An Adaptive Step Toward the Multiphase Conjecture and Lower Bounds on Nonlinear Networks

Young Kun Ko* Omri Weinstein†

August 27, 2020

Abstract

In 2010, Pătraşcu proposed a dynamic set-disjointness problem, known as the *Multiphase problem*, as a candidate for proving *polynomial* lower bounds on the operational time of dynamic data structures. He conjectured that any data structure for the Multiphase problem must make n^{ε} cell-probes in either update or query phases, and showed that this would imply similar *unconditional* lower bounds on many important dynamic data structure problems. There has been almost no progress on this conjecture in the past decade since its introduction. We show an $\tilde{\Omega}(\sqrt{n})$ cell-probe lower bound on the Multiphase problem for data structures with general (adaptive) updates, and queries with unbounded but "layered" adaptivity. This result captures all known set-intersection data structures and significantly strengthens previous Multiphase lower bounds, which only captured non-adaptive data structures.

Our main technical result is a communication lower bound on a 4-party variant of Pătraşcu's Number-On-Forehead Multiphase game, using information complexity techniques. We then use this result to make progress on understanding the power of *nonlinear* gates in networks computing *linear* operators, a long-standing open problem in circuit complexity and network design: We show that any depth-d circuit that computes a random $m \times n$ linear operator $x \mapsto Ax$ using gates of degree k (width-k DNFs) must have $\Omega(m \cdot n^{1/2(d+k)})$ wires. Finally, we show that a lower bound on Pătraşcu's original NOF game would imply a polynomial wire lower bound $(n^{1+\Omega(1/d)})$ for circuits with *arbitrary* gates computing a random linear operator. This suggests that the NOF conjecture is much stronger than its data structure counterpart.

*NYU Courant, ykk254@cs.nyu.edu

†Columbia University. Email: omri@cs.columbia.edu. Research supported by NSF CAREER award CCF-1844887.

1 Introduction

Proving unconditional lower bounds on the operational time of dynamic data structures has been a challenge since the introduction of the *cell-probe* model [Yao79]. In this model, the data structure needs to support a sequence of n online updates and queries, where the operational cost is measured only by the number of *memory accesses* ("probes") the data structure makes to its memory, whereas all computations on probed cells are completely free of charge. A natural question to study is the tradeoff between the *update* time t_u and *query* time t_q of the data structure for supporting the underlying dynamic problem. Cell-probe lower bounds provide a compelling answer to this question, as they are purely information-theoretic and independent of implementation constraints, hence apply to any imaginable data structure. Unfortunately, the abstraction of the cell-probe model also comes at a price, and the highest explicit lower bound known to date, on *any* dynamic problem, is merely polylogarithmic ($\max\{t_u, t_q\} \geq \tilde{\Omega}(\lg^2 n)$, see e.g., [Lar12, LWY18] and references therein). In 2010, Pătraşcu [Pat10] proposed the following dynamic set-disjointness problem, known as the *Multiphase problem*, as a candidate for proving *polynomial* lower bounds on the operational time of dynamic data structures. The problem proceeds in 3 "phases":

- **PI**: Preprocess a collection of $m = \text{poly}(n)$ sets $\vec{S} = S_1, \ldots, S_m \subseteq [n]$ in $O(mnt_u)$ time.
- **PII**: A set $T \subseteq [n]$ is revealed, and the data structure can update its memory in $O(nt_u)$ time.
- **PIII**: Given $i \in_R [m]$, the data structure must determine if $S_i \cap T =^? \emptyset$ in $O(t_q)$-time.

Pătraşcu conjectured that any data structure solving the Multiphase problem must make $\max\{t_u, t_q\} \geq n^{\varepsilon}$ cell-probes, and showed that such a polynomial lower bound would imply similar polynomial lower bounds on many important dynamic data structure problems, including dynamic reachability in directed graphs and online matrix multiplication (for the broad implications and further context of the Multiphase conjecture within fine-grained complexity, see [Pat10, HKNS15]). In the same paper, Pătraşcu [Pat10] proposed an approach to prove an unconditional cell-probe lower bound on the Multiphase problem, by reduction to the following 3-party *Number-On-Forehead* (NOF) communication game $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$, henceforth called the *Multiphase Game* :

- Alice receives a collection of sets $\vec{S} = S_1, \ldots, S_m \subseteq [n]$ and a random index $i \in_R [m]$.
- Bob receives a set $T \subseteq [n]$ and the index i.
- Charlie receives $\vec{S}$ and T (but not i).

Thus, one can think of i as being on Charlie's forehead, T being on Alice's forehead, and $\vec{S}$ being on Bob's forehead. The goal of the players is to determine if $S_i \cap T =^? \emptyset$, where communication proceeds in the following way: First, Charlie sends a message ("advice") $U = U(\vec{S}, T)$ *privately to Bob*. Thereafter, Alice and Bob continue to compute $\mathsf{DISJ}_n(S_i, T)$ in the standard 2-party model. Denoting by Π the second stage protocol, Pătraşcu made the following conjecture:

Conjecture 1.1 (Multiphase Game Conjecture, Conjecture 9 of [Pat10])**.** *Any 3-party NOF protocol for the Multiphase game with $|U| = o(m)$ bits of advice must have $|\Pi| > n^{\varepsilon}$ communication.*

The (naïve) intuition for this conjecture is that, since Charlie's advice is independent of i, it can only provide very little useful information about the interesting subproblem $\mathsf{DISJ}_n(S_i, T)$ (assuming $|U| = o(m)$), and hence Alice and Bob might as well solve the problem in the standard 2-party model. This intuition turns out to be misleading, and in fact when S_i's and T are *correlated*, it is simply false – Chattopadhyay, Edmonds, Ellen and Pitassi [CEEP12] showed a deterministic (2-round) NOF protocol for the Multiphase game with a total of $O(\sqrt{n} \log m) = \tilde{O}(\sqrt{n})$

communication, whereas the 2-party communication complexity of set-disjointness is $\Omega(n)$, even randomized. Surprisingly, they also show that Conjecture 1.1 is equivalent, up to $O(\log m)$ communication factor, for deterministic and randomized protocols. Nevertheless, Conjecture 1.1 still stands for *product distributions* [CEEP12] (incidentally, the 2-party communication complexity of set-disjointness under product distributions is $\tilde{\Theta}(\sqrt{n})$ [BFS86, HW07]).

The technical centerpiece of this paper is an $\Omega(\sqrt{n})$ lower bound on the NOF Multiphase game, for (unbounded-round) protocols in which *only the first* message of Alice in Π depends on her entire input $\vec{S} = S_1, \ldots, S_m$ (and i), while in subsequent rounds $j > 1$, Alice's messages can depend only *on S_i, i and the transcript* $\Pi^{<j}$ (No restriction is placed on Bob and Charlie). Note that Alice's messages in subsequent rounds still heavily depend on *all sets* $\vec{S}$, but only through the *transcript* of $\Pi = \Pi(\vec{S})$ so far (this feature better captures data structures, which can only adapt based on cells probed so far). There is a natural way to view such restricted 3-party NOF protocol in terms of an additional player (Megan, holding $\vec{S}, i$), who, in addition to Charlie's private advice to Bob, can broadcast a single message to *both* Alice and Bob (holding S_i, T respectively), who then continue to communicate in the standard 2-party model (see Figure 5). We define this *4-party NOF* model formally in Section 3.2. Our main technical result is the following tight lower bound on such NOF protocols:

Theorem 1.2 (Informal)**.** *Any Restricted NOF protocol $\Gamma = (U, \Pi)$ for the Multiphase game with $|U| = o(m)$ bits of advice must have $|\Pi| > \Omega(\sqrt{n})$ communication.*

This lower bound is tight up to logarithmic factors, as the model generalizes the upper bound of [CEEP12] (See Section A.2). This suggests that the NOF model we study is both subtle and powerful. Indeed, while the aforementioned restriction may seem somewhat technical, we show that removing it by allowing as little as *two rounds* of Alice's messages to depend on her entire input $\vec{S}$, would lead to a major breakthrough in circuit lower bounds – see Theorem 1.6 below. Interestingly, the Multiphase conjecture itself does not have this implication, since dynamic data structures only have limited and local access to $\vec{S}$, through the probes ("transcript") of the query algorithm, and hence induce weaker NOF protocols.

1.1 Implications to dynamic data structure lower bounds

In contrast to the *static* cell-probe model, *adaptivity* plays a dramatic role when it comes to dynamic data structures. In [BL15], Brody and Larsen consider a variation of the Multiphase problem with $(\lg n)$-bit updates (i.e., the 2nd phase set is of cardinality $|T| = 1$), and show that any dynamic data structure whose query algorithm is *non-adaptive*[1] must make $\max\{t_u, t_q\} \geq \Omega(n/w)$ cell-probes when the word-size is w bits. Nevertheless, such small-update problems have a trivial ($t_q = O(1)$) adaptive upper bound and therefore are less compelling from the prospect of making progress on general lower bounds. (By contrast, proving polynomial cell-probe lower bounds for dynamic problems with *large* $\text{poly}(n)$-bit updates, like Multiphase, even against non-adaptive query algorithms, already seems beyond the reach of current techniques[2]).

We prove a polynomial lower bound on the Multiphase problem, against a much stronger class of data structures, which we call *semi-adaptive*, defined as follows:

[1] An algorithm is non-adaptive, if the addresses of probed memory cells are predetermined by the query itself, and do not depend on content of cells probed along the way.

[2] While intuitively larger updates $|T| = \text{poly}(n)$ only make the problem harder and should therefore only be easier to prove lower bounds against, the *total* update time of the data structure in Phase II is also proportional to $|T|$ and hence the data structure has potentially much more power as it can "amortize" its operations. This is why encoding-style arguments fail for large updates (enumerating all $\binom{n}{|T|} = \exp(n)$ possible updates is prohibitive).

Definition 1.3 (Semi-Adaptive Data Structures)**.** *Let D be a dynamic data structure for the Multiphase problem with general (adaptive) updates. Let $\mathcal{M}(\vec{S})$ denote the memory state of D after the preprocessing Phase I, and let $\Delta(\mathcal{M}, T)$ denote the set of ($\leq |T| \cdot t_u$) cells updated in Phase II. D is* semi-adaptive *if its query algorithm in Phase III operates in the following stages ("layers"):*

- *Given the query $i \in [m]$, D may first read S_i* free of charge.
- *D (adaptively) reads at most t_1 cells from $\mathcal{M}$.*
- *D (adaptively) reads at most t_2 cells from $\Delta(\mathcal{M}, T)$, and returns the answer $S_i \cap T =^? \emptyset$.*

The update time of D is t_u, and the query time is $t_q := t_1 + t_2$.

Thus, the query algorithm has unbounded but "layered" adaptivity in Phase III, in the sense that the model allows only a single alternation between the two layers of memory cells, $\mathcal{M}$ and $\Delta(\mathcal{M}, T)$. While this restriction may seem somewhat technical, all known set-intersection data structures are special cases of the semi-adaptive model (see [DLOM00, BK02, BY04, BPP07, CP10, KPP15] and references therein). But more importantly, in this model there is actually a nontrivial *upper bound* for the Multiphase problem – A semi-adaptive data structure solving it in $t_u = t_q = O(\sqrt{n} \lg m)$ time (based on [CEEP12]'s communication protocol, see Section A.1), indicating that the model is powerful. We remark that even though the set of modified cells $\Delta(\mathcal{M}, T)$ may be *unknown* to D at query time, it is easy to implement a semi-adaptive data structure by maintaining $\Delta(\mathcal{M}, T)$ in a dynamic dictionary [MPP05] that checks membership of cells in Δ, and returns $\perp$ if the cell is $\notin \Delta(\mathcal{M}, T)$.

Our main result is an essentially tight $\tilde{\Omega}(\sqrt{n})$ cell-probe lower bound on the Multiphase problem against semi-adaptive data structures. This follows from Theorem 1.2 by a simple variation of the reduction in [Pat10] :

Theorem 1.4 (Multiphase Lower Bound for Semi-Adaptive Data Structures)**.** *Let $m > \omega(n)$. Any semi-adaptive data structure that solves the Multiphase problem, must have either $n \cdot t_u \geq \Omega(m/w)$ or $t_q \geq \Omega(\sqrt{n}/w)$, in the dynamic cell-probe model with word size w.*

1.2 Implications to Network and Circuit Lower Bounds

Can *non-linear* gates help compute linear operators? A long-standing open problem in networks and circuit complexity is whether *non-linear* gates can help computing linear operators ([Lup56, JS10, Dru12]). Specifically, the challenge is to prove a polynomial lower bound on the number of *wires* of constant-depth circuits with *arbitrary gates* for computing any $m \times n$ linear operator $x \mapsto Ax$ [Juk12]. A random matrix $A \in \{0,1\}^{m \times n}$ easily gives a polynomial $\Omega(mn/\lg m)$ lower bound against *linear circuits* [Lup56, JS10] (this restricted the interest to finding *explicit* hard matrices A, see [Val77]). In contrast, for general circuits, the highest lower bound on the number of wires, even for computing a *random* matrix A, is near-linear [Dru12, GHK$^+$13]. Indeed, the current state of affairs cannot even rule out the possibility that nonlinear networks with $O(m \cdot \operatorname{poly} \log n)$ wires suffice to compute **all** $m \times n$ linear operators [Dru12]. As such, a perplexing open question is whether one can prove the existence of a matrix A which requires a polynomial $m^{1+\varepsilon}$ number of wires for some constant $\varepsilon > 0$ when $m = \operatorname{poly}(n)$.

Motivated by this question, we study an intermediate model of non-linear circuits, where each gate computes a *degree-k polynomial* on its input wires (more precisely, a width-k DNF) and may have unbounded fan-in and arbitrary depth. Note that even in this intermediate model, proving existential lower bounds against linear operators falls short of a counting argument: There are only

2^{n^2} possible $n \times n$ linear operators, but at least $\sim 2^{n^k}$ possible gates/functions of degree k (width-k DNFs) on n inputs[3], hence the counting argument breaks even for $k = 3$ (!).

We use Theorem 1.2 to prove that *most* linear operators require a polynomial number of wires, unless the network computing them is using highly nonlinear gates ($k = \omega(1)$). More formally:

Theorem 1.5. *There are linear operators $A \in \{0,1\}^{m \times n}$ such that any depth-d circuit with width-k DNF gates computing Ax must have $W \geq \Omega\left(m \cdot n^{\frac{1}{2(d+k)}}\right)$ wires.*

Finally, building on a recent reduction of Viola [Vio18], we show that Pătraşcu's NOF Conjecture 1.1, even for *3-round protocols*, would prove a polynomial wire lower bound against networks with *arbitrary* gates (i.e., $k = n$), resolving this longstanding open question. This indicates that Conjecture 1.1 may be much stronger than the Multiphase conjecture itself.

Theorem 1.6 (NOF Game Implies Circuit Lower Bounds)**.** *Suppose Conjecture 1.1 holds, even for 3-round protocols. Then for $m = \omega(n)$, there exists a linear [4] operator $A \in \{0,1\}^{m \times n}$ such that any depth-d circuit computing $x \mapsto Ax$ (with* arbitrary gates and unbounded fan-in*) requires $n^{1+\Omega(\varepsilon/d)}$ wires. In particular, if $d = 2$, the conjecture implies that computing Ax for some A requires $n^{1+\frac{\varepsilon}{2}-o(1)}$ wires.*

Comparison to previous work. The aforementioned work of Brody and Larsen [BL15] proves essentially optimal ($\Omega(n/w)$) dynamic lower bounds on variations of the Multiphase problem, when either the update or query algorithms are *nonadaptive* (or in fact "memoryless" in the former, which is an even stronger restriction). Proving lower bounds in the semi-adaptive model is a different ballgame, as the $\tilde{O}(\sqrt{n}/w)$ upper bound suggests (Section A.1). We also remark that [BL15] were the first to observe a (similar but different) connection between nonadaptive data structures and depth-2 circuit lower bounds.

A more recent result of Clifford et. al [CGL15] proves a "threshold" cell-probe lower bound on *general* dynamic data structures solving the Multiphase problem, asserting that fast queries $t_q = o(\log m / \log n)$ require very high $t_u > m^{1-o(1)}$ update time. This result does not rule out data structures with $t_u = t_q = \text{poly}\log(n)$ time for the Multiphase problem (For general data structures, neither does ours).

As far as the Multiphase NOF Game, the aforementioned work of Chattopadhyay et. al [CEEP12] proves a tight $\tilde{\Theta}(\sqrt{n})$ communication lower bound against so-called *"1.5-round"* protocols, in which Bob's message to Alice is *independent of the index i*, hence he is essentially "forwarding" a small ($o(n)$) portion of Charlie's message to Alice (this effectively eliminates Bob from communicating, making it similar to a 2-party problem). While our restricted NOF model is formally incomparable to [CEEP12] (as in our model, Alice is the first speaker), Theorem 1.2 in fact subsumes it by a simple modification (see Appendix A.2). The model we study seems fundamentally stronger than 1.5-protocols, as it inherently involves multiparty NOF communication.

To best of our knowledge, all previous lower bounds ultimately reduce the Multiphase problem to a *2-party* communication game, which makes the problem more amenable to **compression-based** arguments. This is the main departure point of our work. We remark that most of our information-theoretic tools in fact apply to general dynamic data structures. We discuss this further in Section 6 at the end of this paper.

[3] Indeed, even though the *degree* of each gate is bounded (k), it may have *unbounded* fan-in.

[4] Over the boolean Semiring, i.e., where addition are replaced with $\vee$ and multiplication are replaced with $\wedge$. We note that there is evidence that computing Ax over the boolean Semiring is easier than over $\mathbb{F}_2$ [CGL15], hence in that sense our lower bound is stronger than over finite fields.

2 Technical Overview

Here we provide a streamlined overview of our main technical result, Theorem 1.2. As discussed earlier in the introduction, a naïve approach to the Multiphase Game $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ is a "round elimination" approach: Since Charlie's advice consists of only $|U| = o(m)$ bits and he has no knowledge of the index of the interesting subproblem $i \in_R [m]$, his advice U to Bob should convey $o(1)$ bits of information about the interesting set S_i and hence Alice and Bob might hope to simply "ignore" his advice U and use such efficient NOF protocol Γ to generate a too-good-to-be-true *2-party* protocol for set disjointness (by somehow "guessing" Charlies's message which appears useless, and absorbing the error). The fundamental flaw with this intuition is that Charlie's advice is a function of *both* players' inputs, hence conditioning on $U(\vec{S}, T)$ *correlates* the inputs in an arbitrary way, extinguishing the standard "rectangular" (Markov) property of 2-party protocols in the second phase interaction $\Pi^{A \leftrightarrow B}$ between Alice and Bob (This is the notorious feature preventing "direct sum" arguments in NOF communication models). In particular, Chattopadhyay et. al [CEEP12] show that a small advice ($|U| = \tilde{O}(\sqrt{n})$) can already decrease the communication complexity of the multiphase problem to at most the 2-party complexity of set-disjointness under *product* distributions, yielding a surprising 2-round $\tilde{O}(\sqrt{n})$ upper bound on the Multiphase game. (This justifies why our hard distribution for $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ will be a product distribution to begin with, i.e., $\vec{S} \perp T$). Alas, even if the inputs are originally independent ($I(\vec{S}; T) = 0$), they may *not* remain so throughout Π, and it is generally possible that $I(S_i; T|\Pi) \gg 0$. This means that, unlike 2-party protocols, $\Pi = \Pi(U, \vec{S}, T)$ *introduces correlation* between the inputs, and as such, is not amenable to the standard analysis of 2-party communication techniques.

Nevertheless, one might still hope that if the advice U is small enough, then this correlation will be small for an average index S_i when the inputs are independently chosen. At a high level, our proof indeed shows that if only the *first* message of Alice can (directly) depend on her entire input $\vec{S} = S_1, \ldots, S_m$ (whereas her subsequent messages Π^τ are only a function of S_i, i and the transcript history $[\Pi(\vec{S}, T, i)]^{<\tau}$), then it is possible to *simultaneously* control the information cost and correlation of Π, so long as the advice U is small enough ($o(m)$). This in turn facilitates a "robust" direct-sum style argument for *approximate protocols*. More formally, our proof consists of the following two main steps:

• **A low correlation and information random process for computing AND.** The first part of the proof shows that an efficient Restricted NOF protocol Γ for the Multiphase game $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ (under the natural *product* distribution on $\vec{S}, T$) can be used to design a certain random process $Z(X, Y)$ computing the 2-bit AND function (on 2 independent bits $\sim \mathcal{B}_{\Theta(1/\sqrt{n})}$), which *simultaneously* has *low information cost* w.r.t X, Y and *small correlation*, meaning that the input bits remain roughly independent at any point during the process, i.e, $I(X; Y|Z) = o(1/n)$. Crucially, Z is *not* a valid 2-party protocol (Markov chain) – if this were the case, then we would have $I(X; Y|Z) = 0$ since a deterministic 2-party protocol cannot introduce any additional correlation between the original inputs (this is also the essence of the the celebrated *"Cut-and-Paste" Lemma* [BYJKS02]). Nevertheless, we show that for restricted NOF protocols Γ (equivalently, unrestricted protocols in our 4-party model, cf. Figure 5), it is possible to design such random variable $Z(X, Y)$ from Γ, which is *close enough* to a Markov chain. The design of Z requires a careful choice conditioning variables (to ensure that the correlation $\sim |U|/m$ doesn't accumulate over rounds) as well as a "coordinate sampling" step for reducing entropy, though the analysis of this part ultimately uses standard tools (the chain rule and subadditivity of mutual information).We first design a Z' with similar guarantees for single-copy *disjointness*, and then use (a variation of) the standard

direct-sum information cost argument to "scale down" the information and correlation of Z so as to extract the desired random process for 2-bit AND. An important observation in this last step is that the direct sum property of information cost holds not just for communication protocols, but in fact for more general random variables.

• **A "robust" Cut-and-Paste Lemma.** The second part is proving that such random variable $Z(X,Y)$ cannot exist, i.e., ruling out a random process $Z(X,Y)$ computing AND (with 1-sided error under $X,Y \sim^{iid} \mathcal{B}_{\Theta(1/\sqrt{n})}$) which simultaneously has *low information cost* and *small correlation* ($o(1/n)$). The high-level intuition is that, if $Z(X,Y)$ introduces little correlation, then the distribution over X and Y conditioned on $Z(X,Y)$ should remain *approximately* a product distribution, i.e., close to a rectangle. By the correctness guarantee of Z, the distribution on $\{0,1\}^2$ conditioned on $Z(X,Y) = \mathsf{AND}(X,Y) = 0$ must have 0 mass on the $(1,1)$ entry. But if this conditional distribution does not contain $(1,1)$ in its support and *close to a rectangle*, a KL-divergence calculation shows that $Z(X,Y)$ must reveal a lot of information about either X or Y (this calculation crucially exploits the fact that Pinsker's inequality is loose "near the ends", i.e., $\mathsf{D}_{KL}(p||q) \approx \|p-q\|_1$ for $p,q = o(1)$, and there is no quadratic loss). Our argument can be viewed as a generalization of the Cut-and-Paste Lemma to more robust settings of random variables ("approximate protocols"). We remark that while the proof of the original Cut-and-Paste Lemma [BYJKS02] heavily relies on properties of the Hellinger distance, this technique does not seem to easily extend to small-correlation random variables. This forces us to find a more direct argument, which may be of independent interest.

Sketch of nonlinear network lower bound (Theorem 1.5) We prove Theorem 1.5 via reduction from our communication lower bound (Theorem 1.2). Let $A \in \{0,1\}^{m\times n}$ be the random matrix where every entry is $\mathcal{B}_{\Theta(1/\sqrt{n})}$ (i.e., Alice's input in Theorem 1.2), and respectively, let $x = T \in \{0,1\}^n$ be Bob's input in the Multiphase game. We claim that a cheap circuit C_A for computing Ax (over the boolean semiring), with only W wires (where gates compute width-k DNFs), implies an efficient *Restricted* NOF protocol for computing $(Ax)_i = \mathsf{DISJ}(A_i, x)$, which would violate Theorem 1.2. The key point in this reduction is using Charlie's advice to *kill the high fanin gates in* C_A: By a standard averaging argument, there can be $o(m)$ gates with fanin $\omega(W/m)$. Since in Theorem 1.2 Charlie is allowed to send $o(m)$ bits of advice and he sees *both* A and x (but not i), his advice to Bob will consist of the *outputs* of C_A on these $o(m)$ high fanin gates (see Figure 1). Now, all the remaining gates of C_A have fanin $O(W/m)$, and since C_A has depth d, there can be at most $O((W/m)^d)$ such gates in the induced sub circuit whose root is the ith output gate (as this is a tree of hight d with branching-factor $O(W/m)$). Since each gate computes a degree-k polynomial but has *low fan-in*, Bob can afford to send the explicit function computed at this gate using $O((W/m)^k)$ bits. This induces a (1-round) Restricted protocol for the Multiphase game, with $|\Pi| = O((W/m)^{k+d})$ communication, after which Bob can compute the output $(Ax)_i$. By Theorem 1.2, $|\Pi| > \Omega(\sqrt{n})$, which gives the desired bound on the number of wires W.

3 Preliminaries

3.1 Information Theory

In this section, we provide necessary backgrounds on information theory and information complexity that are used in this paper. For further reference, we refer the reader to [CT06].

First we define entropy of a random variable, which intuitively quantifies how "random" a given random variable is.

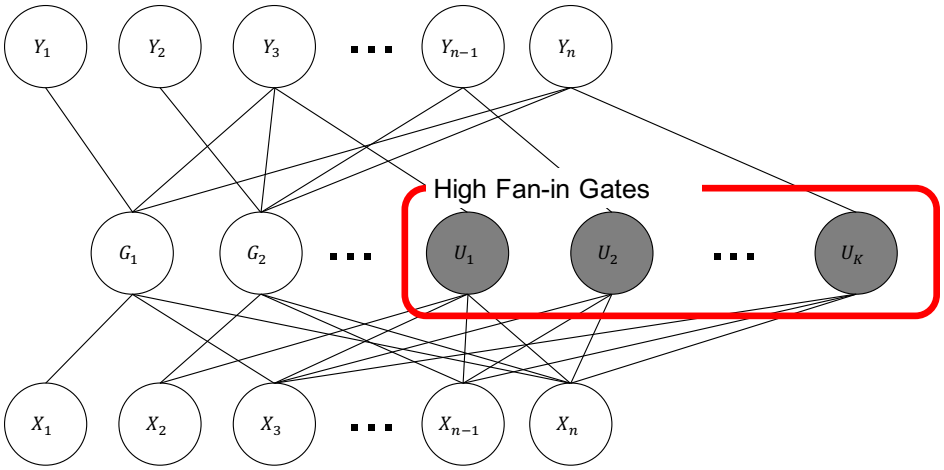

Figure 1: Overview of the reduction

Definition 3.1 (Entropy)**.** *The entropy of a random variable X is defined as*

$$H(X) := \sum_x \Pr[X = x] \log \frac{1}{\Pr[X = x]}.$$

Similarly, the conditional entropy is defined as

$$H(X|Y) := \mathbb{E}_Y \left[\sum_x \Pr[X = x|Y = y] \log \frac{1}{\Pr[X = x|Y = y]} \right].$$

Fact 3.2 (Conditioning Decreases Entropy)**.** *For any random variable X and Y*

$$H(X) \geq H(X|Y)$$

With entropy defined, we can also quantify correlation between two random variables, or how much information one random variable conveys about the other.

Definition 3.3 (Mutual Information)**.** *Mutual information between X and Y (conditioned on Z) is defined as*

$$I(X; Y|Z) := H(X|Z) - H(X|YZ).$$

Similarly, we can also define how much one distribution conveys information about the other distribution.

Definition 3.4 (KL-Divergence)**.** *KL-Divergence between two distributions μ and ν is defined as*

$$\mathsf{D}_{KL}(\mu||\nu) := \sum_x \mu(x) \log \frac{\mu(x)}{\nu(x)}.$$

In order to bound mutual information, it suffices to bound KL-divergence, due to following fact.

Fact 3.5 (KL-Divergence and Mutual Information)**.** *The following equality between mutual information and KL-Divergence holds*

$$I(A; B|C) = \mathbb{E}_{B,C} \left[\mathsf{D}_{KL}(A|_{B=b,C=c}||A|_{C=c}) \right].$$

We can expect the following inequality near 0 distributions.

Fact 3.6 (Divergence Bound)**.** *If* $\mathsf{D}_{KL}(B_q||B_p) < o(p)$ *with* $p = o(1)$*, then* $q \in [0.99p, 1.01p]$

Proof. Since B_q is decreasing as q goes to p, we show that if $q = 0.99p$ or $q = 1.01p$, $D(B_q||B_p) \geq \Omega(p)$. First if $q = 1.01p$, then the term is

$$\begin{aligned} D(B_q||B_p) &= 1.01p \log 1.01 + (1 - 1.01p) \log \frac{1 - 1.01p}{1 - p} \\ &\geq 0.01449p + (1 - 1.01p) \log \left(1 - \frac{0.99p}{1 - p}\right) \\ &\geq 0.01449p + (1 - 1.01p)\,(-0.01443p) \geq \Omega(p) \end{aligned}$$

Similarly, if $q = 0.99p$, we have

$$\begin{aligned} D(B_q||B_p) &= 0.99p \log 0.99 + (1 - 0.99p) \log \frac{1 - 0.99p}{1 - p} \\ &\geq -0.01436p + (1 - 0.99p) \log \left(1 - \frac{0.99p}{1 - p}\right) \\ &\geq -0.01436p + (1 - 0.99p)\,(0.01442p) \geq \Omega(p). \end{aligned}$$

□

We also make use of the following facts on Mutual Information throughout the paper.

Fact 3.7 (Chain Rule)**.** *For any random variable* A, B, C *and* D

$$I(AD; B|C) = I(D; B|C) + I(A; B|CD).$$

Fact 3.8. *For any random variable* A, B, C *and* D*, if* $I(B; D|C) = 0$

$$I(A; B|C) \leq I(A; B|CD).$$

Proof. By the chain rule and non-negativity of mutual information,

$$I(A; B|C) \leq I(AD; B|C) = I(B; D|C) + I(A; B|CD) = I(A; B|CD).$$

□

Fact 3.9. *For any random variable* A, B, C *and* D*, if* $I(B; D|AC) = 0$

$$I(A; B|C) \geq I(A; B|CD).$$

Proof. By the chain rule and non-negativity of mutual information,

$$I(A; B|CD) \leq I(AD; B|C) = I(A; B|C) + I(B; D|AC) = I(A; B|C).$$

□

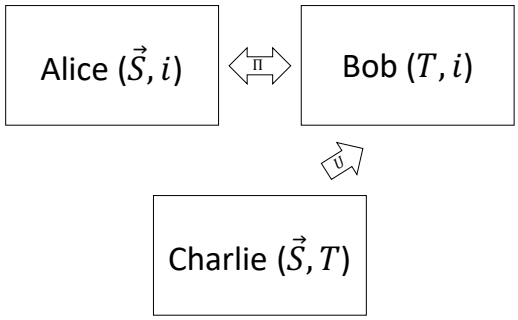

Figure 2: 3-party NOF Communication

3.2 NOF Communication Models

In Pătraşcu's NOF Multiphase Game SEL_f^m, there are 3 players with the following information on their foreheads: **Charlie:** an index $i \in [m]$; **Bob:** a collection of sets $\vec{S} := S_1, \ldots, S_m \subseteq [n]$; **Alice:** a set $T \subseteq [n]$. I.e., Charlie has access to both $\vec{S}$ and T, but not to i. Alice has access to $\vec{S}$ and i, and Bob has access to T and i. The goal is to compute

$$\mathsf{SEL}_f^m(\vec{S}, T, i) := f(S_i, T).$$

The communication proceeds as follows: In the first stage of the game, Charlie sends a message ("*advice*") $U = U(\vec{S}, T)$ *privately to Bob*. In the second stage, Alice and Bob continue to communicate in the standard 2-party settting to compute $f(S_i, T)$ (see Figure 2). We denote such protocol by $\Gamma := (U, \Pi_i)$ where Π_i is the second stage transcript, assuming the index of the interesting subproblem is i.

Unfortunately, lower bounds for general protocols in Pătraşcu's 3-party NOF model seem beyond the reach of current techniques, as we show in Section 5 that Conjecture 1.1, even for *3-round* protocols, would resolve a major open problem in circuit complexity. Fortunately, for dynamic data structure applications, weaker versions of the NOF model suffice (this is indeed one of the main messages of this paper).

We consider the following restricted class of protocols. We say that $\Gamma = (U, \Pi)$ is a **restricted** NOF protocol if Alice is the first speaker in Π (in the second stage of the game) and only her *first* message Π_i^1 to Bob depends on her entire input $\vec{S}$ and i, whereas in subsequent rounds, Alice's messages Π^τ may depend only on S_i, i and the history of the transcript $\Pi^{<\tau}$ with Bob. Note that the latter means that Alice and Bob's subsequent messages can still heavily depend on S_{-i}, but only through the transcript (this feature better captures data structures, since the query algorithm can only adapt based on the information in cells it already probed, and not the entire memory).

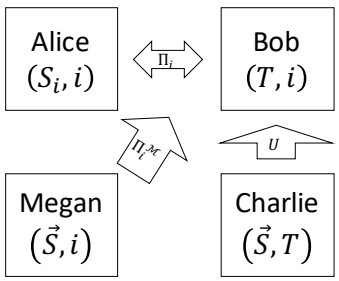

Figure 3: 4-party NOF Communication

An equivalent 4-party NOF model. Restricted 3-party NOF protocols are more naturally described by the following 4-party NOF model. **Alice** has access only to S_i and i, **Bob** has access to T and i. **Charlie** has access to $\vec{S}$ and T, but no access to i. **Megan** has access to $\vec{S}$ and i. In the first stage of the protocol Γ, in addition to Charlie's *private* message to Bob $U(\vec{S}, T)$, Megan can *broadcast* a message $\Pi^{\mathcal{M}} = \Pi^{\mathcal{M}}(\vec{S}, i)$ to both Alice and Bob. Thereafter, Alice and Bob proceed to communicate in the 2-party model as before, denoted $\Pi^{A \leftrightarrow B}$. See Figure 5. We denote 4-party protocols by $\Gamma = (U, \Pi)$ where $\Pi := (\Pi^{\mathcal{M}}, \Pi^{A \leftrightarrow B})$. We write $\Pi_i := (\Pi_i^{\mathcal{M}}, \Pi_i^{A \leftrightarrow B})$ to denote the transcript of Π when the index of the interesting subproblem is $i \in [m]$.

It is straightforward to see that Restricted 3-party NOF protocols for the Multiphase Game are equivalent to (unrestricted) 4-party protocols (by setting Alice's first message as Megan's message $\Pi_i^{\mathcal{M}}$, and Charlie remains unchanged). As such, our main technical theorem (Theorem 1.2) can be restated as follows.

Theorem 3.10 (4-party NOF Lower Bound)**.** *Let $m > \omega(n)$. For any 4-party NOF protocol $\Gamma = (U, \Pi)$ that solves $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ with $|U| < o(m)$, there exists $i \in [m]$ such that $|\Pi_i| > \Omega(\sqrt{n})$.*

4 Lower Bound for 4-party NOF Protocols

Notations. We denote by S_i^j the j-th entry of the set S_i and analogously for T. We write $S_i^{<j} := S_i^1, \ldots, S_i^{j-1}$, similarly $S_i^{-j} := S_i^1, \ldots, S_i^{j-1}, S_i^{j+1}, \ldots, S_i^n$, and analogously for T.

For technical reasons, we shall need to carry out the proof on a restricted subset $\mathcal{P}$ of the original $[m]$ coordinates, where $\mathcal{P} = (i_1, \ldots, i_p) \in [m]^p$. We write $i \in \mathcal{P}$ if there exists some index $\ell \in [p]$ such that $i_\ell = i$. We write $(i_1, \ldots, i_{\ell-1})$ and $(i_1, \ldots, i_\ell)$ in short hand as $i_{<\ell}$ and $i_{\leq \ell}$ respectively. $S_{i_{<\ell}}$ refers to $S_{i_1}, \ldots, S_{i_{\ell-1}}$ and $\Pi^{\mathcal{M}}_{i_{<\ell}}$ refers to $\Pi^{\mathcal{M}}_{i_1}, \ldots, \Pi^{\mathcal{M}}_{i_{\ell-1}}$. Furthermore, $S_{\mathcal{P}}$ and $\Pi^{\mathcal{M}}_{\mathcal{P}}$ refers to $S_{i_1}, \ldots, S_{i_p}$ and $\Pi^{\mathcal{M}}_{i_1}, \ldots, \Pi^{\mathcal{M}}_{i_p}$ respectively. Also Π_i^{ans} denotes the output of $\Gamma = (U, \Pi)$ when the index of interest is $i \in [m]$.

Let $C := \max_{i \in [m]} |\Pi_i|$ be the maximal number of bits exchanged between Megan, Alice and Bob over all $i \in [m]$. Then in particular, for every $i \in [m]$,

$$|\Pi_i^{\mathcal{M}}| \leq |\Pi_i| \leq C. \tag{1}$$

Observe that since Megan does not have access to T, for any subset $\mathcal{P}$ of coordinates it holds that

$$I(T; S_{\mathcal{P}} \Pi^{\mathcal{M}}_{\mathcal{P}}) = 0 \tag{2}$$

assuming $\vec{S} \perp T$, since Megan's message only depends on i and $\vec{S}$. Indeed, $I(T; S_{\mathcal{P}} \Pi^{\mathcal{M}}_{\mathcal{P}}) \leq I(T; \vec{S}, \Pi^{\mathcal{M}}_{\mathcal{P}}) = I(T; \vec{S}) + I(T; \Pi^{\mathcal{M}}_{\mathcal{P}} \mid \vec{S}) = 0$. It is noteworthy that, by contrast, $I(T; \Pi^{\mathcal{M}}_{\mathcal{P}} \mid \vec{S}, \mathbf{U}) \neq 0$, since conditioning on U correlates Megan's message with Bob's input. Indeed, dealing with this subtle feature will be the heart of this section and will later explain the choice of Z^{DISJ}.

Hard Distribution. We consider the natural hard product distribution for set-disjointness, extended to $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$: For all $i \in [m]$ and $j \in [n]$, S_i^j and T^j are i.i.d. Bernoulli $\mathcal{B}_\gamma$ for $\gamma = \frac{1}{1000\sqrt{n}}$.

4.1 A Low Correlation Random Process for DISJ_n

The goal of this section is to show that an efficient 4-party NOF protocol Γ for $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ implies a *low-correlation, low-information* random process for computing a single copy of set-disjointness (under the hard product distribution). For technical reasons, we restrict the proof to a random subset $\mathcal{P} = (I_1, \ldots, I_p) \in_R [m]^p$ of p coordinates, with the constraint that for any $k_1, k_2 \in [p]$, if

$k_1 \neq k_2$, then $I_{k_1} \neq I_{k_2}$ where p is a parameter that will be chosen as $o(m)$.[5] Let $\boldsymbol{\ell} \in_R [p]$ be a uniformly random index. We shall prove the following Lemma:

Lemma 4.1. *Let $\Gamma = (U, \Pi)$ be a 4-party NOF protocol for $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ with $|\Pi_i| < C$ for all $i \in [m]$. Then for $p = o(m)$, there exists a random variable Z^{DISJ} containing $\mathcal{P}$ and $\boldsymbol{\ell}$ such that*

- *If $\mathsf{DISJ}_n(S_{I_{\boldsymbol{\ell}}}, T) = 0$, then $Z^{\mathsf{DISJ}}_{ans} = 0$.*
- *If $\mathsf{DISJ}_n(S_{I_{\boldsymbol{\ell}}}, T) = 1$, then $Z^{\mathsf{DISJ}}_{ans} = 1$.*
- *Satisfies the following information cost bound*

$$I(Z^{\mathsf{DISJ}} T; S_{I_{\boldsymbol{\ell}}}) \leq C + o(C) \tag{3}$$
$$I(Z^{\mathsf{DISJ}}; T) \leq C \tag{4}$$

- *Satisfies the following correlation bound*

$$I(S_{i_\ell}; T | Z^{\mathsf{DISJ}}) \leq O\left(\frac{|UT|}{p}\right). \tag{5}$$

Intuitively, Lemma 4.1 states that an efficient 4-party NOF protocol for $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ can be used to design a random process $Z(S_{I_{\boldsymbol{\ell}}}, T)$ which for a random $\mathcal{P}$ and $\boldsymbol{\ell}$ computes DISJ_n on inputs $S_{I_{\boldsymbol{\ell}}}$ and T, in a way that simultaneously: (i) Z reveals small information on average about both $S_{I_{\boldsymbol{\ell}}}$ and T, and (ii) creates small correlation between $S_{I_{\boldsymbol{\ell}}}$ and T assuming $|UT| = o(p) \leq o(m)$ (i.e., it is in some sense "close" to a 2-party communication protocol). The choice of Z is set to

$$Z^{\mathsf{DISJ}} := \Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}}, \mathcal{P}, \boldsymbol{\ell} = \Pi_{I_{\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}}, \mathcal{P}, \boldsymbol{\ell}$$
$$Z^{\mathsf{DISJ}}_{ans} := \Pi^{ans}_{I_{\boldsymbol{\ell}}}.$$

where the equality holds for Z^{DISJ} since $\Pi^{\mathcal{M}}_{I_{\boldsymbol{\ell}}}$ is included in $\Pi_{I_{\boldsymbol{\ell}}}$. Since we do not bound the number of rounds, we can without loss of generality assume that $\Pi^{ans}_{I_{\boldsymbol{\ell}}}$ is included in $\Pi_{I_{\boldsymbol{\ell}}}$. Note that $\Pi_{I_{\boldsymbol{\ell}}}, \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}}$ are random variables that depend on Charlie's advice U, but importantly U is not included explicitly in Z^{DISJ}. We begin with the following claim, which morally states that $Z^{\mathsf{DISJ}} \setminus \Pi_{I_{\boldsymbol{\ell}}}$ reveals little information on an average set S_i:

Claim 4.2.

$$I(S_{I_{\boldsymbol{\ell}}}; S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T, \mathcal{P}, \boldsymbol{\ell}) \leq \frac{p \cdot C}{m - p}. \tag{6}$$

Proof. First, note that we have

$$I(S_{I_{\boldsymbol{\ell}}}; S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}}, \mathcal{P}, \boldsymbol{\ell}) = I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} | \mathcal{P}\boldsymbol{\ell}) = \mathbb{E}_{\mathcal{P}, \boldsymbol{\ell}} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} | \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell) \right]$$

where the first equality holds since $I(S_{I_{\boldsymbol{\ell}}}; \mathcal{P}, \boldsymbol{\ell}) = 0$ since $\mathcal{P}, \ell$ are independent of $\vec{S}$ and T along with Fact 3.7. Furthermore, for any setting of $\mathcal{P}$ and $\boldsymbol{\ell}$, we have

$$I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} | \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell) = I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}).$$

since the choice of $\mathcal{P}$ and $\boldsymbol{\ell}$ are independent of entries in $\vec{S}$ and T. Therefore, we have

$$I(S_{I_{\boldsymbol{\ell}}}; S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} \mathcal{P}\boldsymbol{\ell}) = I(S_{I_{\boldsymbol{\ell}}}; S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} | \mathcal{P}\boldsymbol{\ell}) = \mathbb{E}_{\mathcal{P}, \boldsymbol{\ell}} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}) \right]. \tag{7}$$

[5] This is equivalent to picking a random ordering over a random p-sized subset in $[m]$

Now consider fixed $\boldsymbol{\ell} = \ell$ and $I_{<\ell} = i_{<\ell}$. Then we prove the following inequality.

$$\mathbb{E}_{\substack{i_\ell, \\ \forall r<\ell, i_\ell \neq i_r}} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T) \right] \leq \frac{p \cdot C}{m-p}. \tag{8}$$

First observe that

$$\begin{aligned} &I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T | I_{<\ell} = i_{<\ell}, I_{\geq \ell}) = I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T | I_\ell, I_{<\ell} = i_{<\ell}) \\ &= \mathbb{E}_{i_\ell} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T | I_\ell = i_\ell, I_{<\ell} = i_{<\ell}) \right] = \mathbb{E}_{\substack{i_\ell, \\ \forall r<\ell, i_\ell \neq i_r}} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T) \right] \end{aligned}$$

Now we have that for all $i \in [m]$ such that $i \notin (i_1, \ldots, i_{\ell-1})$, S_i's are i.i.d. Therefore we get

$$\begin{aligned} &\mathbb{E}_{\substack{i_\ell, \\ \forall r<\ell, i_\ell \neq i_r}} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T) \right] = \frac{1}{m-(\ell-1)} \sum_{i \notin i_{<\ell}} I(S_i; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T) \\ &\leq \frac{I(S_{i \notin i_{<\ell}}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T)}{m-(\ell-1)} = \frac{I(S_{i \notin i_{<\ell}}; \Pi^{\mathcal{M}}_{i_{<\ell}} | S_{i_{<\ell}}, T)}{m-(\ell-1)} \leq \frac{H(\Pi^{\mathcal{M}}_{i_{<\ell}} | S_{i_{<\ell}}, T)}{m-(\ell-1)} \\ &\leq \frac{H(\Pi^{\mathcal{M}}_{i_{<\ell}})}{m-(\ell-1)} \leq \frac{(\ell-1)C}{m-(\ell-1)} \end{aligned}$$

where the last equality holds since $I(S_{i \notin i_{<\ell}}; S_{i_{<\ell}}, T) = 0$ from our assumption on the hard distribution. Now since we have $\ell \leq p$, we get

$$\mathbb{E}_{\substack{i_\ell, \\ i_\ell \notin i_{<\ell}}} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, T) \right] \leq \frac{(\ell-1)C}{m-(\ell-1)} \leq \frac{p \cdot C}{m-p}.$$

Therefore, we have that (8) holds for any fixed $\boldsymbol{\ell} = \ell$ and $I_{<\ell} = (i_1, \ldots, i_{\ell-1})$. Taking expectation over $\mathcal{P}$ and $\boldsymbol{\ell}$, we get

$$\mathbb{E}_{\mathcal{P}, \boldsymbol{\ell}} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T) \right] = \mathbb{E}_{\boldsymbol{\ell}} \left[\mathbb{E}_{I_{<\ell}} \mathbb{E}_{I_\ell} \left[I(S_{i_\ell}; S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T) \right] \right] \leq \frac{p \cdot C}{m-p}.$$

□

The next claim, which is another direct application of the chain rule, asserts that for a random coordinate $i \in \mathcal{P}$, Megan's messages in Π_i ($\Pi^{\mathcal{M}}_i$) do not heavily depend on T, U, conditioned on previous coordinate transcripts.

Claim 4.3. *For any fixed $\mathcal{P} = i_{\leq p}$, if $\boldsymbol{\ell}$ is uniformly distributed over $[p]$*

$$\mathbb{E}_l \left[I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \boldsymbol{\ell} = \ell, \mathcal{P} = i_{\leq p} \right] \leq O\left(\frac{|UT|}{p} \right) \tag{9}$$

Proof. Again, since $\boldsymbol{\ell}$ is picked independently at random (UT is independent of ℓ), we get

$$I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \boldsymbol{\ell} = \ell, \mathcal{P} = i_{\leq p}) = I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \mathcal{P} = i_{\leq p})$$

Then taking expectation over $\boldsymbol{\ell}$, we get

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{\ell}} \left[I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \mathcal{P} = i_{\leq p}) \right] \\ &= \frac{1}{p} \sum_{\ell \in [p]} I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \mathcal{P} = i_{\leq p}) \\ &= \frac{I(S_{i_{\leq p}} \Pi^{\mathcal{M}}_{i_{\leq p}}; UT | \mathcal{P} = i_{\leq p})}{p} \leq \frac{|UT|}{p}. \end{aligned}$$

□

Recall that $\Pi_i := (\Pi_i^{\mathcal{M}}, \Pi_i^{A\leftrightarrow B})$ is the transcript between Megan, Alice and Bob when the index of the interesting subproblem is i. We now turn to establish the fact that conditioning on Π_i cannot introduce too much correlation between the (originally independent) S_i and T. As discussed in the introduction, if Π_i were a standard (deterministic) 2-party protocol, then this would have indeed been the case (as the *rectangle* property of communication protocols ensures that independent inputs S_i, T remain so throughout the protocol: $I(S_i; T|\Pi_i) = 0$). Alas, Π_i no longer has the rectangle property anymore (as Charlie's message $U(\vec{S}, T)$ correlates the inputs in an arbitrary way). Fortunately, we will be able to show that if Megan's messages only depend on $\vec{S}$ and i, and Alice's response only depend on S_i, i and previous transcript then we can control the correlation introduced by Z^{DISJ} (by adding the aforementioned extra variables in the definition of Z^{DISJ}) *without increasing the information cost* of Z^{DISJ} with respect to S_i and T. We begin with the following claim, which shows that the effect of conditioning on Z^{DISJ} can be upper bounded by the following term:

Claim 4.4. *For any fixed $\ell \in [p]$, and $\mathcal{P} = i_{\leq p}$*

$$\begin{aligned}&I(S_{i_\ell}; T|\Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell) = I(S_{i_\ell}; T|\Pi_{i_\ell}^{\mathcal{M}} \Pi_{i_\ell}^{A\leftrightarrow B}, S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell)\\ &\leq I(S_{i_\ell} \Pi_{i_\ell}^{\mathcal{M}}; UT|S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell).\end{aligned}$$

Proof. The proof is by induction on the number of rounds of $\Pi_{i_\ell}^{A\leftrightarrow B} := \Pi_{i_\ell}^1, \ldots, \Pi_{i_\ell}^C$. If at $\tau \in [C]$, it is Alice's turn to speak, then since Alice's message is a function of $S_{i_\ell} \Pi_{i_\ell}^{\mathcal{M}}$ and $\Pi_{i_\ell}^{<\tau}$, it holds that

$$I(\Pi_{i_\ell}^{\tau}; UT|S_{i_\ell} \Pi_{i_\ell}^{\mathcal{M}} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \Pi_{i_\ell}^{<\tau}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell) \leq H(\Pi_{i_\ell}^{\tau}|S_{i_\ell}, \Pi_{i_\ell}^{\mathcal{M}}, \Pi_{i_\ell}^{<\tau}) = 0 \tag{10}$$

(Note that this would not have been true had Alice's message been a function of all $\vec{S}$, because U correlates $\vec{S}$ and T. This is where we use the fact that only Megan's message $\Pi_{i_{<\ell}}^{\mathcal{M}}$ can depend on all $\vec{S}$). If it is Bob's turn to speak at round $\tau \in [C]$, then it still holds that

$$I(\Pi_{i_\ell}^{\tau}; S_{i_\ell}|UT, S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \Pi_{i_\ell}^{\mathcal{M}}, \Pi_{i_\ell}^{<\tau}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell) = 0 \tag{11}$$

since Bob's message is determined by $UT, \Pi_{i_\ell}^{\mathcal{M}}$ and $\Pi_{i_\ell}^{<\tau}$ or equivalently

$$\begin{aligned}&I(\Pi_{i_\ell}^{\tau}; S_{i_\ell}|UT, S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \Pi_{i_\ell}^{\mathcal{M}}, \Pi_{i_\ell}^{<\tau}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell)\\ &\leq H(\Pi_{i_\ell}^{\tau}|UT, \Pi_{i_\ell}^{\mathcal{M}}, \Pi_{i_\ell}^{<\tau}) = 0.\end{aligned}$$

Then applying Fact 3.9 iteratively with (10) and (11) for any $\ell \in [p]$, we get

$$\begin{aligned}&I(S_{i_\ell}; UT|\Pi_{i_\ell}^{\mathcal{M}} \Pi_{i_\ell}^{\leq C} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell)\\ &\leq I(S_{i_\ell}; UT|\Pi_{i_\ell}^{\mathcal{M}} \Pi_{i_\ell}^{<C} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell) \leq \ldots\\ &\leq I(S_{i_\ell}; UT|\Pi_{i_\ell}^{\mathcal{M}} \Pi_{i_\ell}^{1} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell)\\ &\leq I(S_{i_\ell}; UT|\Pi_{i_\ell}^{\mathcal{M}} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell).\end{aligned}$$

We get the final inequality by non-negativity of mutual information or

$$I(S_{i_\ell}; UT|\Pi_{i_\ell}^{\mathcal{M}} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell) \leq I(S_{i_\ell} \Pi_{i_\ell}^{\mathcal{M}}; UT|S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}, \mathcal{P} = i_{\leq p}, \boldsymbol{\ell} = \ell).$$

□

We are finally ready to prove Lemma 4.1 using Claim 4.2, Claim 4.3 and Claim 4.4.

Proof of Lemma 4.1. Recall the definition of $Z^{\mathsf{DISJ}}(S_{I_{\boldsymbol{\ell}}}, T) := \Pi_{I_{\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}}, \mathcal{P}, \boldsymbol{\ell}$. The correctness guarantee of $\mathsf{DISJ}_n(S_{I_{\boldsymbol{\ell}}}, T)$ holds since we set $\Pi^{ans}_{I_{\boldsymbol{\ell}}}$ as Z^{DISJ}_{ans}, and the original NOF protocol $\Gamma = (U, \Pi)$ was assumed to have 0 error.

To establish (3), we get from Claim 4.2 and Fact 3.7 that

$$\begin{aligned} I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T, \mathcal{P}, \boldsymbol{\ell}; S_{I_{\boldsymbol{\ell}}}) &= I(S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T, \mathcal{P}, \boldsymbol{\ell}; S_{I_{\boldsymbol{\ell}}}) + I(\Pi_{i_\ell}; S_{i_\ell} | S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T, \mathcal{P}, \boldsymbol{\ell}) \\ &\leq \frac{p \cdot C}{m - p} + C. \end{aligned}$$

Since we set $p := o(m)$, we get $\frac{p \cdot C}{m-p} = o(C)$.

Next for (4), we write

$$\begin{aligned} I(Z^{\mathsf{DISJ}}; T) = I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}}, \mathcal{P}, \boldsymbol{\ell}; T) &= \underbrace{I(\mathcal{P}, \boldsymbol{\ell}; T)}_{=0} + I(S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}; T | \mathcal{P}, \boldsymbol{\ell}) + \underbrace{I(\Pi_{i_\ell}; T | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \mathcal{P}, \boldsymbol{\ell})}_{\leq C} \\ &\leq \underbrace{I(S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}; T | \mathcal{P}, \boldsymbol{\ell})}_{=0} + C \leq C. \end{aligned} \tag{12}$$

where we used (2) for $I(S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}; T | \mathcal{P}, \boldsymbol{\ell}) = I(S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}; T) = 0$ (recall that $\mathcal{P}, \boldsymbol{\ell}$ are chosen independently of the inputs, so conditioning on them does not change things). We remark that here we (crucially) used the fact that Megan is only allowed to send a single message (i.e., no further interaction with the player holding $\vec{S}$ is allowed).

Finally for (5), from Claim 4.4, we have that for every $\ell \in [p]$ and $\mathcal{P}$,

$$I(S_{i_\ell}; T | \Pi_{i_\ell}, S_{i_{<\ell}}, \Pi^{\mathcal{M}}_{i_{<\ell}}, \boldsymbol{\ell}, \mathcal{P}) \leq I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}}, \Pi^{\mathcal{M}}_{i_{<\ell}}, \boldsymbol{\ell}, \mathcal{P}). \tag{13}$$

But from Claim 4.3, we know that over random $\boldsymbol{\ell} \in_R [p]$, we have

$$\begin{aligned} &I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \boldsymbol{\ell}, \mathcal{P}) \\ &= \mathbb{E}_{\boldsymbol{\ell}, \mathcal{P}} \left[I(S_{i_\ell} \Pi^{\mathcal{M}}_{i_\ell}; UT | S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}}, \mathcal{P}) \right] \leq O\left(\frac{|UT|}{p} \right). \end{aligned} \tag{14}$$

□

4.2 A Random Process for AND with Low Information & Correlation

We now show how to "scale down" the random process Z^{DISJ} (obtained from the NOF protocol Γ in Lemma 4.1) so as to generate another random process (*not* a 2-party protocol) that "approximately" computes the 2-bit $\mathsf{AND}(X, Y)$ function (on independent random bits) under $X, Y \sim \mathcal{B}_\gamma$, with information correlation smaller by a factor of n. This follows the standard "direct sum" embedding (e.g., [BBCR10, BR11, Bra15]) – the important observation here is that the direct sum property of the information cost function apply to general interactive processes and not just to communication protocols. This is the content of the next lemma.

Lemma 4.5. *Let $\Gamma = (U, \Pi)$ be a 4-party NOF protocol that solves $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ with $|U| < o(p) < o(m)$ and $|\Pi_i| < C$ for all $i \in [m]$. Then there exists a random variable $Z^{\mathsf{AND}} = Z^{\mathsf{AND}}(X, Y)$ such that*

- *If $\mathsf{AND}(X, Y) = 1$, then Z^{AND}_{ans} outputs 0.*
- *If $\mathsf{AND}(X, Y) = 0$, then Z^{AND}_{ans} outputs 0 with probability at most 0.001.*

- *Has following information cost guarantees*

$$I(Z^{\mathsf{AND}}; X) \leq \frac{C + o(C)}{n} \tag{15}$$

$$I(Z^{\mathsf{AND}}; Y) \leq \frac{C}{n} \tag{16}$$

- *Has following correlation guarantee*

$$I(X; Y | Z^{\mathsf{AND}}) < o(1/n) \tag{17}$$

Proof. Consider the following embedding of bits X and Y to Γ.

1. Select $\mathcal{P}$, $\boldsymbol{\ell} \in_R [p]$, $\boldsymbol{j} \in_R [n]$ uniformly at random.
2. Set $S^{\boldsymbol{j}}_{I_{\boldsymbol{\ell}}} = X$ and $T^{\boldsymbol{j}} = Y$.
3. Sample the rest of the coordinates all i.i.d. $\mathcal{B}_\gamma$.
4. Run $\Gamma = (U, \Pi)$ with $I_{\boldsymbol{\ell}}$ as the index. Then return the output.

Protocol 1: Embedding $\mathsf{AND}(X, Y)$

Now we claim that

$$Z^{\mathsf{AND}} := \Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T^{<\boldsymbol{j}} \boldsymbol{j} \mathcal{P} \boldsymbol{\ell} = Z^{\mathsf{DISJ}} T^{<\boldsymbol{j}} \boldsymbol{j}$$

with Z^{AND}_{ans} set as $\Pi^{ans}_{I_{\boldsymbol{\ell}}}$ satisfies the conditions of Lemma 4.5.

Correctness Suppose $\mathsf{AND}(X, Y) = 1$. Then $\Pi_{I_{\boldsymbol{\ell}}}$ must output 0 since $\mathsf{DISJ}(S_{I_{\boldsymbol{\ell}}}, T) = 0$ from embedding. Also note that given $\mathsf{AND}(X, Y) = 0$, $\mathsf{DISJ}(S_{I_{\boldsymbol{\ell}}}, T) = 0$ with at most 0.001 probability since $\mathsf{DISJ}(S^{-\boldsymbol{j}}_{I_{\boldsymbol{\ell}}}, T^{-\boldsymbol{j}}) = 0$ with at most 0.001 probability on our distribution on $\vec{S}$ and T for any setting of $\boldsymbol{j}, \mathcal{P}, \boldsymbol{\ell}$. Therefore $\Pi_{I_{\boldsymbol{\ell}}}$ outputs 0 with at most probability 0.001.

Information Cost Recall that we have from Lemma 4.1, (3), (4), (5) or

$$I(Z^{\mathsf{DISJ}} T; S_{I_{\boldsymbol{\ell}}}) = I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T, \mathcal{P}, \boldsymbol{\ell}; S_{I_{\boldsymbol{\ell}}}) \leq C + o(C)$$

$$I(Z^{\mathsf{DISJ}}; T) = I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T, \mathcal{P}, \boldsymbol{\ell}; T) \leq C$$

$$I(S_{i_\ell}; T | Z^{\mathsf{DISJ}}) = I(S_{i_\ell}; T | \Pi_{i_\ell} S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T, \mathcal{P}, \boldsymbol{\ell}) \leq O\left(\frac{|UT|}{p}\right)$$

Now since we embed X and Y in random $\boldsymbol{j} \in [n]$, we get

$$I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T^{<\boldsymbol{j}}, \boldsymbol{j}, \mathcal{P}, \boldsymbol{\ell}; X) = \mathbb{E}_{\mathcal{P},\ell} \mathbb{E}_j \left[I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T^{<j}; X | \boldsymbol{j} = j, \mathcal{P}, \boldsymbol{\ell}) \right]$$

$$\leq \mathbb{E}_{\mathcal{P},\ell} \mathbb{E}_j \left[I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T; S^j_{i_\ell} | \boldsymbol{j} = j, \mathcal{P}, \boldsymbol{\ell}) \right] = \mathbb{E}_{\mathcal{P},\ell} \mathbb{E}_j \left[I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T; S^j_{i_\ell} | \mathcal{P}, \boldsymbol{\ell}) \right]$$

$$= \mathbb{E}_{\mathcal{P},\ell} \left[\frac{1}{n} \sum_{j \in [n]} I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T; S^j_{i_\ell} | \mathcal{P}, \boldsymbol{\ell}) \right] \leq \mathbb{E}_{\mathcal{P},\ell} \left[\frac{1}{n} \sum_{j \in [n]} I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T; S^j_{i_\ell} | S^{<j}_{i_\ell}, \mathcal{P}, \boldsymbol{\ell}) \right]$$

$$\leq \mathbb{E}_{\mathcal{P},\ell} \left[\frac{I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi^{\mathcal{M}}_{i_{<\ell}} T; S_{i_\ell} | \mathcal{P}, \boldsymbol{\ell})}{n} \right] = \frac{I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi^{\mathcal{M}}_{I_{<\boldsymbol{\ell}}} T, \mathcal{P}, \boldsymbol{\ell}; S_{I_{\boldsymbol{\ell}}})}{n}$$

$$= \frac{I(Z^{\mathsf{DISJ}} T; S_{I_{\boldsymbol{\ell}}})}{n} \leq \frac{C + o(C)}{n}$$

where the second inequality holds from $I(S_{i_\ell}^j; S_{i_\ell}^{<j}|\mathcal{P}, \boldsymbol{\ell}) = 0$. The second to last equality holds from $I(\mathcal{P}, \boldsymbol{\ell}; S_{I_{\boldsymbol{\ell}}}) = 0$ or (7). For (16), analogously, we get

$$
\begin{aligned}
&I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi_{I_{<\boldsymbol{\ell}}}^{\mathcal{M}} T^{<\boldsymbol{j}}, \boldsymbol{j}, \mathcal{P}, \boldsymbol{\ell}; Y) = \mathbb{E}_{\mathcal{P},\ell}\mathbb{E}_j \left[I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}} T^{<j}; T^j | \boldsymbol{j} = j, \mathcal{P}, \boldsymbol{\ell}) \right] \\
&\leq \mathbb{E}_{\mathcal{P},\ell}\mathbb{E}_j \left[I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}} T^{<j}; T^j | \mathcal{P}, \boldsymbol{\ell}) \right] = \mathbb{E}_{\mathcal{P},\ell}\mathbb{E}_j \left[I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}; T^j | T^{<j}, \mathcal{P}, \boldsymbol{\ell}) \right] \\
&= \mathbb{E}_{\mathcal{P},\ell} \left[\frac{1}{n} \sum_{j \in [n]} I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}; T^j | T^{<j}, \mathcal{P}, \boldsymbol{\ell}) \right] \leq \mathbb{E}_{\mathcal{P},\ell} \left[\frac{I(\Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}}; T | \mathcal{P}, \boldsymbol{\ell})}{n} \right] \\
&= \frac{I(\Pi_{I_{\boldsymbol{\ell}}} S_{I_{<\boldsymbol{\ell}}} \Pi_{I_{<\boldsymbol{\ell}}}^{\mathcal{M}}, \mathcal{P}, \boldsymbol{\ell}; T)}{n} = \frac{I(Z^{\mathsf{DISJ}}; T)}{n} \leq \frac{C}{n}.
\end{aligned}
$$

where the second equality holds from $I(T^{<j}; T^j | \mathcal{P}, \boldsymbol{\ell}) = I(T^{<j}; T^j) = 0$.

Low Correlation Now for (17), we again take expectation over $\boldsymbol{j}$.

$$
\begin{aligned}
&I(X; Y | \Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}} T^{<j}, \boldsymbol{j}, \mathcal{P}, \boldsymbol{\ell}) = \mathbb{E}_{\mathcal{P},\ell}\mathbb{E}_j \left[I(S_{i_\ell}^j; T^j | \Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}} T^{<j}, \mathcal{P}, \boldsymbol{\ell}, \boldsymbol{j} = j) \right] \\
&= \mathbb{E}_{\mathcal{P},\ell}\mathbb{E}_j \left[I(S_{i_\ell}^j; T^j | \Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}} T^{<j}, \mathcal{P}, \boldsymbol{\ell}) \right] \\
&\leq \mathbb{E}_{\mathcal{P},\ell} \left[\frac{1}{n} \sum_j I(S_{i_\ell}; T^j | \Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}} T^{<j}, \mathcal{P}, \boldsymbol{\ell}) \right] \\
&= \mathbb{E}_{\mathcal{P},\ell} \left[\frac{I(S_{i_\ell}; T | \Pi_{i_\ell} S_{i_{<\ell}} \Pi_{i_{<\ell}}^{\mathcal{M}} \mathcal{P}, \boldsymbol{\ell})}{n} \right] = \frac{I(S_{I_{\boldsymbol{\ell}}}; T | Z^{\mathsf{DISJ}})}{n} \leq O\left(\frac{|UT|}{pn} \right) \leq o(1/n).
\end{aligned}
$$

where the last bound follows from Lemma 4.1 with $|UT| < o(p)$. □

4.3 An Information Lower Bound for Low-Correlation AND Computation

In this section, we rule out a random process that computes the 2-bit AND function with *simultaneously* low information cost and small correlation. The proof can be viewed as a generalization of the classic Cut-and-Paste Lemma [BYJKS02] from 2-party protocols to a more general setting of low-correlation random variables.

A Robust "Cut-and-Paste" Lemma for Product Distribution. Instead of restricting to a two-party protocol, we generalize celebrated "cut-and-paste" lemma to any random process. We prove the following general structural bound for any random process Z when X and Y are distributed B_γ independently with $\gamma = o(1)$.

Lemma 4.6 (Robust Cut and Paste)**.** *Suppose X and Y are distributed i.i.d. $\mathcal{B}_\gamma$ with $\gamma = o(1)$. Then consider a random variable Z containing Z_{ans} such that*

- *If $\mathsf{AND}(X, Y) = 1$ then $Z_{ans} = 0$. Otherwise $\Pr[Z_{ans} = 0] < 0.001$.*
- $\Pr[Z_{ans} = 1] \geq 1/2$
- *Satisfies following two inequalities*

$$I(X; Y | Z) \leq o(\gamma^2) \tag{18}$$
$$I(Z; X) \leq o(\gamma) \tag{19}$$

Then it must be the case $I(Y;Z) \geq \Omega(\gamma)$.

We remark that in the usual cut-and-paste setting introduced in [BYJKS02], one requires $I(X;Y|Z) = I(X;Y)$. And this follows from Z being a 2-party communication protocol. It is crucial in standard cut-and-paste argument that $I(X;Y|Z) = I(X;Y)$. Otherwise, one cannot make a connection between Hellinger distance and KL-divergence. For communication model in Figure 2 which crucially introduces correlation between two inputs conditioned on the protocol, the usual cut-and-paste will not give the desired bound.

Instead, in Lemma 4.6, we show that if X and Y are under product distribution, one can actually relax this condition and make the argument robust to small correlation between inputs (even when Z is not a communication protocol). In particular Z can be an arbitrary random process!

Proof of Lemma 4.6. Towards the proof we make use of the following notation for distribution on X, Y conditioned on $Z = z$ assuming $Z_{ans} = 1$.

$$\begin{aligned} a_z &:= \Pr[X = 0, Y = 0 | Z = z] \\ b_z &:= \Pr[X = 0, Y = 1 | Z = z] \\ c_z &:= \Pr[X = 1, Y = 0 | Z = z] \end{aligned}$$

Note that $\Pr[X = 1, Y = 1 | Z = z] = 0$, since we have the guarantee that if $\mathsf{AND}(X, Y) = 1$ then $Z_{ans} = 0$. We prove the following claim on a_z, b_z, c_z.

	$Y = 0$	$Y = 1$
$X = 0$	a_z	b_z
$X = 1$	c_z	0

Figure 4: Resulting prior conditioned on $Z = z$ with $Z_{ans} = 1$

Claim 4.7. *If* $\mathsf{D}_{KL}(X_z||X) \leq o(\gamma)$, *and* $I(X;Y|Z = z) \leq o(\gamma^2)$, *then* $\mathsf{D}_{KL}(Y_z||Y) \geq \Omega(\gamma)$.

Proof. First, observe that $c_z = \Theta(\gamma)$. If $\mathsf{D}_{KL}(X_z||X) \leq o(\gamma)$, then from Fact 3.6 we have

$$\Pr[X = 1 | Z = z] = c_z = \Theta(\gamma). \tag{20}$$

Next, we expand and lower bound the term $I(X;Y|Z = z)$.

$$\begin{aligned} I(X;Y|Z = z) &= \Pr[Y = 0|Z = z] \cdot \mathsf{D}_{KL}(X_{Y=0,z}||X_z) \\ &+ \Pr[Y = 1|Z = z] \cdot \mathsf{D}_{KL}(X_{Y=1,z}||X_z) \\ &\geq (1 - b_z) \cdot \mathsf{D}_{KL}(\mathcal{B}_{\frac{c_z}{1-b_z}}||\mathcal{B}_{c_z}) + b_z \cdot \mathsf{D}_{KL}(\mathcal{B}_0||\mathcal{B}_{c_z}) \\ &\geq b_z \cdot \mathsf{D}_{KL}(\mathcal{B}_0||\mathcal{B}_{c_z}) = b_z \log \frac{1}{1 - c_z} \geq \Omega\left(b_z c_z\right). \end{aligned}$$

where the last bound holds from $-\log(1 - x) \geq x/2$ for $x < 1/2$. Rewriting the inequality we get,

$$b_z c_z \leq O\left(I(X;Y|Z = z)\right) \tag{21}$$

Now we have that $c_z = \Theta(\gamma)$ from (20). Therefore we can rewrite (21) as

$$b_z \leq O\left(\frac{I(X;Y|Z=z)}{\gamma}\right) \tag{22}$$

Since we assumed $I(X;Y|Z=z) \leq o(\gamma^2)$, we obtain

$$b_z \leq o(\gamma) \tag{23}$$

Then combining (23) with Fact 3.6, we get

$$\mathsf{D}_{KL}(Y_z||Y) \geq \Omega(\gamma).$$

□

To complete the proof of Lemma 4.6, take "good" z such that

- $z_{ans} = 1$
- $\mathsf{D}_{KL}(X_z||X) \leq o(\gamma)$
- $I(X;Y|Z=z) \leq o(\gamma^2)$

By union bound, the mass on such Z must be at least $\Omega(1)$. Furthermore, for these z, Claim 4.7 holds, and $\mathsf{D}_{KL}(Y_z||Y) \geq \Omega(\gamma)$. Then

$$I(Z;Y) = \mathbb{E}_{z\sim Z}\left[\mathsf{D}_{KL}(Y_z||Y)\right] \geq \Pr[z \text{ is good}] \cdot \Omega(\gamma) = \Omega(\gamma)$$

where the expectation sum is only taken over good z. □

4.4 Proof of the Main Theorem

The proof of Theorem 3.10 now follows easily by combining Lemma 4.5 and Lemma 4.6.

Theorem 4.8 (Restated)**.** *Let $\Gamma = (U, \Pi)$ be a 4-party NOF protocol (c.f. Figure 5) that solves $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ with $|T| < |U| < o(m)$. Then there exists some $i \in [m]$ such that $|\Pi_i| \geq \Omega(\sqrt{n})$.*

Proof. Suppose otherwise. Then for all $i \in [m]$, we have $|\Pi_i| < o(\sqrt{n})$. Then by Lemma 4.5, there is some random variable $Z^{\mathsf{AND}}(X,Y)$ for solving $\mathsf{AND}(X,Y)$ with the following guarantees by setting $C = o(\sqrt{n})$:

$$\begin{aligned} I(Z^{\mathsf{AND}};X) &< o(1/\sqrt{n}) \\ I(Z^{\mathsf{AND}};Y) &< o(1/\sqrt{n}) \\ I(X;Y|Z^{\mathsf{AND}}) &< o(1/n). \end{aligned}$$

Furthermore, $Z^{\mathsf{AND}}_{ans} = 0$ whenever $\mathsf{AND}(X,Y) = 1$. But $\Pr[\mathsf{AND}(X,Y)=1] = O(1/n)$. While if $\mathsf{AND}(X,Y) = 0$, $\Pr[Z^{\mathsf{AND}}_{ans} = 0] < 0.001$. Therefore, $\Pr[Z^{\mathsf{AND}}_{ans} = 1] \geq 1/2$. But this is in direct contradiction to Lemma 4.6 with $\gamma := \frac{1}{1000\sqrt{n}}$. Hence such Z^{AND} cannot exist.

□

4.5 Multiphase Lower Bound for Semi-Adaptive Data structures

Here we prove Theorem 1.4, asserting a polynomial lower bound on the Multiphase problem against Semi-Adaptive dynamic data structures (Definition 1.3):

Theorem 4.9 (Main Result)**.** *Let $m = \omega(n)$. Any semi-adaptive data structure solving the Multiphase problem must have (total) update time $t_u n \geq \Omega(\frac{m}{w})$ or query time $t_q \geq \Omega(\sqrt{n}/w)$, in the cell-probe model with word size w.*

Proof. We use a simple variation of the reduction from [Pat10] to show that an efficient semi-adaptive data strucutre implies a too-good-to-be-true 4-party NOF protocol, contradicting Theorem 4.8. To this end, suppose we have a semi-adaptive dynamic data structure with $t_u n < o\left(\frac{m}{w}\right)$ and $t_q < o(\sqrt{n}/w)$.

We argue that this implies a cheap 4-party NOF protocol. First we set the update transcript as U – which can be generated by Charlie who has access to both $\vec{S}$ and T. We then have $|U| = O(t_u n w) < o(k)$. Set $\Pi_i^{\mathcal{M}}$ as addresses and the contents of cells in $\mathcal{M}$ that are accessed by the query algorithm. Set subsequent Alice's message Π_i^{τ} as the cell address of $\Delta(\mathcal{M}, T)$ that are accessed. This is determined by $\Pi_i^{\mathcal{M}}, \Pi_i^{\leq \tau}, i, S_i$. Set Bob's message $\Pi_i^{\tau+1}$ as the cell content. This is determined by $\Pi_i^{\mathcal{M}}, \Pi_i^{\leq \tau}, UT, i$. It is immediate that this is a valid 4-party protocol. We also have the length of 4-party NOF protocol $\Gamma = (U, \Pi)$ as

$$|U| + |T| \leq o(m)$$
$$|\Pi_i| = |\Pi_i^{\mathcal{M}}| + |\Pi_i^{A \leftrightarrow B}| \leq 4 t_q w < o(\sqrt{n})$$

But this is in contradiction to Theorem 4.8. □

5 Lower Bounds on Nonlinear Networks for computing Linear Operators

Circuits with arbitrary gates As mentioned in Section 1, a long-standing open problem in circuit complexity is whether *non-linear* gates can significantly (polynomially) reduce the number of wires of circuits computing linear operators [JS10]. We consider Valiant's depth-2 circuit model [Val77] with arbitrary gates, and its generalizations to arbitrary depths. More formally, consider a circuit computing a linear operator $x \mapsto Ax$ where A is an $m \times n$ matrix with $m = \text{poly}(n)$, using unbounded fan-in, and where gates are allowed to be *arbitrary* functions. Clearly, such circuits can trivially compute any f with m gates. As such, the interesting complexity measure in this model is the minimum number of **wires** (W) to computing the function f. This measure captures how much "information" needs to be transferred between different components of the circuit, in order to compute the function. For a more thorough exposition and motivation on circuits with arbitrary gates, we refer the reader to [Juk12].

Previous Works In contrast to arithmetic circuit models (e.g., [Val77] where allowed functions are simple functions such as AND, OR or PARITY), it is a long-standing open problem [JS10, Juk12, Dru12] whether non-linear circuits can compute *any* linear operator A with near-linear ($\tilde{O}(m)$) number of wires. Indeed, for linear circuits, this is a simple counting argument. Counting argument shows that $\Omega(mn/\log(mn))$ wires are necessary for linear circuits, and this is tight

[6] This is for depth 2 circuit. For any depth, [Dru12] gives $\Omega(m \log n)$ lower bound.

[7] This is for depth 2 circuit, square matrix

Circuit Model	Non-Explicit Matrix	Explicit Matrix
Linear ($\oplus$) Gates	$\Omega(mn/\log n)$ [Lup56]	$\Omega\left(n\log^{3/2} n\right)$ [Fri93, PR94, SSS97]
Arbitrary Gates	$\Omega(n(\log n/\log\log n)^2)$ [GHK+12][6]	$\Omega(n\log n/\log\log n)$ [Juk10][7]
Width-k DNFs	$\Omega\left(mn^{\frac{1}{2(d+k)}}\right)$ **[This Work]**	Same as above [Juk10]

Table 1: Wire lower bounds for computing $m \times n$ linear operators $x \mapsto Ax$ in various circuit models.

[Lup56]. But once again, for arbitrary circuits, counting argument completely fails. The number of possible functions over n bits is already doubly exponential in n. In fact, counting argument fails even when we consider only width-3 DNFs. While there are only 2^{mn} different linear operators, there are at least 2^{n^3} different width-3 DNF gates.

[Juk10] initiated works on analyzing the complexity of *representing* a random matrix, that is computing Ax when x is restricted to having only one 1. In other words, compute Ae_i for $i \in [n]$. [Dru12] showed that when restricted to *representing* a matrix, $\Omega(m \log m)$ is necessary for depth 2 circuit, complementing the previous upper bound of $O(m \log m)$ by [Juk10].

The lower bound of [Dru12] immediately implies $\Omega(m \log m)$ lower bound for computing a matrix using depth 2 circuit, since any circuit that computes a matrix must represent a matrix as well. But no better bounds were known in case of **computing** the linear operator A. Table 1 summarizes known results on wire lower bounds for circuits computing linear operator.

Our Result First, we show that our communication lower bound (i.e. Theorem 3.10) implies a trade-off between degree of gates, depth and number of wires required to compute a random linear operator (over the Boolean semi-ring) using a variant of a reduction in [Vio18]. To the best of our knowledge, this is the first polynomial lower bound on the number of wires for any non-linear circuit model.

We then show that Conjecture 5.4 implies a polynomial lower bound on nonadaptive *static* data structures against a "semi-explicit" static problem—computing set-disjointness queries w.r.t n^2 *random* sets. Using the reduction of [Vio18] once again, we show that this static data structure lower bound implies a polynomial wire lower bound on depth-d circuits with *arbitrary* gates ($k = n$) for computing random linear operators. This is the content of the next two subsections.

5.1 Width-k DNF Lower Bound

First we show that if there exists a circuit with width-k DNF gates and small number of wires, then there exists a good 4-party communication protocol.

Lemma 5.1. *If there exists a depth-d circuit with width-k DNF gates with W-wires for computing Ax where $A \in \{0,1\}^{m\times n}$ with $m = \text{poly}(n)$. Then there exists a 4-party communication protocol for computing $A_i x$ (for any given $i \in [m]$) with $|U| \le o(m)$, $|\Pi_i^{\mathcal{M}}| \le O\left((W/m)^{k+d}\right)$ and $|\Pi_i| \le O\left((W/m)^d \log n\right)$*

Proof. Suppose we have a depth-d circuit with width-k DNF gates with W wires. Then we argue that this induces an efficient 4-party communication protocol.

First set Megan's input as the linear operator A and index in question i, Bob's input as x and i, Merlin's input as A and x, Alice's input as A_i and i.

Merlin and Megan's message We set Merlin and Megan's message in the following manner. Consider the set of gates G with fan-in $\omega(W/m)$. Since there are total of W wires, we know that $|G| < o(m)$. Also note that G has no dependence on i.

Set Megan's message $\Pi_i^{\mathcal{M}}$ as the description of circuit computing $A_i x$, without gates in G. Note that $\Pi_i^{\mathcal{M}}$ contains at most $O((W/m)^d)$ gates since each gate not in G has fan-in at most $O(W/m)$. Furthermore, description of each gate requires at most $O\left((W/m)^k + (W/m)\log(m + (W/m)^d)\right)$ bits, $O\left((W/m)^k\right)$ bits for describing the function and $O\left((W/m)\log(m + (W/m)^d)\right)$ bits for describing the inputs.

Therefore with $m = \text{poly}(n)$, we get

$$|\Pi_i^{\mathcal{M}}| \leq O((W/m)^{k+d} + (W/m)^{d+1}\log(m + (W/m)^d)) = O((W/m)^{k+d}). \tag{24}$$

Furthermore we set Merlin's message U as the value of gates in G. Therefore we have $|U| \leq o(m)$.

Alice and Bob's message Alice queries Bob the gate values of x_i's and G required for computing $A_i x$ as given by $\Pi_i^{\mathcal{M}}$. Bob can answer Alice's query since Bob knows x and gate value of G from U. Since fan-in is bounded as $O(W/m)$ and it is a depth-d circuit, we know there are at most $O((W/m)^d)$ gate values required for computing $A_i x$. Therefore we have

$$|\Pi_i^{A\leftrightarrow B}| \leq O((W/m)^d \log(n + |G|)) = O((W/m)^d \log n) \tag{25}$$

With Bob's message, Alice can compute $A_i x$ by using the circuit from $\Pi_i^{\mathcal{M}}$ with Bob's response as the input. □

Now our 4-party communication lower bound (Theorem 3.10) yields the following circuit lower bound via the reduction given in Lemma 5.1.

Theorem 5.2. *There exists a linear operator $A \in \{0,1\}^{m\times n}$ such that any depth-d circuit with width-k DNF gates computing Ax must have wire $W \geq \Omega\left(m \cdot n^{\frac{1}{2(d+k)}}\right)$.*

Proof. Suppose for all A, there exists a circuit for A with wire $W \leq o(m \cdot n^{\frac{1}{2(d+k)}})$. Then this yields a 4-party communication for $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ with $|\Pi_i^{\mathcal{M}}| < o(\sqrt{n})$, $|U| < o(m)$ and $|\Pi_i^{A\leftrightarrow B}| < o(\sqrt{n})$ from Lemma 5.1. But this is a contradiction to Theorem 4.8. □

Remark 5.3. *We remark that it is possible to obtain a weaker $\Omega(m+n^{1+1/(k+d)})$ lower bound more directly* without *using our communication lower bound Theorem 4.8. The proof was discovered in personal communication with Swastik Kopparty and Sepehr Assadi, and is based on a certain random-restriction argument for eliminating high fan-in gates, combined with the error-correcting properties of random linear operators. This argument, however, can only eliminate $o(n)$ gates (since every elimination shrinks the remaining input space $\{0,1\}^n$ by half), and thus quickly becomes trivial when the number of outputs is $m > n^{1.1}$ (say). By contrast, our lower bound can eliminate $o(m)$ gates, which allows us to prove a polynomial lower bound in the* number of outputs *so long as $m = \text{poly}(n)$.*

5.2 NOF conjecture implications

In this section, we show that Pătraşcu's NOF Conjecture on the original Multiphase Game, even against 3-round protocols, would imply a breakthrough in circuit complexity. This complements our restricted NOF model, as it shows that allowing even *two* of Alice's messages to depend arbitrarily on her entire input $\vec{S}, i$, would resolve a decades-old open problem in circuit lower bounds. It also

suggests that attacking the Multiphase conjecture for (general) dynamic data structures via the the NOF Game, should exploit the fact that data structures induce highly restricted NOF protocols.

First, note that Conjecture 1.1 in particular implies the following special case:

Conjecture 5.4 (3-round NOF Conjecture)**.** *Any 3-round NOF protocol for the 3-party Multiphase Game with $|U| = o(m)$ bits of advice must have $|\Pi| > n^{\varepsilon}$ communication for some constant $\varepsilon > 0$.*

Static Data Structure Lower Bound First, we consider the following class of static data structure problems $\mathcal{P}_A^f(x)$ defined by a query matrix $A \in \{0,1\}^{m \times n}$ and a function $f : \{0,1\}^{2n} \to \{0,1\}$:

1. Given a *fixed* matrix A with rows $A_1, \ldots, A_m$, preprocess an input database $x \in \{0,1\}^n$.
2. Given $i \in [m]$ as a query, the data structure needs to output $f(A_i, x)$.

Note that A is *hard-wired* to the problem, i.e., the data structure can access A for free during both preprocessing and query stage[8]. In particular with word size w, with $s = m/w$ cells, one can store the (boolean) answers for all possible queries and the problem becomes trivial ($t = 1$), whereas without any preprocessing, the query algorithm needs to read x (but not A) to compute the answer $f(A_i, x)$, giving (worst case) query time $t \sim n/w$. Accordingly, the query algorithm is *non-adaptive* if the cell addresses that are probed only a function of $i \in [m]$ and A.

We show that Conjecture 5.4 implies the following lower bound on $\mathcal{P}_A^f(x)$ where $f := \mathsf{DISJ}_n$.

Lemma 5.5 (Polynomial Static Lower Bound for Random Set Disjointness Queries)**.** *Suppose Conjecture 5.4 holds. Let $m = \omega(n)$. Then there exists a collection of m sets $A := A_1, \ldots, A_m \subseteq [n]^m$ such that any non-adaptive static data structure solving $\mathcal{P}_A^{\mathsf{DISJ}_n}$ must either use $s \geq \Omega(\frac{m}{w})$ space, or have $t \geq \Omega(n^{\varepsilon}/w)$ query time, in the cell-probe model with word size w.*

Proof. Suppose for any $A \in \{0,1\}^{m \times n}$ there is a non-adaptive static data structure D_A computing $\mathcal{P}_A^{\mathsf{DISJ}_n}$ with $s \leq o(\frac{m}{w})$ space and $t \leq o(n^{\varepsilon}/w)$ cell probes. We show that this induces a too-good-to-be-true (3-round) NOF protocol for the Multiphase game $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$, violating Conjecture 5.4. Indeed, consider the following simple 3-party protocol for simulating D_A:

Charlie's "advice" U in Phase 1 of the Multiphase game will be the contents of the s memory cells of D_A, Alice's message during Phase 2 (i.e., $\Pi_i^{A \to B}$) will be the memory addresses probed by the D_A for answering $\mathsf{DISJ}(A_i, x)$, and Bob's messages $\Pi_i^{B \to A}$ are the contents of cells probed by Alice. Note this protocol is well defined: Indeed, by the definition of $\mathcal{P}_A^{\mathsf{DISJ}_n}$, U only depends on A and x at preprocessing time, and if D_A is non-adaptive, then $\Pi_i^{A \to B}$ is only a function of A and i; Finally, $\Pi_i^{B \to A}$ depends on the previous transcript and $U = U(A, x)$ which Bob possesses. We therefore have a valid 3-round NOF protocol for $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$ with

$$|U| + |x| \leq sw + n \leq o(m) + n = o(m), \quad \text{and}$$
$$|\Pi_i| = |\Pi_i^{A \to B}| + |\Pi_i^{B \to A}| \leq 2tw \leq o(n^{\varepsilon})$$

bits, which contradicts Conjecture 5.4. □

We consider the following parameter for the circuit lower bound.

Corollary 5.6. *If $m = \omega(n)$, and $s = o(m/w)$ then $t \geq \Omega(n^{\varepsilon}/w)$.*

[8]This is a generalization of Valiant's model [Val77] in that the circuit itself is allowed to depend arbitrarily on A.

Circuit Lower Bound Now we show that Conjecture 5.4 implies circuit lower bounds using reduction by [Vio18] from Lemma 5.5. We use the following translation theorem for lower bounds in arbitrary depths.

Theorem 5.7 ([Vio18])**.** *Suppose function $f : \{0,1\}^n \to \{0,1\}^m$ has a circuit of depth d with W wires, consisting of unbounded fan-in, arbitrary gates. Then for any r there exists a static data structure (with non-adaptive query) with space $s = n + r$, query time $(W/r)^d$, and word size $\max\{\log n, \log r\} + 1$ which solves the following problem*

1. *Preprocess input x depending on x and f*
2. *Given $i \in [m]$, output $f_i(x)$.*

For completeness, we attach the proof here. Though [Vio18] did not remark on queries being non-adaptive, we remark that data structures derived from circuits are intrinsically non-adaptive, i.e. the memory cells probed only depends on query itself and f (but not on the content of cells probed along the way).

Proof of Theorem 5.7. Consider circuit C_f which computes f. Now set G as the set of gates with fan-in $> W/r$. Since the number of wires are bounded by w, $|G| < r$. Now given x, store the values of these gates G in space r.

Now we argue inductively on the level of the gate. We show that if the gate is at level ℓ, that the number of non-adaptive queries made to compute the value of the gate is at most $(W/r)^\ell$. As base case, suppose if it were a level 1 gate. If it lies in G, then the non-adaptive query required is 1, by calling to its address in G, which requires $\log |G|$-bits. Otherwise, the query required is at most W/r, each of which requires $\log n$-bits for the address, and they are non-adaptive. Therefore, the word size required is $\max\{\log n, \log r\} + 1$.

Now as induction step, suppose for any $j < \ell$, level j gates require at most $(W/r)^j$ non-adaptive queries to compute its value with word size $\max\{\log n, \log r\}+1$. Consider a level ℓ gate. If the gate lies in G, again it only requires 1 non-adaptive query, by calling to its address in G. Otherwise, it can be answered by computing at most W/r level $\ell - 1$ gates, and these queries are non-adaptive. Each of these level $\ell - 1$ gates require $(W/r)^{\ell-1}$ non-adaptive queries by induction hypothesis, and the word size required is $\max\{\log n, \log r\} + 1$. Therefore, the total non-adaptive query required is at most $(W/r) \cdot (W/r)^{\ell-1} = (W/r)^\ell$.

Since the final output gate that we are interested in is at level d, we get $(W/r)^d$ as the upper bound on the number of queries. □

Here we crucially used the fact that wirings are fixed if we fix C_f. Note that G is determined by C_f. Then a contrapositive of Theorem 5.7 then states that data structure lower bound with non-adaptive query implies circuit lower bounds. When restricted to $o(m)$ additional space usage (i.e. $r = o(m)$, $f(n)$ query time lower bound translates to $\Omega\left(mf(n)^{1/d}\right)$ lower bound for W. Using the contrapositive along with Corollary 5.6, we show the following circuit lower bound.

Corollary 5.8. *Assuming Conjecture 5.4 and $m = \omega(n)$, there exists a matrix $A \in \{0,1\}^{m \times n}$ such that any depth-d circuit that computes[9] Ax requires $m \cdot n^{\Omega(\varepsilon/d)}$ wirings. In particular, if $d = 2$, there exists A that requires $m \cdot n^{\frac{\varepsilon}{2} - o(1)}$ wirings.*

Proof. We use contrapositive of Theorem 5.7, that is data structure lower bound implies circuit lower bounds. Consider Corollary 5.6. Setting $r = o(m)$, and $m = \omega(n)$. Then Corollary 5.6 with Theorem 5.7 implies that for some A, $(W/r)^d \geq \Omega(n^\varepsilon / \log n)$. Since $r = o(m)$, rewriting in terms of W, we get

$$W \geq \Omega\left(m \cdot n^{\frac{\varepsilon}{d} - \frac{\log\log n}{d \log n}}\right) = m \cdot n^{\Omega\left(\frac{\varepsilon}{d}\right)}.$$

[9] Over the Boolean semi-ring, with addition as OR and multiplication as AND

Now setting $d = 2$, we get $W \geq m \cdot n^{\frac{\varepsilon}{2}-o(1)}$.

□

6 Discussion

Extending our techniques to fully adaptive queries It is natural to ask whether our technical approach can be extended to fully adaptive queries as well, so as to resolve the Multiphase Conjecture for general data strucutres. Our results suggest that any attack on this problem via NOF communication should exploit the fact that data structures induce *restricted* NOF protocols for the Multiphase Game – namely, that the query algorithm only has limited *local* access to its memory $\mathcal{M}(\vec{S}, T)$, through previously probed cells (unlike Alice in the NOF game, who has general access to $\vec{S}$).

Let Π_D be a NOF protocol induced by an efficient dynamic data structure D for $\mathsf{SEL}^m_{\mathsf{DISJ}_n}$. Recall that the main technical challenge in our proof is to use Π to design a random variable $Z(X, Y)$ computing $\mathsf{AND}(X, Y)$, while *simultaneously* controlling $I(Z;Y)+I(Z;X)$ and $I(X;Y|Z)$. It is not clear to us whether one could hope for such $Z = Z(\Pi_D)$ when D is a *fully adaptive* data structure, hence we believe the following question is interesting:

Is it possible to design a r.v. Z where $I(Z;Y), I(Z;X)$ and $I(X;Y|Z)$ are all small?

Acknowledgement

We are grateful to Huacheng Yu for his valuable comments and insightful discussions on this project, and to Swastik Kopparty and Sepehr Assadi for the discussion on Theorem 5.2.

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010*, pages 67–76, 2010.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, 1986.

[BK02] Jérémy Barbay and Claire Kenyon. Adaptive intersection and t-threshold problems. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '02, pages 390–399, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics.

[BL15] Joshua Brody and Kasper Green Larsen. Adapt or die: Polynomial lower bounds for non-adaptive dynamic data structures. *Theory Comput.*, 11:471–489, 2015.

[BPP07] Philip Bille, Anna Pagh, and Rasmus Pagh. Fast evaluation of union-intersection expressions. In Takeshi Tokuyama, editor, *Algorithms and Computation*, pages 739–750, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*. IEEE, oct 2011.

[Bra15] M. Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015.

[BY04] Ricardo Baeza-Yates. A fast set intersection algorithm for sorted sequences. In Suleyman Cenk Sahinalp, S. Muthukrishnan, and Ugur Dogrusoz, editors, *Combinatorial Pattern Matching*, pages 400–408, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[BYJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 209–218, Washington, DC, USA, 2002. IEEE Computer Society.

[CEEP12] Arkadev Chattopadhyay, Jeff Edmonds, Faith Ellen, and Toniann Pitassi. A little advice can be very helpful. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete algorithms*, pages 615–625. Society for Industrial and Applied Mathematics, 2012.

[CGL15] Raphaël Clifford, Allan Grønlund, and Kasper Green Larsen. New unconditional hardness results for dynamic and online problems. In *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, FOCS '15, pages 1089–1107, Washington, DC, USA, 2015. IEEE Computer Society.

[CP10] Hagai Cohen and Ely Porat. Fast set intersection and two-patterns matching. *Theor. Comput. Sci.*, 411(40-42):3795–3800, 2010.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, New York, NY, USA, 2006.

[DLOM00] Erik D. Demaine, Alejandro López-Ortiz, and J. Ian Munro. Adaptive set intersections, unions, and differences. In *In Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA*, pages 743–752, 2000.

[Dru12] Andrew Drucker. Limitations of lower-bound methods for the wire complexity of boolean operators. In *Proceedings of the 2012 IEEE Conference on Computational Complexity (CCC)*, CCC '12, pages 170–180, Washington, DC, USA, 2012. IEEE Computer Society.

[Fri93] Joel Friedman. A note on matrix rigidity. *Combinatorica*, 13(2):235–239, 1993.

[GHK+12] Anna Gál, Kristoffer Arnsfelt Hansen, Michal Koucký, Pavel Pudlák, and Emanuele Viola. Tight bounds on computing error-correcting codes by bounded-depth circuits with arbitrary gates. In *Proceedings of the 44th symposium on Theory of Computing - STOC'12*. ACM Press, 2012.

[GHK+13] Anna Gál, Kristoffer Arnsfelt Hansen, Michal Koucký, Pavel Pudlák, and Emanuele Viola. Tight bounds on computing error-correcting codes by bounded-depth circuits with arbitrary gates. *IEEE Trans. Information Theory*, 59(10):6611–6627, 2013.

[HKNS15] Monika Henzinger, Sebastian Krinninger, Danupon Nanongkai, and Thatchaphol Saranurak. Unifying and strengthening hardness for dynamic problems via the online matrix-vector multiplication conjecture. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing - STOC 15*. ACM Press, 2015.

[HW07] Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(1):211–219, 2007.

[JS10] S. Jukna and G. Schnitger. Circuits with arbitrary gates for random operators. 2010.

[Juk10] Stasys Jukna. Representing (0, 1)-matrices by boolean circuits. *Discrete Mathematics*, 310(1):184 – 187, 2010.

[Juk12] Stasys Jukna. Boolean function complexity. *Algorithms and Combinatorics*, 2012.

[KPP15] Tsvi Kopelowitz, Seth Pettie, and Ely Porat. Dynamic set intersection. In *Lecture Notes in Computer Science*, pages 470–481. Springer International Publishing, 2015.

[Lar12] Kasper Green Larsen. Higher cell probe lower bounds for evaluating polynomials. In *FOCS 2012*, pages 293–301, 2012.

[Lup56] Oleg Borisovich Lupanov. On rectifier and switching-and-rectifier schemes. *Dokl. Akad. Nauk SSSR*, pages 111:1171–1174, 1956.

[LWY18] Kasper Green Larsen, Omri Weinstein, and Huacheng Yu. Crossing the logarithmic barrier for dynamic boolean data structure lower bounds. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 978–989, 2018.

[MPP05] Christian Worm Mortensen, Rasmus Pagh, and Mihai Patrascu. On dynamic range reporting in one dimension. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 104–111, 2005.

[Pat10] Mihai Patrascu. Towards polynomial lower bounds for dynamic problems. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 603–610, 2010.

[PR94] Pavel Pudlák and Vojtech Rödl. Some combinatorial-algebraic problems from complexity theory. *Discrete Math.*, 136(1-3):253–279, 1994.

[SSS97] Mohammad Amin Shokrollahi, Daniel A. Spielman, and Volker Stemann. A remark on matrix rigidity. *Inf. Process. Lett.*, 64(6):283–285, 1997.

[Val77] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In Jozef Gruska, editor, *Mathematical Foundations of Computer Science 1977*, pages 162–176, Berlin, Heidelberg, 1977. Springer Berlin Heidelberg.

[Vio18] Emanuele Viola. Lower bounds for data structures with space close to maximum imply circuit lower bounds. In *ECCC*, volume 25, 2018.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

A Tight Bounds from [CEEP12]

A.1 Upper Bound

First, we show that there is a non-trivial $o(n)$-query time semi-adaptive data structure for $\mathsf{SEL}^k_{\mathsf{DISJ}_n}$, derived from a protocol of [CEEP12] given that the querier has access to i and S_i from the beginning.

Theorem A.1. *There exists $t_q = \tilde{O}(\sqrt{n})$ semi-adaptive data structure with $|U| \leq \tilde{O}(\sqrt{n})$.*

Proof. Consider $U(\vec{S}, T)$ constructed by Chattopadhyay et al. [CEEP12]. They construct $U(\vec{S}, T)$ with $|U| \leq \tilde{O}(\sqrt{n})$ independent of i that has the following nice property: If Alice learns U, Alice (just with i and S_i) can decode that either $|S_i \cap T| > 1$, therefore $\mathsf{DISJ}_n(S_i, T) = 0$; or learns about potential set of elements in $P \subset (S_i \cap T)$ with $|P| \leq \sqrt{n}$. Alice can transmit this set P using $\sqrt{n} \log n$ bits to Bob, then Bob can announce the answer as $\mathsf{DISJ}_n(S_i, T) = \mathsf{DISJ}_n(P, T)$.

We can translate their protocol to data structure if the querier has access to both i and S_i in the beginning.

1. During the update phase, the data structure writes U and T separately.
2. Querier (i, S_i) reads all cells with U using $|U|/w$ queries, then either learns that $\mathsf{DISJ}_n(S_i, T) = 0$ or learns about the candidate set $P \subset S_i$.
3. Querier checks $\mathsf{DISJ}_n(P, T)$ by accessing corresponding entries for T.

Since $|P| \leq \sqrt{n}$, the total number of queries made is $|P| + (|U|/w) = \sqrt{n} + (|U|/w) = \tilde{O}(\sqrt{n})$. Furthermore, the query is semi-adaptive. Querier *only* accesses updated cells. The total number of alternation is 1. □

A.2 Lower Bound

In this section, we showed that a modified 4-party model subsumes 1.5-round protocol. Recall that a "1.5-round protocol" [CEEP12] for $\mathsf{SEL}^k_{\mathsf{DISJ}_n}$ proceeds in the following way:

1. Charlie sends message $U(\vec{S}, T)$ to Bob privately.
2. Bob *forwards* a message $U'(\vec{S}, T) \subseteq U$ to Alice (hence this message is *independent* of i).
3. Alice sends message $\Pi_i^{A \to B}$ to Bob, which is dependent on $U', \vec{S}$ and i.
4. Bob solves $\mathsf{DISJ}_n(S_i, T)$ from U, T, i and $\Pi_i^{A \to B}$.

Then [CEEP12] (Theorem 1.2) show a 1.5-protocol for $\mathsf{SEL}^k_{\mathsf{DISJ}_n}$ with $\tilde{O}(\sqrt{n})$ overall communication. They then complement this upper bound with a lower bound when restricted to a 1.5-protocol. In general, this model is incomparable to our 4-party model. However, a 1.5-round protocol can be simulated by a modified 4-party protocol where Charlie is allowed to send Megan a message U' prior to Megan sending messages to Alice and Bob. Formally the 4-party protocol proceeds in following manner

1. Charlie sends message $U(\vec{S}, T)$ to Bob privately.
2. Charlie sends message $U'(\vec{S}, T)$ to Megan.
3. Megan broadcasts $\Pi_i^{\mathcal{M}}(\vec{S}, i, U')$ to Alice and Bob
4. Alice and Bob communicates and compute $\mathsf{DISJ}_n(S_i, T)$.

Now Megan's message is allowed to depend on i, $\vec{S}$ and U'.

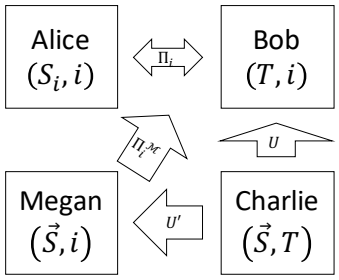

Figure 5: Modified 4-party NOF Communication

Claim A.2. *A 1.5-round protocol where $|U'| < C$ and $|\Pi_i^{A\to B}| < C$ can be simulated by modified 4-party NOF communication with $|\Pi_i^{\mathcal{M}}| < 2C$.*

Proof. Under this model, 4-party can then simulate 1.5-protocol via Megan simulating Alice in 1.5-protocol by sending $\Pi_i^{A\to B}$ and U' as her message $\Pi_i^{\mathcal{M}}$. Bob can then decode $\mathsf{DISJ}_n(S_i, T)$ from Megan's message, with no communication between Alice and Bob. □

Now the question is whether the same lower bound for modified 4-party NOF holds as well. Below we argue that our lower bound (Theorem 1.2) in fact applies to 4-party protocols with Charlie sending a short message to Megan hence also to 1.5-protocols, establishing that our lower bound subsumes [CEEP12] lower bound.

Theorem A.3. *Any modified 4-party NOF protocol $\Gamma = (U, \Pi_i)$ with $\Pi_i = (U', \Pi_i^{\mathcal{M}}, \Pi_i^{A\leftrightarrow B})$ that solves $\mathsf{SEL}^k_{\mathsf{DISJ}_n}$ with $|U| < o(k)$ require $|\Pi_i| > \Omega(\sqrt{n})$.*

Proof Sketch. This follows from the observation that in such 4-party protocols, Lemma 4.1 still holds (up to factor 2) and rest of the proof remains unchanged. To see why, observe that since U' from step 2 does not depend on the index i, Equation (12) in the proof of Theorem 1.2 can be bounded instead by the following inequality:

$$\begin{aligned} &I(T; S_{i_{<\ell}}, \Pi_{i_{<\ell}}^{\mathcal{M}}) \le I(T; \vec{S}, \Pi_{i_{<\ell}}^{\mathcal{M}}, U') \\ &= \underbrace{I(T; \vec{S})}_{=0} + \underbrace{I(T; U'|\vec{S})}_{|U'|} + \underbrace{I(T; \Pi_{i_{<\ell}}^{\mathcal{M}}|\vec{S}, U')}_{=0} \le |U'|, \end{aligned}$$

where $I(T; \Pi_{i_{<\ell}}^{\mathcal{M}}|\vec{S}, U') = 0$ since $U', \vec{S}$ and the index i determine $\Pi_i^{\mathcal{M}}$.

Equation (12) is the only step where the proof used the assumption that Megan speaks first. Now in a 4-party protocol, if we assume further that $|U'| < C$, then by (12), instead of (4) in Lemma 4.1, we get the same bound up to factor 2, i.e., $I(Z^{DISJ}; T) < 2C$. Then rest of the proof remains unchanged, hence it shows a $C \ge \Omega(\sqrt{n})$ lower bound for 4-party protocols. □

As such, the $\tilde{O}(\sqrt{n})$ 1.5-protocol of [CEEP12] shows that our Theorem 1.2 is in fact tight up to logarithmic factors.