-2d-L:h-2d-1},\overline{o}_{h-2d-L+1:h-2d});f_o)(\overline{s}_{h-2d}).$$

Now we verify that the Definition .9 is satisfied.

- Obviously, the $\{\widehat{c}_h\}_{h\in[\overline{H}}$ satisfies Eq.(.7)
- For any $\overline{c}_h$ and corresponding $\widehat{c}_h$ constructed above:

$$\begin{aligned}
&||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot\,|\,\overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot\,|\,\widehat{c}_h)||_1 \\
&\quad \leq ||F^{P(\cdot\,|\,\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{a}_{1:h-2d-1},\overline{o}_{1:h-2d});f_o) - F^{P(\cdot\,|\,\cdot,f_a)}(\overline{\boldsymbol{b}}'_{h-2d}(\overline{a}_{h-2d-L:h-2d-1},\overline{o}_{h-2d-L+1:h-2d});f_o)||_1
\end{aligned}$$

If we choose $L \geq C\gamma^{-4}\log(\frac{S}{\epsilon})$, then for any strategy $\overline{g}_{1:\overline{H}}$ taking expectations over $f_{\tau,h-2d}, f_a, f_o$, from Lemma .19 and Lemma 9 in [15], we have, for any $h \in [\overline{H}]$

$$\mathbb{E}_{\overline{a}_{1:h-1},o_{1:h}\sim\overline{g}_{1:\overline{H}}}||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot\,|\,\overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot\,|\,\widehat{c}_h)||_1 \leq \epsilon.$$

Therefore, such a model is an ϵ-expected-approximate common information model.

Example 4: Baseline sharing of $\mathcal{L}$ is Uncontrolled state process with delayed sharing. Then, the common information should be that, for any $h \in [H], \widehat{c}_h = \{\overline{o}_{1:h-2d}, \{\overline{a}_{2t-1}\}_{t=1}^{\lfloor \frac{h}{2} \rfloor}, \overline{o}_M\}$, where $M = \{(i,t) \mid i \in [n], h-2d+1 \leq t \leq h\}$. Then, still we denote $f_{\tau,h-2d} = \{\overline{o}_{1:h-2d}, \{\overline{a}_{2t-1}\}_{t=1}^{\lfloor \frac{h}{2} \rfloor}\}, f_o = \{\overline{o}_M\}$. We can compute the common information-based belief as

$$\begin{aligned}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \mid \overline{c}_h) &= \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \mid \overline{s}_{h-2d}, f_o)\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d} \mid f_{\tau,h-2d}, f_o) \\ &= \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \mid \overline{s}_{h-2d}, f_o)\frac{\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}, f_o \mid f_{\tau,h-2d})}{\sum_{\overline{s}'_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}'_{h-2d}, f_o \mid f_{\tau,h-2d})}\end{aligned}$$

Denote the probability $P_h(f_o \mid \overline{s}_{h-2d}) := \Pi_{t=1}^{2d}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{o}_{1,h-2d+t}, \overline{o}_{M_{h-2d+t}} \mid \overline{s}_{h-2d})$, where $M_{h-2d+t} = \{(i, h-2d+t) \mid (i, h-2d+t) \in M\}$ denotes the set of observations at timestep $h-2d+t$ and shared through additional sharing. Since the actions do not influence underlying states, here we use the belief notation $\overline{\boldsymbol{b}}_k(\overline{o}_{1:k}), \overline{\boldsymbol{b}}_k(\overline{o}_{k-L:k}) \forall k \in [\overline{H}], L < k$. With such notation, we have

$$\begin{aligned}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d} \mid f_{\tau,h-2d}, f_o) =& \frac{\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d})(\overline{s}_{h-2d})P_h(f_o \mid \overline{s}_{h-2d})}{\sum_{\overline{s}'_{h-2d}} \overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d})(\overline{s}'_{h-2d})P_h(f_o \mid \overline{s}'_{h-2d})} \\ =& F^{P_h(\cdot \mid \cdot)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d}); f_o)(\overline{s}_{h-2d}),\end{aligned}$$

where $F^{P_h(\cdot \mid \cdot)}(\cdot; f_o) : \Delta(\mathcal{S}) \to \Delta(\mathcal{S})$ is a posterior belief update function the same as case 3.
Then, we define the approximate common information as $\widehat{c}_h := \{\overline{o}_{h-2d-L+1:h}, \overline{o}_M\}$ and corresponding approximate common information conditioned belief as

$$\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \mid \widehat{c}_h) = \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h \mid \overline{s}_{h-2d}, f_o)F^{P_h(\cdot \mid \cdot)}(\overline{\boldsymbol{b}}'_{h-2d}(\overline{o}_{h-2d-L+1:h-2d}); f_o)(\overline{s}_{h-2d}).$$

Now we verify that Definition .9 is satisfied.

- Obviously, the $\{\widehat{c}_h\}_{h \in [\overline{H}}$ satisfies Eq.(.7)
- For any $\overline{c}_h$ and corresponding $\widehat{c}_h$ constructed above:

$$\begin{aligned}&||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot, \cdot \mid \overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot, \cdot \mid \widehat{c}_h)||_1 \\ &\quad \leq ||F^{P(\cdot \mid \cdot)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d}); f_o) - F^{P(\cdot \mid \cdot)}(\overline{\boldsymbol{b}}'_{h-2d}(\overline{a}_{h-2d-L:h-2d-1}, \overline{o}_{h-2d-L+1:h-2d}); f_o)||_1\end{aligned}$$

  If we choose $L \geq C\gamma^{-4}\log(\frac{S}{\epsilon})$, then for any strategy $\overline{g}_{1:\overline{H}}$ taking expectations over $f_{\tau,h-2d}, f_o$, from Lemma .19 and Lemma 9 in [15], we have, for any $h \in [\overline{H}]$

$$\mathbb{E}_{\overline{a}_{1:h-1}, o_{1:h} \sim \overline{g}_{1:\overline{H}}} ||\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot, \cdot \mid \overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot, \cdot \mid \widehat{c}_h)||_1 \leq \epsilon.$$

Therefore, such a model is an ϵ-expected-approximate common information model.

Combining **Parts I, II, III**, we complete the proof. □

We introduce the notion of *perfect recall* [23]:

Definition .20 (Perfect recall)**.** We say that agent i has perfect recall if $\forall h \in 2, \cdots, H$, it holds that $\tau_{i,h-1} \cup \{a_{i,h-1}\} \subseteq \tau_{i,h}$. If for any $i \in [n]$, agent i has perfect recall, we call that the Dec-POMDP has a perfect recall property.

12) Proof of Theorem V.1:

Proof. sQC $\Rightarrow$ SI:
Let $\mathcal{D}$ be the Dec-POMDP with sQC information structure. Then, we need to prove $\mathcal{D}$ is SI, i.e., for any $\forall h_1 \in [h-1], i_1 \in [n], \overline{g}_{1:h-1} \in \overline{\mathcal{G}}_{1:h-1}, \overline{g}'_{i_1,h_1} \in \overline{\mathcal{G}}_{i_1,h_1}$, let $\overline{g}'_{1:h-1} = (\overline{g}_{1,1}, \cdots, \overline{g}_{i_1-1,h_1}, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h-1})$, the following holds

$$\mathbb{P}(\overline{s}_h, \overline{p}_h \mid \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \mid \overline{c}_h, \overline{g}'_{1:h-1}). \tag{.24}$$

If there exists $i_3 \neq i_1$ such that agent (i_1, h_1) influences agent (i_3, h), then from $\mathcal{D}$ is sQC, we know that $\sigma(\overline{\tau}_{i_1,h_1}) \cup \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_3,h})$. And then from Assumption III.9, we know that $\sigma(\overline{\tau}_{i_1,h_1}) \cup \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h)$. Then, $\exists$ a unique $\overline{\tau}'_{i_1,h_1}$ and a unique $\overline{a}_{i_1,h_1}$ that $\mathbb{P}(\overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1} \mid \overline{c}_h) = 1$, then

$$\begin{aligned}\mathbb{P}(\overline{s}_h, \overline{p}_h \mid \overline{c}_h, \overline{g}_{1:h-1}) &= \sum_{\substack{\overline{\tau}_{i_1,h_1} \in \overline{\mathcal{T}}_{i_1,h_1} \\ \overline{a}_{i_1,h_1} \in \overline{\mathcal{A}}_{i_1,h_1}}} \mathbb{P}(\overline{s}_h, \overline{p}_h, \overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1} \mid \overline{c}_h, \overline{g}_{1:h-1}) \\ &= \mathbb{P}(\overline{s}_h, \overline{p}_h, \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1} \mid \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \mid \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1}, c_h, g_{1:h-1})\mathbb{P}(\overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1} \mid \overline{c}_h, \overline{g}_{1:h-1}) \\ &= \mathbb{P}(\overline{s}_h, \overline{p}_h \mid \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \mid \overline{\tau}'_{i_1,h_1}, \overline{a}'_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1} \backslash \overline{g}_{i_1,h_1}).\end{aligned}$$

The last equality is because the input and output of $\overline{g}_{i_1:h_1}$ are $\overline{\tau}'_{i_1,h_1}$ and $\overline{a}'_{i_1,h_1}$. Then we know $\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1})$.

If for any $i_2 \neq i_1$ that $\sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h})$ or $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h})$, then agent (i_1, h_1) does not influence agent $(i_2, h), \forall i_2 \neq i_1$ because $\mathcal{D}$ is sQC. Firstly, agent (i_1, h_1) does not influence $\overline{s}_{h_1+1}$, otherwise from Assumption III.7, there exists $i_3 \neq i_1$ that agent (i_1, h_1) influences $\overline{o}_{i_3,h_1+1}$. Since $\overline{o}_{i_3,h_1+1} \in \overline{\tau}_{i_3,h}$, we know that agent (i_1, h_1) influences agent (i_3, h), contradiction! Therefore, for any $h_2 > h_1$, agent (i_1, h_1) does not influence $\overline{s}_{h_2}$. Also, from Assumption III.5, for any $h_2 > h_1, \overline{a}_{i_1,h_1} \notin \overline{\tau}_{h_2}$. Therefore, for any $i_2 \in [n], h_2 > h_1$, agent (i_1, h_1) does not influence the $\overline{\tau}_{i_2,h_2}$ and $\overline{a}_{i_2,h_2}$. Therefore, agent (i_1, h_1) does not influence $\overline{c}_h = \cap_{i_2=1}^n \overline{\tau}_{i_2,h}$. Finally, we can conclude

$$\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}).$$

SI $\Rightarrow$ sQC:

Let $\mathcal{D}$ be a Dec-POMDP with perfect recall, and $\mathcal{D}$ is strategy independent, which means $\forall i \in [n], h \in [\overline{H}], \forall \overline{g}_{1:h-1}, \overline{g}'_{1:h-1} \in \overline{\mathcal{G}}_{1:h-1}, \overline{c}_h \in \overline{\mathcal{C}}_h, \overline{s}_h \in \overline{\mathcal{S}}, \overline{p}_h \in \overline{\mathcal{P}}_h$, the following holds

$$\mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h \,|\, \overline{c}_h, \overline{g}'_{1:h-1}).$$

If $\mathcal{D}$ is not strictly QC, then $\exists\ i_1 \in [n], h_1 \in [H]$, agent (i_1, h_1) influences agent (i_2, h_2), but $\sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h_2})$ or $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h_2})$. Since $\mathcal{D}$ has perfect recall, we can assume $i_2 \neq i_1$.

We know $\exists \overline{g}_{1:h_2-1} \in \overline{\mathcal{G}}_{1:h_2-1}, \overline{c}_{h_2} \in \overline{\mathcal{C}}_{h_2}, \overline{p}_{h_2} \in \overline{\mathcal{P}}_{h_2}, \overline{s}_{h_2} \in \overline{\mathcal{S}}$ satisfies

$$\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \,|\, \overline{c}_{h_2}, \overline{g}_{1:h_2-1}) = \mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \,|\, \overline{c}_{h_2}) \neq 0. \tag{.25}$$

Then, we prove $\mathcal{D}$ is sQC by discussing several different cases.

1) $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{c}_{h_2})$:
Then, there is at least another different action $\overline{a}'_{i_1,h_1}$ such that

$$\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \backslash \{\overline{a}_{i_1,h_1}\} \cup \{\overline{a}'_{i_1,h_1}\} \,|\, \overline{c}_{h_2}) \neq 0. \tag{.26}$$

Then, consider another strategy $\overline{g}'_{i_1,h_1}$ such that

$$\forall \overline{\tau}_{i_1,h_1} \in \overline{\mathcal{T}}_{i_1,h_1}, \overline{g}'_{i_1,h_1}(\overline{a}'_{i_1,h_1} \,|\, \overline{\tau}_{i_1,h_1}) = 1. \tag{.27}$$

Let $\overline{g}'_{1:h_2-1} = (\overline{g}_{1,1}, \cdots, \overline{g}_{i_1-1,h_1}, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h_2-1})$, then we get

$$\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \,|\, \overline{c}_{h_2}, \overline{g}'_{1:h_2-1}) = 0 \neq \mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \,|\, \overline{c}_{h_2}, \overline{g}_{1:h_2-1}), \tag{.28}$$

which leads to a contradiction.

2) $\sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{c}_{h_2})$, then we know $\sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h_2})$. Let $\tau = \overline{\tau}_{i_1,h_1} \backslash \overline{c}_{h_2}$. Still, there is at least another $\overline{\tau}'_{i_1,h_1}$ and $\tau' = \overline{\tau}'_{i_1,h_1} \backslash \overline{c}_{h_2}$ that

$$\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \backslash \{\tau\} \cup \{\tau'\} \,|\, \overline{c}_{h_2}) \neq 0. \tag{.29}$$

Otherwise, we can infer τ from $\overline{c}_{h_2}$, then $\sigma(\overline{\tau}_{i_1,h_1}) \in \sigma(\overline{c}_{h_2})$. Then we consider another strategy $\overline{g}'_{i_1,h_1}$ that

$$\overline{g}'_{i_1,h_1}(\overline{a}_{i_1,h_1} \,|\, \overline{\tau}_{i_1,h_1}) = 0, \overline{g}'_{i_1,h_1}(\overline{a}_{i_1,h_1} \,|\, \overline{\tau}'_{i_1,h_1}) = 1. \tag{.30}$$

Let $\overline{g}'_{1:h_2-1} = (\overline{g}_{1,1}, \cdots, \overline{g}_{i_1-1,h_1}, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h_2-1})$, then we get

$$\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \,|\, \overline{c}_{h_2}, \overline{g}'_{1:h_2-1}) = 0 \neq \mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \,|\, \overline{c}_{h_2}, \overline{g}_{1:h_2-1}). \tag{.31}$$

This completes the proof. □

C. Collections of Algorithm Pseudocodes

Here we collect both our planning and learning algorithms as pseudocode in Algorithms 1, 2, 3, 4, 5, and 6.

Algorithm 1 Planning in QC LTC Problems

```
Require: LTC $\mathcal{L}$, accuracy level $\epsilon_r, \epsilon_z > 0$
  Reformulate $\mathcal{L}$ to $\mathcal{D}_\mathcal{L}$ based on Eq. (IV.1).
  Expand $\mathcal{D}_\mathcal{L}$ to $\mathcal{D}_\mathcal{L}^\dagger$ based on Eq. (IV.2).
  Refine $\mathcal{D}_\mathcal{L}^\dagger$ to $\mathcal{D}_\mathcal{L}'$ based on $\mathcal{L}$ and Eq. (IV.4).
  Construct expected Approximate Common-information Model [Kaiqing: hard to understand if we never defined what this is, and how to construct?] $\mathcal{M}$ from $\mathcal{D}_\mathcal{L}'$ with error $\epsilon_r, \epsilon_z$.
  $\overline{g}^*_{1:\widetilde{H}} \leftarrow$ Algorithm [Kaiqing: worried that this may not be seen if u truncate the appendix – how about just calling the alg in OUR paper? can comment by adding this Algorithm 6 here.] 6($\mathcal{M}$)
  $\widetilde{g}^*_{1:\widetilde{H}} \leftarrow \varphi(\overline{g}^*_{1:\widetilde{H}}, \mathcal{D}_\mathcal{L})$
  $g^{m,*}_{1:H} \leftarrow \{\widetilde{g}^*_1, \widetilde{g}^*_3, \cdots, \widetilde{g}^*_{2H-1}\}$
  $g^{a,*}_{1:H} \leftarrow \{\widetilde{g}_2, \widetilde{g}_4, \cdots, \widetilde{g}_{2H}\}$
  return $(g^{m,*}_{1:H}, g^{a,*}_{1:H})$
```

Algorithm 2 Learning in QC LTC Problems

```
Require: Underlying environment LTC $\mathcal{L}$, iterations $K$.
  Reformulate $\mathcal{L}$ to $\mathcal{D}_\mathcal{L}$ based on Eq. (IV.1).
  Refine $\mathcal{D}_\mathcal{L}$ to $\mathcal{D}_\mathcal{L}'$ based on Eq. (IV.2).
  Obtain $\{\overline{g}^j_{1:\overline{H}}\}_{j=1}^K$ by calling Algorithm 3 of [33].
  for $j = 1$ to $K$ do
    Construct $\widehat{\mathcal{M}}(\overline{g}^j_{1:\overline{H}})$ by calling Algorithm 5 of [15] with the underlying environment $\mathcal{D}_\mathcal{L}'$ and $\overline{g}^j_{1:\overline{H}}$.
    $\overline{g}^{j,*}_{1:\overline{H}} \leftarrow$ Algorithm 6($\widehat{\mathcal{M}}(\overline{g}^j_{1:\overline{H}})$)
  end for
  $\overline{g}^*_{1:\widetilde{H}} \leftarrow$ Algorithm 8($\{\overline{g}^{j,*}_{1:\overline{H}}\}_{j=1}^K$) of [15].
  $\widetilde{g}^*_{1:\widetilde{H}} \leftarrow \varphi(\overline{g}^*_{1:\overline{H}}, \mathcal{D}_\mathcal{L})$
  $g^{m,*}_{1:H} \leftarrow \{\widetilde{g}^*_1, \widetilde{g}^*_3, \cdots, \widetilde{g}^*_{2H-1}\}$
  $g^{a,*}_{1:H} \leftarrow \{\widetilde{g}_2, \widetilde{g}_4, \cdots, \widetilde{g}_{2H}\}$
  return $(g^{m,*}_{1:H}, g^{a,*}_{1:H})$
```

Algorithm 3 Vanilla Realization of $\varphi(\breve{g}_{1:\breve{H}}, \mathcal{D}_\mathcal{L})$

```
Require: Strategy $\breve{g}_{1:\breve{H}}$, QC Dec-POMDP $\mathcal{D}_\mathcal{L}$
  $\widetilde{g}_{1:\breve{H}} \leftarrow \emptyset$
  for $h_2 = 1$ to $\breve{H}$, $i_2 = 1$ to $n$, $\widetilde{\tau}_{i_2,h_2} \in \widetilde{\mathcal{T}}_{i_2,h_2}$ do
    $\breve{\tau}_{i_2,h_2} \leftarrow \widetilde{\tau}_{i_2,h_2}$
    for $h_1 = 1$ to $h_2 - 1$, $i_1 = 1$ to $n$ do
      if $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$ in $\mathcal{D}_\mathcal{L}$ then
        $\widetilde{a}_{i_1,h_1} \leftarrow \widetilde{g}_{i_1,h_1}(\widetilde{\tau}_{i_1,h_1})$
        $\breve{\tau}_{i_2,h_2} \leftarrow \breve{\tau}_{i_2,h_2} \cup \{\widetilde{a}_{i_1,h_1}\}$
      end if
    end for
    $\widetilde{g}_{i_2,h_2}(\widetilde{\tau}_{i_2,h_2}) \leftarrow \breve{g}_{i_2,h_2}(\breve{\tau}_{i_2,h_2})$
  end for
  return $\widetilde{g}_{1:\widetilde{H}}$
```

Algorithm 4 Efficient Implementation of $\varphi(\breve{g}_{1:\breve{H}}, \mathcal{D}_\mathcal{L})$

```
Require: Strategy $\breve{g}_{1:\breve{H}}$, QC Dec-POMDP $\mathcal{D}_\mathcal{L}$
  for $h = 1$ to $\breve{H}$ do
    for $i = 1$ to $n$ do
      Agent $i$ receives $\widetilde{\tau}_{i,h}$
      $\breve{\tau}_{i,h} \leftarrow$ Recover($\widetilde{\tau}_{i,h}, \breve{g}_{1:h-1}, \mathcal{D}_\mathcal{L}$) \\ Defined in Algorithm 5
      Agent $i$ chooses $\breve{g}_{i,h}(\breve{\tau}_{i,h})$ as $\widetilde{a}_{i,h}$
    end for
  end for
```

Algorithm 5 Recover $\breve{\tau}_{i,h}$ from $\widetilde{\tau}_{i,h}$

```
Require: Information $\widetilde{\tau}_{i,h}$, Strategy $\breve{g}_{1:h-1}$, QC Dec-POMDP $\mathcal{D}_{\mathcal{L}}$
  $\breve{\tau}_{i,h} \leftarrow \widetilde{\tau}_{i,h}$
  for $j = 1$ to $n$, $h' = 1$ to $h-1$ do
    if $\sigma(\widetilde{\tau}_{j,h'}) \subseteq \sigma(\widetilde{c}_h)$ in $\mathcal{D}_{\mathcal{L}}$ and $\widetilde{a}_{j,h'} \notin \breve{\tau}_{i,h}$ then
      $\breve{\tau}_{j,h'} \leftarrow$ Recover$(\widetilde{\tau}_{j,h'}, \breve{g}_{1:h'-1}, \mathcal{D}_{\mathcal{L}})$
      $\widetilde{a}_{j,h'} \leftarrow \breve{g}_{j,h'}(\breve{\tau}_{j,h})$
      $\breve{\tau}_{i,h} \leftarrow \breve{\tau}_{j,h} \cup \{\widetilde{a}_{j,h'}\}$
    end if
  end for
  return $\breve{\tau}_{i,h}$
```

Algorithm 6 Planning in Dec-POMDP with expected Approximate Common-information Model

```
Require: Expected Approximate Common-information Model $\mathcal{M}$.
  for $i \in [n]$ and $\widehat{c}_{\overline{H}+1} \in \widehat{\mathcal{C}}_{\overline{H}+1}$ do
    $V^{*,\mathcal{M}}_{i,\overline{H}+1}(\widehat{c}_{\overline{H}+1}) \leftarrow 0$
  end for
  for $h = \overline{H}$ to $1$ do
    for $\widehat{c}_h \in \widehat{\mathcal{C}}_h$ do
      Define $Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h}) := \widehat{\mathcal{R}}^{\mathcal{M}}_h(\widehat{c}_h, \gamma_h) + \mathbb{E}^{\mathcal{M}}\left[V^{*,\mathcal{M}}_{h+1}(\widehat{c}_{h+1}) \,|\, \widehat{c}_h, \gamma_h\right]$
          $\left(\widehat{g}^*_{1,h}(\cdot \,|\, \widehat{c}_h, \cdot), \cdots, \widehat{g}^*_{n,h}(\cdot \,|\, \widehat{c}_h, \cdot)\right) \leftarrow \underset{\gamma_{1:n,h} \in \Gamma_h}{\operatorname{argmax}}\, Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h})$     (.32)
    end for
    $V^{*,\mathcal{M}}_h(\widehat{c}_h) \leftarrow \max_{\gamma_{1:n,h}} Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h})$
  end for
  return $\widehat{g}^*_{1:\overline{H}}$
```

D. Decentralized POMDPs (with Information Sharing)

A Dec-POMDP with n agents and potential information sharing can be characterized by a tuple

$$\mathcal{D} = \langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathbb{T}_h\}_{h\in[H]}, \{\mathbb{O}_h\}_{h\in[H]}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}\rangle,$$

where H denotes the length of each episode, $\mathcal{S}$ denotes state space, and $\mathcal{A}_{i,h}$ denotes the *control action* spaces of agent i at timestep h. We denote by $s_h \in \mathcal{S}$ the state and by $a_{i,h}$ the control action of agent i at timestep h. We use $a_h := (a_{1,h}, \cdots, a_{n,h}) \in \mathcal{A}_h := \mathcal{A}_{1,h} \times \mathcal{A}_{2,h} \times \cdots \mathcal{A}_{n,h}$ to denote the joint control action for all the n agents at timestep h, with $\mathcal{A}_h$ denoting the joint control action space at timestep h. We denote $\mathbb{T} = \{\mathbb{T}_h\}_{h\in[H]}$ the collection of transition functions, where $\mathbb{T}_h(\cdot \,|\, s_h, a_h) \in \Delta(\mathcal{S})$ gives the transition probability to the next state s_{h+1} when taking the joint control action a_h at state s_h. We use $\mu_1 \in \Delta(\mathcal{S})$ to denote the distribution of the initial state s_1. We denote by $\mathcal{O}_{i,h}$ the observation space and by $o_{i,h} \in \mathcal{O}_{i,h}$ the observation of agent i at timestep h. We use $o_h := (o_{1,h}, o_{2,h}, \cdots, o_{n,h}) \in \mathcal{O}_h := \mathcal{O}_{1,h} \times \mathcal{O}_{2,h} \times \cdots \mathcal{O}_{n,h}$ to denote the joint observation of all the n agents at timestep h, with $\mathcal{O}_h$ denoting the joint observation space at timestep h. We use $\{\mathbb{O}_h\}_{h\in[H]}$ to denote the collection of emission matrices, where $o_h \sim \mathbb{O}_h(\cdot \,|\, s_h) \in \Delta(\mathcal{O}_h)$ at timestep h under state $s_h \in \mathcal{S}$. For notational convenience, we adopt the matrix convention, where $\mathbb{O}_h$ is a matrix with each row $\mathbb{O}_h(\cdot \,|\, s_h)$ for all $s_h \in \mathcal{S}$. Also, we denote by $\mathbb{O}_{i,h}$ the marginalized emission for agent i at timestep h. Finally, $\{\mathcal{R}_h\}_{h\in[H]}$ is a collection of reward functions among all agents, where $\mathcal{R}_h : \mathcal{S} \times \mathcal{A}_h \to [0,1]$.

At timestep h, each agent i in the Dec-POMDP has access to some information $\tau_{i,h}$, a subset of historical joint observations and actions, namely, $\tau_{i,h} \subseteq \{o_1, a_1, o_2, \cdots, a_{h-1}, o_h\}$, and the collection of all possible such available information is denoted by $\mathcal{T}_{i,h}$. We use τ_h to denote the *joint* available information at timestep h. Meanwhile, agents may *share* part of the history with each other. The *common information* $c_h = \cup_{t=1}^h z_t$ at timestep h is thus a subset of the joint history τ_h, where z_h is the information shared at timestep h. We use $\mathcal{C}_h$ to denote the collection of all possible c_h at timestep h, and use $\mathcal{T}_{i,h}$ to denote the collection of all possible $\tau_{i,h}$ of agent i at timestep h. Besides the common information c_h, each agent also has her *private information* $p_{i,h} = \tau_{i,h} \backslash c_h$, where the collection of $p_{i,h}$ is denoted by $\mathcal{P}_{i,h}$. We also denote by p_h the *joint* private information, and by $\mathcal{P}_h$ the collection of all possible p_h at timestep h. We refer to the above the *state-space model* of the Dec-POMDP (with information sharing).

1) Intrinsic Model: In an intrinsic model [24], we regard the agent i at different timesteps as *different agents*, and each agent only acts *once* throughout. Any Dec-POMDP $\mathcal{D}$ with n agents can be formulated within the intrinsic-model framework, and can be characterized by a tuple $\langle(\Omega, \mathscr{F}), N, \{(\mathbb{U}_l, \mathscr{U}_l)\}_{l=1}^N, \{(\mathbb{I}_l, \mathscr{I}_l)\}_{l=1}^N\rangle$ [12], where $(\Omega, \mathcal{F})$ is a measurable space of the environment, $N = n \times H$ is the number of agents in the intrinsic model. By a slight abuse of notation, we write $[N] := [n] \times [H]$, and write $l := (i, h) \in [N]$ for notational convenience. This way, any agent $l \in [N]$ corresponds to an agent $i \in [n]$ at timestep $h \in [H]$ in the state-space model. We denote by $\mathbb{U}_l$ the measurable action space of agent l and by $\mathscr{U}_l$ the σ-algebra over $\mathbb{U}_l$. For $A \subseteq [N]$, let $\mathbb{H}_A := \Omega \times \prod_{l\in A} \mathbb{U}_l$ and $\mathbb{H} := \mathbb{H}_{[N]}$. For any σ-algebra $\mathscr{C}$ over $\mathbb{H}_A$, let $\langle\mathscr{C}\rangle$ denote the cylindrical extension of $\mathscr{C}$ on $\mathbb{H}$. Let $\mathscr{H}_A := \langle\mathscr{F} \otimes (\otimes_{l\in A}\mathscr{U}_l)\rangle$ and $\mathscr{H} = \mathscr{H}_{[N]}$. We denote by $\mathbb{I}_l$ the space of *information available* to agent l, and by $\mathscr{I}_l$ the σ-algebra over $\mathbb{H}$. For $l \in [N]$, we denote by I_l the information of agent l, and U_l the action of agent l. The spaces and random variables of agent $l = (i, h)$ in the intrinsic model are related to those in the state-space model as follows: $\forall l = (i, h) \in [N], \mathbb{U}_l = \mathcal{A}_{i,h}, \mathbb{I}_l = \mathcal{T}_{i,h}, U_l = a_{i,h}, I_l = \tau_{i,h}$.

2) Information Structures of Dec-POMDPs: An important class of IS is the *quasi-classical* one, which is defined as follows [24], [12], [13].

Definition .21 (Quasi-classical Dec-POMDPs)**.** We call a Dec-POMDP problem *QC* if each agent in the intrinsic model knows the information available to the agents who influence her, directly or indirectly, i.e. $\forall l_1, l_2 \in [N], l_1 = (i_1, h_1), l_2 = (i_2, h_2), i_1, i_2 \in [n], h_1, h_2 \in [H]$, if agent l_1 influences agent l_2, then $\mathscr{I}_{l_1} \subseteq \mathscr{I}_{l_2}$.

Furthermore, *strictly* quasi-classical IS [24], [25], as a subclass of QC IS, is defined as follows.

Definition .22 (Strictly quasi-classical Dec-POMDPs)**.** We call a Dec-POMDP problem *sQC* if each agent in the intrinsic model knows the information *and* actions available to the agents who influence her, directly or indirectly. That is, $\forall l_1, l_2 \in [N], l_1 = (i_1, h_1), l_2 = (i_2, h_2), i_1, i_2 \in [n], h_1, h_2 \in [H]$, if agent l_1 influences agent l_2, then $\mathscr{I}_{l_1} \cup \langle\mathscr{U}_{l_1}\rangle \subseteq \mathscr{I}_{l_2}$.

3) Intrinsic Model of LTC Problems: Given any LTC $\mathcal{L}$ of the state-space-model form defined in §II-A, we define the intrinsic model of $\mathcal{L}$ as a tuple $\langle(\Omega, \mathscr{F}), N, \{(\mathbb{U}_l, \mathscr{U}_l)\}_{l=1}^N, \{(\mathbb{M}_l, \mathscr{M}_l)\}_{l=1}^N, \{(\mathbb{I}_{l^-}, \mathscr{I}_{l^-})\}_{l=1}^N,$
$\{(\mathbb{I}_{l^+}, \mathscr{I}_{l^+})\}_{l=1}^N\rangle$, where $(\Omega, \mathscr{F})$ is the measure space representing all the uncertainty in the system; $N = n \times H$ is the number of agents in the intrinsic model. By a slight abuse of notation, we write $[N] := [n] \times [H]$, and write $l := (i, h) \in [N]$ for convenience. This way, any agent $l \in [N]$ corresponds to an agent $i \in [n]$ at timestep $h \in [H]$ in the state-space model, and we thus define $l^- := (i, h^-)$ and $l^+ := (i, h^+)$ accordingly. We denote by $\mathbb{U}_l$ and $\mathbb{M}_l$ the measurable control and communication action spaces of agent l, and by $\mathscr{U}_l$ and $\mathscr{M}_l$ the σ-algebra over $\mathbb{U}_l$ and $\mathbb{M}_l$, respectively. For any $A \subseteq [N]$, let $\mathbb{H}_A := \Omega \times \prod_{l\in A}(\mathbb{U}_l \times \mathbb{M}_l)$ and $\mathbb{H} := \mathbb{H}_{[N]}$. For any σ-algebra $\mathscr{C}$ over $\mathbb{H}_A$, let $\langle\mathscr{C}\rangle$ denote the cylindrical extension of $\mathscr{C}$ on $\mathbb{H}$. Let $\mathscr{H}_A := \langle\mathscr{F} \otimes (\otimes_{l\in A}\mathscr{U}_l) \otimes (\otimes_{l\in A}\mathscr{M}_l)\rangle$, $\mathscr{H} = \mathscr{H}_{[N]}$. We denote by $\mathbb{I}_{l^-}$ and $\mathbb{I}_{l^+}$ the spaces of *information available* to agent l *before* and *after* additional sharing, respectively, and by $\mathscr{I}_{l^-} \subseteq \mathscr{H}$ and $\mathscr{I}_{l^+} \subseteq \mathscr{H}$ the associated σ-algebra. The spaces and random variables of agent $l = (i, h)$ in the intrinsic model are related to those in the state-space model as follows: $\forall l = (i, h) \in [N], \mathbb{U}_l = \mathcal{A}_{i,h}, \mathbb{M}_l = \mathcal{M}_{i,h}, \mathbb{I}_{l^-} = \mathcal{T}_{i,h^-}, \mathbb{I}_{l^+} = \mathcal{T}_{i,h^+}, U_l = a_{i,h}, M_l = m_{i,h}, I_{l^-} = \tau_{i,h^-}, I_{l^+} = \tau_{i,h^+}$. For notational convenience, for any random variable B in LTC and the σ-algebra $\mathscr{B}$ generated by B, we overload $\sigma(B)$ to denote the cylindrical extension of $\mathscr{B}$ on $\mathbb{H}$, i.e., $\sigma(B) = \langle\mathscr{B}\rangle$.

E. Conditions Leading to Assumption IV.7

As a minimal requirement for computational tractability (for both Dec-POMDPs and LTCs), Assumption IV.7 is needed for the one-step tractability of the team-decision problem involved in the value iteration in Algorithm 6. We now adapt several such structural conditions from [15] to the LTC setting, which lead to this assumption and have been studied in the literature. Note that since we need to do planning in the approximate model $\mathcal{M}$, which is oftentimes constructed based on the original problem $\mathcal{L}$ and approximate belief $\{\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \mid \widehat{c}_h)\}_{h\in[\overline{H}]}$, we necessarily need assumptions on these two models $\mathcal{L}$ and $\mathcal{M}$, for which we refer to as the **Part (1)** and **Part (2)** of the conditions below, respectively.

- **Turn-based structures. Part (1):** At each timestep $h \in [H]$, there is only one agent, denoted as $ct(h) \in [n]$, that can affect the state transition. More concretely, the transition dynamics take the forms of $\mathbb{T}_h : \mathcal{S} \times \mathcal{A}_{ct(h)} \to \Delta(\mathcal{S})$. Additionally, we assume the reward function admits an additive structure such that $\mathcal{R}_h(s_h, a_h) = \sum_{i\in[n]} \mathcal{R}_{i,h}(s_h, a_{i,h})$ for some functions $\{\mathcal{R}_{i,h}\}_{i\in[n]}$. Meanwhile, since only agent $ct(h)$ takes the action, we assume the increment of the common information $z_{h+1}^b = \chi_{h+1}(p_{h^+}, a_{ct(h),h}, o_{h+1})$. **Part (2):** No additional requirement. Such a structure has been commonly studied in (fully observable) stochastic games and multi-agent RL [34], [35].
- **Nested private information. Part (1):** No additional requirement. **Part (2):** At each timestep $h \in [\overline{H}]$, all the agents form a *hierarchy* according to the private information after $a_{i,h}$ they possess, in the sense that $\forall\ i, j \in [n], j < i, \overline{p}_{j,h} = Y_h^{i,j}(\overline{p}_{i,h})$ for some function $Y_h^{i,j}$. More formally, the approximate belief satisfies that $\mathbb{P}_h^{\mathcal{M},c}(\overline{p}_{j,h} = Y_h^{i,j}(\overline{p}_{i,h}) \mid \overline{p}_{i,h}, \widehat{c}_h) = 1$. Such a structure has been investigated in [36] with heuristic search, and in [15] with finite-time complexity analysis.
- **Factorized structures. Part (1):** At each timestep $h \in [H]$, the state s_h can be partitioned into n local states, i.e., $s_h = (s_{1,h}, s_{2,h}, \cdots, s_{n,h})$. Meanwhile, the transition kernel takes the product form of $\mathbb{T}_h(s_{h+1} \mid s_h, a_h) =$

$\prod_{i=1}^n \mathbb{T}_{i,h}(s_{i,h+1} \,|\, s_{i,h}, a_{i,h})$, the emission also takes the product form of $\mathbb{O}_h(o_h \,|\, s_h) = \prod_{i=1}^n \mathbb{O}_{i,h}(o_{i,h} \,|\, s_{i,h})$, and the reward function can be decoupled into n terms such that $\mathcal{R}_h(s_h, a_h) = \sum_{i,h} \mathcal{R}_h(s_{i,h}, a_{i,h})$. **Part (2):** At each even timestep $h \in [\overline{H}]$, the approximate common information is also factorized so that $\widehat{c}_h = (\widehat{c}_{1,h}, \widehat{c}_{2,h}, \cdots, \widehat{c}_{n,h})$ and its evolution satisfies that $\widehat{c}_{i,h+1} = \widehat{\phi}_{i,h+1}(\widehat{c}_{i,h}, \overline{z}_{i,h})$ for some function $\widehat{\phi}_{i,h+1}$. Correspondingly, the approximate belief need to satisfy that $\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \,|\, \widehat{c}_h) = \Pi_{i=1}^n \mathbb{P}_{i,h}^{\mathcal{M},c}(\overline{s}_{i,h}, \overline{p}_{i,h} \,|\, \widehat{c}_{i,h})$ for some functions $\{\mathbb{P}_{i,h}^{\mathcal{M},c}\}_{i\in[n],h\in[\overline{H}]}$ Such a structure, under general information sharing protocols, can lead to non-classical IS. In this case, it can be viewed an example of non-classical ISs where the agents have no incentive for signaling [13, §3.8.3].

Lemma .23. Given any LTC problem $\mathcal{L}$ and $\mathcal{D}'_{\mathcal{L}}$ is the Dec-POMDP after reformulation and expansion. For any $\mathcal{M}$ to be the approximate model of $\mathcal{D}_{\mathcal{L}}$ and $\{\mathbb{P}_h^{\mathcal{M},c}\}_{h\in[\overline{H}]}$ to be the approximate belief, if they satisfy any of the 3 conditions above, then Eq. (.32) in Algorithm 6 can be solved in polynomial time, i.e., Assumption IV.7 holds.

Proof. We prove the result case-by-case:

- **Turn-based structures:** For any $h = 2t, t \in [H], \gamma_{ct(h),h} \in \Gamma_{ct(h)}, \gamma_{-ct(h),h}, \gamma'_{-ct(h),h} \in \Gamma_{-ct(h),h}$, where $ct(h)$ is the controller, it holds for any $\widehat{c}_h$ that

$$\begin{aligned}
&Q_h^{*,\mathcal{M}}(\widehat{c}_h, \gamma_{ct(h),h}, \gamma_{-ct(h),h})\\
&= \sum_{\overline{s}_h, \overline{p}_h, \overline{s}_{h+1}, \overline{o}_{h+1}} \mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \,|\, \widehat{c}_h) \overline{\mathbb{T}}_h(\overline{s}_{h+1} \,|\, \overline{s}_h, \gamma_{ct(h),h}(\overline{p}_{ct(h),h}) \gamma_{-ct(h),h}(\overline{p}_{-ct(h),h}))\\
&\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1} \,|\, \overline{s}_{h+1})[\overline{\mathcal{R}}_h(\overline{s}_h, \gamma_{ct(h),h}(\overline{p}_{ct(h),h})) + V_{h+1}^{*,\mathcal{M}}(\widehat{c}_{h+1})]\\
&= \sum_{\overline{s}_h, \overline{p}_h, \overline{s}_{h+1}, \overline{o}_{h+1}} \mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \,|\, \widehat{c}_h) \overline{\mathbb{T}}_h(\overline{s}_{h+1} \,|\, \overline{s}_h, \gamma_{ct(h),h}(\overline{p}_{ct(h),h}) \overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1} \,|\, \overline{s}_{h+1})[\overline{\mathcal{R}}_h(\overline{s}_h, \gamma_{ct(h),h}(\overline{p}_{ct(h),h})) + V_{h+1}^{*,\mathcal{M}}(\widehat{c}_{h+1})],
\end{aligned}$$

where the last step is due to the fact $\widehat{c}_{h+1} = \widehat{\phi}_{h+1}(\widehat{c}_h, \overline{z}_{h+1})$. And $\overline{z}_{h+1} = z^b_{\frac{h}{2}+1} = \chi_{\frac{h}{2}+1}(\overline{p}_h, \overline{a}_{ct(h),h}, \overline{o}_{h+1})$. Therefore, right-hand side does no depend on $\gamma_{-ct(h),h}$. Therefore, Eq. (.32) with complexity poly$(\overline{\mathcal{S}}, \overline{\mathcal{P}}_{ct(h)}, \overline{\mathcal{A}}_{ct(h)})$.
- **Nested private information:** For any $i \in [n], h = 2t, t \in [H]$, we first define the $u_{i,h} \in \mathcal{U}_{i,h} := \{(\times_{j=1}^i \mathcal{P}_{j,h}) \times (\times_{j=1}^{i-1} \mathcal{A}_{j,h}) \to \mathcal{A}_{i,h}\}$ and slightly abuse the notation for $Q_h^{*,\mathcal{M}}$ as follows

$$\begin{aligned}
Q_h^{*,\mathcal{M}}(\widehat{c}_h, u_{1,h}, \cdots, u_{n,h}) := &\sum_{\overline{s}_h, \overline{p}_h, \overline{a}_h, \overline{s}_{h+1}, \overline{o}_{h+1}} \mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \,|\, \widehat{c}_h) \Pi_{i=1}^n \mathbb{1}[\overline{a}_{i,h} = u_{i,h}(\overline{p}_{1:i,h}, \overline{a}_{1:i-1,h})] \overline{\mathbb{T}}_h(\overline{s}_{h+1} \,|\, \overline{s}_h, \overline{a}_h)\\
&\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1} \,|\, \overline{s}_{h+1})[\overline{\mathcal{R}}_h(\overline{s}_h, \overline{a}_h) + V_{h+1}^{*,\mathcal{M}}(\widehat{c}_{h+1})]
\end{aligned}$$

  Since the space of $\mathcal{U}_{i,h}$ covers the space $\Gamma_{i,h}$, then for the $u^*_{1:n,h}$ be an optimal one that maximize the $Q_h^{*,\mathcal{M}}$, we have

$$Q_h^{*,\mathcal{M}}(\widehat{c}_h, u^*_{1,h}, \cdots, u^*_{n,h}) = \max_{\{u_{i,h} \in \mathcal{U}_{i,h}\}_{i\in[n]}} Q_h^{*,\mathcal{M}}(\widehat{c}_h, u_{1,h}, \cdots, u_{n,h}) \geq \max_{\{\gamma_{i,h} \in \Gamma_{i,h}\}_{i\in[n]}} Q_h^{*,\mathcal{M}}(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h}).$$

  Meanwhile, due to the nested private information condition, for any $\overline{p}_h \in \overline{\mathcal{P}}_h$, there must exists $\gamma'_{1:n,h}$ such that $\gamma'_{1:n,h}$ output the same actions as $u^*_{1:n,h}$ under $\overline{p}_h$. Therefore, we can conclude that

$$\max_{\{u_{i,h} \in \mathcal{U}_{i,h}\}_{i\in[n]}} Q_h^{*,\mathcal{M}}(\widehat{c}_h, u_{1,h}, \cdots, u_{n,h}) = \max_{\{\gamma_{i,h} \in \Gamma_{i,h}\}_{i\in[n]}} Q_h^{*,\mathcal{M}}(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h})$$

  Therefore, we can solve Eq. (.32) and compute $\gamma^*_{1:n,h}$ from computing $u^*_{1:n,h}$, which can be solved with complexity poly$(\overline{\mathcal{P}}_h, \overline{\mathcal{A}}_h, \overline{\mathcal{S}})$.
- **Factorized structures:** For any $h \in [\overline{H}], t \in [H]$, for any [Kaiqing: what?] we use backward induction to prove that, there exist n functions $\{F_{i,h}\}_{i\in[n]}$ such that

$$Q_h^{*,\mathcal{M}}(\widehat{c}_h, \gamma_h) = \sum_{i=1}^n F_{i,h}(\widehat{c}_{i,h}, \gamma_{i,h})$$

It holds for $h = \overline{H} + 1$ obviously. For any $h \leq \overline{H}$, it holds that

$$
\begin{aligned}
&Q_h^{*,\mathcal{M}}(\widehat{c}_h, \gamma_h) \\
&\quad = \sum_{\overline{s}_h, \overline{p}_h, \overline{s}_{h+1}, \overline{o}_{h+1}} \mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h \,|\, \widehat{c}_h) \overline{\mathbb{T}}_h(\overline{s}_{h+1} \,|\, \overline{s}_h, \gamma_h(\overline{p}_h)) \overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1} \,|\, \overline{s}_{h+1}) \\
&\qquad\qquad [\sum_{i=1}^{n} \overline{\mathcal{R}}_{i,h}(\overline{s}_{i,h}, \gamma_{i,h}(\overline{p}_{i,h}) + F_{i,h+1}(\widehat{c}_{i,h+1}, \widehat{g}^*_{i,h+1}(\widehat{c}_{i,h+1}))] \\
&\quad = \sum_{i=1}^{n} \sum_{\overline{s}_{i,h}, \overline{p}_{i,h}, \overline{s}_{i,h+1}, \overline{o}_{i,h+1}} \mathbb{P}_{i,h}^{\mathcal{M},c}(\overline{s}_{i,h}, \overline{p}_{i,h} \,|\, \widehat{c}_{i,h}) \overline{\mathbb{T}}_h(\overline{s}_{i,h+1} \,|\, \overline{s}_{i,h}, \gamma_{i,h}(\overline{p}_{i,h})) \\
&\qquad\qquad \overline{\mathbb{O}}_{i,h+1}(\overline{o}_{i,h+1} \,|\, \overline{s}_{i,h+1}) [\overline{\mathcal{R}}_{i,h}(\overline{s}_{i,h}, \gamma_{i,h}(\overline{p}_{i,h}) + F_{i,h+1}(\widehat{c}_{i,h+1}, \widehat{g}^*_{i,h+1}(\widehat{c}_{i,h+1}))] \\
&\quad =: \sum_{i=1}^{n} F_{i,h}(\widehat{c}_{i,h}, \gamma_{i,h}).
\end{aligned}
$$

Then, by induction, we know that it holds for any $h \in [\overline{H}]$. We can define $\widehat{g}^*_{i,h}(\widehat{c}_h) \in \text{argmax}_{\gamma_{i,h} \in \Gamma_{i,h}} F_{i,h+1}(\widehat{c}_{i,h+1}, \gamma_{i,h})$, and thus solve Eq.(.32) with complexity $\sum_{i=1}^{n} \text{poly}(\overline{\mathcal{S}}_i, \overline{\mathcal{A}}_{i,h}, \overline{\mathcal{P}}_{i,h})$.

This completes the proof. □

F. *Examples in the Venn Diagram Fig. 1b*

Here, we show some examples of the areas ①-⑤ in the Venn diagram in Fig. 1b.

- **①: Multi-agent MDP [37] with historical states.** The Dec-POMDPs satisfying that for any $h \in [H], i \in [n], \mathcal{O}_{i,h} = \mathcal{S}, \mathbb{O}_{i,h}(s \,|\, s) = 1, c_h = s_{1:h}, p_h = \emptyset$ lie in the area ①.
- **②: Uncontrolled state process without any historical information.** The Dec-POMDPs satisfying that for any $h \in [H], i \in [n], s_h, a_h, a'_h, \mathbb{T}_h(\cdot \,|\, s_h, a_h) = \mathbb{T}_h(\cdot \,|\, s_h, a'_h), c_h = \emptyset, p_{i,h} = \{o_{i,h}\}$ lie in the area ②.
- **③: Dec-POMDPs with sQC information structure and perfect recall, and satisfying Assumptions III.5 and III.7.** This class is what we mainly considered in §V.
- **④: State controlled by one controller with no sharing and only observability of controller.** We consider a Dec-POMDP $\mathcal{D}$. The state dynamics are controller by only one agent (, for convenience, agent 1), and only agent 1 has observability, i.e. $\mathbb{T}_h(\cdot \,|\, s_h, a_{1,h}, a_{-1,h}) = \mathbb{T}_h(\cdot \,|\, s_h, a_{1,h}, a'_{-1,h})$ for all $s_h, a_{1,h}, a_{-1,h}, a'_{-1,h}$, and $\mathcal{O}_{-1,h} = \emptyset$. There is no information sharing, i.e. $c_h = \emptyset, p_{1,h} = \{o_{1:h}, a_{1:h-1}\}, p_{j,h} = \{a_{j,1:h-1}\}, \forall j \neq 1$. Then $\forall j \neq 1, h_1 < h_2 \in [H]$, agent $(1, h_1)$ does not influence (j, h_2), since $\tau_{j,h_2} = \{a_{j,1:h_2-1}\}$ is not influenced by agent $(1, h_1)$. Therefore, $\mathcal{D}$ is sQC and has perfect recall, $\mathcal{D}$ is not SI (underlying state s_h influenced by $g_{1,1:h-1}$). This is because $\mathcal{D}$ does not satisfy Assumption III.7. Then $\mathcal{D}$ lies in the area ④.
- **⑤: One-step delayed observation sharing and two-step delayed action sharing.** The Dec-POMDPs satisfying that for any $h \in [H], i \in [n], c_h = \{o_{1:h-1}, a_{1:h-2}\}, p_{i,h} = \{a_{i,h-1}, o_{i,h}\}$ lie in the area ⑤.

G. *Additional Experimental Details and Results*

a) Experimental setup: We conduct our experiments on two popular and modest-scale partially observable benchmarks, Dectiger [38] and Grid3x3 [39]. To fit the setting of LTC in our paper. We regularize the reward between [0,1] and set the base information structure as one-step-delay. As for the communication cost function, we set $\mathcal{K}_h(Z_h^a) = \alpha|Z_h^a|$, and set $\alpha \in [0.01, 0.05, 0.1]$ for the purpose of ablation study. Also, we study 2 baselines under the same environment with information structure of one-step delay and fully-sharing, respectively. The one-step-delay baseline can be regarded as an LTC problem with extremely high communication cost, thus no additional sharing. On the other hand, the fully-sharing baseline is the LTC problem with no communication cost. Additionally, the results of different horizons and communications costs over 20 random seeds are shown in Tables I and II.

H. *Additional Figures*

We present two figures that illustrate the paradigm and the timeline of the Learning-to-Communicate problem considered in this paper.

Horizon/Cost	No Sharing	Cost=0.1	Cost=0.05	Cost=0.01	Fully Sharing
H=4 w/ cost	1.32±0.025	1.33±0.044	1.44±0.034	1.54±0.013	1.57±0.004
H=4 w/o cost	-	1.36±0.032	1.48±0.034	1.59±0.002	-
H=6 w/ cost	1.95±0.009	1.97±0.07	2.08±0.068	2.26±0.012	2.29±0.002
H=6 w/o cost	-	2.01±0.047	2.14±0.072	2.27±0.011	-
H=8 w/ cost	2.56±0.041	2.64±0.078	2.74±0.118	2.96±0.021	3.0±0.002
H=8 w/o cost	-	2.7±0.044	2.83±0.117	2.98±0.02	-
H=10 w/ cost	3.31±0.024	3.37±0.135	3.51±0.153	3.69±0.029	3.87±0.007
H=10 w/o cost	-	3.46±0.069	3.63±0.152	3.71±0.026	-

TABLE I: Experimental results for Dectiger.

Horizon/Cost	No Sharing	Cost=0.1	Cost=0.05	Cost=0.01	Fully Sharing
H=4 w/ cost	0.14±0.003	0.14±0.019	0.15±0.002	0.26±0.028	-0.48±0.023
H=4 w/o cost	-	0.14±0.019	0.21±0.007	0.33±0.023	-
H=6 w/ cost	0.33±0.02	0.32±0.025	0.4±0.009	0.48±0.059	-0.38±0.075
H=6 w/o cost	-	0.32±0.025	0.54±0.02	0.62±0.075	-
H=8 w/ cost	0.52±0.084	0.52±0.051	0.58±0.072	0.67±0.031	-0.4±0.022
H=8 w/o cost	-	0.52±0.051	0.72±0.035	0.82±0.074	-
H=10 w/ cost	0.73±0.02	0.73±0.037	0.9±0.169	1.03±0.019	-0.15±0.188
H=10 w/o cost	-	0.73±0.037	1.08±0.14	1.25±0.062	-

TABLE II: Experimental results for Grid3x3.

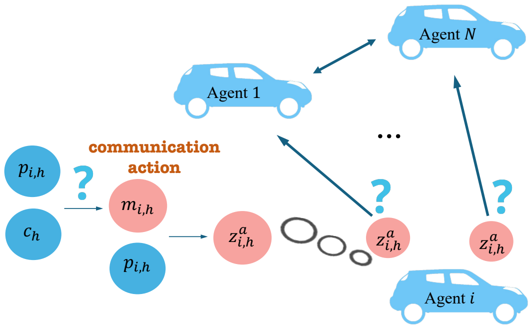


Fig. 4: Illustrating the paradigm of the Learning-to-Communicate problem considered in this paper..

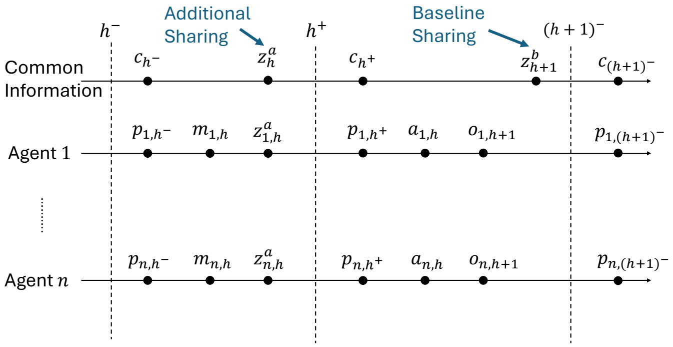


Fig. 5: Timeline of the information sharing and evolution in the Learning-to-Communicate problem considered in this paper.