HARVARD | BUSINESS | SCHOOL

9-215-036
APRIL 1, 2015

LAUREN COHEN
CHRISTOPHER MALLOY
MATTHEW FOREMAN

Domeyard: Starting a High-Frequency Trading (HFT) Hedge Fund

We aim to capture the fast moving alpha of order book events which typically disappears in less than 15 milliseconds.

-- Luca Lin, Co-Founder of Domeyard

Luca Lin, Christina Qi and Jonathan Wang looked at each other gloomily in their office on the 15th floor of a building overlooking the MIT campus, the Charles River, and downtown Boston. Michael Lewis' new book *Flash Boys* had just become a *New York Times* bestseller and there was a general outcry against high frequency trading (HFT)—the basis of their new hedge fund called Domeyard.

"Do you think there would be regulatory changes that could erase our business model or limit our ability to raise more capital?" asked Christina. "I don't think so. Being new, we may be able to react to the changes faster than entrenched players," answered Luca, "besides, the type of trading we do is not what Lewis complains about in his book. We should stop worrying about this book because we other have pressing business decisions to make, such as finding the top software engineers we need."

Perhaps the more pressing question was whether Domeyard should spend roughly $4 million to buy a seat on the CME. "It ties up lots of precious capital in a relatively static asset, but it gives us much lower trading costs," reasoned Luca. In reality Domeyard would need a large volume of trades per day for this purchase to make sense, but the question of how fast they could ramp up given that they had not even started trading yet was still unanswered.

Jonathan, Luca and Christina were facing a myriad of other important decisions as well, such as which markets to trade on, how to raise capital, and from whom to raise capital. Many of these decisions were standard for startups, but some were unique to hedge funds: among them choosing the right fund structure, as well as how to raise capital for the management company versus raising capital to invest. Also given the recent backlash against HFT, the question of whether they should delay any launch also loomed large.

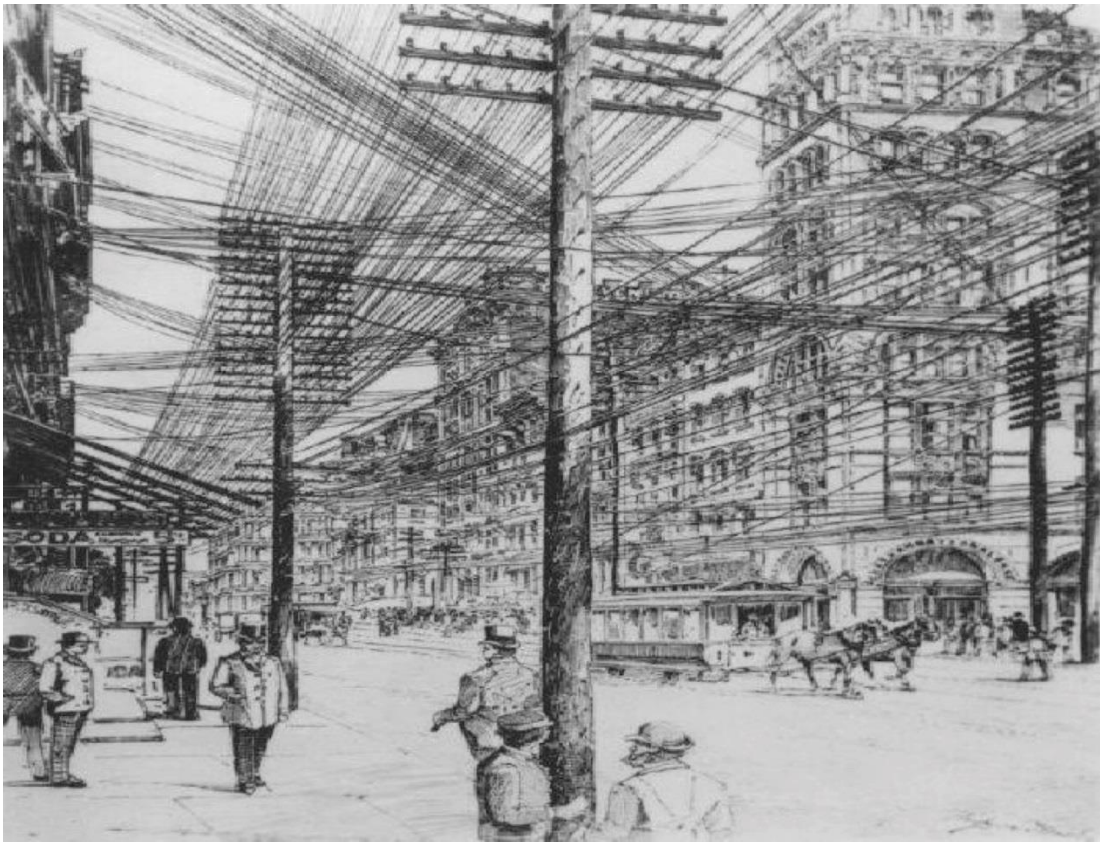
HBS Professors Lauren Cohen and Christopher Malloy and Professor Matthew Foreman (University of California at Irvine) prepared this case. It was reviewed and approved before publication by a company designate. Funding for the development of this case was provided by Harvard Business School and not by the company. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

A Short History of Information Flow

Receiving superior information fast has always been a critical advantage when trading in the stock market. A famous example was Rothschild's use of racing pigeons to bring the news of Napoleon's defeat at Waterloo, and his subsequent purchase of British government bonds. Multiple other anecdotal examples exist, from the development of fast cutters—ostensibly to be the first to deliver harbor pilots to freighters arriving in port but in reality, economically relevant as the first to bring the news of the cargo back to market—to the Pony Express, designed to bring news of the markets in San Francisco back to the East Coast in about 11 days, much faster than it would take for a Clipper to sail around Cape Horn.[1] Indeed the Pony Express was put out of business in less than a year by the advent of the telegraph.

The first practical telephone was invented in 1876, by Alexander Graham Bell. Less than 10 years later the telephone (**Figure 1**) had widespread commercial acceptance driven, in part, by its widespread adoption by the investment community.

Figure 1 Telephone wires in Manhattan in 1890



Source: Henry C. Brown, Book of Old New York (New York City, NY: 1913).

[1] See also Koudijs, Peter, "Those who know most: Insider trading in 18th c. Amsterdam" (Journal of Political Economy, 2014, forthcoming) for evidence of informed trading via information transmitted from England to Amsterdam on board of official mail packet boats.

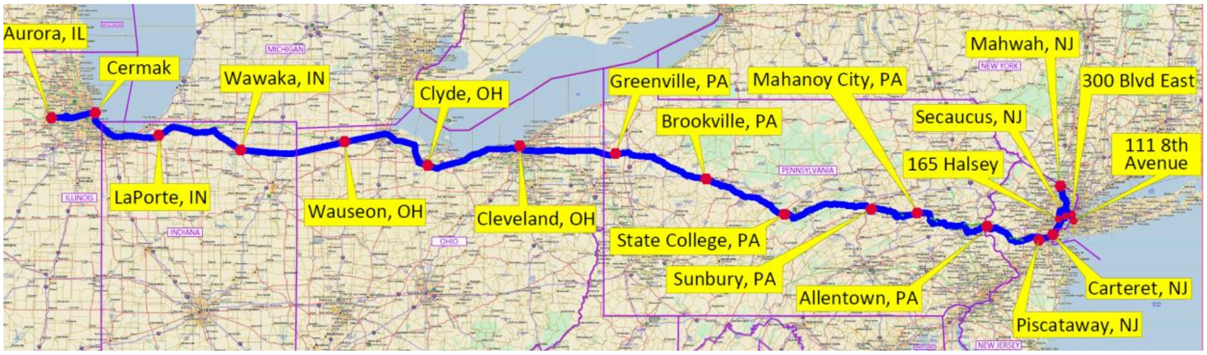
The question of why information was so important in markets, and why it was economically valuable, had a variety of answers. One of the simplest answers was due to so-called "*front-running*" in the stock market. If one had material information affecting the price of a stock (or that allowed one to effectively predict someone else's actions) before others did, then one could simply act first and collect the profits (i.e., "front-run" everyone else's trades). A canonical example would be trading before the release of an acquisition announcement, if one had advance news on the timing and likelihood of such an event.

The Proliferation of Stock Markets, Fiber-Optic Tunnels, and Microwaves

The evolution of information transmission, across the globe and purely in terms of speed, has continued unabated over the past 100 years. Early in the 20th century, there were few stock exchanges and few stocks were listed on more than one exchange. But over the past century there has been a proliferation in the number of exchanges. As of 2014, there were over 40 different exchanges with a great many stocks cross-listed on more than one exchange. These cross-listings allowed the possibility of "inter-venue low latency arbitrage" which referred to attempts to trade the same stock on one exchange before the information was reflected in the price of that same stock on another exchange.

This natural evolution and continuing emphasis on minimizing timing dislocations across markets, even seemingly microscopic ones, led one company called Spread Networks to build a high-speed fiber-optic cable tunnel (see **Figure 2** below) from Chicago to New York along the straightest and shortest route possible (825 fiber miles).

Figure 2 The Spread Networks fiber tunnel from Chicago to New Jersey in 2014



Source: Spread Networks, "Network Map: Chicago to New Jersey," http://spreadnetworks.com/media/11285/1-chicago-newjersey_dec_2013.pdf, accessed March 2015.

Not to be outdone, after the completion of the Spread Networks "tunnel," competitors such as the McCay Brothers began exploiting low-latency microwave technology, which was even faster than fiber-optic cable (but more limited in terms of bandwidth).

How Markets are Made

Markets existed to facilitate the exchange of goods for money or services. In real markets buyers and sellers did not arrive simultaneously. Thus there was a need for intermediaries who were willing to buy or sell inventory on a short-term basis.

Though the financial markets were often held out as the closest approximation to "perfect markets," this problem existed there as well, even for very liquid instruments. Buy-side agents in financial markets expected to be able to buy or sell their positions at the moment of their choice. For this to happen there had to be "*liquidity providers.*"

Traditional Liquidity Provision

The traditional liquidity providers were known as *specialists*.[2] Their responsibilities required them to continuously post both bid and ask prices and be willing to back up these prices by making purchases or sales. Specialists were free to set the bid and ask prices subject to limits on the "spread" set by particular exchanges. If the specialists set the bid-ask levels too far off of the market, they risked losing money from participants who drove prices against them. To mitigate this risk they had to set the prices and spreads with a buffer against the possibility of informed traders moving the market against them.

When a specialist received an order, she immediately matched the order with one on the order book (or available on the floor) if that was possible. For example if she received an order to sell at no less than 60, then the specialist completed the transaction with the highest combination of buy orders above 60. Completing the sale might require matching several buy orders at different prices and might result in only a partial fill. In some exchanges specialists conducted open outcry auctions where floor traders could complete their transactions.

If a buy order came in that was at or above the ask price posted by the specialist, and there was no seller to match it, the specialist was obligated to fill that order with their own inventory. Similarly if there was a sell order at or below the buying price posted by the specialist, the specialist was obligated to make the purchase. In this way specialists were required to hold long or short inventory and carry the risk of concomitant price movements. Conflicts of interest were mitigated by rules requiring that specialists give priority to orders placed by market participants and that required them to execute those orders ahead of their own.

Balancing the risk inherent in buying and selling inventory were the informational advantages that specialists received about order flow and imbalances between buyers/sellers. They were entitled to know the state of the order book (or all known interest in the stocks they represented) and thus whether there were more buyers or sellers at a given moment. Moreover they got a good idea of the proportion of *informed traders* that existed at any given moment. This informational advantage allowed them to manage their inventories and make profits by predicting short-term price changes as well as collecting the bid-ask spread.

The specialists typically had wider bid-ask spreads for less liquid stocks, commensurate with the risk of holding those stocks. They adjusted the bid-ask spreads in a manner that allowed them to keep inventories within acceptable limits. Nonetheless the risks were real: it is estimated that specialists were forced to buy more than $700 million in stocks during the October 1987 market crash.

One study (Madhavan and Sofianos (1998)) suggested that in July 1993, specialists participated in 26.8% of NYSE transactions and almost 53% of transactions of low-volume stocks. Another study (Sofianos and Werner (1997)), using confidential data provided by the NYSE, stated that in January

[2] There were 443 specialists on the NYSE for about 2800 listed stocks in 2014.

and February of 1997 specialists traded 10.8% of the volume of stocks on the NYSE. That number was highly variable, with markets for less liquid stocks being heavily dependent on specialists for trading.

Estimates of the profits made by specialists varied widely. One estimate stated the after-tax profits made by specialists in 2001 were around $414m.

Making Money Without Thinking

Another way that liquidity providers could make money was simply by knowing *who* they were trading with. The rationale was that informed traders, by virtue of greater effort or expertise, had a better than average chance of making money on a stock. Meanwhile, so-called “noise traders” (often viewed to be non-sophisticated investors, such as retail investors), in aggregate were believed to trade in a way that was essentially random. Because informed traders were thought to be making money on average, then the aggregated noise traders had to be losing money on average. Thus, if one knew that he or she was trading with noise traders, one would have a high chance of making money simply by matching all of their trades!

Numerous anecdotal episodes suggested that this information was quite important in practice. For example, Ameritrade’s customers were overwhelmingly retail and thus statistically more likely to be noise traders. And it turned out that the hedge fund Citadel *paid* Ameritrade $.0021 per trade for the right to execute the Ameritrade’s customers’ orders.

Similarly, specialists conducting auctions on the floor had visibility into who was placing large orders—and thus could be reasonably expected to have insight into whether the orders were from noise traders or not.

Aggressors versus Liquidity Providers

The distinction between ordinary market participants and liquidity providers was critical, and required an understanding of the dynamics of the order book (or “book,” for short). A typical order book had several “levels,” reflecting buy and sell limit orders. The best bid and ask prices formed the *inside quote,* or best bid and offer. Orders at a given price level were filled in the temporal order they arrived. (**Exhibit 1** illustrates two examples of an order book).

For example, suppose that the inside quote rested at $40.52 bid and $40.55 ask, and a new market participant arrived. She decided that she was willing to buy at $40.55. She was said to be the *aggressor* and the person selling at $40.55 was the passive party. The passive party was said to be “providing liquidity” to the aggressive party, and the aggressive party was said to be consuming or destroying liquidity. These parties were also sometimes referred to as *makers* and *takers* of liquidity.

Thus, the specialists described above took on this role of liquidity providers. Because markets could not exist without liquidity, and due to the intense competition between exchanges, many exchanges provided “rebates” to liquidity providers. These were fees (or reduced transaction costs, in less extreme cases) that were paid back to the passive parties in transactions.

The Development of Electronic Markets

The arrival of computer networks (much like the arrival of telephones in the late 1800’s) led to fundamental changes in the way that stocks were traded. Electronic communication networks (ECNs) were created and employed automated matching engines—when there were compatible bids and asks, they were optimally matched by deterministic rules. These ECN’s were free from conflict of interest issues and enabled 24-hour trading. Moreover they greatly reduced trading costs, typically charging fees of $0.002 to $0.003 per share. Rules restricting the development of equities markets were relaxed in the late 1990's and the first decade of the 2000's, resulting in a proliferation of exchanges. For equities alone there were over 40, not counting specialized venues such as dark pools (described below).

These markets offered a variety of business arrangements with brokers and traders, such as varied commission structures and rebates for liquidity provision. To prevent brokers from simply trading at the exchange that gave them the best deal, brokers were theoretically obligated to place their orders at the venues giving their clients the best prices. There was a standardized price feed giving the National Best Bid and Offer (NBBO). Markets giving a bid-ask spread that was *crossed* (i.e., where the bid was above the NBBO ask, or where the ask was above the NBBO bid) were prohibited from trading these situations.

High Frequency Trading

A key to profitable trading was the ability to take advantage of relevant information about the value of an asset. The first party able to receive that information and effectively act on it was able to capture the profit implied by it. With the advent of electronic trading and communications, the speed at which information could be assimilated into stock prices increased dramatically.

These technological factors led to the advent of "high frequency trading" (HFT). The speed of modern computers and communication equipment began running into physical limits. For example the speed of light through fiber optic cables led to lag times that were "excessive" beyond a few meters. As a result, many HFT firms insisted on placing the computers that actually executed their algorithms in very close proximity to the matching engines run by the exchanges, a practice known as *co-location*. In many locations, such as the CME, the fiber optic cables connecting the traders "boxes" with the gateways were precisely measured to be of exactly equal length. The CME even had independent "auditors" to ensure these equal lengths.

To give some perspective on the speed involved, HFT firms spoke in terms of "milliseconds," "microseconds," and "nanoseconds." A millisecond (ms or msec) was one thousandth of a second and was commonly used in measuring the time to read to or write from a hard disk or a CD-ROM player. A microsecond was one millionth (10^{-6}) of a second. And a nanosecond was one billionth (10^{-9}) of a second and was a common measurement of read or write access time to random access memory (RAM). Meanwhile, human reaction time to stimulus was on the order of 200 milliseconds (i.e., the time between when humans could see a color flash on their screen and then click with their mouse).[3]

The HFT Landscape

As a consequence of their tremendous speed, HFT firms quickly began to dominate the trading volume on the major stock exchanges. Estimates varied as to what percentage of trading that HFT made up, but it was generally thought to be in the ballpark of 50-60% of stock trading by dollar volume.

[3] See, for example, Human Benchmark, Reaction Time Test, http://www.humanbenchmark.com/tests/reactiontime.

High frequency traders proved to be very profitable. Typical annualized Sharpe ratios ranged from 5 to around 25, depending on the nature of the strategy. (For comparison, the historical Sharpe ratio of the S&P 500 was between 0.5 and 0.6).

As an example of the high returns to high frequency trading, a company called *Virtu Financial* made preliminary attempts to go public in March 2014. In the filing papers it reported that in the previous 5 years it had only one day when it lost money (out of 1278 trading days). Virtu postponed its IPO in the face of controversy generated by Michael Lewis' *Flash Boys* book.

Many high frequency strategies, by their nature, had limited capacity. However, because of the ability to achieve very high returns with low risk, they did not require large amounts of initial capital to be extremely profitable.

HFT firms did face significant technical challenges relating both to speed and the size of the data sets they sometimes employed. For example, "tick data" (which contained data on each order book event) for a day's trading on one instrument could contain around 60 million lines, each line reflecting the state of the order book. NASDAQ alone had about 3100 equities. Supposing that an HFT firm was trading 10,000 instruments in various locations, that meant a day's worth of trading data contained roughly 6 x 10^11 lines. Files of this size were much too big to be handled in software programs like Excel, and the file from a single instrument could take hours to load in other programs like MATLAB, even when running on a reasonably sized computer. And importantly, this data had to be analyzed and acted on in microseconds as it arrived in real-time, not merely loaded for analysis at the end of the day.

And of course after overcoming these data processing problems, a new HFT firm also had to find a profitable trading strategy. Many HFT strategies had limited capacity, meaning that not many firms could use the same idea, and moreover many of the strategies were "winner take all" in nature: the fastest firm got the entire prize.

Having said this, high frequency trading was gradually being commoditized to some extent as hardware costs shrunk and commercial software solutions strengthened. Firms such as Redline Trading Solutions, QuantHouse, Orc Group and others provided high-speed data feeds, APIs (application programming interfaces), and low latency gateways to trading firms not willing or able to develop their own proprietary systems. These commercial solutions allowed traders to focus on developing their core trading strategies without being distracted by the "plumbing" of writing and maintaining numerous feed handlers for global exchanges or certifying in-house order execution gateways with all these exchanges.

A Taxonomy of High Frequency Trading Strategies

In the world of high frequency trading, an important distinction existed between strategies that were "*low latency*" versus truly "*high frequency.*" Low latency referred to the ability to get information fast and act on it immediately. High frequency strategies were strategies that, by design, had very high turnover and very frequent trading, on the order of tens of thousands a trades per day on a single market.

The distinction was not completely clear cut—one had to be very fast to use high turnover strategies, and companies that had low latency capabilities clearly could trade frequently. Roughly speaking one could characterize the difference as follows: low latency strategies involved sprints to "pick up dollar bills lying on the ground," while high frequency strategies were attempts to be the "house" in a casino where the odds were slightly tilted in one's favor.

Strategies of the latter ilk might have a slim advantage in each play. However, the *Central Limit Theorem* stated that for **independent** bets, the expected return increased with **n,** whereas the standard deviation of the return increased with $\sqrt{n}$. Thus, the Sharpe ratio increased with $\sqrt{n}$**.** (However even in HFT, finding truly independent bets was very difficult.)

As a rule, in designing high frequency trading algorithms, there was a tradeoff between being *fast* and being *smart.* The additional processing time required to do statistical analysis could slow execution down sufficiently, such that valuable opportunities were missed. On the other hand, the availability of vast quantities of high frequency data could provide meaningful statistics that were impossible to estimate accurately using only daily or hourly data.

Low Latency Strategies

A fundamental principle in finance was the "*Law of one price*," which stated that two assets with the same cash flows and risk must sell for the same price. This law was enforced by market participants who corrected instances of mispricing by buying or selling the appropriate instruments. In idealized "perfect markets" this process happened "instantly." But in reality, time lags and frictions prevented the exact pricing of equivalent instruments. However, high frequency trading infrastructure allowed traders to recognize and correct arbitrage opportunities at nearly the speed of light.

Thus high frequency trading strategies included all of the conventional "arbitrage" strategies. Examples included exploiting the relationships between options and their delta hedges, between futures and spot prices, between ETF's and their underlying components, and many others. In many of these cases, the first entity to recognize and act on the arbitrage opportunity took the whole prize.

Exploiting Technology Advantages

The US NBBO described earlier was provided by the *Securities Information Processor.* Historically run by NASDAQ, it aggregated every exchange's best bid and offer in real time and provided the information to brokers and traders, who used it for quoting and trading. The intent was to provide uniform pricing at different venues. Regulations against locking and crossing the market were based on the SIP.

However the SIP was also "slow" by HFT standards. Proprietary data feeds were much faster, allowing those with access to more current price information that could be used for trades. Of course, proprietary feeds also had tradeoffs between speed and accuracy; the fastest were also noisier and error prone. At the end of the day, the ability to get accurate price information quickly was the basis for many HFT strategies.

Another opportunity for technology arbitrage came with news events, both scheduled and unscheduled. Scheduled news events included the release of financial data, for example from the Federal Reserve Bank or regarding unemployment data. This information was held by news agencies such as Reuters and released exactly at a specified time.[4] Some HFT hedge funds used sophisticated

[4] There are both known and unknown exceptions to this practice. Some agencies accepted fees for receiving data a few milliseconds early, but this practice has been curtailed. In another case there was an unexplained burst of trading within 0.2 seconds preceding the release of an US unemployment report.

techniques in Natural Language Processing (NLP) to analyze the contents of the releases in milliseconds and place trades based on this analysis.

In addition, during short time intervals after news releases, exchange feeds sometimes became overwhelmed during news events and were unable to keep up with demand in a fast and accurate way, providing opportunities for those with better technology to profit on better information.

Multi-venue Low Latency Strategies

Now consider the situation where a large truck pulled up to a local farmer's market and immediately bought up all the tomatoes. What would be the smart thing to do in this situation? One idea would be to race over to the next market further up the road, buy up all the tomatoes, raise the price and wait for the truck to arrive.

This is essentially a description of what "multi-venue low latency" strategies aimed to achieve. They attempted to front-run large orders appearing on one market, expecting the orders to appear on other markets as well. If they were faster to the next market they could buy up all the stock at the current ask price and wait for the large order to arrive—at least in theory. This idea was the motivation for the construction of the Spread Networks tunnel described earlier. By laying a straight-line optical fiber cable (which entailed building a tunnel underneath the Allegheny Mountains), Spread Networks was able to shave out a few milliseconds in transit time from the trading center in Aurora, Illinois to New Jersey. The tunnel proved worth the initial $300 million investment because of the willingness of subscribers to pay $10 million each for the resulting data feeds; it soon became mandatory to use this tunnel in order to avoid being front-run by competitors.

High Frequency Strategies and Statistical Arbitrage

Old-fashioned specialists proved no match for high frequency trading and were increasingly being displaced in the markets. Several HFT firms focused on playing the role of the specialist by providing liquidity on exchange venues. For example, many venues gave rebates to firms who played this role, thereby increasing profit opportunities.

The strategies employed by these firms attempted to collect the rebate and the bid-ask spread by anticipating price movements. They used statistics "about the book" (e.g., Level 2 price information) in order to analyze the balance between potential buyers and potential sellers. Simplistically, if there happened to be more buyers than sellers, the price would usually go up and vice versa. However, orders resting on the book could be cancelled quickly, giving opportunities for the book to be manipulated.

These strategies were intended to work the majority of the time. Winners would be mixed with losers, with the net profit being thin. However, with enough repetitions of these plausibly independent events, HFT firms aimed to make attractive returns with high Sharpe ratios by trading very frequently.

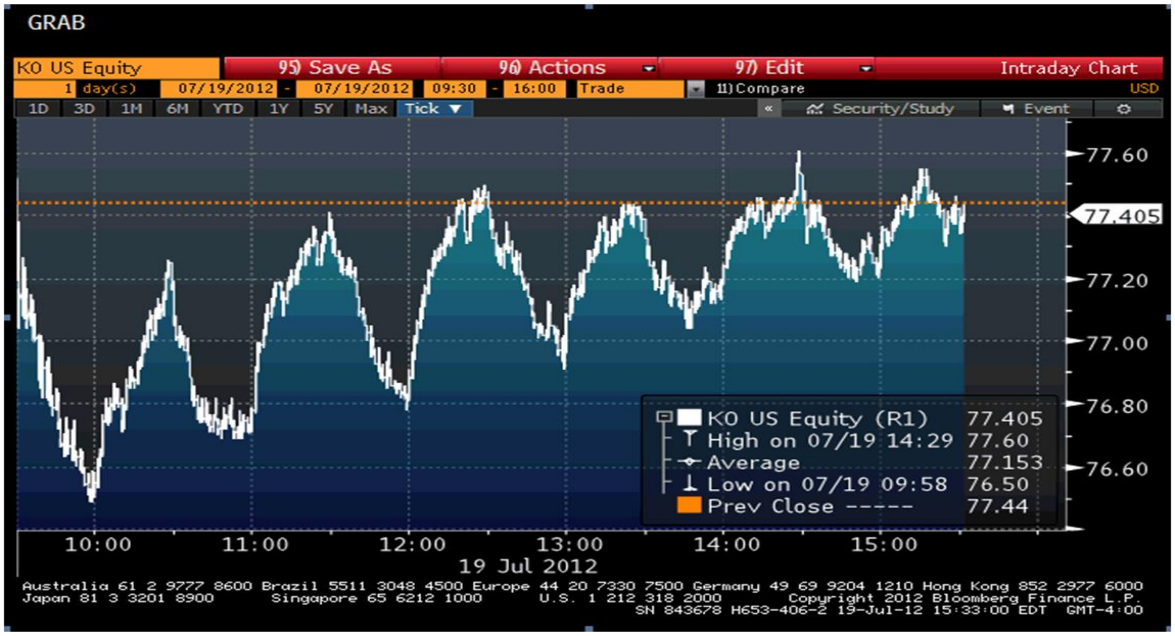
Related strategies used "statistical arbitrage" trades based on certain relationships gleaned from analysis of high frequency data between pairs of stocks or pairs of instruments. The enormous amount of data available for HFT firms to analyze provided ample opportunities to find stable relationships in the data that might be expected to continue in the future.

The Controversy Surrounding High Frequency Trading

Aside from systemic risk issues, discussed below, the biggest complaint about HFT was the ability of fast participants to front-run large orders across venues. These criticisms focused on the common experience of traders seeing a price quote on their terminals, but being unable to execute a trade at or near that quote. Often this was the result of the fact that they placed trades too large to be executed on one market, leaving evidence that allowed faster participants to front-run them on other markets.

How hard was it to front-run large orders? Below in **Figure 3** is an example of price changes for Coca Cola on July 19, 2012. (There were similar price histories for AAPL and a couple of other stocks on that day.)

Figure 3 Coca-Cola stock price graph on July 19, 2012



Source: Daniel Fisher, "Don't Blame High-Frequency Traders For The Mistakes Of Dumb Traders," Forbes, (June 5, 2014): http://www.forbes.com/sites/danielfisher/2014/06/05/dont-blame-high-frequency-traders-for-the-mistakes-of-dumb-traders/, accessed March 2015.

Note the large price jumps that started each hour. The most likely explanation for this was that orders were being placed for a large market participant. This type of footprint encouraged high frequency traders to jump on board when they saw evidence of a "buy cycle." The sell-off after peaks might be due to traders taking profits on the price impact of the initial orders. In the last cycle there was an abortive price increase before the hour, perhaps due to traders anticipating another buy cycle that did not materialize.

Another factor helping HFT firms was that certain “no lock” rules required brokers to execute in markets in order to give their clients the best prices, which meant that brokers routed their orders to several venues, again leaving footprints that allowed their trades to be front-run by HFT traders. (The rules were initially designed to prevent brokers from simply executing trades at venues with fee structures that benefited the brokers but did not have the best prices for their clients).

Finally, critics questioned the social benefit of spending hundreds of millions of dollars in an arms race for faster speeds. To them this was simply a tax on investors contributing little or no value to the underlying economics of investing.

“Unfair” Advantages of HFT Firms

Recognizing the importance of liquidity on their markets, competing exchanges offered a variety of inducements to high frequency traders to be present. In addition to the rebates for providing liquidity, different exchanges had a variety of rules about order types that went beyond the basic limit and market orders. It was claimed that some of these order types were designed to favor HFT firms.

One such example was the so-called “hide-and-light” order types. As noted earlier, orders that crossed the markets could not be placed on an order book. A hide-and-light order avoided this issue by resting “hidden” orders (i.e., orders that were not displayed on the order book) until/unless the order was not crossing the markets at which point it was instantly displayed on the book, being first in line.

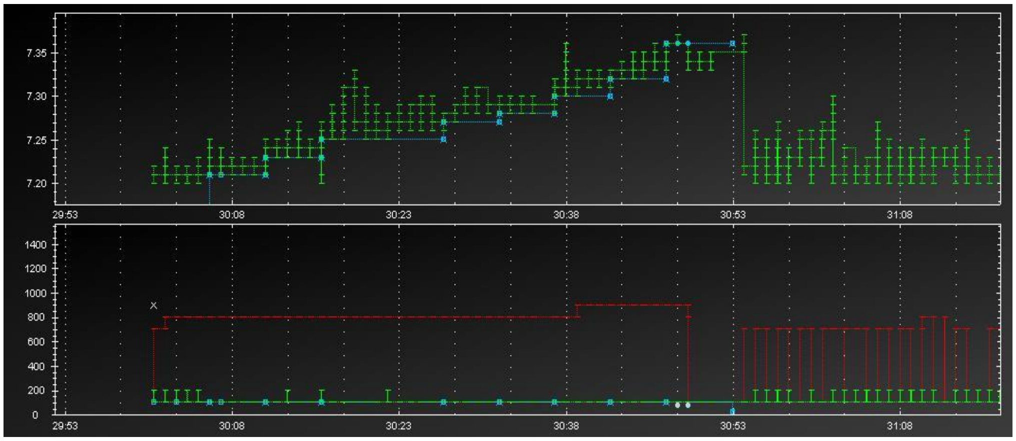
Another advantage that HFT firms had involved exceptions to the so-called “no-cross rule.” Recognizing that some market participants had access to order book data faster than the SIP, the rules allowed qualified participants to cross the NBBO (as displayed on the SIP) provided that they certified that their orders did not cross the markets at the time they were placed. These rules clearly favored HFT firms in that they were the only participants who could take advantage of the relatively slow SIP feed.

Market Manipulation

Critics also charged HFT firms with outright market manipulation, including using tactics such as “spoofing” and “quote dangling.” These involved placing orders with the intention of later cancelling them in order to give the appearance of depth in the order book. An example of spoofing can be seen in **Exhibit 2**.

Quote dangling is illustrated in the following screen-shot in **Figure 4**. In the top screen, the ask part of the order book gave quotes close to the bid, showing some depth. The buyer lifted the best offer believing that she could buy more shares at or near that offer. The orders on the book above that offer were then cancelled, resulting in the best offer increasing. Note that the apparent liquidity was withdrawn at each transaction (blue dots). The liquidity being withdrawn was the result of cancellations from the second or third level of the order book as well as the first level, suggesting that these orders were not intended to be filled. After chasing the bids upward the buyer finally gave up, and the illusory bids were replaced on the order book with the spread returning to the pre-event size.

Figure 4 An Example of "Quote Dangling" for a Given Stock



Source: Marcos Lopez de Prado, "Advances in High Frequency Strategies," PowerPoint Presentation, March 2012, Cornell University, http://www.orie.cornell.edu/engineering2/customcf/iws_events_calendar/files/cfem_20120314_0.pdf, accessed March 2015.

Both quote dangling and spoofing involved the placing of orders with the explicit intent of cancelling them at a later time. Some critics argued that traders actually created inter-venue low latency opportunities by placing large volumes of orders and cancelling them, thereby overwhelming the informational capabilities of the exchanges and of the SIP.

Dark Pools

Responding to HFT firms' use of order book data and the increasingly obvious footprints of large orders, dozens of "dark pools" were subsequently created by broker dealers, electronic market makers and others.[5] These were proprietary venues where prices and liquidity were entirely opaque, as were the details of order matching. They were designed to protect the anonymity and hide the intentions of institutional investors that were buying or selling large volumes. The opacity was designed to allow hidden liquidity to be quietly executed at prices that would not be possible were the demand they represented to be public before execution.

Dark pools were only partially successful in their goals. Some HFT firms quickly learned how to discover the prices in dark pools and the depth of demand by "pinging": placing many small orders and learning from which orders are filled. (For example if placing small buy orders made the price move up, it was likely that there was no large block of stock available for sale). Having deduced the likely internal demand, HFT firms could then still anticipate the orders of dark pool participants.

Moreover, dark pools had their own sets of issues. Did the prices in dark pools track public prices accurately? Was there a risk that there would be tiered pricing with advantages to large clients in dark pools that were not available on the public markets to other investors? A large execution might be achieved at favorable prices for the investor placing it, but what would be the consequences when that order then became public? Because the matching was completely confidential, how confident could an investor be that they were getting the best price? Were brokers more likely to place orders

[5] It was estimated that 40% of equity volume soon became traded on dark pools, according to Reuters.

in the dark pools run by their own company? And if so, did they manage to get the best prices in that dark pool? These difficult questions made dark pools also the subject of controversy.

Responses to Criticism

HFT firms responded to many of these criticisms by framing the issue with a key question: who has benefited and who has been hurt by high frequency trading? According to many market participants (and not just the HFT firms themselves), the surface evidence seemed unambiguous that small investors had benefited. Before the advent of electronic trading, individuals trading their private accounts usually paid over $100 per trade in commissions. Moreover the *minimum* bid-ask spread was 1/8 of a point. Now the minimum bid-ask spread was $.01. Empirical studies confirmed that, even for less liquid stocks, the actual bid-ask spreads were considerably smaller than in the days before high frequency trading.

In addition, HFT firms did seem to play the role of market makers, particularly in less liquid stocks. One academic survey article concluded that "Algorithmic traders are more likely to be at the inside quote when spreads are high than when spreads are low, suggesting that algorithmic traders supply liquidity when it is expensive and demand liquidity when it is cheap." (See Goldstein, Shkilko, Van Ness, and Van Ness (2008)).

More abstractly, high frequency trading plausibly led to better approximations to "perfect markets." Relevant economic information was incorporated much faster, and price information was available to all participants nearly instantly—at least from the point of view of fundamental traders. Price distortions were corrected by arbitrage nearly instantly. Thus, according to its defenders, high frequency trading had democratized the markets, allowing small players with technical skills to participate on an equal footing in the markets.

Responses to Claims of Abuse

According to its defenders, the victims of HFT front running and those who suffered from outsized price impact were simply behind the technological curve. Techniques such as randomizing order placement, "slicing and dicing" (as well as more sophisticated proprietary algorithms) could hide footprints and provide defense against predatory practices.

To the claim that ordinary people were hurt when their pension funds or savings were placed in mutual funds affected by HFT trading, HFT firms argued that these firms were simply behind the times in terms of their order execution techniques. And further, that once these firms learn to play better defense, the problem would be mitigated.

Regarding spoofing and price manipulation, HFT firms argued that any new technology invariably led to new techniques for cheating, and that the rules of the markets would adapt and catch up to prevent these abuses from continuing. Limiting cancellations could represent one such rule; however HFT firms pointed out that there were legitimate investing techniques (such as delta-hedging of options) that involved placing and cancelling large numbers of orders. Moreover, legitimate market-makers, who were setting the inside of the spread and adjusting it frequently in response to market conditions, needed the ability to cancel orders as market conditions changed.

Overall, HFT firms, and many large non-HFT asset managers (such as Dimensional Fund Advisors, AQR, etc.) argued that modern electronic markets had leveled the playing field for the small investor by democratizing information access and reducing costs. Commissions had gone down by over 90 percent and the bid-ask spread had been significantly reduced. Meanwhile large mutual funds were adapting to changes in execution conditions by adopting defensive order execution methods and mitigating price impact. And finally, HFT firms argued that the millions of dollars of profit earned by HFT traders should not be viewed as a "tax" on investors, but rather a simple shift of rents from one group of actors (the specialists and broker/dealers) to another (the tech-savvy HFT traders).

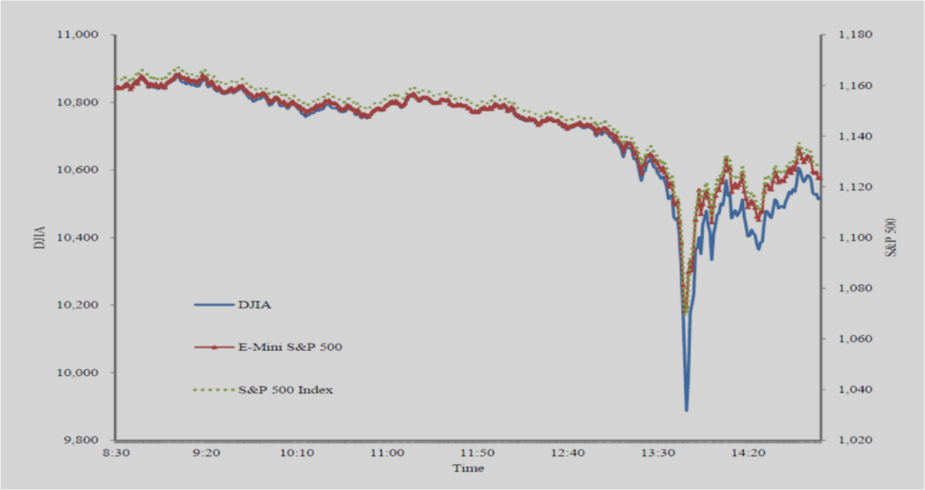
The Risks of High Frequency Trading

Due to the speed at which events could take place, high frequency trading also presented special kinds of risk control issues. These were of concern both to individual firms and their trading strategies, but also to regulators because of the possible systemic effects of HFT "mistakes."

Systemic Effects: The Flash Crash

On May 6, 2010, over the course of 36 minutes, as a result of a large sale of E-mini S&P 500 futures, the DJ industrial average fell over 1000 points and then nearly completely recovered (see **Figure 5** below).

Figure 5 Market price chart for the Flash Crash of May 6, 2010



Source: Kilrenko, Andrei, et al. "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," 2010, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1686004, accessed March 2015.

According to a report originally sponsored by the CFTC:

> "At 2:32 p.m., against [a] backdrop of unusually high volatility and thinning liquidity, a large fundamental trader (a mutual fund complex) initiated a sell program to sell a total of 75,000 E-Mini [S&P 500 futures] contracts (valued at approximately $4.1 billion) as a hedge to an existing equity position. [...] This large fundamental trader chose to execute this sell program via an automated execution algorithm ("Sell Algorithm") that was programmed to feed orders into the June 2010 E-Mini market to

> target an execution rate set to 9% of the trading volume calculated over the previous minute, but without regard to price or time."[6]

The illiquidity resulting from the crash led to a near complete breakdown of pricing. Kirilenko and Lo (2013) report that the stock of Accenture traded as low as a penny a share and that Apple shares soared to $100,000 per share; they compared the period of the crash to "a game of "hot potato" among high frequency traders" as they bought and then sold inventory into a rapidly declining market.

The causes of the crash remained somewhat controversial—the CFTC report blamed the execution structure of the trade as the trigger. However there was agreement that the actions of HFT firms amplified the result, as programmatic trading detected and quickly acted on the order book imbalances.

Defenders of HFT responded that this kind of crash was not unknown even before HFT. An oft-cited example is the 1987 Black Monday crash, where the Dow Jones Industrial Average lost over 22% in a single day. Moreover the Black Monday crash took several months to recover from, as compared to the Flash Crash, where most of the ground was made up during the same day.

Risks for Individual Firms

In July of 2012, Knight Capital was one of the biggest market makers in U.S. equities with a market share of 17.3% on the NYSE and 16.9% on the NASDAQ. On August 1, 2012, Knight was apparently updating its software as part of a modernization effort by the NYSE, and put an untested piece of code into production prematurely. The result was $460 million in trading errors in the first hour of trading.

In the wake of the Knight Capital episode, it became clear that algorithmic trading at the speed of micro-seconds without direct human supervision created risks for HFT firms. The risk of losing one's entire capital base in less than 30 seconds clearly required special risk control measures. But introducing risk monitoring systems that verified trades would slow down the trading process—perhaps eliminating profits in a winner-take-all environment.

Meanwhile, clearing firms provided much of the back office support for HFT firms, but were not in a position to intermediate trade orders, as they might in a conventional trading setting. Because they did not have the infrastructure for HFT, they provided direct access trading to HFT firms that were certified by the exchanges. This limited their ability to monitor positions and reduce leverage and other risks based on realized volatility. They did monitor the positions and results of trading at the end of day, but this was far too slow for effective risk monitoring in the context of high frequency trading.

This left the responsibility for risk monitoring entirely up to HFT firms. Typically HFT firms did not take any overnight exposure, eliminating one source of risk for conventional hedge funds. And to monitor real-time returns, HFT hedge funds usually set up a system distinct from the one executing the trading algorithms in order to record the results of executed trades. Extensive back-testing was done to develop return profiles for each strategy. When actual returns differed significantly from these return profiles—either negatively or positively--trading in a particular

[6] CFTC-SEC Staff Report, Findings Regarding the Market Events of May 6, 2010, quoted from Kirilenko, Kyle, Samadi and Tuzun.

strategy was generally temporarily shut down until humans could intervene and understand what was happening.

Regulatory Responses

Regulatory responses focused primarily on a few key issues: 1) the perception of possible systemic threats to the markets in the wake of the Flash Crash, 2) the regulation of dark pools, and 3) the claims that the markets are somehow "rigged" in favor of HFT firms (as claimed by the Michael Lewis book).

Circuit Breakers

After the Flash Crash most markets implemented automatic circuit breakers that halted trading in the markets after certain thresholds were met (whereby drops of 7%, 13% and 20% triggered increasingly longer trading halts).

The circuit breakers for individual stocks had more complicated triggers. Due to concerns about shutting down trading due to one erroneous or outlying transaction, the individual stock circuit breakers were triggered when a stock failed to return to a pre-specified trading band within 15 seconds.

Dark Pool Regulation

New rules were also put in place to require dark pools to disclose their trading volume on a daily basis with the expectation that this will soon be required at an even higher frequency. In addition, dark pools are now also required to match trades because of a NBBO constructed from the direct feeds rather than the consolidated feeds/SIPs.

Cancellations

The majority of complaints involving single-venue HFT were related to price manipulation on the order book, and specifically due to the practices of spoofing or quote-stuffing. As noted earlier, these strategies entailed placing orders that rested on the order books with the explicit intention of cancelling them before they were executed. One academic paper suggested that over 74% of US listed equities experienced quote stuffing (Egginton et al. (2016)).

In order to discourage this practice, several proposals had been put forth, such as taxing cancellations and putting a minimum time interval that orders could rest on the book before being cancelled.

Various exchanges in turn adopted their own measures in order to reduce market manipulation. The CME implemented rules about what kinds of orders could be made and cancelled on opposite sides of the spread. It also used "messaging ratios" to discourage cancellations by a point system; depending on the type of order, cancellations accrued points, and the accumulation of 100 points without a trade incurred a penalty onto that trader.

In order to facilitate legitimate market-making activities, the CME designated certain firms as official "market makers." These firms were allowed to engage freely in order placement and cancellation subject to the requirement of continuous quoting on both sides of the spread. Critics, however, pointed out that many of these market makers' quotes were not on the inside of the spread (i.e. not giving the best bid or offer) and hence their role in market making was somewhat suspect to begin with. Nonetheless they were permitted to place orders at their own discretion.

Auction Timing

HFT traders that practice inter-venue front-running take advantage of continuous trading in order to arrive ahead of entities trading large blocks. Several mechanisms were proposed to counter this activity. For example, one simple solution was to hold auctions at pre-determined intervals, such as exactly on the second. This proposal of course suffered from the defect that it merely changed the game from "who is fastest" to "who is able to arrive most closely to the exact second." Thus here the low-latency would not be measured in absolute speed, but rather in speed relative to one-second intervals. Critics of this proposal pointed out that this involved the same technology, and the same corresponding faults.

An alternative to this proposal involved holding the auctions (and windows for placing orders) at randomized times, to occur such that there was a pause of no longer than one second. Advocates for this proposal claimed that the randomness would level the playing field for slower participants.

Code Verification/Licensing

Various markets, particularly those in the EU, began requiring HFT firms to verify their code according to certain protocols. These measures, which amounted to licensing HFT traders, attempted to ensure that HFT firms' code did not send erroneous messages, and were robust under various market conditions. They essentially involved codifying minimum risk management standards.

Finally, individual exchanges began implementing various measures designed to make their trading platforms "fairer." One such technique entailed artificially introducing delays in order to make sure that orders originated and arrived at exterior exchanges simultaneously. Other venues, such as the PARFX, introduced random pauses in trading in an attempt to confuse front-running efforts. And still others began to fill best-bid and best-offer quotes randomly (rather than with time priority) or used proportional fill rules—thus somewhat negating the relative benefits of front-running.

Domeyard

In the spring of 2013, Luca Lin and Christina Qi and Jonathan Wang came together and began putting together a company that they felt could be an innovator in the cutthroat world of high frequency trading. Their belief stemmed from the fact that they had assembled a team that spanned all of the key requisite areas for success in this domain. Christina had known Luca and Jonathan through her classes and network during her undergraduate years at MIT, and brought the two together.

Luca was a physicist by training and had spent several years in research labs after college before moving to Asia to begin trading on his own. According to Luca, "I soon realized that many of the concepts I had been working on in applied physics labs could be translated into profitable trading strategies in the market; conceptually the problems were very similar."

Meanwhile, Christina came from a trading background and had interned at Goldman Sachs Asset Management and UBS during her undergraduate days at MIT. After working on these large desks, Christina came to the following realization: "When I saw how consistently Luca was making money trading on his own account, I realized that small companies really could be big players in the market, which was something I had not previously believed was possible."

Rounding out the founding team was Jonathan Wang, who was an experienced computer scientist and previously a software developer at Apple, and had a reputation for being one of the best programmers in his graduating cohort at MIT. Together the three believed they had all the skills necessary for success.

Through contacts in classes, they soon assembled an impressive team of advisors and mentors, ranging from former hedge fund managers to the Nobel laureate Robert Merton. This allowed them to raise a substantial amount of operating capital for their fund management company, and also a significant amount of trading capital, and have a commitment for $5m of early trading money for their hedge fund.

They decided to set up their offices in Cambridge. "We looked around and found a wonderful spot for a small company—both for its proximity to the MIT campus and for its view." Investors were uniformly impressed when they came in. "When we moved in we built our own desks and assembled our own servers and network to our specifications. Some of the major early challenges were physical: we had to arrange our own redundant backup power; our servers produce so much heat that we had to crawl through the ceiling to redo the ductwork in our office and pipe the heat to a spare room we weren't using. The constant roar of the servers was annoying in the conference room, so we installed sound insulation and a special door. These were the fun things. Now we have too many servers so we've mostly moved them to data centers instead."

Bigger challenges involved writing data feed handlers and execution gateways against the exchange APIs so that they could collect data and place orders on their co-located servers. One particularly thorny problem was that each individual exchange had their own protocols and their own APIs, and these systems were constantly changing. As Luca noted, "If you can build a system that is optimized for each individual exchange, you can have a big latency advantage."

Initial Tasks

The mechanics of setting up a hedge fund in general, and specifically a hedge fund in the HFT space, were hardly trivial. There were the legalities associated with forming the various fund and management company entities, as well as the tax consequences associated with setting up the management company/general partner (GP) as either a limited liability company (LLC) or as a limited partnership (LP). Plus there were was the task of finding appropriate prime broker and clearing firm relationships. Finally, there was the difficult decision of how to structure fees for investors into both the management company and the fund itself.

Early Successes

An advantage that Domeyard had was that they had built their own software internally, and had precisely targeted it to their own hardware. This allowed them to quickly detect errors, diagnose them, and fix them in real-time. By contrast, many other HFT traders used off-the-shelf commercial software for some portion of their platform, leaving them at the mercy of vendors.

With Jonathan and Luca's state-of-the-art software and hardware guidance, Domeyard's developers put together some impressive technical achievements. For example, they developed a messaging software library that broke known world records for low latency interprocess communication. This allowed them to utilize distributed computing resources more effectively and trade approximately ten times as many instruments on the same server footprint as other HFT firms that they were familiar with. They developed a "full stack" of proprietary software that put them among the leading firms: from their own feed handlers that interfaced with the proprietary exchange feeds, to their own petabyte-scale storage system, to their own drivers for specialized network cards.

Fundraising Challenges

Raising Money for the Management Company

As Christina pointed out:

> "Tech investors typically don't 'get' quant hedge funds or a technological trading firm. Investors in tech companies expect some kind of product with growth, be it revenue, adoptions or some other metric. A quant hedge fund isn't like that. Typically there is a fixed amount of investment in the LP, assets under management (AUM) is limited by the number of investors and the amount of capital that is brought in. It is a totally different game. On the other hand, when pitching to fund investors we have no trouble at all standing out—we're the only company here with a product like this."

One challenge was that when angel investors and venture capital firms tried to evaluate Domeyard's product, they typically wanted to see at least a prototype. To them this meant, at the very least, having back-tested results. "They push hard for simulated trading results on a live feed, so that we can demonstrate that our system can keep up with the necessary speeds. What they really want is for us to engage in live trading." On the flip side, of course, once Domeyard was engaged in live trading, they were going to have less need for investments into their management company because high frequency trading strategies had the potential to be self-financing even on relatively small amounts of trading capital.

One helpful counter-argument to these concerns was the $5 million that an early investor had put up to seed Domeyard's early trading. The fact that this well-known hedge fund manager had the confidence to invest in Domeyard proved valuable in fundraising. Also attractive was the perk that early fund investors (also referred to as "limited partners" or "LPs") got: their capital would be traded without management fees or "carry" (i.e., performance fees).

Raising Investment Capital

Many conventional hedge funds invested with the intention of holding for a significant period of time and making bets that depended on various "catalysts" to bring the investment to its fundamental value. These funds tended to have liquidity issues—if investors demanded their funds back before the catalyst occurred, dramatic losses could ensue. As a result, many hedge funds required long lock-up periods from their investors.

In contrast, Domeyard planned to be "flat" at the end of each day (i.e., they would not hold any overnight exposure). Barring crisis situations, their investments would be completely liquid. For this reason Domeyard decided not to require lock-ups. Domeyard's structure would offer monthly visibility and redemption opportunities, and only 7 days notice for redemption. Luca, Christina and Jonathan were completely aware that money flowed towards short-term returns, but felt that their potential investors would be attracted to the liquidity that their fund offered.

Other Early Decisions

In addition to the afore-mentioned challenges associated with starting up the hedge fund, the principals at Domeyard continued to mull over a series of other important practical decisions. For example, what should be the right benchmark for the fund, and how would they articulate this to clients? And which clients would be ideal for a product like the one Domeyard was offering? More fundamentally, how could they be confident in their true expected return and target Sharpe ratio without a live track record?

Business Strategy

Luca summarized Domeyard's early strategy as follows:

> "Our initial strategy is to start in the CME Group markets (CME, CBOT, NYMEX, COMEX), move to equity trading in Singapore (on the SGX), and then expand horizontally from there. We don't plan to trade North American equities until 2016Q2 at the earliest.
>
> Many of the strategies that we are using have limits of scale; our initial approach to surmounting this is to expand to a large number of markets. After gaining experience in many markets and a critical mass we believe we will have the expertise to grow the capacity of our strategies, expanding into medium-frequency trading if need be.
>
> We favor starting at the CME because the interface is among the most mature and transparent. They have a high level of familiarity with and openness to high-frequency, electronic market makers and are straightforward and helpful to new participants. An important factor in our decision is the speed at which we can go-live."

The principals at Domeyard felt that starting with the CME, which was a mature, domestic market, would enable Domeyard to quickly connect to live data feeds for simulated trading and have live trading results faster than they could in Singapore. In addition, it was more expensive and difficult to find a source of historical data feed, and to find a colocation vendor that met their latency requirements, in a less mature market such as Singapore.

Of course the downside of this decision was that the CME was a very sophisticated and efficient market. Domeyard knew they could reach a rebate agreement with the Singaporean market, which meant they would be receiving revenue early, and the market participants were far less advanced. However, the interface with their market was less mature and documentation was less accessible.

Trading Costs

The costs and fees associated with trading were absolutely critical to understand and minimize for an HFT firm like Domeyard. As Luca noted, "For us to cover our fixed costs we need to trade on the CME we probably have to do around 10,000 trades per day and capture anywhere between 0.1 and 0.9 basis points of revenue from every dollar of trading volume."

One effect of this large number of trades with razor thin margins was a large sensitivity to trading costs. This emphasis on trading costs led to an important, early decision: whether or not to buy an equity membership on the CME. Buying a membership on the CME was valuable because it enabled one to lower trading costs; but the tradeoff of course was that this purchase required a significant upfront investment.

In the end, the decision came down to spending roughly $4 million upfront to become a so-called CME "106J" member versus a $3,800 per month charge to become a CME "106H" member. The 106J membership would lower transactions costs even more than a 106H membership for a firm like Domeyard, since in addition to lowering clearing and platform fees, it also enabled members to receive rebates if a firm's trading volume reached a certain level (which would likely be the case for Domeyard). On the flip side, in addition to the large upfront investment, there were some additional regulatory and bureaucratic hurdles associated with 106J membership, for example the requirement that the membership be registered in the name of the actual fund and paid for by the fund instead of the management company.

Barriers to Entry

Domeyard felt that their ability to overcome some substantial barriers to entry in the HFT space left them well-positioned to succeed, and maintain that success going forward.

Chief among these barriers was the need for exceptional technical proficiency. Christina noted:

> "A perfect example is our co-founder Jonathan Wang, whose coding talent is second to none. Plus we are wonderfully positioned—our network includes the best recent graduates in computer science and engineering from MIT and Harvard; moreover there is a lot of interest from local tech entrepreneurs in participating in our company as investors."

The costs and complexity of the necessary equipment (both software and hardware) was also substantial. HFT firms needed sophisticated clusters of servers and switches between them, and Domeyard had the internal expertise necessary to assemble these in-house. Potential breakdowns due to abstract forces such as cosmic rays presented a formidable, but often ignored, challenge: "Such errors scale proportionally with the amount of data that you handle or the surface area of your hardware. Most traders overlook this because they can squeeze years of their data onto a single hard disk drive and might not be affected for decades. We expected to encounter hundreds to thousands of single errors each year based on the volume of data that we would handle. And in the HFT domain, there isn't much time for the correcting algorithms to take place; even miniscule errors can be very expensive. Conventional solutions to maintain data integrity are slow and expensive. We designed our own solution," noted Luca.

Colocation

Another important barrier was the need to "co-locate" one's operations as close as possible to the trading venue. As Luca pointed out:

> "To be competitive one must co-locate your trading engine at the exchange you are trading on. This process is easiest to understand at the CME. The speed of the matching process inside the CME isn't very relevant to single venue strategies—it's the same for all of our competitors and us.
>
> What we can control is what goes on inside Domeyard. After the information leaves CME and connects to us, it goes through some very high speed network switches that replicate the data streams and send those downstream to two places: execution servers

> and data servers. Execution servers run our trading algorithms, while data servers monitor everything. Combining the two functions would make the algorithms too slow. The switches are quite fast and most HFT firms already use the same switches. What differentiates one firm from another is the internal "wire-to-wire" processing time of their execution servers. Because of the caliber of our software engineers, we are only interested in trading on markets where we think we've managed to make it to the top twenty in wire-to-wire latency."

(**Exhibit 3** illustrates Domeyard's connection setup.)

The Decision

Awaiting potential angel investors in their conference room, with the faint whiff of ozone from the heat of the server in the air, Luca contemplated what to say to them. Their investors already were advising them to change the phrase "high frequency trading" to "electronic market making" in their pitch books and other literature. This seemed like a good idea and quite easy to do, although, ironically Michael Lewis' book had also alerted some investors to a profit-making opportunity that they hadn't previously known about.

Luca, Christina and Jonathan had already put out a white paper about their positions on the HFT controversy. This helped make clear that Domeyard was not planning on using any of the more controversial strategies; in fact they were making sure that their investors were clearly aware that some of the proposals to "level the playing field" would in fact benefit newcomers like Domeyard.

But what about membership in the CME? It was a large expense relative to the money they had already raised and represented a rather static investment. On the other hand perhaps their investors would be comfortable with putting their money into this; after all, it represented some amount of risk-reduction, being a concrete asset that was unlikely to lose its value in the event of a liquidation of Domeyard. But of course it only made economic sense if they could get trading volumes to the level necessary that the reduced costs paid off. Could they do this in the very short term?

Exhibit 1 Order Books

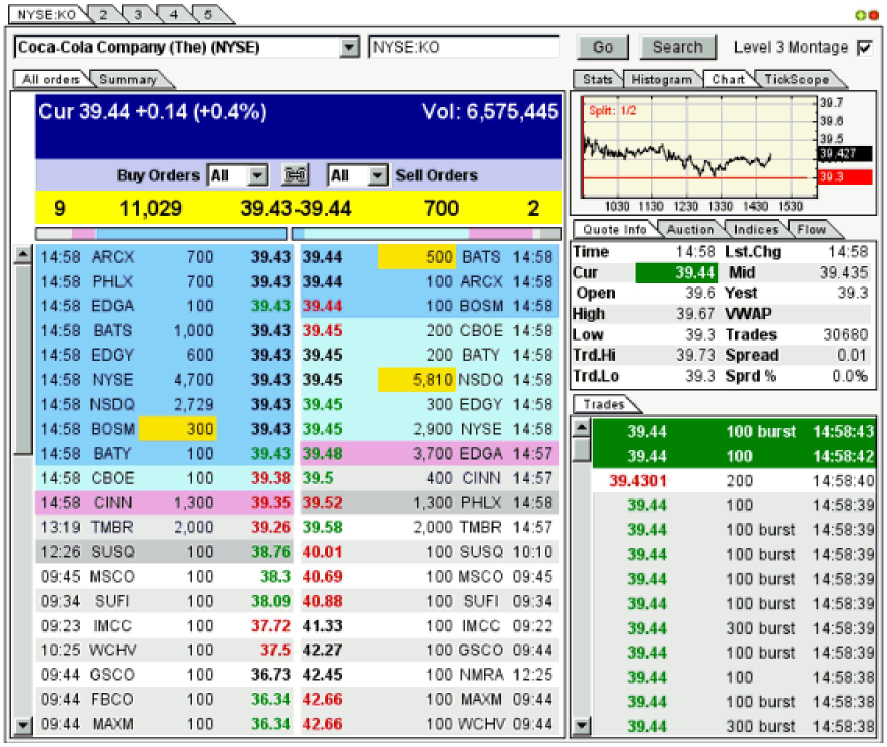
The order book shows information about limit orders on both sides of the spread. Different exchanges display different amounts of information on their order book. For example the CME E-mini's just show the number of orders and the total "depth" at each level.

	Bid			Ask	
Number'of' orders	Total'Depth	Price	Price	Total'Depth	Number'of' Orders
5	25	49.97	50.01	20	3
7	10	49.96	50.02	30	2
2	20	49.95	50.03	32	10
9	5	49.94	50.04	10	1

Source: Casewriters.

In this example the bid-ask spread is 50.01-49.97=0.04. The depth at the inside of the spread was 25 units on the bid side and 20 units on the ask side.

The next picture gives the **Level 2** data for the equity Coca-Cola from the website Investors Hub.



Coca-Cola Company (The) (NYSE) NYSE:KO

Cur 39.44 +0.14 (+0.4%) Vol: 6,575,445

Buy Orders All All Sell Orders

9	11,029	39.43-39.44	700	2

14:58	ARCX	700	39.43	39.44	500	BATS	14:58
14:58	PHLX	700	39.43	39.44	100	ARCX	14:58
14:58	EDGA	100	39.43	39.44	100	BOSM	14:58
14:58	BATS	1,000	39.43	39.45	200	CBOE	14:58
14:58	EDGY	600	39.43	39.45	200	BATY	14:58
14:58	NYSE	4,700	39.43	39.45	5,810	NSDQ	14:58
14:58	NSDQ	2,729	39.43	39.45	300	EDGY	14:58
14:58	BOSM	300	39.43	39.45	2,900	NYSE	14:58
14:58	BATY	100	39.43	39.48	3,700	EDGA	14:57
14:58	CBOE	100	39.38	39.5	400	CINN	14:57
14:58	CINN	1,300	39.35	39.52	1,300	PHLX	14:58
13:19	TMBR	2,000	39.26	39.58	2,000	TMBR	14:57
12:26	SUSQ	100	38.76	40.01	100	SUSQ	10:10
09:45	MSCO	100	38.3	40.69	100	MSCO	09:45
09:34	SUFI	100	38.09	40.88	100	SUFI	09:34
09:23	IMCC	100	37.72	41.33	100	IMCC	09:22
10:25	WCHV	100	37.5	42.27	100	GSCO	09:44
09:44	GSCO	100	36.73	42.45	100	NMRA	12:25
09:44	FBCO	100	36.34	42.66	100	MAXM	09:44
09:44	MAXM	100	36.34	42.66	100	WCHV	09:44

Time	14:58	Lst.Chg	14:58
Cur	39.44	Mid	39.435
Open	39.6	Yest	39.3
High	39.67	VWAP	
Low	39.3	Trades	30680
Trd.Hi	39.73	Spread	0.01
Trd.Lo	39.3	Sprd %	0.0%

39.44	100 burst	14:58:43
39.44	100	14:58:42
39.4301	200	14:58:40
39.44	100	14:58:39
39.44	100 burst	14:58:39
39.44	100 burst	14:58:39
39.44	100 burst	14:58:39
39.44	100 burst	14:58:39
39.44	300 burst	14:58:39
39.44	100 burst	14:58:39
39.44	100	14:58:38
39.44	100 burst	14:58:38
39.44	300 burst	14:58:38

Source: InvestorsHub, http://investorshub.advfn.com/, accessed March 2015.

The inside of the book in this example is: buy at 39.43 (11,029 orders) and sell (700 orders) at $39.44. The left side contains bids at $39.43 (11029 orders), $39.38 (100 orders), 39.35 (1300 orders) and so on. The leftmost column gives the times that orders were placed (the bid at $36.34 was placed at 09:44 and never filled). The second column gives the identity of the market maker or exchange holding the bid.

Exhibit 2 Spoofing

Here is an example of "spoofing" that was penalized by the SEC.[7] Here is what the inside spread of the order book looked like initially (11:08:55.151 am), before the trader took any action:

Prior to Trader

101.27	101.28	101.29	101.3	101.31	101.32	101.33	101.34 ...	101.37
Best bid	🟩	🟩	🟩	🟩	🟩	🟩	🟩	best offer

The best bid is a 101.27 and the best offer is at 101.37. One millisecond later the trader placed an offer to **sell** 1000 GWW shares at $101.34 per share, .03 less than the previous best offer.

Trader's first move

101.27	101.28	101.29	101.3	101.31	101.32	101.33	101.34 ...	101.37
Best bid	🟩	🟩	🟩	🟩	🟩	🟩	offer 1000	XX

Over the next .159 seconds the trader placed 11 orders offering to **buy** 2600 GWW shares at prices ranging from 101.29 to 101.33. The order book now looked like:

Trader's second move

101.27	101.28	101.29	101.3	101.31	101.32	101.33	101.34 ...	101.37
Best bid		trader bids 2600 shares in this region					offer 1000	XX

Note that the price momentum seems to be upwards as the spread has narrowed from below to just a penny.

A second party, seeing this bought all 1000 shares the original trader had on offer at $101.34. This execution was completed ten milliseconds after the initial trader filled the spread.

Market reaction

101.27	101.28	101.29	101.3	101.31	101.32	101.33	101.34 ...	101.37
Best bid		trader bids 2600 shares in this region					offer 1000-lifted	XX

The trader then cancelled all of his bids between $101.27 and $101.33. The total time these orders rested on the book was less than a second.

Trader's third move

101.27	101.28	101.29	101.3	101.31	101.32	101.33	101.34 ...	101.37
Best bid		BIDS CANCELLED						XX

After the trader cancelled his bids the book returned to its initial state.

Final/State

101.27	101.28	101.29	101.3	101.31	101.32	101.33	101.34 ...	101.37
Best%bid	🟩	🟩	🟩	🟩	🟩	🟩	🟩	best%offer

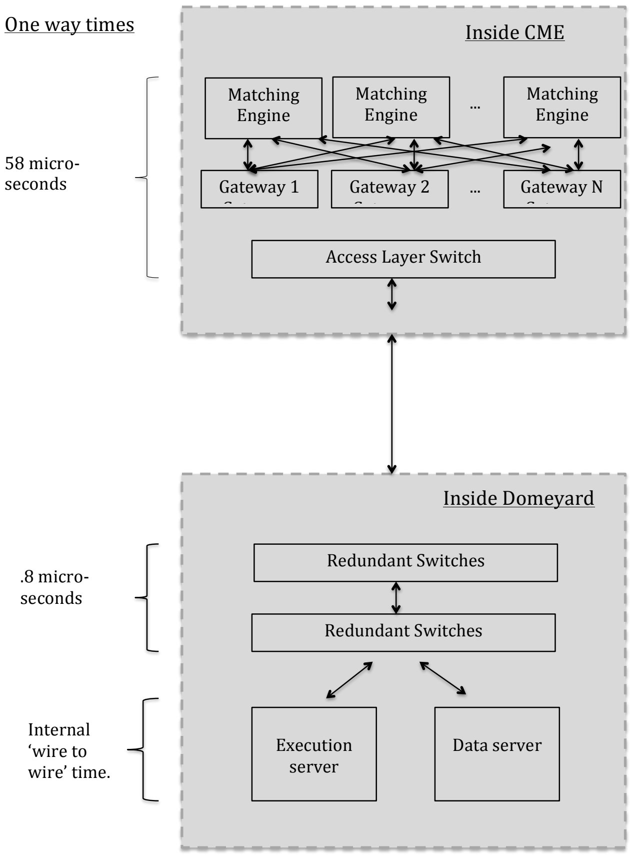
Source: Casewriters.

The interpretation of this situation is that the offending trader placed deceptive buy bids in order to lure a victim into thinking that there was price competition and inducing the victim to purchase the stock at the inflated offer ($101.34). The bids were placed with the intention of cancelling them before they could be executed.

[7] SEC. Administrative Proceeding, File No. 3-15064. September 25, 2012.

Exhibit 3 High Frequency Trading Connection Setup

Below is a conceptual diagram of how a high frequency trader connects to the CME. The biggest distinction is between what happens inside the CME (which is uniform for all participants) and what happens inside the HFT box at the colocation center. Inside the CME, there is a matching engine consisting of multiple servers that keep track of the state of the book and match buyers and sellers that agree on a price. Information comes into the exchange and is routed by a switch to various gateways which in turn choose the least busy matching engine and route the order to that matching engine. The whole process is designed to take very close to 58 micro-seconds. The HFT traders interact with the CME via fiber optic cables. Latency is determined by several variables, including the lengths of these cables. Colocating at a trading center reduced this latency and in venues like the CME create a level playing field. What happens next varies from HFT firm to HFT firm. In Domeyard, there are a couple of layers of redundant switches which route the incoming and outgoing information to two servers, an execution server and a data server. The execution server handles the algorithmic trading protocols. The data server records the results of transactions and handles any on the fly diagnostics of the computational system. The layers of switches come from off-the-shelf hardware and add about .8 microseconds. The latency of the execution server and the data server (the wire-to-wire time) depend on the algorithms and implementations specific to an HFT trading company. In the set up in the diagram the fixed latency is 64 micro-seconds. To this one adds the wire-to-wire time.



Source: Casewriters.

References

Egginton, JF, Van Ness, BF , Van Ness, RA. (2016). Quote stuffing. *Financial Management* 45 (3), 583-608.

Goldstein, MA, Shkilko, AV , Van Ness, BF, Van Ness, BF. (2008). Competition in the market for NASDAQ securities. *Journal of Financial Markets* 11 (2), 113-143.

Kirilenko, Andrei and Andrew W. Lo. (2013). Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents. *Journal of Economic Perspectives* 27 (2), 51-72.

Madhavan, Ananth & Sofianos, George. (1998). An Empirical Analysis of NYSE Specialist Trading. *Journal of Financial Economics*. 48 189-210.

Sofianos, G., Werner, I. (1997.) The trades of NYSE floor brokers. *Journal of Financial Markets* 3, 139-176.