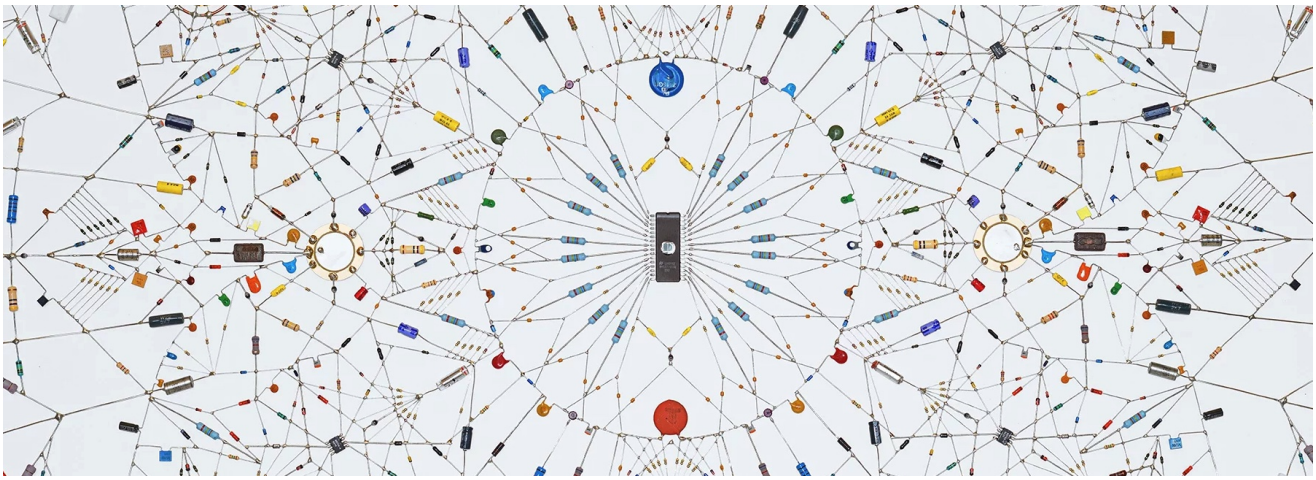

Market Structure Primer

Get Started!

The purpose of this document is to provide an overview of the US equity trading landscape. It started as an onboarding document to help new hires who aren't as well versed in market structure. It can be read start to finish in order, or you can jump around. The visual Life Cycle of an Order also provides a handy view of the various events and entities involved in the process of a trade - from conception to completion. For those with little to no prior knowledge, the three core sections on market participants, mechanisms of communication between participants, and a birds-eye view of overall market activity should provide a working understanding of trading in the US equities market as it is today. Over time, we will be adding material that delves more deeply into further topics, especially into market stucture history and the conflicts of interest that evolve along with market structure.

Mechanisms of Communication between Participants

The When and the How

Now that we have laid out the different types of participants in the US equities market and what their goals might be, we'll move on to discussing the logistics of communications between them. Any full description of communication and interaction in a complex ecosystem should start with the core, established rules that participants all operate under and assume. These do not need to be actively communicated, but rather form the shared context underlying communications. In US equities trading, there are core rules concerning time and price. After discussing the basics of these, we'll go on to the discuss the logistics and content of active communications between the various participants.

The Trading Day

Trading in the US equities market is confined to certain hours of the day. The primary concept of the "trading day" refers to the hours from 9:30 am to 4:00 pm, eastern standard time, Monday through Friday, except for certain declared market holidays. At the start of each trading day, there is a single auction in each security that sets the official opening price for the day. The opening auction is facilitated by a single exchange, which is called the "listing exchange" for that security. The listing exchange collects orders from would-be buyers and sellers, and matches many of them at a single opening price. This typcially happens precisely at 9:30 am. From then on throughout the rest of the day until the end, the security can trade at any venue (not just the listing venue). Similarly, leading up to the 4:00 pm the listing exchange will collect orders from would-be buyers and sellers for a closing auction, which will set the official closing price for that security for the day. The listing exchange for a security has a few other duties as well, such as halting trading in that security when certain conditions arise, and monitoring the security for compliance with its listing standards.

Many trading venues offer "pre-market" and "post-market" hours, allowing participants to trade securities ahead of the opening auction or after the closing auction. These are still typically limited in scope (e.g. a pre-market session might open at 8:30 am eastern). One main difference between these supplemental trading sessions and the official "trading day" is that some regulatory constraints do not apply to pre- and post-market trading.

Pricing Increments

Trading in the US equities market is also confined to a discrete standard of pricing. Would-be buyers and sellers of securities on exchanges must advertise prices over $1 in whole pennies, and trades must be executed at whole penny increments, with the exception of trades at the "midpoint" of advertised buy and sell prices, which sometimes lands on a half-penny increment. The rules are a bit looser for dark pools and internalizers, or for trades that are consummated by participants themselves outside of venues, but there are still constraints of this spirit.

The main goal of imposing a coarse discretization of price is to avoid participants snatching trades out of each other's hands by offering essentially meaningless improvement on price. For example, a would-be buyer who is willing to pay $10 a share for a stock being out-bid by someone willing to pay $10.0000001 a share for a stock is not very civilized. This sort of thing would undermine confidence in the health of the market system, and incentivize some rather silly behavior. Though this motivation of discretization is relatively clean and clear, there is a strange quirk of the way the rules for exchanges are currently written and enforced: participants often *trade* at half-penny increments on exchanges as mid-point trades, but can only *express interst to trade* at full-penny increments.

Communication between Investors and Brokers

Once an investor has decided what they'd like to buy or sell, how much, and some goals in terms of price and/or timing, they need a way to communicate this to a broker who will manage the process of formulating individual orders, routing those orders to various trading venues, and keeping track of the trades that are produced. There is a lot of variety in exactly how granular the investor's instructions are and how they are communicated. Often the communication is accomplished or assisted through an OMS and/or EMS:

EMS/OMS Providers

The current standard buy-side trading workflow incorporates certain core pieces of technology, including an order management system (OMS) and an execution management system (EMS).

An OMS tracks the life cycle and status of orders across the firm, often across multiple asset classes. The "life cycle" of an order refers to all of the relevant events that can occur over the "lifetime" of an order: events like the order being placed at a venue, partially filled, modified, routed away to another venue, fully filled, or canceled, etc.

An EMS is a tool that focuses on individual order execution and allows more fine-grained control, such as slicing a "parent" order into "child" pieces to be sent to a specific counterparty or venue. The EMS also generally has more context on what's happening in the market in real-time to allow the trader to make informed decisions.

A typical work flow is: a portfolio manager at an institutional investor decides to put on a position in US equities, creates an order in the OMS and assigns it to a trader. The trader then stages the order in their EMS which allows them to send pieces (child orders) externally to their brokers. As these child orders get executed at various trading venues, those fills (aka trades) flow back to the EMS, and subsequently to the OMS. The OMS will likely perform or

enable various compliance and risk functions, although in some cases, the EMS might have some built-in risk check logic as well.

These two sets of functionalities are quite related and overlapping, and some vendors try to combine them into a single offering. Overall, the general quality of these technology products is unimpressive, and like many other services in this industry, their pricing models and specific offerings are extremely opaque.

It's an extremely sticky business as the process to transition from one provider to another without losing any custom functionality is a painful one.

Further Reading: New Plateaus for OMS EMS Integration, EMS Consolidation, Liquidnet Trading Solutions

High touch vs. low touch

Within a broker, there are different models for how investor orders are handled. One main axis of differentiation is "high touch" versus "low touch."

In this context, high touch refers to execution services managed by a human, whereas low touch refers to automated execution services, such as trading algorithms. High touch equities teams are sometimes called "cash trading" and low touch "electronic trading." High touch desks often provide additional services such as shopping orders to other clients to directly match a large amount (especially large trades are often called "blocks"), committing capital to take down some or all of the trade, and providing market color and commentary.

Roles on a high touch desk

A high touch desk will typically have employees organized into the following roles:

1. **Trader** - a person who directly manages the execution of orders.
2. **Sales-trader** - a hybrid trader role and sales role that performs duties like shopping flow.
3. **Sales** - account management and business development.
4. **Desk analyst** - a person who conducts independent research and provides more in-depth color than a salesperson or sales trader.

Roles on a low touch desk

A low touch desk will typically have employees organized into the following roles:

1. **Sales-trader/sales** - same as the analogous roles above.
2. **Algo developer/product management/client implementation** - there are several front-office technology roles that work together to build the actual trading platform, manage customizations, and connect to customers and the street.
3. **Operations/support** - the low touch desk's technology is often the backbone of the high touch desk, and when software/hardware/networking issues come up, people in these roles lead the effort to resolve them.

Algo Customizations

One of the core features of most institutional trading platforms is the ability for individual clients to customize the behavior of their chosen trading algorithms. For example, a buy-side trader who is sending orders to be executed by a sell-side broker's electronic trading algorithm might have the option to specify that "the algo" (as it is often lovingly called) should exclude a subset of dark pools, and/or maintain a minimum percentage of volume, and/or always utilize a given minimum execution quantity. These customizations are generally configured or hard coded behind the scenes in the broker's system.

Transaction Cost Analysis (TCA)

Buy-side investors may reasonably want to track and compare performance of brokers across many orders. They might use such data to inform decisions about what kinds of orders to give to which brokers, etc. Getting the relevant data in an actionable form can be challenging, as brokers are sometimes resistant to providing granular detail about their operations to the buy-side clients, and it is not an easy task to isolate the effects of broker decisions from the cummulative biases of data samples and broader market activity. Brokers providing their own analysis of their own trading is clearly problematic, as they are naturally incentivized to choose metrics and data sets that make them look good.

This particular category of communication from brokers back to investors is not very standardized, and the results are often not very satisfying. Getting enough relevant data into

a usable format alongside sufficiently granular market data in order to draw meaningful conclusions remains a non-trivial task. This is especially true for relatively smaller investors who have less resources and leverage to devote to this. As a result, several third-party providers have arisen to provide transaction cost analysis and help buy-side traders evaluate sell-side broker performance. For more discussion of transaction cost analysis and what it entails, see our related blog post.

Communication between Brokers/Proprietary Traders and Trading Venues

There is a lot of information that flows back and forth between the trading venues and their members (the brokers and proprietary traders who are trading on their platform). There is the flow of orders from members into the venue, but there is also a lot of information that must travel the opposite way. The members want to receive timely information from the venue about the status of their own orders and trades. Members also want timely information on all quotes and trades happening on the venue, so they can base their own decisions on an up-to-date view of the market.

At a technical level, information passing from one party to another involves several different layers of physical and logical infrastructure. There is the physical equipment that carries data packets between machines, a protocol specification for the format and meaning of each individual packet, and a high level specification of what data will be communicated, when it will be communicated, and how it is organized. Below we'll give a brief overview of these components for the communications between trading venues and their members. It is also worth noting that these components in aggregate account for the latency of information between brokers and trading venues and between proprietary traders and trading venues. Though these brokers and proprietary traders often use similar means of communicating with venues, their actual needs in terms of latency might vary considerably. We discuss this further in our blog.

Microwave/laser networks

The fastest way to send packets of information across miles of geographic distance is via microwave towers. As a result, market participants seeking the lowest latency in their connections to and from exchanges use microwave as their primary communication medium.

However, microwave can be unreliable (particularly in bad weather), so these participants typically also have fiber optic connections to fall back on.

Extranet providers/dark fiber

Participants who are slightly less latency senstive might use fiber optic connections as their primary way of connecting to trading venues. There are many "extranet" providers who are in the business of leasing their extablished fiber optic connections to market participants.

Data centers

Whether they are traveling via microwave or fiber, data packets orginate and terminate at machines that are housed in data centers. The time it takes a data packet to be formulated at its origin, travel from its origin to its destination, and be processed at its destination is referred to as "latency," and it depends upon the medium of transport (microwave vs. fiber), the distance traversed, any intermediary routing technology (switches etc.), congestion, and the hardware and software that handles the low-level details of packet processing.

The spaces in data centers nearest to where the machines belonging to a NYSE, Nasdaq, or CBOE family stock exchange reside are controlled by those same companies, and they allow market participants to purchase "colocation" - the right to place their machines very close to the exchange's machines. This gives participants an (expensive) option to minimize the distance that data packets must travel between their machines and the exchange's machines, thereby shortening the latency.

FIX

FIX stands for the Financial Information Exchange protocol. It was invented in 1992 as a more reliable and unambiguous replacement for communication between investors and brokers that previously had taken place over the phone. FIX defines a format for each data packet that allows the recipient to interpret its contents, and is tailored to the kind of information that is relevant to financial transactions. At this point, FIX is a common organization of communication between investors and brokers as well as between stock exchanges and their members. Despite its ubiquity and sufficient expressivity, however, FIX

is not a universal choice among financial participants today because it is not particularly optimized for minimal latency.

Proprietary protocols

Most stock exchanges offer their members multiple choices of protocols that control packet formatting. FIX is one choice, but typically the choice that is closest to the native implementation of the exchange software will produce the shortest processing times at the exchange, and hence the lowest latencies. Each exchange family offers the choice of access in a proprietary format that is designed to be highly compatiable with its internal implementation, and it is not surprising that use of these proprietary protocols tends to minimize the packet processing times, and hence the overall latency. As a side effect, this means exchange members have to customize their communication software to the various exchange families, and typically build translation modules to go between the various proprietary formats and FIX or any other formats they use internally.

Orders and Market Data Content

Now that we've touched upon the infrastructure that physically transports packets, and the protocols that are used to format and process them, we'll zoom out a level and talk about what pieces of information go into data packets and how these pieces of information are commonly structured.

Order Types

Brokers communicate their desires to buy/sell stocks to a venue through a menu of order types and associated parameters that the venue defines. A would-be buyer or seller of stock is typically torn between two goals: getting a favorable price, and getting the trade done quickly. Different order types allow participants to express different trade-offs and constraints relating to these competing goals.

There are a lot of important differences in the menus provided by different venues, but most are variations on a few common themes:

Market Orders

A market order is used to communicate an immediate desire to buy or sell, irrespective of price. When a market order to buy enters a venue, it can be immediately matched with any open orders to sell. Conversely, when a market order to sell enters a venue, it can be immediately matched with any open orders to buy. This order type represents an extremal point in the trade-off between price and time: full priority is given to executing immediately, completely insenstive to price.

Limit Orders

Limit orders are used to express a constraint on price. The presumed goal is to execute a trade as soon as possible, within the band of acceptable prices. A limit order to buy, for example, will specify a ceiling for price, and will trade at the first opportunity to buy at or below its limit price. A limit order to sell will specify a floor for price, and will trade at the first opportunity to sell at or above its limit price.

The combination of limit orders and market orders is enough to build up some illustrative examples of trading dynamics. A market maker might have active limit orders to buy 200 shares of a stock for $10.00 a share and to sell 200 shares of the same stock for $10.02 a share. While waiting and available to be filled, these limit orders are referred to as "resting" on the order book kept by the trading venue. A buyer who wants to immediately buy 200 shares might enter a market order to buy, which will be matched against the limit order to sell at a price of $10.02.

Midpoint Orders

A midpoint order is a way of delegating the determination of price to the broader market. Instead of declaring a specific limit for price, a midpoint order will execute as soon as possible at a price that is equal to or more favorable than the midpoint of the current spread. In other words, it behaves like a limit order, but where the limit is a dynamically adjusting price calculated by taking the midpoint between the highest open buy limit orders and the lowest open sell limit orders. This calculation is typically over the price limits being advertised across all exchanges (which does involve some latency as the relevant information travels from the origin exchange to the venue processing the midpoint order).

If we return to our toy example of a single market maker offering to buy 200 shares for $10.00 a share or to sell 200 shares for $10.02 dollars a share, and we assume there are no better prices available across all exchanges, the midpoint price is therefore $10.01. If a midpoint order to buy enters a venue under these circumstances, it will be willing to buy at $10.01 or lower, and therefore will not be matched against the limit order to sell at $10.02. It will wait to interact with a seller willing to sell at $10.01 or lower (which might be a market order to sell, a midpoint order to sell, etc.)

Order Attributes

Many order types allow those submitting orders to specify additional parameters that control more of the fine-grained mechanics of how the orders behave. Limit orders can be "lit" (aka "displayed"), meaning that their sizes and limit prices are visible to other market participants who might want to take the other side of the trade, or "dark," which means they are not visible. Market makers typically use lit orders to advertise and generate trades.

Dark orders can often specify a minimum quantity, meaning that they will not execute for less than that specified number of shares. For example, an order for 1000 shares that has a minimum quantity of 500 shares can not be matched in a trade for 100 shares. It will wait until at least 500 shares can be traded at once.

Some orders can also have pushy designations like "immediate or cancel" (IOC), meaning that they must be executed immediately or not at all, or "fill or kill" (FOK), meaning that they must be executed in full or not all. There is a virtual alphabet soup of overall options, so we will not try to give a comprehensive accounting here. We will discuss more order types later to the extent that they are relevant to higher level issues of market interactions.

Now that we've given a brief overview of how brokers communicate their desires to venues, we'll move to discussing how venues communicate information about trades and prices to their members or other financial participants:

Tapes

Publicly traded stocks in the US are organized into 3 groups called Tapes. Tape A has NYSE-listed securities, Tape C has Nasdaq-listed securities, and Tape B has everything else. As discussed above, a security is "NYSE-listed" if NYSE is the venue responsible for

running its official opening and closing auctions (among other duties), but it can trade on all venues at all times.

For a particular stock, these are the types of information that participants are most interested in:

Top of book and depth of book data

"Top of book" and "depth of book" both refer to data about the best prices and associated quantities that would-be buyers and would-be sellers are currently advertising on a particular venue. The prices and quantities that buyers are advertising are called "bids," and the prices and quantities that sellers are advertising are called "asks" or "offers." For example, a bid might be "would buy 100 shares of security X for $40 a share." This is of course a wordier version than the shorthand formats that participants use, but this is what they mean. Similarly, a sample offer means: "would sell 200 shares of security X for $40.02 a share."

At a particular trading venue, if a buyer was advertising a price that was equal to or higher than a price being advertised by a seller, the buyer and the seller could just trade with each other (at least for the minimum of the two share quantities). Since the venue will catch this and match the trade, the highest bid price at a given venue should always be lower than the lowest offer price. The difference between these is called the "spread" - we've mentioned this before as a driver for compensation to market makers who are constantly willing to buy or sell. The collection of bids at the current highest bid price and the collection of offers at the current lowest offer price are referred to as the "top of book." Data feeds that are categorized as "top of book" include only information about these advertised orders, and not any bids at lower prices or offers at higher prices. Naturally, the current highest bid price and current lowest offer price are moving targets as trades happen and advertised bids and offers are canceled or adjusted. Top of book data feeds are typically event-driven, meaning that as soon as the top of book information changes, the new information is immediately disseminated. A top of book data feed might aggregate the sizes of bids or offers at a particular price, or it might give granular information about each bid/offer.

"Depth of book" refers to information about bids and offers that goes beyond the highest priced bids and the lowest priced offers. A typical depth of book data feed will include information about all bids and offers, but there are some variations on this. Some depth of

book data feeds aggregate sizes of the bids or offers at a particular price, some given granular information on each bid/offer, and some place a limit on the number of distinct prices (sometimes called "levels") that are included. For instance, the CBOE exchange family offers some data feeds that include information only about the 5 highest price levels for bids and the 5 lowest price levels for offers.

As an interesting side note, the highest bid price and the lowest offer price *across venues* can be equal, a state which is referred to as a "locked market." The highest bid price can even be strictly higher than the lowest offer price when we look across venus, which is referred to as a "crossed market." These situations do arise fairly frequently, but are typically quickly resolved, as someone recognizes an opportunity. You might expect that if a seller on one venue and a buyer on another have compatible desires, then both would likely consume the relevant data, recognize the compatability, and one of them would send an order to the other venue to trade. But even if both participants are tracking the bids and offers on the other venue and see each other, billing practices at venues can complicate this story. Many of the exchanges are "maker-taker" venues, meaning that they charge a fee to the entity who submits an order that executes against an advertised bid/offer (aka the "taker") and they give a rebate to the entity whose advertisement led to the trade (aka the "maker"). If a buyer is advertising a price P on a maker-taker venue and a seller is advertising that same price P on a different maker-taker venue, both entities might be banking on receiving the rebate instead of the fee to view that price P as worthwhile. This sometimes creates an amusing standoff until someone changes their view of the price they are willing to pay, or another participant intercedes. This scenario couldn't explain a crossed market though, since the range of rebates/fees is capped to be smaller than the 1 penny price increment allowed for bids and offers.

Trades

In addition to quotations of what would-be buyers and sellers are advertising, market participants also want data about trades as they occur. Once a trade is executed at a venue, the venue sends confirmations to the involved parties, as well as disseminating some information to the larger market ecosystem about the trade. The broader population can learn what security was traded, how many shares, at what price, and at what time. The identities of the buyer and seller are not revealed. In fact, even the trading parties themselves will

typically not know each other's identities. If a trade happens on an exchange, the identity of the exchange will be publicly visible as well. If it happens on a dark pool, the trade will be identifiable as having occurred off exchange, but the identity of the specific pool will not be publicly visible.

Auction Data and Alerts

Leading up to an auction, an exchange may choose to make some data available concerning the possible price or the imbalance between orders to buy and orders to sell. This kind of data may be used by participants to adjust their orders leading up to the auction, and may motivate new buyers and/or sellers to participate.

There are also alerts dissimentated through data feeds that participants might need or want to consume. For example, alerts may be sent out when trading is halted in a stock.

Market Data Dissemination and Trade Reporting

Now that we've talked about what kind of market data participants typically seek and receive, we'll discuss the data feeds that disseminate these types of data and the entities that control them.

SIPs

The SEC has mandated that some market data be made publicly available via Securities Information Processors (SIPs). Currently there are two SIPs, which are operated by NYSE and Nasdaq, respectively. The NYSE SIP (CTA) publishes trades and quotes for Tape A and Tape B securities, and the Nasdaq SIP (UTP) publishes trades and quotes for Tape C securities. In terms of quotes, the SIPs provide only top-of-book information from each exchange. A change in the top-of-book at a given exchange is first communicated by the exchange to the SIP and then from there disseminated to recipients of the SIP data feeds. This two hop approach is unsuprisingly slower than sending the new data directly from each exchange to the ultimate recipient.

Even though these are "public" data feeds that lack depth-of-book or auction imbalance information, they are still extremely expensive to connect to, consume, and redistribute,

especially for real-time data.

Proprietary data feeds

Exchanges also offer multiple kinds and tiers of proprietary data feeds. These products run the gamut from trades and top-of-book quotations to full depth-of-book at the particular exchange. Because they can be consumed directly from the exchange without the data first traveling to a central processor as is the case for the SIPs, these proprietary feeds can deliver data at the lowest latencies (especially to recipients who are co-locating their equipment in data center space provided by the exchange). Some of the data products are free (e.g. all of the data feeds from IEX or NYSE National), some of them are a bit cheaper or comparable in pricing to the SIPs (e.g. Nasdaq Basic), but the full depth-of-book feeds at the main exchanges with the highest market share are an order of magnitude more expensive.

For more details on market data pricing for the real-time data products offered by exchanges, see our market data series of blog posts (part 1, part 2, part 3) and our market data pricing visualization tool.

TRF

The FINRA Trade Reporting Facilities are where information about trades that happen off of exchanges is reported and then relayed to the appropriate SIP for public dissemination. There are currently two TRFs, also operated by NYSE and Nasdaq, but unlike the SIPs the TRFs are not partitioned by Tape. Instead, each dark pool may strike an agreement with the TRF of its choosing to report its trades. Such trades are ultimately visible to others via the SIPs, though with potentially greater latency than exchange trades and with the precise venue not specified. In other words, the TRFs allow other market participants to learn the price, size, and time of off-exchange trades, but only the binary attribute of "off-exchange" is visible/inferrable, not further granularity on which off-exchange venue produced the trade.

Communications among Brokers/Proprietary Traders

In the realm of retail trading, some communication happens directly between brokers and proprietary traders. In fact, most retail trading occurs through direct relationships between retail brokers and proprietary trading firms. The process typically works like this: a retail trader submits an order to a retail broker (e.g. Robinhood, ETrade, etc.) The retail broker sends it to a proprietary trading firm that is acting as a retail wholesaler. The proprietary trading firm decides whether they want to execute the trade with themselves as the counterparty or route it elsewhere.

This is an attractive arrangement for the retail broker, who typically gets payment for order flow (PFOF) and/or price improvement from the retail wholesalers. In this context, "price improvement" means that the end customer (the retail trader) will get a slightly more favorable trade price than what market makers are currently advertising on the exchanges. The retail wholesaler is willing to do all of this (PFOF and/or price improvement) because retail traders typically have small orders (hence limited risk) and trade without significant short-term alpha. (This is a polite way of saying that on a short-term time scale, retail traders don't tend to buy low and sell high.) For this reason, market makers can make an excellent profit off of retail traders in aggregate, so they are willing to grant better prices and pay for order flow when they know they are trading against a less sophisticated counterparty.

Regulatory Constraints and Surveillance

Here we will discuss the main regulations that serve as constraints on market activity, and the communications and data that regulators rely upon to enable their enforcement. These regulatory constraints should be thought of as part of the assumed common knowledge underlying market interactions. We did not discuss all of them above when we discussed the logistics of the trading day and discrete price increments only because some of them require a pretty deep in the weeds perspective to describe. Only now after we have discussed more of the details of communications and interactions between various participants do we have the semantic tools to summarize some of these regulations and their current effects on the market ecosystem.

Major Regulatory Constraints

In the wake of the stock market crash of 1929, new federal laws were passed with the intent of protecting investors and the integrity of markets. Prior to this, stock sales were regulated only under state laws. The goal of the Securities Act of 1933 was to set consistent standards for public companies to protect investors from fraud. The Securities Exchange Act of 1934 created the SEC (the Securities and Exchange Commission) as the regulator of the public markets and established the concept of self-regulation, making exchanges share responsibility for policing their own markets.

Since then, many more regulations have been put in place with the intent of constraining potentially harmful trading activities, incentivizing healthy trading, and ensuring reliability and availability of market infrastructure. We'll address much of this in more detail in a market history section later, so for now we will just highlight the regulations that have a major impact on current trading mechanics.

Securities Act Amendments

In 1975, the securities act amendments were passed, creating a more organized and centralized system to connect exchanges together. The SIPs were created to be a central source of the most relevant information across disparate exchanges.

Reg NMS

In 2005, Regulation National Market System (aka Reg NMS) was introduced which established the Order Protection Rule, among other regulatory changes. The order protection rule mandates that trading venues cannot match trades at prices that are inferior to the best prices currently available on other venues. For example, if a market order to buy arrives at a venue V where the lowest limit a seller has set is $11.00 per share, but some other venue W has a displayed sell order available at $10.50 per share, the buyer and seller cannot be matched at venue V for $11.00 per share. The buy order must be routed to venue W instead to be matched with the seller there at the lower advertised price.

Obviously this is intended to protect the buy order from being executed at a sub-optimal price, but there are a lot of devils hidden in the details. For instance, what does "currently available on other venues" technically mean? For venue V to fulfill its obligations under Reg NMS, it must now track the best available prices on venue W (and everywhere else). Since information takes time to travel from venue W to venue V, the view that venue V has of the "best prices currently available on venue W" is necessarily delayed. The amount of this delay is a dynamic function of all of the communication infrastructure between W and V. V is this regard is similar to any broker connecting to venue W, and it must choose how much to invest in reducing latency.

Reg SHO

Regulation SHO concerns the practice of short-selling, which is selling a stock without owning it. Reg SHO concerns "locate" and "close out" requirements which try to assure that stocks which are sold by short sellers can be reliably borrowed and delivered on time. Reg SHO also puts limits on short selling in the midst of a substantial intraday price decrease. This is intended to address concerns that short-selling can be used manipulatively to artifically depress the price of a stock.

Reg SCI

Regulation Systems Compliance and Integrity (or Reg SCI as it is commonly called) was established in 2014 with the goal of reducing the likelihood of significant to catastrophic technology failures in the financial markets. This regulation created more formal requirements for the operations and testing of core components of trading technology and applies broadly to venues and clearing agents.

LULD bands and circuit breakers

Limit Up Limit Down bands (LULD) prescribe a certain price window that a stock can trade in, as a function of the stock's prices over a preceding five minute window. The bands ensure that a price can't increase or decrease too suddenly by too large of a percentage. Trades outside the band are prohibited, and if prices do not revert back into the band within 15 seconds, a 5-minute trading pause in the stock is triggered.

While LULD bands function individually for each stock, there are also market-wide circuit breakers, which are triggered when the S&P Index decreases by certain percentages, as compared to the previous day's closing price. There are three levels at which circuit breakers can be triggered: Level 1 is triggered by a 7% relative decrease, and will halt trading across the market for 15 minutes (though only if it is before 3:25 pm). Level 2 is triggered by a 13% relative decrease and will also halt trading broadly for 15 minutes (again only before 3:25 pm). Level 3 is triggered by a 20% relative decrease, and halts trading across the market for the remainder of the day, no matter the time.

Logistics of Regulatory Surveillance

To monitor for compliance with regulations, to investigate potential cases of market manipulation, and to more generally keep a finger on the pulse of the markets, regulators needs access to granular data about trading mechanics and behaviors. Here we discuss some of the primary data sources that FINRA and the SEC collect from market participants to enable regulatory surveillance.

605/606 reports

SEC Rules 605 and 606 require venues and brokers to make periodic public disclosures of basic aggregate information about executions. Rule 605 mandates that venues make

monthly reports of basic execution quality metrics on publicly accessible websites. Rule 606 mandates that brokers disclose information about which venues they route orders to and in what proportions.

OATS

OATS stands for the "Order Audit Trail System" that is used by FINRA. It was established by FINRA rules in 1998 (approved by the SEC). All FINRA member firms are required to collect and submit the data according to the technical specficiations of OATS, and FINRA uses the data for market surveillance. The data spans orders, quotes, and trades for listed stocks as well as over-the-counter securities (securities that are not listed on a stock exchange). In aggregate, this data is intended to allow the re-creation of order lifecycles, so that the chain of events from order entry to order modification or cancellation or execution can be followed.

CAT

In 2012, the SEC called for the creation of a "Consolidated Audit Trail" (CAT) to serve as a key data source for market surveillance and regulatory investigation. This is intended to replace OATS and to enable superior surveillance capabilities. In the longterm, all venues and brokers will report data into the CAT processor in prescribed format. The CAT processor will be responsible for assembling this data into easily query-able form and linking together the different pieces of an order life-cycle, so events like order submission, modification, routing, execution, and cancellation can be stitched together and in some cases attributed to the originating investor. Granular identifiers of venues, brokers, and end customers will be maintained so that ultimately regulators can identify actors in involved in any questionable activity. Regulators such as the SEC and FINRA will be given access to the CAT in prescribed ways to enable market surveillance.

The CAT is intended to represent a more complete and sophisticated tool for regulators (particularly at FINRA and the SEC) to get effective information about behaviors in the marketplace. However, it is a complex project with many different stakeholders who are often at odds about exactly what data should be collected, how quickly reporters should have to adapt their systems to report such data, and who should have access to what. Long delays and disagreements have plagued the CAT since its inception.