Transaction Fee Mechanism Design for Leaderless Blockchain Protocols

Pranav Garimidi[1], Lioba Heimbach[2⋆], and Tim Roughgarden[1,3⋆⋆]

[1] a16z crypto pgarimidi@a16z.com
[2] ETH Zurich hlioba@ethz.ch
[3] Columbia University tim.roughgarden@gmail.com

Abstract. We initiate the study of transaction fee mechanism design for blockchain protocols in which multiple block producers contribute to the production of each block. Our contributions include:

- We propose an extensive-form (multi-stage) game model to reason about the game theory of multi-proposer transaction fee mechanisms.
- We define the *strongly BPIC* property to capture the idea that all block producers should be motivated to behave as intended: for every user bid profile, following the intended allocation rule is a Nash equilibrium for block producers that Pareto dominates all other Nash equilibria.
- We propose the *first-price auction with equal sharing (FPA-EQ)* mechanism as an attractive solution to the multi-proposer transaction fee mechanism design problem. We prove that the mechanism is strongly BPIC and guarantees at least a 63.2% fraction of the maximum-possible expected welfare at equilibrium.
- We prove that the compromises made by the FPA-EQ mechanism are qualitatively necessary: no strongly BPIC mechanism with non-trivial welfare guarantees can be DSIC, and no strongly BPIC mechanism can guarantee optimal welfare at equilibrium.

1 Introduction

1.1 Transaction Fee Mechanisms

A *transaction fee mechanism* is the component of a blockchain protocol responsible for deciding which pending transactions should be included for processing, and what the creators of those transactions should pay for the privilege of execution in the blockchain's virtual machine. For example, the Bitcoin protocol [25] launched with a first-price auction as its transaction fee mechanism (which remains in use to this day): users submit bids along with their transactions; should a transaction be included in a block, its bid is transferred from the user to the

⋆ Work performed in part during an internship at a16z crypto.
⋆⋆ Author's research at Columbia University supported in part by NSF awards CCF-2006737 and CNS-2212745.

producer of that block. Block producers are then expected to assemble blocks that maximize their revenue (i.e., the sum of the bids of the included transactions) subject to a block size constraint. The Ethereum protocol also launched with a first-price auction as its transaction fee mechanism [41] but, in order to achieve stronger incentive-compatibility guarantees, the protocol's first-price auction was swapped out in August 2021 in favor of a more sophisticated transaction fee mechanism known as EIP-1559 [6]. Since the initial economic analysis of EIP-1559 [28], a large body of research has been developed to explore the design space of transaction fee mechanisms and to assess different designs through the lenses of incentive-compatibility (both for users and for block producers), collusion-resistance, welfare, revenue, and more; see Section 1.4 for an overview.

The entire literature on transaction fee mechanisms considers only *leader-based* blockchain protocols in which each block is assembled unilaterally by a single block producer (like a Bitcoin miner or an Ethereum validator) with monopoly power over the contents of its block. This focus reflects the fact that the vast majority of the major blockchain protocols deployed to-date are leader-based in this sense. For example, all longest-chain protocols in the spirit of Bitcoin and PBFT-type protocols in the vein of Tendermint [5] are leader-based. But the state-of-the-art in consensus protocol design is evolving, and the design of transaction fee mechanisms must evolve with them.

1.2 Leaderless Blockchain Protocols

A new generation of consensus protocols, known as *DAG-based consensus*, is exploring *leaderless* protocol designs (where "DAG" stands for "directed acyclic graph"). In DAG-based consensus protocols, multiple validators build and propose blocks concurrently. Together, the validators build a DAG: whenever a block is proposed by a validator, the block references blocks from previous rounds, effectively voting on these referenced blocks. In each round, some of the blocks (sometimes referred to as anchor blocks) are used as checkpoints in the DAG structure for consensus. When an anchor block is finalized, transactions from all blocks in its causal history that have not been executed previously are deterministically ordered and staged for execution.

Recently, DAG-based consensus protocols have experienced a rise in the blockchain ecosystem, with Sui running Mysticeti in production [36] and other projects such as Aptos planning to transition to DAG-based consensus. The main reason for the rise in popularity of DAG-based consensus protocols is the significant throughput improvements they achieve in comparison to single-leader BFT consensus protocols [4,17,9,34,35]. These throughput improvements stem primarily from two design choices: (1) the separation of the communication and consensus layers, and (2) the use of simultaneous block proposals by all validators to overcome the bottlenecks that arise with the single-leader approach (in effect, spreading what had been a concentrated workload for the leader across all validators). Further, while DAG-based protocols initially suffered from increased latency, current protocols achieve almost optimal latency (up to one extra communication round) [2,1]. Finally, DAG-based protocols have the advantage that

they generally recover quickly from crash failures of leaders given that they have backup leaders in place [2].

1.3 Our Contributions

This paper initiates the study of transaction fee mechanism design for blockchain protocols in which multiple block producers contribute to the production of each block. To reason about such mechanisms, several new modeling and design challenges must be addressed:

- Transaction fee mechanism design with a single block producer can focus on equilibria purely from the perspective of users, with the block producer best responding to the resulting bids; with multiple block producers, the "game within the game," meaning the interaction between the incentives of different block producers, must be explicitly modeled and analyzed.
- The design of a transaction fee mechanism must now specify how proposals from multiple block proposers are aggregated into a single block of confirmed transactions.
- The design of a transaction fee mechanism must now specify how any unburned fee revenue from users is distributed between the different block proposers.

This paper offers the following contributions:

- We formally model the game theory of multi-proposer transaction fee mechanisms via extensive-form (multi-stage) games. Further, we define incentive-compatibility for block producers in a multi-proposer transaction fee mechanism, focusing on a condition we call *strongly BPIC*. Intuitively, a transaction fee mechanism is strongly BPIC if, no matter what the user bids, following the intended allocation rule is a Nash equilibrium for block producers that Pareto dominates all other Nash equilibria. While we use our model specifically to study the welfare guarantees achievable by multi-proposer transaction fee mechanisms, it should serve as the appropriate starting point to study a number of other potential benefits of such mechanisms (see Section 1.4 below for examples).
- We propose the *first-price auction with equal sharing (FPA-EQ)* mechanism as an attractive solution to the multi-proposer transaction fee mechanism design problem.[4] We prove that the mechanism is strongly BPIC and guarantees near-optimal welfare. Precisely, for every joint distribution over (possibly correlated) user valuations, for every subgame perfect equilibrium of the mechanism in which block producers play only Pareto-dominant Nash

[4] In Sui's current transaction fee mechanism, users pay their bids, and fee revenue is shared with validators pro rata, proportional to validator stake weights (Alberto Sonnino, personal communication, October 2024). The FPA-EQ mechanism can be viewed as the refinement of this mechanism in which transaction fees are shared with validators in proportion to the number of blocks they have contributed to.

equilibria, the expected equilibrium welfare is at least $1 - \frac{1}{e} \approx 63.2\%$ of the maximum possible. Our analysis here brings, for the first time, the powerful toolbox on "price of anarchy" bounds to bear on the analysis of transaction fee mechanisms. A simple example shows that the bound of 63.2% is tight in the worst case.
- We prove that the compromises made by the FPA-EQ mechanism are qualitatively necessary: no strongly BPIC mechanism with non-trivial welfare guarantees can be DSIC (i.e., with truthful bidding a dominant strategy for users), and no strongly BPIC mechanism can guarantee optimal welfare at equilibrium.

1.4 Related Work

General TFM literature. There is a long line of work studying transaction fee mechanisms for single-leader protocols, particularly focusing on Ethereum and Bitcoin. Our model of transaction fee mechanism design closely follows the line of work initiated by Roughgarden [29] to analyze the EIP-1559 mechanism [6]. Before this, research on Bitcoin's fee market focused on monopolistic pricing mechanisms [20,44]. More recent work in this area includes [26] and [13]. Building off Roughgarden's model, Chung and Shi [8] show that achieving an ideal TFM is impossible. They attempt to address these impossibilities using cryptography [31,42], but even with cryptographic methods, perfect TFMs remain unachievable. Furthermore, Chung et al. [7] and Gafni and Yaish [14] show that no mechanism can be incentive compatible for the users and the block producers while also being collusion-resistant. All of these impossibility results carry over to our context, as a single-leader protocol is a special case of a multiple-leader protocol. Although the majority of this work is prior-free, Zhao et al. [46] consider a Bayesian setup, demonstrating ways to circumvent these impossibility results in cases where bidders have i.i.d. valuations. Other works explore TFM dynamics over multiple blocks [10,21] and incorporate maximal extractable value (MEV) into traditional TFM models [3].

DAG-based consensus. Hashgraph [4] was the first protocol to introduce a DAG-based consensus protocol. It separated the communication layer and the consensus logic, with the communication layer constructing a DAG of messages which is then used by the consensus protocol. Later protocols adopted a round-based structure within the DAG to design more efficient asynchronous BFT protocols [15,17,9,34]. Among these, Bullshark's partially synchronous variant became the first widely deployed DAG-based consensus protocol, notably used in the Sui blockchain [35]. Since Bullshark's deployment, a number of papers have focused on reducing the latency of DAG-based consensus protocols [18,23,33,2,32,1]. The designs in these papers generally either move towards uncertified DAGs (which do not require explicit certification) or interleave multiple instances of the Bullshark protocol on a shared DAG. Mysticeti [2] has replaced Bullshark as the consensus protocol used by the Sui blockchain [36].

Economics of multiple leaders. There is a modest amount of work concerning the incentives faced by validators in multi-leader protocols. Zhang and Kate [45] show how DAG-based consensus protocols can be manipulated for MEV, while Malkhi et al. [24] propose MEV protection for such protocols. Fox et al. [11] look at the cost of censorship in single-leader protocols and show how TFMs specific to multi-leader protocols could potentially be used to significantly increase the cost of censorship. The Solana community has been considering whether to introduce multiple leaders to promote competition between block producers for the benefit of users [43], and Ethereum is planning on incorporating some of the ideas from multi-proposer architectures through FOCIL [39] to increase Ethereum's censorship resistance. The present work does not directly address questions around MEV, censorship, or explicit competition between block producers, but we believe that the model that we introduce in the next section can serve as the starting point for a formal study of these questions.

2 The Model

This section defines our game-theoretic model, the design space of transaction fee mechanisms, several notions of incentive-compatibility, and approximate welfare guarantees.

2.1 The Players

Games have three ingredients: players, strategy spaces, and payoffs. For transaction fee mechanisms (TFMs), there are two types of self-interested players, users and block producers (BPs). We discuss each in turn.

We assume that the set $I = \{1, 2, \ldots, n\}$ of users is known, and that each is identified with a single transaction; we refer to users and transactions interchangeably. We assume that user i has a private valuation v_i for the inclusion of its transaction in the next block, and that transaction validity does not depend on transaction ordering. When discussing Bayes-Nash equilibria (as is necessary when discussing TFMs without dominant strategies, such as variants of first-price auctions), we assume that user valuations $\mathbf{v}$ are drawn from a prior distribution $\mathbf{D}$ that is common knowledge among the users.[5] User valuations may be correlated; that is, $\mathbf{D}$ need not be a product distribution.

We consider TFMs in which each user attaches a nonnegative bid b_i to its transaction (thus, the strategy space of user i is the possible choices of b_i). We assume that each user has a quasi-linear utility function, meaning that its payoff is the value it receives (v_i if its transaction is included in the next block and 0 otherwise) minus the payment it makes. (Utilities functions will be stated more formally following the definition of TFMs; see Section 2.4.)

[5] We allow valuation distributions to have atoms at zero (or at other values), in which case the number of (non-null) players can be thought of as stochastic rather than known.

We also consider a set $J = \{1, 2, \ldots, m\}$ of BPs. BP strategies correspond to blocks, where for a known block size k, a *block* is a set of at most k transactions (together with the bids of those transactions). We assume that each BP $j \in J$ has an associated subset S_j of transactions that it can include in its block; we refer to the special case in which $S_j = I$ for all $j \in J$ as the *BP-symmetric setting* and the case of general S_j's as the *BP-asymmetric setting*. A block is *feasible* for BP j if it includes only transactions of S_j and, possibly, additional transactions created by j itself (e.g., in order to manipulate a TFM's payment rule). We assume that the S_j's are common knowledge. The payoff of a BP is defined as the revenue it earns from transactions other than its own minus.

2.2 The Game

TFM outcomes are, intuitively, determined by a two-stage process: users decide which bids to attach to their transactions, and BPs then decide which transactions to include. Previous work on TFMs, with a single BP, could essentially model the process with one stage (with the understanding that the BP will respond to users' bids with its favorite block). With multiple BPs best responding to each other (in addition to users' bids), it is important to explicitly model the block formation process as a two-stage game. We do this next, using the standard formalism for extensive-form games (e.g. [12]).

Game trees. To review, an extensive form game is defined by a rooted tree (the *game tree*). Each node represents a single action to be taken by a single player, with the node labeled with that player and edges leading to the node's children labeled with the possible actions. Each leaf of the tree corresponds to an outcome of the game, and is labeled with players' payoffs in that outcome. Thus, root-leaf paths of the game tree correspond to action sequences that terminate in an outcome of the game. It is a convenient tradition to allow nodes that are labeled with a non-strategic "Nature" player, indicating that the action at that node is chosen at random from a distribution that is common knowledge. Finally, for each player, the nodes labeled with that player are partitioned into information sets. An information set represents a set of nodes that are indistinguishable to the player at the time it must take an action (and thus, the same action must be taken by a player at all nodes in the same information set).

To model behavior in TFMs with multiple BPs, we consider a game tree with $n + m + 1$ levels. (The outcomes and payoffs at the leaves of this tree will depend on the choice of the TFM, but the tree structure is independent of the particular TFM.) At level 0, Nature moves and chooses valuations $\mathbf{v}$ for all users from the assumed prior $\mathbf{D}$. At each level $l = 1, 2, \ldots, n$, user l selects its bid b_l. Information sets are defined for user l so that its choice of bid depends only on its own valuation v_l (and not on the other valuations $\mathbf{v}_{-l}$ determined at level 0 or the bids chosen by users $i \in \{1, 2, \ldots, l-1\}$ at earlier levels). At each level $l = n+1, n+2, \ldots, n+m$, BP $j = l - n$ selects its block B_j. Information

sets are defined for a BP so that its choice of block can depend on the bids $\mathbf{b}$ chosen by users but not on the blocks chosen by the other BPs. [6]

Subgame perfect equilibria. Our analysis uses what is arguably the most canonical equilibrium concept in extensive-form games, namely subgame perfect equilibria. In such a game, a strategy for a player is defined by a mapping from each of its information sets to one of the actions available at that information set. In our model of TFMs, a user has one information set for each realization of its valuation, and a BP has one information set for each user bid vector. Thus, a user strategy is simply a bidding strategy, meaning a mapping $v_i \mapsto b_i$ from valuations to bids. A BP strategy is a mapping $\mathbf{b} \mapsto B_j$ from user bid vectors to feasible blocks. Thus, leaves of the game tree are effectively labeled by $\mathbf{v}$ (Nature's action at level 0), $\mathbf{b}$ (users' actions at levels 1 through n), and $\mathbf{B}$ (BPs' actions at levels $n+1$ through $n+m$); these, in conjunction with the choice of a TFM, will define the player payoffs at this outcome.

A strategy profile in an extensive-form game is called a Nash equilibrium if the usual best-response condition holds: no player can strictly improve its expected payoff through a unilateral deviation to a different mapping of its information sets to actions. That is, each player is best responding to the strategies chosen by the other players.

Every node of a game tree induces a rooted subtree that can be regarded as an extensive-form game in its own right. Similarly, every strategy of an extensive-form game induces a strategy in each of its subgames. A strategy profile of an extensive-form game is called a *subgame perfect equilibrium (SPE)* if, for each if its subgames, the induced strategy profile is a Nash equilibrium. Intuitively, even after "fast forwarding" to an arbitrary node of the game tree, play from then on constitutes a Nash equilibrium.[7]

Intuitively, in our model of TFMs with multiple BPs, the SPE condition translates to (i) users play a Bayes-Nash equilibrium relative to the BP equilibrium strategies; (ii) BPs play a Nash equilibrium relative to the user bids.[8]

[6] Thus, BPs engage in a complete-information game, with the full bid vector $\mathbf{b}$ and the S_j's known to all BPs. A good (though possibly difficult) direction for future work is to consider an incomplete-information generalization of our model. With our assumptions, users can effectively treat BPs as carrying out the welfare-maximizing allocation rule. In an incomplete-information setup, users would effectively be submitting bids to a randomized allocation rule induced by some (perhaps impossible-to-characterize) Bayes-Nash equilibrium played by the BPs.

[7] Without the subgame perfect refinement, Nash equilibria of extensive-form games allow players to play arbitrary strategies in subgames that are reached with probability 0.

[8] We do not model how BPs coordinate on a given equilibrium. Microfounding the assumption that BPs reach an equilibrium (e.g., through experience from repeated play, explicit coordination based on transaction hashes, or other means) is an interesting direction for future research.

2.3 Transaction Fee Mechanisms

A TFM is specified by four ingredients: an inclusion rule (the blocks of transactions that the BPs are expected to contribute), a confirmation rule (given the proposed blocks, which transactions are confirmed for execution), a payment rule (given the proposed blocks, what the creators of confirmed transactions pay), and a distribution rule (given the proposed blocks, the revenue received by BPs). Because BPs have unilateral control over the transactions they include, the inclusion rule can only be viewed as a recommendation to BPs; the other three rules are hard-coded into the code of a blockchain protocol and cannot be manipulated by BPs.

We next define these four ingredients formally, along with a number of examples that illustrate the definitions and demonstrate the richness of the TFM design space with multiple BPs. (Many of the examples are deferred to Appendix A.1.) These rules are all defined with respect to a commonly known *game structure*, meaning a player set I, a BP set J, BP transaction sets $S_1, \ldots, S_m$, and a block size k.[9] Recall that a block B_j is *feasible for* j if it includes only transactions of S_j and, possibly, transactions that j itself created (along with the bids attached to the included transactions). When we are concerned only with the transactions included in a block and not the attached bids, we sometimes abuse notation and treat a block as a subset of I. We call a profile $\mathbf{B} = (B_1, \ldots, B_m)$ of block choices an *allocation*, and call an allocation *feasible* if each of its blocks B_j is feasible for the corresponding BP j. We call an allocation *shill-free* if, for each of its blocks, only user-submitted transactions are included (i.e., $B_j \subseteq S_j$ for every BP j). Note that the same transaction may be included in more than one block of an allocation. We denote by $T(\mathbf{B}) = \cup_{j \in J} B_j$ the transactions that are included (at least once) in an allocation $\mathbf{B}$.

Inclusion rules. An inclusion rule can be thought of as a recommendation of the strategies that BPs should play in each information set of the extensive-form game described in Section 2.2. Formally, with respect to a game structure, an *inclusion rule* is a function $\mathbf{y} : \mathbf{b} \mapsto \mathbf{B}$ mapping user bids vectors to feasible allocations.

For example, the *welfare-maximizing (WM)* inclusion rule maps each bid vector to a feasible shill-free allocation that maximizes the sum of the bids of the included transactions (breaking ties using some consistent rule). For TFMs with first-price payment rules (see below), this inclusion rule can be interpreted as maximizing the total fees paid by users.[10]

Confirmation rules. A confirmation rule specifies which of the included transactions are confirmed for execution. Formally, with respect to a game structure, a *confirmation rule* is a function $\mathcal{C} : \mathbf{B} \mapsto \mathcal{B}$ that maps each feasible allocation $\mathbf{B}$

[9] The valuation distribution $\mathbf{D}$ is not part of the game structure; in this sense, a TFM is by definition prior-free.

[10] For another example, the *serial dictatorship* inclusion rule is described in Appendix A.1.

to a set $\mathcal{B} \subseteq T(\mathbf{B})$ of confirmed transactions. Note that while a transaction may be included in multiple blocks, it can only be confirmed once.

For example, the *first-price auction (FPA) confirmation rule* confirms every transaction that is included at least once: $\mathcal{C}(\mathbf{B}) = T(\mathbf{B})$.[11]

Payment rules. A payment rule specifies the transaction fee paid by the creator of an included transaction. Formally, with respect to a game structure, a *payment rule* is a function $\mathbf{p}$ that maps each feasible allocation $\mathbf{B}$ to a set of n nonnegative numbers (one per user).

For example, the *first-price auction (FPA) payment rule* charges the creator of an included transaction its bid: $p_i(\mathbf{B}) = b_i$ if $i \in T(\mathbf{B})$ and $p_i(\mathbf{B}) = 0$ otherwise.[12]

Distribution rules. A distribution rule specifies the revenue earned by each BP from the set of included transactions. Formally, with respect to a game structure, a *distribution rule* is a function π that maps each feasible allocation $\mathbf{B}$ to a set of m nonnegative numbers (one per BP).

For example, the *equal-share* distribution rule (FPA version) splits the bid of each included transaction equally between the BPs: for all j,

$$\pi_j(\mathbf{B}) = \frac{1}{m} \sum_{i \in T(\mathbf{B})} b_i. \tag{1}$$

Many other examples are possible (e.g., splitting the bid of a transaction between only the BPs that include it in their blocks); see Appendix A.1 for details.

TFMs. A *transaction fee mechanism (TFM)* is then a tuple $(\mathbf{y}, \mathcal{C}, \mathbf{p}, \pi)$. We restrict attention to TFMs that satisfy the following properties (which are also shared by all TFMs that have been deployed in practice to-date): (i) deterministic, meaning that $\mathbf{y}$, $\mathcal{C}$, $\mathbf{p}$, and π are all deterministic functions of their inputs; and (ii) ex post individually rational, meaning that $p_i(\mathbf{B}) = 0$ if user i's transaction is not confirmed by the TFM (i.e., $i \notin \mathcal{C}(\mathbf{B})$) and $p_i(\mathbf{B}) \leq b_i$ otherwise; (iii) weakly budget-balanced, meaning that users' payments always cover BP revenue: $\sum_{j \in J} \pi_j(\mathbf{B}) \leq \sum_{i \in I} p_i(\mathbf{B})$ for every feasible allocation $\mathbf{B}$.[13] We do allow the user payments to exceed the BP revenue, in which we case the remaining user payments are burned (or otherwise redirected away from BPs, for example to a foundation).

TFMs can be assembled from different inclusion, confirmation, payment, and distribution rules in many natural ways; see Appendix A.1 for an incomplete list.

[11] One reason to include unconfirmed transactions is to use their bids to set prices for the confirmed transactions, in the spirit of a second-price auction. For more details, see the *second-price auction (SPA)* confirmation rule described in Appendix A.1.

[12] For another example, the *second-price auction (SPA)* payment rule is described in Appendix A.1.

[13] As an extension to (iii), money-printing in the form of inflationary rewards (like a block reward) can be added to a TFM without affecting its incentive or welfare properties, provided the rewards are the same no matter which feasible allocation $\mathbf{B}$ is chosen by the BPs.

2.4 Incentive Compatibility

Intuitively, a mechanism is incentive-compatible if its participants are motivated to behave in a prescribed way, such as by bidding truthfully (in the case of users) or by choosing blocks as instructed by a TFM's inclusion rule (in the case of BPs). We next formalize these two incentive-compatibility properties (one for users, one for BPs).

Dominant-strategy incentive-compatibility (DSIC). We first observe that the composition of an (intended) inclusion rule $\mathbf{y}$ and confirmation rule $\mathcal{C}$ of a TFM induce an (intended) *allocation rule* $\mathbf{x}$, with $x_i(\mathbf{b}) = 1$ if $i \in \mathcal{C}(\mathbf{y}(\mathbf{b}))$ and $x_i(\mathbf{b}) = 0$ otherwise. That is, $\mathbf{x}(\mathbf{b})$ is the characteristic vector of the confirmed transactions with user bids $\mathbf{b}$, assuming that the BPs carry out the intended inclusion rule. Under the same assumption, the payoff of user i under bid vector $\mathbf{b}$ in the TFM $(\mathbf{y}, \mathcal{C}, \mathbf{p}, \pi)$ is

$$u_i(\mathbf{b}) = v_i \cdot x_i(\mathbf{b}) - p_i(\mathbf{y}(\mathbf{b})). \tag{2}$$

A TFM is then *dominant-strategy incentive-compatible (DSIC)* if, for every user i, valuation v_i, and bid vector $\mathbf{b}$, $u_i(v_i, \mathbf{b}_{-i}) \geq u_i(\mathbf{b})$. That is, after fixing the BP strategies to be those recommended by the TFM's inclusion rule, truthful bidding is a dominant strategy for every user. For example, in the BP-symmetric setting (with $S_j = I$ for all $j \in J$), the SPA-EQ and SPA-Shapley TFMs (see Appendix A.1) are DSIC. TFMs that use the FPA payment rule are never DSIC, as users are incentivized to shade their bids.

Block producer incentive-compatibility (BPIC). In an outcome of a TFM $(\mathbf{y}, \mathcal{C}, \mathbf{p}, \pi)$, specified by the bids $\mathbf{b}$ chosen by users and the feasible allocation $\mathbf{B}$ chosen by BPs, the payoff of BP j is $\pi_j(\mathbf{B})$. A TFM is then *block producer incentive-compatible (BPIC)* if, for every bid vector $\mathbf{b}$ with corresponding intended allocation $\mathbf{y}(\mathbf{b}) = \mathbf{B} = (B_1, \ldots, B_m)$, every BP j, and every block $B_j{}'$ feasible for j, $\pi_j(\mathbf{B}) \geq \pi_j(B_j{}', \mathbf{B}_{-j})$. That is, after fixing the user bids to $\mathbf{b}$, the feasible allocation recommended by the TFM's inclusion rule is a Nash equilibrium among the BPs.

For example, the SPA-EQ and SPA-Shapley TFMs from Appendix A.1 are not BPIC, as BPs generally have an incentive to deviate from the WM allocation rule by including their own transactions in order to boost their overall revenue. The FPA-Shapley TFM (see Appendix A.1) fails to satisfy BPIC for a different reason: BPs are generally incentivized to redundantly include a high-bid transaction multiple times rather that following the WM allocation rule (in which each transaction is included at most once).

Strong BPIC. Despite the fact that many natural TFMs fail to satisfy it, the BPIC condition is relatively weak. For example, any TFM that uses the null distribution rule (with all transaction fees burned) is trivially BPIC, with all BPs indifferent across all outcomes. Thus, the BPIC condition does not generally provide much force toward BPs carrying out the intended inclusion rule.

The next condition, a strengthening of BPIC, states that the intended allocation should not merely be a Nash equilibrium, but should also be strictly

superior to all non-equivalent Nash equilibria. Formally, a TFM $(\mathbf{y}, \mathcal{C}, \mathbf{p}, \pi)$ is *strongly BPIC* if, for every user bid vector $\mathbf{b}$, the following conditions hold:

1. the recommended feasible allocation $\mathbf{B} = \mathbf{y}(\mathbf{b})$ is a Nash equilibrium among the BPs (holding user bids fixed at $\mathbf{b}$);
2. every Nash equilibrium $\mathbf{B}'$ among the BPs (again, with fixed bids $\mathbf{b}$) is either equivalent to or Pareto dominated by $\mathbf{B}$.

Intuitively, two feasible allocations are "equivalent" if they are the same up to tie-breaking and the inclusion of zero-bid transactions. Formally, for a TFM $(\mathbf{y}, \mathcal{C}, \mathbf{p}, \pi)$, feasible allocations $\mathbf{B}$ and $\mathbf{B}'$ are *equivalent* if the multi-sets of the positive bids of the confirmed transactions $\mathcal{C}(\mathbf{B})$ and $\mathcal{C}(\mathbf{B}')$ are identical. We say that one allocation $\mathbf{B}$ *Pareto dominates* another allocation $\mathbf{B}'$ if: (i) $\pi_j(\mathbf{B}) \geq \pi_j(\mathbf{B}')$ for all $j \in J$; and (ii) $\pi_j(\mathbf{B}) > \pi_j(\mathbf{B}')$ for some $j \in J$. We'll see in Section 3.2 an example of a strongly BPIC TFM (the FPA-EQ TFM).

2.5 Approximate Welfare Guarantees

We assess the outcome quality of different TFMs using the welfare objective $W(\cdot)$, defined as the total value of the confirmed transactions. That is, for a TFM $(\mathbf{y}, \mathcal{C}, \mathbf{p}, \pi)$ and feasible allocation $\mathbf{B}$, $W(\mathbf{B}) = \sum_{i \in \mathcal{C}(\mathbf{B})} v_i$. TFMs can suffer from welfare loss for three distinct reasons. First, even if all participants behave as desired, a TFM's inclusion rule may result in a suboptimal feasible allocation. Second, even with the WM allocation rule and truthful bids, BPs may coordinate on a suboptimal Nash equilibrium. Third, even with the WM allocation rule and BPs that coordinate on the intended Nash equilibrium, non-truthful bidding by users can lead to suboptimal allocations. (See Appendix A.2 for examples of all three types.) Thus, a equilibrium welfare approximation guarantee is a guarantee that the welfare loss *from all three of these sources combined* is relatively modest.

3 FPA-EQ: A Strongly BPIC and Near-Optimal TFM

3.1 What Can We Hope For?

We have highlighted three desirable properties of TFMs (in addition to our standing requirements that TFMs be deterministic and ex post individually rational): (i) DSIC; (ii) strong BPIC; and (iii) optimal or near-optimal welfare at equilibrium. In this work, we take the strong BPIC condition (ii) as a hard constraint. (If BPs are not properly motivated to carry out the intended inclusion rule, which in turn determines the confirmed transactions and their payments, it's unclear how to interpret a proposed TFM.) In Theorem 4 in Appendix B.1, we prove that no DSIC and strongly BPIC TFM can achieve a non-trivial equilibrium welfare guarantee. This result implies that we have no choice but to consider non-DSIC TFMs. In Theorem 5 in Appendix B.2, we prove that no (possibly non-DSIC) strongly BPIC TFM can guarantee optimal welfare at equilibrium.

In light of these negative results, the best-case scenario is a strongly BPIC TFM that guarantees near-optimal welfare at equilibrium. We present such a TFM next.

3.2 The FPA-EQ TFM

The rest of this section analyzes the *first-price auction with equal sharing (FPA-EQ)* TFM. The ingredients of this TFM were all introduced in Section 2.3:

- the welfare-maximizing (WM) inclusion rule (i.e., with $\mathbf{y}(\mathbf{b}) = \mathbf{B}$ chosen to maximize the sum of the bids $\sum_{i \in T(\mathbf{B})} b_i$ of the included transactions, with ties broken according to some consistent rule);
- the FPA confirmation rule (with all included transaction confirmed: $\mathcal{C}(\mathbf{B}) = T(\mathbf{B})$);
- the FPA payment rule (with each user of a confirmed transaction paying its bid);
- the equal share (FPA version) distribution rule (with the payment for each confirmed transaction split equally between the m block producers, as in (1)).

Because of its FPA payment rule, the FPA-EQ TFM is not DSIC; bidders are incentivized to shade their bids. Unlike many other natural TFMs, however, the FPA-EQ TFM is strongly BPIC. The proof of this fact leans heavily on the choice of the equal-share distribution rule, and also on the matroid structure of feasible allocations.

Proposition 1 (FPA-EQ Is Strongly BPIC) *For every game structure, the FPA-EQ TFM is strongly BPIC.*

Proof. Fix a game structure and a user bid vector $\mathbf{b}$. The payoff of every BP is proportional to the total amount paid by users (due to the equal-share distribution rule), and therefore to the sum of the bids of the confirmed transactions (due to the FPA payment rule), and therefore to the sum of the bids of the included transactions (due to the FPA confirmation rule). Because the WM allocation rule instructs BPs to maximize the sum of the bids of the included transactions over feasible allocations, the intended allocation $\mathbf{B}^*$ is a Nash equilibrium among the BPs (holding user bids fixed at $\mathbf{b}$). By the same reasoning, $\mathbf{B}^*$ Pareto dominates every Nash equilibrium allocation that fails to maximize the sum of the bids of the included transactions. Finally, because the subsets of transactions that can be included in a feasible allocation form a matroid (see Proposition 6) and due to the lexicographic optimality property of matroids (see Proposition 7), every feasible allocation $\mathbf{B}$ that maximizes the sum of the included bids is equivalent to $\mathbf{B}^*$ (i.e., after ignoring zero-bid transactions, the multi-sets of bids of transactions in $\mathcal{C}(\mathbf{B})$ and $\mathcal{C}(\mathbf{B}^*)$ are identical).

3.3 An Approximate Welfare Guarantee for FPA-EQ

Our main result in this section is that the FPA-EQ TFM, in addition to satisfying the strong BPIC property (Proposition 1), achieves near-optimal welfare at equilibrium. Precisely, in the extensive-form game induced by this TFM $(\mathbf{y}, \mathcal{C}, \mathbf{p}, \pi)$ (see Section 2.2), call a strategy profile *inclusion-rule respecting (IRR) at* $\mathbf{b}$ if, in the subgame corresponding to $\mathbf{b}$, the BPs choose a feasible allocation that is equivalent to $\mathbf{y}(\mathbf{b})$. (As in Section 2.4, two feasible allocations are equivalent if the resulting sets of confirmed transactions share the same multi-sets of positive bids.) A subgame-perfect equilibrium is then called inclusion-rule respecting if it is IRR at every user bid vector $\mathbf{b}$. For a strongly BPIC TFM like FPA-EQ, there is good reason to focus on its IRR SPE—in any other SPE, there are bids vectors for which BPs inexplicably coordinate on a subgame equilibrium that is Pareto dominated by the one suggested by the TFM's inclusion rule.

Theorem 2 (FPA-EQ Is Approximately Welfare-Optimal). *For every game structure and valuation distribution* $\mathbf{D}$*, every inclusion-rule-respecting subgame perfect equilibrium of the FPA-EQ TFM has expected welfare at least* $1 - \frac{1}{e} \approx 63.2\%$ *of the maximum possible.*

The proof of Theorem 2 proceeds in two steps. The first step establishes an equivalence between the IRR SPE of the FPA-EQ TFM and the Bayes-Nash equilibria of a (single-shot) winner-pays-bid matroid auction. Intuitively, with the BP behavior fixed (up to allocation equivalence) in an IRR SPE, we can analyze users as if they are competing in a single-stage game. See Lemma 3 in Appendix B.4 for details.

The second step of the proof applies the theory of smooth games (see e.g. [30]) to prove a worst-case bound on the expected welfare of the Bayes-Nash equilibria of winner-pays-bid matroid auctions.[14] The details are fairly technical and deferred to Appendix B.4.

We can obtain stronger guarantees if we impose symmetry conditions on the BPs and users. In the BP-symmetric setting (see Section 2.1), a simple exchange argument shows that *every* SPE of the FPA-EQ TFM is IRR. Thus:

Corollary 1. *In the BP-symmetric setting, for every game structure and valuation distribution* $\mathbf{D}$*, every subgame perfect equilibrium of the FPA-EQ TFM has expected welfare at least* $1 - \frac{1}{e} \approx 63.2\%$ *of the maximum possible.*

Adapting an example of Syrgkanis [37] for first-price auctions to the present setting gives a lower bound showing that the approximation factor of $1 - \frac{1}{e}$ in Theorem 2 and Corollary 1 is tight (see Appendix B.4 for details).

Proposition 3 (Theorem 2 Is Tight) *There exists a game structure, valuation distribution* $\mathbf{D}$*, and an inclusion-rule-respecting subgame perfect equilibrium of the FPA-EQ TFM with expected welfare* $1 - \frac{1}{e}$ *times the expected maximum welfare.*

[14] Such a bound was proved in [16] for the special case of independent user valuations; the bound here for correlated user valuations appears to be new.

If we further assume that users are symmetric, meaning that their valuations are drawn i.i.d. from a common distribution, then every SPE of the FPA-EQ TFM is in fact fully efficient. The following corollary follows from Lemma 3 (in Appendix B.4) and the full efficiency of Bayes-Nash equilibria in multi-unit auctions with symmetric unit-demand bidders (see e.g. [19]):

Corollary 2 (Optimal Welfare in Symmetric Settings). *In the BP-symmetric setting, for every game structure and i.i.d. valuation distribution, every subgame perfect equilibrium of the FPA-EQ TFM achieves the maximum-possible expected welfare.*

As noted in Section 2.2, these positive results assume that BPs are capable of coordinating on an equilibrium of the appropriate type. It would be interesting to investigate how our guarantees would change under weaker versions of this assumption.

References

1. Arun, B., Li, Z., Suri-Payer, F., Das, S., Spiegelman, A.: Shoal++: High throughput dag bft can be fast! arXiv preprint arXiv:2405.20488 (2024)
2. Babel, K., Chursin, A., Danezis, G., Kokoris-Kogias, L., Sonnino, A.: Mysticeti: Low-latency dag consensus with fast commit path. arXiv preprint arXiv:2310.14821 (2023)
3. Bahrani, M., Garimidi, P., Roughgarden, T.: Transaction fee mechanism design in a post-mev world. In: Böhme, R., Kiffer, L. (eds.) 6th Conference on Advances in Financial Technologies, AFT 2024, September 23-25, 2024, Vienna, Austria. LIPIcs, vol. 316, pp. 29:1–29:24. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2024). https://doi.org/10.4230/LIPICS.AFT.2024.29, https://doi.org/10.4230/LIPIcs.AFT.2024.29
4. Baird, L.: The swirlds hashgraph consensus algorithm: Fair, fast, byzantine fault tolerance. Swirlds Tech Reports SWIRLDS-TR-2016-01, Tech. Rep **34**, 9–11 (2016)
5. Buchman, E.: Tendermint: Byzantine fault tolerance in the age of blockchains. Ph.D. thesis, University of Guelph (2016)
6. Buterin, V., Conner, E., Dudley, R., Slipper, M., Norden, I., Bakhta, A.: Eip-1559: Fee market change for eth 1.0 chain. https://github.com/ethereum/EIPs/blob/master/EIPS/eip-1559.md (2024), accessed: 2024-10-10
7. Chung, H., Roughgarden, T., Shi, E.: Collusion-resilience in transaction fee mechanism design. arXiv preprint arXiv:2402.09321 (2024)
8. Chung, H., Shi, E.: Foundations of transaction fee mechanism design. In: Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). pp. 3856–3899. SIAM (2023)
9. Danezis, G., Kokoris-Kogias, L., Sonnino, A., Spiegelman, A.: Narwhal and tusk: a dag-based mempool and efficient bft consensus. In: Proceedings of the Seventeenth European Conference on Computer Systems. pp. 34–50 (2022)
10. Ferreira, M.V.X., Moroz, D.J., Parkes, D.C., Stern, M.: Dynamic posted-price mechanisms for the blockchain transaction-fee market. In: Proceedings of the 3rd ACM Conference on Advances in Financial Technologies. pp. 86–99 (2021)

11. Fox, E., Pai, M.M., Resnick, M.: Censorship resistance in on-chain auctions. In: Bonneau, J., Weinberg, S.M. (eds.) 5th Conference on Advances in Financial Technologies, AFT 2023, October 23-25, 2023, Princeton, NJ, USA. LIPIcs, vol. 282, pp. 19:1–19:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2023). https://doi.org/10.4230/LIPICS.AFT.2023.19, https://doi.org/10.4230/LIPIcs.AFT.2023.19
12. Fudenberg, D., Tirole, J.: Game Theory. MIT press (1991)
13. Gafni, Y., Yaish, A.: Greedy transaction fee mechanisms for (non-) myopic miners. arXiv preprint arXiv:2210.07793 (2022)
14. Gafni, Y., Yaish, A.: Barriers to collusion-resistant transaction fee mechanisms. arXiv preprint arXiv:2402.08564 (2024)
15. Gkagol, A., Leśniak, D., Straszak, D., Świętek, M.: Aleph: Efficient atomic broadcast in asynchronous networks with byzantine nodes. In: Proceedings of the 1st ACM Conference on Advances in Financial Technologies. pp. 214–228 (2019)
16. Hartline, J., Hoy, D., Taggart, S.: Price of Anarchy for Auction Revenue. In: Proceedings of the 15th ACM conference on Economics and Computation. pp. 693–710 (2014)
17. Keidar, I., Kokoris-Kogias, E., Naor, O., Spiegelman, A.: All you need is dag. In: Proceedings of the 2021 ACM Symposium on Principles of Distributed Computing. pp. 165–175 (2021)
18. Keidar, I., Naor, O., Poupko, O., Shapiro, E.: Cordial miners: Fast and efficient consensus for every eventuality. arXiv preprint arXiv:2205.09174 (2022)
19. Krishna, V.: Auction Theory. Academic press (2009)
20. Lavi, R., Sattath, O., Zohar, A.: Redesigning bitcoin's fee market. ACM Transactions on Economics and Computation **10**(1), 1–31 (2022)
21. Leonardos, S., Monnot, B., Reijsbergen, D., Skoulakis, S., Piliouras, G.: Dynamical analysis of the EIP-1559 Ethereum fee market. In: Proceedings of the 3rd ACM Advances in Financial Technologies (2021)
22. Lucier, B., Paes Leme, R.: GSP Auctions with Correlated Types. In: Proceedings of the 12th ACM Conference on Electronic Commerce. pp. 71–80 (2011)
23. Malkhi, D., Stathakopoulou, C., Yin, M.: Bbca-chain: One-message, low latency bft consensus on a dag. arXiv preprint arXiv:2310.06335 (2023)
24. Malkhi, D., Szalachowski, P.: Maximal extractable value (mev) protection on a dag. In: 4th International Conference on Blockchain Economics, Security and Protocols. p. 1 (2023)
25. Nakamoto, S.: A peer-to-peer electronic cash system (2008)
26. Nisan, N.: Serial monopoly on blockchains (2023)
27. Oxley, J.G.: Matroid Theory, vol. 3. Oxford University Press, USA (2006)
28. Roughgarden, T.: Transaction Fee Mechanism Design for the Ethereum Blockchain: An Economic Analysis of EIP-1559. arXiv preprint arXiv:2012.00854 (2020)
29. Roughgarden, T.: Transaction Fee Mechanism Design. ACM SIGecom Exchanges **19**(1), 52–55 (2021), full version at https://arxiv.org/abs/2106.01340
30. Roughgarden, T., Syrgkanis, V., Tardos, E.: The Price of Anarchy in Auctions. Journal of Artificial Intelligence Research **59**, 59–101 (2017)
31. Shi, E., Chung, H., Wu, K.: What can cryptography do for decentralized mechanism design. arXiv preprint arXiv:2209.14462 (2022)
32. Shrestha, N., Shrothrium, R., Kate, A., Nayak, K.: Sailfish: Towards improving latency of dag-based bft. Cryptology ePrint Archive (2024)
33. Spiegelman, A., Arun, B., Gelashvili, R., Li, Z.: Shoal: Improving dag-bft latency and robustness. arXiv preprint arXiv:2306.03058 (2023)

34. Spiegelman, A., Giridharan, N., Sonnino, A., Kokoris-Kogias, L.: Bullshark: Dag bft protocols made practical. In: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. pp. 2705–2718 (2022)
35. Spiegelman, A., Giridharan, N., Sonnino, A., Kokoris-Kogias, L.: Bullshark: the partially synchronous version. arXiv preprint arXiv:2209.05633 (2022)
36. Sui Foundation: Sui Consensus Architecture. `https://docs.sui.io/concepts/sui-architecture/consensus` (2024), accessed: 2024-10-09
37. Syrgkanis, V.: Efficiency of Mechanisms in Complex Markets. Ph.D. thesis, Cornell University (2014)
38. Syrgkanis, V., Tardos, E.: Composable and efficient mechanisms. In: Proceedings of the 45th annual ACM Symposium on Theory of Computing. pp. 211–220 (2013)
39. Thomas, Barnabe, Francesco, Julian: Fork-choice enforced inclusion lists (focil): A simple committee-based inclusion list proposal. `https://ethresear.ch/t/fork-choice-enforced-inclusion-lists-focil-a-simple-committee-based-inclusion-list-proposal/19870` (2024)
40. Vickrey, W.: Counterspeculation, auctions, and competitive sealed tenders. The Journal of finance **16**(1), 8–37 (1961)
41. Wood, G.: Ethereum: A secure decentralised generalised transaction ledger. `https://ethereum.github.io/yellowpaper/paper.pdf` (2014), accessed: 2024-10-10
42. Wu, K., Shi, E., Chung, H.: Maximizing miner revenue in transaction fee mechanism design. In: Guruswami, V. (ed.) 15th Innovations in Theoretical Computer Science Conference, ITCS 2024, January 30 to February 2, 2024, Berkeley, CA, USA. LIPIcs, vol. 287, pp. 98:1–98:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2024)
43. Yakovenko, A.: Multiple concurrent leaders. `https://x.com/aeyakovenko/status/1810222589991583922` (2024), accessed: 10, 11, 2024
44. Yao, A.C.C.: An incentive analysis of some Bitcoin fee designs. In: Proceedings of the 47th International Colloquium on Automata, Languages, and Programming (ICALP) (2020)
45. Zhang, J., Kate, A.: No fish is too big for flash boys! frontrunning on DAG-based blockchains. Cryptology ePrint Archive, Paper 2024/1496 (2024), `https://eprint.iacr.org/2024/1496`
46. Zhao, Z., Chen, X., Zhou, Y.: Bayesian-Nash-incentive-compatible mechanism for blockchain transaction fee allocation. arXiv preprint arXiv:2209.13099 (2022)

A Supplementary Material for Section 2

A.1 Further Examples of TFMs

All of the following rules are defined with respect to a game structure (a user set I, a BP set J, BP transaction sets $S_1, \ldots, S_m$, and a block size k). In all cases, ties are broken according to some consistent tie-breaking rule.

Example 1 (Serial Dictatorship Inclusion Rule). This inclusion rule is defined, for every user bid vector $\mathbf{b}$, by $\mathbf{y}(\mathbf{b}) = (B_1, \ldots, B_m)$, where B_j is chosen to maximize the sum of the bids of the included transactions, subject to disjointness with $B_1, \ldots, B_{j-1}$ (and feasibility). That is, B_j is the k highest-bidding transactions in $S_j \setminus \cup_{h=1}^{j-1} B_h$. (Or, if there are less than k such transactions, all of them are included.)

Example 2 (Second-Price Auction (SPA) Confirmation Rule). This confirmation rule confirms all but the lowest-bidding included transaction. That is, $\mathcal{C}(\mathbf{B}) = T(\mathbf{B}) \setminus \{t\}$, where t is the transaction of $T(\mathbf{B})$ with the lowest bid.

Example 3 (Second-Price Auction (SPA) Payment Rule). This payment rule charges 0 to the lowest-bidding included transaction t, and b_t to the other included transactions. That is, $p_i(\mathbf{B}) = b_t$ if $i \in T(\mathbf{B})\setminus\{t\}$ and $p_i(\mathbf{B}) = 0$ otherwise.

Example 4 (The Null Distribution Rule). This distribution rule burns all transaction fees: $\pi_j(\mathbf{B}) = 0$ for all $\mathbf{B}$ and j.

Example 5 (Shapley Distribution Rule (FPA Version)). This distribution rule splits the bid of each included transaction equally among the BPs that included it. That is,

$$\pi_j(\mathbf{B}) = \sum_{i \in B_j} \frac{b_i}{m_i(\mathbf{B})},$$

where $m_i(\mathbf{B}) = |\{h \in J : i \in B_h\}$ denotes the number of BPs that included i in their block.

The distribution rule above is intended for use with the FPA payment rule. The SPA version of the Shapley distribution rule is defined similarly, except with b_i replaced by the lowest bid of an included transaction and with no BP earning any revenue from the lowest-bidding transaction.

Example 6 (Serial Dictatorship Distribution Rule (FPA Version)). This distribution rule passes on revenue earned from an included transaction to the lexicographically first BP that included it. That is,

$$\pi_j(\mathbf{B}) = \sum_{i \in B_j \setminus \cup_{h=1}^{j-1} B_h} b_i.$$

The distribution rule above is intended for use with the FPA payment rule. The SPA version of the rule is defined similarly, except with b_i replaced by the lowest bid of an included transaction and with no BP earning any revenue from the lowest-bidding transaction.

There are numerous ways to combine these rules or the rules described in Section 2.3 to produce natural TFMs. The FPA-EQ TFM is analyzed at length in Section 3.3. Other examples include:

1. *SPA-EQ*: WM inclusion rule, SPA confirmation rule, SPA payment rule, equal-share distribution rule (SPA version). (The SPA version of the equal-share distribution rule in (1) replaces b_i by the lowest bid of an included transaction and sums only over the transactions of $T(\mathbf{B})$ other than the lowest-bidding one.)
2. *FPA-Shapley*: WM inclusion rule, FPA confirmation rule, FPA payment rule, Shapley distribution rule (FPA version).

3. *SPA-Shapley*: WM inclusion rule, SPA confirmation rule, SPA payment rule, Shapley distribution rule (SPA version).
4. *FPA-Serial*: serial dictatorship inclusion rule, FPA confirmation rule, FPA payment rule, serial dictatorship distribution rule (FPA version).
5. *SPA-Serial*: serial dictatorship inclusion rule, SPA confirmation rule, SPA payment rule, serial dictatorship distribution rule (SPA version).

A.2 Examples of TFM Welfare Loss

Welfare losses in TFM can have multiple sources. First, a TFM's inclusion rule may result in a suboptimal feasible allocation even if all participants behave as desired. Examples include:

- In the BP-symmetric setting and with known valuations, and a FPA-Shapley TFM which we redefine to be BPIC. This would mean replacing the WM inclusion rule with an inclusion rule whereby the BPs maximize their personal revenue. Then, let $k = 1$ be the block size, there is one transaction with bid $b_1 = m + \epsilon$, where $\epsilon \to 0$, and $m - 1$ transaction with bids $b_i = 1$ for $i \in [2, \ldots, m]$. Then all BPs would all include the first transaction, resulting in a welfare ≈ 2 worse than optimal.
- Consider the following setting, known user valuations, and a TFM serial dictatorship inclusion rule. The block size is $k = 1$, there are $m = 2$ BPs, and two transactions with bids $b_1 = b_2 = 1$ of which the first BP can include both, while the second BP can only include the first transaction. The first BP would include the first transaction and there would be no transactions for the second BP to include – the resulting welfare is a factor of 2 worse than optimal.

Second, TFM welfare losses can result from BPs coordinating on a suboptimal Nash equilibrium even with the WM allocation rule and truthful bids as the following example demonstrates.

- Consider a FPA-EQ TFM in the BP-asymmetric setting. Again, the block size is $k = 1$, there are $m = 2$ BPs, and two transactions with bids $b_1 = b_2 = 1$ of which the first BP can include both, while the second BP can only include the first transaction. The first BP including the first transaction and the second BP including no transaction is a Nash equilibrium that again results in a welfare ≈ 2 worse than optimal.

Third, non-truthful bidding in stage 1 by users can lead to suboptimal allocations even with the WM allocation rule and BPs that coordinate on the as demonstrated by Vickrey [40] by showing that the equilibria for a FPA (i.e., $k = m = 1$) are not generally efficient.

B Supplementary Material for Section 3

B.1 No DSIC and Strongly BPIC TFM Guarantees Non-Trivial Welfare

We show here that insisting on DSIC and strong BPIC implies that the TFM must output the empty set for some bid vectors, precluding it from getting any welfare guarantees. Since we are considering DSIC mechanisms, we consider the welfare achieved when bidders bid truthfully.

Theorem 4. *Any TFM that is DSIC and strongly-BPIC has a worst case welfare approximation of 0.*

The theorem follows immediately from the following lemma.

Lemma 1. *For any DSIC and strongly-BPIC TFM, for all $l > 0$ and $\mathbf{S}$, there exists a valuation vector $\mathbf{v}$ where $v_l > 0 \ \ \forall i \in [l]$ and $v_i = 0$ otherwise, such that $x_i(\mathbf{v}, \mathbf{S}) = 0 \ \forall i \in I$ i.e. the TFM confirms no transactions.*

Proof. We proceed by induction on l. For the base case of $l = 1$, let $\mathbf{v}^1 = (v_1, 0, ..., 0)$ and consider an arbitrary $\mathbf{S}$. Assume for the sake of contradiction that $x_1(\mathbf{v}^1, \mathbf{S}) = 1$ for all $v_1 > 0$. Since the TFM is DSIC, by Myerson's Theorem, we have that $p_1(\mathbf{y}(\mathbf{v}^1)) = 0$. It follows that $\pi_j(\mathcal{C}(\mathbf{y}(\mathbf{v}^1, \mathbf{S}))) = 0$ for all $j \in J$ since no transactions make any non-zero payments. However, then we have that all the block producers are indifferent between the equilibria $B_j = \emptyset$ for all $j \in J$ and $B_j = y_j(\mathbf{v}^1, \mathbf{S})$ contradicting the TFM being strongly-BPIC since these equilibria confirm different sets of bids. Hence for the TFM to be DSIC and strongly-BPIC there exists a $v_1 > 0$ s.t. $x_i(\mathbf{v}^1, \mathbf{S}) = 0 \ \forall i \in I$.

For the inductive hypothesis assume for all $\mathbf{S}$, there exists a valuation vector $\mathbf{v}^l = (v_1, ...v_l, 0, ..., 0)$ s.t. $x_i(\mathbf{v}^l, \mathbf{S}) = 0 \ \forall i \in I$. We then show for all $\mathbf{S}'$ there exists a $v' > 0$ s.t. for $\mathbf{v}^{l+1}$ with $\mathbf{v}_i^{l+1} = \mathbf{v}_i^l$ for $i \neq l + 1$ and $\mathbf{v}_{l+1}^{l+1} = v'$, $x_i(\mathbf{v}^{l+1}, \mathbf{S}') = 0 \ \forall i \in I$

Given a $\mathbf{S}'$ let $\mathbf{v}^l$ be a valuation vector such that $\mathbf{x}(\mathbf{v}^l, \mathbf{S}) = \emptyset$ where $\mathbf{S}$ is the projection of $\mathbf{S}'$ to transactions $i \neq l+1$. Now assume for the sake of contradiction that $\mathbf{x}(\mathbf{v}^{l+1}, \mathbf{S}') \neq \emptyset$ for all $v' > 0$. We claim this implies that we must have $l+1 \in \mathbf{x}(\mathbf{v}^{l+1}, \mathbf{S}') \ \forall v' > 0$. This is because if there is a $v' > 0$ where $l+1 \notin \mathbf{x}(\mathbf{v}^{l+1}, \mathbf{S}')$, we have two cases, either $\sum_j \pi_j(\mathbf{y}(\mathbf{v}^{l+1}, \mathbf{S}')) > 0$ or $\sum_j \pi_j(\mathbf{y}(\mathbf{v}^{l+1}, \mathbf{S}')) = 0$. The former case would imply the TFM is not strongly-BPIC, since in the instance with valuation function $\mathbf{v}^l$ and $\mathbf{S}$, the BPs could censor transaction $l + 1$ and replace it with a transaction with bid v'. Then the BP's following $\mathbf{y}$ under this modified valuation vector would pareto dominate following $\mathbf{y}$ under $\mathbf{v}^l$ since $l + 1$ isn't confirmed under $\mathbf{v}^{l+1}$, hence paying 0 fees, and some BPs get strictly positive compared to 0 revenue. Otherwise when $\sum_j \pi_j(\mathbf{y}(\mathbf{v}^{l+1}, \mathbf{S}')) = 0$, the BPs are indifferent between playing $\mathbf{y}$ or all proposing $B_j = \emptyset$ also violating strong-BPIC.

However, $l+1 \in \mathbf{x}(\mathbf{v}^{l+1}, \mathbf{S}') \ \forall v' > 0$ implies that $p_{l+1}(\mathbf{v}^{l+1}) = 0$ by Myerson's Theorem. Now again we have the same two cases, either $\sum_j \pi_j(\mathbf{y}(\mathbf{v}^{l+1}, \mathbf{S}')) > 0$

or $\sum_j \pi_j(\mathbf{y}(\mathbf{v}^{l+1}, \mathbf{S}')) = 0$. Since we still have that $l+1$ can costlessly be included, the same reasoning applies contradicting the TFM being strongly-BPIC. Thus for any $\mathbf{S}'$ there must exist a $v' > 0$ such that $x_i(\mathbf{v}^{l+1}, \mathbf{S}') = 0 \ \forall i \in I$.

B.2 No Strongly BPIC TFM Guarantees Optimal Welfare

We now show that no TFM can always be fully efficient at equilibrium when bidders draw their values from asymmetric distributions. We effectively reduce our setting to the case of auctioning a single item where the payment rule is forced to only be a function of the winning bidder's bid. We use revenue equivalence with a second price auction to show that efficient equilibrium can't be implemented with these types of payment rules.

Theorem 5. *For any TFM, there exists a game structure and a valuation distribution for which there is a Bayes-Nash equilibrium with expected welfare strictly less than the minimum possible.*

Consider the case where $n = 2$, $m = 1$, and $k = 1$. In this case, the TFM is equivalent to a single item auction with two bidders. Furthermore, $m = 1$ implies that the TFM's payment rule can only be a function of the winning bid. Thus the theorem follows immediately from the following lemma.

Lemma 2. *For any mechanism where the payment rule is a function only of the winning bid, there exists a valuation distribution for which there is a Bayes–Nash equilibrium whose expected welfare is strictly less than the maximum possible.*

Proof. Suppose, for the sake of contradiction, that there is a mechanism (x, p) whose payment rule depends only on the winning bid, i.e. $p_i(\mathbf{b}) = f(b_i)$ for some function f, such that *every* Bayes-Nash equilibrium in this mechanism is efficient. We will exhibit two different valuation instances and argue that the mechanism cannot have an efficient BNE in both instances simultaneously.

Instance 1: Consider an instance $\mathcal{I}_1$ with two bidders where

$$v_1 \sim \text{Uniform}\big([0, 100]\big) \quad \text{and} \quad v_2 \sim \text{Uniform}\big([0, 1]\big).$$

Let $\sigma(\cdot) = \big(\sigma_1(\cdot), \sigma_2(\cdot)\big)$ be a Bayes–Nash equilibrium under (x, p) that implements the *efficient outcome* for all realizations of (v_1, v_2). Since σ implements the efficient outcome for all values, it matches the allocation of a second-price auction. By revenue equivalence, the bidders' expected payments under σ in this mechanism must match their expected payments when bidding truthfully in a second-price auction.

- For bidder 1: Under the assumption that σ is efficient, we have that the probability bidder 1 wins is $Pr[v_1 > v_2]$. Thus $E_{v_2}[p_1\big(\sigma_1(v_1), \sigma_2(v_2)\big) \mid v_1] = Pr[v_1 > v_2] \cdot f\big(\sigma_1(v_1)\big)$. On the other hand, bidder 1's expected payment in

a second price auction is $Pr[v_1 > v_2] \cdot E[v_2 \mid v_2 < v_1]$. Note that $E[v_2 \mid v_2 < v_1] = \min\{\frac{1}{2}, \frac{v_1}{2}\}$ giving us

$$f\big(\sigma_1(v_1)\big) = E\big[v_2 \mid v_2 < v_1\big] = \min\{\frac{1}{2}, \frac{v_1}{2}\} \implies \sigma_1(v_1) \in f^{-1}\big(\min\{\frac{1}{2}, \frac{v_1}{2}\}\big)$$

– For bidder 2: By a symmetric argument, for bidder 2 we have

$$f\big(\sigma_2(v_2)\big) \;=\; E\big[v_1 \mid v_1 < v_2\big] = \frac{v_2}{2} \implies \sigma_2(v_2) \;\in\; f^{-1}\Big(\tfrac{v_2}{2}\Big)$$

Instance 2: Now consider another two-bidder instance $\mathcal{I}_2$, where

$$v_1 \sim \text{Uniform}\big([0, \tfrac{3}{2}]\big) \quad \text{and} \quad v_2 \sim \text{Uniform}\big([0, 100]\big).$$

Let $\sigma'(\cdot) = \big(\sigma'_1(\cdot), \sigma'_2(\cdot)\big)$ be an efficient BNE for this second instance. A parallel revenue-equivalence argument tells us that:

– For bidder 1,

$$f\big(\sigma'_1(v_1)\big) \;=\; E\big[v_2 \mid v_2 < v_1\big] \;=\; \tfrac{v_1}{2} \implies \sigma'_1(v_1) \;\in\; f^{-1}\Big(\tfrac{v_1}{2}\Big),$$

– For bidder 2,

$$f\big(\sigma'_2(v_2)\big) \;=\; E\big[v_1 \mid v_1 < v_2\big] = \min\Big\{\tfrac{3}{4}, \tfrac{v_2}{2}\Big\} \implies \sigma'_2(v_2) \;\in\; f^{-1}\Big(\min\{\tfrac{3}{4}, \tfrac{v_2}{2}\}\Big)$$

We claim that it is impossible for both σ to be an efficient BNE in $\mathcal{I}_1$ *and* σ' to be an efficient BNE in $\mathcal{I}_2$. To see why, consider the following deviation arguments:

1. *Deviation of bidder 2 in instance* $\mathcal{I}_1$. Suppose in $\mathcal{I}_1$ that bidder 2, whenever $v_2 > \frac{3}{4}$, chooses a random sample $a \sim \text{Uniform}([0, 100])$ and then plays $\sigma'_2(a)$ instead of $\sigma_2(v_2)$. Call this strategy $\tilde{\sigma}_2(\cdot)$ If σ is indeed a BNE in $\mathcal{I}_1$, this deviation cannot increase bidder 2's expected utility for *any* v_2.
   Now consider the specific value $v_2 = \frac{7}{8}$. Under σ, bidder 2's expected utility is
   $$E_{v_1}\big[u_2\big(\sigma_1(v_1), \sigma_2(\tfrac{7}{8})\big)\big] \;=\; \Pr\big[v_1 < \tfrac{7}{8}\big] \times \frac{7}{16} \;<\; \frac{1}{200},$$
   By deviating to the $\tilde{\sigma}_2$ by sampling a and playing $\sigma'_2(a)$, bidder 2's expected utility when $v_2 = \frac{7}{8}$ is
   $$E_{v_1,a}\big[u_2\big(\sigma_1(v_1),\, \sigma'_2(a)\big)\big] \;=\; \Pr\big[x\big(\sigma_1(v_1),\, \sigma'_2(a)\big) = 2\big] \;\times\; \tfrac{1}{8}.$$
   For σ to remain an equilibrium, we must therefore have
   $$\Pr\big[x\big(\sigma_1(v_1),\, \sigma'_2(a)\big) = 2\big] \;<\; \frac{1}{25}.$$

2. *Deviation of bidder 1 in instance $\mathcal{I}_2$.* Next, suppose in $\mathcal{I}_2$ that bidder 1, whenever $v_1 > \frac{1}{2}$, samples $b \sim \text{Uniform}([0, 100])$ and plays $\sigma_1(b)$ from instance $\mathcal{I}_1$. From the probability bound above, whenever $v_1 > \frac{1}{2}$ the probability that bidder 1 wins against $\sigma_2'(v_2)$ is at least $\frac{24}{25}$. In particular, at $v_1 = \frac{3}{4}$, bidder 1's expected utility from this deviation is at least

$$\left(\frac{3}{4} - \frac{1}{2}\right) \times \frac{24}{25} \;=\; \frac{6}{25},$$

which is substantially larger than the at most $\frac{1}{100}$ expected utility bidder 1 achieves under playing σ_1', contradicting σ' being a Bayes–Nash equilibrium.

Since we have derived a profitable deviation in one instance assuming that the other instance has an efficient BNE, it follows that σ and σ' cannot both be equilibria of their respective instances under the same payment rule f. Thus any mechanism where the payment rule is only a function of the winning bid must have an inefficient Bayes–Nash equilibrium either in instance $\mathcal{I}_1$ or $\mathcal{I}_2$.

B.3 Review of Relevant Matroid Theory

The matroid structure of feasible allocations play an important role in the incentive-compatibility and welfare guarantees of the FPA-EQ mechanism in Section 3. We review in this appendix the properties of matroids that are relevant to our results.

Definition 1 (Matroid). *A* matroid *is a set system $(X, \mathcal{I})$ with* ground set X *and* independent sets $\mathcal{I} \subseteq 2^X$ *that satisfies:*

1. *$\mathcal{I}$ is non-empty.*
2. *(Downward closure) If $A' \in \mathcal{I}$ and $A \subseteq A'$, then $A \in \mathcal{I}$.*
3. *(Exchange property) If $A, A' \in \mathcal{I}$ with $|A'| > |A|$, then there exists $x \in A' \setminus A$ such that $A \cup \{x\} \in \mathcal{I}$.*

For a game structure $(I, J, \mathbf{S})$, call a subset $A \subseteq I$ of transactions *feasible* if there exists a feasible allocation $(B_1, \ldots, B_m)$ that includes precisely the transactions in A.

Proposition 6 *For every game structure $(I, J, \mathbf{S})$, the subset of feasible transactions forms a matroid over I.*

Proof. (Sketch.) Non-emptyness holds because the empty set of transactions is feasible. Downward closure holds because removing transactions from a feasible allocation cannot destroy feasibility. The exchange property holds from an alternating path argument in the spirit of transversal matroids (see [27, Theorem 1.6.2]).

Matroids have a long list of nice properties.

Proposition 7 (Lexicographic Optimality) *Let $(X, \mathcal{I})$ be a matroid for which each ground set element $x \in X$ has a nonnegative weight w_x. If $A, A' \in \mathcal{I}$ are two maximum-weight independent sets, then the multi-sets of non-zero element weights of A and A' are identical.*

Proof. (Sketch.) We can assume that A, A' are maximal independent sets, extending them with (necessarily zero-weight) elements if necessary. Due to the matroid structure (see [27, Corollary 1.2.5]), there is a sequence $A = A_0, A_1, \ldots, A_l = A'$ such that: (i) each set in the sequence belongs to $\mathcal{I}$; and (ii) each set in the sequence is derived from the previous one by swapping one element for another. Because both A and A' are maximum-weight independent sets, so are all the intermediate sets of the sequence. Thus, each swap of the sequence exchanges one element for another with equal weight. Thus, the multi-sets of element weights of A and A' are identical.

The following proposition establishes a "revenue covering" property (in a sense similar to Hartline et al. [16]) for matroids.

Proposition 8 (Revenue Covering) *Let $(X, \mathcal{I})$ be a matroid for which each ground set element $x \in X$ has a nonnegative weight w_x, and let A^* denote a maximum-weight independent set. Let $t_x(\mathbf{w}_{-x})$ denote the minimum value of x's weight such that, holding the weights $\mathbf{w}_{-x}$ of the other elements fixed, x belongs to a maximum-weight independent set. Then, for every independent set A,*

$$\sum_{x \in A^*} w_x \geq \sum_{x \in A} t_x(\mathbf{w}_{-x}). \tag{3}$$

Proof. (Sketch.) By the optimality of the greedy algorithm for matroids (see [27, Theorem 1.2.6]), A^* remains a maximum-weight independent set even after the weight of each element $x \notin A^*$ is increased to $t_x(\mathbf{w}_{-x})$. Given that $w_x \geq t_i(\mathbf{w}_{-x})$ for all $x \in A^*$ (by the definition of the t_i's), the inequality (3) follows from the optimality of A^*.

B.4 Proof of Theorem 2

Tightness of Theorem 2. Before providing the proof of Theorem 2, we show that the welfare bound of the theorem is tight.

Proposition 9 (Theorem 2 Is Tight) *There exists a game structure, valuation distribution $\mathbf{D}$, and an inclusion-rule-respecting subgame perfect equilibrium of the FPA-EQ TFM with expected welfare $1 - \frac{1}{e}$ times the expected maximum welfare.*

Proof. Take $I = \{1, 2, 3\}$, $J = \{1\}$, and $S_1 = \{1, 2, 3\}$. The support of the joint distribution $\mathbf{D}$ is the valuation vectors of the form $(1, x, x)$ for $x \in [0, 1 - \frac{1}{e}]$. The marginal distribution of the common value of v_2 and v_3 is given by the

CDF $F(x) = \frac{1}{e}\frac{1}{1-x}$ on $[0, 1-\frac{1}{e}]$. Thus, with probability 1, the maximum-possible welfare is 1 (achieved by including the first transaction).

One can check that the following is an IRR SPE. The BP includes the highest-bidding transaction, breaking ties in favor of the first transaction. The first user always bids 0. The second and third users always bid truthfully. A calculation shows that the expected welfare of this IRR SPE is exactly $1 - \frac{1}{e}$.

Equivalence of IRR SPE with BNE of Matroid Auctions. The first step of the proof of Theorem 2 is to establish a correspondence between the IRR SPE of the FPA-EQ TFM and the Bayes-Nash equilibria of winner-pays-bid matroid auctions. Here's what we mean by the latter: For a set of users U and a matroid $(U, \mathcal{I})$ (see Definition 1), the corresponding winner-pays-bid matroid auction is defined by:

1. Simultaneously, each user $i \in U$ submits a nonnegative bid b_i.
2. The mechanism chooses an independent set $A \in \mathcal{I}$ that maximizes the sum $\sum_{i \in A} b_i$ of the bids of the included users, breaking ties arbitrarily. Users of A win and the other users lose.
3. Each winner $i \in A$ pays its bid b_i.

For example, a first-price single-item auction corresponds to the special case of a winner-pays-bid matroid auction in which the set $\mathcal{I}$ contains only the empty set and all the singleton sets.

Every strategy of a user $i \in I$ in the extensive-form game induced by a TFM (see Section 2.2) induces a bidding strategy σ_i, with $\sigma_i(v_i)$ defined as the action (or distribution over actions) taken by user i in the information set corresponding to the realization v_i of its valuation. Meanwhile, every profile of BP strategies induces an allocation rule $\mathbf{x}$, where $x_i(\mathbf{b})$ denotes the probability (over any randomness in BPs' strategies) that user i's transaction is confirmed when the user bid vector is $\mathbf{b}$.

For an arbitrary allocation rule $\mathbf{x}$, the corresponding (single-shot) *winner-pays-bid mechanism* $(\mathbf{x}, \mathbf{p})$ accepts nonnegative bids from users; chooses a feasible allocation from a probability distribution such that each user $i \in U$ is allocated with probability $x_i(\mathbf{b})$; and charges b_i to each allocated user and 0 to each unallocated user. Fron the discussion above, we have:

Proposition 10 *Every strategy profile in the extensive-form game induced by the FPA-EQ TFM is user-outcome-equivalent to the induced bidding strategies $\sigma_1(v_1), \ldots, \sigma_n(v_n)$ in the winner-pays-bid mechanism induced by the allocation rule that is induced by BP's strategies.*

By "user-outcome-equivalent" we mean that, for each user, the probability of allocation and the payment conditional on allocation are identical in the two scenarios. Note that this notion of equivalence preserves the expected welfare.

We now specialize Proposition 10 to the case of IRR SPE. First, the IRR condition means that the allocation rule $\mathbf{x}$ induced by the BP strategies is the one that, given users' bids, selects the feasible allocation with the maximum-possible

sum of bids (breaking ties arbitrarily). Thus, the winner-pays-bid mechanism induced by an IRR SPE is a matroid auction. Second, the equilibrium condition for users' strategies in the IRR SPE translate to the Bayes-Nash equilibrium conditions for the induced bidding strategies $\sigma_1(v_1), \ldots, \sigma_n(v_n)$ in this matroid auction.

Lemma 3. *For every game structure and valaution distribution, every IRR SPE of the FPA-EQ TGM is user-outcome-equivalent to a Bayes-Nash equilibrium of a winners-pay-bid matroid auction (with the same valuation distribution).*

As noted above, user-outcome-equivalence implies that the expected welfare of an IRR SPE and the corresponding Bayes-Nash equilibrium are the same.

The Price of Anarchy of Winner-Pays-Bid Matroid Auctions. Given the equivalence established above, we can complete the proof of Theorem 2 by showing the following:

Theorem 11 (Matroid Auctions Have Only Near-Optimal Equilibria). *For every matroid $(U, \mathcal{I})$ and valuation distribution, every Bayes-Nash equilibrium of the corresponding winner-pays-bid matroid auction has expected welfare at least $1 - \frac{1}{e}$ times the expected maximum welfare.*

In turn, proving Theorem 11 reduces to showing that winner-pays-bid matroid auctions are "smooth" in a suitable sense. The following definition and theorem are essentially due to Lucier and Paes Leme [22]; we follow the formalism in Roughgarden et al. [30, Definition 4.5; Theorem 4.6].

Definition 2 (Smooth Auction with Private Deviations [22,30]). *For parameters $\lambda \geq 0$ and $\mu \geq 1$, an auction with allocation rule $\mathbf{x}$ and payment rule $\mathbf{p}$ is* (λ, μ)-smooth with private deviations *if for every valuation profile $\mathbf{v}$ there exist probability distributions $D_1^*(v_1), \ldots, D_n^*(v_n)$ over bids such that, for every bid profile $\mathbf{b}$,*

$$\sum_i \mathbf{E}_{b_i{}^* \sim D_i^*(v_i)}[u_i(b_i{}^*, \mathbf{b}_{-i})] \geq \lambda \cdot \sum_i v_i \cdot x_i^*(\mathbf{v}) - \mu \cdot Rev(\mathbf{b}). \tag{4}$$

In (4), $u_i(\mathbf{b}) = v_i \cdot x_i(\mathbf{b}) - p_i(\mathbf{b})$ denotes quasi-linear utility (as in (2)), $x^*(\mathbf{v})$ denotes the characteristic vector of a welfare-maximizing feasible solution with respect to valuation profile $\mathbf{v}$, and $Rev(\mathbf{b}) = \sum_i p_i(\mathbf{b})$ denotes the auction's revenue when the bid vector is $\mathbf{b}$. The "private deviations" qualifier refers to the fact that each bid distribution D_i^* is permitted to depend only on user i's valuation v_i, and not on the full valuation profile $\mathbf{v}$.

Theorem 12 (Smoothness Implies Price-of-Anarchy Bounds [22,30]). *If an auction is (λ, μ)-smooth, then for every distribution $\mathbf{D}$ over players' valuations, every Bayes-Nash equilibrium of the auction has expected welfare at least λ/μ times the expected maximum welfare.*

In light of Theorem 12, the following lemma implies Theorem 11 (and hence, by Lemma 3, Theorem 2).

Lemma 4 (Matroid Auctions Are Smooth). *For every matroid $(U, \mathcal{I})$, the corresponding winner-pays-bid matroid auction is $(1 - \frac{1}{e}, 1)$-smooth with private deviations.*

Proof. The proof incorporates elements of the smoothness analysis of first-price auctions by Syrgkanis and Tardos [38] and the revenue covering analysis of matroid auctions by Hartline et al. [16]. Fix a matroid $(U, \mathcal{I})$; let $\mathbf{x}$ and $\mathbf{p}$ denote the allocation and payment rules of the corresponding winner-pays-bid auction. Fix a valuation profile $\mathbf{v}$ for the users of U. For each $i \in U$, define D_i^* as the distribution with density $1/(v_i - x)$ on support $[0, (1 - 1/e)v_i]$.

To verify the smoothness inequality (4), fix a bid vector $\mathbf{b}$. Denote by $t_i(\mathbf{b}_{-i})$ the minimum value z for i's bid such that $x_i(z, \mathbf{b}_{-i}) = 1$. To bound $\mathbf{E}_{b_i{}^* \sim D_i^*}[u_i(b_i{}^*, \mathbf{b}_{-i})]$, we consider two cases. First, if $v_i \cdot (1 - 1/e) \leq t_i(\mathbf{b}_{-i})$, then because $b_i{}^* \sim D_i^*$ is at most v_i with probability 1 and $(\mathbf{x}, \mathbf{p})$ is ex post individually rational, $\mathbf{E}_{b_i{}^* \sim D_i^*}[u_i(b_i{}^*, \mathbf{b}_{-i})] \geq 0$. Second, if $v_i \cdot (1 - 1/e) > t_i(\mathbf{b}_{-i})$, then by similar reasoning,

$$\mathbf{E}_{b_i{}^* \sim D_i^*}[u_i(b_i{}^*, \mathbf{b}_{-i})] \geq \int_{t_i(\mathbf{b}_{-i})}^{(1-1/e)v_i} (v_i - z) \cdot \frac{dz}{v_i - z} = \left(1 - \frac{1}{e}\right) v_i - t_i(\mathbf{b}_{-i}).$$

In this case, because the left-hand side is nonnegative and $x_i(\mathbf{v}) \in [0, 1]$, we also have

$$\mathbf{E}_{b_i{}^* \sim D_i^*}[u_i(b_i{}^*, \mathbf{b}_{-i})] \geq \left(1 - \frac{1}{e}\right) v_i \cdot x_i^*(\mathbf{v}) - t_i(\mathbf{b}_{-i}) \cdot x_i^*(\mathbf{v}).$$

Summing this inequality over all $i \in I$ and applying Proposition 8, we have

$$\begin{aligned}\mathbf{E}_{b_i{}^* \sim D_i^*}[u_i(b_i{}^*, \mathbf{b}_{-i})] &\geq \left(1 - \frac{1}{e}\right) \sum_{i \in I} v_i \cdot x_i^*(\mathbf{v}) - \sum_{i \in I} t_i(\mathbf{b}_{-i}) x_i^*(\mathbf{v}) \\ &\geq \left(1 - \frac{1}{e}\right) \sum_{i \in I} v_i \cdot x_i^*(\mathbf{v}) - \sum_{i \in I} p_i(\mathbf{b}),\end{aligned}$$

which shows that (4) holds with $\lambda = 1 - \frac{1}{e}$ and $\mu = 1$, completing the proof.