Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30

Brian F. Tivnan,[1, 2, 3] David Rushing Dewhurst,[1, 2, 4, 5] Colin M. Van Oort,[1, 2, 4, 5] John H. Ring IV,[1, 2, 4, 5] Tyler J. Gray,[2, 3, 5] Brendan F. Tivnan,[2] Matthew T. K. Koehler,[1] Matthew T. McMahon,[1] David Slater,[1] Jason Veneman,[1] and Christopher M. Danforth[2, 3, 5]

[1] *The MITRE Corporation, McLean, VA 22102*
[2] *Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405*
[3] *Department of Mathematics and Statistics, University of Vermont, Burlington, VT 05405*[*]
[4] *Department of Computer Science, University of Vermont, Burlington, VT 05405*
[5] *Computational Story Lab, University of Vermont, Burlington, VT 05405*

Using the most comprehensive source of commercially available data on the US National Market System, we analyze all quotes and trades associated with Dow 30 stocks in 2016 from the vantage point of a single and fixed frame of reference. Contrary to prevailing academic and popular opinion, we find that inefficiencies created in part by the fragmentation of the equity marketplace are widespread and potentially generate substantial profit for agents with superior market access. Information feeds reported different prices for the same equity—violating the commonly-supposed economic behavior of a unified price for an indistinguishable product—more than 120 million times, with "actionable" latency arbitrage opportunities totaling almost 64 million. During this period, roughly 22% of all trades occurred while the SIP and aggregated direct feeds were dislocated. The current market configuration resulted in a realized opportunity cost totaling over $160 million when compared with a single feed, single exchange alternative—a conservative estimate that does not take into account intra-day offsetting events.

I. INTRODUCTION

The Dow Jones Industrial Average, colloquially known as the Dow 30, is a group of 30 equity securities (or stocks) selected by S&P Dow Jones Indices that reflects a broad cross-segment of the US economy (all industries except for utilities and transportation) [1]. The Dow 30 is one of the best known indices in the US and is used as a barometer of the economy by talk shows and financial publications alike. Thus, while the group of securities that composes the Dow 30 is in some sense an arbitrary collection, it derives economic import from its ascribed characteristics. We study the behavior of these securities as traded in modern US equity markets, known as the National Market System (NMS). Contrary to popular perception, where the stock market may be viewed as a monolithic entity, the NMS is comprised of 13 networked exchanges that are coupled by regulation and form a canonical example of a complex system. Adding another layer of depth and complexity the NMS supports a diverse ecosystem of market participants, ranging from small retail investors to institutional financial firms, such as investment and custodial banks, to high-frequency trading (HFT) firms and designated market makers.

We do not attempt to unravel and attribute the activity of each of these actors here; several others have attempted to classify such activities with varying degrees of success in diverse markets [2–4]. We take a simpler, first-principles approach. We compile an exhaustive catalog of every dislocation and latency arbitrage opportunity, defined as a nonzero pairwise difference between the National Best Bid and Offer (NBBO) as observed via the Securities Information Processor (SIP) feed, which displays the best prices in the lit market, and Direct Best Bid and Offer (DBBO) as observed via the consolidation of all direct feeds. The SIP and consolidation of all direct feeds are representative of the displayed quotes from the national exchanges (lit market). Additionally, we catalog every trade that occurred in the National Market System among the Dow 30 in 2016 in order to investigate the relationship between trade execution and latency arbitrage opportunities; we compile a dataset of all trades that may lead to a non-zero realized opportunity cost, a so-called "differing trade". Counter to popular (e.g., [5]) and prevailing academic opinion [6, 7], the number of dislocations, the resulting potential arbitrage opportunities, and differing trades are far from zero. Indeed, we tally more than 120 million latency arbitrage opportunities, an event derived from dislocations between the SIP NBBO and DBBO, on the Dow 30 in 2016, as shown in Table I. Approximately 65 million of those opportunities are what we term *actionable*, meaning that we estimate there is a high likelihood that an appropriately equipped market participant could realize arbitrage profits due to the existence of such a latency arbitrage opportunity. (We discuss actionability in detail in Sec. III C.) The realized opportunity cost associated with these opportunities is no less impressive in magnitude; an estimated $160 million USD was lost in opportunity cost by the totality of market participants due to information asymmetry between the SIP and Direct feeds in the stocks of the Dow 30 in 2016. We calculate the ROC using the NBBO price as the baseline; deviations from this price contribute to the ROC with positive sign (if the direct

[*] Corresponding authors: Brian Tivnan and Christopher Danforth. Emails: btivnan@mitre.org and chris.danforth@uvm.edu.

feed gives a worse price than the SIP) or with negative sign (if the direct feed gives a better price than the SIP).

In what follows, we show that:

a. There is no other study able to authoritatively claim accurate statistics of both latency arbitrage opportunity (frequency, number, and magnitude) and realized opportunity cost. All other studies have used less comprehensive data [8, 9] or have not sought to answer both questions [10]. On the contrary, we use data effectively identical to that used by the Securities and Exchange Commission, the most exhaustive data available for purchase (see Sec. III C below). In addition to its comprehensive nature, this data was collected from the viewpoint of a unified observer, a single and fixed frame of reference co-located from within the Nasdaq data center in Carteret, N.J.

b. The National Market System features inherent inefficiencies. The fractured nature of the auction mechanism—continuous double auction operating on no fewer than 13 heterogeneous exchanges, to say nothing of Alternative Trading Systems (ATSs), also known as dark pools—is a consistent generator of market inefficiency as measured by the systematic ability of a particular class of market participants (those with access to the direct feed information) to reliably make higher returns than other market participants. Such information is not private as such (it is available for purchase) but it also may not drive prices of equities toward a new equilibrium, as differences in equity prices on differing information feeds often persist and recur on a daily basis.

c. These inefficiencies are not merely a theoretical construct. As mentioned above, the realized opportunity cost arising from these inefficiencies are enormous, and at least $160 million USD of realized opportunity cost was incurred during the period of study at which one of two differing information feeds (SIP or direct feed) quoted a better price for the liquidity consuming market participant.

II. THEORY AND LITERATURE REVIEW

A. Theory

The general theory of financial markets is a vast subject; we can do it no justice here. We thus restrict our attention to a subject that has a direct bearing on the empirical results presented in this work: the so-called "efficiency" of markets and the ability of market participants to make sustained economic profits through market activity. Much had been written on the theory of financial markets before the efficient markets hypothesis (EMH) proposed by Fama [11], but this work has left an indelible mark upon the entire body of thought. Extensive econometric analysis of financial transaction data in the finest granularity available at the time (late 1960s and early 1970s) strongly suggested that individual equity prices, and thus equity markets, fully incorporated all relevant publicly available past information—the typical definition of market efficiency. (A stronger version of this hypothesis proposes the incorporation of private information as well *vís a vís* insider trading and other mechanisms.) While there are notable exceptions to this hypothesis in both the empirical finance and econophysics literature [12], including (but not limited to) price characteristics of equities in emerging markets [13], the apparent existence of momentum in the trajectories of equity prices [14], and speculative asset bubbles (problematic from the strong-EMH point of view), recent work by Fama and French has demonstrated that this hypothesis is largely still valid [14] when price time series are

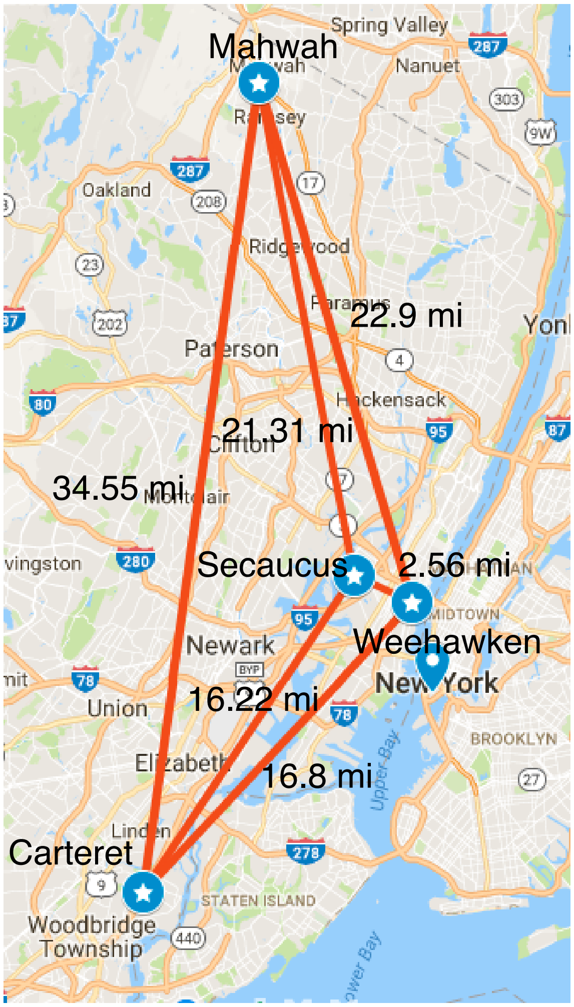

FIG. 1. A geographic depiction of the four major NMS data centers and approximate straight-line distances between them.

1	Total Opportunity Cost	$160,213,922.95
2	SIP Opportunity Cost	$122,081,126.40
3	Direct Opportunity Cost	$38,132,796.55
4	Trades	392,101,579
5	Differing Trades	87,432,231
6	Traded Value	$3,858,963,034,003.48
7	Differing Traded Value	$900,535,924,961.72
8	Fraction of differing trades	0.2230
9	Fraction of differing notional	0.2334
10	Ratio of (9) over (8)	1.0465

TABLE I. Market participants stood to gain $160 million via the use of additional data feeds, with SIP exclusive subscribers missing out on $122 million and Direct exclusive subscribers missing out on $38 million. The SIP feed consistently offered worse prices than the aggregate direct feed for liquidity demanding market participants during periods of dislocation, with a $84 million net difference in opportunity cost. Statistics 9, 10, and 11 indicate that trades occurring during dislocations involve approximately 5% more value per trade on average than those that occur while feeds are synchronized. Thus, market participants may be more heavily impacted by the existence of arbitrage opportunities than a mean-field analysis would suggest. The values reported above are sums of daily observations, except for statistics 9-11. We note that, since positive (favoring the SIP) and negative (favoring the direct feeds) ROC can cancel out in the above summary due to intraday effects, the above opportunity cost figures may be underestimates of the true values.

examined at timescales on the order of 20 minutes (or longer) over a sufficiently long period of time. These conditions are patently not expressed within the NMS, which has been shown to operate at speeds far beyond that of normal human cognition [15] and consists of fragmented exchanges [16] that engender some amount of arbitrage opportunities.

The modern U.S. stock market observed at its native operating frequency does not satisfy even the most lenient preconditions for weaker forms of the EMH, with some research indicating that large and surprising deviations occur from what might otherwise be predicted by proponents of the EMH [17]. Other theories of market efficiency have been developed, such as the adaptive markets hypothesis (AMH) presented by Lo [18] in which market agents are conjectured to adapt to evolving market conditions using learned heuristics. More permissive theories on market efficiency allow for the existence of phenomena such as dislocations and arbitrage opportunities due to reaction delays, faulty heuristics, etc, along with nontrivial effects of information asymmetry [19].

At the opposite end of the human-automation spectrum, if all market participants employed rationally-programmed, artificially intelligent strategies with similar processing capabilities and market latencies—that is, if the market was algorithmically-saturated—we might expect to observe perfectly rational behavior in financial markets at all times in the case of a non-fractured market. There is already evidence for this claim in automated auctions used in Internet advertising [20].

But the market of today is a mixture of these pure states. While many financial firms do use algorithmic trading as a core component of their day-to-day operations, and still others have algorithmic trading as their *raison d'être*, many more yet have no algorithmic aspect to their trading activity at all, to say nothing of the legions of retail investors who trade for leisure or attempt to grow their retirement savings.

The proposition that increased density of HFT is associated with, and may lead to, less-efficient markets is not merely conjecture. Here, we adopt a prevailing definition of HFT, which is strategy driven and done by computers at extremely fast speeds [16]. Indeed, recent work by O'Hara [16], Bloomfield [21], and others [22] has provided evidence that relatively well-informed actors, such as HFT and other algorithmic traders, are able to consistently beat market returns as a result of both structural advantages and the actions of less-informed actors—so called "noise traders" [23]. This compendium of results points to a synthesis of the competing viewpoints of market efficiency: that financial markets do indeed incorporate all relevant publicly available past information—when the market mechanisms are observed at roughly the speed of human cognition and reaction—and so in this sense the EMH is validated. However, when these mechanisms are observed at timescales on the order of $\sim 1s$ or less, this hypothesis fails to hold due to asymmetries in processing capability or HFT's exploitation of market design.

B. Empirical studies of equity markets

Several authors have considered the questions of calculating price dislocations, defined as a best bid and offer (BBO) that differs either based on exchange or on information source, and the potential resultant arbitrage. There is general agreement that price dislocations do not have a substantial effect on retail and other small investors, as these investors tend to trade infrequently and in relatively small quantities. Conclusions differ on

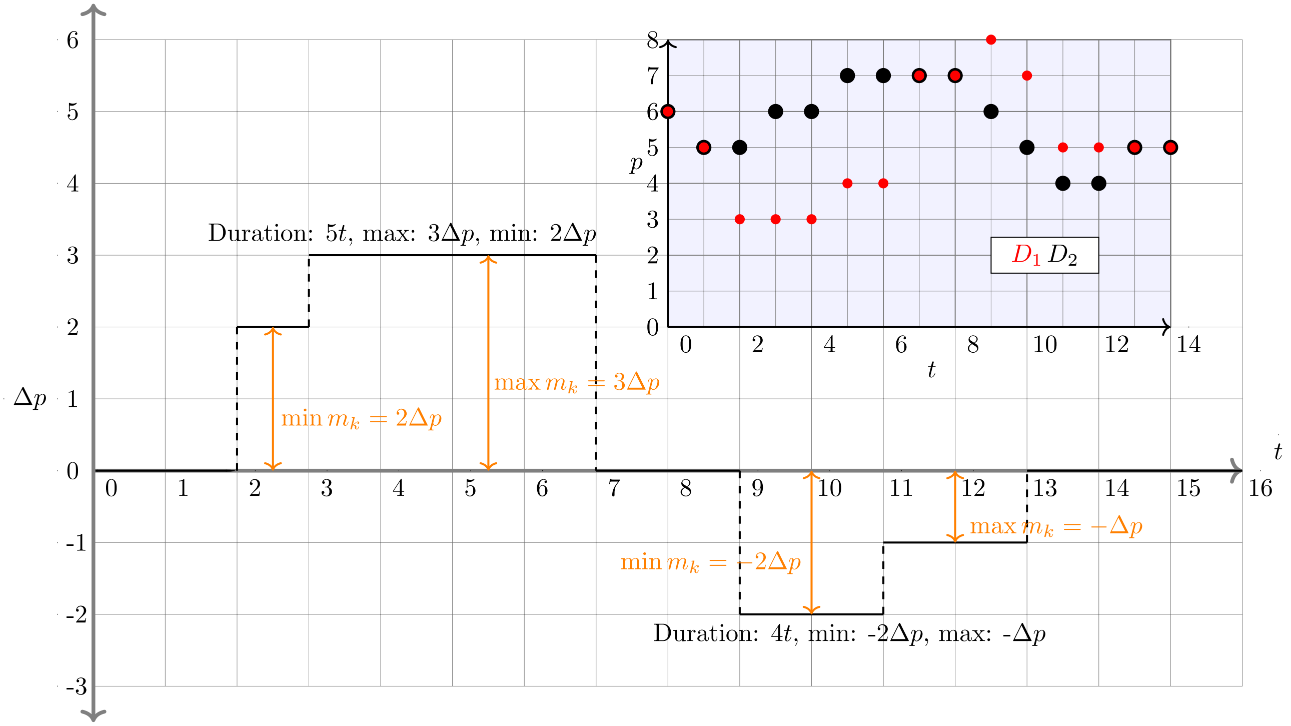

FIG. 2. Diagram of two latency arbitrage opportunities. The inset plot shows the time series of best quotes that generate the arbitrage opportunities. Where the time series diverge from the same value, a latency arbitrage opportunity occurs. We have deliberately not placed units on t, Δp, and p to indicate that latency arbitrage opportunities can occur in any market in which there are differing information feeds, not just in the NMS, though we do assume that these quantities are quantized. In the case of the NMS, we take t in units of μs and Δp in units of \$0.01. The reader will note that this diagram represents one side (either bid or offer) of the book. The marker sizes in the inset time series subplot do not denote any particular aspect of the time series; the size of the D_2 marker is larger than that of the D_1 marker simply for visual distinction.

the effect of dislocations on investors who trade more frequently and/or in larger quantities, such as institutional investors and trading firms.

While we are not aware of any study that seeks to answer precisely the questions considered here, several do attempt to quantify price dislocations in the NMS. Ding, Hanna, and Hendershot (DHH) [9] considered the effect of differential information speed on price signals and resultant arbitrage opportunities, although in the context of correlated features of multiple assets rather than pure latency arbitrage. They found that dislocations (and hence latency arbitrage opportunities), far from being rare, occur multiple times per second and tend to last between one and two milliseconds. In addition, DHH find that dislocations are associated with higher prices, volatility, and trading volume. Bartlett and McCrary [8] also attempted to quantify the frequency and magnitude of dislocations. However, direct feed data was not used and so the existence of dislocations was estimated using Securities Information Processor (SIP) data; their results cannot be directly compared to those presented here due to their lack of comprehensive data. Wah [10] calculated the potential arbitrage opportunities generated by latency arbitrage on the S&P 500 in 2016 using data from the SEC's MIDAS platform [24]. A study by the TABB Group of trade execution quality on midpoint orders in ATSs also noted the existence of latency between the SIP and direct data feeds, as well as the existence of intra-direct feed latency, due to differences in exchange and ATS software and other technical capabilities [25].

Other authors have also analyzed the effect of high-frequency trading on market microstructure. O'Hara [16] provides a high-level overview of the modern-day equity market and in doing so outlines both the possibility of latency arbitrage opportunities arising from differential information speed. Contrary to many authors, Angel [6, 7] states that price dislocations and associated latency arbitrage opportunities are relatively rare occurrences. Carrion [26] also provides evidence of high-frequency trading strategies' effectiveness in modern-day equity markets via successful, intra-day market timing. Budish [22] notes that high-frequency trading firms successfully perform statistical arbitrage (pairs trading) in the equities market, and ties this phenomenon to the continuous double auction price discovery mechanism omnipresent in the current market structure. Menkveld [27] analyzed the role of HFT in market making, finding that HFT market making activity correlates negatively with long-run price movements and providing some evi-

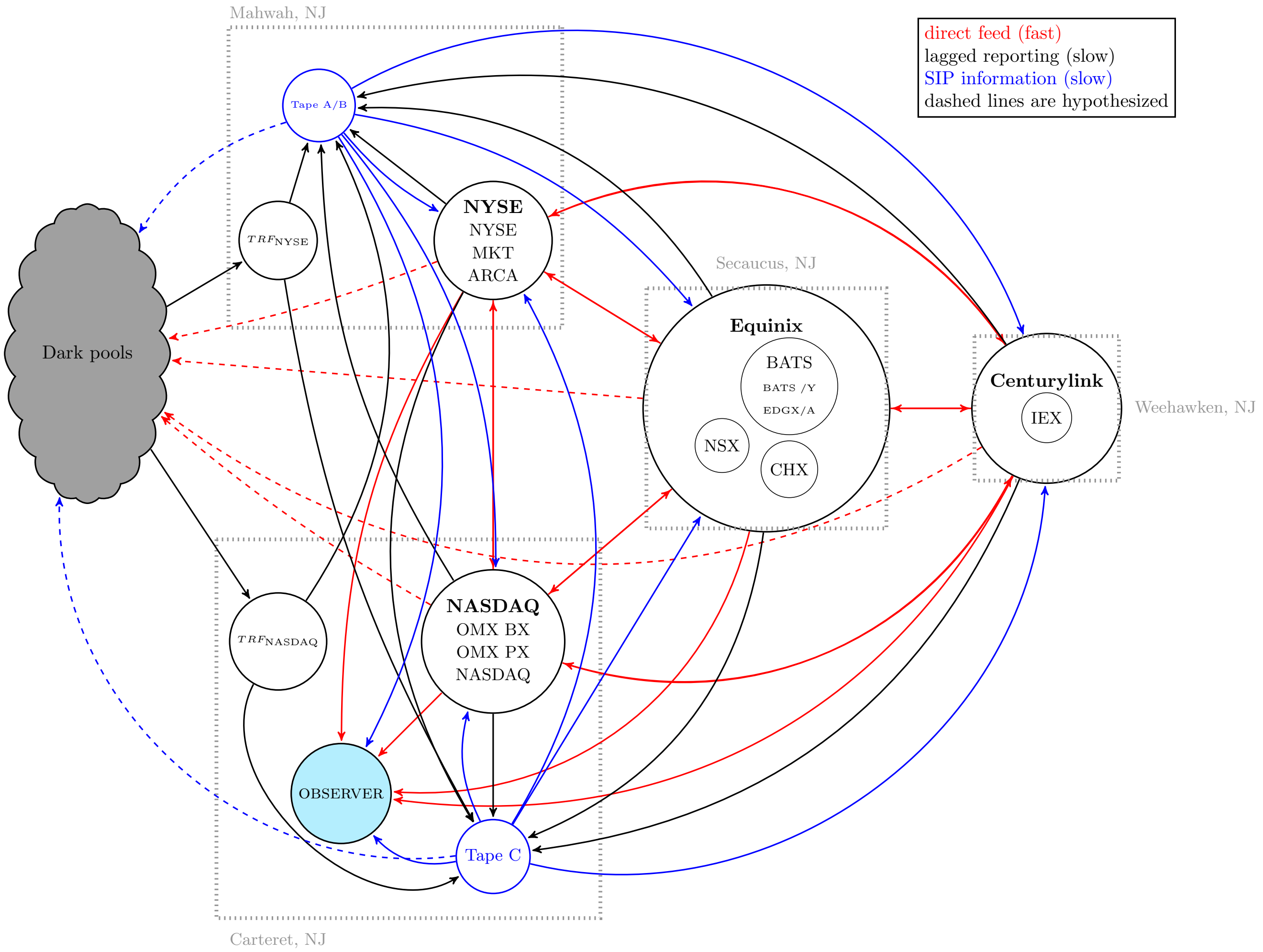

FIG. 3. The NMS (lit market and ATSs / dark pools) as implied by the comprehensive market data. As we do not have the specifications of inter-market center communication mechanisms and have minimal knowledge of intra-market center communication mechanisms, we simply classify information as having high latency, as the SIP and lagged information heading to the SIP do, or low latency, as the information on the direct feeds does. Note the existence of the central observer, in our case based in Carteret NJ. Without this central observer it would be impossible to rectify the differing timestamps originating from each exchange and so any statement made about dislocations / latency arbitrage could not be taken at face value; see [8] for an example of the traps into which one can fall.

dence that HFT market makers may exacerbate price fluctuations. Kirilenko [2] provided an important classification of active trading strategies on the Chicago Mercantile Exchange E-mini futures market, which can be useful in creating statistical or agent-based models of market phenomena.

For a more comprehensive review of the literature on high frequency trading and modern market microstructure more generally, we refer the reader to Goldstein et al. [28] or Chordia et al. [29]. Arnuk and Saluzzi [30] provide a monograph-level overview of the subject from the viewpoint of industry practitioners.

III. DESCRIPTION OF EXCHANGE NETWORK AND DATA FEEDS

Here we provide a brief overview of the National Market System (NMS), including a description of infrastructure components and some varieties of market participants. In particular, we note the information asymmetry between participants using the legally-mandated Securities Information Processor to receive quote and trade information and participants using proprietary, direct information feeds.

A. Market participants

There are, broadly speaking, three classes of agents involved in the NMS: traders, of which there exist essentially four subclasses (retail investors, institutional investors, brokers, and market-makers) that are not mutually exclusive; exchanges and ATSs, to which orders are routed and on which trades are executed; and regulators, which oversee trades and attempt to ensure that the behavior of other market participants abides by market regulation. We note that Kirilenko *et al.* claim the existence of six classes of traders based on technical attributes of their trading activity [2]. This classification was derived from activity in the S&P 500 (E-mini) futures market, not the equities market, but is an established classification of trading activity. We are unable to create such an exhaustive classification because attribution of trades in the NMS is not available for purchase [31].

1. *Traders*

The four broad classes of traders have different market objectives, and thus generally have different mechanisms for interacting with the market. Retail investors generally have a small amount of capital and thus interact with the market indirectly, usually through their broker. Since their orders are so small in relation to both the value of total market transactions and the size of the inventory of the executing broker-dealer, their orders are often *internalized* (*i.e.* matched against the inventory of their broker-dealer rather than finding a counter-party in the lit market) [32]. For example, if investor A wishes to sell 100 shares of AAPL at the market price to sell, it is likely that, within a large brokerage, investor B wishes to buy 100 shares of AAPL at the market price to buy; these orders can then be executed at the midpoint of the prices. Alternatively, the broker-dealer may choose to be the counter-party to both traders, using its own inventory of equities and capital.

Institutional investors represent an institution, such as a large corporation, university, or state pension fund. They are typically far more highly capitalized than a retail investor, and thus their orders are likely to interact more directly (whether through their brokerage or with a market maker) with the NMS.

Brokerages execute orders on behalf of their clients. They may do this by contacting market-makers, who will execute trades on behalf of the brokerage, or they may themselves be a market-maker or broker-dealer. Brokerages may enter into contracts with market-makers, who agree to buy some percentage of the brokerage's order flow [33]. Securities and Exchange Commission (SEC) (the chief regulatory body of equity markets in the US) regulation requires brokers to guarantee their clients the "best execution" for their trades, which may include most competitive price for their trades (i.e., highest possible bid price and lowest possible offer price) [34]. As has been previously identified elsewhere [35, 36], this regulatory requirement for "Best Price" execution fails to consider implications from special relativity [37]; namely, that it is impossible to determine whether two distinct events occur at the same time if those events are geographically separated in space.

Market-makers are responsible for ensuring the market's liquidity. They quote a buy and sell price for a set of traded assets at all times, and stand ready to buy or sell an amount of those assets at their respective prices [38]. Exchanges can establish designated market-makers, who are responsible for"making the market" in a specific asset[27].

2. *Market centers*

Exchanges in the NMS are privately-owned venues on which securities are traded. They are extensively regulated by the SEC and are required by law to provide the best possible execution price (under most circumstances) to their customers [34, 39]. For each equity, each exchange maintains a local order book that aggregates the orders submitted by market participants. These local order books contain information about resting limit orders, updated by order flow, including the side (buy/sell), limit price, size, and execution modifiers that give the market participant greater control over how and when their order is executed. Using their local book and proprietary matching software, exchanges match buyers with sellers.

There are currently 13 major stock exchanges:

a) NYSE (3): main exchange; ARCA, primarily for trading exchange-traded funds (ETFs); and MKT, the smallest of the NYSE family

b) NASDAQ (3): main exchange; BX, the Boston stock exchange; and PSX, the Philadelphia stock exchange

c) BATS (4): BATS and BATS Y; EDGX and EDGA. These exchanges are now owned by CBOE, effective early 2017.

d) IEX: the Investors Exchange, which was a ATS until 17 June 2016

e) CHX: the Chicago stock exchange

f) NSX: the National Stock Exchange and by far the smallest stock exchange in terms of shares traded. It has a long history of trading intermittently, with pauses in operation of duration longer than a year. NSX was purchased by NYSE on 2018-01-12 and re-branded as NYSE National, and has at time of writing (2018-07-16) begun trading again.

Though each exchange keeps offices in its namesake city, trading actually occurs (via each exchange's matching engine) in one of three data centers in northern New Jersey; see Section III B.

ATSs, colloquially known as "dark pools", are market centers on which invited participants may trade equity and other securities. While regulated by the SEC, dark pools are not required to publish quotes and are subject to less scrutiny than are the exchanges. Dark pools are not required to publish the location of their matching engine(s), and as a rule their location is generally not known to the public. Public SEC filings contain a location for each registered ATS, though it may simply be an office and not the location of the matching engine.

3. *Regulatory mechanisms*

The National Market System is regulated primarily by the SEC. The equities industry also self-regulates through the Financial Industry Regulatory Authority (FINRA), which charges itself with regulating member brokerages and exchanges. While an authoritative institution, it does not have law enforcement power itself and must refer suspected violations of securities law to the SEC for enforcement. (FINRA has some ability to provide incentives and penalties to member organizations, such as expulsion.) The Securities Information Processor (SIP), mandated by SEC regulation, is a digital information processor on which all quotes, trades, and administrative messages such as trading halts and limit-up / limit-down (LULD) messages are recorded and through which information can be disseminated to exchanges, dark pools, and other market participants. The SIP constructs the NBBO from this data, which forms the basis of the notion of "best price" for the National Market system. There are three SIP data collection "tapes", two of which (A and B) are located at the NYSE data center in Mahwah, NJ, and one of which (C) is located at the NASDAQ data center in Carteret, NJ.

In addition to the SIP tapes mentioned above, there are two FINRA-operated Trade Reporting Facilities (TRFs), one each in Mahwah and Carteret. Dark pools are required to report trades to the TRFs, which in turn report the trades to the correct SIP tape [40, 41].

Other regulatory machinery exists to prevent the "overheating" of markets in the form of price changes deemed excessive [42]. There are two types of these mechanisms: individual-stock limit-up, limit-down (LULD) mechanisms and market-wide circuit-breakers. Individual stock LULD mechanisms set price bands of 5%, 10%, and 20% for each individual stock based on prices in the immediate trailing five-minute trading period. If the stock's price exits the bands and does not return within a fifteen second time period, a five-minute trading halt for that stock is initiated. Similarly, market-wide circuit breakers (set at 7%, 13%, and 20%) initiate halts in trading if the S&P 500 breaches these bands. A breach of the first two levels results in a market-wide trading halt for 15 minutes, while a breach of the last band results in a trading halt for the rest of the trading day.

Regulatory influence on the market is not limited to price reporting and circuit-breaker mechanisms. Beginning in 2016, the SEC instituted a live-market experiment in which some securities would be quoted in minimum increments greater than a penny (which is the current minimum increment at which prices are quoted for all stocks with a share price greater than $1.00) [43]. Known as the tick-size pilot program (or tick pilot), this program directly alters the pricing mechanism and fundamental price quantization and thus may have an effect on market dynamics.

B. Physical considerations

Contrary to its moniker, "Wall Street" is actually centered around northern New Jersey. Figure 1 shows the geographic locations of the three major data centers where the bulk of trading activity occurs. The matching engines for the three NYSE exchanges is based in Mahwah, NJ, while the matching engines for the three NASDAQ exchanges is based in Carteret, NJ. The other major exchange families base their matching engines at the Equinix data center, located in Secaucus, NJ, except for IEX, which is based close to Secaucus in Weehawken, NJ. The location of individual dark pools is not public information. However, since there is a great incentive for dark pools to be located close to data centers (see sections II and VI), we believe it is likely that many dark pools are located near to the data centers at Mahwah, Carteret, and / or Secaucus. Since matching engines perform the work of matching buyers with sellers in the NMS, we hereafter refer to the locations of the exchanges by the geographic location of their matching engine. For example, IEX has its point of presence in Secaucus, but its matching engine is based in Weehawken; we locate IEX at Weehawken.

This geographic decentralization has a profound effect on the operation of the NMS. Minimum propagation delays between exchanges may be calculated and are shown in Table II. In reality, the time for a message to travel between exchanges will be strictly greater than these lower bounds, since light is slowed by transit through a fiber optic cable, and further slowed by any curvature in the cable itself. The two-way estimates in Table II give a lower bound on the minimum duration required for a latency arbitrage opportunity to be "actionable" and a more realistic estimate derived by assuming propagation through a fiber optic cable with a refractive index of 1.47 [44]. These estimates do not account for computing delays, which may occur at either

NMS Propagation Delay Estimates

	Carteret-Mahwah	Mahwah-Secaucus	Carteret-Secaucus	Secaucus-Weehawken
Straight-line Distance	34.55 mi	21.31 mi	16.22 mi	2.56 mi
	55.6 km	34.3 km	26.1 km	4.12 km
Light speed, one-way	185.75 μs	114.57 μs	87.2 μs	13.76 μs
Light speed, two-way	371.5 μs	229.14 μs	174.4 μs	27.52 μs
Fiber, one-way	272.44 μs	168.07 μs	127.89 μs	20.19 μs
Fiber, two-way	544.88 μs	336.14 μs	255.78 μs	40.38 μs
Hybrid laser, one-way	-	-	94.5 μs	-
Hybrid laser, two-way	-	-	189 μs	-

TABLE II. The speed of light is approximated by $186,000$ mi/s (or $300,000$ km/s) and fiber propagation delays are assumed to be 4.9μs/km [44]. These propagation delays form the basis for estimates of the duration required for a latency arbitrage opportunity to be considered actionable, though these figures do not account for any computing delays and thus are lower bounds for the definition of actionable. Datacenter locations, distances between datacenters, and one-way hybrid laser propagation delay are obtained from Anova Technologies [45].

end of the communication lines, in order to avoid speculative guesses. In practice such computing delays will also have a material effect on which arbitrage opportunities are truly actionable and will depend heavily on the performance of available computing hardware.

Connecting the exchanges are two basic types of data feeds: SIP feeds, containing quotes, trades, LULD messages, and other administrative messages complied by the SIP; and direct data feeds, which contain quotes, trades, order-flow messages (add, modify, etc), and other administrative messages. The direct data feeds operate on privately-funded and installed fiber optic cables that may have differential information transmission ability from the fiber optic cables on which the SIP data feeds are transmitted. The latency of the SIP may also be introduced by additional propagation delays and computation delays involved in consolidation and dissemination. Due to the observed differential latency between the direct data feeds and the SIP data feed and the heterogeneous distance between exchanges, arbitrage opportunities are created solely by the macro-level organization of the market system.

Our understanding of the physical layout of the NMS is depicted in Figure 3 at a relatively high level. There are three basic types of information flow within the NMS:

a. Direct feed information, which flows to anyone who subscribes to it. Practically speaking, direct feed information is very expensive (on the order of $\$130,000$ USD per month, see Appendix VI for details) and so is used primarily by the exchanges themselves, large financial firms, and dark pools. Direct feed information thus flows to and from the exchanges (and the major exchange participants). We hypothesize that direct feed information also flows to the dark pools, since the dark pools require some type of price signal in order for the market mechanism to function. (We do not test this hypothesis here since we cannot directly observe the internals or locations of dark pools.) The direct feeds provide the fastest means by which to acquire a price signal, and thus may provide the best economic value to traders dependent on frequent information updates; this provides the economic foundation for our hypothesis.

b. SIP information, which is considerably cheaper than direct feed information and exists by regulatory mandate. Since the direct feeds update much faster than the SIP, *a priori* it is difficult to understand why any market participant for whom purchasing direct feed data is rational would subscribe to the SIP. However, market participants may still subscribe to the SIP as a tool for use in arbitrage; see Section II for discussion of this possibility. Market participants that cannot afford the direct feed data also purchase the SIP data for use as a price signal, etc.

c. Lagged reporting data that is not yet collated by the SIP. Regulation requires that exchanges report all local quote and trade activity, and that dark pools report all trade activity. This information is collected by the appropriate SIP tapes and then disseminated through the SIP data feeds. It is the responsibility of the exchanges to report their quote and trade information to the SIP, and of the dark pools to report their trade information to the FINRA Trade Reporting Facilities. Thus, though this information will be eventually visible to all subscribers to SIP or direct feed data, it differs qualitatively from that data due to its lagged nature.

For example, suppose a trade occurs at NYSE MKT on a NASDAQ-listed security that updates the NBBO for that security. Since this trade occurs at Mahwah, it takes a non-negligible amount of time for the information to propagate to SIP Tape C, located in Carteret. However, traders located at Mahwah will have access to this information much more quickly, allowing them an information advantage over their Carteret-based competitors.

C. Data

We used data effectively identical to that used by the SEC for its Market Information Data Analytics System (MIDAS) program [24]. Every day, MIDAS collects more than one billion records from the direct feeds of all national exchanges. These records represent the exhaustive set of (1) posted orders and quotes on national exchanges, (2) modifications and cancellations of those orders, (3) trades executed against those orders, numpy and (4) administrative messages. We obtained the data from the sole data provider for the MIDAS program; Thesys Group Inc., formerly known as Tradeworx [46]. Prior to awarding Thesys Group the MIDAS contract [47], the SEC conducted a competitive source selection [48], thereby designating Thesys Group as the authoritative source for NMS data.

In addition to being the authoritative data source for the SEC's MIDAS program, another significant attribute of the Thesys data is that it is collected by a single observer from a consistent location in the NMS (i.e., the Nasdaq data center in Carteret, NJ) as depicted in Figure 1. The single observer not only allows the user to account for the relativistic effects described above but also to directly observe latency arbitrage opportunities and realized opportunity cost instead of compiling estimates of these quantities as has been done in previous studies. In collating dislocation data, we record the maximum and minimum value of each latency arbitrage opportunity only; we do not record a time-weighted average of dislocation value or other statistic. In much of our analysis we take the absolute values of the maximum and minimum values of each latency arbitrage opportunity as the fundamental object of study as any dislocation, regardless of which feed is favored, presents an opportunity for arbitrage.

IV. LATENCY ARBITRAGE

Market inefficiencies, dislocations, and latency arbitrage opportunities are all closely related, thus for the benefit of the reader we must clearly differentiate these constructs. See Figure 4 for a depiction of the relationship between these three concepts. See Appendix VII for more details. We provide a brief definition of a latency arbitrage opportunity as calculated and used in this work. Each latency arbitrage opportunity can be represented by a 4-tuple:

$$v_n = (t_n^{\text{start}},\ t_n^{\text{end}},\ \min \Delta p,\ \max \Delta p). \qquad (1)$$

The maximum (resp. minimum) value of the latency arbitrage opportunity are simply the maximum (resp. minimum) difference in the prices that are generating the latency arbitrage opportunity over the time period $[t_n^{\text{start}}, t_n^{\text{end}})$. The time period $[t_n^{\text{start}}, t_n^{\text{end}})$ is determined by identifying a contiguous period of time where $\Delta p > 0$ or $\Delta p < 0$. From the above quantities the duration of the latency arbitrage opportunity can also be calculated. We define $\Delta p(t)$ as the difference in the price transmitted by the information feeds at time t. From the definitions of $\max \Delta p$ and $\min \Delta p$ the reader will note that arbitrage opportunities will tend to feature $\min(|\min \Delta p|) \geq \$0.01$, since the minimum tick size in the NMS is set at one penny for securities with a share price of at least \$1.00, though mid-point orders can occur at half penny increments and thus disrupt this trend.

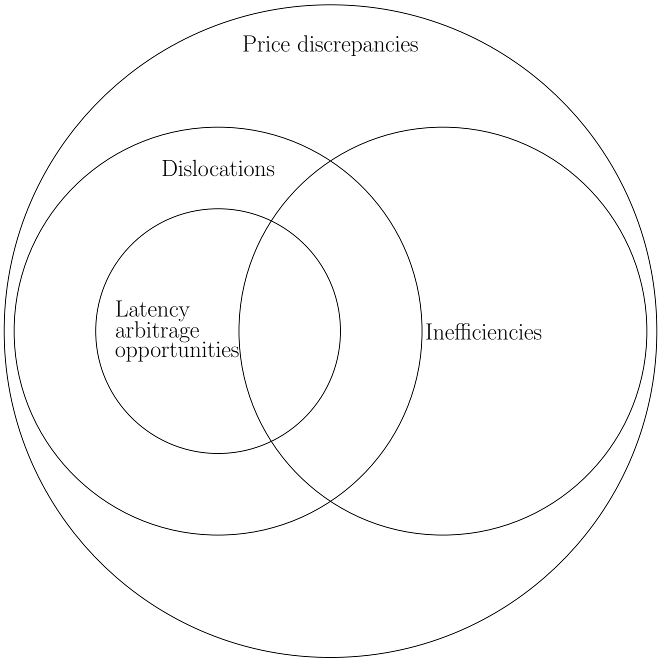

FIG. 4. The relationship between latency arbitrage opportunities, dislocations, market inefficiencies, and price discrepancies. All latency arbitrage opportunities are necessarily dislocations, but the converse is not true.

Fundamentally, latency arbitrage is the ability to generate risk-free profit due to price discrepancies between two (or more) trading locations that exist because of both a market's physical configuration and existence of information sources of differing speeds. It is best explained through a toy example.

Suppose there are two exchanges, A and B, which trade a single security, S. In this toy market system, there is a consolidating entity that constructs and disseminates a global best bid and offer (GBBO), this consolidator and the GBBO that it produces are similar to the SIP and NBBO in the NMS.

Assume that a trader is located at A, the local best bid and offer (LBBO) for S is initially bid @ \$100.00 - offer @ \$100.01 at both exchanges, and the GBBO is also bid @ \$100.00 - offer @ \$100.01. At some future time, the LBBO at A changes to bid @ \$99.98 - offer @ \$99.99 while the LBBO at B and the GBBO both remain fixed. The trader purchases shares of S at \$99.99 from A by placing a bid that will execute against the new LBO, this provides the active trader with price improvement over the global best offer (GBO) that remains at \$100.01. Finally, the trader sells shares at B for \$100.00 by placing an offer that will execute against the LBB, which is currently in sync with the global best bid (GBB), obtaining a gross profit of \$0.01 per share.

This toy example assumes that the trader is able to observe the LBBO at A, the LBBO at B, and the GBBO, then submit the appropriate orders faster than all other market participants, otherwise the arbitrage opportunity may have been consumed by another trader located at A or even a trader located at B. Note that the construction of the GBBO necessarily introduces latency when compared to the characteristic speed of market activity, thus the GBBO may diverge from the LBBO at either exchange or a synthesized DBBO.

Though we have abstracted away many of the complexities of the NMS in order to produce a clear and concise example, notice that this market state may be reached while remaining in compliance with Reg. NMS via certain sequences of orders at exchange A. The trivial sequence involves cancellation on the bid side and the submission of orders on the offer side, though some of the change in the bid side may be induced via inter-market sweep orders. Refer to Appendix VII for more details on latency arbitrage, and other strategies that may be able to leverage these market events.

The above example demonstrates conditions that are necessary, and occasionally sufficient, for latency arbitrage:

1. Two or more distinct trading locations

2. Two or more information feeds with differing latency

3. A price discrepancy. In particular, Appendix VII considers trading strategies that profit from crossed markets. If the price discrepancy is in fact a crossed market, then these conditions are both necessary and sufficient.

In assuming that the trader at A acted on the updated information faster than it was transmitted to and executed at B and faster than other market participants at A, we assume that there are two or more information feeds operating at different speeds. Conversely, consider an example in which any one of the above conditions is not satisfied. If there is no price discrepancy, there clearly can be no latency arbitrage. From first principles, for a price discrepancy to exist, there must be two or more locations at which prices are discovered; hence, two or more exchanges. And two differing information speeds are also required, as without these, there is no structural mechanism by which a market participant can have differential access to information.

V. REALIZED OPPORTUNITY COST

In finding potential profit opportunities via analysis of trade data, we utilized the following decision procedure: for each trade that occurred on the NMS at the NBBO, we checked the data feeds to see if a discrepancy between the SIP and consolidated direct feeds was present at the time the trade executed and counted each as a *differing trade*. If the differing trade executed at a price offered by the SIP feed then a price difference was calculated, i.e. $p_{\text{SIP}} - p_{\text{direct}}$ if the liquidity-demanding order was a offer and $p_{\text{direct}} - p_{\text{SIP}}$ if the liquidity-demanding order was a bid, and a cost, termed the realized opportunity cost (ROC), was assigned to the trade using the number of shares multiplied by the price difference. The sum total of all ROC occurrences over a day was calculated and recorded. Since intra-day events can offset—e.g, two dislocations can occur at the same time, with one dislocation favoring a direct data feed and one dislocation favoring a SIP data feed—these so-called upstream offsetting events imply that our calculation of ROC is a conservative lower bound for ROC that actually occurred. With this construction, positive opportunity costs indicate an incentive for liquidity demanding market participants to use the SIP feed while negative opportunity costs indicate an incentive to use the aggregated direct feeds. By ignoring the sign of the opportunity costs, and thus which feed is favored, an aggregate or total realized opportunity cost may be constructed. Precise definitions of quantities described here are located in Appendix VII.

As above, we provide a brief toy example of how realized opportunity cost can arise and a description of its ' calculation. A minimal example involves two traders, each of which is in the market to buy the security XYZ. One trader buys against the SIP NBO and the other buys against the best offer from a direct feed. If a trade for 100 shares of XYZ executes against the direct best offer quote of \$100.00 per share when there was a stale SIP quote of \$100.01 per share, the trader who buys exclusively on the SIP would have realized an opportunity cost of \$0.01 per share, or \$1.00 in total. Because this opportunity cost favored the direct feed, this portion of ROC would be assigned a negative value. If, during another trade on the same day, another trade for 100 shares of XYZ executes when the direct feed offer price is \$101.02 and the SIP quotes at \$101.00 per share, the trader who buys exclusively on the direct feeds would have experienced a realized opportunity cost of \$0.02 per share, or \$2.00 in total. This ROC is assigned a positive value because it favors the SIP feed. Summing these two together produces a net ROC of \$1.00, hence the conservative nature of our estimates. If, instead, our calculation summed the absolute value of each ROC-generating event, the figure above would instead be \$3.00.

VI. RESULTS

A. Dislocations and latency arbitrage opportunities

Contrary to a body of academic work [6–8] we find that dislocations and latency arbitrage opportunities are widespread and may have qualitative welfare effects on NMS participants, particularly large investors or investors that interact with the NMS directly on a frequent basis. The combined number of latency arbitrage opportunities in 2016 among Dow 30 securities was $120,355,462$, or $\frac{120,355,462}{252\times 6.5\times 60^2} \approx 20.4$ latency arbitrage opportunities *per second* on the NMS. When restricting our attention to what we term *actionable* latency arbitrage opportunities (those that last at least 545 μs), we find that there were $\frac{65,073,196}{252\times 6.5\times 60^2} \approx 11$ actionable latency arbitrage opportunities every second. Even when inspecting actionable latency arbitrage opportunities with a minimum magnitude of at least 2 cents we find that there were $\frac{2,872,734}{252\times 6.5\times 60^2} \approx .49$ latency arbitrage opportunities per second, or almost one large actionable latency arbitrage opportunity every two seconds.

We focus much of our subsequent analysis on the latency arbitrage opportunity distribution conditioned on both duration and magnitude, as we estimate it unlikely that there is much potential profit to be made on latency arbitrage opportunities that are shorter. From an academic point of view, arbitrage opportunities with a minimum magnitude greater than one cent are more interesting since one might expect all arbitrage opportunities to feature a magnitude that corresponds with the minimum tick size ($0.01 in this case). There are several aspects of the conditional distribution that bear special notice. First, the distribution of each attribute is exceptionally heavy-tailed. In absolute value, the 75%-ile of the minimum and maximum magnitude are a not-insignificant three cents—but the mean in absolute value of the minimum magnitude (resp. maximum magnitude) is 3.05 (resp. 8.23) cents! A similar phenomena is true for the duration distribution, where the 75%-ile is 4231 μs, while the mean is an astounding 0.389 *seconds*, almost two orders of magnitude longer. The max magnitude, min magnitude, and duration distributions are all highly skewed, while the distributions of the maximum and minimum magnitudes are nearly identical.

Figure 5 shows the distribution of latency arbitrage opportunities modulo day, binned by minute. Intraday latency arbitrage opportunity distributions are markedly nonuniform, with a majority of the probability mass concentrated toward the beginning of the trading day. There is also a notable spike in the number of latency arbitrage opportunities occurring in mid-afternoon and at the very end of the trading day. Additionally, note that the sawtooth pattern in the distribution of latency arbitrage opportunity starts has a spike roughly every half hour.

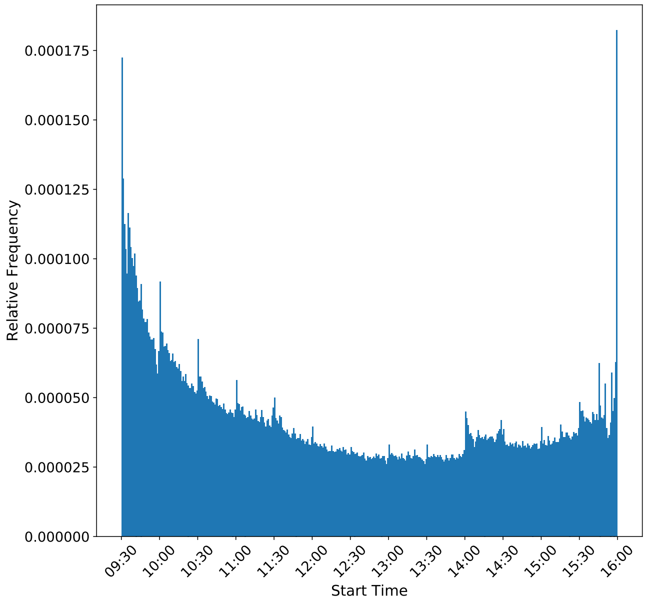

FIG. 5. Distribution of latency arbitrage opportunity start times binned by minute.

To further unpack the relationship between time of day, length, and magnitude of latency arbitrage opportunities, we created a representation of latency arbitrage opportunities modulo day as an ordered network, which we hereafter refer to as a circle plot and display for AAPL on an arbitrary day in Figures 8 and 7. Circle plots are constructed using the following algorithm. Starts and stops of latency arbitrage opportunities at time t are termed events $v(t)$ and denoted by black nodes. More than one event can occur at each time t; all events are represented by the same node. Events $v_i(t)$ and $v_j(s)$ where $t < s$ are connected by an edge e_{ij} when a latency arbitrage opportunity starts at $v_i(t)$ and ends at $v_j(s)$. It is not necessarily the case that latency arbitrage opportunities start and stop in order as seen above; for example, an opportunity could start at v_i, another opportunity could start at v_j, the first opportunity could end at v_k, and then the second opportunity could end at v_ℓ. When N events occur "out of order" in this way, we identify the events as a single component (even though, as in the above example, the component decomposes into two two-tuples of events) and term it an N-component for reasons we state below; the above example is a 4-component. Nodes are plotted in rays that spread outward from the geometric center of the plot in a modulo 10 relation: in the case of the real time representation, an event represented by a node on a fixed but arbitrary circle of the graph occurred at a multiple of $10\mu s$ from all other events represented by nodes on the ring; in the case of the event-time representation, an event represented by a node on a fixed but arbitrary circle of the graph and another event represented by a node on the same circle are separated by

Filter	Statistic	Duration	Min. Value	Max. Value	Min. Mag.	Mean Mag.	Max. Mag.
None	count	120,355,462					
	mean	0.073712	-0.0012	0.0013	0.0112	0.0124	0.0137
	std	5.519033	0.1698	0.4815	0.0529	0.2581	0.5075
	min	0.000000	-141.49	-63.21	0.01	0.01	0.01
	25%	0.000216	-0.01	-0.01	0.01	0.01	0.01
	50%	0.000624	0.01	0.01	0.01	0.01	0.01
	75%	0.001190	0.01	0.01	0.01	0.01	0.01
	max	10,789.83	372.69	4,905.69	372.69	2,452.85	4,905.69
Duration	count	65,073,196					
	mean	0.136142	-0.0020	0.0022	0.0109	0.0130	0.0151
	std	7.505197	0.2233	0.6511	0.0653	0.3474	0.6850
	min	0.000546	-141.49	-63.21	0.01	0.01	0.01
	25%	0.000751	-0.01	-0.01	0.01	0.01	0.01
	50%	0.001103	0.01	0.01	0.01	0.01	0.01
	75%	0.002391	0.01	0.01	0.01	0.01	0.01
	max	10,789.83	372.69	4,905.69	372.69	2,452.85	4,905.69
Duration	count	2,872,734					
&	mean	0.387866	-0.0250	0.0267	0.0305	0.0564	0.0823
Min. Mag.	std	29.566716	0.9046	1.0021	0.3102	0.7116	1.3115
	min	0.000546	-141.49	-63.21	0.02	0.02	0.02
	25%	0.000724	-0.02	-0.02	0.02	0.02	0.02
	50%	0.001207	0.02	0.02	0.02	0.02	0.02
	75%	0.004231	0.02	0.02	0.03	0.03	0.03
	max	10,789.83	372.69	593.43	372.69	372.84	593.43

TABLE III. Latency arbitrage opportunity attributes where the first section is unconditioned, the middle section is restricted to opportunities with a duration greater than $545\mu s$, and the final section is restricted to opportunities with a duration greater than $545\mu s$ and a minimum magnitude greater than \$0.01. Of the approximately 120 million opportunities observed, more than 54% of them have a duration that would allow them to be considered actionable, and about 2.4% of opportunities are both actionable and feature a minimum magnitude greater than \$0.01. This makes the magnitude of the realized opportunity cost even more remarkable. Additionally, note that observed durations of "0" are the result of opportunities that begin and end within the same microsecond, the maximum precision used for the majority of market data timestamps.

an integer multiple of events that occurred between them. Edges between nodes v_i and v_j are weighted according to the quantity

$$\sum_{(v_i,v_j)} \max(|\Delta p_{\max}|, |\Delta p_{\min}|), \qquad (2)$$

where the sum is taken over all events that started at node v_i and ended at node v_j and $\Delta p_{\max}$ and $\Delta p_{\min}$ are the largest (resp. smallest) positive change in value that occurred during each event. Figure 8 displays the circle plot for AAPL for an arbitrary day (2016-01-07) of trading. There is high time-space event density near the beginning of the day, as we have shown to be typical above, and there is another spike in time-space density near noon-12:30 PM. This clustering can make interpretation of the fine event structure difficult to discern, so we conduct a renormalization into event space with a simple method: consecutive events $v_i(t)$ and $v_j(s)$ are plotted in order, but at a uniform distance so that the measure on the graph becomes a Stieltjes-type instead of a Lebesgue-type measure. Figure 7 displays the circle graph in this renormalized space, where it is easier to see that the usual behavior of latency arbitrage opportunities is a regular cyclic, on-off (start-stop) pattern. However, there are multiple deviances from this pattern—any component other than a 2-component is structurally different from a purely sequential pattern. In fact, there is an injection from an N-component and a tied, non-negative random walk $\{x_n\}_{n=0}^{N}$, $x_0 = x_{N+1} = 0$, $x_n \geq 0$ for all n. This injection is defined by the relationships

$$\text{start of } k \text{ events } \cong k \text{ steps up}$$

and

$$\text{end of } k \text{ events } \cong k \text{ steps down}.$$

As a concrete example, the 4-component described above maps to the random walk steps $\{1, 1, -1, -1\}$, with values $x_0 = 0,\ x_1 = 1,\ x_2 = 2,\ x_3 = 1,\ x_4 = 0$. Figure 6 displays a toy example of the injection between N-components in a circle graph and a tied positive random walk, as outlined above.

When aggregated over all trading days, evidence of per-

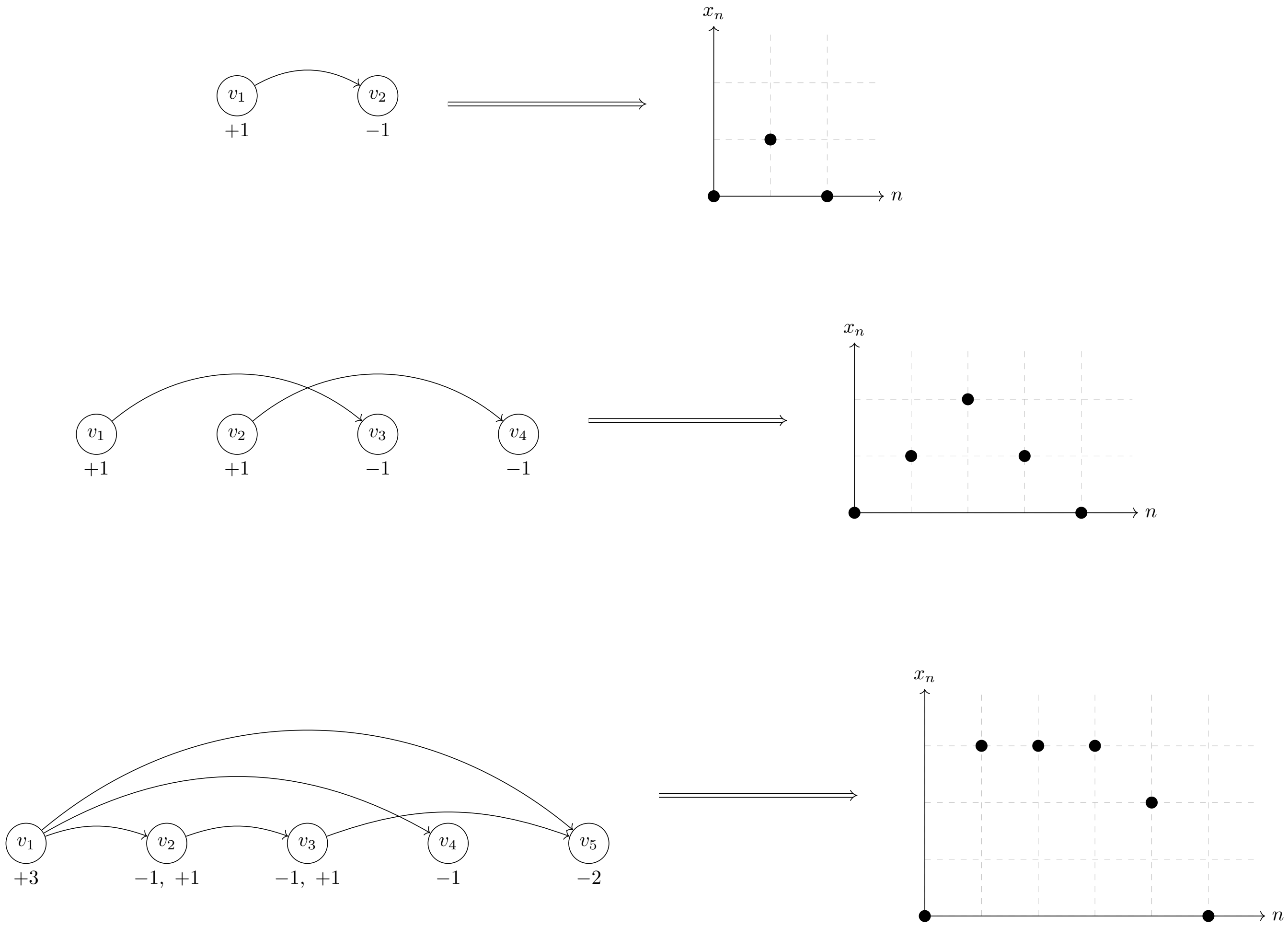

FIG. 6. A graphic displaying the injection mapping from an N-component in a circle graph to a tied positive random walk of length $N + 1$. The injection is given by j outgoing edges $\cong j$ steps up and likewise k incoming edges $\cong k$ steps down. The total number of steps up or down is given by $x_{n+1} - x_n = $ # of steps up + # of steps down. The top row displays a simple 2-component, where an equity begins a dislocation at time t_i and ends it at time t_{i+1}. The corresponding walk on the line starts at zero, moves up a step, and then moves down. The second row displays a 4-component identical to that described in the text of the article. This 4-component demonstrates the geometric nature of the circle graph—in purely graph-theoretical language, this component is clearly separable into two disconnected pieces, but since an ordering is imposed on the nodes, the crossing of the edges implies the staggered starts and stops of the two dislocations.

sistent nontrivial structure in the event-space density of N-tuples emerges. Figures 8 and 7 display the aggregate of events in AAPL modulo day since it is likely that this is the longest timescale on which HFT firms maintain a long or short position [49, 50]. For a visual comparison between all Dow 30 tickers, we include tiles of these directed networks in Figures 20 and 21.

B. Realized opportunity cost

The large number of actionable latency arbitrage opportunities likely has a direct effect on the potential profitability of latency arbitrage strategies and opportunity cost market participants may incur by using one information source over the other. The aggregate of this realized opportunity cost can be estimated by cataloging the quantity and characteristics (average price difference, etc.) of differing trades. Table I summarizes many of these findings. In the time period studied (01-01-2016 through 31-12-2016) there were a total of 392,101,579 trades of stocks in the Dow 30, with a traded value of \$3,858,963,034,003.48 USD. Of those trades, we classified 87,432,231 trades, or 22.3% of the total number of trades, as *differing* trades, defined as follows: if the trade is on the buy side, it is a differing trade if the SIP bid is not equal to the direct bid; if the trade is on the sell side, it is a differing trade if the SIP offer is not equal to the direct offer. These differing trades had a traded value of \$900,535,924,961.72 USD, or 23.34% of the total traded value. We estimate that there was a total of \$83,948,329.85 USD to be gained over this time period

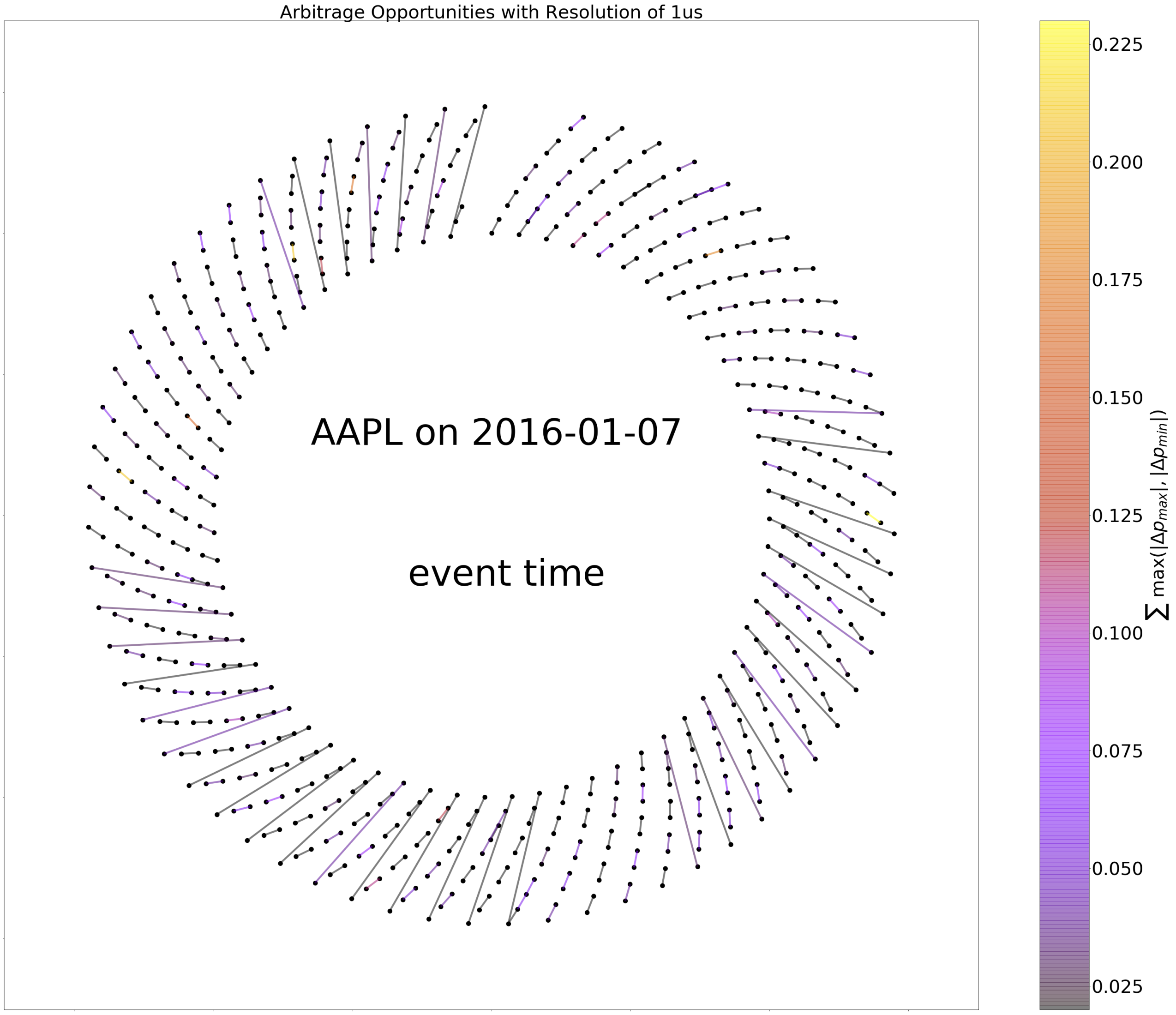

FIG. 7. Distribution of latency arbitrage opportunities with minimum magnitude greater than \$0.01 and duration longer than $545\mu s$ for one arbitrary day of AAPL (2016-01-07) ordered with respect to event time. Nodes are placed in rings modulo 10; nodes zero through 9 are in the first ray from the origin, then the angle in the plot is incremented and nodes 10 through 19 are in the second ray, etc. A link e_{ij} connects two nodes—latency arbitrage opportunity events—v_i and v_j if a latency arbitrage opportunity starts at v_i and stops at v_j. This view of the latency arbitrage opportunity network preserves time ordering while defining a nonlinear transformation between uniform time ordering, as shown below in Figure 8, and uniform event-space ordering, as shown here. As noted in the text, it is not necessary for only one latency arbitrage opportunity to exist at the same point in time t. For example, there are many instances of new latency arbitrage opportunities starting while another is still ongoing—the first starts at v_i and then another starts at v_j and ends at v_k, followed by the first latency arbitrage opportunity ending at v_ℓ. Irregular behavior such as this generates the irregular banding of the edge distribution. Interested readers may wish to have some more context for the selected date. For AAPL, 2016-01-07 ranked 8th out of 252 trading days when considering ROC. \$106,990.23 in ROC was accumulated, which lies between the minimum of \$2,773.35 and the maximum of \$138,331.08. This day of AAPL also ranked 15th when considering the number of LAOs. A total of 108,843 ocurred, falling between the minimum of 9,256 and the maximum of 188,656.

by using the aggregated direct feeds instead of the SIP feeds, and \$160,213,922.95 USD to be gained by using a combination of both the SIP and aggregated direct feeds. The ratio of the fraction of differing notional values to the fraction of differing trades,

$$f = \frac{D_{\text{notional}}/T_{\text{notional}}}{D_{\text{trades}}/T_{\text{trades}}},$$

is $f = 1.046$. Figures 22 displays the daily net opportunity cost aggregated over all tickers in our sample.

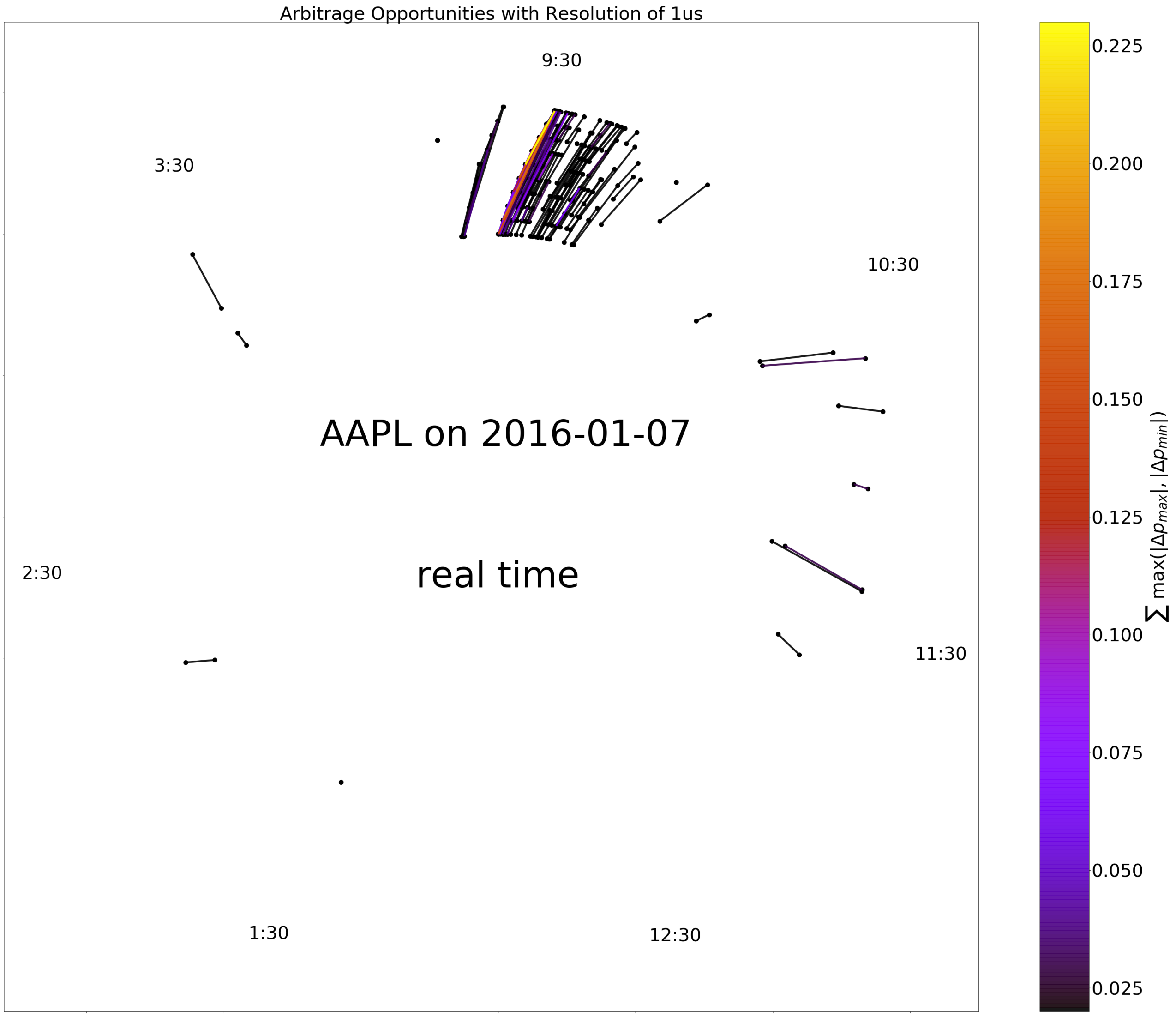

FIG. 8. Latency arbitrage opportunities in the same day of AAPL (2016-01-07) are plotted, but are not transformed to event space. This displays the obvious nonuniform density, with a large number of latency arbitrage opportunities occurring near the beginning of the trading day and another spike in activity near noon - 12:30 PM.

	Trades	Traded Value	Diff. Trades	Diff. Traded Value	ROC	ROC/Share
mean	1,555,958.65	15,313,345,373.03	346,953.30	3,573,555,257.78	635,769.54	0.011804
std	463,558.93	3,891,299,900.31	146,677.85	1,234,882,079.43	655,911.15	0.008592
min	579,206	6,664,671,053.15	89,564	1,035,855,029.71	145,205.65	0.008848
25%	1,278,813.25	12,915,031,172.08	262,209	2,804,569,367.64	417,485.73	0.009613
50%	1,429,062	14,431,597,662.02	309,158	3,274,390,601.60	514,856.64	0.010154
75%	1,715,351.25	16,829,521,684.38	387,772	3,993,470,514.97	666,268.27	0.011213
max	3,596,006	30,999,914,293.66	1,073,029	9,428,952,387.10	7,817,684.58	0.098303

TABLE IV. Summary statistics of realized opportunity cost and related statistics for Dow 30 stocks, aggregated over the 252 trading days in 2016.

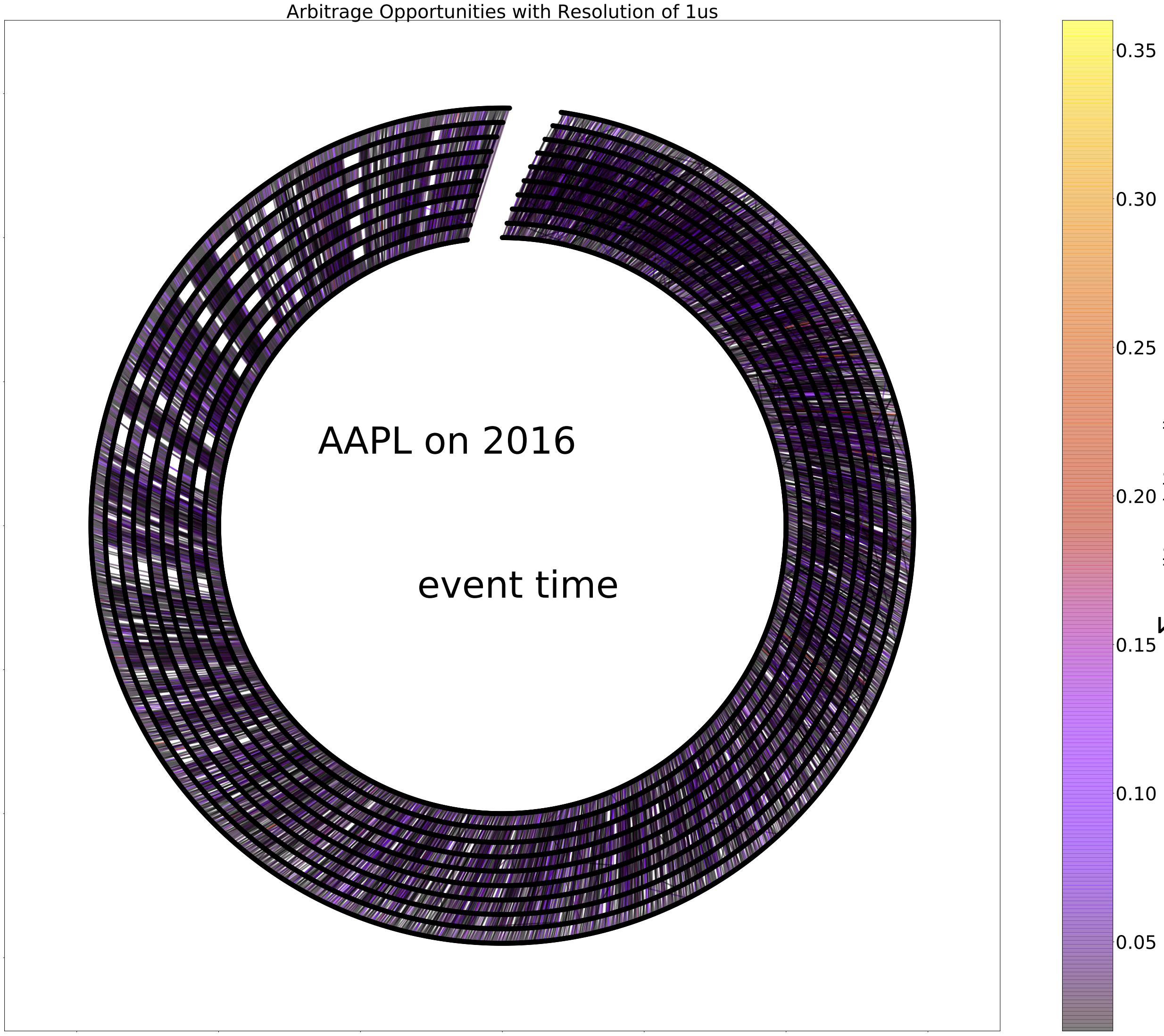

FIG. 9. Latency arbitrage opportunities are plotted as above, but aggregated over an entire year and plotted modulo day, as this is likely the longest timescale over which HFT strategies are used. Here latency arbitrage opportunities are plotted in event space, where density is uniform between events v_i and v_{i+1}. Note the presence of irregular structure even here, evidence of higher-order structure in the ordering of starts and stops of latency arbitrage opportunities.

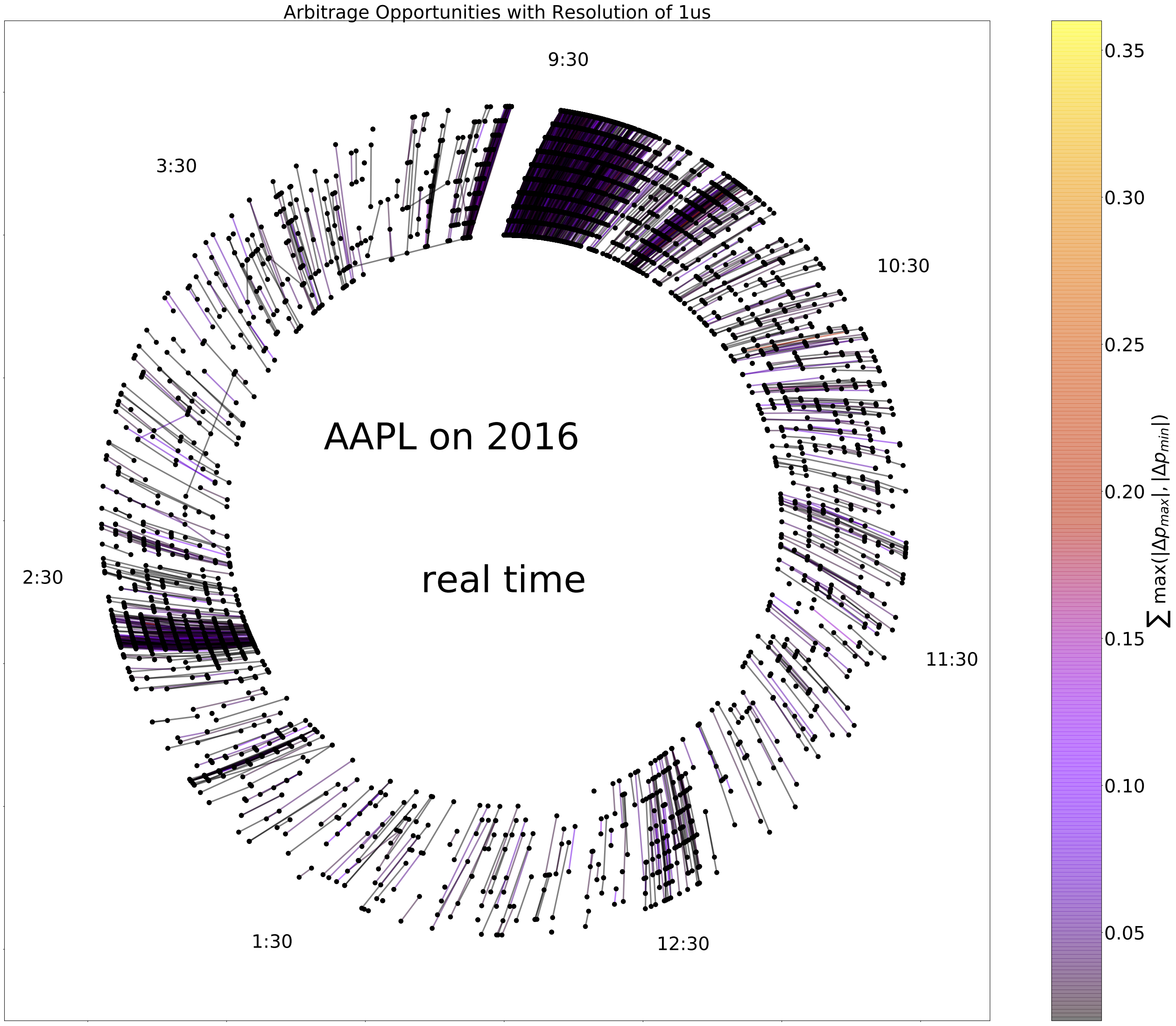

FIG. 10. Latency arbitrage opportunities are here aggregated over a year and plotted modulo day, as above, but not transformed to event space. The high density of latency arbitrage opportunities at the beginning of the trading day, near noon - 12:30. and near 2:15 - 2:30 is readily apparent.

Figure 11 provides further insight into the joint distribution of total and differing trades. While we might *a priori* expect that the ratio of total to differing trades would remain roughly a fixed constant, we see that this is not observed empirically.

VII. CONCLUDING REMARKS

Using the most comprehensive set of NMS data available for purchase, we have shown that market inefficiencies in the form of dislocations and arbitrage opportunities were common in the Dow 30 in 2016. Contrary to prevailing academic and popular opinion, we find that inefficiencies due to the physical fragmentation of the market are widespread and potentially generate massive profit for agents with access to superior market information. Actionable latency arbitrage opportunities—those we estimate can provide real latency arbitrage opportunities—occur more than ten times every second. Correspondingly, the total potential differing traded value calculated from the volume of differing trades exceeded $900B USD in 2016 on the Dow 30, while total realized opportunity cost exceeded $160M USD on the same group of equities during the same time period. These figures are entirely at odds with a body of academic research [6–8], while corresponding well with figures reported in other bodies of work [9, 10].

We briefly remark on the significance of the above results. At first blush, it may appear that ∼$900B in differing trades should not cause a material effect on the security and functioning of the NMS; comparing this figure to

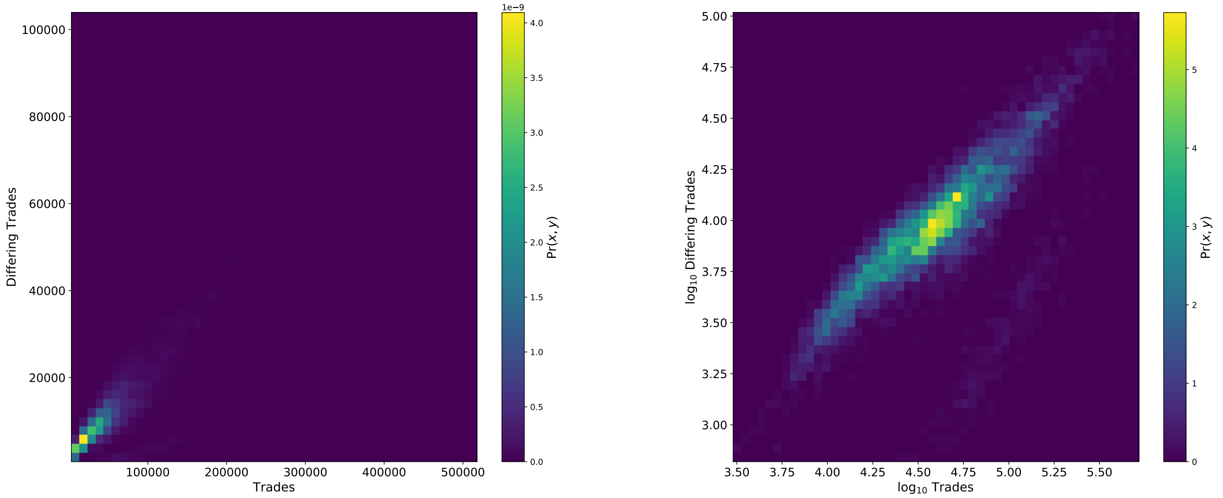

FIG. 11. Left: A bivariate empirical distribution function for total trades and number of differing trades. Right: The same distribution, but with logged axes. We might expect *a priori* that they are related by a constant proportion and hence should observe a fit $\log_{10}$ total trades $= c + \log_{10}$ differing trades, where $c < 0$. Though there is good evidence of this linear relationship, we see there is a non-negligible area of higher total trades with markedly sub-linear scaling of differing trades.

the total amount traded on the NMS in 2016 ($3.858T USD) may lead one to the conclusion that the arbitrage opportunities pale correspondingly in importance. Nothing could be farther from the truth. The mere perpetual existence of such opportunities provides a counterpoint to the weak efficient markets hypothesis: these opportunities can be repeatedly and reliably exploited by a particular class of market participants—those with access to faster direct feed information. In addition, the existence of dislocations and actionable arbitrage opportunities has bearing on security aspects of the NMS. If firms whose strategies rely on the ability to leverage direct information are suddenly faced with a lack of liquidity or another market phenomenon leading to their inability to perform what they consider normal trading activity, systemic activation of semi- or fully-automated risk management systems may occur, possibly resulting in spreading of financial contagion and concomitant adverse economic effects (e.g., large and sustained price level drops, effects on the real economy, etc.)

Though our work is empirical, our results do have implications for theoretical results on the efficiency of financial markets. We do not direct examine price time series and so do not comment on whether or not these data appear to incorporate all publicly-available past information, as hypothesized by weak-form EMH. However, the discovery of systematically-different prices as measured in geographically-distinct locations that can be routinely observed by agents with access to higher-speed information flows—and cannot be routinely observed by agents without this access—has a logical bearing on questions of distributional effects of asymmetric information and market design. More fundamentally, detailing the nuances of the current NMS infrastructure begs increased precision in common definitions used in the theoretical study of finance. For example, a reasonable criticism to our results from the point of view of microeconomic theory would proceed as follows: since a share of AAPL trading at NYSE and a share of AAPL trading at IEX are different products, differentiated as they are by exchange, it does no violence to the "law of one price" that these shares may quote at different prices. Such a criticism overlooks the assumed indistinguishability of exchanges as implied by the SIP price discovery mechanism, but taken at face value, beggars belief in the utility of such a definition of differentiated product.

As the first study to use entirely comprehensive data in this field, we focused our attention on the (admittedly limited) data set of the Dow 30 in 2016. Future work should focus on longer time periods, larger groups of equities, and other exchange traded products such as Exchange Traded Funds (ETF). For example, an extension of the current work to larger groups of equities, such as the S&P 500 and Russell 3000, or a time series analysis of the latency arbitrage opportunities and realized opportunity cost series over several years would be useful extensions of the current work.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge helpful discussions with Anshul Anand, Yosry Barsoum, Aaron Blow, Lashon Booker, David Bringle, Peter Carrigan, Gary

Comparetto, Bryanna Dienta, Peter Sheridan Dodds, Jordan Feidler, Andre Frank, Chase Garwood, Bill Gibson, Frank Hatheway, Emily Hiner, Chuck Howell, Eric Hunsader, Neil Johnson, Connie Lewis, Wade Shen, Jonathan Smith, Thomas Wilk and all of the participants of the "Flash Mob Research in Computational Finance" hosted by the Complex Systems Center at the University of Vermont. All opinions and remaining errors are the sole responsibility of the authors and do not reflect the opinions nor perspectives of their affiliated institutions nor that of the funding agencies. B.F.T., D.R.D, C.M.V.O., J.H.R., and J.V. were supported by DARPA award #W56KGU-17-C-0010. B.F.T., M.T.T.K., M.T.M., D.S., and J.V. were supported by HSARPA award #HSHQDC-14-D-00006. C.M.D. was supported by NSF grant #1447634. C.M.D., and T.J.G. were supported by a gift from Massachusetts Mutual Life Insurance Company. The authors declare no conflicts of interest. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

[1] SP Dow Jones Indices. Dow jones industrial average, 2018.

[2] Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: The impact of high frequency trading on an electronic market. *Available at SSRN*, 1686004, 2011.

[3] Michael A Goldstein and Kenneth A Kavajecz. Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*, 7(3):301–333, 2004.

[4] Mark Grinblatt and Matti Keloharju. The investment behavior and performance of various investor types: a study of finland's unique data set. *Journal of financial economics*, 55(1):43–67, 2000.

[5] John Authers. Momentum and value keep markets randomly efficient. *The Financial Times*, 2017.

[6] James J Angel, Lawrence E Harris, and Chester S Spatt. Equity trading in the 21st century. *The Quarterly Journal of Finance*, 1(01):1–53, 2011.

[7] James J Angel, Lawrence E Harris, and Chester S Spatt. Equity trading in the 21st century: An update. *The Quarterly Journal of Finance*, 5(01):1550002, 2015.

[8] Robert P Bartlett III and Justin McCrary. How rigged are stock markets?: Evidence from microsecond timestamps. Technical report, National Bureau of Economic Research, 2016.

[9] Shengwei Ding, John Hanna, and Terrence Hendershott. How slow is the nbbo? a comparison with direct exchange feeds. *Financial Review*, 49(2):313–332, 2014.

[10] Elaine Wah. How prevalent and profitable are latency arbitrage opportunities on us stock exchanges? 2016.

[11] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[12] Jean-Philippe Bouchaud. Econophysics: Still fringe after 30 years? *arXiv preprint arXiv:1901.03691*, 2019.

[13] James Foye, Dusan Mramor, and Marko Pahor. The persistence of pricing inefficiencies in the stock markets of the eastern european eu nations. 2013.

[14] Eugene F Fama and Kenneth R French. Size, value, and momentum in international stock returns. *Journal of financial economics*, 105(3):457–472, 2012.

[15] Neil Johnson, Guannan Zhao, Eric Hunsader, Hong Qi, Nicholas Johnson, Jing Meng, and Brian Tivnan. Abrupt rise of new machine ecology beyond human response time. *Scientific reports*, 3:2627, 2013.

[16] Maureen OHara. High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270, 2015.

[17] Neil Johnson, Guannan Zhao, Eric Hunsader, Jing Meng, Amith Ravindar, Spencer Carran, and Brian Tivnan. Financial black swans driven by ultrafast machine ecology. *arXiv preprint arXiv:1202.1448*, 2012.

[18] Andrew W Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. 2004.

[19] George A Akerlof. The market for lemons: Quality uncertainty and the market mechanism. In *Uncertainty in Economics*, pages 235–251. Elsevier, 1978.

[20] Mervyn King, Jill Atkins, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American economic review*, 97(1):242–259, 2007.

[21] Robert Bloomfield, Maureen Ohara, and Gideon Saar. How noise trading affects markets: An experimental analysis. *The Review of Financial Studies*, 22(6):2275–2302, 2009.

[22] Eric Budish, Peter Cramton, and John Shim. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621, 2015.

[23] Fischer Black. Noise. *The Journal of finance*, 41(3):528–543, 1986.

[24] U.S. Securities and Exchange Commission. Midas: Market information data analytics system, 2013.

[25] Jeff Alexander, Linda Giordano, and David Brooks. Dark pool execution quality: A quantitative view. *http://blog.themistrading.com/wp-content/uploads/2015/08/Dark-Pook-Execution-Quality-Short-Final.pdf*, 2015.

[26] Allen Carrion. Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets*, 16(4):680–711, 2013.

[27] Albert J Menkveld. High frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712–740, 2013.

[28] Michael A Goldstein, Pavitra Kumar, and Frank C Graves. Computerized and high-frequency trading. *Financial Review*, 49(2):177–202, 2014.

[29] Tarun Chordia, Amit Goyal, Bruce N Lehmann, and Gideon Saar. High-frequency trading. 2013.

[30] Sal Arnuk and Joseph Saluzzi. *Broken markets: how high frequency trading and predatory practices on Wall Street are destroying investor confidence and your portfolio*. FT Press, 2012.

[31] The Consolidated Audit Trail (CAT) is an SEC initiative (SEC Rule 613) that will require such attribution to be made available [51]. At the time of writing this framework was not yet constructed.

[32] U.S. Securities and Exchange Commission. Trade execution, 2013.

[33] U.S. Securities and Exchange Commission. Payment for order flow, 2007.

[34] U.S. Securities and Exchange Commission. Best execution, 2011.

[35] Nanex. Einstein and the great fed robbery, 2013.

[36] James J Angel. When finance meets physics: The impact of the speed of light on financial markets and their regulation. *Financial Review*, 49(2):271–281, 2014.

[37] Albert Einstein. On the electrodynamics of moving bodies. 1905.

[38] U.S. Securities and Exchange Commission. Market maker, 2000.

[39] U.S. Securities and Exchange Commission. Regulation nms, 2005.

[40] U.S. Securities and Exchange Commission. Regulation alternative trading system, 1998.

[41] U.S. Securities and Exchange Commission. Regulation alternative trading system amendments, 2015.

[42] U.S. Securities and Exchange Commission. Measures to address market volatility, 2012.

[43] U.S. Securities and Exchange Commission. Tick size pilot program, 2016.

[44] Kevin Miller. Calculating optical fiber latency. http://www.m2optics.com/blog/bid/70587/Calculating-Optical-Fiber-Latency. Accessed: 2017-07-31.

[45] Anova Technologies. Anova technologies network map, 2018.

[46] Thesys Group Inc. Thesys group inc., 2018.

[47] Federal Business Opportunities. Midas contract award notice, 2012.

[48] Nathaniel Popper and Ben Protess. To regulate rapid traders, s.e.c. turns to one of them, 2012.

[49] Douglas J Cumming, Feng Zhan, and Michael J Aitken. High frequency trading and end-of-day manipulation. 2012.

[50] Albert J Menkveld. High-frequency traders and market structure. *Financial Review*, 49(2):333–344, 2014.

[51] U.S. Securities and Exchange Commission. Consolidated audit trail, 2012.

GLOSSARY AND DEFINITIONS

Market Architecture

Definition VII.1 (*Market System*)**.** *A market system may be defined as a network or graph which consists of a set of one or more market centers connected by a set of communication channels or (links), i.e. system = (centers, links).*

Definition VII.2 (*Market Center*)**.** *A market center is a location, physical or digital, where agents may interact with a market system. A market center may be defined as a tuple containing a local order book, a set of valid actions, and a set of traded financial instruments, i.e. center = (book, actions, instruments).*

Definition VII.3 (*Local Order Book*)**.** *The local order book contains information about the unfulfilled orders that have been submitted to a market center, allowing it to accumulate and maintain state. One possible representation of a local order book for a single financial instrument is two ordered lists of queues, where each list is associated with a side of the marked (bid/offer) and each queue is associated with a price.*

Definition VII.4 (*Action Set*)**.** *The action set defines the valid actions at a market center. No requirements are imposed on the action set, though a simple real world action set might allow for the submission of limit orders (which guarantee price), market orders (which guarantee execution), modification of resting orders, and cancellation of resting orders; i.e. actions = {limit order, market order, modify, cancel}.*

Definition VII.5 (*System Activity*)**.** *Let the system activity, $\mathbb{A}$, be a chronological list of all actions that are performed in a market system. This includes actions performed by market participants, administrative messages transmitted by regulators, and messages transmitted by the exchange(s).*

Definition VII.6 (*Data Feed*)**.** *A data feed, D, is defined to be any subset of the system activity of a market system (i.e. $D \subseteq \mathbb{A}$). Note that recorded occurrence times of identical events may vary between distinct data feeds due to physical considerations such as the finite speed of information propagation, desynchronized clocks, etc.*

Financial Instruments

Definition VII.7 (*Security*)**.** *A security is a financial instrument that represents partial or total ownership of an object or entity. Securities are fungible; securities belonging to the same "class" have the same value, and therefore are interchangeable. Additionally, the exact value of a security is negotiable. Common varieties of securities include stocks, bonds, and options, all of which may be traded on electronic markets, such as the NMS.*

Definition VII.8 (*Stock*)**.** *Stocks, which are also called equities or equity securities, are a variety of security that represents partial ownership of a publicly traded company. Stocks are a vehicle by which companies can acquire the capital necessary to grow and the secondary market for stocks is the basis of a large portion of the U.S. financial industry.*

The Best Bid/Offer

The following definitions assume the existence of a market system, *system* = (*centers*, *links*). Each *center* $\in$ *centers* has an *action* set that allows for limit orders and trades a financial instrument i. Additionally, there exists a data feed, D, that contains information about the top of the book at each market center (i.e., a consolidated quote feed).

Definition VII.9 (*Local Best Bid/Offer*)**.** *The local best bid and offer (LBBO) is a tuple composed of the local best bid and the local best offer at a particular market center.*

The local best bid for i at a particular center $\in$ centers, at a time, t, is given by the tuple (p, q), where p is the maximum price among all active bids for i in the book at center (as observed via data feed D) and q is the quantity of shares of i available at that price at center (i.e. $LBB(D, center, i, t) = (p, q)$). The local best offer is defined similarly, but uses the minimum price among active offers at center along with the number of shares associated with that order (i.e. $LBO(D, center, i, t) = (p', q')$).

Definition VII.10 (*Global Best Bid/Offer*)**.** *The global best bid and offer (GBBO) is a tuple composed of the global best bid and the global best offer at a particular market center.*

The global best bid is similar to the local best bid, but is formed by the maximum price (and the quantity associated with that order) among resting bids for i among all market centers, i.e. $GBB(D, i, t) = (p'', q'')$. Similarly, the global best offer is formed by the minimum price among resting offers and the number of shares at that price (i.e. $GBO(D, i, t) = (p''', q''')$).

The NBBO, provided by the SIP, is an example of a GBBO in the NMS. Note that any real implementation of a GBBO necessitates the introduction of some amount of latency from propagation delays between the market centers and consolidating entity. This latency can have material implications in electronic markets where information propagation approaches the speed of light.

Market Inefficiencies

The following definitions assume the existence of a market system, *system* = (*centers*, *links*), containing two market centers, two data feeds, D_1 and D_2, and a financial instrument i that is traded at each *center* $\in$ *centers*. D_1 and D_2 are assumed to contain quote information from each market center, though they may have additional information that contributes to their uniqueness. Additionally, the distribution of reporting latency and timestamps associated with each event may differ between the feeds.
Note that these definitions are phrased for the best bid, but apply similarly to the best offer.

Definition VII.11 (*Price Discrepancy*)**.** *A bid price discrepancy is said to occur when the best bid price differs between D_1 and D_2, i.e.*

$$\Delta BB(i,t) = BB(D_1, i, t).price - BB(D_2, i, t).price \neq 0.$$

Definition VII.12 (*Market Inefficiency*)**.** *A market inefficiency occurs whenever a market participant is able to systematically profit from a price discrepancy, usually via the purchase and immediate sale of i.*

Definition VII.13 (*Dislocated Data Feeds*)**.** *D_1 and D_2 are dislocated with respect to the best bid of i at a time t if there is a bid price discrepancy between D_1 and D_2.*

Definition VII.14 (*Dislocation*)**.** *A dislocation between D_1 and D_2 occurs whenever they are dislocated with respect to the best bid of i over a half-open interval of time $[a, b)$.*

Definition VII.15 (*Differing trade*)**.** *A trade is referred to as a differing trade if it occurs during the lifetime of a dislocation.*

Definition VII.16 (*Latency Arbitrage Opportunity*)**.** *A latency arbitrage opportunity with respect to the best bid of i is any half-open interval of time, $[a, b)$, where D_1 and D_2 are dislocated with respect to the best bid of i and $sgn(\Delta BB(i,t)) = sgn(\Delta BB(i,a))$ $\forall t \in [a, b)$.*

Definition VII.17 (*Direction*)**.** *The direction of a latency arbitrage opportunity over an interval $[a, b)$ is defined as $sgn(\Delta BB(i, a))$.*

Definition VII.18 (*Duration*)**.** *The duration of a dislocation or latency arbitrage opportunity over an interval $[a, b)$ is defined as $b - a$.*

Definition VII.19 (*Magnitude*)**.** *The magnitude of a dislocation or latency arbitrage opportunity over an interval $[a, b)$ may be defined as one of the following:*

$$max\_mag = \max_{t \in [a,b)} \{|\Delta BB(i,t)|\}$$
$$min\_mag = \min_{t \in [a,b)} \{|\Delta BB(i,t)|\}$$
$$mean\_mag = \frac{max\_mag + min\_mag}{2}$$

Definition VII.20 (*Realized Opportunity Cost*)**.** *The Realized Opportunity Cost (ROC) experienced by market participants over a period of time $[a, b]$ is defined as:*

$$\sum_{t \in T} |p_{D_1}(time(t), side(t)) - p_{D_2}(time(t), side(t))|,$$

where T are all trades that occurred at the NBBO in the period $[a, b]$, $time(t)$ is a function that returns the time that trade t executed, $side(\cdot)$ returns the opposite side (bid or offer) of the order that instigated the trade, $p_{D_1}(time, side)$ returns the best price offered on feed D_1 at the given time and on the given side, and $p_{D_2}(time, side)$ provides the same information for feed D_2.

Market Actions

The following definitions provide a high-level description of the purpose and details of some common order types, but are not necessarily representative of implementations at NMS market centers.

Definition VII.21 (*Limit Order*). *Guarantees market participants an execution price no worse than a provided limit price, but does not provide any guarantees about the timeliness of execution. This may be implemented by placing a received limit order into the price queue associated with the provided limit price on the correct side of the book (bid or offer, as specified by the order), assuming that it did not match with a resting order at a better price.*
Fields: Instrument identifier, bid/offer, limit price, desired quantity.

Definition VII.22 (*Market order*). *Guarantees instant execution on a best effort basis, but does not provide any guarantees about the execution price. This may be implemented by matching the market order with the best resting orders on the opposite side of the book until the desired quantity is obtained. A market order may be thought of as a limit order with the limit price set in order to guarantee execution (i.e. 0 for a market offer or infinity for a market bid).*
Fields: Instrument identifier, bid/offer, desired quantity

Definition VII.23 (*Modify*). *Allows market participants to update values associated with resting orders and allows for adaptation to changing market conditions. The main usage of this order is to change the number of shares required to fulfill a particular order, since modifying the limit price of order may cause it to lose its place in its current price queue.*
Fields: Order identifier, field(s) to modify, new value(s)

Definition VII.24 (*Cancel*). *Allows market participants to remove resting orders from the local book prior to execution.*
Fields: Order identifier

Definition VII.25 (*Immediate Or Cancel*). *Often shortened to IOC, this is a modifier which may be applied to any order rather than a stand alone order type. The modifier indicates that the associated order should be executed immediately upon receipt or canceled if immediate execution is not possible.*

Definition VII.26 (*Non-Displayed Orders*). *Orders may be marked with a conditional flag which indicates that they should not be displayed on an exchanges order book, in part or whole. Such orders are sometimes referred to as hidden orders, since market participants can not identify active non-displayed orders in an order book from publicly available information.*

Non-displayed orders may come with some negative consequences including increased fees and decreased execution priority in comparison with displayed orders with identical attributes.

Definition VII.27 (*Midpoint Peg*). *A variety of hidden order that executes at the midpoint of the NBBO, i.e.* $0.5(NBB.price + NBO.price)$.

Latency Arbitrage Strategies

Crossed Market: Exchange - Exchange

The simplest latency arbitrage opportunities occur when the highest bid at one exchange crosses the lowest offer at another, and the prices of both orders are within the NBBO. In this case the arbitrageur may purchase the shares provided by the offer and immediately sell them to the bid, profiting from the difference between the two.

Exchanges are prohibited from publicly displaying quotations which would lock or cross the NBBO (see Reg. NMS Rule 610 / Access Rule), so effective application of this strategy requires access to direct feeds at both market centers.

Crossed Market: Exchange - ATS

In a similar fashion, it is possible for the highest bid of an Alternative Trading System (ATS) to cross the lowest offer of an exchange, and vice-versa. In this case a similar strategy may be applied, where the arbitrageur purchases the shares provided by the offer and immediately sells them to the bid.

This strategy requires that the arbitrageur has access to at least one direct feed to an exchange and at least one direct feed to an ATS.

Crossed Midpoint

If the lowest bid or highest offer at one exchange crosses the midpoint of another exchange, then a midpoint peg order may be used to capture latency arbitrage by purchasing shares if the midpoint was crossed by a bid or selling shares if crossed by an offer, assuming the crossing order and midpoint both fall within the NBBO. Additionally, if the midpoint of one exchange crosses the midpoint of another exchange then the arbitrage may be completed via two midpoint peg orders.

REGULATION NATIONAL MARKET SYSTEM

Regulation National Market System (Reg. NMS) is the set of regulations which defines much of the macro-level organization of the U.S. NMS. The primary goal of Reg. NMS is the creation a unified National Stock Market, additionally it has two secondary goals: to promote competition between markets and between orders, and to serve the interests of long-term investors and listing companies [39]. Reg. NMS is composed of several rules and regulations, the most important of which are summarized below. See [39] for more details.

Order Protection Rule

The Order Protection Rule (Rule 611), also known as the Trade-through Rule, is meant to protect orders from trade-throughs, which occur when a market center matches an order against a local counter-party when a better price is available via a protected quotation displayed by an alternative market center. Note that a "better" price in this context is defined from the perspective of the new order entering the market, a.k.a. a liquidity demanding or liquidity consuming order. Therefore a lower execution price is be considered better for an entering bid (offer to buy), while a higher execution price is be considered better for an entering offer (offer to sell).

A protected quotation is defined in Reg. NMS as a bid or offer quotation that satisfies the following properties: the quotation must be automated, the quotation must be displayed by an automated trading center, and the quotation must offer the lowest offer price or highest bid price among all publicly displayed quotations.

A quotation is considered automated if it may be executed without human intervention (up to the full listed quantity), allows for the correct execution of Immediate-Or-Cancel (IOC) orders against the quotation, immediately provides a response to the sender of an Immediate-Or-Cancel order indicating the execution status of that order, and immediately updates the quotation to reflect any changes to its status.

A trading center is considered automated if it implements systems and procedures that allow it to display automated quotations as defined above, and quotations that do not satisfy the requirements of an automated quotation are identified as manual quotations as quickly as possible.

Trade-throughs are prohibited Under Rule 611, however exceptions are allowed for Intermarket Sweep Orders (ISO), quotations displayed by markets that fail to meet the reporting requirements for automated quotations, and flickering quotations with multiple prices displayed in a single second.

Access Rule

The Access Rule (Rule 610) concerns itself with setting standards for access to quotations in NMS stocks, and caps the fees that an exchange may charge for accessing its protected quotations at $0.003 per share. Rule 610 allows for the creation and usage of private data feeds, often referred to as direct feeds by market participants since they are offered directly by exchanges rather than through a third party. Rule 610 also prohibits trading centers from displaying quotations which would lock or cross a protected quotation from a different trading center.

A market is said to be locked if the bid-offer spread of that market is zero, in other words there exists a resting bid and a resting offer with identical limit prices. A market is said to be crossed if the bid-offer spread of that market is negative, i.e. there exists a resting bid whose limit price is greater than the limit price of a resting offer, or equivalently a resting offer exists whose limit price is less than the limit price of a resting bid. These effects are the result of coupling geographically fragmented exchanges, since an order that may lock or cross a market would immediately find a counter-party if the two orders were present on the same exchange.

Sub-Penny Rule

The Sub-Penny Rule (Rule 612) prohibits market participants from displaying or accepting quotations for NMS stocks priced in an increment less than $0.01 unless the quotation price is less than $1.00, in which case the minimum increment is $0.0001. Rule 612 is meant to prohibit the practice of "sub-pennying" in which market participants could "step ahead" of a protected quotation by providing a negligible amount of price improvement, allowing the "sub-pennied" order faster execution at effectively no extra cost.

The significance of this rule, with respect to geographic fragmentation and market inefficiencies, is that the minimum increment for the quoted price of a traded instrument sets the minimum magnitude of all arbitrage opportunities.

Market Data Rules

Rules 601 and 603 are referred to as Market Data Rules and are meant to promote wide availability of market data, thus providing all market participants with an accurate and reliable source of information on the best prices in NMS stocks. These rules cover the organization of a consolidated data feed for NMS stocks, the reward structure for contributing information to the consolidated data feed, and establishes standards for quote and trade information provided to and provided by the consolidated data feed.

In particular these rules concern the Consolidated Tape Association (CTA) plan which disseminates transaction information for NYSE listed securities, the Consolidated Quotation (CQ) plan which disseminates quote information for NYSE listed securities, and the Nasdaq UTP plan which disseminates quote and trade data for Nasdaq listed securities. The information provided by the CTA plan and CQ plan forms Consolidated Tape A, and the information provided by the UTP plan forms Consolidated Tape C. There also exists a Consolidated Tape B which reports trade information for stocks listed on regional exchanges. The aggregation of Consolidated Tapes A, B, and C form what is commonly referred to as the SIP feed.

FIGURES

TABLES

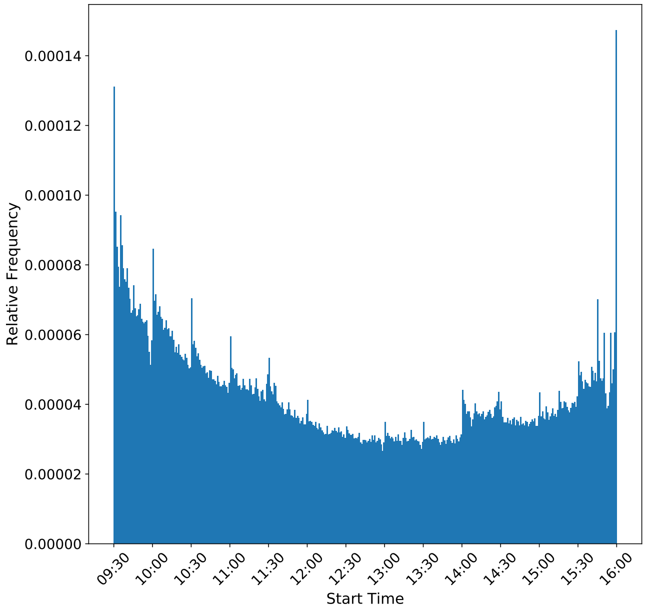

FIG. 12. Distribution of latency arbitrage opportunities with duration longer than 545μs.

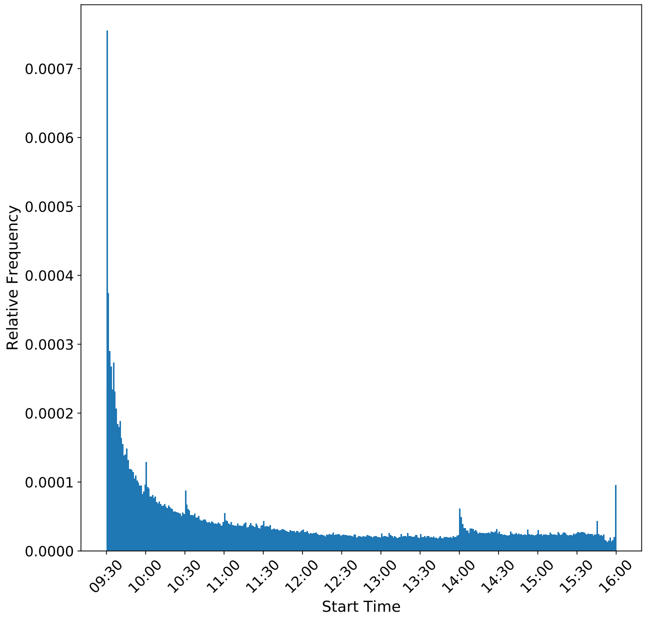

FIG. 13. Distribution of latency arbitrage opportunities with duration longer 545μs and minimum magnitude greater than $0.01. Note that the distribution is heavily skewed right; a plurality of actionable latency arbitrage opportunities occur in the half-hour following the opening bell when compared to any other half-hour during the day. There is also a spike in the number of latency arbitrage opportunities in the middle of the afternoon; we do not speculate regarding the reasons for this phenomenon.

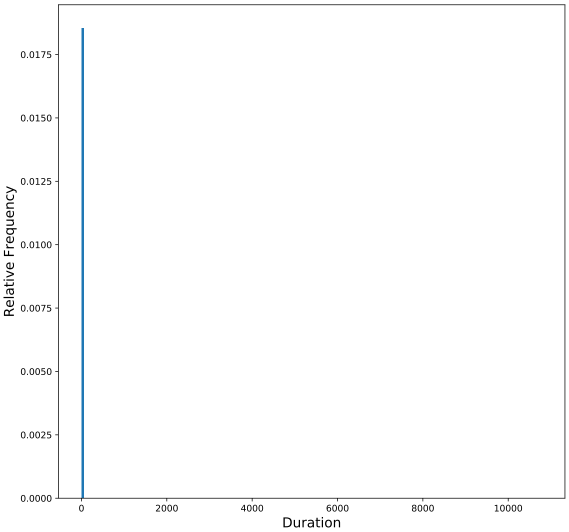

FIG. 14. Distribution of latency arbitrage opportunity durations.

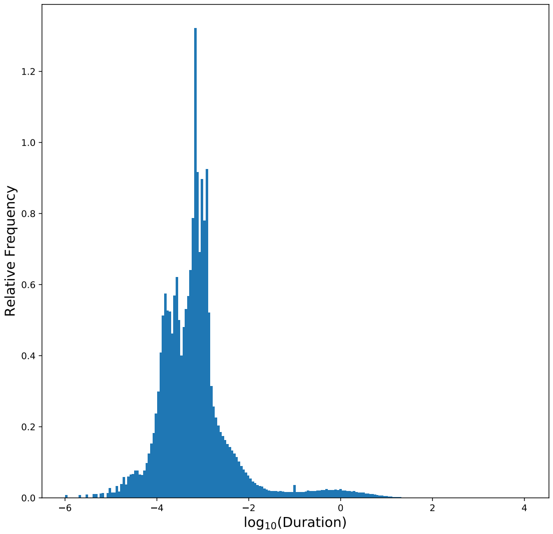

FIG. 15. Distribution of latency arbitrage opportunity durations with a logged x-axis.

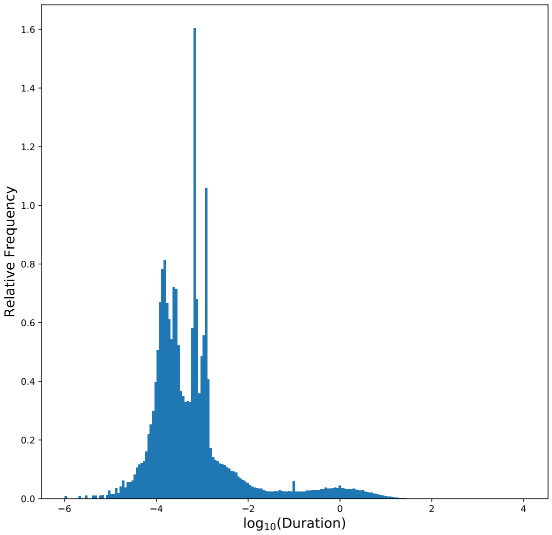

FIG. 16. Distribution of arbitrage opportunity durations with a logged x-axis, conditioned on opportunities with magnitude greater than $0.01.

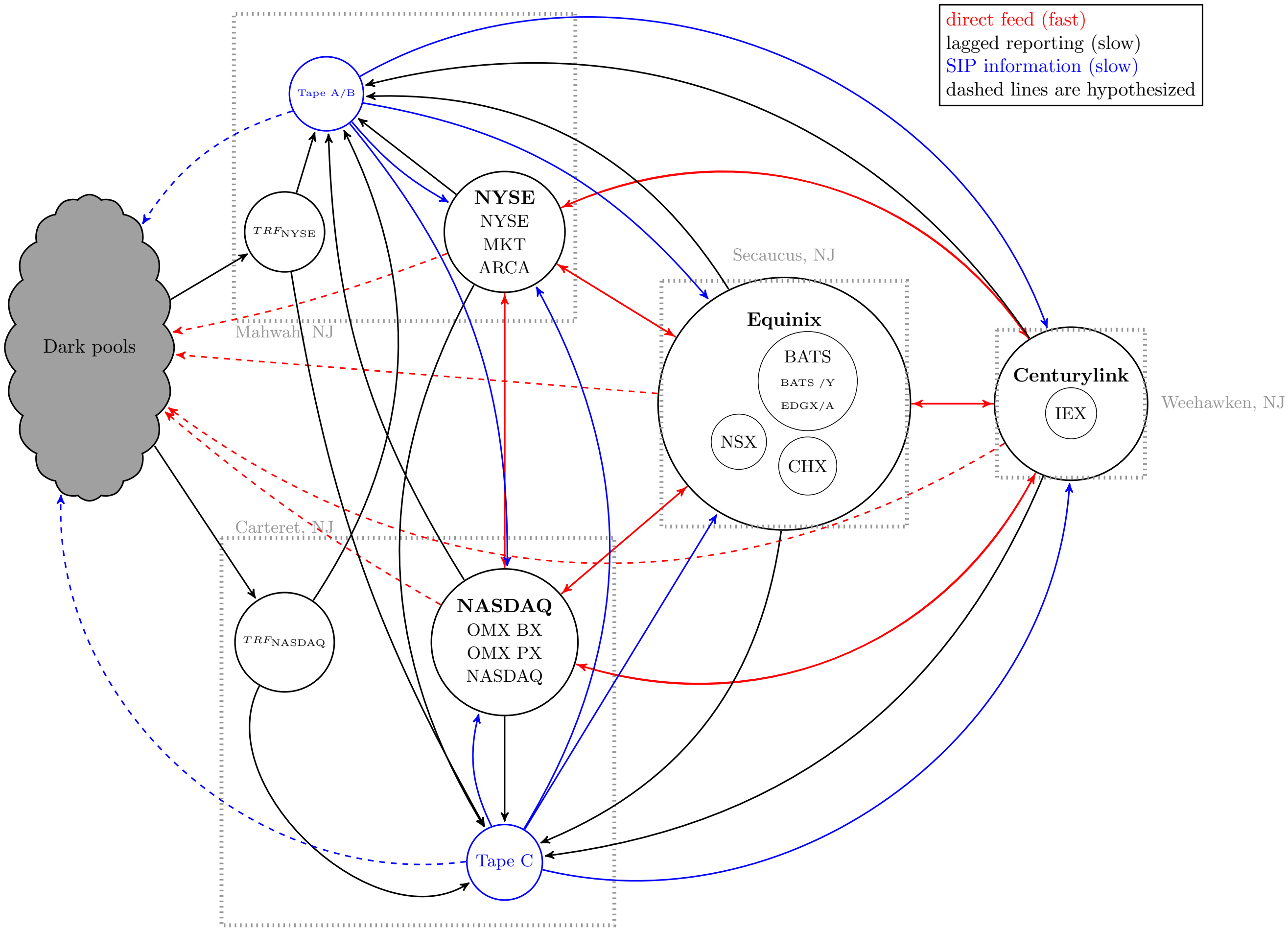

FIG. 17. The full NMS with no centralized observer.

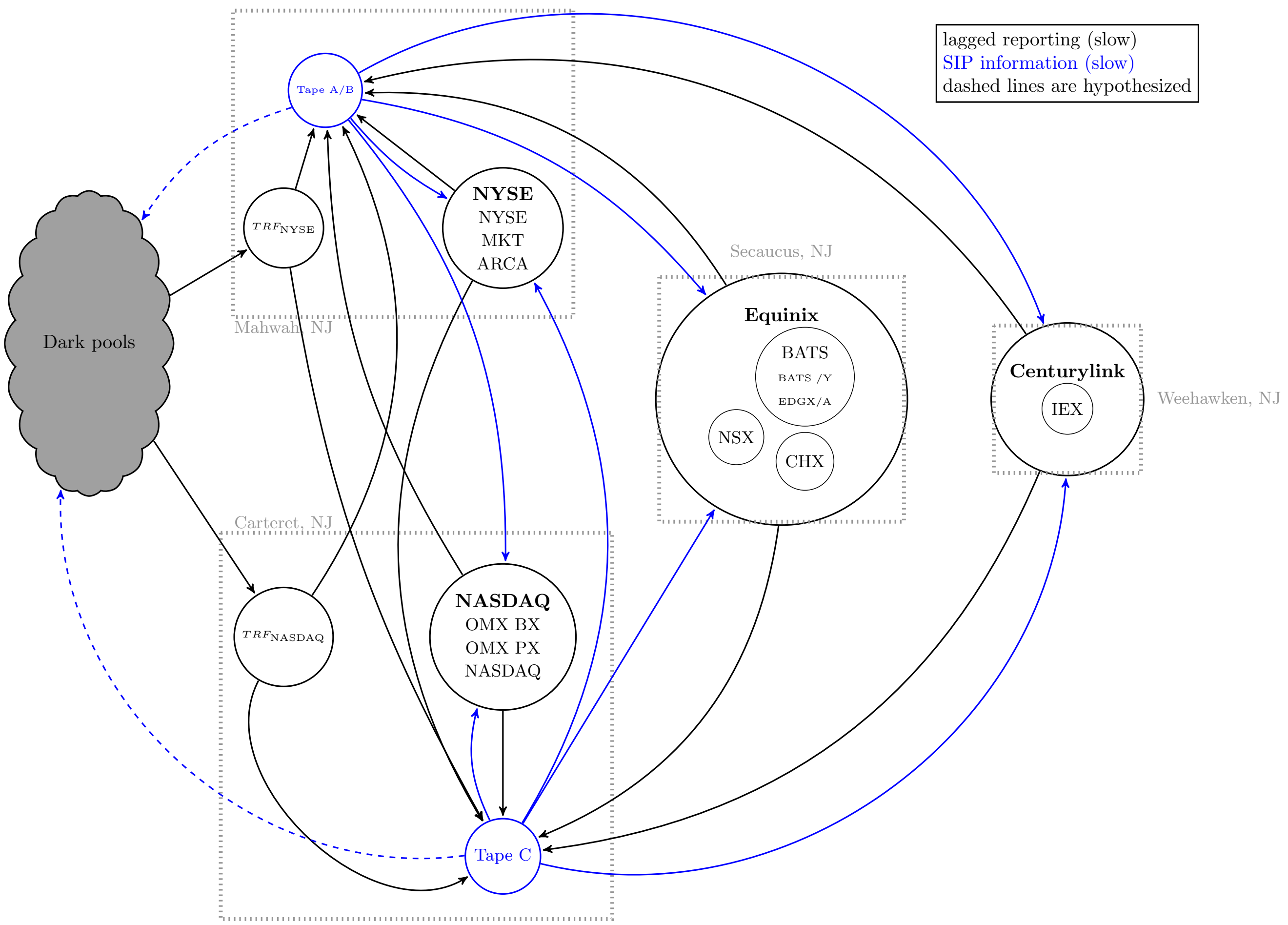

FIG. 18. The NMS as viewed by a market participant with access to only the SIP data feed. Aside from the obvious absence of the direct data feeds, note the lack of a central observer, requiring the synthesis of information from both locations of the SIP tapes.

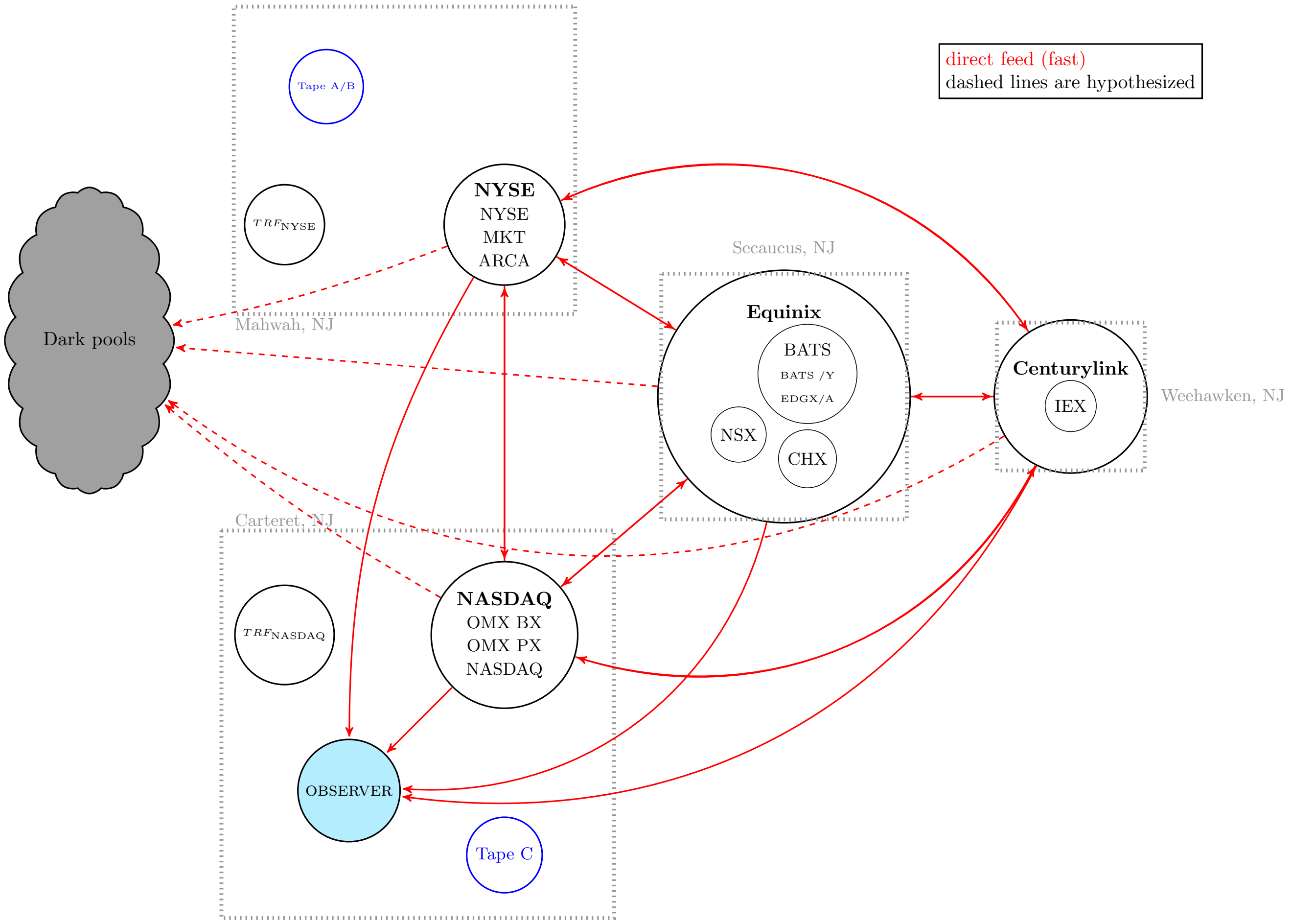

FIG. 19. The NMS as viewed from an observer in Carteret without access to the SIP data feeds. This avoids incurring the cost of the SIP data feeds but removes the possibility of fully understanding the structure and distribution of dislocations on a global (market-wide) scale.

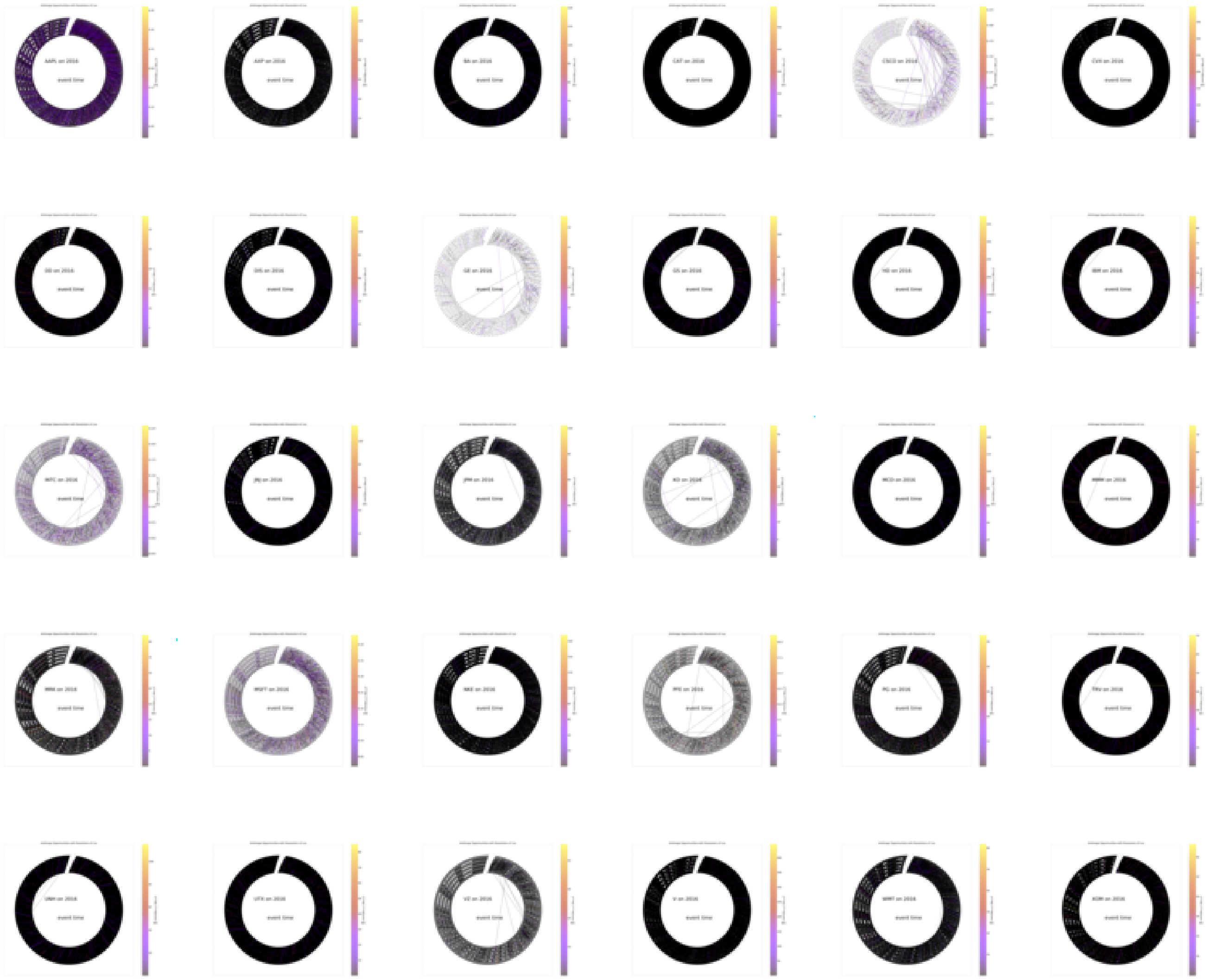

FIG. 20. Event-time circle plots of all Dow 30 tickers.

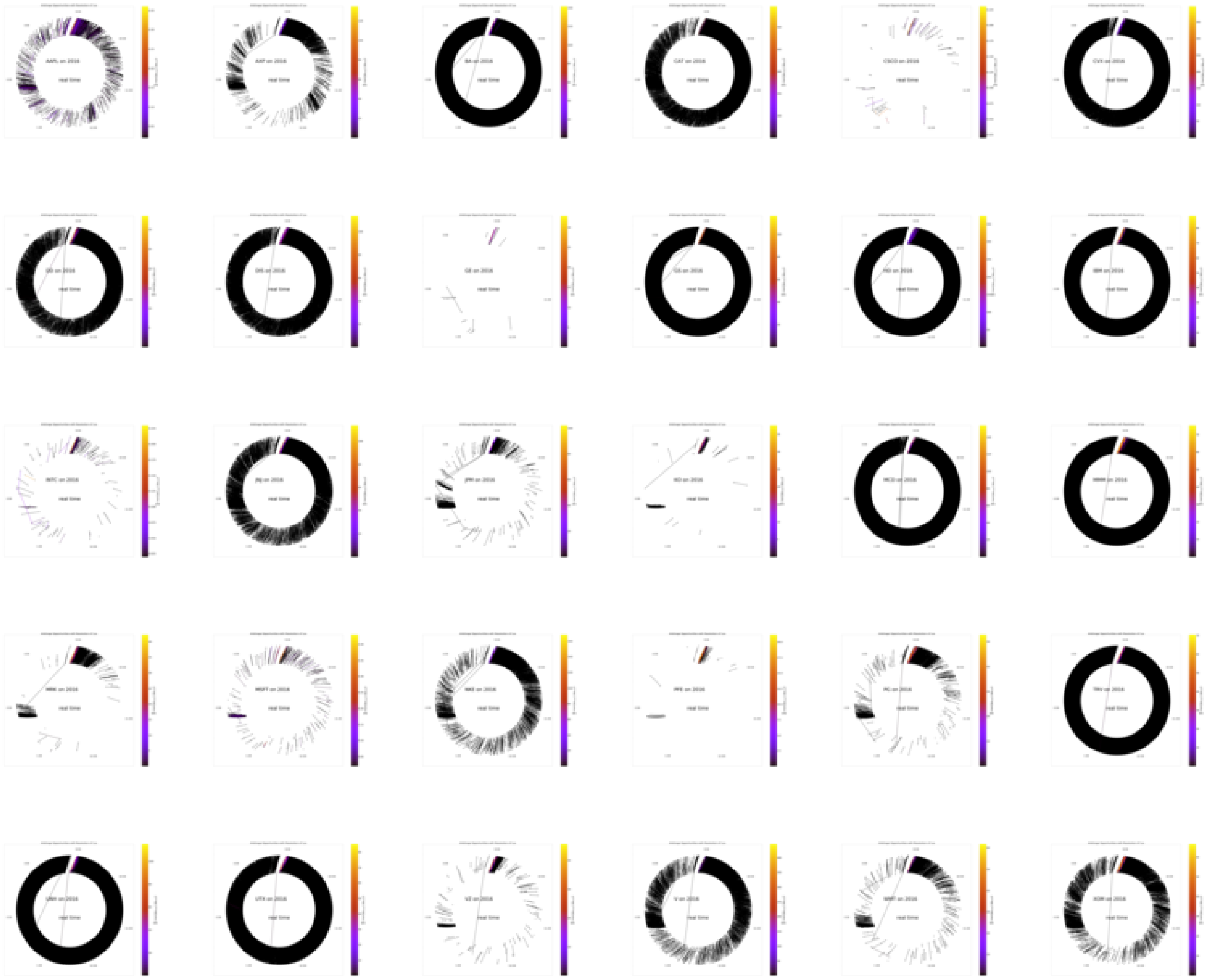

FIG. 21. Real-time circle plots of all Dow 30 tickers.

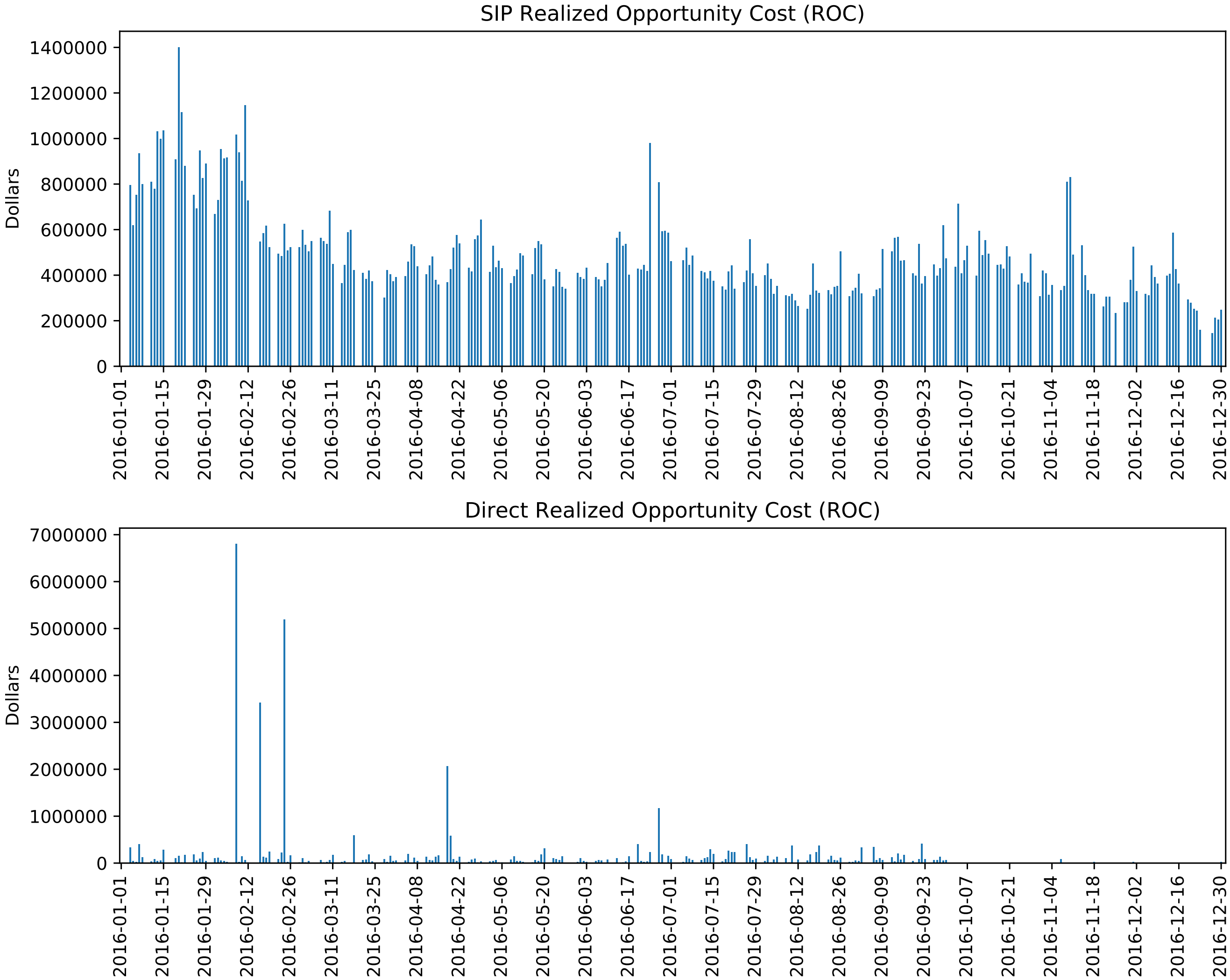

FIG. 22. Daily ROC during calendar year 2016. A large majority of days favored the direct data feeds when aggregated across all tickers.

Symbol		Trades	Traded Val	Diff Trades	Diff Traded Val	ROC	ROC/Share
AAPL	mean	174,820.85	2,542,188,952.00	34,316.58	483,265,898.89	45,852.81	0.007569
	std	68,897.09	1,040,923,482.82	20,556.77	280,321,422.73	27,275.35	0.001224
	min	54,824	983,856,430.54	2,112	35,009,317.38	2,773.35	0.004007
	25%	129,830	1,872,512,861.35	23,498.75	340,459,129.85	32,088.96	0.007307
	50%	156,198.50	2,272,037,106.11	32,741	452,246,993.35	43,204.10	0.007682
	75%	199,793.25	2,870,019,105.28	42,674.75	599,146,631.96	57,647.05	0.008351
	max	517,270	8,280,915,338.59	103,885	1,596,912,962.05	138,331.08	0.011924
AXP	mean	32,348.46	250,614,304.97	9,086.69	71,464,081.61	11,622.14	0.008811
	std	16,110.77	143,031,721.64	4,434.64	36,283,858.01	7,156.73	0.000757
	min	11,095	90,438,986.65	2,219	19,241,382.52	2,666.91	0.007285
	25%	22,756.50	168,209,590.34	5,999.50	49,149,197.52	7,672.38	0.008324
	50%	26,835	207,178,850.49	7,476	57,481,058.84	8,987.37	0.008723
	75%	37,067.75	277,456,051.09	10,905.75	86,485,488.61	13,792.36	0.009248
	max	159,135	1,468,245,304.80	31,507	302,294,385.78	75,473.73	0.013393
BA	mean	20,749.26	288,851,358.58	7,071.25	100,312,506.47	10,955.15	0.012812
	std	10,435.29	154,859,396.19	3,027.87	41,626,820.43	6,235.13	0.008003
	min	4,220	60,869,511.81	1,209	22,712,059.97	2,404.84	0.008497
	25%	14,825.75	202,629,761.69	4,865	69,859,447.55	6,607.66	0.010588
	50%	18,904	260,864,798.97	6,613.50	95,165,851.81	9,608.32	0.011730
	75%	24,641.25	339,733,518.51	8,877.75	123,131,081.82	13,061.48	0.013252
	max	101,159	1,496,951,020.26	19,630	303,000,376.46	47,010.92	0.131181
CAT	mean	30,586.73	269,579,023.84	9,239.74	81,143,988.26	11,986.17	0.010142
	std	11,384.23	107,296,519.47	3,721.88	30,953,831.34	8,988.83	0.005617
	min	7,660	72,342,016.91	2,283	24,025,499.19	2,847.50	0.007044
	25%	22,670.50	204,956,948.21	6,684.75	58,633,048.01	7,680.36	0.008842
	50%	28,267	245,802,664.58	8,451	76,013,433.31	10,301.82	0.009394
	75%	36,304.25	323,347,949.42	10,730	95,123,308.69	13,455.58	0.010105
	max	77,886	964,799,514.35	22,381	222,261,612.89	100,244.92	0.084153
CSCO	mean	77,364.30	493,693,519.98	11,555.12	74,134,548.33	26,409.30	0.008899
	std	33,235.82	207,062,395.07	8,173.36	50,695,319.75	19,401.61	0.001000
	min	31,865	182,535,557.30	660	4,502,758.25	1,461.77	0.005896
	25%	58,015	367,489,467.23	6,881	46,381,850.87	15,394.81	0.008321
	50%	68,328.50	444,190,912.86	10,643	70,264,638.23	23,922.43	0.009015
	75%	86,368.50	548,980,902.84	14,364.50	92,558,544.11	32,439.85	0.009212
	max	307,808	1,702,786,754.09	58,922	316,907,129.91	130,317.79	0.019144
CVX	mean	44,441.79	462,460,384.39	12,439.81	134,648,014.78	17,036.07	0.012786
	std	17,816.08	164,739,606.44	6,693.09	56,874,558.81	16,228.50	0.034249
	min	13,879	144,582,207.22	2,377	28,830,654.89	2,456.15	0.006135
	25%	32,594.50	346,722,417.91	8,344.75	97,784,127.50	9,772.81	0.008240
	50%	39,655.50	430,819,298.93	10,794.50	122,659,276.18	12,993.21	0.008796
	75%	53,123.50	538,846,282.98	14,257.25	158,567,573.75	18,798.64	0.009602
	max	148,515	1,263,782,534.87	50,186	423,871,063.95	190,901.32	0.531465
DD	mean	18,036.06	132,521,012.09	4,913.74	37,476,052.09	6,342.15	0.009882
	std	8,759.67	63,295,360.33	2,764.13	19,045,796.05	4,403.69	0.004642
	min	5,262	40,582,912.43	773	6,491,584.69	832.8600	0.004446
	25%	12,017	89,557,470.32	3,123.50	24,611,595.27	3,912.60	0.008929
	50%	15,462	114,690,819.55	4,104.50	32,687,710.05	5,066.52	0.009511
	75%	20,793.50	155,332,165.60	5,499	44,016,133.46	7,243.21	0.010036
	max	52,298	418,605,566.86	15,217	113,435,890.12	42,392.91	0.080300
DIS	mean	41,156.78	495,392,306.10	10,535.97	129,495,234.75	39,331.64	0.024431
	std	15,686.14	208,901,188.83	4,550.24	53,409,308.00	323,220.91	0.153256
	min	17,030	203,854,389.74	2,633	36,156,641.31	3,221.94	0.006826
	25%	31,892.25	374,803,933.38	7,657.50	97,517,442.34	10,284.53	0.008200
	50%	36,745.50	430,039,189.49	9,220.50	114,691,273.09	13,055.17	0.008814
	75%	45,623.25	558,118,900.42	11,989	144,599,501.14	17,818.25	0.010073
	max	124,145	1,659,028,038.95	32,212	369,007,239.36	5,138,897.26	2.4261
GE	mean	83,963.26	741,830,493.33	12,828.05	119,789,470.34	44,606.60	0.011460
	std	35,661.52	309,010,418.40	8,012.21	69,234,760.11	71,027.27	0.009108
	min	27,905	290,466,991.56	2,653	33,035,264.68	7,844.38	0.005032
	25%	59,365.25	514,268,974.25	7,603	74,660,777.81	23,089.88	0.008105
	50%	74,767.50	670,261,948.27	10,589	96,944,717.80	29,655.10	0.009088
	75%	96,517	876,527,780.45	14,826	141,143,821.44	45,447.71	0.009994
	max	236,395	1,961,985,442.37	49,675	427,596,291.36	1,020,533.87	0.074863
GS	mean	16,072.52	266,630,735.82	6,039.60	100,455,871.70	12,632.51	0.018917
	std	6,759.46	124,491,943.62	2,299.09	38,813,844.47	7,817.49	0.008519
	min	5,914	106,821,197.38	1,908	43,864,040.44	4,126.93	0.009892
	25%	11,672.50	178,117,450.68	4,400.50	72,797,279.72	8,094.88	0.015293
	50%	14,285	224,601,809.66	5,593.50	89,806,560.53	10,478.40	0.017610
	75%	18,995.50	329,800,789.60	7,207.25	123,468,835.06	14,081.00	0.020349
	max	50,816	857,877,495.97	14,393	247,177,637.53	72,612.29	0.117195
HD	mean	27,728.62	366,840,862.69	8,920.89	123,442,984.04	12,744.17	0.011027
	std	8,963.39	127,799,636.38	3,551.59	45,816,593.86	13,953.86	0.005344
	min	13,006	165,434,810.90	2,515	36,439,575.01	2,864.63	0.007334
	25%	21,668.50	276,025,739.12	6,473	92,772,210.99	7,812.18	0.009211
	50%	25,747	339,935,686.44	8,234	115,952,305.04	9,863.20	0.009837
	75%	31,341.50	416,697,720.23	10,762	146,870,527.07	13,201.84	0.010717
	max	64,114	1,031,531,952.92	22,597	291,592,154.20	186,403.78	0.059372
IBM	mean	19,503.60	283,053,487.10	6,540.58	97,629,157.53	10,322.91	0.023045
	std	7,762.60	121,204,978.80	3,031.09	41,935,403.95	11,852.61	0.155151

Continued on next page

Symbol		Trades	Traded Val	Diff Trades	Diff Traded Val	ROC	ROC/Share
	min	6,168	83,951,134.35	1,493	24,252,638	2,042.64	0.007853
	25%	14,595	209,597,644.74	4,586	71,732,705.50	5,972.78	0.010419
	50%	17,729	252,167,134.97	5,852.50	89,826,517.09	7,844.79	0.011260
	75%	22,431.25	328,204,236.66	7,532.75	111,719,286.16	10,385.29	0.012460
	max	59,625	972,131,459.03	21,810	299,050,973.50	111,628.46	2.4712
INTC	mean	88,012.92	539,061,461.61	13,623.27	80,485,200.80	24,652.76	0.008581
	std	32,133.18	218,280,102.40	8,604.73	50,349,950.02	16,048.53	0.001366
	min	25,392	174,808,926.57	668	3,512,129.76	906.3800	0.003979
	25%	66,319.50	409,452,090.75	8,564.50	53,076,046.07	15,370.02	0.008123
	50%	81,767	493,100,646.52	13,526	79,046,604.94	23,962.08	0.008955
	75%	100,219.25	601,791,580.25	17,608.50	104,796,930.94	32,243.13	0.009165
	max	233,578	1,765,833,707.79	48,079	318,483,188.44	91,380.43	0.017641
JNJ	mean	41,248.16	516,784,968.61	10,117.01	132,739,127.27	15,971.14	0.011066
	std	13,010.19	163,195,302.28	4,751.53	54,033,725.20	24,562.10	0.009799
	min	15,606	194,794,413.45	2,156	34,113,674.01	3,046.94	0.006887
	25%	32,847.50	413,846,348.61	7,231.25	98,042,130.74	8,347.87	0.008033
	50%	38,411.50	483,292,741.16	8,718	117,582,458.80	10,975.76	0.008545
	75%	45,961.50	586,813,347.06	11,288	153,593,921.79	16,623.58	0.009356
	max	94,603	1,244,615,527.23	32,165	338,562,051.69	362,771.34	0.091514
JPM	mean	88,003.57	801,423,694.85	21,356.75	193,852,644.59	29,550.37	0.008671
	std	39,466.22	360,958,601.04	11,483.72	91,730,373.77	14,749.77	0.001427
	min	30,040	331,806,293.97	4,953	58,788,624.46	7,089.25	0.006291
	25%	61,325.75	565,821,050.04	13,638	130,610,914.09	19,065.50	0.007994
	50%	77,139	711,684,130.50	17,913.50	171,373,698.77	25,663.01	0.008471
	75%	101,690.75	948,789,239.22	25,153	232,200,018.53	34,390.20	0.008981
	max	256,973	3,004,137,079.38	70,052	646,651,792.53	92,386.71	0.019550
KO	mean	52,120.74	406,264,869.51	10,086.25	81,371,474.40	18,263.90	0.009458
	std	19,287.46	161,269,975.11	4,577.72	33,628,799.23	8,429.36	0.003791
	min	19,958	185,384,176.07	3,156	30,732,830.23	7,111.74	0.006435
	25%	39,138.50	301,353,437.57	7,209.25	59,076,462.53	13,153.95	0.008509
	50%	47,536.50	368,857,020.57	8,995	74,612,460.50	16,482.05	0.008996
	75%	58,796	463,324,316.41	11,326.50	92,283,870.91	20,688.83	0.009567
	max	151,901	1,308,364,552.46	30,895	222,649,014.88	88,890.33	0.059954
MCD	mean	28,809.30	380,847,318.26	7,442.77	103,499,997.57	10,822.55	0.010045
	std	9,250.20	146,529,362.71	2,681.91	39,288,571.37	10,847.35	0.004427
	min	9,911	117,553,924	2,479	32,522,381.94	2,926.51	0.007534
	25%	22,526.25	277,305,454.74	5,422.50	75,412,153.02	6,484.23	0.008638
	50%	26,999.50	355,968,666.10	7,088.50	98,050,825.62	8,795.32	0.009242
	75%	33,173.25	455,898,847.39	8,601.50	121,862,623.83	11,289.40	0.009876
	max	72,028	1,044,773,633.09	20,018	265,940,261.14	114,279.57	0.055288
MMM	mean	11,365.37	167,307,657.17	3,636.52	57,734,183.69	12,063.44	0.017206
	std	3,901.98	56,357,516.03	1,769.07	23,315,655.68	102,963.80	0.055266
	min	3,704	42,376,029.54	852	12,737,335.17	1,268.98	0.008572
	25%	8,870.50	128,614,072.44	2,564	42,198,399.30	3,302.92	0.011656
	50%	10,484	156,620,977.62	3,148	53,116,097.50	4,412.75	0.012771
	75%	13,011	192,588,435.88	4,113	67,265,920.46	6,182.75	0.014374
	max	27,168	374,180,512.59	11,339	141,420,561.60	1,638,916.42	0.888354
MRK	mean	52,065.45	404,241,094.10	12,269.51	97,773,420.29	17,435.31	0.008974
	std	21,247.82	198,964,179.96	6,450.80	49,864,763.12	10,578.05	0.002336
	min	18,727	139,953,296.94	4,541	37,156,365.39	5,935.82	0.005243
	25%	39,157.50	302,275,577.81	7,789	64,970,853.90	10,991.95	0.008151
	50%	46,619.50	360,181,487.45	10,518	84,618,791.63	15,107.30	0.008574
	75%	58,293.50	460,791,595.64	13,950.25	113,701,288.72	19,811.20	0.008998
	max	232,717	2,584,131,245.57	46,595	456,348,016.09	112,089.37	0.027925
MSFT	mean	141,856.07	1,190,901,402.50	24,761.04	203,129,267.74	36,706.48	0.008303
	std	63,588.22	533,057,860.34	17,480.90	139,593,661.04	25,629.62	0.001006
	min	37,036	459,917,664.02	1,070	8,596,209.02	1,253.33	0.004649
	25%	102,602.75	837,915,412.53	14,050.50	122,690,779.06	22,117.84	0.007944
	50%	124,327.50	1,053,837,918.53	22,263	180,976,495.62	32,474.83	0.008414
	75%	156,482	1,373,674,878.44	32,070.75	262,037,990.81	48,649.14	0.008943
	max	456,106	4,125,126,448	98,307	950,946,403.87	138,913.71	0.010702
NKE	mean	46,386.10	377,535,172.78	10,935.36	89,164,054.37	18,227.11	0.009535
	std	15,357.30	145,806,631.19	3,796.12	32,750,534.77	20,652.07	0.006259
	min	13,818	84,721,641.01	2,885	18,676,669.13	3,523.46	0.005848
	25%	37,737.50	295,226,871.17	8,613.25	69,144,244.49	12,031.69	0.008070
	50%	42,544	344,601,219.52	9,822	80,592,736.50	14,390.93	0.008480
	75%	51,532.50	424,862,753.77	12,534	102,100,476.22	18,212.90	0.008969
	max	121,962	1,195,681,284.35	28,410	232,923,873.16	280,266.40	0.084753
PFE	mean	91,040.68	692,324,391.87	13,862.73	110,715,986.10	31,625.70	0.009084
	std	49,256.08	473,362,104.74	6,672.49	60,406,629.99	16,222.08	0.002189
	min	32,599	212,898,806.65	4,422	28,855,501.39	8,447.34	0.005328
	25%	59,097.50	426,001,630.46	9,726.75	74,658,424.30	21,093.39	0.008060
	50%	80,628	611,356,656.02	13,270.50	103,824,254.59	29,745.09	0.008872
	75%	109,044.50	783,454,707.50	16,379	129,458,335.77	36,810.46	0.009218
	max	474,221	5,427,524,575.47	56,238	602,885,333.66	145,936.99	0.021475
PG	mean	50,438.27	570,844,223.65	11,760.85	134,139,256.91	17,786.87	0.011319
	std	26,464.80	419,733,122.26	5,828.25	70,501,433.31	13,441.76	0.027016
	min	19,980	185,431,171.67	3,696	40,926,831.50	4,789.61	0.005871
	25%	34,682.50	362,299,048	7,530.75	87,831,864.68	10,158.57	0.007830

Continued on next page

Symbol		Trades	Traded Val	Diff Trades	Diff Traded Val	ROC	ROC/Share
	50%	43,215.50	456,219,304.50	10,168.50	113,664,173.39	14,134.84	0.008337
	75%	57,796	612,257,033.06	14,163	160,102,759.35	20,335.25	0.008869
	max	181,697	3,330,428,860.98	38,467	460,594,145.16	111,040.42	0.427532
TRV	mean	10,544.19	106,389,400.10	3,568.88	39,506,286.77	4,441.92	0.011206
	std	3,416.58	36,241,051.32	1,447.62	15,393,415.99	2,794.22	0.002964
	min	3,018	27,592,851.46	771	7,628,101.68	964.9800	0.007730
	25%	8,487.25	82,492,360.44	2,705	29,702,903.75	2,990.61	0.009813
	50%	9,965.50	101,071,670.21	3,334.50	37,475,398.81	3,837.38	0.010699
	75%	12,010.25	122,831,584.80	4,172.25	46,650,233.97	4,933.57	0.011895
	max	27,468	294,476,802.95	11,339	107,591,813.81	28,594.17	0.048296
UNH	mean	17,446.67	228,660,097.56	5,642.65	77,377,042.02	7,680.73	0.011369
	std	5,246.70	81,435,935.28	2,011.09	27,032,487.15	4,216.99	0.001956
	min	6,412	89,234,548.68	1,849	26,225,274.13	2,378.89	0.008077
	25%	14,129	173,512,633.11	4,413.50	59,089,553.72	5,357.03	0.010139
	50%	16,636	214,637,619.41	5,371.50	75,046,042.50	6,539.35	0.010909
	75%	19,932.50	260,900,529.94	6,717.75	92,546,473.56	8,912.00	0.012157
	max	41,842	725,532,688.10	15,652	218,550,591.96	30,826.07	0.020873
UTX	mean	24,903.26	263,375,122.23	8,217.23	88,158,366.16	17,510.92	0.011823
	std	12,739.37	141,586,417.29	4,913.50	47,017,041.73	109,897.32	0.025491
	min	5,358	49,595,310.95	1,315	13,549,323.03	1,579.70	0.007425
	25%	16,977	182,655,118.69	4,942.75	59,862,917.81	6,260.83	0.009197
	50%	21,463	229,710,235.45	7,034.50	77,766,133.91	8,629.70	0.009769
	75%	27,806.50	295,642,614.43	9,447.25	101,084,369.56	11,780.77	0.010558
	max	86,284	1,144,629,181.20	29,297	275,444,139.17	1,749,683.12	0.413092
V	mean	48,950.33	460,497,961.48	13,097.62	122,925,302.52	16,818.68	0.009524
	std	17,793.66	170,963,876.45	6,162.95	50,951,708.77	9,603.52	0.007991
	min	23,142	162,781,451.21	3,273	30,311,313.92	3,873.41	0.005686
	25%	36,797	351,926,962.18	9,092.75	88,092,316.21	11,157.63	0.007977
	50%	44,660	411,627,578.41	11,552	112,507,821.59	14,624.21	0.008478
	75%	56,347	531,962,904.82	14,735	139,117,677.04	18,457.99	0.009089
	max	128,775	1,261,830,529.49	42,661	355,125,487.75	85,584.79	0.120466
VZ	mean	62,098.01	494,149,523.80	13,525.58	109,544,287.76	51,450.08	0.013465
	std	23,339.73	185,520,474.01	5,963.55	44,308,281.09	427,124.11	0.030589
	min	29,671	204,408,079.42	5,039	41,836,595.67	6,539.31	0.005940
	25%	46,137.50	362,469,762.67	9,445.50	77,277,337.22	14,070.98	0.008312
	50%	55,823.50	449,943,857.51	11,922	100,842,957.95	19,546.38	0.008925
	75%	71,345	574,383,307.99	15,340.50	128,753,322.82	27,179.16	0.009833
	max	147,919	1,264,130,771.03	36,340	266,067,716.19	6,798,041.07	0.469146
WMT	mean	49,823.30	448,218,124.11	11,786.63	109,524,107.26	19,815.12	0.011010
	std	20,042.63	187,614,765.33	5,642.63	49,138,231.17	26,412.77	0.013753
	min	20,706	211,540,076.99	3,709	34,219,678.86	4,605.33	0.006303
	25%	36,156.25	325,522,820.91	7,935	74,826,489.33	10,770.16	0.008199
	50%	44,622.50	399,048,171.16	10,657	99,148,394.58	14,630.73	0.008728
	75%	57,546.50	520,989,325.68	13,105.25	125,786,171.91	19,936.97	0.009294
	max	156,021	1,562,166,750.41	36,698	361,429,655.92	246,675.56	0.176158
XOM	mean	64,074.02	670,862,447.91	17,774.64	188,657,442.78	35,104.83	0.013337
	std	28,483.97	265,569,760.30	10,924.37	93,450,720.08	127,162.28	0.021539
	min	21,646	201,555,090.63	4,205	46,296,094.35	4,953.61	0.005362
	25%	46,888	496,895,704.13	11,816.25	129,373,837.81	15,072.46	0.007813
	50%	55,080.50	593,690,988.09	14,020	162,976,539.44	18,862.46	0.008369
	75%	74,045.75	786,147,285.48	19,397.50	211,792,668.60	31,372.43	0.010116
	max	209,816	1,761,362,028.61	75,421	613,405,517.24	2,003,841.58	0.238129

TABLE V: Summary ROC statistics for Dow 30 stocks, aggregated by day and trading symbol.

CALCULATING REALIZED OPPORTUNITY COST

Calculating Realized Opportunity Cost (ROC) For each trade of interest: - Obtain the Securities Information Processor (SIP) National Best Bid and Offer (NBBO) prices and the Direct Best Bid and Offer (DBBO) at the time of the trade. - Check if the trade executed at one of the NBBO prices. - If yes, then the difference between the execution price and the corresponding price from the DBBO, multiplied by the number of shares transacted, becomes the ROC associated with that trade. Note: Depending on the side of the active order (bid or offer), and the relationship between the NBBO and DBBO, the ROC may be identified as favoring the SIP or a Direct feed. In other words, when the active order could receive price improvement by executing at the price offered by the DBBO, then the ROC becomes associated with the SIP (SIP ROC). Likewise, if the active order received a price improvement by executing at the NBBO rather than the DBBO, then the ROC becomes associated with the Direct feeds (Direct ROC). If no, then the trade is discarded from the analysis, since it is difficult to accurately determine the side of the active order in this situationand knowing the side of the active order is required in order to accurately calculate the directional ROC. Note: ROC experienced on both sides of the book (bid and offer) are aggregated over each day, ticker, and exchange; thus, there may be some cancellation between positive ROC (Direct ROC) and negative ROC (SIP ROC) during the aggregation to determine net ROC for that day-ticker-exchange. The net ROC is therefore

Direct Feed and Historical Data Pricing

Data Provider	Feed(s)	One-time Cost	Monthly Fee
CTA	CQS, CTS		$11,002
UTP	UQDF, UTDF		*$6,000
NYSE	Integrated		$27,500
	Historical	$60,000	
NYSE ARCA	Integrated		$10,000
	Historical	$36,000	
NYSE MKT	Integrated		$7,500
	Historical	$18,000	
National Stock Exchange NSX	Integrated	$?	$?
(Now NYSE National)	Historical	$?	$?
NASDAQ	TotalView-ITCH		$25,000
	Historical		$1,250
NASDAQ BX	TotalView-ITCH		$20,000
	Historical		$500
NASDAQ PSX	TotalView-ITCH		$17,000
	Historical		$500
BATS BZX	Depth		$2,000
	Historical	$8,500	
BATS BYX	Depth		$2,000
	Historical	$8,500	
Direct Edge EDGA	Depth		$1,000
	Historical	$8,500	
Direct Edge EDGX	Depth		$2,000
	Historical	$8,500	
The Investors Exchange	TOPS	$0	$0
	DEEP	$0	$0
	Historical	$0	$0
Chicago Stock Exchange	CHX Book	$0	$0
	Historical	$?	$?
Total		$148,000	$133,252

TABLE VI. The pricing presented in this table assumes a single consumer with a non-display, non-trading use case aiming to construct a dataset similar to what was used in this analysis. Strictly speaking, one need not pay for live direct feeds since historical data is sufficient for replicating the analysis presented in this paper. However, highlighting the monthly cost for comprehensive direct feed access shines a light on one of the reasons for the lack of academic participation in the analysis of modern stock markets. This does not include costs which may be incurred while curating the data, fulfilling potential co-location requirements, ISP/networking costs, computing hardware acquisition and maintenance, etc. Additionally, historical data for NYSE National/NSX and CHX are not included since they are not directly available from the exchange and must be purchased from a third party. This list is not guaranteed to be comprehensive, additional fees/costs may exist. *UTP access fees may be waived for academic institutions, see UTP Feed Pricing for more info. The sources used to construct this table involve CTA feed pricing, UTP feed pricing via the Data Policies document, NYSE feed pricing, NYSE historical data pricing, NASDAQ feed pricing, BATS/DirectEdge feed pricing, and CHX feed pricing

a conservative measure, since it is possible that investors could experience both SIP and Direct ROC for that day-ticker-exchange.

Example: In particular, see the 79th trade in Table VII, where 100 shares of AAPL transacted at $99.13 at 9:48:55.398386. The NBBO at that time was (bid @ $99.13, offer @ $99.15), while the DBBO was (bid @ $99.16, offer @ $99.17). Since the trade executed at $99.13, the best bid displayed by the SIP, we infer that the resting order was a bid and the active order was an offer. The ROC is then calculated as ($99.13 per share - $99.16 per share) * 100 shares = (-$0.03 per share) * 100 shares = -$3.00 in favor of the Direct feeds (i.e. SIP ROC). From this example, one can note that when the active order is an offer, then the formula for ROC is (SIP National Best Bid (NBB) - Direct Best Bid (DBB)) * shares. This results in a positive value when the NBO provides price improvement for the active bid and a negative value when the DBO provides price improvement for the active bid. Additionally, see the 95th trade in Table VII where 100 shares transacted at $99.14 at 9:48:55.398560. The NBBO at that time was (bid @ $99.14, offer @ $99.14) and the DBBO was (bid @ $99.16, offer @ $99.17). Since the SIP was locked at the time of execution the active order could have been from either side of the book. For this example, we will focus on the situation where we assume the active order is a bid and the resting order is an offer. The ROC is then calculated as ($99.17 per share - $99.14 per share) * 100 shares = ($0.03 per share) * 100 shares = $3.00 in favor of the SIP (i.e. Direct ROC). Note that in this example, the formula used to calculate the ROC reverses the position of the SIP and Direct prices since the active order is a bid instead of an offer. Thus, the formula for ROC is (Direct Best Offer (DBO) - SIP Best Offer (NBO)) * shares. This maintains the meaning of the sign, where positive values indicate price improvement featured by the NBB and negative values indicate price improvement featured by the DBB (from the perspective of the active order). Thus, ROC from both sides of the book may be treated uniformly in that positive values favor the SIP feed and negative values favor the consolidated Direct feeds. We aggregate the ROC by date, stock, and venue. Since these two trades occurred at the same trading venue, they would be summed, resulting in a net ROC of $0.00. Similar cancelations occur for every date-stock-venue combination resulting in these conservative measures of ROC. Another aspect of conservative measures of ROC In the example dislocation, there were almost 100 differing trades (i.e., trades that occurred while the NBBO and DBBO are dislocated) as contained in Table VII. Yet, our ROC measures only include trade executions at the NBBO. Therefore, we only consider a total of 11 trades (6 on the offer and 5 on the bid) during this dislocation, thus providing additional evidence that our ROC measures are conservative.

Dislocations and Latency Arbitrage Opportunities: From paper 1, Figure 8, we see all dislocations in AAPL on January 7, 2016. We select an arbitrary dislocation to investigate which existed on the offer side from 9:48:55.396886 to 9:48:55.398749 (a duration of 1863 microseconds). This dislocation features a maximum value of $0.06, which occurs between 9:48:55.397644 and 9:48:55.398027 (a duration of 383 microseconds or 20.56% of its lifetime). During this time where the dislocation featured its maximum value, the SIP best offer remained at $99.11 and the Direct best offer remained at $99.17. Thus, any bid orders submitted during this period stood to save $0.06 per share by transacting at the SIP BO rather than the Direct BO, assuming that they could actually locate resting offers at $99.11, either in the lit or dark markets. Note that this dislocation started and ended while the Limit-up Limit-down (LULD) mechanism was in effect (this is engaged at 9:45 each day, following the first 15 minutes of trading), featured a duration greater than 545 microseconds (what we consider to be the minimum duration in order to be actionable) and featured a maximum magnitude greater than $0.01. Note: you can find more info on LULD here: http://www.luldplan.com/index.html.

Connecting Realized Opportunity Cost and Latency Arbitrage Opportunities: The ROC statistic captures events that occurred (i.e. trades) and assigns an opportunity cost to them based on the state of the SIP and Direct feeds at the time of the trade. Hopefully the above example has illustrated the extreme sparsity of our ROC approach, which only considers trades that execute at either side of the prevailing NBBO, features cancellation effects due to aggregation, and does not consider duration/actionability (e.g., could the agent who entered the active order have reasonably reacted to the state of the two feeds?). Latency arbitrage opportunities are constructed to capture the relative states between the NBBO and DBBO through time, observing the dislocations between the two feeds and collecting information about their duration and magnitude. With our approach, we capture the inefficiencies and opportunity costs that actually occurred (i.e., realized), and what inefficiencies and opportunity costs could have occurred (i.e., latency arbitrage opportunities). For illustrative purposes only, if the NBBO and DBBO were tightly synchronized, then the ROC statistic would tend towards $0.00. [Note: there are specific policy reasons in RegNMS that SIP reporting will always lag reporting on the direct feeds, independent of technological infrastructure]. Thus, constructing LAOs so that they only consider the NBBO and DBBO allows us to isolate one component of the ROC statistic and investigate it in greater detail. Additionally, the ROC statistic does not account for duration/actionability, while LAOs allow for such considerations of duration / actionability to be addressed in a simple and direct way. These two measurements, ROC and LAOs, were constructed to investigate similar phenomena from slightly different perspectives to provide complementary and synergistic views of NMS dynamics.

Trade number	Date and time	delta	symbol	size	price	Exchange Number	extra
0	2016-01-07 09:48:55.396951	255	AAPL	100	99.11	1	-651
1	2016-01-07 09:48:55.396951	227	AAPL	100	99.12	1	-651
2	2016-01-07 09:48:55.396978	237	AAPL	100	99.12	1	-678
3	2016-01-07 09:48:55.396978	222	AAPL	100	99.12	1	-678
4	2016-01-07 09:48:55.396978	204	AAPL	100	99.13	2	-852
5	2016-01-07 09:48:55.396998	207	AAPL	100	99.13	2	-872
6	2016-01-07 09:48:55.396998	190	AAPL	100	99.13	2	-872
7	2016-01-07 09:48:55.397064	239	AAPL	100	99.13	2	-938
8	2016-01-07 09:48:55.397064	216	AAPL	100	99.12	1	-764
9	2016-01-07 09:48:55.397068	204	AAPL	50	99.13	2	-942
10	2016-01-07 09:48:55.397196	316	AAPL	200	99.13	2	-1070
11	2016-01-07 09:48:55.397196	296	AAPL	100	99.16	1	-1013
12	2016-01-07 09:48:55.397196	279	AAPL	100	99.13	2	-1070
13	2016-01-07 09:48:55.397196	262	AAPL	395	99.11	3	-1044
14	2016-01-07 09:48:55.397297	344	AAPL	100	99.13	1	-997
15	2016-01-07 09:48:55.397297	327	AAPL	100	99.16	4	-1114
16	2016-01-07 09:48:55.397297	309	AAPL	100	99.13	1	-997
17	2016-01-07 09:48:55.397297	292	AAPL	100	99.14	2	-1171
18	2016-01-07 09:48:55.397297	275	AAPL	100	99.13	1	-997
19	2016-01-07 09:48:55.397297	259	AAPL	100	99.12	3	-1145
20	2016-01-07 09:48:55.397361	306	AAPL	100	99.14	2	-1235
21	2016-01-07 09:48:55.397431	358	AAPL	100	99.13	3	-1279
22	2016-01-07 09:48:55.397431	317	AAPL	100	99.14	2	-1305
23	2016-01-07 09:48:55.397431	298	AAPL	100	99.13	1	-1131
24	2016-01-07 09:48:55.397431	268	AAPL	100	99.13	3	-1279
25	2016-01-07 09:48:55.397499	316	AAPL	50	99.13	1	-1199
26	2016-01-07 09:48:55.397499	299	AAPL	100	99.13	3	-1347
27	2016-01-07 09:48:55.397499	284	AAPL	100	99.13	1	-1199
28	2016-01-07 09:48:55.397504	272	AAPL	100	99.14	2	-1378
29	2016-01-07 09:48:55.397504	255	AAPL	100	99.13	1	-1204
30	2016-01-07 09:48:55.397565	299	AAPL	50	99.13	3	-1413
31	2016-01-07 09:48:55.397565	281	AAPL	100	99.14	1	-1265
32	2016-01-07 09:48:55.397565	266	AAPL	200	99.13	3	-1413
33	2016-01-07 09:48:55.397604	290	AAPL	100	99.14	1	-1304
34	2016-01-07 09:48:55.397604	276	AAPL	100	99.13	3	-1452
35	2016-01-07 09:48:55.397604	260	AAPL	100	99.14	1	-1304
36	2016-01-07 09:48:55.397685	325	AAPL	100	99.14	3	-1533
37	2016-01-07 09:48:55.397685	309	AAPL	100	99.14	3	-1533
38	2016-01-07 09:48:55.397685	293	AAPL	100	99.14	1	-1385
39	2016-01-07 09:48:55.397731	323	AAPL	100	99.14	3	-1579
40	2016-01-07 09:48:55.397731	309	AAPL	100	99.14	1	-1431
41	2016-01-07 09:48:55.397731	294	AAPL	100	99.14	3	-1579
42	2016-01-07 09:48:55.397731	279	AAPL	50	99.15	1	-1431
43	2016-01-07 09:48:55.397767	300	AAPL	100	99.14	3	-1615
44	2016-01-07 09:48:55.397767	285	AAPL	100	99.14	3	-1615
45	2016-01-07 09:48:55.397767	269	AAPL	900	99.15	3	-1615
46	2016-01-07 09:48:55.397824	310	AAPL	100	99.15	1	-1524
47	2016-01-07 09:48:55.397824	294	AAPL	100	99.15	3	-1672
48	2016-01-07 09:48:55.397824	280	AAPL	100	99.14	2	-1698
49	2016-01-07 09:48:55.397824	266	AAPL	100	99.15	2	-1698
50	2016-01-07 09:48:55.397870	298	AAPL	100	99.15	2	-1744
51	2016-01-07 09:48:55.397870	282	AAPL	100	99.15	1	-1570
52	2016-01-07 09:48:55.397894	290	AAPL	100	99.15	3	-1742
53	2016-01-07 09:48:55.397894	275	AAPL	100	99.15	1	-1594
54	2016-01-07 09:48:55.397894	260	AAPL	50	99.15	3	-1742
55	2016-01-07 09:48:55.397973	323	AAPL	50	99.15	1	-1673
56	2016-01-07 09:48:55.397973	307	AAPL	100	99.15	3	-1821
57	2016-01-07 09:48:55.397973	293	AAPL	100	99.15	1	-1673
58	2016-01-07 09:48:55.397994	299	AAPL	50	99.15	2	-1868
59	2016-01-07 09:48:55.398058	346	AAPL	50	99.16	1	-1758
60	2016-01-07 09:48:55.398058	331	AAPL	200	99.15	3	-1906
61	2016-01-07 09:48:55.398058	313	AAPL	100	99.16	1	-1758
62	2016-01-07 09:48:55.398125	366	AAPL	100	99.15	2	-1999
63	2016-01-07 09:48:55.398128	354	AAPL	100	99.15	3	-1976
64	2016-01-07 09:48:55.398147	357	AAPL	100	99.14	1	-1422
65	2016-01-07 09:48:55.398147	342	AAPL	200	99.15	2	-2021
66	2016-01-07 09:48:55.398158	339	AAPL	100	99.16	3	-2006
67	2016-01-07 09:48:55.398177	342	AAPL	100	99.15	2	-2051
68	2016-01-07 09:48:55.398225	375	AAPL	100	99.16	3	-2073
69	2016-01-07 09:48:55.398225	359	AAPL	5	99.16	3	-2073
70	2016-01-07 09:48:55.398225	345	AAPL	100	99.14	1	-1500
71	2016-01-07 09:48:55.398267	373	AAPL	100	99.15	1	-1542
72	2016-01-07 09:48:55.398267	358	AAPL	100	99.16	1	-1542
73	2016-01-07 09:48:55.398267	342	AAPL	100	99.17	1	-1542
74	2016-01-07 09:48:55.398267	327	AAPL	100	99.17	1	-1542
75	2016-01-07 09:48:55.398267	312	AAPL	50	99.17	1	-1542
76	2016-01-07 09:48:55.398272	300	AAPL	400	99.11	5	-1967
77	2016-01-07 09:48:55.398272	285	AAPL	100	99.12	5	-1967
78	2016-01-07 09:48:55.398386	384	AAPL	100	99.13	5	-2081
79	2016-01-07 09:48:55.398414	397	AAPL	100	99.13	5	-2109
80	2016-01-07 09:48:55.398414	381	AAPL	100	99.13	5	-2109
81	2016-01-07 09:48:55.398414	365	AAPL	100	99.13	5	-2109
82	2016-01-07 09:48:55.398444	381	AAPL	50	99.13	5	-2139
83	2016-01-07 09:48:55.398444	366	AAPL	100	99.13	5	-2139
84	2016-01-07 09:48:55.398444	352	AAPL	100	99.13	5	-2139
85	2016-01-07 09:48:55.398444	337	AAPL	100	99.13	5	-2139
86	2016-01-07 09:48:55.398444	322	AAPL	100	99.14	5	-2139
87	2016-01-07 09:48:55.398532	395	AAPL	50	99.14	5	-2227
88	2016-01-07 09:48:55.398532	369	AAPL	100	99.15	2	-1507
89	2016-01-07 09:48:55.398532	354	AAPL	100	99.16	2	-1507
90	2016-01-07 09:48:55.398537	344	AAPL	50	99.17	2	-1512
91	2016-01-07 09:48:55.398537	330	AAPL	100	99.17	2	-1512
92	2016-01-07 09:48:55.398560	339	AAPL	100	99.17	2	-1535
93	2016-01-07 09:48:55.398560	324	AAPL	50	99.17	3	-1282
94	2016-01-07 09:48:55.398560	309	AAPL	100	99.14	5	-1434
95	2016-01-07 09:48:55.398571	305	AAPL	50	99.15	5	-1445
96	2016-01-07 09:48:55.398571	291	AAPL	100	99.15	5	-1445

TABLE VII: Trades that occurred during a dislocation in AAPL on 2016-01-07 at approximately 0948, more than three minutes after the trading “guardrails” are enforced.

timestamp	Exchange number	price	shares	direct_bid	direct_ask	sip_bid	sip_ask	roc
2016-01-07 09:48:55.396951	1	99.11	100	99.14	99.14	99.10	99.11	3.0
2016-01-07 09:48:55.397196	3	99.11	395	99.14	99.15	99.10	99.11	15.8
2016-01-07 09:48:55.398147	1	99.14	100	99.16	99.17	99.12	99.14	3.0
2016-01-07 09:48:55.398225	3	99.14	100	99.16	99.17	99.12	99.14	3.0
2016-01-07 09:48:55.398532	2	99.15	100	99.16	99.17	99.14	99.15	2.0
2016-01-07 09:48:55.398560	5	99.14	100	99.16	99.17	99.14	99.14	3.0

TABLE VIII. These trades, a subset of the trades noted above in Table VII, resulted in realized opportunity cost. Positive ROC here means that these trades favored the SIP data feeds.

timestamp	Exchange number	price	shares	direct_bid	direct_ask	sip_bid	sip_ask	roc
2016-01-07 09:48:55.398272	5	99.12	100	99.16	99.17	99.12	99.14	-4
2016-01-07 09:48:55.398386	5	99.13	100	99.16	99.17	99.13	99.15	-3
2016-01-07 09:48:55.398444	5	99.14	100	99.16	99.17	99.14	99.15	-2
2016-01-07 09:48:55.398532	5	99.14	50	99.16	99.17	99.14	99.15	-1
2016-01-07 09:48:55.398560	5	99.14	100	99.16	99.17	99.14	99.14	-2

TABLE IX. These trades, a subset of the trades noted above in Table VII, resulted in realized opportunity cost. Negative ROC here means that these trades favored the direct data feeds.