Information Complexity of the AND Function in the Two-Party and Multi-party Settings

Yuval Filmus\textsuperscript{1} · Hamed Hatami\textsuperscript{2} · Yaqiao Li\textsuperscript{2} · Suzin You\textsuperscript{2}

Received: 13 September 2017 / Accepted: 16 July 2018 / Published online: 20 July 2018 © Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

In a recent breakthrough paper Braverman et al. (in: STOC’13, pp 151–160, 2013) developed a local characterization for the zero-error information complexity in the two-party model, and used it to compute the exact internal and external information complexity of the 2-bit AND function. In this article, we extend their results on AND function to the multi-party number-in-hand model by proving that the generalization of their protocol has optimal internal and external information cost for certain natural distributions. Our proof has new components, and in particular, it fixes a minor gap in the proof of Braverman et al.

Keywords Information complexity · AND function · Multi-party number-in-hand model · Concavity condition

1 Introduction

Although communication complexity has been witnessing steady and rapid progress since its birth, it was not until recently that a focus on an information-theoretic approach resulted in new and deeper understanding of some of the classical problems in the area. This gave birth to a new area of complexity theory called information complexity. Communication complexity is concerned with minimizing the amount of communication
required for players who wish to evaluate a function that depends on their private inputs. Information complexity, on the other hand, is concerned with the amount of information that the communicated bits reveal about the inputs of the players to each other, or to an external observer.

One of the important achievements of information complexity is the recent breakthrough of [2] that determines the exact asymptotics of the randomized communication complexity of one of the oldest and most studied problems in communication complexity, set disjointness:

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{R_\varepsilon(\text{DISJ}_n)}{n} \approx 0.4827.$$  (1)

Here $R_\varepsilon(\cdot)$ denotes the randomized communication complexity with an error of at most $\varepsilon$ on every input, and $\text{DISJ}_n$ denotes the two-party set disjointness problem in which two players, each holds a subset of $[n]$, need to verify if their subsets are disjoint. Prior to the discovery of these information-theoretic techniques, proving the lower bound $R_\varepsilon(\text{DISJ}_n) = \Omega(n)$ had already been a challenging problem, and even Razborov’s [14] short proof of that fact is intricate and sophisticated.

Let $\text{AND}_k : \{0, 1\}^k \to \{0, 1\}$ be the $k$-party AND function where the input has $k$ bits. We usually denote $\text{AND}_2$ simply by AND. Note that the $\text{DISJ}_n$ is nothing but an OR of AND functions. More precisely, for $i = 1, \ldots, n$, if $x_i$ is the Boolean variable which represents whether $i$ belongs to Alice’s set or not, and $y_i$ is the corresponding variable for Bob, then $\bigvee_{i=1}^n (x_i \land y_i)$ is true if and only if Alice’s input intersects Bob’s input. Braverman et al. [2] exploited this fact to prove (1).

Roughly speaking, they first determined the exact information complexity of the AND function for any underlying distribution $\mu$ on the set of inputs $\{0, 1\} \times \{0, 1\}$, and then used the fact that amortized communication equals information complexity [4] to relate this to the communication complexity of $\text{DISJ}_n$. The constant 0.4827 in (1) is indeed the maximum of the information complexity of the 2-bit AND function over all distributions $\mu$ that assign a zero mass to $\{1,1\} \in \{0,1\} \times \{0,1\}$. That is

$$\max_{\mu: \mu(11) = 0} \text{IC}_\mu(\text{AND}) \approx 0.4827,$$  (2)

where $\text{IC}_\mu(\text{AND})$ denotes the information complexity of the 2-bit AND function with respect to the distribution $\mu$ with no error (see Definition 3 below).

A similar phenomenon also happens in the multi-party setting. Consider the $k$-party promised set disjointness problem $\text{PDISJ}_{n,k} : \{0, 1\}^n \times \{0, 1\}^n \times \cdots \times \{0, 1\}^n \to \{0, 1\}$ in the number-in-hand model, where $k$ players each holds a subset of $[n]$, and it is promised that either these $k$ subsets are all pairwise disjoint, or they intersect at exactly one element of $[n]$. Similarly, the $k$-party set disjointness function can also be expressed as an OR of $\text{AND}_k$ functions. Let $e_i \in \{0, 1\}^k$ be the usual $i$th basis vector $(0, \ldots, 0, 1, 0, \ldots, 0)$, and $0, 1$ denote the all-0 and all-1 vectors, respectively. For the promised set disjointness function $\text{PDISJ}_{n,k}$, the fact that those subsets either are pairwise disjoint or intersect at exactly one element shows the input to $\text{AND}_k$ must lie in the subset $\{0, 1, e_1, \ldots, e_k\}$. Using this fact, [5,9] determined $R_\varepsilon(\text{PDISJ}_{n,k}) = \Omega(n/k)$ by showing a lower bound of (a variant of) the information complexity of
the multi-party $\text{AND}_k$ function. These works show the importance of studying the information complexity of simple functions such as the $\text{AND}_2$ and $\text{AND}_k$ functions.

Return to the two-party setting, although obtaining the asymptotics of $R_{\varepsilon}(\text{DISJ}_n)$ from the information complexity of the AND function is not straightforward and a formal proof requires overcoming some technical difficulties, the bulk of [2] is dedicated to computing the exact information complexity of the 2-bit AND function. This rather simple-looking problem had been studied previously by Ma and Ishwar [11, 12], and some of the key ideas of [2] originate from their work. In [2] Braverman et al. introduced a protocol to solve the AND function, and proved that it has optimal internal and external information cost. Interestingly this protocol is not a conventional communication protocol as it has access to a continuous clock, and the players are allowed to “buzz” at randomly chosen times. Indeed, it is known [2] that no protocol with a bounded number of rounds can have optimal information cost for the AND function, and hence the infinite number of rounds, implicit in the continuous clock, is essential. We shall refer to this protocol as the buzzers protocol.

1.1 Our Contributions

Fixing the argument of [2] In order to show that the buzzers protocol has optimal information cost, inspired by the work of Ma and Ishwar [11,12], Braverman et al. came up with a local concavity condition, and showed that if a protocol satisfies this condition, then it has optimal information cost. This condition, roughly speaking, says that it suffices to verify that one does not gain any advantage over the conjectured optimal protocol if one of the players starts by sending a bit $B$. In the original paper [2], it is claimed that it suffices to verify this condition only for signals $B$ that reveal arbitrarily small information about the inputs. As we shall see, however, this is not true, see Sect. 3.1.

In Theorem 2 we prove a variant of the local concavity condition that allows one to consider only signals $B$ with small information leakage, this fixes the argument in [2]. We have been informed through private communication that Braverman et al. have also independently fixed this error.

Another minor issue in [2] is that one notation is used for two different notions of information complexity, this sometimes causes confusion. In order to emphasize difference between the two different notions of information complexity, we define both notions and apply different notations. We discuss this in Sect. 2.2.

Extension of [2] to the multi-party setting We then apply Theorem 2 to extend the information complexity result of [2] on two-party $\text{AND}_2$ function, to the multi-party $\text{AND}_k$ function in the number-in-hand model. To achieve this, we proceed by defining a generalization of the buzzers protocol, and then prove in Theorem 3 that it has optimal internal and external information cost when the input distribution $\mu$ for $\text{AND}_k$ satisfies the following assumption:

Assumption 1 The support of $\mu$ is a subset of $\{0, 1, e_1, \ldots, e_k\}$. 
Note that in the two-party setting, every distribution satisfies this assumption and thus our results are complete generalizations of the results of [2] of the two-party setting. As we discussed before, the distributions in Assumption 1 arise naturally in the study of the multi-party promised set disjointness problem, and they have been considered previously in [5,9].

This extension is not straightforward since in [2], a large part of the calculations for verifying the local concavity conditions are carried out by the software Mathematica, however, in the number-in-hand model, the number of players $k$ can be arbitrary, one cannot simply rely on a computer for those calculations. Indeed, we had to look into the protocol and show that it suffices to analyze the behaviour of the information cost in a small time interval that is determined by the distribution of the inputs, and furthermore one can reduce all distributions into a class of essentially 3-parameter distributions. Hence the problem can be reduced to one that has only a constant number of variables, thus allowing us to use Mathematica to verify the concavity condition. We believe our analysis may provide new insights even for the two-party setting.

2 Preliminaries

2.1 Notation

We typically denote random variables by capital letters (e.g., $A$, $B$, $C$, $X$, $Y$, $I$), and write $A_1 \ldots A_n$ to denote the random variable $(A_1, \ldots, A_n)$. We simply use the word distribution to mean a probability distribution. Let $[k] := \{1, \ldots, k\}$. Given a finite space $\Omega$, the notation $\Delta(\Omega)$ denotes the set of all probabilistic distributions on $\Omega$.

Let $X$, $Y$ be two random variables distributed on a finite space $\Omega$. As usual, $H(X) := -\sum_{x \in \Omega} \Pr[X = x] \log \Pr[X = x]$ is the Shannon entropy of $X$, where here and throughout the paper $\log(\cdot)$ is in base 2, and $0 \log 0 = 0$. $H(X|Y) := \sum_{y \in \Omega} \Pr[Y = y] H(X|Y = y)$ is the (conditional) Shannon entropy of $X$ conditioned on $Y$, and $I(X; Y) := H(X) - H(X|Y)$ is the mutual information between $X$ and $Y$. Given three random variables $X$, $Y$, $Z$, the conditional mutual information $I(X; Y|Z)$ is defined as $I(X; Y|Z) := H(X|Z) - H(X|Y|Z)$. For every $\varepsilon \in [0, 1]$, let $H(\varepsilon) = -\varepsilon \log \varepsilon - (1 - \varepsilon) \log (1 - \varepsilon)$ denote the binary entropy.

Let $\text{supp}(\mu)$ denote the support of a distribution $\mu$. Let $\mu$, $\nu$ be two distributions on a finite space $\Omega$. Define $|\mu - \nu| := \frac{1}{2} \sum_{a \in \Omega} |\mu(a) - \nu(a)|$ to be the statistical distance (a.k.a. total variation distance) between $\mu$ and $\nu$. The distance between two distributions in this paper is always the statistical distance. The notation $D(\mu||\nu) := \sum_{a \in \Omega} \mu(a) \log \frac{\mu(a)}{\nu(a)}$ denotes the Kullback-Leibler divergence (a.k.a. relative entropy) from $\nu$ to $\mu$. We also use $D(X||Y)$ to denote the divergence from the distribution of a random variable $Y$ to the distribution of a random variable $X$. Given three random variables $X$, $Y$, $Z$, one has $I(XY; Z) = \E_{z \sim Z} D(XY|Z=z||XY)$, and $I(X; Y|Z) = \E_{y \sim Y, z \sim Z} D(X|y=z, z=z||X|z=z)$. For more details on entropy, mutual information and divergence, see [6].

A random binary bit is sometimes also called a signal.
2.2 Communication Complexity and Information Complexity

We briefly review the notion of two-party communication complexity which was introduced by Yao [15], a standard reference on communication complexity is [10]. In this model there are two players (with unlimited computational power), often called Alice and Bob, who wish to collaboratively compute a given function \( f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} \). Alice receives an input \( x \in \mathcal{X} \) and Bob receives \( y \in \mathcal{Y} \). Neither of them know the other player’s input, and they wish to communicate in accordance with an agreed-upon protocol \( \pi \) to compute \( f(x, y) \). The protocol \( \pi \) specifies as a function of (only) the transmitted bits whether the communication is over, and if not, who sends the next bit. Furthermore \( \pi \) specifies what the next bit must be as a function of the transmitted bits and the input of the player who sends the bit. The transcript \( \Pi \) of a protocol \( \pi \) is the list of all the transmitted bits during the execution of the protocol. In the randomized communication model, the players have access to private random strings. These random strings are independent, and they can have any desired distributions individually. The communication cost of a protocol \( \pi \) is the maximal number of transmitted bits among all possible inputs \( (x, y) \in \mathcal{X} \times \mathcal{Y} \). The communication complexity of a function \( f \) is defined to be the minimal communication cost among all protocols that solve the function \( f \) correctly.

While communication complexity cares about the number of transmitted bits, the information complexity cares about the information revealed by the communicated bits. To be able to measure information, we need to assume a prior distribution \( \mu \) on the input space \( \mathcal{X} \times \mathcal{Y} \). Once we assumed the input distribution \( \mu \), the transcript \( \Pi \) becomes a random variable: it denotes the random transcript of the protocol \( \pi \) that runs on a randomly sampled input \( (x, y) \) according to distribution \( \mu \). The information cost of a protocol \( \pi \) is, intuitively, how much information one can learn about the input \( X \) and \( Y \) by observing the random transcript \( \Pi \). There are two different notions of information cost, depending on who are the observers.

**Definition 1 (Information cost)** The external information cost and the internal information cost of a protocol \( \pi \) with respect to a distribution \( \mu \) on \( \mathcal{X} \times \mathcal{Y} \) are defined as

\[
IC_{\mu}^\text{ext}(\pi) := I(\Pi; XY) \tag{3}
\]

and

\[
IC_{\mu}(\pi) := I(\Pi; X|Y) + I(\Pi; Y|X) \tag{4}
\]

respectively, where \( \Pi = \Pi_{XY} \) is the transcript of the protocol when it is executed on the input \( XY \) that is randomly sampled according to \( \mu \).

Intuitively, the external information cost, as the name suggests, specifies what an external observer can learn about the input \( XY \) by knowing the random transcript \( \Pi \). In contrast, the internal information cost is the amount of information that Alice and Bob can learn about each other’s input by looking at the random transcript.

Usually when we simply say information cost, we mean the internal information cost.
Next we define the information complexity of a function. In order to distinguish two distinct notions of information complexity (as we shall see shortly), we introduce the communication tasks.

**Definition 2** (*Communication task*) Given a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, we say a communication protocol $\pi$ performs the communication task $[f, 0]$ if it computes $f$ correctly for every input $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Given a distribution $\nu$ on input space $\mathcal{X} \times \mathcal{Y}$. We say a protocol $\pi$ performs the communication task $[f, \nu, 0]$ if it computes $f$ correctly for every input $(x, y) \in \text{supp}(\nu)$.

One could replace the number 0 in the above definition by an $\varepsilon \in [0, 1]$ to allow error, e.g., a protocol $\pi$ performs the task $[f, \nu, \varepsilon]$ if $\Pr_{(X,Y) \sim \mu}[\pi(X, Y) \neq f(X, Y)] \leq \varepsilon$. Obviously $[f, \nu, 0]$ is the same as $[f, 0]$ when $\nu$ has full support. However, $[f, \nu, 0]$ can be strictly easier than $[f, 0]$ when $\nu$ is not fully supported. Since in the present paper we focus on zero error setting, sometimes we simply use $[f], [f, \nu]$ to denote $[f, 0]$ and $[f, \nu, 0]$, respectively.

**Definition 3** (*Information complexity*) Given two distributions $\mu, \nu$ on the input space $\mathcal{X} \times \mathcal{Y}$. The **external information complexity** and the **internal information complexity** of a function $f$ with respect to a distribution $\mu$, are defined as the infimum information cost with respect to $\mu$ among all (randomized) protocols that perform the task $[f]$, i.e.,

$$
IC^\text{ext}_\mu(f) := \inf_{\pi : \pi \text{ performs } [f]} IC^\text{ext}_\mu(\pi)
$$

and

$$
IC^\text{int}_\mu(f) := \inf_{\pi : \pi \text{ performs } [f]} IC^\text{int}_\mu(\pi)
$$

respectively.

The **distributional external information complexity** and the **distributional internal information complexity** of a function $f$ with respect to a distribution $\mu$, are defined as the infimum information cost with respect to $\mu$ among all (randomized) protocols that perform the task $[f, \nu, 0]$, i.e.,

$$
IC^\text{ext}_\mu(f, \nu) := \inf_{\pi : \pi \text{ performs } [f, \nu]} IC^\text{ext}_\mu(\pi)
$$

and

$$
IC^\text{int}_\mu(f, \nu) := \inf_{\pi : \pi \text{ performs } [f, \nu]} IC^\text{int}_\mu(\pi)
$$

respectively.

Our focus will be on the (non-distributional) $IC^\text{int}_\mu(f)$ and $IC^\text{ext}_\mu(f)$, that is, the communication task that we shall focus on will be exclusively on $[f]$, that is, to compute the function $f$ correctly for all inputs.
Remark 1 (Remarks on Definition 3)

1. The infimum is essential in the definition. In other words, there are functions (e.g., the 2-bit AND function) for which there is no protocol that achieves ICμ(f) while there is a sequence of protocols whose information costs converge to ICμ(f).

2. In defining the distributional information complexity, the distribution μ is used to measure the information cost while the distribution ν is used to measure the error of a protocol. In practice, often ν is chosen to be the same as μ, hence the notations ICextμ(f,μ) and ICμ(f,μ), respectively.

3. In the literature of information complexity it was common to use “ICμ(f)” to denote its distributional counterpart, i.e., what we denote by ICμ(f,μ). Unfortunately, this has become the source of some confusions in the past, as sometimes “ICμ(f)” is used to denote both (non-distributional) ICμ(f) and the distributional information complexity ICμ(f,μ).

As an example to show the difference between ICμ(f) and ICμ(f,μ): let f = AND be the two-bit AND function, and μ be the uniform distribution on (1, 0), (0, 1) and (0, 0). Obviously ICextμ(f,μ) = 0 since a protocol that simply outputs 0 without any communication solves the task [AND, μ]. On the other hand, ICμ(f) = log 3 by [2].

Lastly, we also need the uniform continuity of the information complexity with respect to μ.

Lemma 1 ([3, Lemma 4.4]) ICμ(f) is uniformly continuous with respect to μ, under the statistical distance of distributions.

2.3 The Multi-party Number-In-Hand Model

The number-in-hand model is the most straightforward generalization of Yao’s two-party model to the settings where more than two players are present. In this model there are k players who wish to collaboratively compute a function f: X1 × ... × Xk → Z. Each player i knows only an input xi ∈ Xi, and the communication is in the shared blackboard model, which means that all the communicated bits are visible to all the players. Let μ be a distribution on X1 × ... × Xk, and let X = (X1, ..., Xk) be sampled from X1 × ... × Xk according to μ. Definition 1 generalizes in a straightforward manner to ICextμ(π) = I(Π; X), and ICμ(π) = ∑k i=1 I(Π; X−i|Xi) where X−i := (X1, ..., Xi−1, Xi+1, ..., Xk). Note also that I(Π; X|Xi) = I(Π; X−i|Xi), and thus we have ICμ(π) = ∑k i=1 I(Π; X|Xi). The information complexity ICμ(f), ICμ(f, ν), and the continuity also generalize in a straightforward manner to this setting.

3 The Local Characterization of the Optimal Information Cost

Fix a function f: X × Y → Z, the information complexity ICμ(f) (or ICextμ(f)) can be viewed as a function of μ, and it is continuous (Lemma 1). Let π be a protocol that solves [f, 0], by definition one has ICμ(π) ≥ ICμ(f), i.e., the information cost.
of $\pi$ is an upper bound of the information complexity. How to show a conjectured optimal protocol $\pi$ is indeed optimal? This is equivalent to show that the information cost of $\pi$ is also a lower bound for any other protocols. Perhaps surprisingly, it turns out that essentially one can show the optimality of a protocol by verifying a simple “local” condition.

In order to state the local condition, we define some notions related to a single random bit. Consider the two-party communication model. Let $B$ be a random bit sent by one of the players in some protocol, and let $\mu_0 = \mu|_{B=0}$ and $\mu_1 = \mu|_{B=1}$, or in other words
\[
\mu_b(xy) := \Pr[XY = xy|B = b] = \frac{\Pr[B = b|XY = xy]}{\Pr[B = b]} \mu(xy),
\]
for $b = 0, 1$. In particular, $\text{supp } \mu_b = \text{supp } \mu$. Denote $\Pr[\cdot|xy] := \Pr[\cdot|XY = xy]$. Note we use $(x, y)$ and $xy$ interchangeably for the sake of convenience depending on the context.

**Definition 4** Let $\mu$ be a distribution and $B$ be a bit sent by one of the players.
- $B$ is called unbiased with respect to $\mu$ if $\Pr[B = 0] = \Pr[B = 1] = \frac{1}{2}$.
- $B$ is called non-crossing if for all $(x, y) \neq (x', y')$, $\mu(x, y) < \mu(x', y')$ implies that $\mu_b(x, y) \leq \mu_b(x', y')$ for $b = 0, 1$.
- $B$ is called $\varepsilon$-weak if $|\Pr[B = 0|xy] - \Pr[B = 1|xy]| \leq \varepsilon$ for every input $xy$.

A protocol is said to be in normal form with respect to $\mu$ if all bits sent in it are unbiased and non-crossing with respect to $\mu$.

Obviously, the above notions can be generalized into the multi-party setting in a straightforward manner.

**Remark 2** (A discussion on non-crossing bits) The notion of non-crossing bits defined in Definition 4 is slightly different from the one defined in [2, Section 7.4].

The notion of non-crossing bits in [2, Section 7.4] is specifically defined for AND$_2$ function. Let $\mu$ be a distribution on the input space for the 2-bit AND$_2$ function, that is $\mu = (\alpha \beta \gamma \delta)$ is a distribution on $\{0, 1\} \times \{0, 1\}$ where $\alpha = \mu(00)$, $\beta = \mu(01)$, $\gamma = \mu(10)$ and $\delta = \mu(11)$. Suppose Alice is the row player and Bob is the column player. If $\beta < \gamma$, then $\Pr[Y = 0] = \alpha + \gamma > \alpha + \beta = \Pr[X = 0]$, that is, for a randomly sampled input according to $\mu$, it is more likely that Bob has a 0 than Alice does. Intuitively, for $\mu$ satisfies $\beta < \gamma$, it is then better for Bob to speak first in the optimal protocol for AND$_2$ to reduce information leakage.

Based on such intuition, in [2], a distribution $\mu$ with $\beta < \gamma$ is called in Bob’s region, and $\beta > \gamma$ in Alice’s region, and a bit $B$ is said to be non-crossing if $\mu|_{B=0}$ and $\mu|_{B=1}$ stay in the same region as $\mu$. It is easy to see that any non-crossing bit in Definition 4 is non-crossing in the sense in [2]. We choose our definition to allow generality for functions other than AND.

Finally we point out that the above discussion shows an optimal protocol can depend on the input distribution: the players act differently in an optimal protocol if the input
distribution changes. The non-crossing property simplifies the task of verifying the local condition, as we shall see shortly.

Recall $\Delta(\mathcal{X} \times \mathcal{Y})$ denotes the set of distributions on $\mathcal{X} \times \mathcal{Y}$. A distribution $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$ is said to be internal-trivial (resp. external-trivial) for $f$ if $IC_\mu(f) = 0$ (resp. $IC_\mu^{ext}(f) = 0$). These distributions are characterized in [7].

3.1 The Local Characterization

Let us first understand the properties that $IC_\mu(f)$ satisfies as a function of $\mu$. Given the input distribution $\mu$, consider a protocol where one player first sends a random bit $B$, and if $B = 0$, the players continue by running a protocol that is optimal for $\mu_0$, and if $B = 1$, they run a protocol that is optimal for $\mu_1$. Denote

$$I_B^{ext} := I(B; XY), \quad I_B := I(B; X|Y) + I(B; Y|X).$$

That is, they are the information cost revealed by $B$ to the external observer and internal players, respectively. By definition, the information cost of this protocol is an upper bound of the information complexity,

$$IC_\mu(f) \leq I_B + \mathbb{E}_B[IC_{\mu B}(f)]. \quad (11)$$

Similarly, for the external case,

$$IC_\mu^{ext}(f) \leq I_B^{ext} + \mathbb{E}_B[IC_{\mu B}^{ext}(f)]. \quad (12)$$

Note that (11) is a local condition: it concerns only the bit $B$. If a protocol $\pi$ is optimal, then $IC_\mu(\pi)$ should also satisfy (11). In [2, Section 5.2] it is shown that this simple condition suffices for our purpose: it characterizes the optimal information cost as a function of $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$.

Theorem 1 ([2]) Suppose that $C : \Delta(\mathcal{X} \times \mathcal{Y}) \to [0, \log(|\mathcal{X} \times \mathcal{Y}|)]$ satisfies

(i) $C(\mu) = 0$ for every distribution $\mu$ such that $IC_\mu(f) = 0$, and
(ii) for every signal $B$ that can be sent by one of the players,

$$C(\mu) \leq I_B + \mathbb{E}_B[C(\mu_B)]. \quad (13)$$

Then $C(\mu) \leq IC_\mu(f)$.

Similarly if $IC_\mu(f)$ is replaced by $IC_\mu^{ext}(f)$, and $I_B$ is replaced by $I_B^{ext}$, then $C(\mu) \leq IC_\mu^{ext}(f)$.

Furthermore, in both of the external and the internal cases, it suffices to verify (ii) only for non-crossing unbiased signals $B$.

In light of Theorem 1, in order to determine the values of $IC_\mu(f)$, one has to first prove an upper bound by constructing a protocol (or a sequence of protocols) for every
distribution \( \mu \). Then it suffices to verify that the information cost of this protocol, as a function of \( \mu \), satisfies the conditions of Theorem 1.

A small gap in verifying Condition (ii) in [2]: Let \( \pi \) be the buzzers protocol for AND_2 function, in [2] \( \text{IC}_\mu(\pi) \) is computed to be an explicit formula for any given \( \mu \) (hence also for \( \mu_0 \) and \( \mu_1 \)), and thus the Condition (ii) in Theorem 1 becomes an explicit inequality that one should verify. However, in [2, Section 7.6, 7.8] Condition (ii) is only verified for non-crossing, unbiased, and \( \varepsilon \)-weak signals where \( \varepsilon \) depends on \( \mu \). This is not sufficient.

For example, set \( C(\mu) = K \geq 0 \) for distributions \( \mu \) such that \( \text{IC}_\mu(f) > 0 \), and otherwise set \( C(\mu) = 0 \). Obviously Condition (13) holds if \( C(\mu) = 0 \). On the other hand, if \( C(\mu) = K \), equivalently \( \text{IC}_\mu(f) > 0 \), then by continuity of \( \text{IC}_\mu(f) \) as a function of \( \mu \), if we choose the signal \( B \) to be \( \varepsilon \)-weak where \( \varepsilon \) is sufficiently small depending on \( \mu \), one also has \( \text{IC}_{\mu_0}(f) > 0 \) and \( \text{IC}_{\mu_1}(f) > 0 \), hence \( C(\mu_0) = C(\mu_1) = K \), and thus (13) holds in this case as well. Obviously, taking \( K \) to be sufficiently large violates the desired conclusion that \( C(\mu) \leq \text{IC}_\mu(f) \).

The function \( C \) in the previous example is not continuous. It turns out that, essentially, adding the continuity condition suffices. The proof of the following Theorem 2 is presented in Sect. 3.3.

**Theorem 2** Let \( w : (0, 1] \to (0, 1] \) be a non-decreasing function, \( \Omega \subseteq \Delta(\mathcal{X} \times \mathcal{Y}) \) be a subset of distributions containing the internal trivial distributions for function \( f \). Let \( \delta(\mu) \) denote the statistical distance of \( \mu \) from \( \Omega \). Suppose that \( C : \Delta(\mathcal{X} \times \mathcal{Y}) \to [0, \log(|\mathcal{X} \times \mathcal{Y}|)] \) satisfies

(i) \( C(\mu) = \text{IC}_\mu(f) \) if \( \delta(\mu) = 0 \),

(ii) for every non-crossing, unbiased, \( w(\delta(\mu)) \)-weak signal \( B \) that can be sent by one of the players,

\[
C(\mu) \leq I_B + \mathbb{E}_B[C(\mu_B)].
\]  

(14)

(iii) \( C(\mu) \) is uniformly continuous with respect to \( \mu \).

Then \( C(\mu) \leq \text{IC}_\mu(f) \).

Similarly, if we replace \( \Omega \) by a subset containing the external trivial distributions for function \( f \), and replace \( \text{IC}_\mu(f) \) by \( \text{IC}_\mu^{\text{ext}}(f) \), \( I_B \) by \( I_B^{\text{ext}} \), then \( C(\mu) \leq \text{IC}_\mu^{\text{ext}}(f) \).

Firstly, how to apply Theorem 2? Recall we discussed in Remark 2 that a protocol can depend on the input distribution \( \mu \) (such as the buzzers protocol). Suppose we have a protocol \( \pi_\mu \) (that depends on \( \mu \)) that we suspect is optimal. Think of the information cost \( \text{IC}_\mu(\pi_\mu) \) as a function of \( \mu \), and we apply Theorem 2 with \( C(\mu) := \text{IC}_\mu(\pi_\mu) \). The Condition (iii) is automatically true (since information cost is defined via entropy).

Next one verifies Condition (i). For example, if we choose \( \Omega \) to be the set of internal trivial distributions, i.e., distributions \( \nu \) such that \( \text{IC}_\nu(f) = 0 \), one needs to verify \( \text{IC}_\nu(\pi_\nu) = 0 \) for all such \( \nu \). The main work is to verify the concavity Condition (ii): given an input distribution \( \mu \) and suppose a non-crossing, unbiased, and \( \varepsilon \)-weak signal \( B \) is sent which results two new distributions \( \mu_0 := \mu|_{B=0} \) and \( \mu_1 := \mu|_{B=1} \), one needs to verify \( \text{IC}_\mu(\pi_\mu) \leq I_B + \mathbb{E}_B[\text{IC}_\mu_B(\pi_{\mu_B})] \). Indeed, [2] verified Condition (ii) in Theorem 2 for non-crossing, unbiased, and \( \varepsilon \)-weak signals.
Secondly, in the statement of Theorem 2, we explicitly allow the set $\Omega$ to be possibly larger than the set of trivial distributions. Consequently, the Condition (i) changes accordingly comparing to Condition (i) in Theorem 1. As we shall see in Sect. 4.5, this flexibility is useful when we verify the optimality of the multi-party protocol.

Thirdly, Theorem 2 is stated for functions $C$ instead of information costs of communication protocols directly, this allows to verify the optimality of possibly unconventional protocols directly instead of discretizing it first.

Finally, as we shall see in Sect. 3.3, the proof of Theorem 2 does not depend on the two-party model, i.e., Theorem 2 can be straightforwardly generalized into the multi-party setting. In fact, Condition (i) and (iii) does not depend on whether the underlying model is two-party or multi-party. For Condition (ii), we remarked before that one can generalize the non-crossing, unbiased, and weak signal into multi-party setting, and obviously one can also generalize the definition of $I_B$ in (10) naturally into the multi-party setting. We do not state separately a counterpart of Theorem 2 in the multi-party setting, as the modification will be only notational.

### 3.2 Communication Protocols as Random Walks on $\Delta(\mathcal{X} \times \mathcal{Y})$

One can view a communication protocol as a random walk on $\Delta(\mathcal{X} \times \mathcal{Y})$ (see [2,7]), and this sometimes simplifies our view towards a protocol and its information cost. We will need to apply it in proof of Theorem 2.

Consider a protocol $\pi$ and a prior distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$. Suppose that firstly Alice sends a random signal $B$ to Bob, we can interpret this as a random update of the prior distribution $\mu$ to a new distribution $\mu_0 := \mu|_{B=0}$ or $\mu_1 := \mu|_{B=1}$ depending on the value of $B$. Things are similar when Bob sends a signal. Therefore, we can think of a protocol as a random walk on $\Delta(\mathcal{X} \times \mathcal{Y})$ that starts at $\mu$, and every time that a player sends a signal, it moves to a new distribution with a step of random walk in $\Delta(\mathcal{X} \times \mathcal{Y})$.

Let $\Pi$ denote the random transcript of the protocol $\pi$. The random walk corresponding to the protocol $\pi$ starts at $\mu$ and terminates at $\mu_\Pi := \mu|_\Pi$. Since $\Pi$ is a random variable, $\mu_\Pi$ is also a random variable taking values in $\Delta(\mathcal{X} \times \mathcal{Y})$. Therefore, the distribution of $\mu_\Pi$ is a distribution on $\Delta(\mathcal{X} \times \mathcal{Y})$. For example, let $\Pi = t$ be a specific transcript (one can also think of $t$ as a leaf of the protocol tree, as they are one-to-one correspondent), then $\mu_t(xy) := \Pr[XY = xy|\Pi = t]$ is the updated distribution (from the prior distribution $\mu$) on the input space $\mathcal{X} \times \mathcal{Y}$ when an external observer sees the transcript $t$ (alternatively, when the protocol reaches the leaf $t$). The notation $\mu_\Pi$ is hence also called as “distribution on distributions” in [2, Section 7.2].

Recall $\text{IC^{ext}}_\mu(\pi) := I(XY; \Pi) = \mathbb{E}_{\pi \sim \Pi} D(XY|_{\Pi=\pi} \parallel XY)$. Note that the distribution of $XY|_{\Pi=t}$ is simply $\mu_t$. The above discussion shows the external information cost of the protocol $\pi$ is completely determined by the prior distribution $\mu$ and the final distribution of $\mu_\Pi$. To avoid confusion, let us emphasize again that different from $\mu$ which lies in $\Delta(\mathcal{X} \times \mathcal{Y})$, $\mu_\Pi$ is distributed on $\Delta(\mathcal{X} \times \mathcal{Y})$. Obviously, same argument also applies to the internal case via the similar identity $I(X; \Pi|Y) = \mathbb{E}_{\pi \sim \Pi, y \sim Y} D(X|_{\Pi=\pi, Y=y} \parallel X|_{Y=y})$ that defines the internal information cost.
**Proposition 1** \( ([2]) \) Let \( \pi \) and \( \tau \) be two communication protocols with the same input set \( X \times Y \) endowed with a distribution \( \mu \). Let \( \Pi \) and \( T \) denote the transcripts of \( \pi \) and \( \tau \), respectively. If \( \mu_{\Pi} \) has the same distribution as \( \mu_{T} \), then \( IC_{\mu}(\pi) = IC_{\mu}(\tau) \) and \( IC_{\mu}^{ext}(\pi) = IC_{\mu}^{ext}(\tau) \).

The proposition shows that it is only the final distribution \( \mu_{\Pi} \) on \( \Delta(X \times Y) \) that matters to the information cost. Let \( C_{T}^{\mu}(\Delta(X \times Y)) \) denote the set of all distributions on \( \Delta(X \times Y) \) that can be obtained, starting from the distribution \( \mu \), through communication protocols that perform a given communication task \( T \). The information cost of performing the task \( T \) is the infimum of the information costs of the distributions in \( C_{T}^{\mu}(\Delta(X \times Y)) \). Although this infimum is not always attained (see Remark 1), if one takes the closure of \( C_{T}^{\mu}(\Delta(X \times Y)) \) (under weak convergence) then one can replace the infimum with minimum. For the 2-bit AND function, the buzzers protocol of \([2]\) yields the distribution in the closure of \( C_{T}^{\mu}(\Delta(X \times Y)) \) that achieves the minimum information cost. We believe that the following is an important open problem.

**Problem 1** Define a paradigm such that for every communication task \( T \) and every distribution \( \mu \) on an input set \( X \times Y \), the set of distributions on \( \Delta(X \times Y) \) resulting from the protocols performing the task \( T \) in this paradigm is exactly equal to the closure of \( C_{T}^{\mu}(\Delta(X \times Y)) \).

### 3.3 Proof of Theorem 2

#### 3.3.1 A High-Level Discussion of the Proof Strategy

Note that to show \( C(\mu) \leq IC_{\mu}(f) \) is to show \( C(\mu) \leq IC_{\mu}(\pi) \) for any protocol \( \pi \) that performs the task \([f, 0]\). The local condition (14) naturally suggests an inductive proof of Theorem 2 by induction on the number of bits of the protocol.

**An inductive pseudo-proof** The base case: if a protocol performs \([f, 0]\) without communication, obviously \( IC_{\mu}(f) = 0 \), hence \( C(\mu) = 0 \) by Condition (i) in Theorem 2.

The induction hypothesis: suppose \( C(\mu) \leq IC_{\mu}(\pi) \) for any protocol \( \pi \) that performs \([f, 0]\) and sends at most \( k \) bits.

Now let \( \pi \) be a protocol that performs \([f, 0]\) and sends \( k + 1 \) bits. Consider the first bit \( B \) in \( \pi \), and assume \( B \) is non-crossing, unbiased, and weak (let us for now ignore the parameter of weakness), by Condition (ii) in Theorem 2,

\[
C(\mu) \leq I_{B} + E_{B}[C(\mu_{B})]. \tag{15}
\]

Let \( \pi_{0} \) and \( \pi_{1} \) denote the sub-protocol of \( \pi \) that starts at \( B = 0 \) and \( B = 1 \), respectively. Then \( \pi_{0} \) and \( \pi_{1} \) satisfy the inductive hypothesis. Hence

\[
C(\mu_{0}) \leq IC_{\mu_{0}}(\pi_{0}), \quad C(\mu_{1}) \leq IC_{\mu_{1}}(\pi_{1}). \tag{16}
\]

Note

\[
IC_{\mu}(\pi) = I_{B} + E_{B}[IC_{\mu_{B}}(\pi_{B})]. \tag{17}
\]
These together imply $C(\mu) \leq IC_\mu(\pi)$, complete the proof.

An immediate problem is: actually, what we proved is only for protocols in which all bits sent are non-crossing, unbiased, and weak. This is obviously insufficient. Hence we will first show a signal simulation lemma (Lemma 3) which says that any bit sent in a protocol can be simulated by a sequence of non-crossing, unbiased, and weak bits. This seems resolve the problem. However, a new issue is: this simulation might be infinite. In particular, if we use our signal simulation lemma to simulate a finite protocol, we might get a protocol that sends infinitely many bits. This hinders the induction. As we shall see, we will need to use the uniform continuity assumption of $C(\mu)$ and of $IC_\mu(f)$ with respect to $\mu$ to resolve this issue, and complete the proof.

3.3.2 The Signal Simulation Lemma

We start by restating a splitting lemma from [2]. Let $[\mu_0, \mu_1]$ denote the set of all convex combinations $\alpha \mu_0 + (1-\alpha) \mu_1$, where $\alpha \in [0, 1]$. Recall a signal simply means a random bit.

Lemma 2 (Splitting Lemma, [2]) Let $\mu \in \Delta(X \times Y)$ and a signal $B$ sent by one of the players, and let $\mu_b = \mu|_{B=b}$ for $b = 0, 1$. For any $\rho_0, \rho_1 \in [\mu_0, \mu_1]$ and $\rho \in (\rho_0, \rho_1)$, there exists a signal $B'$ that the same player can send starting at $\rho$ such that $\rho_b = \rho|_{B'=b}$ for $b = 0, 1$.

We emphasize that the simulation is done by the same player, which is obviously required in order to control the information leakage. Lemma 2 is proved in [2, Lemma 7.11], there is a minor mistake in the original statement as it is claimed that the lemma holds for $\rho \in [\rho_0, \rho_1]$ where the interval is closed.

We are now ready to prove the signal simulation lemma, which says any signal sent by one player can be perfectly simulated by the same player using a sequence of non-crossing, unbiased, and $\varepsilon$-weak signals. In terms of random walk, it shows that an one-step random walk can be simulated by a random walk consists of a sequence of possibly infinitely many smaller steps where each smaller step satisfies good properties. It generalizes [2, Lemma 5.2].

Although the proof of Lemma 3 is long, it is not difficult: it simply verifies that one can force the signal in each simulation step to be non-crossing, unbiased and weak, by appropriately choosing how far the smaller step in the simulation should jump from the current to the next.

Lemma 3 (Signal simulation) Given $0 < \varepsilon < 1$ and $\mu \in \Delta(X \times Y)$. Let $B$ be a signal sent by one of the players. Then there exists a sequence of non-crossing, unbiased, and $\varepsilon$-weak signals $B = (B_1 B_2 \ldots)$ that the same player can send, such that it terminates with probability 1, and furthermore $\mu|_B$ has the same distribution as $\mu|_B$.

**Proof** Let $\mu_0 = \mu|_{B=0}$ and $\mu_1 = \mu|_{B=1}$. Let $\Pr[B = 0] = t \in (0, 1)$, then $\mu = t\mu_0 + (1-t)\mu_1$. The following protocol explains how the sequence $(B_1 B_2 \ldots)$ is constructed from the signal $B$. 

 Springer
1. Set $\mu_c = \mu$ and $i = 1$;
2. Repeat until $\mu_c = \mu_0$ or $\mu_c = \mu_1$;
3. If $\mu_c \in [\mu_0, \mu_1]$, then
   
   Set $\lambda$ to be the largest value in $[0, \min\{1, \frac{t}{1-t}\}]$ satisfying
   
   $\lambda \max_{x,y} \frac{\mu_c(x,y) - \mu_0(x,y)}{\mu_c(x,y)} \leq \varepsilon$, and
   
   $\lambda |\mu_0(x,y) - \mu_0(x',y') - \mu_c(x,y) + \mu_c(x',y')| \leq \mu_c(x',y') - \mu_c(x,y)$, for every $(x,y) \neq (x',y')$ such that $\mu_c(x,y) < \mu_c(x',y')$.

   Send a signal $B_i$ that splits $\mu_c$ to $(1 - \lambda)\mu_c + \lambda \mu_0$ and $(1 + \lambda)\mu_c - \lambda \mu_0$;
4. If $\mu_c \notin (\mu, \mu_1]$, then
   
   Set $\lambda$ to be the largest value in $[0, \min\{1, \frac{1-t}{t}\}]$ satisfying
   
   $\lambda \max_{x,y} \frac{\mu_c(x,y) - \mu_1(x,y)}{\mu_c(x,y)} \leq \varepsilon$, and
   
   $\lambda |\mu_1(x,y) - \mu_1(x',y') - \mu_c(x,y) + \mu_c(x',y')| \leq \mu_c(x',y') - \mu_c(x,y)$, for every $(x,y) \neq (x',y')$ such that $\mu_c(x,y) < \mu_c(x',y')$.

   Send a signal $B_i$ that splits $\mu_c$ to $(1 - \lambda)\mu_c + \lambda \mu_1$ and $(1 + \lambda)\mu_c - \lambda \mu_1$;
5. Update $\mu_c$ to the current distribution;
6. Increase $i$;

Consider Step 3 The existence of the signal $B_i$ is guaranteed by the splitting Lemma 2, which requires both $(1 - \lambda)\mu_c + \lambda \mu_0$ and $(1 + \lambda)\mu_c - \lambda \mu_0$ lie inside $[\mu_0, \mu_1]$. This is obviously true for $(1 - \lambda)\mu_c + \lambda \mu_0$. We verify $(1 + \lambda)\mu_c - \lambda \mu_0 \in [\mu_0, \mu_1]$. Suppose $\mu_c = \alpha \mu_0 + (1 - \alpha)\mu_1$ for some $\alpha \in (0, 1)$. Recall $\mu = t\mu_0 + (1-t)\mu_1$. The assumption that $\mu_c \in [\mu_0, \mu]$ implies $\alpha \geq t$. Now

\[
(1 + \lambda)\mu_c - \lambda \mu_0 = ((1 + \lambda)\alpha - \lambda)\mu_0 + (1 + \lambda)(1 - \alpha)\mu_1. \tag{18}
\]

It suffices to verify $0 \leq (1 + \lambda)(1 - \alpha) \leq 1$. The nonnegativity is trivial. For the upper bound, one has

\[
1 + \lambda \leq 1 + \frac{t}{1-t} \leq 1 + \frac{\alpha}{1-\alpha} = \frac{1}{1-\alpha} \tag{19}
\]

as desired. Hence we have shown the existence of the signal $B_i$ in Step 3.

We proceed to verify the properties of $B_i$. Obviously it is unbiased. It is also $\varepsilon$-weak. Indeed,

\[
|\Pr[B_i = 0|x,y] - \Pr[B_i = 1|x,y]| = \left| \frac{\mu_c(x,y) \mathbb{1}[B_i = 0]}{2\mu_c(x,y)} - \frac{\mu_c(x,y) \mathbb{1}[B_i = 1]}{2\mu_c(x,y)} \right| = \lambda \frac{|\mu_c(x,y) - \mu_0(x,y)|}{\mu_c(x,y)},
\]

and the choice of $\lambda$ in (3.1) guarantees that this is bounded by $\varepsilon$. $B_i$ is also non-crossing. Indeed, if $\mu_c(x,y) < \mu_c(x',y')$, by our choice of $\lambda$, we have

\[
\lambda \left| \mu_0(x,y) - \mu_0(x',y') - \mu_c(x,y) + \mu_c(x',y') \right| \leq \mu_c(x',y') - \mu_c(x,y). \tag{20}
\]
Hence
\[
\lambda \left( \mu_0(x, y) - \mu_0(x', y') - \mu_c(x, y) + \mu_c(x', y') \right) \leq \mu_c(x', y') - \mu_c(x, y), \tag{21}
\]
or equivalently, \((1 - \lambda)\mu_c(x, y) + \lambda \mu_0(x, y) \leq (1 - \lambda)\mu_c(x', y') + \lambda \mu_0(x', y').\) Expanding the absolute value in (20) with a negative sign gives the other inequality.

Obviously the above analysis can be carried over the signal \(B_i\) in Step 4 analogously. Next to show the sequence \(B\) terminates with probability 1. Define
\[
\Omega = \{ \nu \in [\mu_0, \mu_1] : \exists (x, y), (x', y') \text{ s.t. } \nu(x, y) = \nu(x', y'), \text{ while } \mu_0(x, y) \neq \mu_0(x', y') \text{ or } \mu_1(x, y) \neq \mu_1(x', y'). \}
\]
Note that \(\Omega\) is a finite set. Indeed, we first observe that if \(\nu \neq \mu_0, \mu_1\), then “or” in the definition can be replaced by “and”, this is because \(\nu\) is a convex combination of \(\mu_0\) and \(\mu_1\), and if \(\mu_0(x, y) \neq \mu_0(x', y')\) but \(\mu_1(x, y) = \mu_1(x', y')\), then it is impossible that \(\nu(x, y) = \nu(x', y')\). Now suppose \(\nu \neq \mu_0, \mu_1\). Let \((x, y) \neq (x', y')\) such that \(\mu_0(x, y) \neq \mu_0(x', y')\) and \(\mu_1(x, y) \neq \mu_1(x', y')\), let \(\nu = \alpha \mu_0 + (1 - \alpha)\mu_1\) for some \(\alpha \in (0, 1).\) As \(\nu(x, y) = \nu(x', y')\), equivalently,
\[
\alpha \mu_0(x, y) + (1 - \alpha)\mu_1(x, y) = \alpha \mu_0(x', y') + (1 - \alpha)\mu_1(x', y'). \tag{22}
\]
This linear equation uniquely determines \(\alpha\) for the given \((x, y)\) and \((x', y')\), if any. This shows every pair \((x, y) \neq (x', y')\) corresponds to at most one \(\nu\), hence \(\Omega\) is finite.

Consider Step 3, i.e., \(\mu_c \in [\mu_0, \mu]\). If the value of \(\lambda\) is set by (3.1), then there is a uniform lower-bound for \(\lambda\):
\[
\lambda \geq \lambda_0 := \varepsilon \max_{xy} \frac{|\mu(x, y) - \mu_0(x, y)|}{\mu(x, y)} > 0. \tag{23}
\]
Indeed, since \(\mu_c \in [\mu_0, \mu]\), that is, \(\mu_c\) is closer to \(\mu_0\) than \(\mu\), it can be easily verified that
\[
\max_{xy} \frac{|\mu_c(x, y) - \mu(x, y)|}{\mu_c(x, y)} \leq \max_{xy} \frac{|\mu(x, y) - \mu_0(x, y)|}{\mu(x, y)} \quad \text{(here we also use the fact that } \sup \mu_c = \sup \mu, \text{ see (9)).}
\]
If \(\lambda\) is set by (3.2), i.e., there exist \((x, y) \neq (x', y')\) such that \(\mu_c(x, y) < \mu_c(x', y')\) and the equality in (3.2) holds for our choice of \(\lambda\) (since we choose the largest \(\lambda\)), equivalently, either \(\mu_c|_{B_i=0}(x, y) = \mu_c|_{B_i=0}(x', y')\) or \(\mu_c|_{B_i=1}(x, y) = \mu_c|_{B_i=1}(x', y')\), then at least one of \(\mu_c|_{B_i=0}\) and \(\mu_c|_{B_i=1}\) belongs to \(\Omega\).

One can analyze Step 4 analogously. Suppose the uniform lower bound corresponding to (23) that comes from Step 4 is \(\lambda_0'.\) Let \(\bar{\lambda} := \min\{\lambda_0, \lambda_0', \frac{1}{\min \nu}, \frac{1}{\max \nu}\}\).

Hence starting at any point \(\mu_c\), the random walk terminates with probability at least \(2^{-[(1/\bar{\lambda}) + |\Omega|]}\) after \(\lceil 1/\bar{\lambda} \rceil + |\Omega|\) steps. It follows that with probability 1, the random walk terminates.

Finally, since the sequence \(B\) is generated by an unbiased random walk in the interval \([\mu_0, \mu_1]\) that starts from \(\mu\) and eventually either reaches \(\mu_0\) or \(\mu_1\), the distributions of \(\mu|_B\) and \(\mu|_{\overline{B}}\) are the same. \(\square\)
3.3.3 Proof of Theorem 2

As we sketched in Sect. 3.3.1, the highlevel proof structure is a simulation (of an arbitrary protocol) followed by an induction. We first write the induction step as a technical lemma as follows.

**Lemma 4** Let \( w, \delta(\mu) \) and \( C \) be as in Theorem 2, and suppose \( C \) satisfies Conditions (i), (ii) and (iii). Let \( \tau \) be a protocol that terminates with probability 1, and further assume \( \tau \) is in normal form and every signal sent in \( \tau \) is \( \varepsilon \)-weak. Given a distribution \( \mu \in \Delta(\mathcal{X} \times \mathcal{Y}) \), for every node \( u \) in the protocol tree of \( \tau \), let \( \mu_u \) be the distribution conditioned on the event that the protocol reaches \( u \). If \( \mu \) satisfies \( w(\delta(\mu_u)) \geq \varepsilon \) for every internal node \( u \), then

\[
C(\mu) \leq IC_{\mu}(\tau) + \mathbb{E}_{\ell}[C(\mu_{\ell})],
\]

where the expected value is over all leaves \( \ell \) of \( \tau \) chosen according to the distribution (on the leaves) when the inputs are sampled according to \( \mu \).

**Proof** For every internal node \( u \), the assumption in the statement of the lemma implies that the signal sent from \( u \) is \( w(\delta(\mu_u)) \)-weak. Hence Condition (ii) of Theorem 2 shows the claim is true if \( \tau \) is a 1-bit protocol, and thus by a simple induction as we did in Sect. 3.3.1, it is true if \( \tau \) is a \( c \)-bit protocol for any \( c < \infty \).

Now assume \( \tau \) has infinite depth. Consider a large integer \( c \), obtain \( \tau_c \) by truncating \( \tau \) after \( c \) bits of communication, trivially \( IC_{\mu}(\tau_c) \leq IC_{\mu}(\tau) \). Let \( G_c \) denote the set of the leaves of \( \tau_c \) in which the protocol is forced to terminate. Let \( L_c \) be the set of leaves in \( \tau \) with depth at most \( c \). Clearly, the set of leaves in \( \tau_c \) is exactly \( G_c \cup L_c \). As \( \tau_c \) has a bounded depth, we have

\[
C(\mu) \leq IC_{\mu}(\tau_c) + \mathbb{E}_{\ell \in L_c \cup G_c}[C(\mu_{\ell})].
\]

Let \( \Pi_c \) denote the transcript of \( \tau_c \). As \( \tau \) terminates with probability 1, given any \( \alpha > 0 \), one can guarantee \( \Pr[\Pi_c(xy) \in G_c] < \alpha \) for every \( xy \) by choosing \( c \) to be sufficiently large. Hence

\[
C(\mu) \leq IC_{\mu}(\tau) + \mathbb{E}_{\ell \in L_c}[C(\mu_{\ell})] + \alpha \log(|\mathcal{X} \times \mathcal{Y}|).
\]

Taking the limit \( \alpha \to 0 \) shows \( C(\mu) \leq IC_{\mu}(\tau) + \mathbb{E}_{\ell \in \mathcal{L}}[C(\mu_{\ell})] \) where \( \mathcal{L} \) is the set of all leaves of \( \tau \).

**Proof of Theorem 2** Firstly by (i), \( \delta(\mu) = 0 \) implies \( C(\mu) = IC_{\mu}(f) \leq IC_{\mu}(f) \).

Hence assume \( \delta(\mu) > 0 \). Let \( \pi \) be an arbitrary protocol that performs \([f, 0]\), our aim is to show \( C(\mu) \leq IC_{\mu}(\pi) \).

For any \( \eta \) such that \( 0 < \eta < \delta(\mu) \), applying Lemma 3 one can get a new protocol \( \tilde{\pi} \) by replacing every signal sent in \( \pi \) with a random walk consisting of \( w(\eta) \)-weak, non-crossing and unbiased signals, and \( \tilde{\pi} \) terminates with probability 1. Moreover, since \( \tilde{\pi} \) is a perfect simulation of \( \pi \), by Proposition 1 we have \( IC_{\mu}(\pi) = IC_{\mu}(\tilde{\pi}) \).
For every node $v$ in the protocol tree of $\tilde{\pi}$, let $\mu_v$ be the distribution $\mu$ conditioned on the event that the protocol reaches the node $v$. Obtain $\tau$ from $\tilde{\pi}$ by terminating at every node $v$ that satisfies $\delta(\mu_v) \leq \eta$. Note that by the construction, Condition (ii) is satisfied on every internal node $v$ of $\tau$, as every such node satisfies $\delta(\mu_v) \leq \eta$, thus $w(\delta(\mu_v)) \geq w(\eta)$ implying the signal sent on node $v$ is $w(\delta(\mu_v))$-weak. Hence by Lemma 4,

$$C(\mu) \leq \text{IC}_\mu(\tau) + \mathbb{E}_\ell[C(\mu_\ell)], \quad (24)$$

where the expected value is over all leaves of $\tau$. For every $\mu_\ell$, let $\mu_\ell' \in \Omega$ be a distribution such that $\delta(\mu_\ell) = |\mu_\ell - \mu_\ell'|$. By Conditions (i) and (iii), and the uniform continuity of $\text{IC}_\mu(f)$, for every $\varepsilon > 0$ there exists $\eta > 0$, such that for all $\mu_\ell$, as long as $\delta(\mu_\ell) = |\mu_\ell - \mu_\ell'| \leq \eta$, then

$$C(\mu_\ell) \leq C(\mu_\ell') + \varepsilon = \text{IC}_{\mu_\ell'}(f) + \varepsilon \leq \text{IC}_{\mu_\ell}(f) + \varepsilon \leq \text{IC}_{\mu}(f) + 2\varepsilon. \quad (25)$$

As a result, $C(\mu) \leq \text{IC}_\mu(\tau) + \mathbb{E}_\ell[C_{\mu_\ell}(f)] + 2\varepsilon = \text{IC}_\mu(\tau) + \mathbb{E}_\ell[\text{IC}_{\mu_\ell}(f)] + 2\varepsilon$. Since $\mu_\ell$ is generated by truncating $\tilde{\pi}$, we have $\text{IC}_\mu(\tau) = \text{IC}_\mu(\tilde{\pi}) = \text{IC}_\mu(\pi)$. Therefore $C(\mu) \leq \text{IC}_\mu(\pi) + 2\varepsilon$. As this holds for arbitrary $\varepsilon$, we must have $C(\mu) \leq \text{IC}_\mu(\pi)$. \hfill $\Box$

4 The Multi-party AND Function in the Number-In-Hand Model

In [2] it is shown that in the two-player setting, a certain (unconventional) protocol that we refer to as the buzzers protocol, has optimal information and external information cost for the two-party AND$_2$ function. In this section we show that the buzzers protocol can be generalized to an optimal protocol for the multi-party AND$_k$ function in the number-in-hand model (assuming Assumption 1): the support of $\mu$ is a subset of \{0, 1, e, \ldots, e\}_k.

4.1 The Generalized Buzzers Protocol

For the sake of brevity, we denote $\mu_x := \mu(\{x\})$ for every $x \in \{0, 1\}^k$. Furthermore without loss of generality we assume that $\mu_{e_1} \leq \cdots \leq \mu_{e_k}$. The protocol is given by having buzzers with waiting times which have independent exponential distributions, and start at times $t_1, \ldots, t_k$ for players 1, ..., $k$, respectively. Although the protocol $\pi_\mu$ described in Fig. 1 is not a conventional communication protocol, it can be easily approximated by discretization and truncation of time.

For every $i \in [k]$, if $x_i = 0$ then the $i$th player activates a buzzer at time $t_i$ (later we call it the “activation time”) and becomes “active”. The protocol terminates with $\bigwedge_{i=1}^k x_i = 0$ once the first buzz happens, otherwise the time reaches $\infty$ without anyone buzzing, and they decide $\bigwedge_{i=1}^k x_i = 1$. Obviously, the protocol $\pi_\mu^\wedge$ performs the task [AND$_k$, 0], i.e., it computes AND$_k(x)$ correctly for all $x \in \{0, 1\}^k$.

Theorem 3 For every $\mu$ satisfying Assumption 1, the protocol $\pi_\mu^\wedge$ has the smallest external and internal information cost.
There is a clock whose time starts at 0 and increases continuously to $+\infty$.

Let $t_i := \ln \frac{\mu_{e_i}}{\mu_{e_1}}$ for $i = 1, \ldots, k$, and let $t_{k+1} := \infty$.

For every $i = 1, \ldots, k$, if $x_i = 0$ then the $i$-th player privately picks an independent random variable $T_i$ with exponential distribution with parameter $\lambda = 1$, and if time reaches $t_i + T_i$, the player announces that his/her input is 0, and the protocol terminates immediately with all the players knowing that $\bigwedge_{i=1}^k x_i = 0$.

If the clock reaches $+\infty$ without any player announcing their input, the players will know that $\bigwedge_{i=1}^k x_i = 1$.

Fig. 1 The protocol $\pi^\wedge_\mu$ for solving the multi-party AND function on a distribution $\mu$

Note that since $\mu$ is supported on $\{0, 1, e_1, \ldots, e_k\}$, it can be parametrized using $k + 2$ variables. However, the fact that $k$ is itself also a variable makes it impossible to directly compute the information cost $IC_\mu(\pi^\wedge_\mu)$ as what is done in [2] for the two-party case. We will show, through a series of reductions, that we only need to verify Condition (ii) in a small time interval that depends on the distribution $\mu$ and the weakness parameter $\epsilon$ of the signal, instead of the whole range, and furthermore, the distribution $\mu$ can in fact be parametrized by essentially 3 variables, enabling us to verify the concavity Condition (ii) in Theorem 2. See more discussion of the intuition of the reduction in Remark 3.

4.2 Notations and Set-Up

Let $\mu$ be a distribution satisfying Assumption 1, and $X = (X_1, \ldots, X_k) \in \{0, 1\}^k$ be the random $k$-bit input. Let $\Pi$ be the random transcript of protocol $\pi^\wedge_\mu$, and $\Pi_x = \Pi|_{X=x}$.

As we remarked after Theorem 2, the main work is to verify Condition (ii). Given the input distribution to be $\mu$, denote

$$\beta_s := \mu[X_s = 1], \quad \xi_s := \mu[X_s = 0] = 1 - \beta_s.$$  

Consider a signal $B$ sent by one player $s \in \{1, 2, \ldots, k\}$ satisfying,

$$\Pr[B = 0|X_s = 0] = \frac{1 + \varepsilon \beta_s}{2}, \quad \Pr[B = 1|X_s = 1] = \frac{1 + \varepsilon \xi_s}{2}. \quad (27)$$

It is easy to verify that $B$ is unbiased and $\varepsilon$-weak.

Let $\mu^0$ and $\mu^1$ denote the distributions of $X^0 := X|_{B=0}$ and $X^1 := X|_{B=1}$, then $\mu = \frac{\mu^0 + \mu^1}{2}$. Let $\Pi^0$ and $\Pi^1$ denote the random variables corresponding to the transcripts of $\pi^\wedge_{\mu^0}$ and $\pi^\wedge_{\mu^1}$, respectively. Then Condition (ii) is,

$$IC_\mu(\pi^\wedge_\mu) \leq I_B + \mathbb{E}_B IC_{\mu_B}(\pi^\wedge_{\mu_B}) = I_B + \frac{IC_{\mu^0}(\pi^\wedge_{\mu^0}) + IC_{\mu^1}(\pi^\wedge_{\mu^1})}{2} \quad (28)$$
for the internal case, and
\[
IC^{ext}_{\mu}(\pi^\wedge_\mu) \leq I_B^{ext} + \frac{IC^{ext}_{\mu_0}(\pi^\wedge_\mu) + IC^{ext}_{\mu_1}(\pi^\wedge_\mu)}{2}
\] (29)
for the external case.

The distributions \(\mu^0\) and \(\mu^1\) For \(B = 0\), if \(x_s = 0\), then
\[
\mu^0_x = \Pr[X = x|B = 0] = 2\Pr[B = 0|X = x]\mu_x = 2\Pr[B = 0|x_s = 0]\mu_x = (1 + \varepsilon\beta_s)\mu_x.
\]
Similarly one can compute the case when \(x_s = 1\). Hence we have,
\[
\mu^0_x = \begin{cases} 
(1 + \varepsilon\beta_s)\mu_x, & x_s = 0, \\
(1 - \varepsilon\zeta_s)\mu_x, & x_s = 1.
\end{cases}
\] (30)
By the fact \(\mu = \frac{\mu^0 + \mu^1}{2}\), one gets,
\[
\mu^1_x = \begin{cases} 
(1 - \varepsilon\beta_s)\mu_x, & x_s = 0, \\
(1 + \varepsilon\zeta_s)\mu_x, & x_s = 1.
\end{cases}
\] (31)

It is then easy to see that the signal \(B\) sent by player \(s\) is non-crossing if \(\varepsilon\) is sufficiently small. Recall we assumed \(\mu_{e_1} \leq \ldots \leq \mu_{e_k}\), the non-crossing property ensures that this order is also respected by \(\mu^0\) and \(\mu^1\).

The distributions of \(\Pi\), \(\Pi^0\) and \(\Pi^1\), and the intuition of reduction For convenience, we collect the terms and notations that we use frequently regarding the protocol \(\pi^\wedge_\mu\):

- activation time: denoted by \(t_i := \ln \frac{\mu_{e_i}}{\mu_{e_1}}\), player \(i\) becomes active at \(t_i\) if his/her input is 0;
- active: player \(i\) is active if he/she receives a 0 and the current time \(t \geq t_i\);
- active duration: the time duration that the player has been active before time \(t\);
- total active duration: the sum of active duration of all (active) players before time \(t\), denoted by \(\Phi_x(t)\);
- termination time: the time when the protocol terminates;
- \(s\): the player who sends the signal \(B\);
- \(m\): a generic player.

Note the random transcript \(\Pi\) contains the termination time \(t\), and if \(t < \infty\), also the name of the player who buzzed. We denote by \(\pi^\infty\) the transcript corresponding to termination time \(t = \infty\), and by \(\pi^m_t\) the termination time \(t < \infty\) with player \(m\) buzzing.

Suppose player \(i\) receives a bit 0. For \(t \in [0, \infty)\), the time duration that player \(i\) has been active is \(\max\{t - t_i, 0\}\). Let \(\Phi_x(t)\) denote the total active duration spent by all players before time \(t\) if the input is \(x\). For \(t_r \leq t < t_{r+1}\), we have
\[ \Phi_x(t) = \sum_{i \in \{1, 2, \ldots, k\} : x_i = 0} \max(t - t_i, 0) = \sum_{i \in \{1, 2, \ldots, r\}, x_i = 0} (t - t_i). \quad (32) \]

Note that player \( m \) can buzz at time \( t < \infty \), only if \( x_m = 0 \), \( t \geq t_m \), and no player has buzzed before time \( t \). Let \( f_x \) denote the probability density function of the random variable \( \Pi_x \), then

\[
f_x(\pi^m_t) = \begin{cases} 0, & t < t_m \text{ or } x_m = 1, \\ e^{-\Phi_x(t)}, & \text{otherwise}. \end{cases} \quad (33)\]

Also \( \Pr(\Pi_1 = \pi_\infty) = 1 \). Let \( f \) denote the probability density function of the random transcript \( \Pi \), then

\[
f(\pi^m_t) = \sum_x \mu_x f_x(\pi^m_t). \quad (34)\]

Denote \( f^0_x \), \( f^0 \) and \( f^1_x \), \( f^1 \) analogously for \( \Pi^0 \) and \( \Pi^1 \). Consider \( \pi^\wedge_\mu \) that runs on input distribution \( \mu \). Since the input distribution \( \mu \) is different from \( \mu \), the starting times in \( \pi^\wedge_\mu \) for players change accordingly. We call the starting time in \( \pi^\wedge_\mu \) for player \( i \) as new starting time, and denote it by \( t_i^0 := \ln \frac{\mu^0}{\mu_{\varepsilon_1}} \). By (30), we have \( t_i^0 = t_i \) for \( i \neq s \), and \( t_s^0 = t_s - \gamma_0 \) where

\[
\gamma_0 = \ln \left( \frac{1 + \varepsilon_\beta_s}{1 - \varepsilon_\xi_s} \right) > 0. \quad (35)\]

Hence

\[
\mu_x f^0_x(\pi^m_t) = \begin{cases} \mu^0_x f_x(\pi^m_t), & t < t_s - \gamma_0, \\ (1 + \varepsilon_\beta_s) \mu_x f_x(\pi^m_t) e^{-(t - t_s + \gamma_0)}, & t \in [t_s - \gamma_0, t_s), x_s = 0, m \neq s, \\ (1 - \varepsilon_\xi_s) \mu_x f_x(\pi^m_t), & t \in [t_s - \gamma_0, t_s), x_s = 1, m \neq s, \\ \mu^0_x f^0_x(\pi^m_t), & t \in [t_s - \gamma_0, t_s), m = s, \\ (1 - \varepsilon_\xi_s) \mu_x f_x(\pi^m_t), & t \geq t_s. \end{cases} \quad (36)\]

Similarly, the new starting times in \( \pi_\mu^\wedge \) are \( t_i^1 = t_i \) for \( i \neq s \), and \( t_s^1 = t_s + \gamma_1 \) where

\[
\gamma_1 = \ln \left( \frac{1 + \varepsilon_\xi_s}{1 - \varepsilon_\beta_s} \right) > 0. \quad (37)\]

Hence when \( m \neq s \),

\[
\mu_x f^1_x(\pi^m_t) = \begin{cases} \mu^1_x f_x(\pi^m_t), & t \leq t_s, \\ (1 - \varepsilon_\beta_s) \mu_x f_x(\pi^m_t) e^{t - t_s}, & t \in [t_s, t_s + \gamma_1), x_s = 0, \\ (1 + \varepsilon_\beta_s) \mu_x f_x(\pi^m_t), & t \in [t_s, t_s + \gamma_1), x_s = 1, \\ (1 + \varepsilon_\beta_s) \mu_x f_x(\pi^m_t), & t \geq t_s + \gamma_1. \end{cases} \quad (38)\]
and for \( m = s \),
\[
\mu_x f_x^1(\pi_s^s) = \begin{cases} 
(1 + \varepsilon_s)\mu_x f_x(\pi_s^s), & t > t_s + \gamma_1 \text{ and } x_s = 0, \\
0, & \text{otherwise}.
\end{cases}
\tag{39}
\]

**Remark 3** (The numbers \( \gamma_0 \) and \( \gamma_1 \), and the intuition) The number \( \gamma_0 = t_s - t_s^0 > 0 \), by definition, denotes how much earlier that player \( s \) can become active in \( \pi_{\mu_0}^\wedge \) than in \( \pi_{\mu_0}^\cdot \). Similarly, \( \gamma_1 \) denotes how much later that player \( s \) can become active in \( \pi_{\mu_1}^\wedge \) than in \( \pi_{\mu_1}^\cdot \). We emphasize that the activation time for all players except player \( s \) is the same among \( \pi_{\mu_0}^\wedge \), \( \pi_{\mu_0}^\cdot \) and \( \pi_{\mu_1}^\wedge \). Hence, intuitively it seems plausible that the time interval \([-\gamma_0, \gamma_1]\) plays an important role towards verifying Condition (ii) of Theorem 2. This observation as we shall see in Sect. 4.3 is indeed the case, and it is the key to our reduction.

**The Condition (ii) of Theorem 2** Consider now the Condition (ii) given in (29) for external case. Let \( \phi(x) := x \ln(x) \). By definition of information cost,
\[
IC_{\mu}^{\text{ext}}(\pi_\mu^\wedge) = I(\Pi; X) = H(\Pi) - H(\Pi|X) = H(\Pi) - \sum_x \mu_x H(\Pi_x)
\]
\[
= - \int_{-\infty}^{\infty} \sum_m \phi(f(\pi^m_x)) + \sum_x \mu_x \int_{-\infty}^{\infty} \sum_m \phi(f_x(\pi^m_x)).
\tag{40}
\]

One can write down \( IC_{\mu_0}^{\text{ext}}(\pi_\mu^\wedge) \) and \( IC_{\mu_1}^{\text{ext}}(\pi_\mu^\wedge) \) by replacing \( f, f_x \) with \( f^0, f_x^0 \) and \( f^1, f_x^1 \), respectively. Now
\[
I_B^{\text{ext}} = I(B; X) = H(X) - H(X|B) = H(X) - \frac{H(X^0) + H(X^1)}{2}
\]
\[
= - \sum_x \phi(\mu_x) + \frac{\sum_x \phi(\mu_x^0) + \sum_x \phi(\mu_x^1)}{2}.
\tag{41}
\]

Use the fact \( \int_{-\infty}^{\infty} \sum_m f_x(\pi^m_x) = 1 \) for all \( x \in \{0, 1\}^k \) (since \( f_x \) is the PDF function), and \( \phi(x) = x \ln(x) \), we have
\[
\sum_x \phi(\mu_x) + \sum_x \mu_x \int_{-\infty}^{\infty} \sum_m \phi(f_x(\pi^m_x))
= \sum_x \phi(\mu_x) \int_{-\infty}^{\infty} \sum_m f_x(\pi^m_x) + \sum_x \mu_x \int_{-\infty}^{\infty} \sum_m \phi(f_x(\pi^m_x))
= \int_{-\infty}^{\infty} \sum_x \sum_m \mu_x f_x(\pi^m_x) \left( \ln(\mu_x) + \ln(f_x(\pi^m_x)) \right)
= \int_{-\infty}^{\infty} \sum_x \sum_m \phi(\mu_x f_x(\pi^m_x)).
\tag{42}
\]
Combining (40)–(42), the Condition (ii) for external information cost given in (29) is equivalent to,
\[ \int_{-\infty}^{\infty} \text{concav}^{ext}_\mu (t) \, dt \geq 0, \tag{43} \]
in which
\[ \text{concav}^{ext}_\mu (t) := \sum_m \left( \phi(f(\pi_t^m)) - \frac{\phi(f^0(\pi_t^m)) + \phi(f^1(\pi_t^m))}{2} \right) \]
\[ - \sum_m \sum_x \left( \phi(\mu_x f_x(\pi_t^m)) - \frac{\phi(\mu_x^0 f_x^0(\pi_t^m)) + \phi(\mu_x^1 f_x^1(\pi_t^m))}{2} \right). \tag{44} \]

A similar calculation shows for the internal case, the condition (28) is equivalent to,
\[ \sum_{j=1}^{k} \int_{-\infty}^{\infty} \text{concav}_\mu (t, j) \, dt \geq 0, \tag{45} \]
in which
\[ \text{concav}_\mu (t, j) := \sum_m \sum_{b=0}^{1} \left( \phi(f_{x,j=b}(\pi_t^m)) - \frac{\phi(f_{x,j=b}^0(\pi_t^m)) + \phi(f_{x,j=b}^1(\pi_t^m))}{2} \right) \]
\[ - \sum_m \sum_x \left( \phi(\mu_x f_x(\pi_t^m)) - \frac{\phi(\mu_x^0 f_x^0(\pi_t^m)) + \phi(\mu_x^1 f_x^1(\pi_t^m))}{2} \right). \tag{46} \]

where \( f_{x,j=b}(\pi_t^m) := \sum_{X: X_j=b} \mu_X f_X(\pi_t^m) \), and \( f_{x,j=b}^0(\pi_t^m) \) and \( f_{x,j=b}^1(\pi_t^m) \) are similarly defined.

### 4.3 Reductions

Our goal is to verify (43) and (45), which are integrations over the time range from \(-\infty\) to \(\infty\). Using the memoryless property of exponential distribution, one can shift the activation time of all the players by \(-\ln(\mu_{e_s}/\mu_{e_1})\), and assume that \(t_1 = -\ln(\mu_{e_s}/\mu_{e_1}), \ldots, t_s = 0, \ldots, t_k = \ln(\mu_{e_k}/\mu_{e_s})\). It turns out that the integration in (43) and (45) on the ranges \((-\infty, -\gamma_0]\) and \([\gamma_1, \infty]\) are both nonnegative, hence it suffices to focus on the small time interval \([-\gamma_0, \gamma_1]\) which depends on the distribution \(\mu\) and the (weakness) parameter \(\epsilon\). Furthermore, this fact also enables us to parametrize the distribution \(\mu\) using essentially 3 parameters. This makes it possible to verify Condition (ii) of Theorem 2 using the software Wolfram Mathematica.

Firstly we analyze the information cost of the distribution \(\mu\) that is uniform on \(\{e_1, \ldots, e_k\}\), i.e., \(\mu_{e_1} = \cdots = \mu_{e_k} = 1/k\). For the purpose of reducing the number of variables to parametrize the distribution \(\mu\) from arbitrary to 3, this step is not
necessary. However, it turns out that to verify Condition (ii) it is necessary to analyze this distribution separately. Later we will choose $Ω$ (see this notation in Theorem 2) to include this $μ$ together with trivial distributions.

**Statement 1** Let $μ$ be the distribution $μ_{e_1} = \cdots = μ_{e_k} = 1/k$. The internal and external information cost of the protocol $π^∧$ is optimal with respect to $μ$.

Observe that all the information will be revealed if the input is $I$, that is, every player receives a 1. This special case can be removed from the analysis.

**Statement 2** To verify (43) and (45), it suffices to assume $μ$ satisfies $μ(I) = 0$.

Our third major reduction shows that $\int_{−γ_0}^{∞} \text{concav}_{μ}^{ext}(t) dt \geq 0$ and $\int_{γ_0}^{∞} \text{concav}_{μ}^{ext}(t) dt = 0$. Similarly for the internal case. This shows, as we mentioned before, that we can focus on a small time interval $[−γ_0, γ_1]$.

**Statement 3** Assuming $μ(I) = 0$, it suffices to verify $\int_{−γ_0}^{γ_1} \text{concav}_{μ}^{ext}(t) dt \geq 0$ and $\sum_{j=1}^{k} \int_{−γ_0}^{γ_1} \text{concav}_{μ}(t, j) dt \geq 0$.

Lastly we parametrize the distribution $μ$ with essentially 3 parameters (i.e., $s$, $β$ and $γ_0$). Firstly we observe that conditioned on the buzz time $t \in [−γ_0, γ_1]$, we have $μ_{e_1}|t≥t_1−γ_0 = \cdots = μ_{e_{s−1}}|t≥t_1−γ_0$. Secondly, we show that one can transfer the mass on those $e_j$ where $μ_{e_j} > μ_{e_j}$ to $μ_0$.

**Statement 4** To verify $\int_{−γ_0}^{γ_1} \text{concav}_{μ}^{ext}(t) dt \geq 0$ and $\sum_{j=1}^{k} \int_{−γ_0}^{γ_1} \text{concav}_{μ}(t, j) dt \geq 0$, it suffices to assume $μ$ satisfies $μ_{e_1} = \cdots = μ_{e_{s−1}} = β$, $μ_{e_s} = \cdots = μ_{e_k} = e^{γ_0} β$, and $μ_0 = 1 − (s−1)β − (k−s+1)e^{γ_0} β$, where $0 < β < 1$.

### 4.4 Proof of Reductions

**Statement 1** Let $μ$ be the distribution $μ_{e_1} = \cdots = μ_{e_k} = 1/k$. The internal and external information cost of the protocol $π^∧$ is optimal with respect to $μ$.

**Proof** First we present the proof for the external information complexity. Let $\hat{π}$ be any conventional (finite) protocol that solves $\text{AND}_k, 0$, and $\hat{I}$ denote the random transcript of $\hat{π}$. Let $τ$ be a possible transcript of the protocol $\hat{π}$. Firstly note that it is not possible to have $Pr[\hat{π}_{em} = τ] > 0$ for all $1 \leq m \leq k$. Indeed by rectangle property this would imply $Pr[\hat{π}_1 = τ] > 0$, and since the correct output for $I$ is different from that of $e_1, \ldots, e_k$, we would get a contradiction with the assumption that $\hat{π}$ solves $\text{AND}_k$ correctly on all inputs. Hence to every transcript $τ$, we can assign a $j(τ) ∈ \{1, \ldots, m\}$ with $Pr[\hat{π}_{e_j} = τ] = 0$. Now for a random $X ∼ μ$, denote $J = j(\hat{π}_X)$, and notice that conditioned on $J = j$, $X$ is supported on the set $\{e_1, \ldots, e_k\} \setminus \{e_j\}$ of size $k−1$, and thus $H(X|J) ≤ log(k−1)$. Consequently, we have

$$IC_{μ}^{ext}(\hat{π}) = I(X; \hat{I}_X)$$
$$= I(X; \hat{I}_X J) \geq I(X; J) = H(X) − H(X|J) \geq log k − log(k−1).$$
On the other hand, consider our protocol $\pi^\wedge$ and let $\Pi$ denote its random transcript. Note that under $\mu$, all players are activated at the same time, and consequently by symmetry, for every termination time $t$ and player $j \in \{1, \ldots, k\}$, the random variable $X_{\Pi_{X=(t,j)}}$ is uniformly distributed on $\{e_1, \ldots, e_k\} \setminus \{e_j\}$. Hence $H(X_{\Pi_X}) = \log (k - 1)$. We conclude that

$$IC^\text{ext}_\mu (\pi^\wedge) = I(X; \Pi_X) = H(X) - H(X_{\Pi_X}) = \log k - \log (k - 1).$$

Next we turn to the internal case. Again, let $\hat{\pi}$ be any protocol that solves the multi-party AND$_k$ correctly, and let $J$ be defined as above. First note that for $i \in [k]$, $X_{\hat{\pi}=i}$ is supported on the single point $\{e_i\}$ (hence $H(X_{\hat{\pi}=i}) = 0$) and $X_{\hat{\pi}=0}$ is uniformly distributed on $\{e_1, \ldots, e_k\} \setminus \{e_i\}$. Hence

$$H(X_{\hat{\pi}=i}) = \frac{1}{k} H(X_{\hat{\pi}=i} = 1) + \frac{k - 1}{k} H(X_{\hat{\pi}=i} = 0) = \frac{k - 1}{k} \log (k - 1).$$

Moreover for $i, j \in [k]$, $X_{J=j, \hat{\pi}=0}$ is supported on $\{e_1, \ldots, e_k\} \setminus \{e_i, e_j\}$. Hence using $\Pr[J=i] = \Pr[J=i, X_i=0]$ and $H(X_{\hat{\pi}=i}) = 0$, we have

$$H(X|J_{X_i}) = \sum_{j=1}^{k} \Pr[J = j, X_i = 0] H(X_{J=j, X_i=0}) \leq \sum_{j=1}^{k} \Pr[J = j, X_i = 0] \log |\{e_1, \ldots, e_k\} \setminus \{e_i, e_j\}|$$

$$= \frac{k - 1}{k} \log (k - 2) + \Pr[J = i](\log (k - 1) - \log (k - 2)).$$

Summing over $i$, we obtain

$$\sum_{i=1}^{k} H(X|J_{X_i}) \leq (k - 2) \log (k - 2) + \log (k - 1),$$

and thus

$$IC_{\mu}(\hat{\pi}) = \sum_{i=1}^{k} I(X; \hat{\pi}_{X|X_i}) = \sum_{i=1}^{k} I(X; \hat{\pi}_{J|X_i}) \geq \sum_{i=1}^{k} I(X; J|X_i) = \sum_{i=1}^{k} H(X|X_i) - H(X|J_{X_i}) \geq (k - 1) \log (k - 1) - ((k - 2) \log (k - 2) + \log (k - 1))$$

$$= (k - 2)(\log (k - 1) - \log (k - 2)).$$
On the other hand, for the protocol $\pi^\wedge$, by symmetry, for every termination time $t$ and player $j \in \{1, \ldots, k\}$, the random variable $X|_{\Pi_X=(t,j),X_i=0}$ is uniformly distributed on $\{e_1, \ldots, e_k\} \setminus \{e_j, e_t\}$. Hence

$$H(X|\Pi_X X_i) = \frac{1}{k} \log(k - 1) + \frac{k - 2}{k} \log(k - 2).$$

We conclude that

$$\IC_{\mu}(\pi^\wedge) = \sum_{i=1}^{k} I(X; \Pi_X | X_i) = (k - 1) \log(k - 1) - \sum_{i=1}^{k} H(X|\Pi_X X_i)$$

$$= (k - 2)(\log(k - 1) - \log(k - 2)).$$

\[\square\]

**Statement 2** To verify (43) and (45), it suffices to assume $\mu$ satisfies $\mu(1) = 0$.

**Proof** We will show if $\pi^\wedge$ is optimal for distributions that have zero mass on $1$, then $\pi^\wedge$ is also optimal for general distributions that can have non-zero mass on $1$. This will prove the statement.

Let $\mu$ be an arbitrary distribution, and let $\pi$ be a protocol that solves the multi-party AND$_k$ function correctly on all inputs. Let $\Pi$ denote the transcript of this protocol, and let $R_X = 1_{X=1}$. Since $\pi$ solves the AND$_k$ correctly, $\Pi$ determines the value of $R_X$, that is $H(R_X|\Pi_X) = 0$. Hence

$$\IC_{\mu}(\pi) = I(X; \Pi_X) = I(X R_X; \Pi_X) = I(R_X; \Pi_X) + I(X; \Pi_X | R_X)$$

$$= H(R_X) + \Pr[X = 1] I(X; \Pi_X | X = 1) + \Pr[X \neq 1] I(X; \Pi_X | X \neq 1)$$

$$= H(\mu_1) + (1 - \mu_1) \IC_{\mu'}(\pi),$$

where $\mu'$ is the distribution $\mu$ conditioned on the event that the input is not equal to $1$, and $H(\mu_1)$ denotes the binary entropy with the parameter $\mu_1$.

Suppose we have shown that $\pi^\wedge$ is optimal for all distributions $\nu$ that satisfies $\nu_1 = 0$, that is, $\IC_{\nu}(\pi^\wedge) \leq \IC_{\nu}(\pi)$ for any protocol $\pi$. Then, by the equality we just established, one has

$$\IC_{\mu}(\pi^\wedge) = H(\mu_1) + (1 - \mu_1) \IC_{\mu'}(\pi^\wedge)$$

$$\leq H(\mu_1) + (1 - \mu_1) \IC_{\mu'}(\pi) = \IC_{\mu}(\pi).$$

That is, $\pi^\wedge$ will be optimal for distributions that can have non-zero mass on input $1$. The argument for internal case is similar. \[\square\]

**Statement 3** Assuming $\mu(1) = 0$, it suffices to verify $\int_{-\gamma_0}^{\gamma_0} \text{concav}_{\mu}^e(t) dt \geq 0$ and $\sum_{j=1}^{k} \int_{-\gamma_0}^{\gamma_0} \text{concav}_{\mu}(t, j) dt \geq 0$. 

\[\square\]
Proof Firstly we show \(\int_{\gamma_1}^{\infty} \text{concav}^\text{ext}_\mu(t) dt = 0\) and \(\sum_{j=1}^{k} \int_{\gamma_1}^{\infty} \text{concav}_\mu(t, j) dt = 0\). Indeed, as \(t \geq \gamma_1\), one has \(\mu^0_x f^0_x(\pi^m_t) = (1 - \varepsilon \xi_s) \mu_x f_x(\pi^m_t)\) and \(\mu^1_x f^1_x(\pi^m_t) = (1 + \varepsilon \xi_s) \mu_x f_x(\pi^m_t)\). A direct calculation shows for \(t \geq \gamma_1\),

\[
\phi(\mu_x f_x(\pi^m_t)) - \phi(\mu^0_x f^0_x(\pi^m_t)) + \phi(\mu^1_x f^1_x(\pi^m_t))
\]

\[
= - \frac{\phi(1 - \varepsilon \xi_s) + \phi(1 + \varepsilon \xi_s)}{2} \mu_x f_x(\pi^m_t),
\]

and

\[
\phi(f(\pi^m_t)) - \phi(f^0(\pi^m_t)) + \phi(f^1(\pi^m_t)) = - \frac{\phi(1 - \varepsilon \xi_s) + \phi(1 + \varepsilon \xi_s)}{2} f(\pi^m_t).
\]

Plug them into (44) shows \(\text{concav}^\text{ext}_\mu(t) = 0\) for \(t \geq \gamma_1\). Hence \(\int_{\gamma_1}^{\infty} \text{concav}^\text{ext}_\mu(t) dt = 0\). Similarly one can calculate the internal case.

Next we show \(\int_{-\infty}^{-\gamma_0} \text{concav}^\text{ext}_\mu(t) dt \geq 0\) and \(\sum_{j=1}^{k} \int_{-\infty}^{-\gamma_0} \text{concav}_\mu(t, j) dt \geq 0\). Observe that \(\Pi\), \(\Pi^0\) and \(\Pi^1\) are identical up to time \(-\gamma_0\). Let \(\Pi_P\) denote a similar protocol, with the only difference that in \(\Pi_P\) at time \(t = -\gamma_0\) all the players reveal their inputs. Then,

\[
\int_{-\infty}^{-\gamma_0} \text{concav}^\text{ext}_\mu(t) dt = H(X|\Pi_P) - H(X|\Pi_P, B) \geq 0.
\]

The equality holds because both denote the information revealed until time \(\gamma_0\) by the (same) protocol. Similarly,

\[
\sum_{j=1}^{k} \int_{-\infty}^{-\gamma_0} \text{concav}_\mu(t, j) dt = \sum_{j=1}^{k} (H(X|X_j, \Pi_P) - H(X|X_j, \Pi_P, B)) \geq 0.
\]

\(\square\)

Statement 4 To verify \(\int_{-\gamma_0}^{\gamma_1} \text{concav}^\text{ext}_\mu(t) dt \geq 0\) and \(\sum_{j=1}^{k} \int_{-\gamma_0}^{\gamma_1} \text{concav}_\mu(t, j) dt \geq 0\), it suffices to assume \(\mu\) satisfies \(\mu_{e_1} = \cdots = \mu_{e_{s-1}} = \beta, \mu_{e_s} = \cdots = \mu_{e_k} = e^{\gamma_0} \beta\), and \(\mu_0 = 1 - (s - 1)\beta - (k - s + 1)e^{\gamma_0} \beta\), where \(0 < \beta < 1\).

We formulate a technical lemma to facilitate the proof of Statement 4.

Lemma 5 For every \(z\),

\[
\Pr[X = z \wedge t(\Pi^0_z) \in [-\gamma_0, \gamma_1]]
\]

\[
= \frac{\Pr[X^0 = z \wedge t(\Pi^0_z) \in [-\gamma_0, \gamma_1]] + \Pr[X^1 = z \wedge t(\Pi^1_z) \in [-\gamma_0, \gamma_1]]}{2}.
\]
Proof We need to show
\[
\mu_z \sum_m \int_{-\gamma_0}^{\gamma_1} f_z(\pi^m_z) dt = \frac{1}{2} \left( \mu_z^0 \sum_m \int_{-\gamma_0}^{\gamma_1} f_z^0(\pi^m_z) dt + \mu_z^1 \sum_m \int_{-\gamma_0}^{\gamma_1} f_z^1(\pi^m_z) dt \right).
\]

Recall that \(\Phi_z(t)\) denotes the total amount of active duration spent by all players before time \(t\). The probability that \(\Pi_z\) finishes in the interval \([-\gamma_0, \gamma_1]\) is equal to
\[
e^{-\Phi_z(-\gamma_0)} - e^{-\Phi_z(\gamma_1)}.
\]

Denoting by \(\Phi^0_z(t)\) and \(\Phi^1_z(t)\) the total active duration for the protocols \(\pi^0_{\mu_0}\) and \(\pi^1_{\mu_1}\) on the input \(z\), the claim is equivalent to
\[
\mu_z \cdot (e^{-\Phi_z(-\gamma_0)} - e^{-\Phi_z(\gamma_1)})
\]
\[
= \frac{\mu_z^0 \cdot (e^{-\Phi^0_z(-\gamma_0)} - e^{-\Phi^0_z(\gamma_1)}) + \mu_z^1 \cdot (e^{-\Phi^1_z(-\gamma_0)} - e^{-\Phi^1_z(\gamma_1)})}{2}.
\]

Since \(\mu_z = \frac{\mu_z^0 + \mu_z^1}{2}\) and \(\Phi_z(-\gamma_0) = \Phi^0_z(-\gamma_0) = \Phi^1_z(-\gamma_0)\), the equality reduces to
\[
\mu_z e^{-\Phi_z(\gamma_1)} = \frac{\mu_z^0 e^{-\Phi^0_z(\gamma_1)} + \mu_z^1 e^{-\Phi^1_z(\gamma_1)}}{2}.
\]

When \(z = 1\), \(\Phi_z = \Phi^0_z = \Phi^1_z\), and thus \(\mu_z = \frac{\mu_z^0 + \mu_z^1}{2}\) verifies the equality. In the case of \(z = 0\), we have that \(\Phi^0_z(\gamma_1) = \Phi_z(\gamma_1) + \gamma_0\), and \(\Phi^1_z(\gamma_1) = \Phi_z(\gamma_1) - \gamma_1\). Substituting \(\gamma_0 = \ln \left( \frac{1 + \epsilon \beta_s}{1 - \epsilon \beta_s} \right)\), \(\gamma_1 = \ln \left( \frac{1 + \epsilon \beta_s}{1 - \epsilon \beta_s} \right)\), \(\mu_z^0 = (1 + \epsilon \beta_s) \mu_z\) and \(\mu_z^1 = (1 - \epsilon \beta_s) \mu_z\) verifies the equality. \(\square\)

Now we prove Statement 4.

Proof Recall that it is player \(s\) who sends a signal \(B\), and we have shifted the activation time of every player so that \(t_s = 0\) (see the discussion at the beginning of Sect. 4.3). Recall also \(-\gamma_0\) is the time that player \(s\) can become active in the protocol \(\pi^0_{\mu_0}\), and \(\gamma_1\) is the time that player \(s\) can become active in the protocol \(\pi^1_{\mu_1}\). Throughout the proof, we confine ourselves in the time interval \(t \in [-\gamma_0, \gamma_1]\).

Firstly we observe that conditioned on the buzz time \(t \in [-\gamma_0, \gamma_1]\), we have \(\mu_{e_1} \mid t \geq t_s - \gamma_0 = \cdots = \mu_{e_{s-1}} \mid t \geq t_s - \gamma_0\). Hence we can assume \(\mu_{e_1} = \cdots = \mu_{e_{s-1}}\).

Secondly, we will show that we can also assume \(\mu_{e_j} = \mu_{e_s}\) for all \(j \geq s\).

Suppose \(\mu\) satisfies \(\mu_{e_s} = \cdots = \mu_{e_{s-1}} < \mu_{e_{s+a+1}}\) for some \(a \geq 0\) with \(s+a+1 \leq k\). Let \(\mu'\) be a distribution on \(\{0, 1\}^{s+a}\) defined as: \(\mu' = \mu_0 + \sum_{j>s+a} \mu_{e_j}\), and \(\mu_{e_j} = \mu_{e_j}\) for \(1 \leq j \leq s + a\). Note that we have used the notation \(e_j\) to denote the \(j\)th basis vector, and \(0\) to denote the all-0 vector in both \(\{0, 1\}^{s+a}\) and \(\{0, 1\}^k\). Note that \(\mu'\) is a distribution satisfying Assumption 1 where the number of players is \(s + a\), and furthermore \(\mu'_{e_j} = \mu_{e_j}\) for all \(s \leq j \leq s + a\). Hence we can consider to verify the Condition (ii) for the protocol \(\pi^1_{\mu'}\) that solves AND\(_{s+a}\) correctly. We will show,
(1) \[ \int_{-\gamma_0}^{\gamma_1} \text{concave}^{\text{ext},t} \mu (t) dt = \int_{-\gamma_0}^{\gamma_1} \text{concave}^{\text{ext},t} \mu (t) dt; \]

(2) If \[ \sum_{j=1}^{s+a} \int_{-\gamma_0}^{\gamma_1} \text{concave}^{\mu, t} (j) dt \geq 0, \] then \[ \sum_{j=1}^{k} \int_{-\gamma_0}^{\gamma_1} \text{concave}^{\mu, t} (j) dt \geq 0. \]

Consider the protocol \( \pi^\wedge \). Recall \( \gamma_1 \) as defined in (37) depends on \( \epsilon \), let \( \epsilon \) be sufficiently small so that \( \gamma_1 \leq t_{s+a+1}. \) That is to say, in the protocol \( \pi^\wedge \), player \( s+a+1 \) (and hence all players after player \( s+a+1 \)) can become active only after time \( \gamma_1 \). Hence for all \( t \in [-\gamma_0, \gamma_1] \),

\[ f (\pi_t^m) = 0, \quad \forall m \geq s+a+1. \] (48)

Since the activation time for all players except player \( s \) is the same among \( \pi^\wedge, \pi^\wedge_0 \) and \( \pi^\wedge_1 \), we also have \( f^0 (\pi_t^m) = f^1 (\pi_t^m) = 0 \) for every player \( m \geq s+a+1 \).

From now on we assume \( m \leq s+a \). Intuitively, we “forget” players after \( s+a \) in \( \pi^\wedge, \pi^\wedge_0 \) and \( \pi^\wedge_1 \), this enables us to compare \( \pi^\wedge \) (where there are \( k > s+a \) players) and \( \pi^\wedge_{\mu} \) (where there are \( s+a \) players). Let \( f, f', \Pi, \Pi' \) denote the PDFs and protocols for \( \pi^\wedge \) and \( \pi^\wedge_{\mu} \), respectively.

(1) Consider \( \pi^\wedge_{\mu} \), for every \( m = 1, 2, \ldots, s+a \), and \( s+a \leq j \leq k \),

\[ f_0 (\pi_t^m) = f_{e_j} (\pi_t^m). \] (49)

By the definition of \( \mu' \), obviously we have for every \( m = 1, 2, \ldots, s+a \),

\[ f'_0 (\pi_t^m) = f_0 (\pi_t^m), \quad \text{and} \quad f'_{e_j} (\pi_t^m) = f_{e_j} (\pi_t^m), \quad 1 \leq j \leq s+a. \] (50)

Hence (50) and (49) imply that

\[ f (\pi_t^m) = f' (\pi_t^m), \quad \forall m = 1, 2, \ldots, s+a. \] (51)

Clearly similar results hold for \( \Pi^0, \Pi^1 \) and \( \Pi'^0, \Pi'^1 \). This implies that the first integral in (43) does not change from \( \mu \) to \( \mu' \).

It remains to show that the second integral in (43) does not change either. Expanding this integral gives,

\[
\int_{-\gamma_0}^{\gamma_1} \sum_X \sum_m \left( f_X (\pi_t^m) \mu_X \log (\mu_X) - \frac{\mu_X}{X} f_0 (\pi_t^m) \log (\mu_X) + \frac{\mu_X}{X} f_1 (\pi_t^m) \log (\mu_X) \right) dt \\
+ \int_{-\gamma_0}^{\gamma_1} \sum_X \sum_m \left( \mu_X \phi (f_X (\pi_t^m)) - \frac{\mu_X}{X} \left( \phi (f_0 (\pi_t^m)) + f_0 (\pi_t^m) \log (1+\epsilon\beta) + \mu_X \right) \left( \phi (f_1 (\pi_t^m)) + f_1 (\pi_t^m) \log (1-\epsilon\beta) \right) \right) dt.
\] (52)

By Lemma 5 the first integral in (52) is 0. Hence it only remains to show the second integral in (52) does not change. But this is again a direct consequence of (50) and (49) with the corresponding facts for \( \Pi'^0_X \) and \( \Pi'^1_X \).
By definition of the distributions $\mu, \mu'$, one has

$$\sum_{j=1}^{k} \int_{-\gamma_0}^{\gamma_1} \text{concav}_\mu(t, j) \, dt - \sum_{j=1}^{s+a} \int_{-\gamma_0}^{\gamma_1} \text{concav}_\mu'(t, j) \, dt$$

$$= \sum_{j=s+a+1}^{k} \int_{-\gamma_0}^{\gamma_1} \text{concav}_\mu(t, j) \, dt.$$ 

Hence it suffices to show

$$\sum_{j=1}^{k} \int_{-\gamma_0}^{\gamma_1} \text{concav}_\mu(t, j) \, dt \geq 0.$$

Let $\mu_{X_j=b}$ denote the distribution of $X$ conditioned on $X_j = b$, one can check that

$$\int_{-\gamma_0}^{\gamma_1} \text{concav}_\mu(t, j) \, dt = \mathbb{E}_b \int_{-\gamma_0}^{\gamma_1} \text{concav}^\text{ext}_{\mu_{X_j=b}}(t) \, dt. \quad (53)$$

In Sect. 4.5 we show the external concavity condition indeed holds, hence (53) implies $\int_{-\gamma_0}^{\gamma_1} \text{concav}_\mu(t, j) \, dt \geq 0$, as desired.

\[\Box\]

### 4.5 Proof of Theorem 3

**Proof of Theorem 3** We use $\land$ to denote the multi-party AND function. Consider the external case first. Set $\Omega$ to be the set of all external trivial distributions together with the distribution in Statement 1, hence Condition (i) and (ii) are satisfied. Picking $w(x) = c k^{-20} x^4$ for some fixed constant $c > 0$, we verify Condition (iii) in Sect. 4.6.3. Hence $IC^\text{ext}_\mu(\pi^\land) \leq IC^\text{ext}_\mu(\land)$, as $IC^\text{ext}_\mu(\pi^\land)$ is also an upper bound, hence we proved $IC^\text{ext}_\mu(\pi^\land) = IC^\text{ext}_\mu(\land)$.

Similarly for the internal case the concavity Condition (iii) is verified in Sect. 4.6.4.

\[\Box\]

### 4.6 Information Cost of Multi-party AND Function

To simplify the notation, since every function has the argument $\pi^m$, we sometimes omit it, such as we write $f$ to mean $f(\pi^m)$. We will use $\mu_0, \mu_j, f_0, f_j$ instead of $\mu_0, \mu_e, f_0, f_e$ when there is no ambiguity, similar notations are used for distributions $\mu_0, \mu^1$ and functions $f^0, f^1$.

### 4.6.1 Taylor Expansions

Recall $\beta_s = \mu_s$ and $\zeta_s = 1 - \beta_s$. By the reductions in Sect. 4.3, we can assume that

$$\mu_1 = \cdots = \mu_{s-1} = \beta, \quad \mu_s = \cdots = \mu_k = e^{\gamma_0} \beta, \quad \mu_0 = 1 - (s-1) \beta - (k-s+1) e^{\gamma_0} \beta. \quad (54)$$

Observe that $0 < \beta < 1/k$ (the distribution when $\beta = 0$ is both external and internal trivial). Furthermore, viewing $\gamma_0$ and $\gamma_1$ as functions of $\epsilon$, plugging $\beta_s = e^{\gamma_0} \beta$ and $\zeta_s = 1 - \beta_s = 1 - e^{\gamma_0} \beta$ into (35) and (37), by implicit differentiation, we have
Recall Taylor’s theorem with the remainder in Lagrange form says that the error term

\[
\begin{aligned}
\gamma_0(0) &= 0, \; \gamma_0'(0) = 1, \; \gamma_0''(0) = 1 - 2\beta, \; \gamma_0'''(0) = 2 - 10\beta + 8\beta^2; \\
\gamma_1(0) &= 0, \; \gamma_1'(0) = 1, \; \gamma_1''(0) = 2\beta - 1, \; \gamma_1'''(0) = 2 + 6\beta^2.
\end{aligned}
\] (55)

These derivatives are used in the Mathematica computation. To simplify the notation we let \( \xi = \xi_0 + \beta \).

Assuming \( \varepsilon < 1/2 \), then \( \gamma_0 + \gamma_1 \leq 2 \ln(1 + 2\varepsilon) \leq 4\varepsilon \). Also note \(|e^{-1} - 1| \leq x \) for \( x \geq 0 \), which together with the fact that \( \Phi_\mu(t) \leq k(\gamma_0 + \gamma_1) \leq 4k\varepsilon \) implies the following:

- \( \mu_0 f_0(\pi_i^m) \) is either 0 or close to \( 1 - k\beta \) with distance bounded by \( 4k\varepsilon \);
- for \( j \neq 0 \), we have \( \mu_j f_j(\pi_i^m) \) is either 0 or close to \( \beta \) with distance bounded by \( 4(k + 1)\varepsilon \);
- \( f(\pi_i^m) \) is either 0 or close to \( 1 - \beta \) with distance bounded by \( 16k^2\varepsilon \);
- for \( j \neq 0 \) and \( j \neq m \), we have \( f(\pi_i^m) - \mu_j f_j(\pi_i^m) \) is either 0 or close to \( 1 - 2\beta \) with distance bounded by \( 16k^2\varepsilon \).

Note that \( f(\pi_i^m) = 0 \) implies \( \mu_x f_x(\pi_i^m) = 0 \) for every \( x \), hence \( \phi(\cdot) = 0 \) for all functions under consideration. On the other hand when \( f(\pi_i^m) \neq 0 \), then all \( \mu_x f_x(\pi_i^m) \) are nonzero except when \( x_m = 1 \). In this case these functions \( \phi(\cdot) \) have the following Taylor expansions at corresponding points as follows,

\[
\begin{aligned}
\phi(\mu_0 f_0(\pi_i^m)) &= -\frac{1 - k\beta}{2} + (\ln(1 - k\beta))\mu_0 f_0(\pi_i^m) + \frac{(\mu_0 f_0(\pi_i^m))^2}{2(1 - k\beta)} + O(\varepsilon^3), \\
\phi(\mu_j f_j(\pi_i^m)) &= -\frac{\beta}{2} + (\ln\beta)\mu_j f_j(\pi_i^m) + \frac{(\mu_j f_j(\pi_i^m))^2}{2\beta} + O(\varepsilon^3), \quad j \neq 0, j \neq m \\
\phi(f(\pi_i^m)) &= -\frac{1 - \beta}{2} + (\ln(1 - \beta))f(\pi_i^m) + \frac{f(\pi_i^m)^2}{2(1 - \beta)} + O(\varepsilon^3), \\
\phi(f(\pi_i^m) - \mu_j f_j(\pi_i^m)) &= -\frac{1 - 2\beta}{2} + (\ln(1 - 2\beta))f(\pi_i^m) + \frac{f(\pi_i^m)^2}{2(1 - 2\beta)} - \\
&\quad (\ln(1 - 2\beta))\mu_j f_j(\pi_i^m) + \frac{(\mu_j f_j(\pi_i^m))^2}{2(1 - 2\beta)} - \frac{\mu_j f_j(\pi_i^m)f(\pi_i^m)}{1 - 2\beta} + O(\varepsilon^3), \quad j \neq 0, j \neq m.
\end{aligned}
\]

Recall Taylor’s theorem with the remainder in Lagrange form says that the error term \( O(\varepsilon^3) \) in the expansion of \( \phi(\mu_0 f_0(\pi_i^m)) \) equals \( \frac{|\phi^{(3)}(\xi)|}{6} |\mu_0 f_0(\pi_i^m) - (1 - k\beta)|^3 \) for some \( \xi \) between \( \mu_0 f_0(\pi_i^m) \) and \( 1 - k\beta \). Since \( |\mu_0 f_0(\pi_i^m) - (1 - k\beta)| \leq 4k\varepsilon \), we have \( |\xi - (1 - k\beta)| \leq 4k\varepsilon \), hence \( \xi \geq (1 - k\beta) - 4k\varepsilon \) if \( \varepsilon < \frac{1 - k\beta}{4k} \). Furthermore, we have \( 0 < \frac{0}{(1-k\beta)-4k\varepsilon} \leq \frac{2}{1-k\beta} \) as long as \( \varepsilon \leq \frac{1 - k\beta}{8k} \). Therefore,

\[
\begin{aligned}
\frac{|\phi^{(3)}(\xi)|}{6} |\mu_0 f_0(\pi_i^m) - (1 - k\beta)|^3 &= \frac{1}{6\xi^2} |\mu_0 f_0(\pi_i^m) - (1 - k\beta)|^3 \\
&\leq \frac{1}{6(1 - k\beta - 4k\varepsilon)^2} (4k\varepsilon)^3 \\
&\leq \frac{4}{6(1 - k\beta)^2} (4k)^3 \varepsilon^3 \leq \frac{k^{11}}{6(1 - k\beta)^2} \varepsilon^3,
\end{aligned}
\]
when $0 < \varepsilon \leq 1 - \frac{k\beta}{2k'} < \frac{1-k\beta}{8k}$. Denote the constant in this upper bound by $R_1$.

Similarly, let $R_2$, $R_3$ and $R_4$ denote the constants that we can get as upper bounds of the absolute values of error terms in the expansions of $\phi(\mu_j f_j(\pi^m_i))$, $\phi(f(\pi^m_i))$ and $\phi(f(\pi^m_i) - \mu_j f_j(\pi^m_i))$, respectively. We have

$$\begin{align*}
R_1 &\leq \frac{k^{11}}{6(1-k\beta)x}, \quad \text{when } 0 < \varepsilon \leq \frac{1-k\beta}{8k}; \\
R_2 &\leq \frac{k^{14}}{6\beta x}, \quad \text{when } 0 < \varepsilon \leq \frac{\beta}{k'}; \\
R_3 &\leq \frac{k^{20}}{6(1-k\beta)^2}, \quad \text{when } 0 < \varepsilon \leq \frac{1-\beta}{k'}; \\
R_4 &\leq \frac{k^{20}}{6(1-2\beta)x}, \quad \text{when } 0 < \varepsilon \leq \frac{1-2\beta}{k'}.
\end{align*}$$

Observe that $\mu^0_x f_x^0$ and $\mu^1_x f_x^1$ are both close to $\mu_x f_x$ with distance bounded by $3\varepsilon$, hence the corresponding functions $\phi(\mu^0_x f_x^0)$ and $\phi(\mu^1_x f_x^1)$ have the same expansions as above, the same holds for functions $\phi(f^0)$, $\phi(f^1)$ and $\phi(f^0 - \mu^0_x f_x^0)$, $\phi(f^1 - \mu^1_x f_x^1)$.

We continue to use the Taylor expansions to expand the concavity conditions (43) and (45).

– Taylor expansion of external concavity condition (43). When $f(\pi^m_i) \neq 0$, we have the following expansion,

$$\begin{align*}
\phi(f(\pi^m_i)) - \sum_x \phi(\mu_x f_x(\pi^m_i)) \\
= \phi(f(\pi^m_i)) - \phi(\mu_0 f_0(\pi^m_i)) - \sum_{j=1, j \neq m}^k \phi(\mu_j f_j(\pi^m_i)) \\
= (\ln(1-\beta))f + \frac{1}{2(1-\beta)}f^2 - (\ln(1-k\beta))\mu_0 f_0 - \frac{1}{2(1-k\beta)}(\mu_0 f_0)^2 \\
- \ln \beta \sum_{j=1, j \neq m}^k \mu_j f_j - \frac{1}{2\beta} \sum_{j=1, j \neq m}^k (\mu_j f_j)^2 + O(\varepsilon^3),
\end{align*}$$

where constant in $O(\varepsilon^3)$ can be bounded by $R_1 + (k-1)R_2 + R_3$ according to (56). Using Lemma 5, we see that the first, third and fifth terms in (57) become 0 in (43). Let

$$F^\text{ext}_m(t) = \frac{1}{2(1-\beta)}f^2 - \frac{1}{2(1-k\beta)}(\mu_0 f_0)^2 - \frac{1}{2\beta} \sum_{j=1, j \neq m}^k (\mu_j f_j)^2. $$

Define $F^\text{ext,0}_m(t)$ and $F^\text{ext,1}_m(t)$ with $f$ replaced by $f^0$ and $f^1$, respectively, etc. Observe that $F^\text{ext}_m(t) = 0$ when $f(\pi^m_i) = 0$, hence in general $F^\text{ext}_m(t)$ is a correct representation of $\phi(f(\pi^m_i)) - \sum_x \phi(\mu_x f_x(\pi^m_i))$. Therefore, what we want to verify in (43) becomes

$\varepsilon$ Springer
\[
\int_{-\gamma_0}^{\gamma_1} \sum_{m=1}^{k} \left( \frac{F_{m}^{\text{ext}}(t) - F_{m}^{\text{ext},0}(t) + F_{m}^{\text{ext},1}(t)}{2} \right) dt + O(\varepsilon^4). \tag{58}
\]

As \(\gamma_0 + \gamma_1 \leq 4\varepsilon\), by (56), the constant in \(O(\varepsilon^4)\) in (58) can be bounded by

\[
8k(R_1 + (k - 1)R_2 + R_3) \leq 4k^2 \left( \frac{1}{(1 - k\beta)^2} + \frac{1}{\beta^2} \right),
\]

when \(\varepsilon \leq \frac{1}{k} \min\{\beta, 1 - k\beta\} \).

- Taylor expansion of internal concavity condition (45).

A direct calculation gives,

\[
\sum_{j=1}^{k} \sum_{b=0,1} \phi(f_{xj}=b(\pi_i^m)) - k \sum_{x} \phi(\mu x f x(\pi_i^m)) \\
= \phi(f(\pi_i^m)) - k\phi(\mu_0 f_0(\pi_i^m)) \\
+ \sum_{j=1, j \neq m}^{k} \left( \phi(f(\pi_i^m) - \mu_j f_j(\pi_i^m)) - (k - 1)\phi(\mu_j f_j(\pi_i^m)) \right). \tag{59}
\]

As did for the external case, when \(f(\pi_i^m) \neq 0\) the above formula expands as follows,

\[
(59) = (\ln(1 - \beta) + (k - 2)\ln(1 - 2\beta) - (k - 1)\ln(1 - k\beta)) f \\
+ \left( \frac{1}{2(1 - \beta)} + \frac{k - 3}{2(1 - 2\beta)} \right) f^2 \\
+ (\ln(1 - 2\beta) + (k - 1)\ln(1 - 2\beta) - (k - 1)\ln(1 - k\beta)) \mu_0 f_0 \\
- \frac{k}{2(1 - k\beta)}(\mu_0 f_0)^2 + \left( \frac{1}{2(1 - 2\beta)} - \frac{k - 1}{2\beta} \right) \sum_{j=1, j \neq m}^{k} (\mu_j f_j)^2 \\
+ \frac{1}{1 - 2\beta} \mu_0 f_0 f + O(\varepsilon^3). \tag{60}
\]

Lemma 5 implies the first and third terms in (60) become 0 in (45). Let

\[
F_m(t) = \left( \frac{1}{2(1 - \beta)} + \frac{k - 3}{2(1 - 2\beta)} \right) f^2 - \frac{k}{2(1 - k\beta)}(\mu_0 f_0)^2 \\
+ \left( \frac{1}{2(1 - 2\beta)} - \frac{k - 1}{2\beta} \right) \sum_{j=1, j \neq m}^{k} (\mu_j f_j)^2 + \frac{1}{1 - 2\beta} \mu_0 f_0 f.
\]
Define $F_m^0(t)$ and $F_m^1(t)$ similarly. Then $F_m(t)$ is a correct representation for (59). Therefore what we want to verify in (45) becomes

$$\int^{-\gamma_1}_{-\gamma_0} \sum_{m=1}^k \left( F_m(t) - \frac{F_m^0(t) + F_m^1(t)}{2} \right) dt + O(\epsilon^4).$$

(61)

4.6.2 Density Functions in Explicit Form

We continue to calculate functions explicitly that will be used for computing.

– In the protocol $\pi^\wedge_{\mu}$.

Consider the interval $t \in [-\gamma_0, 0)$. Let $A = (s - 1)(t + \gamma_0) + (s - 1)t + (s - 1)\gamma_0$. The total active duration $\Phi_0(t) = \Phi_j(t) = A$ for $s \leq j \leq k$, and $\Phi_j(t) = A - (t + \gamma_0)$ for $1 \leq j \leq s - 1$. Hence for $1 \leq m \leq s - 1$, we have,

$$\mu_j f_j(\pi_t^m) = \begin{cases} 
\mu_0 e^{-A}, & j = 0, \\
\mu_j e^{t+\gamma_0} e^{-A} = e^{\gamma_0} \beta e^{t} e^{-A}, & 1 \leq j \leq s - 1 \text{ and } j \neq m, \\
\mu_j e^{-A} = e^{\gamma_0} \beta e^{-A}, & s \leq j \leq k, \\
0, & j = m.
\end{cases}$$

For $m \geq s$, we have $\mu_x f_x(\pi_t^m) = 0$ for all $x$. Therefore when $t \in [-\gamma_0, 0)$,

$$f(\pi_t^m) = \begin{cases} 
0, & m \geq s, \\
(1 - (s - 1)\beta + (s - 2) e^{\gamma_0} \beta e^{t}) e^{-A}, & 1 \leq m \leq s - 1.
\end{cases}$$

Similarly for the interval $t \in [0, \gamma_1)$, let $D = (s - 1)(t + \gamma_0) + (k - s + 1)t = kt + (s - 1)\gamma_0$, the total active duration is $\Phi_0(t) = D$, $\Phi_j(t) = D - (t + \gamma_0)$ for $1 \leq j \leq s - 1$, and $\Phi_j(t) = D - t$ for $s \leq j \leq k$. Hence for all $1 \leq m \leq k$ we have,

$$\mu_j f_j(\pi_t^m) = \begin{cases} 
\mu_0 e^{-D}, & j = 0, \\
\mu_j e^{t+\gamma_0} e^{-D} = e^{\gamma_0} \beta e^{t} e^{-D}, & 1 \leq j \leq s - 1 \text{ and } j \neq m, \\
\mu_j e^{-D} = e^{\gamma_0} \beta e^{-D}, & s \leq j \leq k \text{ and } j \neq m, \\
0, & j = m.
\end{cases}$$

Therefore when $t \in [0, \gamma_1)$,

$$f(\pi_t^m) = (1 - (s - 1)\beta - (k - s + 1)e^{\gamma_0} \beta + (k - 1)e^{\gamma_0} \beta e^{t}) e^{-D}.$$
– In the protocol $\pi_{\mu_0}$.

Using results from Sect. 4.2, we have,

$$
\mu_x^0 f_x^0(\pi_t^m) = \begin{cases} 
(1 - \varepsilon \xi) e^{-t} \mu_x f_x(\pi_t^m), & t \in [-\gamma_0, 0), x_s = 0, m \neq s, \\
(1 - \varepsilon \xi) \mu_x f_x(\pi_t^m), & t \in [-\gamma_0, 0), x_s = 1, m \neq s, \\
(1 - \varepsilon \xi) \mu_x f_x(\pi_t^m), & t \in [0, \gamma_1).
\end{cases}
$$

For the special case $m = s$ and $t \in [-\gamma_0, 0)$, we have,

$$
\mu_j^0 f_j^0(\pi_t^s) = \begin{cases} 
(1 - \varepsilon \xi) \mu_0 e^{-t} e^{-A}, & t \in [-\gamma_0, 0), j = 0, \\
(1 - \varepsilon \xi) e^{\gamma_0} \beta e^{-A}, & t \in [-\gamma_0, 0), 1 \leq j \leq s - 1, \\
0, & t \in [-\gamma_0, 0), j = s \\
(1 - \varepsilon \xi) e^{\gamma_0} \beta e^{-A}, & t \in [-\gamma_0, 0), s + 1 \leq j \leq k.
\end{cases}
$$

Therefore when $t \in [-\gamma_0, 0)$,

$$
f^0(\pi_t^m) = \begin{cases} 
(1 - \varepsilon \xi) ((1 - e^{\gamma_0} \beta - (s - 1) \beta) e^{-t} + (s - 1) e^{\gamma_0} \beta) e^{-A}, & 1 \leq m \leq s, \\
0, & s + 1 \leq m \leq k.
\end{cases}
$$

When $t \in [0, \gamma_1)$, it is simply,

$$
f^0(\pi_t^m) = (1 - \varepsilon \xi) f(\pi_t^m).
$$

– In the protocol $\pi_{\mu_1}$.

Using results from Sect. 4.2, when $m = s$, then $\mu_x^1 f_x^1(\pi_t^m) = 0$ for all $x$ for $t \in [-\gamma_0, \gamma_1]$. Therefore $f^1(\pi_t^s) = 0$ for all $t \in [-\gamma_0, \gamma_1]$.

When $m \neq s$, we have,

$$
\mu_x^1 f_x^1(\pi_t^m) = \begin{cases} 
\mu_x^1 f_x(\pi_t^m), & t \in [-\gamma_0, 0), \\
(1 - \varepsilon e^{\gamma_0} \beta) e^t \mu_x f_x(\pi_t^m), & t \in [0, \gamma_1), x_s = 0, \\
(1 + \varepsilon \xi) \mu_x f_x(\pi_t^m), & t \in [0, \gamma_1), x_s = 1.
\end{cases}
$$

Hence when $t \in [-\gamma_0, 0)$, we have $f^1(\pi_t^m) = (1 - \varepsilon e^{\gamma_0} \beta) f(\pi_t^m) + \varepsilon \mu_s f_s(\pi_t^m)$, and when $t \in [0, \gamma_1)$ we have $f^1(\pi_t^m) = (1 - \varepsilon e^{\gamma_0} \beta) e^t f(\pi_t^m) + (1 + \varepsilon \xi - (1 - \varepsilon e^{\gamma_0} \beta) e^t) \mu_s f_s(\pi_t^m)$. Plug in $f$ we get, when $t \in [-\gamma_0, 0)$,

$$
f^1(\pi_t^m) = \begin{cases} 
0, & m \geq s, \\
(1 + e^{\gamma_0} \beta (1 - \varepsilon e^{\gamma_0} \beta)((s - 2)e^t - (s - 1)e^{-\gamma_0})) e^{-A}, & 1 \leq m \leq s - 1.
\end{cases}
$$
When $t \in [0, \gamma_1)$,

$$f^1(\pi^m_t) = \begin{cases} 0, & m = s, \\ (1 + e^{\gamma_0} \beta (1 - \varepsilon e^{\gamma_0} \beta)((k - 2)e' - (s - 1)e^{-\gamma_0} - k + s))e' e^{-D}, & m \neq s. \end{cases}$$

### 4.6.3 External Information Cost

Using Wolfram Mathematica (see the supplemented Mathematica code, or download from here http://www.cs.mcgill.ca/~yli252/files/MultiAnd.nb) with results from Sects. 4.6.1 and 4.6.2, we obtain

\begin{equation}
(58) = \frac{(k + 5s - 6)(1 - 2\beta)\beta}{12(1 - \beta)\ln 2} \varepsilon^3 + O(\varepsilon^4).
\end{equation}

Therefore, using the bound of the error term given in Sect. 4.6.1, one finds (62) as long as

$$\varepsilon < \min \left\{ \frac{(k + 5s - 6)(1 - 2\beta)\beta}{12(1 - \beta)\ln 2} \sqrt{4k^{21} \left( \frac{1}{(1 - k\beta)^2} + \frac{1}{\beta^2} \right)}, \frac{1}{k^4} \min\{\beta, 1 - k\beta\} \right\}.$$

Note that $\frac{2}{x + y} \geq \min\{x, y\}$ for all $x, y > 0$. Simplifying the above formula, one obtains (62) as long as

$$\varepsilon < ck^{-20} \min\{\beta, 1 - k\beta\}^3,$$

for some constant $c > 0$. So we have verified the concavity condition (43) is satisfied for all $\varepsilon$-weak signals such that $\varepsilon$ is no greater than $ck^{-20} \min\{\beta, 1 - k\beta\}^3$.

Let $\mu^E$ denote the distribution in Statement 1, we have $|\mu - \mu^E| \leq 1 - k\beta$. Let $\mu'$ be defined as $\mu'_{s - 1} = 0, \mu'_s = e^{\gamma_0} \beta + \beta$, and $\mu'_j = \mu_j$ for all other $j$, then $|\mu - \mu'| = \beta$. Observe that $\mu'$ is external trivial, hence $\mu^E, \mu' \in \Omega$ (the $\Omega$ we chose in Sect. 4.5). Therefore we have $\delta(\mu) \leq \min\{\beta, 1 - k\beta\}$. Thus as we choose $w(x) = ck^{-20} x^4$, the concavity condition (43) is satisfied for all $w(\delta(\mu))$-weak signals because

$$w(\delta(\mu)) \leq ck^{-20} \min\{\beta, 1 - k\beta\}^4 < ck^{-20} \min\{\beta, 1 - k\beta\}^3.$$

By Theorem 2, we have proved the protocol $\pi^\wedge$ in Fig. 1 is optimal for external information cost.

### 4.6.4 Internal Information Cost

Similarly, using Wolfram Mathematica, we obtain

\begin{equation}
(61) = \begin{cases} \frac{(k + 5s - 6)(1 - 2\beta)\beta}{12(1 - \beta)\ln 2} \varepsilon^3 + O(\varepsilon^4), & k = 2, \\ \frac{(k + 5s - 6)((3k - 2)\beta^2 - 4(k - 1)\beta + k - 1)\beta}{12(1 - \beta)(1 - 2\beta)^2 \ln 2} \varepsilon^3 + O(\varepsilon^4), & k \geq 3. \end{cases}
\end{equation}
Similarly one can pick an appropriate function $w$ to verify that the concavity condition (45) is satisfied for all $w(\delta(\mu))$-weak signals when $\varepsilon$ is sufficiently small. Hence by Theorem 2, the protocol is optimal for internal information cost.

5 Discussion

A natural extension of our result is to remove the Assumption 1. This seems difficult without new ideas, as even under Assumption 1 we need to study the protocol in great detail and rely on heavy computation.

Secondly, is it possible to compute explicitly $IC^\mu(\pi^\wedge)$ and $IC^{ext}_\mu(\pi^\wedge)$ as functions that depend on the parametrization of the distribution $\mu$? We have not attempted to do so. Recall we discussed our Assumption 1 on the distribution has a background in communication complexity of the $k$-party promised set disjointness problem $PDISJ_{n,k}$, and the order of which were determined to be $\Omega(n/k)$ in [5,9]. The method in [5,9] is to compute a lower bound of a variant of information complexity of $\AND_k$ introduced in [1]. Since we showed the protocol $\pi^\wedge$ is optimal for $\AND_k$, it would be interesting to see if the optimal information cost $IC^\mu(\pi^\wedge)$ would give us any knowledge about the constant in $\Omega(n/k)$. We defer this to a future work.

We have focused exclusively on the multi-party number-in-hand model. Another natural multi-party communication model is the number-on-forehead (NoF) model. Attempt to define information complexity in NoF model has been considered in some restricted sense in [8], where the communication is restricted to be one way, and furthermore player $i$ can only, roughly, uses information about inputs $x_1, x_2, \ldots, x_{i-1}$. To the knowledge of the authors, no substantial study has been made on multi-party information complexity in the NoF model, even for $\AND_k$ function. In general, extend and determine the exact information complexity in the NoF model is an interesting open problem.

Finally, we mention that inspired by the local concavity characterization (i.e., Theorem 1) of the optimal information cost given in [2], in [13] a set of partial differential equations (PDE) is given to characterize the optimal information cost for $\AND_2$ and for general functions in the two-party setting. The solution of the PDE system in [13] gives the optimal information cost of the two-bit $\AND_2$ function. It would be interesting to see how the PDE characterization would work in the multi-party setting, and in particular, whether it can simplify our argument, or even extract an explicit formula for information cost for $\AND_k$.

Funding Yuval Filmus is supported by Israel Science Foundation (Grant No. 2022103), Hamed Hatami is supported by an NSERC grant.

Appendix

The computational results of the two-party $\AND_2$ in [2, Section 7.7] show the concavity term (the one that we want to verify its non-negativity) can be of order $\varepsilon^2$. We see in Sect. 4.6 that our computation is of order $\varepsilon^3$. This is because we choose to focus our
computation, as allowed by a series of reductions, on the time interval \([-\gamma_0, \gamma_1]\) only. Claim 1 below shows an order \(\varepsilon^2\) term can appear too if the whole range is considered.

**Claim 1** Suppose \(s \geq 2\) and \(L = |t_{s-1}| > 0\). If \(\gamma_0 \leq L/2\), then

\[
\int_{t_{s-1}}^{-\gamma_0} \text{concav}^\text{ext}_\mu(t) \, dt \geq \frac{(1 - e^{-(s-1)L/2})\mu_0 \mu_e \varepsilon^2}{2(s-1)} \geq 0, \tag{64}
\]

and

\[
\sum_{j=1}^{k} \int_{t_{s-1}}^{-\gamma_0} \text{concav}_\mu(t, j) \, dt \geq \frac{(k-1)(1 - e^{-(s-1)L/2})\mu_0 \mu_e \varepsilon^2}{2(s-1)} \geq 0. \tag{65}
\]

**Proof** Consider external case first. Let \(\mu'\) be defined as

\[
\mu'_x = \begin{cases} 
\beta \mu_x, & x_s = 0, \\
-\zeta \mu_x, & x_s = 1.
\end{cases}
\]

Then \(\mu_0 = \mu + \varepsilon \mu'\) and \(\mu^1 = \mu - \varepsilon \mu'\), hence \(f^0(\pi_t^m) = f(\pi_t^m) + \varepsilon \sum \mu'_x f_x(\pi_t^m)\) and \(f^1(\pi_t^m) = f(\pi_t^m) - \varepsilon \sum \mu'_x f_x(\pi_t^m)\). Note that \(f(\pi_t^m) = 0\) (in our case this happens when \(m \geq s\)) implies \(\text{concav}^\text{ext}_\mu(t) = 0\). On the other hand, when \(f(\pi_t^m) \neq 0\) (i.e., \(1 \leq m \leq s-1\)), using Taylor expansion at the point \(f(\pi_t^m)\) for functions \(\phi(f^0(\pi_t^m))\) and \(\phi(f^1(\pi_t^m))\), and expansion at \(\mu_x f_x(\pi_t^m)\) for functions \(\phi(\mu_0 f^0(\pi_t^m))\) and \(\phi(\mu^1 f^1(\pi_t^m))\), we obtain (note here we won’t have \(\varepsilon^3\))

\[
\text{concav}^\text{ext}_\mu(t) \geq \sum_{m=1}^{s-1} \left( \frac{(\mu'_x f_x(\pi_t^m))^2}{2 \mu_x f_x(\pi_t^m)} - \frac{(\sum \mu'_x f_x(\pi_t^m))^2}{2 f(\pi_t^m)} \right) \varepsilon^2 \\
= \frac{1}{2} \sum_{m=1}^{s-1} \left( \mu_e f_e(\pi_t^m) \left( 1 - \frac{\mu_x f_x(\pi_t^m)}{f(\pi_t^m)} \right) \right) \varepsilon^2 \tag{66} \\
\geq \frac{1}{2} \sum_{m=1}^{s-1} \left( \mu_e f_e(\pi_t^m) \frac{\mu_0 f_0(\pi_t^m)}{f(\pi_t^m)} \right) \varepsilon^2 \geq 0.
\]

By Statement 4 one can assume \(\mu_e = \cdots = \mu_e\) and \(\mu_1 = \cdots = \mu_e = -\mu_e = \mu_0 = 1 - (k - s + 1)\mu_e - (s - 1)\mu_0 \mu_e e^{-L}\). Then \(t_{s-1} = 0\) and \(t_s = L\). As \(\gamma_0 \leq L/2\) implies \(t_s - \gamma_0 = L - \gamma_0 \geq L/2\), and (66) says the integrand is non-negative, hence a lower bound is given by the integration of (66) in the range \([0, L/2]\). For \(t \in [0, L/2]\) and \(1 \leq m \leq s-1\), we have \(f_0(\pi_t^m) = f_e(\pi_t^m) = \cdots = f_e(\pi_t^m) = e^{-(s-1)t}\), and \(f_1(\pi_t^m) = \cdots = f_e(\pi_t^m) = e^{-(s-2)t}\), thus \(f(\pi_t^m) = (1 - (s-1)\mu_e e^{-L} + (s-2)\mu_0 \mu_e e^{-L} e^{(-s-1)t}) = (s-1)e^{-(s-1)t}\). Hence,

\[
(66) \geq \frac{s-1}{2} \mu_e f_e(\pi_t^m) \frac{\mu_0 f_0(\pi_t^m)}{(s-1)e^{-(s-1)t}} \varepsilon^2 = \frac{\mu_0 \mu_e e^{-(s-1)t}}{2} \varepsilon^2.
\]

\(\Box\) Springer
Integrating in the range $[0, L/2]$ with respect to $t$ gives the desired bound \((64)\).

Similarly, in the internal case one has

\[
\sum_{j=1}^{k} \text{concav}_{\mu}(t, j) \geq \frac{1}{2} \sum_{m=1}^{s-1} \sum_{j=1, j \neq s}^{k} \left( \mu_{e, e_x} \left( \pi_{t}^{m} \right) \frac{\mu_{0, 0} \left( \pi_{t}^{m} \right) f(e_j) \left( \pi_{t}^{m} \right)}{f(e_x) \left( \pi_{t}^{m} \right)} \right) \varepsilon^2
\]

\[
\geq \frac{k - 1}{2} \sum_{m=1}^{s-1} \left( \mu_{e, e_x} \left( \pi_{t}^{m} \right) \frac{\mu_{0, 0} \left( \pi_{t}^{m} \right) f(e_j) \left( \pi_{t}^{m} \right)}{f(e_x) \left( \pi_{t}^{m} \right)} \right) \varepsilon^2.
\]

Hence we get the bound \((65)\) after integration. \(\Box\)

References

1. Bar-Yossef, Z., Jayram, T.S., Kumar, R., Sivakumar, D.: An information statistics approach to data stream and communication complexity. J. Comput. Syst. Sci. 68(4), 702–732 (2004)
2. Braverman, M., Garg, A., Pankratov, D., Weinstein, O.: From information to exact communication (extended abstract). In: STOC’13, pp. 151–160. ACM (2013)
3. Braverman, M., Garg, A., Pankratov, D., Weinstein, O.: Information lower bounds via self-reducibility. Theory Comput. Syst. 59(2), 377–396 (2016)
4. Braverman, M., Rao, A.: Information equals amortized communication. IEEE Trans. Inf. Theory 60(10), 6058–6069 (2014)
5. Chakrabarti, A., Khot, S., Sun, X.: Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In: Proceedings of 18th IEEE Annual Conference on Computational Complexity, 2003., pp. 107–117. IEEE (2003)
6. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, Hoboken (2012)
7. Dagan, Y., Filmus, Y., Hatami, H., Li, Y.: Trading information complexity for error. Theor. Comput. 14(6), 1–73 (2018)
8. Gronemeier, A.: Nof-multiparty information complexity bounds for pointer jumping. In: International Symposium on Mathematical Foundations of Computer Science, pp. 459–470. Springer (2006)
9. Gronemeier, A.: Asymptotically optimal lower bounds on the nih-multi-party information. arXiv preprint arXiv:0902.1609 (2009)
10. Kushilevitz, E., Nisan, N.: Communication Complexity. Cambridge University Press, Cambridge (1997)
11. Ma, N., Ishwar, P.: Some results on distributed source coding for interactive function computation. IEEE Trans. Inf. Theory 57(9), 6180–6195 (2011)
12. Ma, N., Ishwar, P.: The infinite-message limit of two-terminal interactive source coding. IEEE Trans. Inf. Theory 59(7), 4071–4094 (2013)
13. Pankratov, D.: Communication complexity and information complexity. Ph.D. thesis, The University of Chicago (2015)
14. Razborov, A.A.: On the distributional complexity of disjointness. Theor. Comput. Sci. 106(2), 385–390 (1992)
15. Yao, A.C.C.: Some complexity questions related to distributive computing(preliminary report). STOC’79, pp. 209–213. ACM (1979)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.